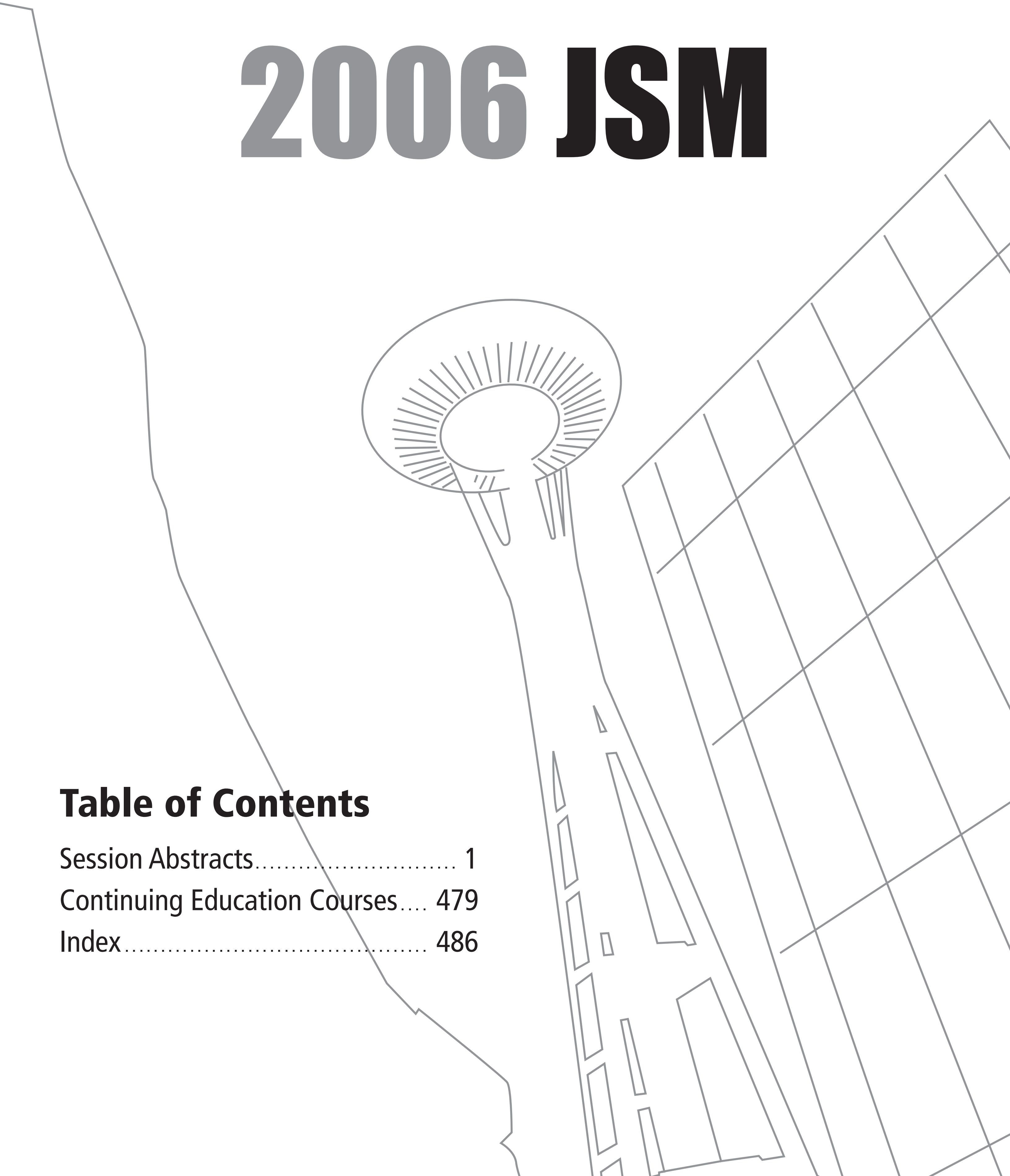
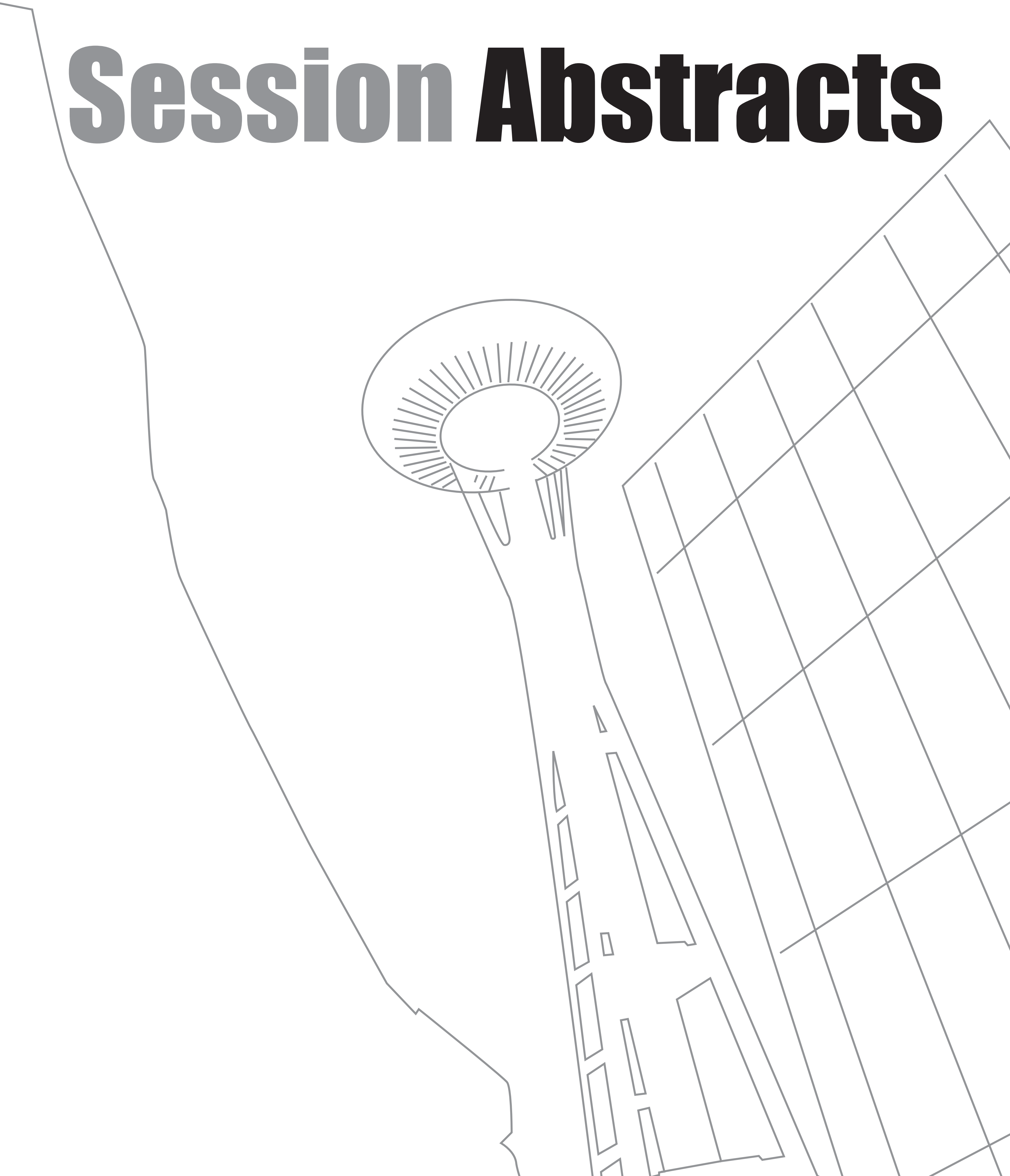
# 2006 JSM



## Table of Contents

# Session Abstracts



# 1 Introductory Overview Lectures: Genetic Association Studies

The ASA, ENAR, IMS, SSC, WNAR

**Sunday, August 6, 2:00 pm–3:50 pm**

## Association Mapping of Human Disease Genes

❖ Bruce S. Weir, University of Washington, Department of Biostatistics, 1705 NE Pacific Street, Box 357232, Seattle, WA 98195-7232, *bsweir@u.washington.edu*

***Key Words:*** genetic markers, case-control test, hapmap project, human disease genes, population structure, multiple testing

Understanding the genetic basis of human diseases and designing therapies or eventual cures rests on knowing the genomic locations of disease susceptibility genes. This task has been greatly simplified recently by completion of Phase I of the human HapMap project: More than 1 million addresses (markers) in the human genome have been identified where there is variation among individuals, and population frequencies of the variants at these locations have been reported. The simplest disease gene mapping procedure is to compare these frequencies for samples of people with or without a disease: The largest differences, or disease-marker associations, are thought to indicate the closest markers. There are several difficulties with this approach, ranging from multiple testing issues to spurious associations caused by population structure. Use of family data can overcome some of these difficulties.

## Statistical Analysis of Haplotype-Disease Association

❖ Danyu Lin, The University of North Carolina at Chapel Hill, 3101E McGavran-Greenberg Building, CB#7420, Department of Biostatistics, Chapel Hill, NC 27599-7420, *lin@bios.unc.edu*

***Key Words:*** association studies, case-control studies, EM algorithm, maximum likelihood, missing data, SNPS

A haplotype is a specific sequence of nucleotides on a single chromosome. The associations between haplotypes and disease phenotypes provide critical information about the genetic basis of complex human diseases. Only the combination of the two haplotypes is directly observable on an individual. In this lecture, we describe various statistical methods for inferring haplotype-phenotype associations. We consider all commonly used study designs, including cross-sectional, case-control and cohort studies, as well as family-based studies. The phenotype can be a disease indicator, a quantitative trait or a potentially censored time to disease variable. Both candidate-gene studies and genomewide scans will be discussed. Illustrations with real studies will be provided. There will also be software demonstration.

# 2 Seasonal Time Series

Business and Economics Statistics Section

**Sunday, August 6, 2:00 pm–3:50 pm**

## Comparing MSEs for Finite X-11 and Model-Based Seasonal Adjustment Filters

❖ William R. Bell, U.S. Census Bureau, Room 3000-4, Stop 91, 4700 Silver Hill Road, Washington, DC 20233-9100, *William.R.Bell@census.gov*; Yea-Jane Chu, SPSS Inc.; George C. Tiao, The University of Chicago

***Key Words:*** signal extraction, canonical decomposition, airline model

Depoutot and Planas (1998) compared weights from X-11 and model-based seasonal adjustment filters, the latter obtained from the approach of Hillmer and Tiao (1982). Chu, Tiao, and Bell (2005) compared the mean-squared errors (MSEs) for infinite symmetric and concurrent model-based filters with those for X-11 filters when estimating nonseasonal components of series following various airline models. Results from these papers show which X-11 filters provide the best approximations to given model-based filters. The results of Chu, Tiao, and Bell also show the MSE from the best X-11 filter may not be much higher than that from the optimal model-based filter. We will review these results and extend them by comparing MSEs from finite X-11 and model-based filters. Accuracy losses from finite X-11 filters should be less than in the infinite sample cases studied previously.

## A Spectral Approach for Locally Assessing Model Misspecification

Tucker S. McElroy, U.S. Census Bureau; ❖ Scott Holan, University of Missouri-Columbia, Department of Statistics, 146 Middlebush Hall, Columbia, MO 65211-6100, *holans@missouri.edu*

***Key Words:*** acceleration, kernel, seasonal adjustment, spectral density

Peaks in the spectrum of a stationary process are indicative of periodic phenomenon, such as seasonality or business cycles. To this end, an important aspect of developing parametric models for periodic processes is proper characterization of spectral peaks. By using an aggregate measure of acceleration, this work proposes to test whether a hypothesized model is supported locally by the data. This technique is useful for detecting model misestimation. More specifically, it can be used to determine the appropriateness of a hypothesized seasonal adjustment model. Finally, the diagnostic is investigated through simulation and application to several seasonal series.

## Nonlinear Seasonal Adjustment in Unobserved Components Models

❖ Siem Jan Koopman, Vrije Universiteit Amsterdam, De Boelelaan 1105, Econometrics Department, Amsterdam, 1081 HV The Netherlands, *s.j.koopman@feweb.vu.nl*; Kai Ming Lee, Vrije Universiteit Amsterdam

***Key Words:*** unobserved components, nonlinear state space, Kalman filter

In unobserved component models, trends, seasonal terms, cycles, and other components are modeled as stochastic processes and can be estimated using the Kalman filter, provided the model is linear. In many economic series, the components combine multiplicatively so the Kalman filter applies with a trivial logarithmic transformation, but more complicated setups. For instance, a model with both additive and multiplicative components, are problematic. In this paper, we will estimate nonlinear UC models with some classical nonlinear state estimation methods, like extended Kalman filtering. Specifically, we will analyze the performance in a model which combines a multiplicative trend and seasonal component with additive noise. As an empirical illustration, we fit a number of U.S. macro-economic time series.

# 3 Recent Advances in the Design and Analysis of Clinical Trials ● ✪

WNAR, Biometrics Section, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

## A Bayesian Seamless Design

❖ Lurdes Y. T. Inoue, University of Washington, F600 Health Sciences Building, Box 357232, Department of Biostatistics, Seattle, WA 98195-7232, *linoue@u.washington.edu*

***Key Words:*** Bayesian sequential design, clinical trials, mixture models, seamless designs

In this talk, we present a seamless phase II/III Bayesian design for comparative clinical trials. The design is based on both survival time and discrete early events related to survival through a mixture model. We discuss simulation studies in the context of a nonsmall cell lung cancer trial and compare operating characteristics of our proposed design to conventional group-sequential phase III designs.

## A Geometric Approach to Comparing Treatments for Rapidly Fatal Diseases

❖ Peter F. Thall, M. D. Anderson Cancer Center, Department of Biostatistics and Applied Math, Box 447, 1515 Holcombe Boulevard, Houston, TX 77030, *rex@mdanderson.org*; Leiko H. Wooten, M. D. Anderson Cancer Center; Elizabeth J. Shpall, M. D. Anderson Cancer Center

***Key Words:*** adaptive design, Bayesian design, clinical trials, competing risks, cord blood transplantation, computer simulation

In therapy of rapidly fatal diseases where the risk of death decreases with response to treatment, it is desirable to achieve a response as quickly as possible. We propose a Bayesian method for comparing treatments in this setting using a competing risks model for response and death. Treatment effect is defined as a two-dimensional parameter: the probability of response and the mean time to response. Several target parameter pairs are elicited from the physician so all pairs embody equally desirable improvements in treatment efficacy compared to a fixed standard. A curve fit to the elicited pairs is used to determine a two-dimensional parameter set where a new treatment is superior to the standard. Posterior probabilities of this set are used to construct rules for treatment comparison and safety monitoring. The method will be illustrated by a cord blood transplantation trial.

## Advances in Simple Phase I Trials: Three Examples

❖ Rick Chappell, University of Wisconsin-Madison, Department of Biostatistics, 600 Highland Ave., K6/430, Madison, WI 53792, *chappell@stat.wisc.edu*

***Key Words:*** toxicities, phase I, clinical trials, dose escalation

Phase I studies in oncology involve the first clinical application of a novel treatment and are designed to assess safety by escalating the dose until severe toxicities are seen. This talk will point out several situations in which information is wasted in current phase I designs and give simple solutions. Bayesian and other model-based methods provide a rich source of extensions to traditional designs and can incorporate a variety of sources of information. However, they are complex and can appear to be unintuitive "black box" methods to clinical investigators who may thus be reluctant to use them. As simple, if not uniformly superior alternatives, I will present generalizations of the traditional algorithmic designs to the following situations: incomplete followup of long-term toxicities, ordered risk group cohorts, and information about the severity of toxicities.

## Decision Analysis and Clinical Trial Design

❖ Donald Berry, The University of Texas, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *dberry@mdanderson.org*

***Key Words:*** Bayesian design, decision analysis, clinical trials

I will describe recent Bayesian innovations in the design of clinical trials. The goals are more efficient clinical trials and clinical development programs and treating patients more effectively---both in and outside of trials. Many innovative designs have been used at my home institution in national oncology studies and industry-sponsored trials. I will provide background on Bayesian designs for clinical trials and give case studies, including the possibility of early stopping and variations on themes such as seamless phases II and III trials with sequential sampling and using early endpoints. The cost savings of such an approach is usually substantial, as are the benefits to patients. I also will describe generally the role of statistical decision analysis in these designs and in the pharmaceutical and device industry.

# 4 Rapid Production of Small-Area Estimates Using the Behavioral Risk Factor Surveillance System

Section on Survey Research Methods, Section on Health Policy Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

## Rapid Response Health Surveillance and the Utility of Small-Area Estimates: Responding to the 2004--05 Influenza Vaccine Shortage

❖ Michael W. Link, Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, MS:K-66, Atlanta, GA 30341-3717, *MLink@cdc.gov*; Ali H. Mokdad, Centers for Disease Control and Prevention

***Key Words:*** survey, small area estimation, health surveillance, GIS, influenza

Rapid development of valid and reliable small-area estimates of health conditions is critical for mounting an effective response to emerging public health emergencies. During the 2004--05 influenza season, U.S. public health officials faced such a situation when the influenza vaccine supply was unexpectedly cut nearly in half due to production problems. The action prompted the issuance of interim recommendations for prioritization of influenza vaccine, asking healthy persons to forego vaccination. Monitoring the effectiveness of these recommendations at the state and local levels was essential to developing effective public health messages and informing vaccine redistribution efforts as the season unfolded. We delineate how the Behavioral Risk Factor Surveillance System was used for rapid health surveillance and highlight the importance of small-area estimates in addressing emerging heal

### Development of Methodology for Production of Rapidly Available, County-Level, Small-Area Estimates To Monitor the Course of Influenza Vaccine Coverage

❖ Haomiao Jia, Mercer University School of Medicine, 1550 College Street, Department of Community Medicine, Macon, GA 31207, *haomia@yahoo.com*; Michael W. Link, Centers for Disease Control and Prevention; Ali H. Mokdad, Centers for Disease Control and Prevention; James Holt, Centers for Disease Control and Prevention; Lei Li, RTI International; Paul S. Levy, RTI International

***Key Words:*** small area estimation, GIS, GLMM, data smooth

This paper summarizes the methodology for the rapid production of county-level, small-area estimates (SAEs) of influenza vaccine coverage from the ongoing Behavior Risk Factor Surveillance System (BRFSS). Our objectives are construction of monthly vaccination rates that could be updated steadily and investigation of the differences between counties during the 2004--05 influenza season. We developed a procedure that utilizes random effects logistic regression followed by "head-banging" (a form of spatial smoothing). This method is suitable for situations such as this, where most of the counties have few respondents or none at all and there is a need for estimates to be generated quickly for purposes of surveillance. It also is adaptable to spatial-temporal mapping using the GIS system, which can provide visual scanning for clusters of hotspots having low vaccination coverage.

### Evaluating the Small-Area Estimates of the 2004--05 County-Level Influenza Vaccination Rates

❖ Lei Li, RTI International, 3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-2194, *lei@rti.org*; Paul S. Levy, RTI International; Akhil Vaish, RTI International; Michael W. Link, Centers for Disease Control and Prevention; Ali H. Mokdad, Centers for Disease Control and Prevention; Lina Balluz, Centers for Disease Control and Prevention; Haomiao Jia, Mercer University School of Medicine

***Key Words:*** small area estimate, influenza vaccination rate, composite estimate, survey weighted HB estimate

The estimates of the influenza vaccination rates were produced for each county by population subgroup from the Behavioral Risk Factor Surveillance System. An expedient small-area estimation method was used. This evaluation focuses on the agreement of the estimates with estimates known to be more reliable from various angles. First, the county-level estimates were aggregated based on population distribution to generate state estimates, which were compared with composite state estimates and the survey-weighted hierarchical Bayes state estimates. Second, random samples of counties were selected, and the estimates in a selected county and its surrounding counties were aggregated to generate neighboring area estimates, which were compared with composite estimates for the areas. Third, design-based estimates were obtained in counties where the sample size was deemed large as a comparison.

## 5 Pipeline Issues in Recruiting Federal Statisticians ✪

Committee on Membership Retention and Recruitment, Section on Statistical Education, Committee on Career Development

**Sunday, August 6, 2:00 pm–3:50 pm**

### JPSM: 14 Years of Training Federal Statisticians

❖ Richard Valliant, University of Michigan, 1218 Lefrak Hall, University of Maryland, College Park, MD 20784, *rvalliant@survey.umd.edu*; Roger Tourangeau, University of Maryland

JPSM is a unique partnership among two universities---the University of Maryland and the University of Michigan---and a private, for-profit corporation---Westat. It provides in-career training to a number of federal and other survey researchers, while offering graduate-level training to recent college graduates. JPSM offers three categories of courses: professional development courses (some taught by non-JPSM faculty and offsite), leading in some cases to nondegree qualifications; a master's program; and a doctoral program. JPSM faces a number of challenges in coming years to meet the needs of federal agencies and the field of survey methodology, in general. This talk will discuss the history of JPSM and its future plans.

### Gulliver Tied down by Red Tape? The Federal Government's Challenges as an Employer of Statisticians

❖ Janice Lent, Research and Innovative Technology Administration, 400 7th Street, SW, Suite 4432, Washington, DC 20590, *janice.lent@dot.gov*

***Key Words:*** federal statistical system

As the largest single employer of statisticians, the federal government plays an influential role in the U.S. statistical community. Encumbered by relatively inflexible rules for employment and career advancement, the government faces unique challenges in its efforts to maintain and modernize the myriad of programs that make up the federal statistical system. Many statisticians who play key roles in federal programs are approaching retirement age, and they will be difficult to replace. Graduate programs in statistics attract a high percentage of international students, while Federal employment requires U.S. citizenship. Moreover, skilled statisticians regularly leave federal employment for the private sector, because advancement opportunities in nonmanagerial roles often are limited.

### Federal Statisticians in the Physical and Engineering Sciences

❖ William Guthrie, National Institute of Standards and Technology, Statistical Engineering Division, 100 Bureau Drive, Stop 8980, Gaithersburg, MD 20899-8980, *will.guthrie@nist.gov*

***Key Words:*** government, physical sciences, engineering, recruitment

Federal statisticians in the physical and engineering sciences make up a small percentage of federal statisticians as a whole. As such, recruitment of statisticians for federal posts in physical science agencies requires extra outreach and a constant effort to maintain visibility. The areas of statistical expertise needed for these jobs also differs from that needed in many other agencies and types of statistics jobs, which can mean recent graduates face a steep learning curve when starting in these positions. On the other hand, statistical work in the physical sciences offers an abundance of interesting and rewarding applications and opportunities for research. This talk will illustrate these and other issues that affect the recruitment of statisticians for federal posts in physical and engineering sciences agencies.

# 6 Surveillance Geoinformatics and Hotspot Dynamics for Prediction, Policy, and Management ✪

Environmental and Ecological Statistics, Section on Statistics and the Environment

**Sunday, August 6, 2:00 pm–3:50 pm**

## Spatiotemporal Geoinformatic Disease Surveillance

❖ Stephen L. Rathbun, University of Georgia, 152 Environmental Health Science Building, Dept. of Health Admn, Biostatistics, Epidemiology, Athens, GA 30602, *rathbun@uga.edu*; Ganapati P. Patil, The Pennsylvania State University

***Key Words:*** upper level set, geosurveillance, generalized linear mixed model, spatial probit model, conditional autoregressive model

Disease surveillance for early warning requires quick and efficient methods for delineation of disease hotspots in space and time. The upper level set (ULS) scan statistic is a computationally efficient approach for delineating hotspots of arbitrary shape. However, the current version of the ULS scan statistic assumes the data are distributed independently, an assumption that may be untenable for georeferenced data. We shall investigate the statistical properties of the ULS scan statistic under a variety of models for spatial dependence, including spatial probit models for the spatial distribution of hotspots and generalized linear mixed models with conditional autoregressive random effects. Simple methods are sought for testing the significance of hotspots, adjusting for the effects of spatial dependence. Our approach is illustrated using zoonotic disease data.

## Crime Mapping and Hotspot Detection

❖ Reza Modarres, The George Washington University, Department of Statistics, 2140 Pennsylvania Avenue, Washington, DC 20052, *reza@gwu.edu*; Ganapati P. Patil, The Pennsylvania State University

***Key Words:*** hotspots, geoinformatic surveillance, scan statistic, upper level set scan, crime mapping

We discuss the Upper Level Scan (ULS) statistic and its theory as well as implementation for univariate and bivariate data. The ULS-tree and ULS-hotspot algorithm that depict and maintain a list of connected components of the rate surface at each level of the ULS tree are examined. The general theory of bivariate hotspot detection is discussed, including the bivariate binomial and Poisson models and the multivariate exceedance approach. We also examine the joint and intersection methods for detecting bivariate hotspots and study the sensitivity of the joint hotspots to the degree of association between the variables. We investigate the applications of ULS detection methodology in mapping of crime hotspots in the counties of several U.S. states.

## Applications of Hotspot Detection Analysis to Large-Scale Plant Disease Forecasting: Case Study of Fusarium Head Blight

❖ Murali Haran, The Pennsylvania State University, 326 Thomas Building, Statistics Department, University Park, PA 16802, *mharan@stat.psu.edu*

***Key Words:*** hotspot geoinformatics, disease mapping, spatial misalignment, crop disease, lattice data, Markov chain Monte Carlo methods

Fusarium Head Blight (FHB) is a disease that affects wheat crops, frequently resulting in enormous financial losses for farmers. Accurate forecasts of FHB are vital for cost effective management practices. Risk predictions for FHB are now available from weather-driven models of disease biology based on experimental data. Some sparse survey information about true disease rates is also available. We propose a hierarchical model that naturally integrates the various sources of information to predict risk. The underlying true disease rates are assumed to be spatially and temporally dependent. The survey data are treated as a function of the true disease rate and measurement error while the risk predictions are assumed to be functions of the true disease risk and terms that account for bias and random error. We conclude with an application of our approach to FHB data from North Dakota.

# 7 Density-Based Clustering ●

Section on Statistical Computing, Section on Statistical Graphics, Section on Nonparametric Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

## Variable Bandwidth Mode Testing

❖ Michael C. Minnotte, Utah State University, Department of Mathematics and Statistics, Logan, UT 84322-3900, *Mike.Minnotte@usu.edu*

***Key Words:*** higher-order density estimation, kernel density estimation, multimodality, Silverman's test, variable bandwidth

We investigate the effects of applying variable bandwidth, fourth-order density estimation to Silverman's (1980, 1981) test of multimodality of a density. Silverman's test uses as a test statistic the "critical bandwidth," the largest bandwidth for a kernel density estimate in which the number of modes is at least the alternative-specified k (often 2). In this study, we show that use of variable-bandwidth, fourth-order (reduced bias) estimates similar to those proposed by Abramson (1982) can produce superior level and power results for mode testing compared to those generated from traditional second-order, normal kernel density estimates. This is not the case for other fourth-order methods of density estimation. In particular, the use of fourth-order kernels for mode testing should be discouraged strongly.

## Generalized Single-Linkage Clustering

❖ Werner Stuetzle, University of Washington, Box 354322, Department of Statistics, Seattle, WA 98195-4322, *wxs@stat.washington.edu*; Rebecca Nugent, University of Washington

***Key Words:*** cluster tree, single linkage, nearest neighbor density estimate, minimal spanning tree, connected components

The goal of clustering is to identify distinct groups in a dataset and assign a group label to each observation. To cast clustering as a statistical problem, we regard the data as a sample from some unknown population density, p(x). The hierarchical cluster structure of p(x) is represented by the cluster tree of its level sets. We estimate the cluster tree of p(x) by the cluster tree of a density estimate. For most types of density estimates, exact computation of the cluster tree is impossible. Generalized single-linkage clustering is a graph-based method for computing an approximation to the cluster tree that can be applied to any density estimate. Application of our method to the nearest-neighbor density estimate is equivalent to single-linkage clustering.

### Mixture Model Building for High-Dimensional and Functional Data

❖ Catherine Loader, The University of Auckland, Auckland, New Zealand, *csea@herine.net*; Ramani S. Pilla, Case Western Reserve University

***Key Words:*** clustering, mixture models, functional data, density estimation

Mixture models provide a conceptually simple tool for modeling data distributions and identifying components, or clusters, within a dataset. Standard fitting procedures have drawbacks ranging from slow convergence to convergence to local maxima to selection of cluster components. These issues are especially acute in high dimensions. We present new tools to address these shortcomings. Our work begins with a large number of candidate components and uses likelihood ratio methods to build a parsimonious model. Use of a penalty function allows us to incorporate model selection criteria into the algorithm, leading to automated decisions about mixture complexity. The technique works from a distance metric defined on the parameter space and a likelihood function. This enables extension to nonstandard problems, including "large p, small n" settings arising from functional data analysis.

## 8 Statistical Methods for Oral Health Research ●

ENAR, Biometrics Section, WNAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### Analysis of Correlated Dental Data: Challenges and Recent Developments

❖ Brian G. Leroux, University of Washington, 1959 NE Pacific Street, Box 357475, Seattle, WA 98195, *leroux@u.washington.edu*

***Key Words:*** estimating equation, clustered data, cluster size, correlation structure

In clinical dental research, outcomes typically are recorded on many sites within each patient's mouth. These outcome data present a unique set of challenges for statistical analysis, including large cluster sizes, multilevel data structures (e.g., teeth within patients, sites or surfaces within teeth), complex correlation structures, informative cluster sizes, missing data at different levels, and a small number of clusters. The impact of these features on bias and precision of statistical analyses will be discussed. Recent methodological developments by the author and colleagues will be presented, with an emphasis on estimating equations methods that address issues arising with large cluster sizes, complex correlation structures, and informative cluster sizes.

### Finding the Right Pair of Genes: Adding a Genetic Component to Existing Oral Health Studies

❖ Deborah V. Dawson, The University of Iowa, N439E Dental Science Building - Dows Institute, College of Dentistry, Iowa City, IA 52242-1010, *deborah-dawson@uiowa.edu*

***Key Words:*** genetic analysis, oral health, candidate genes, genetic linkage analysis, family-based association studies, case-control studies

There is currently considerable interest in the augmentation of existing oral health studies by addition of a genetic component. These extant studies frequently reflect such traditional designs as clinical trials, case-control studies, and longitudinal epidemiologic investigations. In this talk, a number of special considerations associated with the design and analysis of such extensions are addressed and several paradigms for addressing genetic research hypotheses are explored, including those for assessment of gene-gene and gene-environment interactions. Models to address concerns related to bias of ascertainment are presented. These are illustrated using examples from ongoing efforts in areas such as periodontology, enamel defects, and caries research.

### A Flexible Model for Recurrent Event Outcomes in Oral Health

❖ Elizabeth H. Slate, Medical University of South Carolina, Department of Biostatistics, Bioinformatics, and Epidemiology, PO Box 250835, 135 Cannon St., Suite 303, Charleston, SC 29425, *slateeh@musc.edu*; Edsel A. Pena, University of South Carolina

***Key Words:*** oral health, recurrent event, longitudinal biomarker

This talk will describe a flexible model for recurrent event outcomes and discuss its application in oral health research. The model accommodates the effects of interventions following event occurrences, correlation among events, a strengthening or weakening effect on the subject due to accumulating events, and covariate effects. Recent research that incorporates this model in a latent class framework for jointly modeling a longitudinal biomarker and recurrent event outcome also will be presented.

### Statistical Approaches for Dealing with Missing Tooth- and Surface-Level Data in Caries Research

❖ Jason Roy, University of Rochester, 601 Elmwood Ave., Rochester, NY , *jason_roy@urmc.rochester.edu*

***Key Words:*** hierarchical models, censoring, spatial correlation, dentistry

Dental caries is a common oral disease that results in demineralization of the tooth. In caries research, the number of decayed, missing, or filled tooth surfaces (DMFS) are analyzed often. However, missing data due to missing teeth or surfaces (e.g., due to sealants) are common in oral health studies. There is no consensus on how to count these surfaces when coming up with a DMFS score. The problem is that the risk of caries varies dramatically across locations. Often, higher-risk teeth and surfaces are more likely to be missing. We propose a model-based approach for imputing these 'missing' values. Our method exploits the spatial correlation between surfaces and allows for the possibility of informative censoring. This approach should increase efficiency, and in some cases reduce bias. We illustrate the method using data from a study of childhood lead exposure.

## 9 Empirical Likelihood-Based Semiparametric Inference ●

Biometrics Section, Section on Nonparametric Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

### Empirical Likelihood-Based Inference for Comparison of Two Populations with Censored Data

❖ Hua Liang, University of Rochester Medical Center, Department of Biostatistics and Computational Biology, 601 Elmwood Avenue, Box 630, Rochester, NY 14642, *hliang@bst.rochester.edu*

*Key Words:* estimating equation, empirical likelihood ratio, Kaplan-Meier estimation, nuisance parameter, receiver operating characteristic

This session will compare two samples of censored data when the model for one sample is parametric and the other is nonparametric. The parameter of interest may represent, for example, the receiver operating characteristic. The estimating equation function may be discontinuous on a nuisance parameter. The proposed empirical likelihood ratio is shown to be asymptotically chi-squared, and the advantage of the confidence interval derived from the empirical likelihood method over its counterpart---constructed from the common estimating equation---is demonstrated by simulation experiments. A real example is analyzed with the proposed methods.

### Empirical Likelihood-Based Inferences for Receiver Operating Characteristic Curves in the Presence of Verification Bias

❖ Jing Qin, National Institute of Allergy and Infectious Diseases, 6700B Rockledge Drive, MSC 7609, Bethesda, MD 20892, *jingqin@niaid.nih.gov*

*Key Words:* diagnostic medicine, ROC

Diagnostic medicine has progressed tremendously in the last several decades, and the trend promises to continue. The most widely used measures of accuracy are sensitivity and specificity. The receiver operating characteristic (ROC) curve is a plot of one---specificity vs. sensitivity. There is a verification bias problem in ROC estimation if only a selected subset of patients receive the “gold standard” test and the unverified patients are ignored. In this talk, we show that empirical likelihood can be used to correct the bias. The empirical likelihood-based method can produce more efficient estimates by using the unverified data.

### Empirical Likelihood for Accelerated Failure Time Model

❖ Mai Zhou, University of Kentucky, Department of Statistics, Lexington, KY 40506-0027, *mai@ms.uky.edu*

*Key Words:* random censoring, semiparametric regression model, testing hypothesis, Wilks theorem, Buckley-James estimator

By Accelerated Failure Time (AFT) model, we mean log(Y) = b X + e, where the distribution of the random error e is unspecified and Y is subject to censoring. We shall compare several different estimation methods and corresponding empirical likelihoods that generalize Owen (1991 “Empirical Likelihood for Linear Models.” The Annals of Statistics. 19:1725-1747). Special attention is paid to generalizations of empirical likelihood that admits a Wilks theorem, which conclude that the empirical likelihood ratio has a chi square limiting distribution under null hypothesis. This includes (but not limited to) the well known Buckley-James estimator. Examples and simulations of empirical likelihood with AFT model will also be presented.

### Nonparametric Imputation of Missing Values for Estimating Equation-Based Empirical Likelihood Inference

Song X. Chen, Iowa State University; ❖ Dong Wang, University of Nebraska, Department of Statistics, 340 Hardin Hall North, East Campus, Lincoln, NE 68583-0963, *dwang@iastate.edu*

*Key Words:* empirical likelihood, estimating equations, kernel estimation, missing values

We consider semiparametric inference for parameters defined by a set of estimating equations when certain components of the random variable under study are missing. The missing values are imputed from a nonparametric conditional distribution estimator based on the fully observed portion of the data under the notion of Missing at Random. Empirical likelihood is then formulated by utilizing imputed estimating equations. We show that the empirical likelihood inference can achieve the semiparametric efficiency bound, at least when the number of estimating equations is the same as the dimension of parameter. Empirical investigations are reported for cases where either a response variable or covariates are missing.

## 10 New Directions in Bayesian Joint Modeling of Longitudinal and Survival Data ●

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival

❖ Elizabeth Brown, University of Washington, *elizab@u.washington.edu*

*Key Words:* joint longitudinal and survival models, b-splines, MCMC, HIV/AIDS

Often, when jointly modeling longitudinal and survival data, we are interested in a multivariate longitudinal measure that may not fit well with linear models. To overcome this problem, we propose a joint longitudinal and survival model that has a nonparametric model for the longitudinal markers. We use cubic B-splines to specify the longitudinal model and a proportional hazards model to link the longitudinal measures to the hazard. To fit the model, we use a Markov chain Monte Carlo algorithm. We select the number of knots for the cubic B-spline model using the Conditional Predictive Ordinate (CPO) and the Deviance Information Criterion (DIC). We apply this method to examine the link between viral load, CD4 count, and time to event in data from an AIDS clinical trial.

### Extensions of the Standard Joint Model

❖ Jeremy M. G. Taylor, University of Michigan, 1420 Washington Heights, Department of Biostatistics, Ann Arbor, MI 48109, *jmgt@umich.edu*

*Key Words:* joint models, cured fraction, CPO statistics

In this paper, I will consider extensions of the standard joint model for longitudinal and survival data to more complex situations. This work is motivated by applications in radiation therapy of prostate cancer, in which longitudinal PSA and clinical endpoint data are collected. The extensions of the standard model include incorporation of a cured fraction, use of CPO statistic to assess the need for a cured fraction, nonlinear longitudinal models, comparison of non-nested models, hierarchical models for data from multiple institutions, covariate selection, time-dependent hazard models that involve more than just the current value of the longitudinal variable, and use of the models for individualized predictions.

### Bayesian Semiparametric Methods for Joint Modeling of Longitudinal and Survival Data

❖ Adam Branscum, University of Kentucky, College of Public Health, 121 Washington Ave., Ste. 205B, Lexington, KY 40536, *abran3@email.uky.edu*; Timothy Hanson, University of Minnesota; Wesley O. Johnson, University of California, Irvine

***Key Words:*** Polya trees, Bayesian nonparametrics

Development of statistical models for the analysis of survival data with time-dependent covariates is an active area of research. The observed values of time-dependent covariates are often available at discrete time points. To circumvent potential bias inherent in last-value-carried-forward approaches, current methodologies propose modeling longitudinal covariate processes jointly with survival data. Recent approaches have focused primarily on modeling longitudinal processes using mixed effects methods, while the survival component is characterized by a Cox proportional hazards regression. In this talk, we will review briefly Bayesian approaches to joint survival/longitudinal modeling and present a novel Bayesian semiparametric approach, which is illustrated using data relating blood chemistry measurements to survival time of kidney dialysis patients.

## 11 Modern Monte Carlo Methods for Statistical Inference

IMS, Section on Bayesian Statistical Science, Section on Nonparametric Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

### An Overview of SMC and Adaptive MCMC

❖ Anthony Brockwell, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA 15213, *abrock@stat.cmu.edu*

Sequential Monte Carlo and adaptive Markov chain Monte Carlo schemes are discussed at an introductory level. Their use is motivated using examples in the fields of finance, signal processing, and neurobiology. Comparison between these methods and standard MCMC methods will be given, and (time permitting) a new algorithm combining some of the features of both SMC and MCMC will be introduced.

### Simulated Tempering Made Easy

❖ Yves Atchade, University of Ottawa, Ottawa, ON K1N 6N5 Canada, *yatchade@uottawa.ca*

***Key Words:*** simulated tempering, Markov chain Monte Carlo, phylogenetic trees

Simulated tempering is a well-known technique to improve on the mixing rate of MCMC samplers. The method works by introducing a finite set of 'temperatures' at which the Markov chain might converge faster. But choosing these temperatures and the associated weights is difficult in practice. In this talk, I will introduce a new approach to simulated tempering where the temperatures and weights are handled by the algorithm automatically. I will illustrate the performance of the method on the interesting problem of sampling from the posterior distribution of phylogenetic trees in molecular evolution.

### Adaptive Monte Carlo Computing Methods

❖ Christophe Andrieu, University of Bristol, UK, *c.andrieu@bristol.ac.uk*

***Key Words:*** adaptive, Monte Carlo, graphical models, physics

Adaptive Monte Carlo methods have received some attention recently. While the theory is well-developed now, the methodology is still in its infancy. In this talk, applications of adaptive computing methods to problems ranging from inference in graphical models to problems arising in physics will be presented.

## 12 Advances in Item Response Theory

Social Statistics Section

**Sunday, August 6, 2:00 pm–3:50 pm**

### Limited Information Goodness-of-Fit Testing in Multidimensional Contingency Tables

❖ Harry Joe, The University of British Columbia, Department of Statistics, Vancouver, BC V6T 1Z2 Canada, *harry@stat.ubc.ca*

***Key Words:*** multivariate discrete data, item response theory, composite likelihood

We introduce a family of goodness-of-fit statistics for testing composite null hypotheses in multidimensional contingency tables of arbitrary dimensions. These statistics are quadratic forms in marginal residuals up to order r. They are asymptotically chi-square under the null hypothesis when parameters are estimated using any consistent and asymptotically normal estimator. For an item response model, in nonsparse situations when the null distribution of X2 is approximately chi-square, we show empirically that the proposed statistics are also more powerful than X2. The proposed statistics, applied to subtables, also can be used for a piecewise goodness-of-fit assessment to determine the source of misfit in poorly fitting models. This research is joint with Albert Maydeu-Olivares.

### Sensitivity of Latent Trait Analysis to Highly Skewed Ability Distributions

❖ David Dailey, Woodcock-Munoz Foundation, 17810 3rd Ave., NE, Shoreline, WA 98155, *David@PandoraHouse.org*; Elena Erosheva, University of Washington

***Key Words:*** latent trait analysis, skewed ability distributions, marginal maximum likelihood estimation, two-parameter logistic model, simulation study, parameter estimates

In this simulation study, we explore parameter estimation sensitivity to characteristics of the generating ability distribution and choices of the fitting distribution. Using marginal maximum likelihood estimation for the two-parameter logistic response model, we examine the usual choice of the standard normal, a double gamma for a bimodal case, and an exponential for a highly skewed case. We obtain simulated data under these distributions, varying test length and sample size. We examine effects of using incorrect fitting distributions by analyzing recovery of item parameters and item ranking, as well as goodness-of-fit. We find that the exponential distribution has a high impact on parameter estimates. Thus, it is important not to use the normal fitting distribution with exponential-generated data or the exponential fitting distribution with normal-generated data.

### Checking the Appropriateness of the Statistical Model Used in National Assessment of Educational Progress

❖ Sandip Sinharay, Educational Testing Service, MS-12T, Rosedale Road, Princeton, NJ 08541, *ssinharay@ets.org*

***Key Words:*** model fit, bootstrap, psychometrics, educational testing, educational measurement

National Assessment of Educational Progress (NAEP) is an ongoing survey of the academic achievement of the school students in the United States in a number of subject areas, such as reading, writing, and mathematics. NAEP plays an important role in educational policymaking in the United States. To ensure quality control, it is important to rigorously assess the fit of the statistical model used in NAEP. We apply a simulation-based model fit approach to investigate whether the NAEP statistical model can predict simple data summaries, such as proportion-correct scores on the test questions and the average raw scores of different examinee subgroups (e.g., whites, males, etc.), adequately. The suggested technique is applied to operational NAEP data.

### Improve Variance Estimation for the Assessments Based on the Plausible Values Approach

❖ Jiahe Qian, Educational Testing Service, Rosedale Road, T198, Princeton, NJ 08541, *jqian@ets.org*; Shelby Haberman, Educational Testing Service

***Key Words:*** jackknifed variance estimation, assessments, NAEP, conditioning models

The jackknife procedure is used to estimate the variance of plausible values generated by the conditioning process for assessments such as NAEP. The plausible values are generated from a posterior distribution based on conditioning on test scores and demographic and education-associated variables. Then, the plausible values are transformed to reporting scale. Currently, the jackknifed estimation is based on one set of plausible values. This study will estimate jackknifed variance from all sets of the plausible values generated from jackknife replicate samples by conditioning models. Although the process is computationally intensive, it is no longer an obstacle with the growth of modern computing capabilities. Data from the NAEP assessments will be used to compare new results with the old and conduct variance analysis.

### Estimation of Measurement Errors at Observed and Scaled Scores

❖ Michelle Liou, Academia Sinica, Institute of Statistical Science, Academia Sinica, Taipei, 115 Taiwan, *mliou@stat.sinica.edu.tw*; Philip E. Cheng, Academia Sinica

***Key Words:*** arcsine transformation, beta-binomial, standard errors of measurement, log-linear smoothing, scaled scores, true scores

In the placement service and licensure examination, the observed scores or scaled observed scores are used to assign test-takers into competence levels. It is well-known that measurement errors may vary widely according to score levels. We derived computational formulae for estimating bias and error variance at each observed score and extended the use of the formulae to scaled scores that were nonlinear transformations of observed scores. In the empirical study, the formulae were applied to assessing conditional measurement errors at observed scores on an English test administered to 58,054 takers in a national assessment. The scores were transformed using an arcsine function for stabilizing error variances across the score range. The estimated bias and error variance at each observed and scaled score was compared to validate the use of the arcsine transformation.

## 13 Bayesian Modeling of Biomedical Data ●

Biometrics Section, Section on Bayesian Statistical Science, WNAR, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### Analysis of Longitudinal Clinical Trial Data with Informative Dropout

❖ Xiaohong Yan, University of California, Los Angeles, 3170 Sawtelle Blvd., Apt 305, Los Angeles, CA 90066, *xhyan@ucla.edu*; W. John Boscardin, University of California, Los Angeles

***Key Words:*** longitudinal study, missing data, informative dropout, Bayesian

We present a variety of approaches for analyzing data from a clinical trial with incomplete longitudinal data. It is often inappropriate to assume the missingness is at random for such data, especially when dropouts are common. We extend the methodology of Carpenter, Pocock, and Lamm (2002) to simultaneously model multivariate incomplete longitudinal data and the potentially informative dropout (ID) mechanism using a Bayesian approach. The methodology is illustrated through reanalysis of rheumatology clinical trial data from a study of penicillamine treatment for scleroderma patients. We examine two primary longitudinal measures for assessing the outcome of the study. We compare the results for univariate and bivariate informative dropout models to approaches commonly employed, such as complete case analysis, LOCF, and multiple imputation.

### Bayesian Approach to Multiple Changepoint with Application to HIV Immunologic Responses

❖ Pulak Ghosh, Georgia State University, 30 Pryor Street, Department of Math and Stat, Atlanta, GA 30303, *pghosh@mathstat.gsu.edu*; Kaushik Ghosh, New Jersey Institute of Technology; Ram Tiwari, National Institutes of Health

***Key Words:*** changepoint, HIV-AIDS, Dirichlet process

We will assess the time to development of drug resistance in HIV-infected individuals treated with antiviral drugs by using longitudinal HIV viral load measurements. We will assume a piecewise linear changepoint model for log HIV RNA and develop Bayesian procedures to estimate the unknown parameters. We are interested in the rate of decline/rebound and the time of changing points for the viral loads. For example, a viral load trajectory can be divided into three stages: rapid decline, slow decline, and rebound. Usually, in the literature, one assumes each subject has the same number of changepoints; however, we relax this assumption by assuming a different changepoint for each subject. We also consider the clustering of the changepoints using a Dirichlet process mixture prior on the changepoints.

### Bayesian Model Checking for a Longitudinal Binary Variable

❖ Catherine Crespi, University of California, Los Angeles, School of Public Health, CHS 51236D, Los Angeles, CA 90095-1772, *ccrespi@ucla.edu*; W. John Boscardin, University of California, Los Angeles; William G. Cumberland, University of California, Los Angeles

***Key Words:*** longitudinal data, binary data, model fit, Bayesian model, posterior predictive model checking

In many longitudinal studies, the outcome of interest is a binary variable and the data are a series of 0/1 indicators. We develop methods for assessing the fit of a Bayesian model for a longitudinal binary variable using posterior predictive model checking. Application to clinical data demonstrates how the methods can help discriminate among competing models.

### Modeling Multivariate Biomedical Data with Polynomial Smoothing Splines

❖ Hector Lemus, University of California, Los Angeles, 3100 Sawtelle Building 302, Los Angeles, CA 90066, *hlemus@ucla.edu*; W. John Boscardin, University of California, Los Angeles

*Key Words:* smoothing splines, state space models, Bayesian model

Biostatisticians are asked frequently to perform inference for datasets with multivariate repeated or longitudinal measurements. Investigators typically will ask: Are measures X and Y correlated? Did either of measures X or Y exceed clinically important thresholds? We have extended the work of Anderson, Jones, and Swanson (1990) and Brown and MaWhinney et al. (2001) to develop a Bayesian multivariate smoothing spline model in a state-space framework. The key advance is that our model allows for incorporation of substantial intersubject heterogeneity in a parsimonious manner. The model is applied to two datasets from the UCLA Brain Injury Research Center to make statistical inference about correlation of measures and threshold exceedance.

### Real-Time Learning for Heterogeneous Multivariate Longitudinal Data

❖ W. John Boscardin, University of California, Los Angeles, Department of Biostatistics, Los Angeles, CA 90095, *jbosco@ucla.edu*; Hector Lemus, University of California, Los Angeles

*Key Words:* state space models, Bayesian forecasting, smoothing splines

Medical research often involves the longitudinal collection of multivariate measurements on extremely heterogeneous subjects with the goals of estimating the overall and subject-specific time courses of one or more of the measurements, assessing the correlation of pairs of the measurements, and generating short- and long-term predictions for subjects throughout the data collection period. A Bayesian multivariate smoothing spline model in a state-space framework can be used to analyze these data in a prospective real-time setting (i.e., it is possible to efficiently learn about the current subject's recent time course, correlations, and predictions given the available data for the current subject and the ensemble of previous subjects). We demonstrate this methodology using data from severe head trauma subjects.

## 14 Planning Medical Device Studies

●

Biopharmaceutical Section, Biometrics Section, ENAR
**Sunday, August 6, 2:00 pm–3:50 pm**

### Statistical Myths in the Design and Analysis of Clinical Trials

❖ Victor Hasselblad, Duke University, 2400 Pratt Street, Room 0311 Terrace Level, Durham, NC 27705, *hasse002@mc.duke.edu*

*Key Words:* clinical trial, design, assumptions, normality, historical controls

There are common myths in the design and analysis of clinical trial data that continue to be spread by researchers and journal reviewers. These can be demonstrated to be essentially incorrect. They include normality is required for t tests, so nonparametric tests are preferred; knowledge of the distribution (normality) is not required for power calculations---just the means and standard deviations; stratification can improve power for a dichotomous primary endpoint in a randomized clinical trial; device trials don't need active controls---historical controls work just fine; and odds ratios are more appropriate for the analysis of dichotomous data than are relative risks. Myths will be dispelled using real---rather than hypothetical---trial data.

### New Medical Device? When Clinical Data Are Needed for a New Medical Device

❖ Jeng Mah, American Medical Systems Inc., 10700 Bren Road West, Minnetonka, MN 55343, *jeng.mah@ammd.com*

*Key Words:* decision rule, loss, Bayesian hypothesis testing, clinical study

Medical devices are mechanical systems of hardware and software that deliver medical treatments. Compared to drugs, the design of a medical device can be modified gradually and often, the effects being more predictable. The physical nature of medical devices poses a challenge in determining applicability of safety and effectiveness data collected in predicate devices to a new device. In a regulatory setting, decisions about whether new studies are needed before allowing the marketing of new medical devices affect public well-being and the interest of the medical device industry. FDA has provided a guidance document outlining an "evidence model decision schematic," without an attempt to define "reasonable assurance" and "substantial equivalence" criteria for determining when clinical data are needed for new medical devices.

### Statistical Review Quality Assessment for Therapeutic PMA Submissions

❖ Lilly Yue, U.S. Food and Drug Administration, 1350 Piccard Drive, HFZ 550, Rockville, MD 20850, *lilly.yue@fda.hhs.gov*

*Key Words:* therapeutic medical device, planning clinical study, preparing PMA submission

To standardize the structure of statistical review memos and ensure review quality, the Division of Biostatistics (DBS) in FDA's Center for Devices and Radiological Health developed a comprehensive list of items to be visited in an in-depth review of a PMA (premarket approval) application for a therapeutic device. This has served as a valuable source for statisticians in DBS when reviewing PMA submissions for therapeutic devices. In this presentation, I will emphasize elements of the list for a quality statistical review with the hope that such a regulatory perspective is useful to statisticians involved in planning clinical studies and preparing PMA submissions for therapeutic devices.

### What Device Pivotal Studies Have in Common: Recurring Themes in Study Planning

❖ Philip Lavin, Averion Inc., 4 California Ave., Framingham, MA 01701, *plavin@averioninc.com*

*Key Words:* medical device, pivotal study, study design, effectiveness, safety

Medical device approvals require a well-conducted pivotal study to gain FDA approval. There is a common misconception that these pivotal studies are unique and nonstandard. This talk will identify common statistical themes faced in the planning of pivotal device studies inclusive of study design options, historical data use, control group choices, endpoint multiplicity issues, offset choice for a noninferiority study, data pooling across strata and sites, sample-size calculation, prospective analysis plan requirements, population definitions for effectiveness and safety, and adverse events vs. complications. ICH and GCP linkage will be cited. Examples will be provided to depict agency precedents and sponsor objectives.

# 15 Bayesian Student Paper Competition II

Section on Bayesian Statistical Science

**Sunday, August 6, 2:00 pm–3:50 pm**

## Statistical Inference for Nonlinear Models Involving Ordinary Differential Equations

❖ Lovely Goyal, North Carolina State University, 2435 Wycliff Road, Apt. H, Raleigh, NC 27607, *lgoyal@stat.ncsu.edu*; Sujit Ghosh, North Carolina State University

***Key Words:*** Bayesian, differential equations, MCMC, nonlinear models, fixed effect

In the context of nonlinear fixed-effect modeling, it is common to describe the relationship between a response variable and a set of explanatory variables by a system of nonlinear ordinary differential equations (ODEs). More often, such a system of ODEs does not have any analytical closed-form solution, making parameter estimation for these models challenging and computationally demanding. Two new methods based on Euler's approximation are proposed to obtain an approximate likelihood that is analytically tractable, thus making parameter estimation computationally less demanding than other methods. These methods are illustrated using data on growth colonies of Paramecium aurelium. Simulation studies are presented to compare the performances of these new methods to established methods in the literature.

## Weighted Model-Based Clustering for Remote Sensing Image Analysis

❖ Joseph Richards, Carnegie Mellon University, 2718 Allenton Ave., Hacienda Heights, CA 91745, *joeyrichar@gmail.com*; Johanna Hardin, Pomona College

***Key Words:*** weighted likelihood, mixture model, EM algorithm, large data sets, geologic map

We introduce a weighted method of clustering the individual units of a segmented image. Our weighted model-based clustering method (WMBC) employs a weighted likelihood and assigns fixed weights to each unit corresponding to the number of pixels located within the unit. To conquer computing time and space issues, WMBC uses the means and standard deviations of the pixel values within each unit. With both simulated and real datasets, we show WMBC is more accurate than standard model-based clustering.

## On Bayesian Analysis of Generalized Linear Models Using Jacobian Technique

❖ Sourish Das, University of Connecticut, 215 Glenbrook Road, Department of Statistics, Storrs, CT 06269, *sourish.das@uconn.edu*; Dipak Dey, University of Connecticut

***Key Words:*** Bayesian central limit theorem, canonical link, conjugate prior, Jacobian technique, natural exponential family, regression parameter

In this paper, we obtain an exact estimator of the regression parameter for generalized linear model using Jacobian technique. We restrict ourselves within the natural exponential family for response variable and choose the conjugate prior for the natural parameter. Then using Jacobian transformation, we obtain the posterior distribution of the canonical link function and obtain posterior mode for the link. Under the full rank assumption of covariate matrix, we find an exact estimator for regression parameter for the natural exponential family. We then obtain an exact estimator of regression parameter for Poisson model with log link function, Bernoulli responses with logit link function, and Binomial response variable with the same logit link function. Finally, an illustrative, real-life example is given for the Poisson model with log link.

## Using Incompatibility To Build Fast Gibbs Samplers

❖ Taeyoung Park, Harvard University, One Oxford Street, 7th floor, Cambridge, MA 02138, *tpark@stat.harvard.edu*; David A. van Dyk, University of California, Irvine

***Key Words:*** Bayesian hierarchical modeling, Gibbs sampler, functional incompatibility, incompatible Gibbs sampler, partially marginalized Gibbs sampler, spectral analysis

Increasing computational power continues to popularize statistical computing methods and tools at an ever-increasing rate. Among the popular, computationally intensive methods, the Gibbs sampler (Geman and Geman 1984) has been spotlighted because of its simplicity and power to fit highly structured models. Its sometimes slow convergence, however, has been a long-standing complaint, especially when the highly structured models are fitted. In this paper, we present partially marginalized Gibbs sampling strategies that improve the convergence characteristics of an ordinary Gibbs sampler by exploiting a set of functionally incompatible conditional distributions avoided in the construction of the ordinary Gibbs sampler. We illustrate these strategies by fitting a highly structured multi-level spectral model, which an ordinary Gibbs sampler fails to fit.

## Improving Classification When a Class Hierarchy Is Available Using a Hierarchy-Based Prior

❖ Babak Shahbaba, University of Toronto, 55 Ellerslie Ave., Apt. 1717, Toronto, ON M2N 1X9 Canada, *babak@stat.utoronto.ca*; Radford Neal, University of Toronto

***Key Words:*** Bayesian models, hierarchical classification, multinomial logistic regression

We introduce a new method for building classification models when we have prior knowledge of how the classes can be arranged in a hierarchy, based on how easily they can be distinguished. The new method uses a Bayesian form of the multinomial logit (MNL) model, with a prior that introduces correlations between the parameters for classes nearby in the tree. We compare the performance of simulated data on the new method, the ordinary MNL model, and a model that uses the hierarchy in a different way. We also test the new method on page layout analysis and document classification problems and find it performs better than the other methods.

# 16 IT Process Monitoring and Planning ● ✪

Section on Physical and Engineering Sciences

**Sunday, August 6, 2:00 pm–3:50 pm**

### Fourier Domain Estimation for Network Tomography

❖ Jin Cao, Bell Labs, Lucent Technologies, 700 Mountain Ave., Room 2c278, Murray Hill, NJ 07974, *jincao@lucent.com*; Aiyou Chen, Bell Labs, Lucent Technologies; Tian Bu, Bell Labs, Lucent Technologies

***Key Words:*** network tomography, Fourier domain inference, general method of moments, inverse problem, delay tomography, mixture model

Network tomography is a promising methodology for inferring unobservable network behaviors from directly measurable metrics that does not require cooperation between the network internal elements and the end users. In this talk, we will present a novel estimation approach for the network tomography problem based on Fourier domain inference. In addition, we also will obtain identifiability results that apply to general distribution models. We will focus on network delay tomography and develop a Fourier domain inference algorithm based on flexible mixture models of link delays. Through extensive model simulation and simulation using real internet trace, we are able to demonstrate that the new algorithm is computationally more efficient and yields more accurate estimates than previous methods, especially for a network with heterogeneous link delays.

### Robust Estimation for Zero-Inflated Longitudinal Data with Application to IT System Monitoring

❖ Jing Shen, University of Georgia/IBM, 164 Cole Manor Drive, Athens, GA 30606, *jingshen@stat.uga.edu*; Daniel Hall, University of Georgia

***Key Words:*** EM algorithm, m-estimator, Mallows class, observation downweighting, robust GEE

Zero-inflated models comprise an important subclass of finite mixture models useful for data containing many zeros. Recently, several researchers extended these models to the longitudinal data case. Such a problem can arise (e.g., in IT system monitoring) where occurrences of error-related events in some time intervals are tracked. An error event may not happen often, inflating the number of zero counts. But, in rare occasions, an error can occur a many times in an interval, producing a potential outlier. The existing methods for such models provide no particular protection against the presence of outliers, which can influence parameter estimates and inferences substantially. We present a robust estimation approach designed to downweight influential observations. Properties of the proposed method and its comparison to others are discussed.

### Some Statistical Problems in Capacity Management and Planning for On-Demand Computing Services

❖ Ta-Hsin Li, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598-0218, *thl@us.ibm.com*

***Key Words:*** risk, time series, quantile, optimization

Efficient capacity management and adequate capacity planning are vital to the business of on-demand computing services. We will discuss several statistical problems in these applications, including capacity allocation and risk management using a quantile-based method and capacity demand forecasting based on historical time series with different time resolution and time span.

### Modeling Multivariate Time Series with Application to Software Defects Data

❖ Mihaela Serban, Carnegie Mellon University, 5000 Forbes Ave., Baker Hall 132, Pittsburgh, PA 15213, *mserban@stat.cmu.edu*; Wanli Min, IBM T. J. Watson Research Center

***Key Words:*** EM, HMM, Poisson, multinomial

We develop a Poisson hidden Markov model for categorized data with application to software defects data. The numbers of defects are modeled as a state-dependent Poisson process and, conditioning on the total number of defects we specify, a state-dependent multinomial distribution to characterize the defects by type. We fit two-regimes and three-regimes HMMs using an EM algorithm. Model selection is carried out using BIC for in-sample comparison and mean relative error for out-of-sample comparison. We also compare our model with the NHPP model for categorized software data introduced by Liu, Z. et al. (2005).

### Improving Service Delivery Process

❖ Wen-Hua Ju, Avaya Labs Research, 233 Mt. Airy Road, RM 2D28, Basking Ridge, NJ 07920, *whju@avaya.com*; Lorraine Denby, Avaya Labs Research; James M. Landwehr, Avaya Labs Research

A primary goal is to understand the actual flow of multiple processes to resolve problems in a large, complicated service environment from detailed event log data. Based on these flows and their comparisons with the intended---or ideal---process maps, we measure appropriate process components and use the results to monitor and help improve the overall service delivery process. We developed specialized algorithms to process data extracted from the underlying databases and turned the results into useful, relevant information. We also developed many graphical tools to display the information at various levels of granularity appropriate for audiences interested in different aspects of the multiple processes.

## 17 Statistical and Quantitative Literacy 2006 ●

Section on Statistical Education, Section on Teaching Statistics in the Health Sciences

**Sunday, August 6, 2:00 pm–3:50 pm**

### Increasing Quantitative Literacy through the Mathematics Across the Community College Curriculum Project

❖ Rebecca Hartzler, Seattle Central Community College, 1760 N. Northlake Way, 122, Seattle, WA 98103, *rhartzler@sccd.ctc.edu*; Kim Rheinlander, Dartmouth College

The Mathematics Across the Community College Curriculum (MAC3) project is a national dissemination project capitalizing on the work of MAC in Washington and MATC at Dartmouth College. All three efforts are, or have been, funded by the National Science Foundation and support curriculum development and online publication for the integration of mathematics into the undergraduate curriculum. To date, more than 28 disciplines and 20 schools have been involved in the projects. Published curriculum representing the many forms of quantitative literacy strengthening the ability of students to think numerically will be shared, including statistical literacy, number sense, graphical literacy, understanding reasonable quantitative claims, and rates of change. In addition, opportunities for creating or publishing curricula will be presented.

## Quirks of Rhetoric: a Quantitative Analysis of Quantitative Reasoning in Student Writing

❖ Neil Lutsky, Carleton College, 1 North College Street, Northfield, MN 55057, *nlutsky@carleton.edu*; Scott Bierman, Carleton College

***Key Words:*** quantitative reasoning, education, undergraduate, statistics

We analyzed quantitative reasoning (QR) in a random sample of 200 student papers submitted to meet a college writing portfolio requirement. Each paper was coded for the potential relevance of QR; the degree to which QR was employed; the quality of the implementation, communication, and interpretation of QR shown; and other content features (e.g., presence of graphics, uses of terminology). Findings suggested QR was judged centrally relevant to 36% of all papers (primarily lab reports) and peripherally relevant to another 28%. However, QR was used electively to provide peripheral details in only 12% of all potential instances. Instead, students either failed to cite quantitative information or relied on vague quasi-quantitative terms (e.g., many). We identify implications of these results for efforts to address QR across the undergraduate curriculum.

## Common Misconceptions in Statistical Literacy

❖ Marc Isaacson, Augsburg College, 2211 Riverside Ave., Minneapolis, MN 55454, *isaacson@augsburg.edu*

Based on the experiences of teaching an online statistical literacy course at Capella University, this paper will explore the common issues and concepts learners have encountered. While some of these are common to traditional introductory statistics courses, statistical literacy online has unique challenges that must be recognized in order to successfully teach students the skills necessary to become literate in evaluating arguments involving statistics.

## Statistical Literacy: Graphs, Studies, and Related Confounders

❖ Milo Schield, Augsburg College, 1767 Alameda, Roseville, MN 55113, *milo@pro-ns.net*

***Key Words:*** statistical education, observational studies

Graphs commonly are used to communicate information, but graphs of ratios (average, percentages, and rates) can be difficult to read, even when they are presented accurately. Examples are presented. The prevalence of various graphs is reviewed with their strengths and weaknesses. The use of time-based graphs to summarize longitudinal studies is analyzed for both cohort and entry-exit studies. The use of cross-sectional graphs of ratio data to display associations between groups is reviewed (e.g., groups that have more X tend to have more Y). The use of such group-level associations to support cross-level inference conclusions about individuals is reviewed (e.g., Xs are more likely to have Y than are non-Xs: the ecological fallacy). Finally, the ability of longitudinal and cross-sectional studies to resist the influence of various types of confounders is analyzed.

## Pedagogical Challenges of Quantitative Literacy

❖ Bernard Madison, University of Arkansas, Math Sciences SCEN 301, University of Arkansas, Fayetteville, AR 72701, *bmadison@uark.edu*

***Key Words:*** numeracy, quantitative literacy

This paper is based on two years' work in developing and delivering an ever-fresh, real-world--based course that starts students down a path toward quantitative literacy (QL). Numerous pedagogical challenges have been encountered, but none more significant than the habits students have acquired from traditional courses in mathematics and statistics. Discussions of articles from current newspapers and magazines keep these courses fresh but offer significant challenges for the instructors, both in class and assessment. I will discuss a list of characteristics that seem necessary for QL-friendly course, including freshness, absence of formal algorithms, venues for continued practice, and emphasis on number sense.

# 18 SAMSI Program on National Defense and Homeland Security: 2005--2006 ✪

Section on Statisticians in Defense and National Security

**Sunday, August 6, 2:00 pm–3:50 pm**

## A Study of Data Swapping for Categorical Variables

❖ Lisa R. Denogean, SAMSI, 10300 Grovechase Lane, Apt. 307, Raleigh, NC 27617, *ldenogean@samsi.info*

***Key Words:*** data swapping, Markov chains

Data swapping can be used by government agencies to protect the confidentiality of publicly released data files. We study the stochastic process generated by data swapping applied to a data file of categorical variables to understand the effect of swapping and to help the original data owners determine which variables to swap and how much to swap. We will discuss utility measures and introduce the idea of measuring distance from the limit, rather than from the original file. In addition, we will introduce a new type of swapping that we propose is superior to current methods.

## Anomaly Detection

❖ Francisco Vera, National Institute of Statistical Sciences, 19 TW Alexander Drive, Research Triangle Park, NC 27709, *veraf1978@gmail.com*

***Key Words:*** anomaly detection, outliers, sindromic surveillance, clustering, terrorist prediction

An important national defense issue is the ability to detect terrorist attacks as soon as possible. When an attack occurs, many numerical indicators show a behavior different from before. A research group was formed at the Statistical and Applied Mathematical Sciences Institute (SAMSI) to do research on how to detect anomalies from data when this occur. During the weekly activities of our group, we have presented previous work and discussed improvements and the creation of new statistical models for anomaly detection. This talk will be a synthesis of this work.

## New Measures of Data Utility

❖ Mi-Ja Woo, National Institute of Statistical Sciences, 19 TW Alexander Drive, RTP, NC 27709, *mjwoo@niss.org*

***Key Words:*** confidentiality, statistical disclosure limitation, utility, cumulative distribute, clustering, propensity score

Data producers should collect high-quality data and protect confidentiality when releasing data. Protection of confidentiality necessitates altering values of identity information and sensitive attributes. Therefore, it is important to find the data alteration method with high confidentiality that provides satisfactory data quality. This paper focuses on developing methods of measuring data quality when the distribution of data is not assumed to be normal. We treat data utility as a problem of evaluating similarities of original data structure to masked data structure. The data utilities we present are based on the cumulative distribute function, clustering, and propensity score approaches. When the distribution departs from normal, simulations for a variety of data structures show how these measures can be used for evaluating disclosure limitation procedures.

### Agent-Based Methods for Dynamic Social Networks

❖ Eric Vance, Duke University, 114 Old Chemistry, ISDS, Box 90251, Durham, NC 27705, *ervance@stat.duke.edu*; David Banks, Duke University

***Key Words:*** agent-based models, social networks, social space

Agent-based models of social network interactions allow simple encoding of complex social expectations and conventions. For static networks, traditional mathematical formulations (e.g., the p1 model by Holland and Leinhardt (1981) and its generalizations) can be tuned to model almost any multiway relationship, using the clique-decomposition of Frank and Strauss (1986). But this is cumbersome and the formalism does not extend in natural ways to time-varying networks. The notion of actors in a social network occupying positions in a latent social space was introduced by Hoff, Raftery, and Handcock (2002). Our approach uses agent-based models with rules constructed to correspond to terms in a p1 model and rules determining how agents move dynamically in social space. Our talk will illustrate attractive features of agent-based models in such situations.

## 19 Estimation Techniques for Diagnostics Devices ● ✪

Section on Statistics in Epidemiology, Biometrics Section, Section on Physical and Engineering Sciences

**Sunday, August 6, 2:00 pm–3:50 pm**

### Evaluation of a Noninvasive Diagnostic Device Using Weighted Least Squares Approach

❖ Zhen Huang, Duke Clinical Research Institute, P.O. Box 17969, Durham, NC 27715, *zhen.huang@duke.edu*; Huiman Barnhart, Duke University

***Key Words:*** agreement, correlated kappa, correlated sensitivity/specificity, weighted least squares

In CINEMA study, the performance of a noninvasive multi-row detector subsecond computed tomography (MRD-SUCT) is evaluated by comparison with the commonly accepted invasive diagnostic method coronary catheterization. The kappa coefficient and sensitivity and specificity are used as measures of agreement and reliability. Because 28 coronary segments per patient are available for evaluation, there are 28 estimated kappa coefficients, as well as subunit sensitivities and specificities. We use the weighted least squares approach to take into account the correlation between these estimated statistics and share our experience in testing the equality of these parameters and related issues.

### System Accuracy Requirements for Blood Glucose Monitors

❖ Nancy Schatz, Home Diagnostics, Inc., 2400 NW 55th Court, Fort Lauderdale, FL 33309, *nschatz@hdidiabetes.com*

***Key Words:*** medical devices, blood glucose monitors, clinical trials, system accuracy, ISO15197 compliance, Parkes error grid

With the harmonized European directive, ISO EN 15197:2003, manufacturers of blood glucose monitors interested in obtaining CE marking for their products are revising their approach to clinical trials. Compliance to ISO15197 requires a different emphasis on system accuracy performance, which changes existing protocol procedures and analysis. The biggest challenge occurs with achieving the specified sample distribution in the clinical setting. The analysis is modified with specific acceptance criteria in a precise format. The talk will touch on our experience of incorporating this new directive into our clinical studies.

### Bayesian Predictive Probability as a Diagnostic Assessment of the Likelihood of Coronary Artery Disease in Collateral Arteries

❖ Laura Thompson, U.S. Food and Drug Administration, FDA/CDRH/OSB/DBS, HFZ-550, 1350 Piccard Dr., 1st Floor, Rockville, MD 20850, *laura.thompson@fda.hhs.gov*

***Key Words:*** Bayesian, gender differences, angiography, nonparametric, coronary artery disease

We assume a binary tree model for the human coronary artery tree and apply a reinforced random walk model to model movement of lipid particles through the artery tree. Data might consist of a set of angiographic images---with one or more stenoses---morphometric measurements of the vessels, and biological characteristics of the patients from whom the images were obtained. We apply nonparametric Bayesian methods to obtain the predictive probability of plaque, given the presence of plaque in a collateral region of the vessel that has already stenosed. This predictive probability can be used as a diagnostic tool to identify risk of recurrent coronary artery disease (CAD). We also explore the special case of female CAD.

### Disease Diagnosis Maximizing Effectiveness and Minimizing Cost of Health Care

❖ Jeffrey Vaks, Beckman Coulter, Inc., 3 Crivelli Aisle, Irvine, CA 92606, *jevaks@sbcglobal.net*

***Key Words:*** clinical lab test results, stratification, classification, false negatives/positives, minimum total cost

Most existing disease markers do not allow for full separation of healthy and diseased populations. Advanced statistical methods allow for improved diagnosis of disease compared to methods in use. Improvements include use of clinical laboratory test results for both healthy and diseased individuals---with relevant stratification of the population by age, sex, demographics, etc.---for development of classifiers and use of the minimum total cost of false positives and false negatives criteria for optimizing classifier parameters. Analytical process in clinical laboratories needs to be in a long term state of statistical control, providing for consistent patient test results across laboratories and time. Quadratic discriminant, canonical discriminant analysis, and logistic regression were used successfully to develop a classifier for diagnosis of prostate cancer.

### Use of Frequency Domain Measures Instead of Traditional Summary Statistics for Use in Diagnostic Devices

❖ Roseann White, Guidant Corporation, 3200 Lakeside Drive, Santa Clara, CA 95054, *rowhite@guidant.com*

***Key Words:*** FFT, spatial frequency, ROC

Several noninvasive and invasive imaging techniques are being used to assess the progression of vascular disease, particularly whether areas of potential plaque rupture will occur. Traditional analyses have looked at summary statistics in terms of sums, means, medians, maximum, etc. However, it may be that the spatial distribution of key characteristics has better diagnostic value. Therefore, it would be worthwhile to evaluate the data in the frequency domain. Using a simulated dataset based on current available data, the two methodologies will be compared for measures such as total plaque burden versus a particular spatial frequency band and amplitude using ROC curves and associated statistics.

## 20 Types of Modes and Effects on Response Rates and Performance

Social Statistics Section

**Sunday, August 6, 2:00 pm–3:50 pm**

### A Comparison of in-Class and Online Student Evaluations

❖ David Swanson, University of Mississippi, University, MS 38677/1848, *dswanson@olemiss.edu*

***Key Words:*** nonrandom error, response bias, reliability, validity, university

A comparative analysis of results from simultaneous in-class and online student evaluations taught in a social science department at a large public university strongly suggests online evaluations are neither valid nor reliable because of nonrandom error resulting from high levels of nonresponse. As a result of this study, the department decided not to use online evaluations in annual reviews of its faculty until a subsequent and more detailed analysis is found to contradict this finding or average response rates for online evaluations (about 26%) achieve those for in-class evaluations (about 72%).

### A Repeated Measures Design To Investigate Mode Effects in the Center for Epidemiologic Studies Depression Scale

❖ Richard Swartz, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 237, Houston, TX 77030, *rswartz@mdanderson.org*; Carl de Moor, Harvard Medical School; Karon Cook, University of Washington; Rachel T. Fouladi, Simon Fraser University; Karen Basen-Engquist, M. D. Anderson Cancer Center; Cathy Eng, M. D. Anderson Cancer Center

***Key Words:*** mode effects, item response theory, differential functioning of items and tests

Changing a mode of administration for a questionnaire might affect people's responses (a mode-effect). This study investigates mode-effects from administering the Center for Epidemiologic Studies Depression (CES-D) scale by a personal digital assistant (PDA) versus a Paper and Pencil (P&P) administration. Participants took both versions and were randomly assigned to complete the PDA or the P&P version first. An item response theory method identified mode-effects on both the test overall and the individual items. A mixed-effects regression model summarized the mode-effects using CES-D scores. Scores were on average higher on the first administration of the P&P test than all other administrations (PDA second, PDA first, and P&P second). 18 of 20 items exhibited a statistically significant mode effect when the P&P format came first. Also mode-by-age and mode-by-education effects were found.

### Mode of Data Collection and the Foreign Born in the American Community Survey

❖ Alexa Kennedy-Puthoff, U.S. Census Bureau, 9726 53rd Ave., College Park, MD 20740, *akputhoff@gmail.com*

***Key Words:*** foreign born, American Community Survey, data collection

The American Community Survey (ACS) is a primary source for data on the foreign-born population in the U.S. ACS data are collected via three modes: mail (self-enumeration by mail-back forms), telephone (Computer Assisted Telephone Interview: CATI), and personal visit (Computer Assisted Personal Interview: CAPI). Response rates for each mode vary among groups. This paper focuses on the mode of data collection of responses for native compared with foreign-born populations, and naturalized compared with non-citizen respondents. A large proportion of responses from foreign-born respondents are gathered via CATI and CAPI, and mode patterns are more similar between naturalized citizens and natives than between naturalized citizens and noncitizens. Allocation rates and weighting procedures differ among the modes, suggesting that the distribution of the foreign-born among the modes may be of importance.

### Making Item Selection More Efficient in Computerized Adaptive Testing

❖ Hua-Hua Chang, University of Illinois, 2109 Vale Street, Champaign, IL 61822, *hhchang@cyrus.psych.uiuc.edu*; Zhiliang Ying, Columbia University

***Key Words:*** computerized adaptive testing, sequential design, a-stratified design, item selection, test security, item pool usage

Computerized adaptive testing (CAT) has become a popular mode of educational assessment. The most important component in CAT is the item selection procedure for selecting items during the course of the test. According to Lord (1970), an examinee is measured most effectively when test items are neither too difficult nor too easy. Several sequential item selection methods have been proposed, including a version modified from the Robbins-Monro process, Lord's maximum Fisher information method, and Owen's Bayesian method. Recently, Chang and Ying proposed the a-stratified method with an objective to limit the exposure on any given item by using that item at the most advantageous point in testing. One of the advantages to use the a-stratified method is that it attempts to equalize the item exposure rates for all the items in the pool. Various statistical issues will be addressed in this paper

### Infant Malnutrition in the Cape Coast District: a Ghanaian Case

❖ Amakye Agyekum, University of Cape Coast, Math and Statistics Department, Cape Coast, 233100000 Ghana, *daak2g1@yahoo.com*

***Key Words:*** infant, malnutrition, mortality, kwashiorkor, feeding, anaemia

This study is about infant malnutrition and infant mortality in Ghana and uses the Cape Coast District as the main case study area. The study seeks to find the percentage of infant deaths caused by malnutrition and determine the rate of these deaths from 1999 to 2003. The study also looks into the factors that cause malnutrition and suggests ways to reduce the pandemic drastically. Both secondary and primary data were used in the study. Data on infant malnutrition and infant mortality were collected from the Cape Coast District Hospital from 1999 to 2003. The primary data were collected by interviews and direct observation from Apewosika and Amamoma and descriptive statistics such as bar charts and pie charts were used to display the nature of the data looks. It was observed that infant mortality is on the increase and that infant malnutrition contributed the majority of the deaths.

## 21 Functional Data Analysis, Supervised Learning, and Dimension Reduction

Biometrics Section, Section on Nonparametric Statistics, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### Functional Regression Analysis for Longitudinal Data with a Large Number of Repeated Measures

❖ Xiaowei Yang, University of California, Davis, Med Sci 1-C, Davis, CA 95616, *xdyang@ucdavis.edu*; Hongquan Xu, University of California, Los Angeles; Qing Shen, Edmunds.com

***Key Words:*** functional f test, functional data analysis, functional regression analysis, longitudinal data analysis

Longitudinal datasets in biomedical research often consist of variables measured repeatedly on each subject, yielding a large number of observations. This characteristic complicates the use of standard longitudinal modeling strategies where rigorous assumptions on intra-subject correlation structure are required. An innovative way to model the data is to apply functional regression analysis in which observations of the same subject are viewed as a sample from a functional space. Shen and Faraway (2004) introduced an F test for linear models with functional responses. This talk illustrates how to apply this method to the setting of longitudinal data. A smoking cessation study for methadone-maintained tobacco smokers will be analyzed for demonstration.

### Self-Modeling Regression with Application to Arterial Pulse Pressure Waveforms

❖ Lyndia Brumback, University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, *lynb@u.washington.edu*; Doug Tommet, University of Washington; Richard Kronmal, University of Washington

***Key Words:*** functional data, shape invariant model, nonlinear mixed effects model

Self modeling regression, a method for functional data, is based on the simple assumption that the x and y axes can be separately shifted and scaled for each curve so that the data from all curves lie approximately on one typical curve. We propose to model the typical shape with a regression spline, and to treat the shift and scale parameters as random with a restricted covariance structure to ensure computational identifiability. The resulting nonlinear mixed effects model involves random effects that are correlated among experimental units, unlike random effects in usual mixed effects models, but the model can still be implemented using standard software. We describe our approach and apply it to pulse waveforms (where variability of the amplitude shift parameter represents short-term variability in blood pressure) from a large observational study of cardiovascular disease.

### Classification and Gene selection of Cancer Microarrays by nu-Ridge Regression

❖ Jun Luo, Michigan State University, A443 Wells Hall, Michigan State University, East Lansing, MI 48824, *luojun@msu.edu*

***Key Words:*** micro-array, cancer diagnosis, feature selection, support vector machine, ridge regression, penalized logistic regression

Classification of micro-array is an important aspect of cancer diagnosis and treatment. The support vector machine (SVM) and penalized logistic regression (PLR) have been successfully applied to micro-array problems. In this paper, we propose nu-ridge regression (nu-RR) as an alternative method to the SVM and PLR for the micro-array classification. Due to the goal in micro-array cancer diagnosis is mostly to identify the most responsible genes, rather than classification alone, we propose a new ranking method MR besides applying UR and RFE for gene selection. Our simulation on the real dataset, Leukemia diagnosis data, indicates nu-RR combined with either UR, MR or RFE tends to select less significant genes than other methods. Meanwhile, nu-RR performs superior to SVM and PLR with a lower rate in both cv-error and test error.

### Boosting with Missing Predictors

❖ Ching-Yun Wang, Fred Hutchinson Cancer Research Center, P.O. Box 19024, 1100 Fairview Ave N, Seattle, WA 98109, *cywang@fhcrc.org*; Ziding Feng, Fred Hutchinson Cancer Research Center

***Key Words:*** additive model, classification, imputation, incomplete data, nonmonotone missing, pancreatic cancer

Boosting is an important tool in classification methodology. It combines the performance of many weak classifiers to produce a powerful committee, and the validity of it can be explained by additive modeling and maximum likelihood. The method has very general applications, especially to high dimensional predictors. For example, it can be applied to distinguish cancer samples from healthy control samples by using antibody microarray data. Microarray data are often high dimensional, and many of them are incomplete. One natural idea is to impute a missing variable based on observed predictors. The problem itself becomes more challenging when the missing data pattern is not monotone. In this paper, we propose an imputation method based on an iterative algorithm. This method can be applied to the situation when a complete-case subset does not even exist.

### Prediction Based on Two-Stage Modeling

Amita K. Manatunga, Emory University; ❖ Jose N. G. Binongo, Emory University, 1518 Clifton Road, Room 330, Rollins School of Public Health, Atlanta, GA 30322, *jbinong@emory.edu*; Ming Yuan, Georgia Institute of Technology

***Key Words:*** prediction, modeling, logistic regression

One of the problems in nuclear medicine is recognizing obstruction in the kidney. The first stage is to determine whether a diuretic is needed. If it is not needed, the kidney is not obstructed. If it is needed, the second stage is to determine whether the kidney is obstructed. The goal is to identify covariates that predict kidney obstruction. A naive approach is to develop a two-stage model based on ROC curve methodology. In this presentation, we present an alternative method, and we compare this method with the naive approach.

### Steps Toward Individualized Treatment: a Double Supervised Machine-Leaning Method

❖ Steven Y. Cen, University of Southern California, 1000 S. Fremont Ave., Unit # 8 (post), Building A-5, Alhambra, CA 91803, *cen@usc.edu*; Catherine Sugar, University of Southern California; Bryan Langholz, Keck School of Medicine of USC; David Conti, University of Southern California; Doug Stahl, City of Hope National Medical Center; Stanley P. Azen, University of Southern California

***Key Words:*** machine learning, data mining, individualized treatment, CART, interaction, dimension reduction

A double supervised machine learning method is proposed for selecting the best option for a patient among various treatments, taking into account multiple domains i.e. demographic characteristics, disease prognostic factors and genetic factors. This new method, Modified Homogeneity Score Searching Method (MHS-SM) combines a homogeneity score derived from Breslow-Day's homogeneity test, and a searching strategy adopted from Classification and Regression Trees (CART). The result from simulation study showed that compared to CART, the MHS-SM is more adept at detecting a simulated treatment effect modifier in the presence of marginal effects or independent confounding main effects. The comparison was also made using data from a large-scale clinical trial in acute lymphoblastic leukemia. The MHS-SM was able to detect treatment effect modifiers from this complex dataset while CART was not.

### On Reducing Multiple Outcomes into a Single Score

❖ Hui Xie, Boston University, 812 Memorial Drive, Apt 804, Cambridge, MA 02139, *hx30@bu.edu*

***Key Words:*** outcome definition, clinical trial, multiple outcome, data reduction

In clinical studies of complex diseases, the quantity of interest is often measured by multiple outcomes. For example, in rheumatic arthritis (RA), the improvement of disease status after treatments is measured and can be compared between treatments in terms of the changes in 7 core-set measurements. There is a need to summarize these multiple outcomes into a single score that optimally quantify the underlying subject-specific level of improvement. We designed a variety of data reduction schemes that lead to different scoring definitions for improvement. The commonly used scoring method, weighted average of the multiple outcomes, is used as a benchmark for comparison. We tested these scoring definitions in recently collected RA clinical trials. A continuous score that requires consistency of improvement across all measures is shown to be robust to outliers and outperform the benchmark.

## 22 Regression for Censored Data

●

Biometrics Section, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### Regression Analysis for Long Term Survival Rate

❖ Yichuan Zhao, Georgia State University, Department of Mathematics and Statistics, 726 COE, 7th floor, 30 Pryor Street, Atlanta, GA 30303, *matyiz@cantor.mathstat.gsu.edu*

***Key Words:*** confidence region, link function, right censoring, Kaplan-Meier estimator

In recent years, regression models have been shown to be useful for predicting the long term survival probabilities of patients in clinical trials. For inference on the vector of regression parameters, there are semiparametric procedures based on normal approximations. However, the accuracy of such procedures in terms of coverage probability can be low when the censoring rate is heavy. In this paper, we apply an empirical likelihood ratio (ELR) method to the regression model and derive the limiting distribution of the ELR. Based on the result, we develop a confidence region for the vector of regression parameters. Furthermore, we use a simulation study to compare the proposed method with the normal approximation-based method proposed by Jung (1996). Finally, the proposed procedure is illustrated with data from a clinical trial.

### Accelerated Failure Time Model with Random Effects

❖ Yaqin Wang, Iowa State University, 46 Schilletter Village, Apt. B, Statistics Dept., Ames, IA 50010, *wang_yaqin@hotmail.com*; Kenneth Koehler, Iowa State University

***Key Words:*** AFT model, correlated survival data, random effects, Laplace approximation

There is an increasing interest in incorporating multivariate frailties into the analysis of correlated survival data. We propose an accelerated failure time (AFT) model with a multivariate lognormal distributed random effect. It allows for random effects with a complicated dependence structure that may be a function of unknown covariance parameters. This model can be applied to right, left or interval-censored survival data. An estimation procedure is developed for the proposed AFT model that is based on the Laplace approximation to the marginal likelihood. The performance of this approximation is evaluated through several simulation studies.

### Penalized Weighted Least Squares Method for Accelerated Failure Time Models with Gene Expression Data

❖ Simin Hu, Case Western Reserve University, 2235 Overlook Road, Apt 310, Cleveland Heights, OH 44106, *sxh96@case.edu*; J. S. Rao, Case Western Reserve University

***Key Words:*** accelerated failure time model, right censoring, weighted least squares method, regularizations, variable selection, gene expression data

A special feature of survival data is that survival times are frequently censored. Stute's weighted least squares method for accelerated failure time models uses weights derived from the Kaplan-Meier product-limit estimator. In this paper, we propose a new regularized version of Stute's method for model estimation and variable selection. We use censoring as a constraint while solving the weighted least squares objective function. Since our method aims especially to select genes with high dimensional gene expression data, we also include L-1 and L-2 penalties to achieve model parsimony and select important covariates. Simulation studies and real data examples show our method is very effective in variable selection and provides higher prediction accuracy when compared with other available methods.

### On Linear Regression under the Partial Koziol-Green Model of Random Censorship

❖ Ke Wu, California State University, Fresno, Department of Mathematics, 5245 N Backer Ave Mail Stop PB108, Fresno, CA 93740, *kewu@csufresno.edu*

***Key Words:*** linear regression, partial Koziol-Green model, partial ACL estimator, random censorship, Kaplan-Meier estimator

In survival analysis, a model of informative censoring is the Koziol-Green (KG) model, where the survival function of the censoring times is some power of the survival function of the lifetimes. A generalization of the KG model is the partial Koziol-Green (PKG) model which allows the lifetimes to be censored by two types of variables, one of which censors in an informative way and the other one in a non-informative way. In this paper we give a new method of estimating parameters in the linear regression model under the PKG model. The partial ACL estimator for the survival function under the PKG model is used in the proposed method, rather than the Kaplan-Meier (1958) estimator used for the general random censoring model in the Buckley and James's (1979) procedure for estimating parameters in linear regression with censored data. Simulations comparing the two methods are described.

### Inference for Interval-Censored Data with Different Censoring Patterns among Treatment Groups

❖ Guozhi Gao, Amgen Inc., 1 Amgen Center Drive, mail stop 24 2 C, Thousand Oaks, CA 91320, *ggao@amgen.com*; Xiang Zhang, Amgen Inc.; Steven Snapinn, Amgen Inc.; Qi Jiang, Amgen Inc.

***Key Words:*** interval censored data, censoring patterns, proportional hazards model, survival analysis, non-informative censoring

While methods exist for analyzing general interval-censored data, the impact of censoring pattern has not been carefully studied. A naÔve approach to the analysis of interval-censored data is to use a standard Cox proportional hazards model, where an event is treated to occur at the beginning, the middle, or the end of the interval it belongs to. Preliminary research results have shown that this naÔve approach performs well only if the censoring patterns are the same among treatment groups. In this presentation, we will present a study of some existing methods on interval-censored data with different (non-informative) censoring patterns among treatment groups. In addition, we have evaluated the impact of censoring pattern by modeling the (non-informative) censoring patterns using (semi) parametric models. An application to a real dataset will be given for illustration.

### Additive Hazards Model for Case-2 Interval-Censored Failure Time Data

❖ Lianming Wang, University of Missouri-Columbia, 146 Middlebush Hall, Department of Statistics, Columbia, MO 65211, *lwdzc@mizzou.edu*; Jianguo Sun, University of Missouri-Columbia; Xingwei Tong, University of Missouri-Columbia

***Key Words:*** additive hazards model, counting processes, estimating equation, proportional hazards model, regression analysis

Interval-censored failure time data often arise in clinical trials and medical follow-up studies and a few methods have been proposed for their regression analysis using various regression models (Finkelstein, 1986; Huang, 1996; Lin et al., 1998; Sun, 2005). This paper considers the regression analysis using the additive hazards model, for which it seems that there is no inference approach available except for some special cases. To estimate regression parameters of interest, a conditional inference approach is presented that does not involve estimation of the cumulative baseline hazard function. Asymptotic properties of the proposed parameter estimates are established and some simulation results and an illustrated example are provided.

### Statistical Analysis of Multivariate Failure Time Data with Auxiliary Covariates

❖ Zhaozhi Fan, Memorial University of Newfoundland, Department of Mathematics and Statistics, St Johns, NF A1C 5S7 Canada, *zhaozhi@math.mun.ca*

***Key Words:*** correlated failure time, marginal hazards model, auxiliary covariate, kernel smoothing, partial likelihood

In this paper, we consider marginal hazards model of multivariate failure times with continuous auxiliary covariates. We extend the kernel smoothing procedure of Zhou and Wang (2000) to correlated failure time data. We estimate the marginal partial likelihood function with a semiparametric procedure and the regression parameters are estimated based on the estimated partial likelihood. We present asymptotic properties of the induced estimator and a consistent estimator of the covariance matrix of the estimated regression coefficients. We provide small sample simulation results.

## 23 Normalization and Analysis of Microarrays ●

Biometrics Section, ENAR
**Sunday, August 6, 2:00 pm–3:50 pm**

### Two Extensions of the TW-SLM for Systematically Incorporating Control Genes and Spot Quality Information To Improve Normalization of cDNA Microarray Data

❖ Deli Wang, The University of Alabama at Birmingham, 1824 6th Ave., S., WTI 153, Birmingham, AL 35294, *deli.wang@ccc.uab.edu*; Cun-Hui Zhang, Rutgers University; Marcelo B. Soares, Northwestern University; Jian Huang, The University of Iowa

***Key Words:*** microarray, spike, normalization, quality control, semiparametric model

Normalization and data quality control are basic steps in microarray data analysis. Proper normalization and data quality control ensure the intensity ratios provide meaningful and accurate measurement of relative gene expression values. Side information, including control genes (e.g., spike-in genes), and spot quality information is available in a typical microarray dataset, but it is not utilized fully by many normalization and quality-control methods. We propose two extensions of the two-way semi-linear model (TW-SLM) for appropriately combining control genes and spot quality information to improve normalization of cDNA microarray data. A new quality score also is derived and can be used independently for spot quality assessment. We used two datasets produced by specially designed microarray experiments to demonstrate the proposed methods.

### Using Cytogenetics Data To Guide the Normalization of SNP Microarray Signals

❖ Stanley Pounds, St. Jude Children's Research Hospital, 332 N. Lauderdale Street, Memphis, TN 38105, *stanley.pounds@stjude.org*; Cheng Cheng, St. Jude Children's Research Hospital; Charles Mullighan, St. Jude Children's Research Hospital; Salil Goorha, St. Jude Children's Research Hospital; Sheila Shurtleff, St. Jude Children's Research Hospital; Susana C. Raimondi, St. Jude Children's Research Hospital; James R. Downing, St. Jude Children's Research Hospital

***Key Words:*** microarray, SNP, DNA copy number, cytogenetics, normalization, genomics

SNP microarrays simultaneously query genomic DNA at thousands of positions and are used to identify regions of DNA gain or loss in tumor cells. Correct identification of such regions depends on accurately normalizing the array signals. Readily available software implements a median-centering normalization. However, it implicitly requires that the copy number of much of the genome is unaltered; this is not true for several cancers. In some cases, median-centering infers incorrect copy numbers for the entire genome. Therefore, we propose a method that uses cytogenetics data, which identifies microscopic genomic lesions, to guide the normalization of SNP array signals. In our study of 242 pediatric acute lymphoblastic leukemia cases, the proposed method leads to more accurate copy number inferences than does median-centering.

## Category Analysis for Microarray Data

❖ Zhen Jiang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M2 B876, PO Box 19024, Seattle, WA 98109, *zjiang@fhcrc.org*; Robert Gentleman, Fred Hutchinson Cancer Research Center

***Key Words:*** gene set enrichment analysis, microarray analysis, biostatistics, computation biology

A disease or phenotype exists not because the changes in one or a few genes only. It involves much complex biological process. All the genes that involve in this process would be affected at some level. As a result, we expect the functional related genes to have coordinated and moderate changes. Gene Set Enrichment Analysis (GSEA) method has be developed recently to capture this coordinated changes in the sets of functionally related genes. We propose an extension of it, which combines the gene-to-phenotype association and the gene-to-gene-set association into gene-set-to-phenotype association. We present applications of our method in several examples to illustrate the various realizations of our method by using different measures of either associations. We also address the problem of overlap among gene sets in this paper.

## Probe-Level Modeling and Multiple Testing of Microarray Gene Expression

❖ Tao Wang, University of South Florida, 13201 Bruce B Downs Blvd., Tampa, FL 33612, *twang@hsc.usf.edu*; Magali Mouy, deCODE genetics; Jason Hsu, The Ohio State University; Hakon Hakonarson, deCODE genetics; Kari Stefansson, deCODE genetics

***Key Words:*** microarray, gene expression, modeling, multiple testing

In this talk, I present a novel statistical approach for oligonucleotide microarray data analysis that models the probe level data. Comparisons of our probe level analysis with the popular gene level analysis using R package affy and multtest found our approach more powerful in both detecting differences and establishing equivalence in gene expressions. This talk also addressed some important issues in analyzing gene expression data such as the handling for non-normality and unequal variances of the error distribution in the linear model. An real microarray experiment to study the potential effect of an additional cycle of amplification on gene expression signals were analyzed using both probe and gene level analysis for comparison.

## Application of Temporal Association Rules to a cDNA Microarray Experiment

❖ Bruce Southey, University of Illinois, 1207 W. Gregory Drive, Urbana, IL 61801, *southey@uiuc.edu*; Sandra Rodriguez-Zas, University of Illinois; Younhee Ko, University of Illinois; Chengxiang Zhai, University of Illinois

***Key Words:*** microarray, association rules, data mining, cDNA

The analysis of cDNA microarray experiments provides vast information on the magnitude of gene expression differences between multiple experimental conditions for thousands of genes. Traditional association rules were adapted to describe gene expression patterns over 5 times points in a honey bee microarray experiment. After normalization and analysis using a two-stage linear mixed model, association rules were applied simultaneously to the pairwise differences between time points. The pairwise differences were classified into three groups, significantly under-expressed, significantly over-expressed or not differently expressed. Our approach enabled the discovery of relationships between genes that were not obvious from the pairwise comparisons alone. The application of association rules enhanced the interpretation of results from the statistical analysis of microarray experiments.

## Quantitative Association Rules Applied to the Analysis of cDNA Microarray Experiments

❖ Younhee Ko, University of Illinois, 1004 W. Main Street, Apt 201, Urbana, IL 61801, *unygo@hotmail.com*; Bruce Southey, University of Illinois; Chengxiang Zhai, University of Illinois; Sandra Rodriguez-Zas, University of Illinois

***Key Words:*** microarray, gene expression, association rule

cDNA microarray experiments are used to simultaneously measure the expression of thousands of genes across conditions. Association rules were applied to a honey bee microarray experiment that examined the expression of 9,000 cDNAs at 5 time points. After normalization and analysis using a two-stage linear mixed model, least-square mean estimates at each time point were used to characterize the expression patterns of each cDNA. Quantitative association rules were applied to the least-squares means of the cDNAs with differential expression using a predetermined threshold based on the expected adjusted expression of all genes. Several known relationships between genes were confirmed and novel ones were uncovered. Association rules provided an effective approach to extract meaningful information from microarray studies.

## Strategies for Genome-Wide Family-Based Association Analysis for the Study of Integrative Genomics

❖ James Degnan, Harvard University, 655 Huntington Ave., Dept of Biostatistics, Boston, MA 02115, *jdegnan@hsph.harvard.edu*; Jessica Su, Harvard University; Cliona Molony, Rosetta Inpharmatics LLC; Eric Schadt, Rosetta Inpharmatics LLC/Merck Research Laboratories; Benjamin Raby, Harvard University; Christoph Lange, Harvard School of Public Health

***Key Words:*** multiple testing, family-based association test, genetic association, gene expression

Successful strategies to bypass the multiple testing problem in genome-wide family-based association studies have been developed to identify genetic main effects (Van Steen et al., Nat Genet 2005, 37(7): 683) and have been successfully applied to 100k scans (Herbert et al., Science, 2006, in press). To get a deeper insight into the pathway of complex diseases, it will be crucial to identify gene-gene interactions. We therefore extended the testing strategy by Van Steen et al. to an integrative genomics setting using microarray expression data as univariate phenotypes to identify cis-acting genes. We applied our new analysis strategy to a genome-wide association study with expression data in 20 extended CEPH pedigrees. We compare the results of our new analysis strategy to those of Monks, et al. (Am J Hum Genet 2004, 75(6):1094).

# 24 Pharmacokinetics and Crossover Trials ●

Biopharmaceutical Section, Biometrics Section, ENAR
**Sunday, August 6, 2:00 pm–3:50 pm**

## Assessing PK-AE Relationships Using Nonlinear Models

❖ Haiyuan Zhu, Merck Research Laboratories, 126 E. Lincoln Ave., RY34-A304, Rahway, NJ 07065, *haiyuan_zhu@merck.com*

***Key Words:*** PK/PD modeling, adverse events, odds ratio, plateau

Adverse events (AEs) occur in preclinical and clinical trials. Understanding the relationship between a drug's Pharmacokinetic (PK) profile and AEs is crucial to finding the maximum tolerated dose and clinical dose in drug development. One question often asked during dose finding is whether a plateau in the incidence of AEs has been reached at a dose or plasma concentration level. However, commonly used logistic regression models are not capable of addressing this question. For this reason, nonlinear models are proposed to model the PK/AE relationships. Log odds ratio, which is a function of plasma concentration in these nonlinear models, is studied and a testing procedure for plateau hypothesis is constructed based on simultaneous confidence band for the log odds ratio curve. An example is used to illustrate the model fitting and testing procedure.

## Applications of Adapted Crossover Designs To Reduce Study Cost and Length in Phase I Clinical Trials

❖ Fang Liu, Merck Research Laboratories, RY34 A304, PO BOX 2000, Rahway, NJ 07065, *fang_liu@merck.com*

***Key Words:*** crossover design, incomplete block design, reduction in study length and cost, phase I trials

Time and cost are important practical factors in designing phase I trials. Complete and balanced incomplete block crossover designs are able to reduce variance for treatment comparisons and save on sample size. Both designs are well-established, with desirable statistical properties. At times, practical consideration may call for shorter study length and further reduction in study cost. To find common ground between the requirements of empirical sciences and statistical principles, some crossover designs can be adapted without too much sacrifice of their statistical properties (i.e., balancedness). In this talk, adaptations that have been used in practice will be presented. Statistical considerations in power/sample size calculation and data analysis also will be discussed.

## Analysis of Replicated Crossover Designs for Average Bioequivalence

❖ Donna Kowalski, Astellas Pharma Inc., Three Parkway N., Deerfield, IL 60015, *donna.kowalski@us.astellas.com*; Devan V. Mehrotra, Merck Research Laboratories

***Key Words:*** compound symmetry, replicated crossover, empirical option, sandwich estimator, unstructured covariance

Replicated crossover designs (e.g., TRR, RTT) are sometimes used to demonstrate average bioequivalence of a test (T) and reference (R) treatment. Pharmacokinetic responses from such designs (e.g., log(AUC)) are commonly analyzed using a linear mixed effects modeling approach described in an FDA guidance document on bioequivalence analyses. For the "FDA approach", in PROC MIXED terminology, the covariance matrix for each subject (= ZGZ'+R) is modeled using both a RANDOM statement, with TYPE = FA0(2) for G, and a REPEATED statement, with TYPE = VC for R. A simpler approach is to only use a REPEATED statement, with TYPE = CS (with or without the EMPIRICAL option) or UN for R. The type I error rate and power properties of the different covariance structure-modeling approaches will be compared using simulations.

## Assessing Treatment Differences Adjusted by Possible Carryover Effects in Crossover Clinical Trials

❖ Ling Chen, U.S. Food and Drug Administration, 10903 New Hampshire Ave., WO building 22 RM 5241, Silver Spring, MD 20993, *ling.chen@fda.hhs.gov*

***Key Words:*** crossover design, carryover effects, multiple treatments, Williams design

In crossover clinical trials the study model often includes the first-order carryover as a fixed effect. Using conventional methodology we are not able to accurately detect treatment differences if carryover effects exist. A test for no first-order carryover effect is usually performed before testing for the treatment difference. The test is likely to have low power to detect a significant difference. In addition, Freeman (1989) showed that this two-stage procedure of testing for a carryover difference in the first stage and then for a direct treatment difference in the second stage not only inflated the probability of making a type I error, but also produced a biased estimate of the direct treatment difference. In this article, I propose a new test for treatment differences adjusted by possible first-order carryover effects in crossover designed studies for three or more treatments.

## Factorial Crossover Designs with Fewer Periods and Fewer Subjects

❖ Sourav Santra, Northern Illinois University, Division of Statistics, DU 366, DeKalb, IL 60115, *santra@math.niu.edu*

***Key Words:*** crossover design, factorial experiment, confounding

Crossover designs are commonly used in drug development. Two-period designs are the most popular because of their relative ease of implementation in practice. Factorial experiments require designs with larger number of periods. Such designs are useful as long as the number of treatment combinations, and hence the number of periods is within manageable limits from practical point of view. Another problem is that factorial crossover designs require a larger number of subjects as well. Russell and Dean (1998) constructed crossover designs with fewer subjects for two-factor experiments. Factorial crossover designs with fewer periods and fewer subjects are provided in this paper. Furthermore, designs are not restricted to two-factor experiments. An R-algorithm for constructing designs is available from the author upon request.

## Locally D-Optimal Designs for Pharmacokinetics Compartmental Models

❖ Xin Fang, University of Illinois at Chicago, 702 S. Lytle Street, 1 SOUTH, CHICAGO, IL 60607, *xinfang@math.uic.edu*

***Key Words:*** d-optimal, pharmacokinetics, compartmental models

Locally D-optimal designs are investigated for an open one-compartment model with zero order input and first order elimination and a general compartmental model with 2n parameters. These nonlinear models have been used and studied in pharmaceutical industries and other research organizations. We allow an additive error which follows a Normal distribution. The LD design for the two-parameter one compartmental model are found to be saturated with the explicit two design points. An upper bound of the number of points in the support of the LD design for the general compartmental model with 2n parameters is updated. A sufficient condition for the inclusion of the sampling time 0 to be in the LD design support of the general compartmental model is provided. Robust designs are investigated. Simulation results suggest that LD designs are relatively robust designs in regard to the nominal values.

### The Hypothesis Testing behind Steady State Determination in Clinical Pharmacology Trials

❖ Bingming Yi, Merck & Co., Inc., 101 Redwood Ave., Edison, NJ 08817, *bingming_yi@merck.com*; Xun Chen, sanofi-aventis; Patrick Larson, Merck & Co., Inc.

***Key Words:*** steady state, multiple dose, equilibrium, stepwise hypothesis testing, slope test, contrast

Multiple dose studies usually are conducted to assess the drug safety and pk profile after repeated dosing. When the drug concentration or area under concentration reaches and maintains an almost-constant level, the amount of drug in the body is said to achieve steady state. In the pharmaceutical industry, regression slope tests, suggested by FDA, or contrasts in linear models often are conducted in a stepwise fashion to find the steady state point. However, the hypothesis behind the problem of steady state determination was not well-defined for these tests. This research discusses the rationale to change the traditional hypothesis and introduces the comparability bounds. Novel testing procedures are proposed and compared to the traditional methods based on simulated datasets. It is found that common methods should be used with caution, and the best methods are discussed.

## 25 Multiple Trials and Multiple Endpoints ●

Biopharmaceutical Section, Biometrics Section, ENAR
**Sunday, August 6, 2:00 pm–3:50 pm**

### Analysis of a Composite Endpoint with Missing Data in Components

❖ Hui Quan, sanofi-aventis, BX2-416A, 200 Crossing Blvd., P O Box 6890, Bridgewater, NJ 08807, *hui.quan@sanofi-aventis.com*; Daowen Zhang, sanofi-aventis; Ji Zhang, sanofi-aventis; Laure Devlamynck, sanofi-aventis

***Key Words:*** data imputation, estimation bias, correlated endpoints, event rate

Composite endpoints are often used in clinical trials in order to increase the overall event rates, reduce the sizes of the trials and achieve desired power. Just as any endpoints, missing data can occur in the components of the composite endpoint. If a patient has missing data on some of the components but not all the components, this patient may not have complete data but has partial data for the composite endpoint. The patient should not be discarded from analysis. In this talk, we propose approaches for the analysis of composite endpoint with missing data in components. The main idea is first to derive the rates for individual components based on the appropriate model and then to combine them to obtain the overall rate for the composite endpoint. Simulation will be used to compare the approaches and a data example will be used to illustrate the application of the approaches.

### A Multivariate Median-Based Robust Procedure To Analyze Multiple Endpoints

❖ Kao-Tai Tsai, Organon, 34 Baldwin Drive, Berkeley Heights, NJ 07922, *tsai0123@yahoo.com*; Harji Patel, Georgia Southern University

***Key Words:*** multivariate median, clinical trial, Wald statistics

In clinical trials, one often compares two treatment groups with respect to multiple, say d, endpoints whose joint distribution may deviate from multivariate normality. For each endpoint, we apply a robust regression model with common covariates, but without the treatment effect, and compute the residuals. After grouping the residuals by treatment group, we generate all pairwise differences. Corresponding to the d endpoints jointly, we estimate the multivariate median (Chaudhuri 1996). This median vector is unique and possesses nice asymptotic properties. We use permutation test with Wald statistic for simultaneous inference on the treatment differences. Based on clinical importance of individual endpoints, one can a priori partition the overall significance level and compute an asymptotic confidence interval for each endpoint. A numerical example is given to illustrate the methodology.

### On O'Brien's OLS and GLS Tests for Multiple Endpoints

❖ Sergei Leonov, GlaxoSmithKline, 1250 S. Collegeville Road, PO Box 5089, Collegeville, PA 19426, *Sergei.2.Leonov@gsk.com*; James Roger, GlaxoSmithKline; Nigel Dallow, GlaxoSmithKline

***Key Words:*** multiple endpoints, O'Brien's OLS and GLS tests, operational effect size, sample size calculation, adjustment for covariates

O'Brien (1984) proposed two global multivariate tests, OLS and GLS, to compare two treatment groups on multiple endpoints. In this presentation the term "operational effect size", or OES, is introduced, which is the effect size that would be required for a single endpoint to have the same power in an equivalently sized study. We mention the fundamental inequality between OES for GLS and OLS tests which explains why the GLS test is generally more powerful than the OLS test. We explore the effect of different correlation structures and discuss the adjustment of both tests for covariates.

### Tree-Structured Gatekeeping Procedures in Clinical Trials with Multiple Objectives

❖ Alex Dmitrienko, Eli Lilly and Company, 10913 W. 144th Street, Overland Park, KS 66221, *dmitrienko_alex@lilly.com*; Brian L. Wiens, Myogen, Inc.; Ajit C. Tamhane, Northwestern University; Xin Wang, Northwestern University

***Key Words:*** multiplicity, multiple endpoints, control of type I error rate, clinical trials

This talk will discuss a new approach to constructing testing procedures in clinical trials with multiple objectives [e.g., trials involving multiple dose-control tests with logical restrictions (testing of secondary endpoints is restricted to the doses for which the primary endpoint is significant) or multiple primary/secondary endpoints]. The proposed tree gatekeeping procedures account for the hierarchical structure of the testing problem. If a certain test is not significant, other tests may no longer be of interest or might not be logical. The proposed procedures enable clinical researchers to apply the overall alpha initially allocated to those further tests to other tests that retain interest. The proposed methodology will be illustrated using a trial with multiple endpoints and multiple testing strategies (superiority and noninferiority testing).

### Control of Overall Type I Error in Clinical Trials with Both Surrogate and Final Endpoints

❖ Chung-Kuei Chang, Cephalon, Inc., 1378 Beaconfield Lane, Lancaster, PA 17601, *schang@cephalon.com*

***Key Words:*** surrogate endpoint, accelerate approval, overall type I error, interim analysis, modified Simes procedure, adjusted p-value

For serious diseases with an unmet medical need, FDA may grant an accelerated approval based on a surrogate endpoint and stipulate post-marketing studies using the endpoint of primary interest (the final endpoint). FDA may also request the sponsor to design phase 3 studies with enough power for both surrogate and final endpoints and adjust for multiplicity. Assuming a non-negative but unknown correlation of the two endpoints, we show how to control the overall type I error for testing both endpoints with multiple interim analyses. An adjusted p-value for the interim and final analyses is first defined for each endpoint. After showing that these adjusted p-values follow the uniform distribution on (0, 1) under the null hypothesis, we demonstrate that a modified Simes procedure applied to the adjusted p-values controls the overall type I error through a phi-inverse transformation.

### Monitoring Futility in Two-by-Two Factorial Studies

❖ Leslie A. McClure, The University of Alabama at Birmingham, RPHB 327G, 1530 3rd Ave S, Birmingham, AL 35294-0022, *lmcclure@uab.edu*; Christopher S. Coffey, The University of Alabama at Birmingham; George Howard, The University of Alabama at Birmingham

***Key Words:*** monitoring, clinical trials, factorial designs

For studies with 2 x 2 factorial designs, the complexity of determining an appropriate futility analysis plan is increased as compared to studies where patients are randomized to one treatment. Issues that must be addressed include the possibility of a significant interaction and the need to determine how to proceed given evidence of futility in one arm. Suggested approaches include a two-stage plan, which first assesses futility of the interaction term and proceeds to examine the main effects given sufficient evidence that no interaction is present, and variations on one-stage plans, which assume the trial will not be stopped for futility in the interaction. We will summarize issues inherent to monitoring futility in 2 x 2 factorial studies, compare statistical properties of various approaches, and make recommendations as to which approach is preferred.

## 26 Mortgages and Auctions ●

Business and Economics Statistics Section

**Sunday, August 6, 2:00 pm–3:50 pm**

### Loss and Prepayment Modeling in the Context of Subprime Mortgage Loans

Deniz Senturk, GE Global Research; ❖ Huaiyu Ma, GE Global Research, One Research Circle, Bldg K1 4C43, Niskayuna, NY 12309, *mah@research.ge.com*; Greg Ratkovsky, WMC

***Key Words:*** survival analysis, default, hazard curves, subprime mortgage, prepayment, multinomial logistic regression

Developing robust models that better predict prepayments and losses for loans provides a competitive edge to mortgage businesses. Previous work concentrates on mostly default side. However, prepayment is extremely important for subprime lending. Business loses more than 90% of their loans in the first five years and approximately 50% in the first two years of loan age. In this study, we illustrate a new technology that successfully predicts prepayments and defaults simultaneously as competing risks. This new approach also provides means to understand the mortgage sensitivity to economic conditions, loan and borrower information, and default and prepayment probabilities over the loan life cycle. Results are actionable for mortgage businesses in policymaking, risk taking, rate setting, and accounting practices.

### Credit Rating Transition of U.S. Corporate Bonds

❖ Weijian Liang, New York University, KMEC Suite 8-160, 44 West 4th Street, New York, NY 10012, *wliang@stern.nyu.edu*; Halina Frydman, New York University; Stephen Figlewski, New York University

***Key Words:*** credit migration, Cox model, time-dependent covariate, macroeconomic variable, transition history, hazard rate

This research is an extensive study of factors that influence the hazard rate of occurrence of a credit event. Credit events include defaults from different rating categories, and upgrades and downgrades between two different rating categories. We use macroeconomic variables and firm specific variables as covariates in our analysis. The Cox model is employed to estimate the effects of fixed and time-dependent covariates on the hazard rate. The study is based on the corporate credit migration data from Moody's over the period from 1981 to 2002. Our analysis starts with a large number of macroeconomic factors, and for each type of transition (e.g. C to default) selects the model with the relevant subset of macroeconomic variables. Our final models, in addition to a relatively recent rating history of a firm, generally consist of a small subset of macroeconomic variables.

### Credit Risk Ananlysis for Taiwan Electronic Industrial

❖ Yi-Kuan Jong, St. John's University, 499 Sec 4 Tam King Road Tamsui, Department of Finance, Taipei, 25135 Taiwan, *joey.jong@msa.hinet.net*

***Key Words:*** credit rating, logistic regression, credit risk

Credit rating is an instrument used to measure a company's credit risk. It is measured by credit rating companies such as Moody's Investor Services, Standard & Poor's Corporation, Taiwan Rating, TCRI, etc. Institution inventors use this information to evaluate listed companies' credit risks. When news of a company being down graded break out, investors will sell their stocks simultaneously and cause the price down ticked. And it eventually affects investor's financial statement. In this study, we use linear regression analysis and logistic regression to analyze TCRI data for electronic industrial in Taiwan. We want to find the important factors which cause down grading and provide institution investor a method to evaluate a companies credit risk before latest credit rating is published.

### Statistical Validation of a Credit Risk Model

❖ Lydian Medema, University of Groningen, P.O. Box 800, Dep. Econometrics, Groningen, 9700AV The Netherlands, *l.medema@rug.nl*

***Key Words:*** internal statistical validation, Basel II, logit models, credit risk models

The Basel Committee on Banking Supervision published a new proposal (Basel II) for revising the existing Basel I capital adequacy framework from 1999. Basel II forces banks to develop credit risk models to be used in risk management and for calculation of reserves. According to the Basel II guidelines, these models have to be validated. However, regulators have not provided clear guidelines as to how this validation should take place. We propose a simple validation methodology that can be used by banks to validate their credit risk models and apply this methodology to a real-world case. We focus on statistical criteria for validation. Validation is obviously not only a statistical exercise, we also pay attention to managerial judgment and a qualitative analysis of the model. This paper gives a clear overview of the available statistical validation methods.

### A Semiparametric Investigation of the Effect of Reserve Prices on Selling Prices Using Identical Auctioned Items from eBay

❖ Dawit Zerom, University of Alberta, School of Business,
Room 2-43 Business Building, Edmonton, AB T6G2R6 Canada,
*dawit@ualberta.ca*; Peter Popkowski Leszczyc, University of Alberta

***Key Words:*** semi-parametric, auctions, identical items, non-linearity, censored, endogenous

In this paper we investigate the nature of the effect of reserve prices on the auction outcome. In particular, we develop a semi-parametric model with the following properties: a) all continuous covariates that affect the auction outcome are allowed to enter non-parametrically, but additively, without imposing any a priori rigid structure, b) number of bidders is allowed to be endogenous and c) the response variable can be censored. In estimating such model, we assume that sufficient instruments are available for the number of bidders' variable and the censoring points are known. Using five identical items from eBay, and applying the proposed model, our study is able to disentangle the separate effects of censoring, endogeneity and non-linearity. Such a result will guide future auction researchers how to model these key empirical regularities of internet auctions.

### A Statistical Approach to Controlling Sniping in Electronic Auctions

❖ Dawn Porter, Georgetown University, McDonough School of Business, 3700 O St NW, Washington, DC 20057, *dcp22@georgetown.edu*; J. Keith Ord, Georgetown University

***Key Words:*** electronic auctions, exponential distribution, random end-times, sniping, uniform distribution

The online auction market has been growing at a spectacular rate. Most auctions are open-bid auctions where all the participants know the current highest bid. This knowledge has led to a phenomenon known as sniping, whereby some bidders may wait until the last possible moment before bidding, thereby depriving other bidders of the opportunity to respond and also preventing sellers from obtaining the highest price for an item. This is especially true in the case of a second price, fixed deadline auction. We develop a procedure involving a randomly determined, yet finite, end-time, which eliminates the potential benefits to a sniper. The scheme enables all bidders to compete on a more equal footing and promotes an early bidding strategy, which is expected to increase the price received by the seller.

### A New Model for Forecasting Credit Spread Changes: Model Estimation, Prediction, and Inference Procedures

❖ Yang Wang, The Pennsylvania State University, 325 Thomas Building, University Park, PA 16802, *ywang@stat.psu.edu*

***Key Words:*** varying-coefficient model, Wilk's phenomenon, forecasting, finance, local linear, generalized likelihood ratio test

Motivated by both empirical and statistical considerations of potential structural changes of linear regression models, we proposed time-varying-coefficient models, a new modeling technique for financial data, to analyze credit spread data for nine Merrill Lynch corporate bond indexes. Chow test was conducted to motivate the proposed model. Estimation procedures using local linear modeling were explained. We compared the new model fit to that of ordinary least squares regression model. A novel prediction procedure by time-varying-coefficient model was proposed and compared to ordinary least squares regression model. Confidence intervals were constructed and generalized likelihood ratio-type test was applied to test whether the coefficients really vary. Wilks' phenomenon was shown to hold. Monte Carlo simulation studies were used to assess the finite sample performance of the procedure.

## 27 Software

Section on Statistical Computing

**Sunday, August 6, 2:00 pm–3:50 pm**

### The Carapace Environment

❖ Gary Oehlert, University of Minnesota, School of Statistics, 224 Church St SE, Minneapolis, MN 55455, *gary@stat.umn.edu*

***Key Words:*** user interface, dynamic graphics, R

Carapace is a user interface environment suitable for use with statistical software such as R. It uses the wxWidgets library to provide input/output windows, graphics windows, user-specified menus and modal dialog boxes, and some dynamic interactive graphics with native look and feel on Windows, Mac OS X, and Linux-GTK. I will describe some of what Carapace provides and give an overview of how you use it.

### Enterprise Automatons with R

❖ Zubin Dowlaty, InterContinental Hotels Group, P.O. Box 467282, Atlanta, GA 31146, *zubin.dowlaty@ichotelsgroup.com*; Dean Mao, InterContinental Hotels Group; Simon Urbanek, AT&T Labs-Research

***Key Words:*** R, automation, business process modeling, enterprise analytics, workflow

Presently within the software industry their exists a genre with the label of Business Process Modeling or BPM, which is an attempt to create a generic framework for modeling workflow processes. Further, SAS with Enterprise Miner, SPSS with the Clementine product, and others have introduced a process driven approach focusing more on statistical applications. When we scan the open source landscape, there presently did not exist an enterprise capable analytics solution that leverages these BPM concepts. R as a statistical language is extremely robust, we feel R coupled with an enterprise quality open source BPM engine and visual client that can be used to model, persistent and schedule analytics workflows, this combination would elevate R into many new and unique applications. Purpose of our talk is to discuss our solution and release the open source codebase to the community.

### Estimation and Inference in Parametric Stochastic Frontier Models: a SAS/IML Procedure for a Maximum Likelihood Bootstrap Method

❖ Sylvie Tchumtchoua, University of Connecticut, 210 Quinebaug, Storrs, CT 06269, *tchumtchoua@yahoo.fr*

***Key Words:*** linear model, bootstrap, technical efficiency, inference, maximum likelihood, SAS/IML

Parametric Stochastic Frontier Models (PSFM) specify the output or cost of a production unit in terms of a response function and a composite error made of a symmetric noise and a one-sided error representing technical inefficiency. PSFM are widely used in productivity analysis and are commonly estimated using FRONTIER or LIMDEP packages which do not provide inference about the inefficiency term. Moreover usual approach for inference about efficiency in PSFM is based on percentiles of the estimated distribution of the one-sided error term, conditional on the composite error, rather than on the sampling distribution of the inefficiency estimator. We propose a concise program for a maximum likelihood bootstrap method which makes inference about the inefficiency term based on its sampling distribution. The program is written using matrix language SAS/IML with the optimization subroutine NLPQN.

### A New Program for Computing Percentage Points for Pearson Distributions

❖ Wei Pan, University of Cincinnati, Division of Educational Studies, PO Box 210002, Cincinnati, OH 45221-0002, *wei.pan@uc.edu*; Haiyan Bai, University of Cincinnati

***Key Words:*** Pearson distributions, curve fitting

Pearson distribution family (Pearson 1895) provides approximations to a wide variety of observed distributions using the first four moments or the first three moments with a left or right boundary. Curve fitting using the Pearson distributions has been extensively applied in many fields. However, in practice, it is quite unwieldy to obtain the percentage points for the Pearson distributions when consulting the inconvenient tables of percentage points or using the out-of-date computer programs. Therefore, it is essential to have an advanced and efficient computer program to compute percentage points for the Pearson distributions. To obtain the percentage points for Pearson distributions conveniently, the present study develops a new computer program which is compatible with contemporary SAS and SPSS programs.

### Statistical Inference Package (SIP)

❖ Esa Uusipaikka, University of Turku, 200014 Turun Yliopisto, Finland, Turku, 20014 Finland, *esa.uusipaikka@utu.fi*

***Key Words:*** inference, likelihood, profile likelihood, confidence interval, program, Mathematica

SIP is a MATHEMATICA package for statistical inference using concepts and constructs of teaching, writing, and applying. SIP contains procedures for maximum likelihood estimation, likelihood ratio tests of general hypotheses concerning parameters, and profile likelihood based confidence intervals for general interest functions of parameters. SIP contains large collection of discrete and absolutely continuous univariate distributions and also multivariate distributions. SIP contains many sophisticated statistical models. SIP uses a new method for calculation of profile likelihood based confidence intervals for general parameter functions of interest in general parametric statistical models. SIP gives in addition to the statistical analysis procedures easy access to the powerful tools in MATHEMATICA for doing mathematics, graphics, programming, and presentation.

## 28 Testing

Section on Statistical Computing

**Sunday, August 6, 2:00 pm–3:50 pm**

### Testing the Equality of Two Normally Distributed Populations

❖ Charles Dunn, Miami University, Department of Math and Statistics, Oxford, OH 45056, *dunncl@muohio.edu*

***Key Words:*** likelihood ratio, Sukhatme test, squared distribution difference, Welch's t-test, f-test

A new hypothesis test, called the difference test, is presented for testing whether two normal populations have both equal means and equal standard deviations. First, a new parameterization scheme is presented that describes the difference in the populations in terms of a primary 'distance' parameter and a secondary 'orientation' parameter. The type I error and power properties for the difference test for various values of the primary and secondary parameters is then compared to those for four other tests: the likelihood ratio test, the Sukhatme test, the Welch t-test for the equality of means, and the F-test for the equality of variances. The distance test has favorable power properties for large sample sizes and also for small samples when the difference in the populations is due primarily to the difference in the means. A complete table of critical values is included.

### Generation of the Distribution of the Test for a Latin Square Design with Heterogeneous Variances

❖ Miin-Jye Wen, National Cheng Kung University, Management Building, 2nd Floor, No. 1 Ta-Hsueh Road, Department of Statistics, 70101 Tainan, Taiwan, *mjwen@stat.ncku.edu.tw*; Hubert Chen, National Cheng Kung University

***Key Words:*** heteroscedasticity, one-stage procedure, t distribution, Latin square, simulation, critical values

Assuming a Latin square design model whose error terms follow independent and normal distributions with unknown and possibly unequal variances, the interest is to develop a one-stage sampling procedure to test the null hypothesis of equal treatment effects. A weighted sample mean is used to serve as a basis for a test statistic for the Latin square design model. It is found that the distribution of the test statistic based on the weighted sample mean is completely independent of the unknown variances. A SAS simulation program to m by m Latin square design is developed to implement the testing procedure.

### On Testing the Bioequivalence of Several Treatments Using the Measure of Distance

❖ Hubert Chen, National Cheng Kung University, 1 University Road, Department of Statistics, Tainan, 70101 Taiwan, *jpchen@stat.ncku.edu.tw*; Miin-Jye Wen, National Cheng Kung University

***Key Words:*** studentized range, bioequivalence, least favorable, level and power, two-stage procedure, quadrature

A studentized range test based on k (k being larger than two) random samples is proposed to test the hypothesis of equivalence of treatment means in terms of the standardized distance among means. A least favorable configuration (LFC) of the means to guarantee the maximum level at the null hypothesis and a LFC of the means to guarantee the minimum power at an alternative hypothesis are obtained. The level and power of the test are fully independent of the unknown means and variances at their corresponding LFCs. For a given level and power, both the critical values and the required experimental sample sizes can be determined. In situations where the common population variance is unknown, and if the measure of bioequivalence is considered to be purely the distance of normal means, a two-stage sampling procedure can be employed to find the needed solutions.

### An Exact Test for Testing the Equality of Parameter Matrices in Two Multivariate Linear Models

❖ Jinadasa K. Gamage, Illinois State University, Department of Mathematics, Normal, IL 61790-4520, *jina@ilstu.edu*; Malwane M. A. Ananda, University of Nevada, Las Vegas

***Key Words:*** multivariate linear models, generalized p-values, heteroscedasticity

A method of testing the hypothesis of equality of parameter matrices in two independent multivariate linear models is presented using the generalized p-values under the assumption of error matrix variate normality and heteroscedasticity. This method calculates the exact p-value in the generalized sense. Even for the homoscedastic case there are no exact tests available for this testing problem. Nel (1997) provided an approximate degrees of freedom test to test this hypothesis.

### Performance of Robust and Nonrobust Roy-Bargmann Stepdown Follow up to a Significant MANOVA under a Variety of Conditions: a Simulation Study

❖ Holmes Finch, Ball State University, Educational Psychology, Muncie, IN 47306, *whfinch@bsu.edu*

***Key Words:*** MANOVA, post hoc analysis, Roy-Bargmann, stepdown procedures

Following a significant MANOVA, researchers may want to identify individual dependent variables for which group means differ. One approach for examining these univariate differences is the Roy-Bargmann stepdown procedure. In this analysis dependent variables are ordered by theoretical importance, after which sequential univariate testing is conducted in descending order of importance, where more important variables serve as covariates for less important ones. This simulation study examined the ability of the Roy-Bargmann method to correctly identify dependent variables for which group means differ (power) and those for which means do not differ (Type I error). Several factors were manipulated in the simulations. Results suggest that the manipulated variables do influence the performance of the method, and that robust and non-robust approaches perform differently under some conditions.

### Comparisons of Sets of Multivariate Time Series

❖ Jaydip Mukhopadhyay, University of Connecticut, Department of Statistics, 215 Glennbrook Road CLAS Bldg, Storrs, CT 06269, *jaystat@yahoo.com*; Nalini Ravishanker, University of Connecticut; Jonathan Hosking, IBM Research

***Key Words:*** cross-spectra, likelihood ratio test, maximum periodogram ordinate test

We discuss the problem of comparison of several multivariate time series via their spectral properties. For two independent multivariate Gaussian stationary time series, such a comparison is made via a likelihood ratio test based on the estimated cross-spectra of the series. This is an extension of the maximum periodogram ordinate test developed in the literature to compare two independent univariate stationary time series. A simulation based critical value enables effective comparison of several such multivariate time series, and is useful in applications to biomedical time series or marketing or manufacturing time series. The spectral approach is extended for the comparison of non-Gaussian and/or nonlinear time series.

### Iterated BH Procedure

❖ Nasrine Bendjilali, Lehigh University, 4133 Huron Ave., Culver City, CA 90232, *nab4@lehigh.edu*; Wei-Min Huang, Lehigh University

***Key Words:*** false discovery rate, p-value, BH procedure, multiple hypotheses testing

This article introduces a method for multiple hypotheses testing which is an iterated version of the original BH Procedure (the procedure that was introduced by Benjamini and Hochberg (1995) in their paper titled "Controlling the false Discovery Rate: a Practical and Powerful Approach to Multiple Testing"). The jump size between the potential "smallest null" p-value and the potential "largest alternative" p-value is used to determine the stopping rule. It is known that the original BH procedure tends to be conservative in general. The proposed modification seems to have fairly good control of FDR (the false discovery rate). Preliminary study also shows promising gain in power.

## 29 Likelihood-Based Inference

IMS

**Sunday, August 6, 2:00 pm–3:50 pm**

### Testing of Rate Ratio under Inverse Sampling

❖ Hon Keung T. Ng, Southern Methodist University, Department of Statistical Science, 3225 Daniel Ave, Dallas, TX 75275, *ngh@mail.smu.edu*; Man Lai Tang, Hong Kong Baptist University; Yijie Liao, Hong Kong Baptist University; Ping Shing Chan, The Chinese University of Hong Kong

***Key Words:*** negative binomial, score test, Wald test, likelihood ratio test, Monte Carlo simulation

Inverse sampling is considered to be a more appropriate sampling scheme than the usual binomial sampling scheme when the subjects arrive sequentially and when the underlying response of interest is acute. In this talk, we present various test statistics for testing rate ratio in case-control studies under inverse sampling. They include the Wald, unconditional score, Wald-score, likelihood ratio and conditional score test statistics. Three methods, namely the asymptotic, conditional exact and mid-P exact methods, are adopted for P-value computations. We evaluate the performance of different test procedures in terms of their actual type I error rates and powers via an empirical study. Some recommendations under different settings are provided. The methodologies are illustrated by a real example from a heart disease study.

### A Bivariate Interval Censorship Model for Partnership Formation

❖ Qiqing Yu, Binghamton University, Math Department, Binghamton, NY 13902, *qyu@math.binghamton.edu*; Linda Wong, Binghamton University

***Key Words:*** bivariate mixed interval-censored data, missing data, generalized MLE, exact observations, asymptotic properties

We consider a statistical problem of estimating a bivariate age distribution of newly formed partnership. The study is motivated by a type of data that consist of uncensored, right-censored, left-censored, interval-censored, and missing observations in the coordinates of a bivariate random vector. A model is proposed for formulating such data. A feasible algorithm to estimate the generalized MLE (GMLE) of the bivariate distribution function also is proposed. We establish asymptotic properties for the GMLE and apply the method to the dataset.

### The Likelihood Ratio Test of Mixture Hypotheses and the Tube Volume Problem

❖ Yong Lin, University of Medicine & Dentistry of New Jersey, 195 Little Albany Street, Room 5536, New Brunswick, NJ 08901, *linyo@umdnj.edu*; Bruce G. Lindsay, The Pennsylvania State University

***Key Words:*** mixture hypotheses, likelihood ratio test, Hotelling's tube volume, projection onto cone

The testing of the number of components in a mixture model has wide applications. But the distributional properties of likelihood ratio test of the mixture hypotheses have long been an enigma. In this paper, we will consider the testing of s versus t components of mixtures for multinomial models when the parameter is in any finite dimensional space. Based on Lindsay's geometric framework (1995), we will use Hotelling-Naiman's tube volume formula and Weyl's technique to find upper and lower bounds for the upper quantile of the projection of normal vector onto a non-convex cone, and hence give the upper and lower bounds for the upper quantile of the asymptotic distribution of likelihood ratio statistics of the mixture hypotheses. Adequacy of the quantile approximation will be studied. Furthermore, we will apply our results to the mixture of binomial distributions.

### On Hinkley's Estimator: Inference about the Change-Point

❖ Stergios B. Fotopoulos, Washington State University, Department of Management and Operations, Pullman, WA 99164, *fotopo@wsu.edu*; Venkata Jandhyala, Washington State University

***Key Words:*** maximum likelihood estimation, multivariate Gaussian vector, random walks, Wiener-Hopf factorization, ladder epochs

Maximum likelihood method is applied to estimate the change-point of a distribution function associated with a sequence of independent random elements. Fluctuation theory of random walks is applied to show exact expressions for the limiting distribution of the maximum likelihood estimator of a change-point. The derived expressions are computationally accessible in the sense that one may compute the exact distribution of the change-point through an algorithmic approach on the basis of the expressions derived. In showing this, a new formula for the ultimate maximum, the maximum of the sequence of partial maxima of a random walk is exhibited. As an example, the methodology is illustrated for estimating the change-point in the mean vector or/and variance-covariance matrix of the multivariate normal distribution.

### Conditional Properties of a Parametric Bootstrap

❖ Russell Zaretzki, University of Tennessee, 328 Stokely Management Center, Knoxville, TN 37996-0532, *rzaretzk@utk.edu*

***Key Words:*** bootstrap, asymptotics, conditional inference, likelihood theory

DiCiccio, Martin and Stern(2O01) introduced the parametric bootstrap of the signed root statistic as a useful computational alternative to analytic approximations when highly accurate statistical inference is desired. This performance is equivalent to asymptotic techniques such as the r-star formula; see Barndorff-Nielson(1986). In addition, simulation examples contained in DiCiccio (2001) suggest that this bootstrap technique can produce extremely accurate conditional inferences. The present work further investigates these conditional properties. In particular, we prove that, in the presence of nuisance parameters, inferences based on a parametric bootstrap of the signed root are conditionally accurate to order 1/n. This project is joint work with Tom DiCiccio and G.A. Young.

### Summarizing and Interpreting Likelihood Functions as Functions

❖ Michael Brimacombe, University of Medicine & Dentistry of New Jersey, 185 S. Orange Ave., MSB F-647, Newark, NJ 07083, *brimacmb@umdnj.edu*; Bo Peng, University of Medicine & Dentistry of New Jersey

***Key Words:*** likelihood, meta-analysis, functional data analysis

The combining of likelihood-based information in the context of meta-analysis and multiple related experiments is considered. The set of likelihood functions, viewed as a set of functions containing all information of relevance to statistical inference, can be summarized using functional data-related methods to obtain an overall assessment of likelihood-related evidence. In the case of one-dimensional likelihood, the functions can be renormed and centered, providing a useful graphical assessment of overall study variation. This can be combined with a weighted average measure of location to provide a summary interval for estimation of the parameter of interest. In higher dimensions, this approach can be applied to conditional and marginal likelihoods as well as respective Bayesian posterior densities. The approach is compared to empirical Bayes methods.

### Testing for and against a Set of Linear Inequality Constraints in the Product Multinomial Setting

❖ Hammou Elbarmi, Baruch College, One Baruc Way, New York, NY 10010, *hammou_elbarmi@baruch.cuny.edu*

***Key Words:*** chi bar square, likelihood ratio test, orthant probabilities, Lagrange multipliers

A problem that is frequently encountered in statistics concerns testing for equality of multiple probability vectors corresponding to independent multinomials against an alternative they are not equal or that they are stochastically ordered. Our aim here is to generalize the existing results and provide a unified technique for testing for and against a set of linear inequality constraints placed upon on any $r\; (r \ge 1)$ probability vectors corresponding to $r$ independent multinomials. We show how to compute the mles under all hypotheses of interest and obtains the limiting distributions of the LRT statistics. These are of chi bar square type and to illustrate our results, an example is discussed.

## 30 Bayesian Biomedical Modeling

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### Monitoring Event Times in Early-Phase Clinical Trials: Practical Issues

❖ Leiko H. Wooten, M. D. Anderson Cancer Center, Department of Biostatistics and Applied Math, Box 447, 1515 Holcombe Boulevard, Houston, TX 77030, *leiko@mdanderson.org*; Peter F. Thall, M. D. Anderson Cancer Center; Nizar M. Tannir, M. D. Anderson Cancer Center

***Key Words:*** Bayesian statistics, futility, safety monitoring, phase II clinical trial

In many early-phase clinical trials, rather than characterizing patient outcome as a binary variable, it may be more natural to construct decision rules based on time-to-event variables and event rates. This approach often makes better use of available information and avoids logistical problems and loss of information due to replacing an event time by an indicator of whether the event occurred within a fixed observation period. Monitoring event times may involve various complications, however. We describe Bayesian methods to deal with certain complications that may arise when monitoring event times, including interval censoring and effects of intermediate events on survival time. We present several designs, each based on the times to a severe adverse event, disease progression, and death. Simulation results for each design are presented.

## Modeling Long-Term HIV Dynamics: a Bayesian Approach

❖ Dacheng Liu, Boehringer Ingelheim, Biometrics and Data Management, 900 Ridgebury Rd, Ridgefield, CT 06877, *dliu@rdg.boehringer-ingelheim.com*; Hulin Wu, University of Rochester; Yangxin Huang Huang, University of South Florida

***Key Words:*** antiretroviral drug therapy, Bayesian mixed-effects models, drug exposure, drug resistance, long term HIV dynamics, MCMC

HIV dynamics studies have significantly contributed to the understanding of HIV infection and antiviral treatment strategies. But most studies are limited to short-term viral dynamics due to the difficulty of establishing a relationship of antiviral response with multiple treatment factors such as drug exposure and drug susceptibility during long-term treatment. We propose a mechanism-based dynamic model for characterizing long-term viral dynamics with antiretroviral therapy. We directly incorporate drug concentration, adherence, and drug susceptibility into a function of treatment efficacy, defined as an inhibition rate of virus replication. We investigate a Bayesian approach under the framework of hierarchical Bayesian (mixed-effects) models for estimating unknown parameters. We run simulation studies and apply the methodology to a data set from an AIDS clinical trial.

## Prior Structures for Surrogate Endpoint Validation Using PTE

❖ Chunyao Feng, Baylor University, 4701 Staggerbrush Road, Apt. 436, AUSTIN, TX 78749, *Chunyao_Feng@baylor.edu*; John W. Seaman, Baylor University; Stacy Lindborg, Eli Lilly and Company

***Key Words:*** Bayesian, surrogate, endpoint, PTE, prior, posterior

To save time and reduce the size and cost of clinical trials, surrogate endpoints are frequently measured instead of true endpoints. The proportion of the treatment effect explained by surrogate endpoints (PTE) is a widely used validation criteria. Frequentist and Bayesian methods have been developed to facilitate such validation. The existing Bayesian method, due to Cowles (2000), uses a diffuse normal prior structure on the regression parameters needed to define PTE. Consequently, the posterior for PTE can be outside the unit interval. We propose a model using conditional beta priors that precludes this. Furthermore, we derive the induced prior on PTE thereby allowing direct comparison of prior and posterior distributions.

## A Bayesian Multivariate PK/PD Model for Analyzing Cortisol Circadian Rhythm in a Depression Study

❖ Niko Kaciroti, University of Michigan, 300N Ingalls Building, CHGD 10 floor room 1027NW, Ann Arbor, MI 48109, *nicola@umich.edu*; Trivellore E. Raghunathan, University of Michigan; Delia Vazquez, University of Michigan

***Key Words:*** Markov chain Monte Carlo, nonlinear hierarchical models

This paper presents results from a study of depression among new mothers and their infants using their cortisol circadian rhythm (CCR). We construct a multivariate Pharmacokinetic-Pharmacodynamic (PK/PD) model to jointly estimate the CCR for mother and infant as well as the strength of the interdependence between their individual CCR. A non-linear random effects model is used where each individual trajectory has its own parameters. We assume that the subject-specific parameters are normally distributed around the parameters of the overall population-average trajectory. The model is fitted using the Bayesian approach implemented through MCMC. The proposed model is implemented to asses the relationship between alteration of the mother's CCR and depression. Furthermore to assess whether infants whose CCR is strongly related to their mother's CCR are likely to show better neurodevelopment.

## Bayesian Modeling of Correlated Binary Data from the Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) Study

❖ Claudia Pedroza, The University of Texas School of Public Health, 1200 Herman Pressler, RAS E831, Houston, TX 77030, *claudia.pedroza@uth.tmc.edu*; Betty Tung, The University of Texas School of Public Health

***Key Words:*** Bayesian multilevel modeling, correlated binary data, multicenter clinical trials, random effects models, mixed effects

Data arising from multicenter clinical trials can be complex to analyze due to the possibility of correlation between observations within a clinical center. If the correlation is ignored, variance estimates could potentially be incorrect. In this talk, we present a Bayesian multilevel model which accounts for correlation within a study center when estimating the incidence of ROP. We show that this methodology allows for great flexibility both in the choices of distributions and also in the structure of the model, i.e. random effects for the centers and/or patients. Methods presented will be applied to a data set from the CRYO-ROP study. This data set consists of a binary outcome (ROP) and baseline variables (e.g., gestational age, race, birth weight, gender). We will also discuss the ramifications that a random effects study can have on a multicenter clinical trial.

## Bayesian Analysis of Age-Adjusted Cancer Rates Using Joinpoint Regression Model

❖ Ram Tiwari, National Institutes of Health, 6116 Executive Blvd., Bethesda, MD 20892, *tiwarir@mail.nih.gov*; Pulak Ghosh, Georgia State University

***Key Words:*** annual percentage change, Bayes information criterion, deviance information criterion, conditional predictive ordinate, Dirichlet process, permutation test

Bayesian model selection procedures for the joinpoint regression model are developed under different assumptions on the prior distributions for the joinpoints, the regression parameters, and the random errors. The methods are applied to analyze the age-adjusted cancer incidence and mortality rates from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI). The results are compared with the ones resulting from the NCI's permutation test based Joinpoint software.

## Bayesian Modeling of Noncompliance in Folic Acid Dosing Studies

❖ Owen Devine, Centers for Disease Control and Prevention, MS E87, 1600 Clifton Road, Atlanta, GA 30333, *ojd1@cdc.gov*

***Key Words:*** noncompliance, dosing studies, Markov chain Monte Carlo

We use a Bayesian approach to estimate casual effects, as defined by Rubin, in studies comparing folic acid dosing regimes with regard to blood folate level and the occurrence of neural tube defects. In these studies, noncompliance is associated with only a subset of Rubin's classifications, compliers, always takers, never takers and defiers, in that non or partial compliance implies a lowering of true folate levels in all study arms. In addition to estimating causal effects, we propose updating self-reported compliance values using observed folate levels and additional covariates. Potential association between compliance, outcome and missingness is assessed using a sensitivity analysis.

# 31 Consumer Prices and Expenditures

Section on Government Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

## A Micro-Level Latent Class Analysis of Underreporting on the Consumer Expenditure Survey

❖ Brian Meekins, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212, *meekins.brian@bls.gov*; Clyde Tucker, Bureau of Labor Statistics; Paul Biemer, RTI International

***Key Words:*** latent class analysis, measurement error, consumer expenditures

This paper expands upon recent work by the authors using latent class analysis (LCA) to estimate the amount of underreporting on the BLS Consumer Expenditure Quarterly Survey (CEQ). Using micro-level or interview-specific indicators a latent construct is formed that indicates degree of underreporting of expenditures by those reporting a purchase of any item within that category. Previous models were applied to household total quarterly expenditures and a small number of commodity categories. This work expands the number of commodity categories examined using a variety of LC models. Best fitting models are determined from well-known statistical tests as well as substantive diagnostics developed by the authors. Data from the CEQ for the years 1996 to 2003 are used in the analysis. An evaluation of possible causes of underreporting is undertaken using multinomial logistic regression.

## The Use of Geocoding to Locate Outlets Outside of Sample Area Boundaries to Determine Significant Areas of Commerce

❖ John Schilp, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 3655, Washington, DC 20212, *schilp.john@bls.gov*; Fred Marsh, III, Bureau of Labor Statistics

***Key Words:*** geocoding, sampling, index calculation

This paper gives an introduction to the method of sampling retail outlets in the U.S. Consumer Price Index (CPI). Some of these sampled retail outlets are located outside the primary sampling area boundary and are currently priced. To reduce data collection costs, BLS would like to develop a rule to eliminate some outlets from being priced. Currently, BLS prices any selected outlet that is less than 25 miles from the primary sample area boundary or has more than 10 quotes in a cluster. We examine the impact of eliminating certain groups of outlets from the sample on the CPI index and also explain the definition of a spatial cluster and the new "significant area of commerce."

## Internet Portals and Outlet Selection Issues in the Consumer Price Index

❖ Charles Mason, Bureau of Labor Statistics, Room 3615, 2 Massachusetts Avenue, NE, Washington, DC 20212, *mason_c@bls.gov*; Roberta Sangster, Bureau of Labor Statistics; Madeleine Saxton, Bureau of Labor Statistics

***Key Words:*** internet, household survey, RDD

The growth of E-commerce, has brought with it a number of issues for the computation of the Bureau of Labor Statistics Consumer Price Index (CPI). Most significantly, identifying the specific outlet from which price data collectors can collect prices necessary to measure the monthly price change has become more difficult as e-commerce service portals have expanded. These portals, such as Travelocity and Expedia, serve as one stop shops for the purchase of hotel rooms, automobile rentals and airline fares. However, selecting a particular expenditure (e.g., Hilton, Days Inn, or Motel 6 in Peoria) for repeated pricing is more difficult because of a lack of knowledge of the probability of its purchase by the CPI population. This paper describes the cognitive research and the changes proposed for the BLS Point of Purchase survey to address this growing phenomenon.

## A Spatial Analysis of Price Change in CPI Housing Index

❖ William Larson, Bureau of Labor Statistics, 12307 Millstream Drive, Bowie, MD 20715, *larson_w@bls.gov*

***Key Words:*** CPI, spatial statistics, housing

Historically there has been an assumption that geographic proximity is one of the most important predictors of price change behavior in the housing component of the Consumer Price Index (CPI). However there has been limited research examining at what level of geography rent changes are correlated. This spatial data analysis explores the strength of the geographic relationship at various distances down to the Census block level.

## Comparison of Chained CPI-U and Regular CPI-U All-U.S. Indexes in the Housing Sector (2000--2004)

❖ Owen Shoemaker, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 3655, Washington, DC 20212-0022, *shoemaker_o@bls.gov*

***Key Words:*** Tornqvist, superlative index

In February, 2006, the BLS calculated and published its fourth annual set of C-CPI-U indexes --- for the 12 months of 2004. The C-CPI-U (Chained Consumer Price Index - Urban) is calculated and published every year, with a one year lag, using a Tornqvist formula, which is a "Superlative" index formula. By contrast, the regular CPI-U uses a combination (Hybrid) of Geomeans and Laspeyres formulas as its final estimator. For 12-month price changes, the All_US-All_Items chained C-CPI-U index results continue to diverge (significantly lower) from regular CPI-U index results, but in the Housing Sector (Rent plus Owners' Equivalent Rent, or REQ) the Superlative results have begun to track consistently higher than their Regular CPI results. We investigate the anomalous nature of these Housing results, and discover that a mathematical result rather than index theory is determinative.

### Comparison between Newly Proposed Response Rates and Current Response Rates for the TPOP Survey

❖ Fred Marsh, III, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Office 3655, Washington, DC 20212, *marsh.fred@bls.gov*

***Key Words:*** response rates

Currently, the Telephone Point of Purchase Survey (TPOPS) calculates three of the four Bureau of Labor Statistics (BLS) recommended response rates using the American Association for Public Opinion Research's (AAPOR) RR4. These four BLS recommended response rates are Initiation, Collection, Estimation, and Total Survey. As part of an effort to incorporate consistent definitions with other Office of Prices and Living Conditions (OPLC) surveys, TPOPS is developing new response rates that will satisfy all four BLS recommended response rates and still satisfy AAPOR's RR4.

### Effect of Computer-Assisted Personal Interviews in the U.S. Consumer Expenditure Interview Survey

❖ Moon Jung Cho, Bureau of Labor Statistics, 9676 Scotch Haven Drive, Vienna, VA 22181, *Cho.Moon@bls.gov*; Carolyn Pickering, Bureau of Labor Statistics

***Key Words:*** mode effect, data quality, mean expenditure, reporting rate, respondent characteristics, interviewer variables

The U.S. Consumer Expenditure (CE) Interview Survey began utilizing the CAPI instrument for data collection in April 2003. In this paper, the CAPI method is compared to the conventional PAPI method. We compare mean expenditures and reporting rates across groups defined by expenditure categories, respondent demographic variables, and interviewer variables. We will provide the relevant literature review, with special attention to the mode effect. We also will present an overview of the CE Interview Survey and CAPI application. Finally, we will describe the comparison study and present the statistical tools used. The proposed methods are applied to selected subsets of items from the CE Interview Survey. Specifically, we distinguish between parts that were exact translations of the PAPI version and parts with additional changes, such as wording and order changes.

## 32 Applications for Modeling Health Survey Data ●

Section on Health Policy Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

### Modeling of Longitudinal Polytomous Outcomes from Complex Survey Data

❖ Punam Pahwa, University of Saskatchewan, 103 Hospital Drive, Saskatoon, SK S7N 0W8 Canada, *pup165@mail.usask.ca*; Chandima Karunanayake, University of Saskatchewan; Helen H. McDuffie, University of Saskatchewan

***Key Words:*** complex survey design, ordinal logistic regression, generalized estimating equations, National Population Health Survey, mental distress, respiratory diseases

Statistics Canada's longitudinal National Population Health Survey (NPHS) dataset from the first five cycles (1994/95 - 2002/03) was used to investigate the effects of demographic, social, life-style, and health-related factors on the longitudinal changes of mental distress scores among the NPHS participants who self-reported physician diagnosed respiratory diseases, specifically asthma and chronic bronchitis. The NPHS longitudinal sample includes 17,276 persons of all ages. Mental distress, an ordinal outcome variable (categories: no/low, moderate, and high) was examined. Ordered logistic regression models based on the generalized estimating equations approach were fitted to investigate the association between respiratory diseases and mental distress adjusting for other covariates of interest. The final model was used to predict the probabilities of prevalence of mental distress.

### Statistical Modeling of Longitudinal Mental Distress among the National Population Health Survey Participants: Missing Data Analysis

❖ Chandima Karunanayake, University of Saskatchewan, 103 Hospital Drive, Royal University Hospital, Saskatoon, SK S7N OW8 Canada, *cpk646@mail.usask.ca*; Punam Pahwa, University of Saskatchewan; Helen H. McDuffie, University of Saskatchewan

***Key Words:*** National Population Health Survey, generalized estimating equation, bootstrap weights, longitudinal data, mental health, missing data patterns

National Population Health Survey (NPHS) collects longitudinal data on the physical and mental health of Canadians. The main objective of this paper was to investigate how missing data pattern influences the results obtained from the analysis of subpopulation of NPHS participants who reported non-malignant respiratory diseases in order to study longitudinal changes in mental health status of those who are age 15 years and older. From the relatively wide range of mental health indicators available in the NPHS, the distress measure based on a subset of items of the Composite International Diagnostic Interview with six questions was chosen as the ordinal outcome. The analysis was conducted using generalized estimating equation approach accounting for the complexity of multi-stage survey design using bootstrap weights available for incomplete longitudinal data.

### A Two-Phase Model To Study the Health Care--Seeking Behaviors for Common Cold of People in Taiwan

❖ Hsing-Yi Chang, National Health Research Institutes, CHPRD, 35 Keyan Road, A3223, Zhunan Maoli, 350 Taiwan, *hsingyi@nhri.org.tw*; Yu-Wen Wen, National Health Research Institutes

***Key Words:*** spatial, national health insurance, two-phase model, Taiwan

Huge health expenditures have caused an unprecedented financial burden on the National Health Insurance in Taiwan. Understanding the health care seeking behavior of the population is crucial for re-designing the payment system. We proposed to use a two-phase model to examine the behavior of the population. Phase I model estimated the logarithm of odds ratio for a person going out of town for common cold controlling for demographics. Regional variation was treated as random spatial effects. Phase II model explored the regional effects using spatial features. The models were applied to 2001 National Health Interview Survey in Taiwan. In phase I, we found younger age (< 45 yrs), lower education level were more likely to go out of town for medical care. In phase II, we found the number of medical institutes and density of paved road were negatively associated with going out of town for care.

### Alcohol Disorders and Employment Stability: a Longitudinal Study

❖ Richard Bryant, University of Missouri-Rolla, G9 Harris Hall, Rolla, MO 65401, *rrbryant@umr.edu*; V. A. R. Samaranayake, University of Missouri-Rolla

***Key Words:*** alcohol dependence, substance abuse, employment history, panel data

National Longitudinal Survey of Youth data are used to analyze the relationship between measures of alcohol abuse or dependence on employment stability indicated by the number of non-employment spells within two separate two-year periods. Self-reported answers to questions that correspond to diagnostic criteria published by the American Psychological Association are used to construct dummy variables indicating whether or not an individual was dependent on or abusing alcohol in 1989 and in 1994. The number of non-employment spells is modeled as a function of the alcohol abuse/dependence dummy variables and other personal, demographic, and economic variables in a Poisson regression. Weekly data on labor force status are used to construct the number of spells of non-employment over the period 1989-90 and the period 1994-95. Separate analyses are carried out for males and females.

### Significance Analysis of Physician Photo Identification Cards Trial

❖ Ye-Ying Cen, Hennepin County Medical Center, 954 Millwood Ave., Roseville, MN 55113, *ye-ying.cen@co.hennepin.mn.us*; Jennings Ryan Staley, United States Air Force; Baolin Wu, University of Minnesota; Scott F. Davies, Hennepin County Medical Center

***Key Words:*** patient satisfaction, survey, cumulative logistic regression, questionnaire

We undertook a prospective study to examine whether the use of physician photo identification cards could improve patient satisfaction significantly. Physicians were participants of the study who presented the cards during intervention and did not during usual care. Patient responses were collected prior to discharge via an anonymous validated nine-question survey. Univariate significance testing was conducted. Few had marginal significances; however, across all questions presented in each table, there was a consistent trend of the intervention group performing better than the usual care group. We thus performed a multivariate cumulative logistic regression model to aggregate the information across all nine questions. The intervention group produced significantly higher ratings of the nine questions ($p< 0.016$), with an odds ratio of 1.21 and a significant 95% C.I (1.04, 1.41).

### Development, Scaling, and Implementation of a Patient Satisfaction Inventory for Organ Transplant Candidates and Recipients

❖ Irene Feurer, Vanderbilt University Medical Center, 801 Oxford House, Nashville, TN 37232-4753, *irene.feurer@vanderbilt.edu*; Hongxia Liu, Vanderbilt University School of Nursing; Panarut Wisawatapnimit, Vanderbilt University School of Nursing; C. Wright Pinson, Vanderbilt University Medical Center

***Key Words:*** survey, validation, patient satisfaction, organ transplantation

Introduction: This study examined the psychometric properties of an inventory measuring patients' satisfaction with transplantation healthcare delivery. Method: The Vanderbilt Transplant Center Patient Satisfaction Inventory (VTCPSI) is part of a battery of instruments measuring health-related quality of life that includes the Short Form-36(r) Health Survey (SF-36) and a visual analogue scale of overall health (OH). The VTCPSI includes six time point-specific parallel forms, each having 13 core items. A suitable scoring method was determined using latent variable and item analysis methods. Results: Surveys were completed by 1498 adult, solid organ (liver, heart, kidney, lung) transplant patients. A consistent scoring system comprising a summary score and subscales was indicated for each form and organ group. Divergent criterion-related was indicated in relation to SF-36 and OH.

### Factor Analysis with Categorical Data: a Methodological Illustration with the GAZA Child Health Survey Data

❖ Dongguang Li, National Cancer Institute of Canada, 10 Stuart Street, Kingston, ON K7L3N6 Canada, *dli@ctg.queensu.ca*; John D. Pringle, Queen's University; Julio Arboleda-Florez, Queen's University; Heather Stuart, Queen's University

***Key Words:*** factor analysis, categorical data, war-trauma

Factor analysis is designed to identify the latent factors from a large set of variables using certain mathematics models. In contrast to the analysis for continuous data, the methods for categorical data factor analysis are relatively dubious. With the Gaza Child Health Survey data, this work conducts factor analysis using the multiple categorical war-trauma data and the transformed dichotomous data. The results are compared. The influences of different correlation coefficients are tested. The study evaluates the effects of the rotate options and the threshold for number of factors based on eigenvalues and factor loadings. The relevant methodological issues are discussed. The analysis finds the dichotomous transformation makes more appropriate evaluations of the latent factors and the tetrachoric correlation is the best index in implementing a factor analysis with dichotomous data.

## 33 Nonparametric Approaches to Regression and Spatial Modeling ●

Section on Nonparametric Statistics

**Sunday, August 6, 2:00 pm–3:50 pm**

### Regression Model-Fitting with Long Memory

❖ Hongwen Guo, Michigan State University, Department of Statistics and Probability, East Lansing, 48824, *guohongw@stt.msu.edu*; Hira L. Koul, Michigan State University

***Key Words:*** long memory, heteroscedasticity, model fitting, nonparametric regression

This paper considers a simple linear regression model with long memory design and nonparametric heteroscedastic long memory moving average errors. In some cases, the first-order asymptotic distribution of the least square estimator of the slope parameter is degenerate. In these cases, under mild conditions, the second-order asymptotic distributions of this estimator are provided. In addition, an estimator of H, based on pseudo-residuals, is shown to be \log (n)-consistent. All these estimators and findings are then used to carry out a lack-of-fit test of a simple linear regression model using a marked empirical process. The proposed lack-of-fit test is then applied to fit a simple linear regression model to a currency exchange dataset.

### Multivariate Theil-Sen Estimators

❖ Xin Dang, University of Mississippi, Department of Mathematics, P.O. Box 1848, University, MS 38677, *xdang@olemiss.edu*; Hanxiang

Peng, University of Mississippi; Xueqin Wang, Yale University School of Medicine

***Key Words:*** strong consistency, asymptotic normality, nonparametric statistics, multivariate linear regression

As an alternative of Least square estimator (LSE), the Theil-Sen estimator was introduced in a simple linear regression model, which is robust to the outliers and easy to be interpreted in a geometric view. However, the generalization of this estimator to multivariate linear regression have not gained enough attention because of technical difficulty. With the recent development of the depth function which is a method to describe the center of data in the high-dimension space. We use varied depth functions to construct the Theil-Sen estimators for multivariate linear regression model. The strong consistency and the asymptotic behaviors are investigated under modest conditions. The asymptotic relative efficiencies (A.R.E.) and the robustness properties are also compared for different depth functions. simulations are performed to verify the properties of estimators.

## A Goodness-of-Fit Test for Parametric Regression Models When Some Covariates are Missing

❖ Lei Jin, Texas A&M University, 1100 Hensel Drive, Apt. T1L, College Station, TX 77840, *ljin@tamu.edu*; Suojin Wang, Texas A&M University

***Key Words:*** goodness-of-fit, missing data, nonparametric

Several methods have been proposed for testing adequacy of parametric models against nonparametric alternatives. These methods may encounter difficulties when observed data are partially missing. In this paper, we propose a test for linear models when some covariates are partially missing. We investigate its asymptotic properties in comparison to existing methods. Simulation studies are also given to demonstrate the finite sample performances of these methods.

## Estimating Prediction Error in Linear Regression by Cross-Validation

❖ Hui Shen, The University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC V6T 1Z2 Canada, *hshen@stat.ubc.ca*; William J. Welch, The University of British Columbia

***Key Words:*** prediction error, cross validation

Cross validation is commonly used to estimate prediction error. In this paper, we study the properties of the cross validation estimator of mean squared prediction error (CVMSE) in linear regression. We compare the prediction performance for different numbers of folds, v, in cross validation and find that the bias of CVMSE decreases with v increasing. We also propose a correction method to reduce the bias. We compare our correction with that of Burman (1993) through simulated and real examples. Our correction can reduce the bias of CVMSE significantly when the number of parameters in the model is not small relative to the number of observations in the data set. Moreover, we find that the bias correction can help in model selection.

## On Nonparametric Smoothing Methods for Assessing Climate Change

❖ Patricia Menendez Galvan, Swiss Federal Research Institute WSL/ETHZ, Swiss Federal Research Institute WSL, Zuercherstrasse 111, Birmensdorf, CH-8903 Switzerland, *patricia.menendez@wsl.ch*; Sucharita Ghosh, Swiss Federal Research Institute WSL

***Key Words:*** curve estimation, climate change, time series, PALAEO research

The Greenland Ice Core Project (Johnsen et al. 1997) provided data on past environmental conditions of the earth covering the last 250000 years. Such palaeo data are obtained from environmental proxies. For instance, oxygen isotopes trapped in ice are used to estimate past temperature. These long-term records reveal major environmental fluctuations and long-term shifts. Given the scenarios of arbitrary changes in the underlying process, it seems reasonable to use nonparametric smoothing methods to assess different aspects of climate change. In this paper, we expand on the ideas of Ghosh, Beran & Innes (1997) who allow the underlying (conditional) pdf to change over time. We consider one-dimensional transformations of Gaussian processes and examine their ramifications in the context of climate change. We obtain asymptotic results and illustrate our method via examples.

## Evaluation of Spatial Normalization Parameters for SPM: Application to Type 2 Diabetes Data

❖ Bedda Rosario, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, *blr5@pitt.edu*; Scott Ziolko, University of Pittsburgh; Lisa Weissfeld, University of Pittsburgh; Julie Price, University of Pittsburgh

***Key Words:*** type 2 diabetes, statistical parametric mapping, spatial normalization, registration, total entropy

Normalization is the process of warping MRI and PET images of different subjects into a standard anatomical space. Statistical parametric mapping (SPM) software is widely used to perform this process; however, the impact of parameter choice within SPM has not been well studied. We assessed the effects of different parameter value choices (e.g. degree of regularization, elderly or MNI template and number of nonlinear basis functions) on the analysis of MRI and PET image data. Images were spatially normalized for two groups of subjects: Control and Type 2 Diabetics. The effect of parameter choice for SPM t-maps was assessed by location and magnitude of local maxima. Descriptive statistics and paired t tests were used to examine differences in total entropy.

## Statistical Methods for Proportional Hazards Regression with Missing Covariates

❖ Lihong Qi, University of California, Davis, Rowe Program in Human Genetics, One Shields AVE 4458 Tupper Hall, Davis, CA 95616, *lhqi@ucdavis.edu*; Ching-Yun Wang, Fred Hutchinson Cancer Research Center; Ross Prentice, Fred Hutchinson Cancer Research Center

***Key Words:*** case-cohort, kernel smoother, missing covariate data, nested case-control, nonparametric method, weighted estimating equation

Missing covariate data are common in medical studies. In some situations, certain covariates are observed for all study subjects and other covariate data are collected only for a subset. Inconsistent and inefficient estimates can be generated by naively discarding subjects with incomplete data. In this talk, I will present both simple weighted and kernel-assisted fully augmented weighted estimators that use partially incomplete data nonparametrically. The resulting weighted estimators are more efficient than the simple weighted estimator with the inverse of true selection probability as weight. These weighted estimators allow the missing-data mechanism to depend on outcome variables and observed covariates, and they are applicable to various cohort sampling procedures, including case-cohort and nested case-control designs.

# 34 Optimal Experimental Design

Section on Physical and Engineering Sciences

**Sunday, August 6, 2:00 pm–3:50 pm**

## A Catalog of Nonisomorphic Indicator Functions

❖ Shao-Wei Cheng, Academia Sinica, Institute of Statistical Science, 128 Academia Road Section 2, Taipei, 115 Taiwan, *swcheng@stat.sinica.edu.tw*; Chien-Yu Peng, Academia Sinica

***Key Words:*** factorial designs, index sets, j-characteristics

The indicator function of a design A is a polynomial function defined on the design space D such that for each design point x in D, the value of the function is the number of appearances of x in A. For nonregular designs, indicator function plays a role similar to defining contrast subgroup in regular designs. We characterize an indicator function as a combination of two key elements---the index set of monomials with nonzero coefficients and the values of these coefficients---from which the concept of isomorphism for indicator functions is developed. We link the nonzero coefficients and their corresponding contrasts to a system of linear equations whose structure is related to a regular design. By solving different sets of linear equations, a catalog of nonisomorphic indicator functions can be constructed. We will present a catalog obtained from an exhausted computer search.

## Certain Orthogonal Arrays with Generalized Minimum Aberration

❖ Aijun Zhang, University of Michigan, Department of Statistics, 439 West Hall, 1805 South University Ave, Ann Arbor, MI 48109, *ajzhang@umich.edu*

***Key Words:*** experimental design, orthogonal arrays, generalized minimum aberration, optimality

Orthogonal arrays are widely used in industrial experiments. In this talk we discuss some special classes of orthogonal arrays that are optimal under the generalized minimum aberration. The optimality properties include equidistance, weak-equidistance and beyond. Partial existence results of these orthogonal arrays will also be addressed.

## Optimal Fold-over Designs for Three-Level Fractional Factorial Designs

❖ Hong Zhou, University of Memphis, 600 Patterson Street, Apt. 6, MEMPHIS, TN 38111, *hongzhou@memphis.edu*; Manohar L. Aggarwal, University of Memphis; Lih Yuan Deng, University of Memphis; Dennis K. J. Lin, The Pennsylvania State University

***Key Words:*** optimal design, orthogonality, fractional factorial design, optimality criteria, generalized minimum aberration

Fold-over design is standard follow-up experiment commonly used in the practices. Optimal fold-over plans for two-level regular fractional factorial designs were discussed by Li and Lin [2003]. A fold-over design is a design, which combines the initial design and a fold-over plan. A fold-over plan is reversing the signs of one or more columns of the initial design. We extend their ideas into fold-over designs for three-level by adding different fold-over sets to the initial design. The combined designs are the optimal regular designs based on resolution and minimum aberration criteria (Fries and Hunter [1980]). The optimal fold-over designs increases resolution or de-alias the main effect and its interactions. Tables of such fold-over designs are given and some properties are discussed.

## (M,S)-Optimality in Selecting Factorial Designs

❖ Xianggui Qu, Oakland University, Department of Math and Statistics, 2200 Squirrel Road, Rochester, MI 48309, *qu@oakland.edu*; Robert Kushler, Oakland University; Theophilus Ogunyemi, Oakland University

***Key Words:*** fractional factorial designs, orthogonal arrays, minimum aberration, (m,s)-optimality

Use of the (M,S) criterion to select and classify factorial designs is proposed and studied. The criterion is easy to deal with computationally and independent of the choice of treatment contrasts. It can be applied to two-level designs and multilevel symmetrical and asymmetrical designs. An important connection between the (M,S) and minimum aberration criteria is derived for regular fractional factorial designs. Relations between the (M,S) criterion and generalized minimum aberration criteria on nonregular designs also are discussed. The (M,S) criterion is then applied to study the projective properties of nonregular designs.

## Algorithms for Generating Experimental Designs for Irregularly-Shaped Regions

❖ Greg Piepel, Battelle-PNNL, Statistical Sciences K6 08, PO Box 999, Richland, WA 99352, *greg.piepel@pnl.gov*; Nam-Ky Nguyen, University of New England

***Key Words:*** d-optimal, Fedorov exchange algorithm, genetic algorithm, g-optimal, mixture experiment design, optimal design

This presentation focuses on the construction of computer-generated designs on irregularly-shaped regions. Overviews are given of the Fedorov exchange algorithm (FEA) and other exchange algorithms for constructing D-optimal designs. A faster implementation of the FEA is presented, which is referred to as fast-FEA (denoted FFEA). The FFEA was applied to construct D-optimal designs for several published examples with constrained experimental regions. Designs resulting from the FFEA are more D-efficient than published designs, and provide benchmarks for future comparisons of design construction algorithms. The construction of G-optimal designs for constrained regions is also discussed and illustrated with a published example.

## Orthogonal-Maximin Latin Hypercube Designs

❖ Ying Hung, Georgia Institute of Technology, 350187 Georgia Tech Station, Atlanta, GA 30332, *yhung@isye.gatech.edu*; Roshan J. Vengazhiyil, Georgia Institute of Technology

***Key Words:*** computer experiments, kriging, multi-objective optimization, simulated annealing

A randomly generated Latin hypercube design (LHD) can be quite structured: the variables may be highly correlated or the design may not have good space-filling properties. There are procedures to find good LHDs by minimizing correlations or maximizing the inter-site distances. In this article we have shown that these two criteria need not agree each other. In fact, maximization of inter-site distances can result in LHDs where variables are highly correlated and vice versa. Therefore, we propose a multi-objective optimization approach to find good LHDs by combining correlation and distance performance measures. We also propose a new exchange algorithm for efficiently generating such designs. Several examples are presented to show that the new algorithm is fast and the optimal designs are good in terms of both correlations and distances.

## Optimal Design of an Ion Trapping Experiment

❖ Kevin Coakley, National Institute of Standards and Technology, 325 Broadway Street, Statistical Engineering Division, Boulder, CO 80305-3337, *kevin.coakley@nist.gov*

***Key Words:*** asymptotics, dead time, neutron lifetime, stochastic modeling, fundamental physics, optimal design

When a neutron decays, it produces a proton, an electron and an antineutrino. In a neutron lifetime experiment performed at the National Institute of Standards and Technology, a beam of neutrons passes through a detection volume. Each run of the experiment consists of a trapping stage where protons are confined in a trap, and a detection stage where purged protons are detected. In each detection stage, the detector goes dead after the first proton that arrives. I present asymptotically valid formulas for the bias, variance, and mean square prediction error of a maximum likelihood estimate of the proton trapping rate. Based on the value of the flux of neutrons and the duration of the detection stage, I determine the duration of the trapping stage that minimizes the mean square error of the proton trapping rate estimate.

## 35 Causal Models and Causal Effects ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Sunday, August 6, 2:00 pm–3:50 pm**

### On Informative Detection Bias in Screening Studies

❖ Arvid Sjölander, Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, Stockholm, 17177 Sweden, *arvid.sjolander@ki.se*; Juni Palmgren, Karolinska Institutet

***Key Words:*** informative detection bias, directed acyclic graph, counterfactuals, sensitivity analysis, bounds

A problem which has drawn some attention in the causal literature is to estimate the causal effect of hormone replacement therapy (HRT) on breast cancer. Since HRT is considered a risk factor for breast cancer HRT users are more likely to have a mammogram taken than non users. The observed discrepancy in cancer prevalence between users and non users is thus likely to suffer from detection bias. Standard statistical control for this bias, such as stratification on screening pattern, will severe the bias whenever screening pattern and disease status is confounded. We present the problem within the standard counterfactual framework. A sensitivity analysis is performed and bounds on the causal HRT effect are derived. We compare this method with the standard statistical approach.

### Estimating a Class of Causal Treatment Effect for Survival Data

❖ Jing Ning, Johns Hopkins University, Department of Biostatistics, 615 N. Wolfe Street, Biostatistics Department, Baltimore, MD 21205, *jning@jhsph.edu*; Mei-Cheng Wang, Johns Hopkins University; Zhiqiang Tan, Johns Hopkins University

***Key Words:*** treatment effect, proportional hazards model, accelerated failure time model, causal time ratio, causal hazards ratio

To identify a treatment effect by time-to-event data, the proportional hazards model and accelerated failure time model have been widely used, which need strong model assumptions. In randomized clinical trial, the feature of randomization exists only at the initial time. In this talk, we define a class of causal treatment effect to overcome this problem due to the loss of the comparability of two groups. The treatment effects are compared with treatment effects arising out of proportional hazards model and accelerated failure time model. The asymptotic distribution of the estimator is derived. Real data from a randomized clinic trial, which aims to evaluate the treatment effect of brain malignant gliomas, are used for illustration.

### Path Analysis for Ordinal Variables

❖ Haihong Li, University of Florida, P.O. Box 100177, 1329 SW 16th Street Suite 5130, Gainesville, FL 32610, *hli@biostat.ufl.edu*; P. V. Rao, University of Florida

***Key Words:*** path analysis, structural equations modeling, categorical data

Path analysis is one simple kind of the structural equations modeling methods. The main idea is that it decomposes the interrelationships among variables into direct and indirect effects. However, most literatures discuss path analysis only for variables that are normally distributed, while in many applications categorical variables are used. In this study we attempt to quantify the decomposition of effects for ordinal variables and investigate the associated statistical inferences. The maternal child health data in the state of Florida is used for illustration.

### Approaches to Obtaining Standard Errors for Parameter Estimates in Latent Class Analysis

❖ David M. Thompson, The University of Oklahoma, 825 NW 49th Street, Health Sciences Center, Dept. of Biostatistics and Epidemiology, Oklahoma City, OK 73118, *dave-thompson@ouhsc.edu*

***Key Words:*** latent class analysis, standard errors, E-M algorithm, SAS-IML, PROC Catmod

Latent class analysis (LCA) has attracted the interest of clinical professionals who must place clients in diagnostic or prognostic categories when a gold standard for doing so is poorly defined. However, the classic MLE approach to LCA employs an expectation-maximization algorithm that does not yield standard errors. SAS-IML or SAS PROC CATMOD's loglinear modeling facility permit LCA approaches that open up strategies for obtaining standard errors. These include repetitive analyses using a grid of initial estimates, or conversion of CATMOD's loglinear expressions for SE into probabilities. The presentation addresses data in which four binary indicators permit estimation of a two-class latent structure. CATMOD's flexibility in loglinear modeling potentially allows estimation of larger models, including ones that accommodate residual local dependence among indicators.

### An Application of Multivariate Path Models and the Calculus of Coefficients to Describe Effects of Health Behaviors on the Metabolic Syndrome

❖ Youngju Pak, University at Buffalo, Department of Biostatistics, School of Public Health and Health Professions, 249 Farber, 3435 Main st, Bldg 26, Buffalo, NY 14214-3000, *ypak@buffalo.edu*; Randy L. Carter, University at Buffalo

***Key Words:*** causal models, direct and indirect effects, intersection-union tests, union-intersection tests, cardiovascular disease, diabetes

Path analysis is useful to explain the interrelationships among variables in a causal chain. However, path analysis has been underutilized, especially in the health sciences because it assumes a complete causal ordering. We suggest multivariate path modeling as a solution. The Calculus of Coefficients for univariate path models partitions total effects into sums of direct and all possible indirect effects through intermediate variables. We extend this important result to multivariate models when groups of intermediate variables can be causally ordered but variables within groups cannot be. We suggest an application of combinations of union-intersection and intersection-union tests for testing indirect effects. The methods are applied to data from the Western New York Health Study to describe the effects of lifetime alcohol consumption and physical activity on a metabolic syndrome index.

### Signed Directed Acyclic Graphs for Causal Inference

❖ Tyler J. VanderWeele, Harvard School of Public Health, 63 Mt. Vernon Street, 6, Cambridge, MA 02140, *tvanderw@hsph.harvard.edu*; James Robins, Harvard School of Public Health

***Key Words:*** bias, causal inference, confounding, directed acyclic graphs, structural equations

By introducing the notions of a monotonic effect, a weak monotonic effect, and a signed edge, the directed acyclic graph causal framework can be extended to allow not only for the graphical representation of causal relations among variables, but also for the sign of these causal relations. Results are developed relating monotonic effects to the sign of the causal effect of an intervention in the presence of intermediate variables. Further, the incorporation of signed edges into the directed acyclic graph causal framework allows for the development of rules governing the relationship between monotonic effects and the sign of the covariance between two variables and rules governing the sign of the bias that arise when control for confounding is inadequate.

## 36 Unit Nonresponse in Surveys I

●

Section on Survey Research Methods
**Sunday, August 6, 2:00 pm–3:50 pm**

### Nonresponse to a Computer-Assisted Self-Interviewing (CASI) Module

❖ Mick Couper, University of Michigan, 3193 Lakewood Drive, Ann Arbor, MI 48103-2041, *mcouper@umich.edu*; Eleanor Singer, University of Michigan; John Van Hoewyk, University of Michigan

***Key Words:*** nonresponse, CASI, self-administration

The 2004 GSS included a self-administered module with a series of vignettes eliciting attitudes toward genetic versus environmental influences on various attributes and behaviors. Of 2,812 respondents, 15% did not complete the CASI module, while a further 26% completed it with interviewer assistance. All respondents were also asked a series of general questions about genetics in the interviewer-administered part of the survey, allowing us to explore possible nonresponse bias or self-administration effects in the CASI module. We examine the correlates of nonresponse to the CASI module, and fit a series of multinomial logistic regression models to explore the differences between those who self-completed, those whom the interviewer assisted, and those who did not complete the module. We also examine substantive differences between self-completers and those assisted by the interviewer.

### Assessing the Effectiveness of Weighting Cell Adjustments for Longitudinal Nonresponse

❖ Leroy Bailey, U.S. Census Bureau, 4401 Suitland Road, Rm 3134 4, Suitland, MD 20746, *leroy.bailey@census.gov*

***Key Words:*** longitudinal nonresponse, nonresponse weighting cells, model-based reweighting.

In this paper longitudinal respondents are considered as members of a panel survey sample for whom data are collected every wave of a given period under consideration, and members of the panel for whom there is nonresponse for at least one wave of the designated period are termed longitudinal nonrespondents. To compensate for the effects of wave nonresponse and attrition, the weights of the longitudinal respondents can be adjusted. This paper will discuss criteria and strategies for the selection of adjustments for cross-sectional estimates, and provide the theoretical framework for an evaluation of selected weighting alternatives designed to account for wave nonresponse in the estimation of longitudinal variables. Empirical results from the 1996 and 2001 panels of the Survey of Income and Program Participation are presented for two weighting cell adjustment alternatives.

### Sample Reweighting To Reflect an Initial Population

❖ Julia Bienias, Rush University Medical Center, 1645 W. Jackson Blvd., Suite 675, Chicago, IL 60612, *jbienias@rush.edu*; Phillip S. Kott, National Agricultural Statistics Service; Todd L. Beck, Rush University Medical Center; Denis A. Evans, Rush University Medical Center

***Key Words:*** complex surveys, Poisson sampling, finite-population inference, model-based inference, longitudinal

The Chicago Health and Aging Project is a complex longitudinal community-based study which examines risk factors for chronic health problems of older adults. Every three years we draw a stratified Poisson sample for detailed clinical evaluation. We wish to combine these samples and represent the complete population as it existed at the baseline of the study. We propose creating model-dependent sample weights, treating these later sampled observations as though they were actually obtained at baseline but the data were "delayed." Such an approach has an implicit non-response adjustment, and we discuss the ramifications of this as well as using alternative adjustments for non-participation and mortality when creating our weights. These methods are illustrated with several predictors of interest in aging research and for outcomes such as mortality. Supported by NIH grant AG11101.

### Approaches to Nonresponse Bias Analysis in an Adult Literacy Survey

❖ Wendy Van de Kerckhove, Westat, 1650 Research Blvd., RE463, Rockville, MD 20850, *wendyvandekerckhove@westat.com*; Thomas Krenzke, Westat; Leyla Mohadjer, Westat

***Key Words:*** response rate evaluation, outcome, assessment, disposition codes

In a standard nonresponse bias analysis, respondents and nonrespondents are compared on available demographic information. This type of analysis only provides an indication of nonresponse bias to the extent the auxiliary demographic variables are related to the outcome statistics of the survey. This paper describes approaches to performing a more extensive nonresponse bias analysis using data directly related to the survey statistics. The approaches include an assessment of the relationship between auxiliary variables and the survey statistics, a comparison of the statistics for different types of respondents, and an analysis of nonrespondents whose reasons for nonresponse correlate with survey statistics. Specific approaches used in the 2003 National Assessment of Adult Literacy (NAAL) survey are provided for illustration.

### An Application of Propensity Modeling To Adjust Weights for Nonresponse: Effectiveness of Restricting Variables and Propensity Values

❖ Frank Potter, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08543, *FPotter@mathematica-mpr.com*; Nuria Diaz-Tena, Mathematica Policy Research, Inc.; Stephen R. Williams, Mathematica Policy Research, Inc.

***Key Words:*** non-response, weighting, propensity modeling, survey bias and precision, community tracking study

Adjusting for nonresponse can use logistic regression models to predict the probability that a unit will respond. The predicted propensities can reflect more predictive variables than in the weighting class method and the inverse of the propensity can be the weight adjustment factor. Having used this method for rounds two and three of a large physician survey, this paper describes the results from round four. The independent variables used in round four are expanded to include design variables, basic sampling weights, and higher-order interactions. Predictive power of the propensity models were substantially improved but also presented some interesting issues. The more effective models produce more extreme adjustment factors. We consider whether or not to restrict the nonresponse ranges of candidate variables and whether the extreme adjustments or the final weights should be trimmed.

### Adjusting for Nonignorable Missing Data with Nonignorable Sampling Design in Longitudinal Sample Survey

❖ Moh Yin Chang, University of Nebraska-Lincoln, Gallup Research Center, 200 N 11th St, Lincoln, NE 68588-0241, *mohyin@unlserve.unl.edu*

***Key Words:*** nonignorable missing data, informative missingness, sampling weights, sampling design, longitudinal data, survey

Current nonignorable missing data literature has overlooked the issue of complex sampling schemes in longitudinal sample surveys. Considering missing data and sampling designs concurrently has important practical implications. In this study, I will extend current models for nonignorable missing data to incorporate the informative sampling design in longitudinal survey analysis. Empirical and simulation studies will be performed to compare the results of these estimations.

### Estimated Response Propensities as a Means To Evaluate Error Effects Due to Nonresponse

❖ Leela Aertker, The University of North Carolina at Chapel Hill, 401-A Coolidge Street, Chapel Hill, NC 27516, *leela@email.unc.edu*; William D. Kalsbeek, The University of North Carolina at Chapel Hill

***Key Words:*** surveys, nonresponse, bias, response propensity

In addition to providing a basis for the weighting adjustment for nonresponse, individual response propensities obtained from a logistic model may offer a means by which nonresponse bias effects can be routinely estimated in samples where useful auxiliary data are available for response propensity modeling. Using data from the multi-round National Longitudinal Study of Adolescent Health (Add Health), we present measures of nonresponse error effects and direct estimates of round-specific nonresponse bias. We further assess the utility of estimated response propensities as a direct mechanism to estimate nonresponse bias.

## 37 Estimation and Confidentiality

●

Section on Survey Research Methods, Section on Health Policy Statistics
**Sunday, August 6, 2:00 pm–3:50 pm**

### Disclosure Avoidance for the 2007 ACS PUMS: a Model-Based Approach for Group-Quarters Data

❖ Rolando Rodriguez, U.S. Census Bureau, 4401 Suitland Road, RM 3132A, Suitland, MD 20764, *rolando.a.rodriguez@census.gov*

***Key Words:*** confidentiality protection, disclosure avoidance, ACS, microdata, group quarters

In 2007 the Census Bureau will release ACS PUMS data with group quarters (GQ) records. GQ consists of institutional facilities such as prisons and non-institutional facilities such as dormitories. We are developing a model-based approach to create partially-synthetic data for records at risk of identity disclosure. This contrasts with methods like swapping, which substitute or rearrange records. Modeling may more accurately reproduce the data distributions and provide better protection. We can implement models using information from a single geographically-isolated institution. Other methods must pool information from several institutions to insure proper protection. Our approach has several stages: flagging records at risk, modeling, verifying the resulting distributions, and analyzing the improvements in confidentiality. We discuss this process along with GQ-specific issues.

### Reporting to Payers, Regulators, and Managers: Issues and Experiences with Confidentiality and Compliance

❖ Richard Carlson, Medica, Mail Route CP455, 401 Carlson Parkway, Minnetonka, MN 55407, *rick.carlson@medica.com*

***Key Words:*** confidentiality, HIPAA, compliance

When dealing with Protected Health Information care must be maintained to maintain confidentiality. At the same time demands are made for reporting activity and at levels that may cause risk for confidentiality. This paper describes some of the experiences and methods used to maintain compliance.

### The Change-of-Variance Function in Generalized Linear Mixed-Effect Models with Applications to Poisson-Gamma and Beta-Binomial Models

❖ Gabriela Cohen-Freue, The University of British Columbia, 2730 Acadia Road, 203, Vancouver, BC V6T 1R9 Canada, *gcohen@stat.ubc.ca*

***Key Words:*** change-of-variance function, v-robustness, generalized mixed effect models, beta-binomial models, Poisson-gamma model

This paper extends the Change-of-Variance function to Generalized Linear Mixed effect Models. We use this function to examine the sensitivity of the asymptotic variance of the marginal maximum likelihood estimator under a slight contamination of the mixing distribution. The Poisson-Gamma model and two Binomial mixed effects models, attractive for their distributional closed form and applicability, are analyzed in detail. For these particular models, we find that the MLE is V-robust when the mixing distribution is contaminated by any other distribution sharing the first two moments. A simulation study is performed for the Poisson-Gamma model. Although the asymptotic variance remains bounded when the mixing distribution is contaminated, the bias of the estimates of some parameters increases while the variance decreases, leading to erroneous inference about the unknown parameters.

### A Bridge between the Greg and the Linear Regression Estimators

❖ Sarjinder Singh, St. Cloud State University, Department of Statistics, St Cloud, MN 56301, *sarjinder@yahoo.com*; Raghunath Arnab, University of Botswana

***Key Words:*** Greg, linear regression estimator, estimation of total and variance

We discovered there is a choice of weights that builds a bridge between the GREG proposed by Deville and Sarndal (1992) and the linear regression estimator due to Hansen, Hurwitz, and Madow (1953) while using one auxiliary variable. It gives the same result as given in Singh (2003, 2004) and Stearns and Singh (2005) for unequal probability sampling by using two calibration constraints in the presence of one auxiliary variable. Thus, these approaches can be considered as alternative to each other while considering use of one auxiliary variable. The bridge developed here reconfirms that the sum of the calibrated weights should be equal to the sum of the design weights in a given sample. The approach by Singh (2003, 2004) and Stearns and Singh (2005) seems simple while using multiauxiliary information.

### A Generalized Forced Quantitative Randomized Response Model

❖ Oluseun Odumade, St. Cloud State University, Department of Statistics, St Cloud State University, St Cloud, MN 56301, *odol0201@stcloudstate.edu*; Sarjinder Singh, St. Cloud State University

***Key Words:*** randomized response sampling, estimation of population total, sensitive quantitative variable

A generalized forced quantitative randomized response (GFQRR) model for estimating the population total of a sensitive quantitative variable is proposed. The properties of the proposed estimator of the population total are studied under a unified approach. The models, due to Eichhorn and Hayre (1983); Bar-Lev, Bobovitch, Boukai (2004); Liu and Chow (1976a, 1976b); Stem and Steinhorst (1984); and Gjestvang and Singh (2005) are shown as special cases of the proposed GFQRR model. The performance of the proposed GFQRR model has been demonstrated with numerical illustrations.

### Global and Hierarchical Linear Regression in Two-Stage Sampling

❖ Dhirendra Ghosh, Synectics for Management Decisions, Inc., 1901 N. Moore Street, Suite 900, Arlington, VA 22209, *vogta@georgetown.edu*; Andrew Vogt, Georgetown University

***Key Words:*** cluster sampling, least squares, regression estimates

In a two-stage sampling design where the second stage units have a known independent variable and the survey collects data on a dependent variable, we compare different forms of hierarchical linear regression, as well as global linear regression (where the first stage units, or clusters, are ignored). We consider cases where each cluster in the sample has a large sample size with or without an auxiliary cluster variable; where each cluster has a small sample size and an auxiliary cluster variable; and where the cluster variable is unknown for clusters not in the sample.

### Confidentiality in Survey Data: the Lack of Consistent Standards

❖ M. Leeann Habte, University of California, Los Angeles, Center for Health Policy Research, 10911 Weyburn Ave., Los Angeles, CA 90024, *lhabte@ucla.edu*; Hongjian Yu, University of California, Los Angeles; Jenny Chia, University of California, Los Angeles; Brandon Traudt, University of California, Los Angeles

***Key Words:*** privacy, confidentiality, statistical disclosure limitation, disclosure risk estimation, survey data, statistics

Recent advances have been made in the development and refinement of statistical formulas to estimate the risk of disclosure of confidential information. What is lacking, however, is a clear standard for what constitutes adequate confidentiality protection for survey data. Interpretations of state and federal regulations by the Courts and by Institutional Review Boards (IRBs) have been inconsistent on whether the standard for confidentiality is one of anonymity or whether it is sufficient that the identity of a survey respondent is not reasonably ascertainable. The applicability and utility of disclosure limitation methods depends on the nature of the data and the standard to be met. This paper will explore some of the different ways that confidentiality has been interpreted and discuss the applicability of risk estimation methods to divergent confidentiality standards.

## 38 Introductory Overview Lectures: Adaptive Designs/Interim Pilots and Regression Trees

The ASA, ENAR, IMS, SPAIG Committee, WNAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### Regression Trees

❖ Wei-Yin Loh, University of Wisconsin-Madison, Department of Statistics, 1300 University Avenue, Madison, WI 53706, *LOH@STAT.WISC.EDU*

***Key Words:*** prediction, machine learning, recursive partitioning, data mining

This is an introduction to the modern subject of regression tree modeling. A regression tree is a piecewise constant or piecewise linear model constructed by recursively partitioning a dataset. Because the number of partitions automatically adapts to the complexity of the data, these models can possess very good prediction accuracy. More importantly, the tree structure can yield information about the variables and the data. The lecture will introduce the speaker's GUIDE and LOTUS algorithms and illustrate them with examples from small designed experiments to large databases, and from least squares to quantile, Poisson, logistic, and relative risk regression. Comparisons with CART and M5 will also be presented.

### Adaptive and Internal Pilot Designs

❖ Christopher S. Coffey, The University of Alabama at Birmingham, 327M RPHB, Department of Biostatistics, 1665 University Boulevard, Birmingham, AL 35294-0022, *ccoffey@uab.edu*

***Key Words:*** sample size re-estimation, power, clinical trials, group sequential methods

The strong desire to modify the sample size to reflect interim results from a clinical trial has stimulated considerable recent research on adaptive designs. Such designs permit changing the study when interim data disagree with the assumptions upon which the original sample size calculations were based. The rapid proliferation of adaptive designs, and inconsistent use of terminology, has created confusion about the similarities and, more importantly, the differences among the techniques. This introductory overview will clarify the definition of an adaptive design and describe distinctions among unblinded sample size re-estimation based on: 1) interim estimates of the primary effect size and 2) nuisance parameters (internal pilots). Finally, the overview will compare and contrast the motivation for implementing adaptive designs versus information based group sequential methods.

## 39 Statistics in Biotechnology around the Puget Sound

ASA, Puget Sound Chapter, Section on Statistical Graphics
**Sunday, August 6, 4:00 pm–5:50 pm**

### Reference Samples and Other Low-Level Choices for the Design and Analysis of Two-Color Microarray Experiments

❖ Kathleen Kerr, University of Washington, Box 357232, Seattle, WA 98195, *katiek@u.washington.edu*

***Key Words:*** microarray, experimental design, normalization

The reference design is a practical and popular choice for microarray studies using a two-color platform. In this design, the reference RNA uses half of all array resources, leading investigators to ask: What is the best reference RNA? Previous papers on reference RNAs have asserted that the best reference RNA has broad representation of the genes on the array. We question the priority of this quality of a reference RNA and propose alternative criteria. We present the results of an experiment that was specially designed to evaluate three common choices of reference RNA, including a commercial "universal" reference and a pooling strategy. We find that a pooling strategy is very effective and the commercial reference showed no advantage in our data. Our experimental design also allowed some novel evaluation of pre-processing choices for array data.

### Graphs and Networks in Computational Biology

❖ Robert Gentleman, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, *rgentlem@fhcrc.org*

***Key Words:*** computational biology, graph, network

Graphs and networks play important roles in computational biology. In this talk, we consider protein-protein interactions and some of the statistical problems that arise when trying to estimate an interactome. We also consider a number of applications for an in silico interactome.

### Statistical Methods for Integrating High-Dimensional Genotype, Molecular Profiling, and Clinical Data To Elucidate Human Disease

❖ Eric Schadt, Rosetta Inpharmatics LLC/Merck Research Laboratories, Seattle, WA 98109, *eric_schadt@merck.com*

***Key Words:*** genetics, gene expression, integrative genomics, gene networks

The development of statistical methods focused on the analysis of genetic or molecular profiling data has been extensive in the last 10 years. A number of recent studies have demonstrated the power of integrating genotypic, molecular profiling and clinical data to elucidate human disease and drug response and have resulted in the identification of a number of genes and gene networks associated with these complex traits. Underlying these integrative genomics approaches are a number of fairly simple statistical procedures that simultaneously consider orthogonal data from a number of sources. We describe these novel approaches that take a more holistic view of biological systems (versus a reductionist view), characterize the extent of discovery resulting from their application, and discuss ways to enhance the statistical methods to achieve even greater power in such studies.

## 40 Statistical Issues in Genetic Association Studies

General Methodology, Biometrics Section, ENAR
**Sunday, August 6, 4:00 pm–5:50 pm**

### Family Studies in the Age of Association

❖ Nan M. Laird, Harvard School of Public Health, 677 Huntington Ave., Biostatistics, Boston, MA 02115, *laird@hsph.harvard.edu*

***Key Words:*** genome scans, multiple testing, family-based designs

Samples of families have long played a fundamental role in genetic studies of human disease. With the advent of SNPs and the possibility of testing thousands of SNPs in candidate regions, association analysis has taken a front row seat among methods for gene mapping. Case-control and cohort studies are favored for association studies because of their cost effectiveness and ready availability, but recent developments in methods for family studies make them competitive. We present an application to a 100K scan of the Framingham Heart Study.

### Hybrid Vigor: Family-Based and Population-Based Designs Can Work Together

❖ Clarice R. Weinberg, National Institute of Environmental Health Sciences, MD A3-03, P O Box 12233, Research Triangle Park, NC 27709, *weinber2@niehs.nih.gov*; David M. Umbach, National Institute of Environmental Health Sciences

***Key Words:*** genetic association studies, hybrid design, gene-by-environment interaction, case-control studies, case-parent studies

Both family- and population-based designs can provide powerful inference for studying associations between genetic variants and risk, but both have major limitations. A hybrid design, best for rare diseases with onset early in life, augments a set of cases and their parents with population-based controls and their parents. The cases and their parents are genotyped, but only parents of controls are genotyped, while exposures are ascertained for both. A log-linear analysis allows estimation of relative risk parameters, provides more power than either the case-parents approach or a case-mother/control-mother approach, allows for missing genotypes, permits direct testing for bias due to mating asymmetry or population stratification, and permits valid alternative analyses when such bias is in evidence. Gene-by-environment interaction also can be assessed more efficiently using this design.

### Analysis of Complex Pathways in Molecular Epidemiology

❖ Duncan C. Thomas, University of Southern California, Department of Preventive Medicine, 1540 Alcazar St. CHO-220, Los Angeles, CA 90089-9011, *dthomas@usc.edu*

***Key Words:*** pathways, gene networks, molecular epidemiology, pharmacokinetic models, statistical genetics

Molecular epidemiology is evolving rapidly from consideration of single genes (or in pairwise combinations with each other or with a single environmental factor) toward entire pathways or networks involving many factors in combination. Study designs and analysis methods for pathway-driven research are still in their infancy. Several approaches to this problem will be discussed, ranging from purely exploratory tools to hierarchical models to highly mechanistic pharmacokinetic models. A major concern is with uncertainty about model choice (the 'topology' of a network), for which Bayesian model averaging may be useful. Particularly attractive are approaches that allow for incorporation of prior biological knowledge or measurements of intermediate biomarkers. By simulation and application to epidemiologic data, the potential of various modeling approaches will be explored.

## 41 Statistical Effect Assessment of Environmental Exposure ● ✪

ENAR, Biometrics Section, WNAR, Section on Statistics and the Environment

**Sunday, August 6, 4:00 pm–5:50 pm**

### Bayesian Models for Multiple Outcomes Nested within Domains

❖ Sally W. Thurston, University of Rochester Medical Center, Department of Biostatistics, 601 Elmwood Ave # 630, Rochester, NY 14642, *thurston@bst.rochester.edu*; David Ruppert, Cornell University

***Key Words:*** methylmercury, multiple outcomes, Seychelles

Understanding the human health effects of exposure to a toxicant is complicated when the toxicant affects multiple endpoints. Furthermore, the endpoints may cluster into broad domains such that similarity of the exposure effects on outcomes within a domain is expected. An example of this situation is a study of the effects of prenatal methylmercury (MeHg) exposure on children in Seychelles, where the 21 neurodevelopmental outcomes measured at nine years cluster within seven domains. We discuss results of a Bayesian model that allows the effect of MeHg exposure to vary across outcomes, while shrinking the outcome- and domain-specific effects to the extent warranted by the data. Our model allows us to estimate these effects as well as an overall exposure effect. By analyzing several outcomes simultaneously, one gains power to detect small, but potentially important, health effects.

### Analysis of Multivariate Longitudinal Data Using Structural Equation Models

❖ Esben Budtz-Jorgensen, University of Copenhagen, , *ebj@biostat.ku.dk*; Philippe Grandjean, Harvard University; Frodi Debes, University of Southern Denmark; Pal Weihe, Faroese Hospital System

***Key Words:*** environmental epidemiology, measurement error, multiple endpoints, mercury exposure

In prospective cohort studies, multivariate data often are collected longitudinally both on the exposure and the response. Standard statistical procedures are poorly suited for analysis of such complex data, and, typically, a multiple regression or mixed model is fitted to only a subset of the data. A superior analysis may be obtained in structural equation models. In these models, multivariate endpoints are considered to be manifestations of causally related latent variables. This structure may provide a parsimonious and more powerful representation of the exposure effect while properly accounting for measurement errors and the complex correlation structure typically encountered in such data. These methods are illustrated by analyzing the effect of mercury exposure on neurobehavioral test scores in Faroese children at 7 and 14 years.

### Double-Smoothing Local Linear Estimation in Partial Linear Models with Application to Environmental Health Data

❖ Li-Shan Huang, University of Rochester, 601 Elmwood Ave., Box 630, Dept of Biostatistics, Rochester, NY 14642, *Lhuang@bst.rochester.edu*; Christopher Cox, Johns Hopkins University

***Key Words:*** nonparametric regression, partial linear models, local linear regression, hypothesis testing, environmental health

New estimators and F-type hypothesis testing procedures are proposed for partial linear models. The new estimator for the nonparametric component is based on the local linear estimator with "double smoothing." The proposed F-tests formally evaluate significance of the nonparametric component by testing a no-effect null hypothesis and whether the nonparametric function can be simplified to a linear relationship. The Seychelles Child Development Study is an ongoing longitudinal study of child development after prenatal exposure to methylmercury through maternal fish consumption. A possible nonlinear effect of methylmercury exposure was seen in Huang et al. (2005), but its significance was not confirmed. We apply partial linear models to the Seychelles data with the proposed new estimators and formally examine the significance of the nonlinear exposure effect.

### Synthesizing Data from Multiple Sources for Environmental Risk Assessment

❖ Louise Ryan, Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115, *lryan@hsphmail.harvard.edu*

We discuss statistical issues associated with the synthesis of data from multiple studies involving multiple endpoints. Hierarchical models are shown to provide a useful and flexible framework for analysis in such settings and provide a capacity for sensitivity analysis and assessing model uncertainty. The methods are applied for a case study in methylmercury.

## 42 Graphical Models and Variational Methods

IMS, Section on Bayesian Statistical Science

**Sunday, August 6, 4:00 pm–5:50 pm**

### Variational Methods for Dirichlet Process Mixtures

❖ David M. Blei, Princeton University, , *blei@cs.princeton.edu*; Michael I. Jordan, University of California, Berkeley

***Key Words:*** nonparametric Bayes, Gibbs sampling, variational methods, mixture models

Dirichlet process (DP) mixture models are the cornerstone of nonparametric Bayesian statistics, and the development of Monte Carlo Markov chain (MCMC) sampling methods for DP mixtures has enabled the application of nonparametric Bayesian methods to a variety of practical data analysis problems. However, MCMC sampling can be prohibitively slow, and it is important to explore alternatives. One class of alternatives is provided by variational methods, a class of deterministic algorithms that convert inference problems into optimization problems. Thus far, variational methods have been explored mainly in the parametric setting, within the formalism of the exponential family in particular. In this talk, I will present a variational inference algorithm for DP mixtures. I will compare the algorithm to Gibbs sampling and present an application to a large-scale image analysis problem.

### Structured Prediction, Dual Extragradient, and Bregman Projections

❖ Ben Taskar, University of California, Berkeley, 485 Soda Hall,
Computer Science, Berkeley, CA 94720, *taskar@cs.berkeley.edu*

***Key Words:*** Markov random fields, m-estimation, Bregman projections, large-scale optimization

We present a simple and scalable algorithm for M-estimation of structured output models, including important classes of Markov networks and combinatorial models. We formulate the estimation problem as a convex-concave saddle-point problem that allows us to use simple projection methods based on the dual extragradient algorithm (Nesterov, 2003). The projection step can be solved using dynamic programming or combinatorial algorithms for min-cost convex flow, depending on the structure of the problem. We show that this approach provides a memory-efficient alternative to formulations based on reductions to a quadratic program (QP). We analyze the convergence of the method and present experiments on two very different structured prediction tasks---3D image segmentation and word alignment---illustrating the favorable scaling properties of our algorithm.

### A Variational Inference Procedure Allowing Internal Structure for Overlapping Clusters and Deterministic Constraints

❖ Christopher Meek, Microsoft Research, 1 Microsoft Way,
Redmond, WA 98052-8300, *meek@microsoft.com*; Dan Geiger,
Technion-Israel Institute of Technology

***Key Words:*** variational methods, Bayesian networks, genetic linkage analysis

We develop a novel algorithm, called VIP*, for structured variational approximate inference. This algorithm extends known algorithms to allow efficient multiple potential updates for overlapping clusters and overcomes the difficulties imposed by deterministic constraints. The algorithm's convergence is proven and its applicability demonstrated for genetic linkage analysis. This is joint work with Chris Meek and Ydo Wexler.

## 43 Statistical Methods in HIV/AIDS Research ●

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### Methods for Determining the Accuracy of Quantitative PCR for Low Levels of HIV-1

❖ Barbra Richardson, University of Washington, Harborview
Medical Center, Department of Biostatistics, Box 359909, 325 Ninth
Avenue, Seattle, WA 98104-2499, *barbrar@u.washington.edu*

***Key Words:*** sensitivity, zero-inflated distributions, HIV, laboratory assays

Quantitative polymerase chain reaction (PCR)-based tests are used in several scientific fields to determine levels of a target DNA sequence of interest (the target molecule). To assess the accuracy of these tests, a sample with a known concentration of a target DNA sequence is diluted serially into replicate aliquots and tested to determine if the observed quantity of the target is close to the expected quantity. Statistical methods conventionally used to assess the accuracy of these assays do not take into account the variability in the number of target molecules in each aliquot from the original sample. We develop methods that take this into account and determine the accuracy of quantitative PCR-based tests. These methods are applied to data from an experiment to test the accuracy of a real-time PCR assay at low HIV-1 DNA copy levels.

### A Bernoulli/Left-Censored Lognormal Mixture Model for Activity of the Protease of HIV-1 as a Function of Amino Acid Characteristics

❖ Paul W. Stewart, The University of North Carolina at Chapel Hill,
McGavran-Greenberg Hall, Biostatistics Department, Chapel Hill,
NC 27599-7420, *paul_stewart@unc.edu*

***Key Words:*** HIV, mixture model, protease, AIDS, left censored

In a recent study by the UNC CFAR, we developed a mixture model for mean level of HIV-1 protease activity as a function of structural characteristics of the protease. The protease is an enzyme required for HIV replication. Previous protease research resulted in a class of FDA-approved drugs: protease inhibitors. As mutations of the protease occur, various amino acids are substituted for the amino acids usually found in the protease. The choices of amino acids in key positions have been manipulated experimentally, yielding observed levels of protease activity that range from 0% to 100%, subject to a lower limit of detection. Some configurations of the protease rendered activity impossible. A statistical model developed for these data will be described and issues of estimation, inference, and diagnostics will be discussed.

### Evaluating Linked Substitutions in HIV Genomic Sequences

❖ Francoise Seillier-Moiseiwitsch, Georgetown University
Medical Center, Suite 180, Building D, 4000 Reservoir Road, N.W.,
Washington, DC 20057, *seillier@georgetown.edu*; Huwaida Rabie,
Georgetown University Medical Center; Rebecca Slack, Georgetown
University Medical Center; JaeHyung Ahn, The University of North
Carolina at Chapel Hill; Gary Koch, The University of North Carolina
at Chapel Hill

***Key Words:*** HIV, genome, substitutions, independence, conditional independence

The HIV genome evolves rapidly. The inefficiency of its replication process gives rise to many variants. The observed variability reflects both mutant viability and selection pressures from the immune system. Detecting linked mutations may provide insights into protein structures and genome evolution. Statistical methodology applied so far pertains to testing the statistical independence between two positions. The relationship between the lack of statistical independence and mutational linkage will be discussed. Statistical procedures developed specifically to detect linkage and excess of double mutations will be presented. Generalizations of the methodology to triple and quadruple mutations also will be described. We will explore different concepts of independence in higher dimensions. We apply these tests to sequences from the V3 loop of the gene coding for the envelope protein.

### A Comprehensive Mathematical Model of HIV/STD Spread in Communities

❖ Georgiy V. Bobashev, RTI International, 3040 Cornwallis Road, Durham, NC 27709, *bobashev@rti.org*; Michael Goedecke, RTI International; Elizabeth Costenbader, RTI International; Wiliam Zule, RTI International

***Key Words:*** epidemiology, mathematical modeling, community, HIV, intervention, survey sampling

We present an approach that allows one to build a comprehensive mathematical model for simulation of disease spread based on combined survey and published data. We illustrate the model-building process---including the selection process for known and unknown parameters---with a comprehensive model of HIV/STD spread within sexual, drug, and social networks in a North Carolina community. The model is agent-based in as much as microsimulation occurs at the level of each individual in the model. The parameters are obtained partially from peer-reviewed sources and estimated from a 480+ question survey, which currently is being conducted among a high-risk population in Eastern North Carolina. The model provides a basis for answering a number of research questions related to sampling, epidemic course, critical parameters and interventions, interplay of hierarchical factors, etc.

## 44 Global Views on the Role of Statistics in Medical Device Regulation ●

Biopharmaceutical Section, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### Statistical Regulations in the EU: Do They Exist for Medical Devices?

❖ Bart Gerritse, Medtronic, Inc., Endepolsdomein 5, Maastricht, NL-6229GW The Netherlands, *bart.gerritse@medtronic.com*

***Key Words:*** medical devices, regulations, Europe

There are fundamental differences between clinical trials with medical devices and trials researching pharmacological treatment. To name two examples: in device trials, adverse effects often are attributable to the therapy with greater certainty and blinding is not possible. Reflecting these differences, laws and guidelines treat devices separately. In the European Union, market approval for medical devices is governed by directives on medical devices (MDD), active implantable medical devices (AIMDD), and in vitro diagnostic medical devices (IVDMD). Member states are obliged to implement these in national laws. Additionally, ISO 14155 and MEDDEV 2.7 provide guidelines for the conduct of device trials and evaluation of clinical data. This presentation will review statistical aspects of the directives and guidelines.

### Statistics in the Chinese Regulatory Environment of Medical Devices

❖ Li Wei, Cardiovascular Institute and Fu Wai Hospital, No. 167, Bei Li Shi Lu, Xi Cheng District, Division of Biometrics, Beijing, 100037 China, *liwei0325@yahoo.com.cn*; Yao Chen, Peking University First Hospital

In China, the State Food and Drug Administration (SFDA) is the primary agency responsible for regulating medical devices. Presently, there is no statistical guideline for medical devices regulation. However, SFDA pays particular attention to statistics, especially for class III Medical Devices---including implants, life support systems, and devices posing potential risk to the human body---which are controlled strictly for safety and effectiveness. Usually, the Medical Device Evaluation Center of SFDA invites trial statisticians to attend the technical review meeting of a new medical device. The trial statistician will evaluate the device clinical trial for study design, sample size, power, efficacy, and safety. Sponsors tend to cooperate with the trial statistician from the beginning to the end of a trial. Statistics is playing an increasingly more important role in device regulation.

### The Global Harmonization Task Force

❖ Larry G. Kessler, U.S. Food and Drug Administration, 9200 Corporate Blvd., HFZ-100, Rockville, MD 20850, *lgk@cdrh.fda.gov*

Regulation of medical devices, including in vitro diagnostic devices, varies by region. For more than 13 years, efforts to harmonize various requirements for devices have been ongoing under the auspices of the Global Harmonization Task Force. We will discuss the structure and function of the GHTF and note the relevance of statistical and study design differences across regulatory authorities. In one study group, a statistical document was developed to assist in deciding whether to send reports to regulators concerning trends in adverse events. Finally, harmonization efforts for emerging technologies, such as nanotechnology and pharmacogenomic Rx/Dx codeveloped products, will be discussed.

## 45 Statistical Learning and Data Mining

International Chinese Statistical Association, Section on Nonparametric Statistics

**Sunday, August 6, 4:00 pm–5:50 pm**

### Image Denoising via Solution Paths

❖ Ji Zhu, University of Michigan, 439 West Hall, 1085 South University, Ann Arbor, MI 48109-1107, *jizhu@umich.edu*; Li Wang, University of Michigan; Hui Zou, University of Minnesota

***Key Words:*** linear programming, quadratic programming, regularization, variable selection

Image denoising is a problem that arises in many engineering fields, because in practice images can be easily contaminated with noise when they are captured or transmitted. Many image denoising methods can be characterized as minimizing "loss + penalty", where the "loss" measures the fidelity of the denoised image to the data, and the "penalty" measures the smoothness of the denoising function. In this paper, we consider a family of models that use the L1-norm of the pixel updates as the penalty. The L1-norm penalty has the advantage of changing only the noisy pixels, while leaving the non-noisy pixels untouched. We derive efficient algorithms that compute entire solution paths of these L1-norm penalized models, which facilitate the selection of a balance between the "loss" and the "penalty".

### Using Input-Dependent Weights for Model Combination and Model Selection with Multiple Sources of Data

❖ Wei Pan, University of Minnesota, Division of Biostatistics, School of Public Health, A460 Mayo, MMC 303, Minneapolis, MN 55455,

*weip@biostat.umn.edu*; Guanghua Xiao, University of Minnesota; Xiaohong Huang, University of Minnesota

***Key Words:*** classification, microarray data, model mixing, partial least squares, prediction

With constantly accumulating data, the importance of integrative analyses of multiple data sources has been recognized increasingly. A natural approach is to combine multiple models, each built on one source of data. A challenge, however, is to account for different local information contents of different sources of data. The choice of the weight on each candidate model (and thus each source of data) may depend on the input for which a prediction is to be made, suggesting the constant weights used in most existing approaches may not be optimal. Here, we propose an input-dependent weighting (IDW) scheme. We apply IDW to discriminating human heart failure etiology using two sources of gene expression data and to gene function prediction by a combined analysis of gene-expression and protein-protein interaction data.

### Binning in Gaussian Kernel Regularization

❖ Bin Yu, University of California, Berkeley, Statistics Department, 367 Evans Hall, #3860, Berkeley, CA 94720-3860, *binyu@stat.berkeley.edu*; Tao Shi, University of California, Berkeley

Gaussian kernel regularization is used widely in the machine learning literature and has been proven successful in many empirical experiments. The periodic version of the Gaussian kernel regularization has been shown to be minimax rate optimal in estimating functions in any finite order Sobolev spaces. However, for a dataset with n points, the computation complexity of the Gaussian kernel regularization method is of order O(n^3). In this talk, we propose using binning to reduce the computation of Gaussian kernel regularization in both regression and classification. For the periodic Gaussian kernel regression, we show the binned estimator achieves the same minimax rates of the unbinned estimator, but the computation is reduced to O(m^3), with m as the number of bins. To achieve the minimax rate in the kth order Sobolev space, m needs to be in the order of O(kn^{1/(2k+1)}).

## 46 Statistical Graphics: from Playfair to Bertin and Beyond

Section on Statistical Graphics, Section on Statistical Education

**Sunday, August 6, 4:00 pm–5:50 pm**

### Graphics in French Statistical Journals during the 19th Century

❖ Antoine de Falguerolles, University Paul Sabatier (Toulouse III), LSP, 118 route de Narbonne, Toulouse, F-31000 France, *falguero@cict.fr*

***Key Words:*** graphical statistics, history of statistics, French statistical societies

The 19th century saw the birth and sometimes the death of a number of French statistical societies. Recently, the name of the Statistical Society of Marseille came up in discussions published in the newsletters of the ASA and RSS. But there were other well-attended societies. Most of them published journals or, at least, were keen on leaving printed evidence of their activity. While some members of the early societies were aware of the work of William Playfair (and a few other forerunners of graphical methods), the discussion of graphical methods in statistics and the use of graphics developed slowly in the French statistical journals. It culminated with articles written by E. Levasseur and E. Cheysson after 1875. I will present and discuss some noticeable examples taken mostly from the Journal of the Statistical Society of Paris.

### Andre-Michel Guerry and the Rise of Moral Statistics

❖ Michael Friendly, York University, Psychology Department, 4700 Keele Street, Toronto, ON M3J 1P3 Canada, *friendly@yorku.ca*

***Key Words:*** history of graphics, crime mapping, multivariate visualization

Andre-Michel Guerry's (1833) "Essai sur la statistique morale de la France" was one of the foundation studies of modern social science. Guerry assembled data on crimes, suicides, literacy, and other 'moral statistics' in an attempt to shed light on social issues in France. He used tables and maps to analyze these issues and was likely the first to try to use such data to answer questions about the relations among multiple social variables. His analyses showed that rates of crime---broken down by department, gender, and age---remained remarkably stable over time, yet varied greatly from one place to another. In this talk, I trace the development of the study of moral statistics in Guerry's work. I also illustrate recent attempts to ask what light modern methods of statistical graphics and thematic cartography can shed on Guerry's questions and data.

### William Playfair and the Psychology of Graphs

❖ Ian Spence, University of Toronto, Department of Psychology, 100 St George St, Toronto, ON M5S 3G3 Canada, *spence@psych.utoronto.ca*

***Key Words:*** William Playfair, history of statistics, statistical graphs, perception, cognition

The appeal of statistical graphs depends heavily on their ability to exploit our natural perceptual and cognitive capacities, which evolved to meet very different challenges. Although little was known about the brain two centuries ago, the inventors of statistical graphs must have had some intuitive awareness of how the visual system functioned. William Playfair---inventor of the time series line graph, the bar chart, and the pie chart---was particularly astute in his choice of designs. His written justifications, and the constructions themselves, indicate he had an instinctive understanding of our psychological capabilities and, moreover, he understood how to exploit them. He anticipated many ideas that are the focus of work in experimental psychology to this day.

## 47 Statistical Methods in Oral Health Research ●

Biometrics Section, WNAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### A Semiparametric Bayesian Model for Inter-Rater Agreement of Probing Pocket Depth

❖ Elizabeth G. Hill, Medical University of South Carolina, Department of Biostatistics, Bioinformatics, and Epidemiology, PO Box 250835, 135 Cannon St., Suite 303, Charleston, SC 29425, *hille@musc.edu*; Elizabeth H. Slate, Medical University of South Carolina

***Key Words:*** Bayesian modeling, Dirichlet process mixture models, agreement studies, periodontal disease, oral health

Probing pocket depth (PPD) is a measurement used in the diagnosis of periodontal disease. In periodontal research, examiners are calibrated by quantifying reproducibility relative to a reference examiner. Typically, multiple examiners record PPD at multiple sites in a subject's mouth, inducing correlation among measures from the same patient and between measures from the same surface. We adopt a semiparametric Bayesian hierarchical approach using Dirichlet process mixtures to model PPD and as prior distributions on rater effects. This induces marginal correlation among PPD measures at both the subject and examiner levels, and groups surface into classes that reflect varying degrees of difficulty in reproducibility for each examiner. We will analyze data from a PPD calibration study and illustrate how assessments of examiners' agreements with the standard derive naturally from the model.

### Ensemble Models for Risk Prediction with Survey and Multilevel Data

❖ Stuart A. Gansky, University of California, San Francisco, 3333 California Street, Suite 495, San Francisco, CA 94118, *stuart.gansky@ucsf.edu*; Nancy F. Cheng, University of California, San Francisco

***Key Words:*** classification and regression tree, random forests, survey, nesting, logit, hybrid

Classification and regression tree (CART), multiple additive regression tree (MART), and random forests (RF) prediction models are powerful prediction techniques. Multilevel or complex survey sample data need to have their designs taken into account to provide proper inference. Ensemble models to account for these designs with a multilevel hybrid CART survey logit model approach and differential misclassification costs were developed. Extensions for ensemble models with MART and with RF are proposed. A stratified cluster sample health examination survey of early childhood caries is used for illustration. Results show that this hybrid approach may be effective at identifying associations with health disparities. Support: US DHHS NIH/NIDCR, NCMHD U54DE14251.

### Spatial Analyses of Periodontal Data Using Conditionally Autoregressive Priors Having Two Classes of Neighbor Relations

❖ Brian Reich, North Carolina State University, 2501 Founders Drive, Box 8203, Raleigh, NC 27695, *reich@stat.ncsu.edu*; James Hodges, University of Minnesota; Bradley P. Carlin, University of Minnesota

***Key Words:*** conditional autoregressive prior, spatial statistics, periodontal data

Attachment loss often is used to measure the current state of a patient's periodontal disease and to monitor disease progression. It can be analyzed using a conditionally autoregressive prior distribution, which smoothes fitted values toward neighboring values. However, it may be desirable to have more than one class of neighbor relation in the spatial structure so the different classes of neighbor relations can induce different degrees of smoothing. For example, we may wish to allow smoothing of neighbor pairs bridging the gap between teeth to be different from smoothing of pairs that do not bridge such gaps. Adequately modeling the spatial structure may improve monitoring of periodontal disease progression. We develop a two-neighbor-relation CAR model to handle this situation and apply the model to the data of 50 subjects from a recent clinical trial.

## 48 Minorities, Environment, and Statistics ● ✪

Committee on Minorities in Statistics, Section on Statistical Education

**Sunday, August 6, 4:00 pm–5:50 pm**

### Minorities, Environment, and Statistics

❖ Nagambal Shah, Spelman College, 350 Spelman Lane, SW, Atlanta, GA 30314, *nshah@spelman.edu*; ❖ William Hunt, North Carolina State University, Department of Statistics, 10A Patterson Hall, Raleigh, NC 27695, *whunt@stat.ncsu.edu*; ❖ Julia Bader, The University of Texas at El Paso, 137 Bell Hall, UTEP, El Paso, TX 79968, *jbader@utep.edu*; ❖ Kishi Animashaun Ducre, Syracuse University, 200 Sims Hall, Department of African American Studies, Syracuse, NY 13244, *kanimash@syr.edu*

***Key Words:*** minorities, environmental justice, environmental data, environmental health

The economically disadvantaged and minorities in the United States bear a disproportionate share of environmental problems. While environmental justice movements have contributed to the development of environmental policies, there is still work to be done to ensure all Americans have the basic right to live in a healthy environment. Much money, time, and energy are being spent by federal, state, and private organizations in the United States and other countries to collect environmental data, but much too little is invested in the analysis and interpretation of these data. This panel brings experts from academia, government, and industry perspectives to discuss issues of environmental justice and environmental health at the national and global levels. A viable win-win-win model for research in training undergraduates to cultivate tomorrow's problemsolvers will be presented.

## 49 "Bad" Statistical Methods: What Are the Costs? ● ✪

The American Statistician, Section on Statistical Education, Section on Statistical Consulting, Section on Teaching Statistics in the Health Sciences

**Sunday, August 6, 4:00 pm–5:50 pm**

### "Bad" Statistical Methods: What Are the Costs?

❖ David Freedman, University of California, Berkeley, , *census@stat.berkeley.edu*; ❖ S. Stanley Young, National Institute of Statistical Sciences, , *young@niss.org*; ❖ Mary Foulkes, U.S. Food and Drug Administration, CBER, , *foulkes@cber.fda.gov*; ❖ Juliet Shaffer, University of California, Berkeley, , *shaffer@stat.berkeley.edu*

***Key Words:*** research methods, loss, risk, publication, procedures

Statistics are used increasingly in cases where costs and benefits are important, including litigation, credit scoring, fraud and terrorism detection, scientific technology transfer, and scientific policymaking. We have all seen examples of statistical methods, procedures, and protocols that are suspect at best---and completely inappropriate at worst. Yet, in many cases, such methods continue to be used and enjoy popularity in their respective scientific disciplines. If 'bad' methods are used, are they really that 'bad'? What are the real costs of using 'bad' statistical methods?

## 50 Strengths and Weaknesses of a Megatrial ●

Biopharmaceutical Section, Biometrics Section, ENAR
**Sunday, August 6, 4:00 pm–5:50 pm**

### Megatrials: Not Necessarily Either/Or

❖ Lloyd Fisher, University of Washington, 19220 64th Place, NE, Kenmore, WA 98028-3367, *lloyd_fisher@alum.mit.edu*

***Key Words:*** megatrial, adaptive subsets, embedded small studies

Because of the (by definition) large size of megatrials, there are necessarily smaller amounts of information collected on each patient. While this is usually an economic and practical necessity, we note that this need hold only on the average. Thus, features of smaller trials can be incorporated for subsets of patients (e.g., more safety data for the first "n" individuals, added testing on subjects with some specified baseline characteristics). In addition, adaptive designs for prespecified subsets may be incorporated in a partial sense, embedding prespecified smaller trials within the megatrial---albeit with some penalty to be paid.

### Issues in the Use of a Composite Endpoint in Megatrials

❖ Steven Snapinn, Amgen Inc., One Amgen Center Drive, 24-2-C, Thousand Oaks, CA 91320, *ssnapinn@amgen.com*

A megatrial's primary endpoint is often the composite of a number of clinical events. These composite endpoints add complexity to the trial's design and analysis. In this presentation, some of the issues surrounding the use of a composite endpoint---including heterogeneity of the effect of the treatment across components of the endpoint, the appropriate analysis of the components, and considerations for labeling based on a composite endpoint---will be discussed. Several examples of published megatrials will be used.

### Are Megatrials Worth It?

❖ Barry Davis, The University of Texas School of Public Health, 1200 Herman Pressler, Houston, TX 77030, *barry.r.davis@uth.tmc.edu*

***Key Words:*** clinical trials, sample size, subgroups, misclassification

Megatrials are large, randomized, controlled trials. Strengths include the ability to detect moderate but worthwhile treatment benefits for meaningful outcomes (morbidity/mortality), especially for diseases affecting a sizeable proportion of the population; broad recruitment criteria, thus making results more generalizable; and the ability to provide more "real-world" estimates of effectiveness and adverse effects. Weaknesses include enormous cost and effort; increased risk of protocol/treatment nonadherence, poor data quality, and misclassification; and possible inability to assess subgroup effects or address treatment mechanism questions. Given the above, are megatrials worth the massive resources required to conduct them? Examples of such studies, including the Antihypertensive and Lipid-Lowering Treatment to prevent Heart Attack Trial (ALLHAT), will be discussed.

### Strengths and Weaknesses of a Megatrial: Complexity of Designing, Handling, and Implementing Megatrials

❖ Timothy Church, University of Minnesota, 200 Oak Street, SE, Suite 350, Minneapolis, MN 55105, *trc@cccs.umn.edu*

***Key Words:*** megatrial, recruitment, randomization, endpoint, administration, multi-center

This talk will examine challenges associated with size and long-term duration, including enrollment in quantity and multicountry setting, complex disease and endpoint aspects, and many administrative committees (steering, executive, adjudication, data monitoring) and their interplay with sponsor, CRO (if applicable), local/regional IRBs, and health authorities.

## 51 Issues with Open Source Statistical Software in Industry: Validation, Legal Issues, and Regulatory Requirements ● ✪

Section on Statistical Computing, Biopharmaceutical Section, Section on Statistical Consulting, Section on Statistical Graphics
**Sunday, August 6, 4:00 pm–5:50 pm**

### Open-Source Software and Pharma Development: Computer Systems Validation and Value

❖ Anthony Rossini, Novartis Pharma AG, WSJ 27 1 12, Basel, 4002 Switzerland, *anthony.rossini@novartis.com*

***Key Words:*** statistical computing, 21CFR part 11, GXP, computer systems validation, modeling and simulation, open source software

Innovation and conservative behavior are contradictory paradigms surrounding the development phase of the pharma industry, where regulated data processing and computing is a mission-critical aspect. Open-source software has the potential to provide both and neither. In the context of pharmaceutical development, the "free as in beer" aspect of open source is a quality that might seem attractive, though it is the "free as in speech" aspect where the actual benefit to the business can arise. We present issues that arise in the actual computer systems validation process and how they can be addressed. Both R, a general statistical programming language, and SPK, a tool for pharmacological modeling and simulation, will be used as case studies. We will conclude with possible strategies for incorporation and a sense of the cost/benefits that might arise.

### Open-Source Software in Pharmaceutical Discovery

❖ Gregory Warnes, Pfizer Inc., MS 8260 2104, Eastern Point Road, Groton, CT 06340, *gregory.r.warnes@pfizer.com*; A. Max Kuhn, Pfizer Global Research & Development; James Rogers, Pfizer Global Research & Development

***Key Words:*** software, pharmaceutical, R, microarray, genetics, open-source

Open-source statistical software is being used with increasing frequency for the analysis of pharmaceutical data, particularly in support of "omics" technologies within discovery. While it is relatively straightforward to employ open-source tools for basic research, software used in any regulatory context must meet more rigorous requirements for documentation, training, software lifecycle management, and technical support. We will focus on R---a full-featured, open-source statistical software package. We'll briefly outline R's benefits and then discuss the requirements for its use in discovery. Next, we will discuss key issues for organizations to be comfortable using it within regulatory arenas in clinical, safety, or manufacturing. We'll then discuss how well R meets these requirements and ask whether third-party commercial support can resolve some issues.

### Use of Open-Source Software by an Academic Center in a Regulatory Environment

❖ Thomas D. Cook, University of Wisconsin-Madison, 209 WARF, 610 Walnut Street, Madison, WI 53726, *cook@biostat.wisc.edu*

***Key Words:*** open source software, clinical trials, interim analysis

The Statistical Data Analysis Center (SDAC) in the Department of Biostatistics and Medical Informatics at the University of Wisconsin-Madison serves as the independent analysis center supporting data and safety monitoring boards for industry-sponsored, multicenter, randomized clinical trials. Interim analyses conducted at SDAC are subject to scrutiny by various scientific, regulatory, and ethics groups within the United States and internationally. In this environment, the conduct of interim analyses requires a balance between timeliness, completeness, and accuracy. To optimally satisfy these competing goals, SDAC uses a hybrid of propriety and open-source software, using each for the applications in which it is best suited. We will describe the system currently in place at SDAC and the advantages and disadvantages of various software tools.

### Times R A'changin': FDA Perspectives on Use of Open Source

❖ B. Sue Bell, U.S. Food and Drug Administration, WO22 Mailstop 6105, 10903 New Hampshire Ave, Silver Spring, MD 20993-0002, *Sue.Bell@fda.hhs.gov*; Kathleen Morrish, U.S. Food and Drug Administration; Ferrin Harrison, U.S. Food and Drug Administration; David Petullo, U.S. Food and Drug Administration; Laura Thompson, U.S. Food and Drug Administration; Gerry W. Gray, U.S. Food and Drug Administration

***Key Words:*** open source, freeware

In August 2003, the FDA issued "Guidance for Industry Part 11, Electronic Records; Electronic Signatures - Scope and Application." The FDA explains it is re-examining Part 11 and intends to interpret its scope narrowly. It recognized that earlier drafts risked discouraging innovation and technological advances without providing a significant public health benefit. Meanwhile, the National Cancer Institute's caBIG (cancer Biomedical Informatics Grid) initiative applies open-source, open-development principles. This talk will provide a brief history of how FDA's perspectives have evolved over time, discuss the internal environment of how FDA scientists are permitted limited access to software that is not "commercial off the shelf" (COTS), and discuss implications for regulatory submissions based on software that is not COTS.

### Software and Code Evaluation: Risk-Based Approaches to Software Integration

❖ Nicholas J. I. Lewin-Koh, Eli Lilly and Company, Corporate Center, Statistics and Information Sciences, Drop 2233, Indianapolis, IN 46285, *nikko@lilly.com*; Robert A. Myers, Eli Lilly and Company

***Key Words:*** software, open source, clinical reporting, risk assessment

Decisions in the pharmaceutical world are motivated by FDA requirements, especially in clinical phases. However phases of the industry are less regulated and require more flexible approaches to data management. A possible approach to selecting software can be based on competing risks. More regulated environments incur higher risk when security can be compromised and higher cost when the burden of validation must be incurred with company resources. In contrast, in preclinical and discovery phases, risk is determined by the ability to adapt to new technologies and processes. Costs are elevated by extensive validation requirements, which are more severe when regulation is tighter. We see a role for open source tools in areas where validation costs are minimal in relation to the complexity of the software requirements.

## 52 Model-Based Seasonal Adjustment: Algorithms and Applications

Business and Economics Statistics Section

**Sunday, August 6, 4:00 pm–5:50 pm**

### Numerical Implementation of Kalman Filter/Smoother for State Space Models with Partially Diffuse Initial Conditions

❖ Rajesh Selukar, SAS Institute, Inc., 500 SAS Campus Drive, Cary, NC 27513, *rajesh.selukar@sas.com*

***Key Words:*** Kalman filter, state space, non-stationary, smoothing, diffuse initial state, square-root

In the past few years, several alternate forms of Kalman Filter/Smoother algorithms have been published that deal with state space models with partially diffuse initial conditions. We will survey the field briefly and discuss the relative merits of some of these algorithms. We will focus on mainly the issues of numerical stability and ease of implementation.

### Evaluation of Finite-Sample Diagnostics for Model-Based Seasonal Adjustments and Trends

❖ David Findley, U.S. Census Bureau, Statistical Research Division, Washington, DC 20233-9100, *david.f.findley@census.gov*; Richard Gagnon, U.S. Census Bureau; Tucker S. McElroy, U.S. Census Bureau

***Key Words:*** ARIMA, signal extraction, SEATS, X-13A-S

Most seasonal adjustment and trend diagnostics of the SEATS seasonal adjustment program (Gomez and Maravall, 1997, Maravall and Caporello, 2004) are calculated with formulas for infinite-length series. Using matrix formulas of Bell and Hillmer (1988), McElroy and Sutcliffe (2004), and McElroy (2005) for finite-sample ARIMA model-based signal extraction filters and the error covariances of their estimates, a variety of finite-sample-based diagnostics were obtained. Findley and Martin (2003, 2006) and Findley, McElroy, and Wills (2004) show early examples. Now, finite-sample alternatives to the main diagnostics of SEATS and new diagnostics have been developed and implemented in a test version, the U.S. Census Bureau's hybrid X-13A-S program fusing X-12-ARIMA and SEATS. This has enabled an evaluation study of these diagnostics involving many series, results from which we will summarize.

### Aspects of Model Averaging for Seasonal Adjustment

❖ John Aston, Academia Sinica, Institute of Statistical Science, 128 Academia Road Sec 2, Taipei, 11529 Taiwan, *jaston@stat.sinica.edu.tw*

***Key Words:*** time series, seasonal adjustment, Bayesian model averaging, frequentist model averaging

Current methods for model-based seasonal adjustment tend to focus on selecting the optimal model (according to some criteria) and then computing the seasonal adjustment based on that unique model. Here, an alternative approach is proposed where the adjustment is based on weighted averages of the seasonally adjusted series from many models under consideration. Two main types of averaging are considered: frequentist (with likelihood-based weights) and Bayesian (with weights based on the Bayes factors of the models). The resulting averaging is compared in terms of its performance compared with current methods, both in terms of forecasting, and in terms of the revisions and seasonal stability of the adjusted series. This is investigated through both simulated series (where the true model may or may not be among the set considered) and real datasets.

### Assessing Spectral Peaks in Economic Time Series

❖ Thomas D. Evans, Bureau of Labor Statistics, 4561 Sawgrass Court, Alexandria, VA 22312-3152, *evans.thomas@bls.gov*; Stuart Scott, Bureau of Labor Statistics; Scott Holan, University of Missouri-Columbia; Tucker S. McElroy, U.S. Census Bureau

***Key Words:*** spectral analysis, labor force statistics

Analysis of spectra is useful in seasonal adjustment of economic time series for identifying the presence of seasonality in observed series and the presence of residual seasonality in seasonally adjusted series. McElroy and Holan (2005) propose a nonparametric test for the presence of spectral peaks. To test this approach, seasonal monthly labor force time series from the U.S. Current Population Survey and simulated seasonal monthly time series will be evaluated for peaks at the seasonal frequencies. Comparisons will be made against other indicators of seasonality: the “visually significant peaks” from X-12-ARIMA, estimated autocorrelations at seasonal lags, and modified F-tests.

## 53 From Sharks to Salmon: Quantitative Tools in Marine Demography and Management for Puget Sound and Alaska Fisheries ●

Section on Statistics and the Environment

**Sunday, August 6, 4:00 pm–5:50 pm**

### The Management Strategy Evaluation Approach and the Gulf of Alaska Walleye Pollock Fishery

❖ Teresa A’mar, University of Washington, SAFS Box 355020, Seattle, WA 98195-5020, *zta@u.washington.edu*; Andre E. Punt, University of Washington; Martin W. Dorn, National Oceanic & Atmospheric Administration

***Key Words:*** fisheries management, simulation testing, management strategy evaluation, walleye pollock

Management strategy evaluation (MSE) is the process of using simulation testing to examine the robustness of proposed management strategies to error and uncertainty. MSE involves using a model (i.e., “operating model”) to represent the true underlying dynamics of the resource and to generate future data, using an estimation model to assess the state of the stock relative to an agreed target and limit reference points at each time step based on the simulated data, and using a catch control rule to determine management actions (e.g., the Total Allowable Catch, TAC), given the results of the stock assessment. The latter two steps constitute the management strategy. The parameters of the management strategy can be selected to achieve desired (but conflicting) management goals and objectives.

### Using Mixture Models To Estimate Abundance of Patchy Species

❖ Elizabeth Conners, National Oceanic & Atmospheric Administration, 7600 Sand Point Way, NE, Seattle, WA 98115, *liz.conners@noaa.gov*

***Key Words:*** mixture models, delta distribution, fisheries

A common problem in fisheries and many other areas of environmental science is estimating the overall abundance of a resource with an aggregated or patchy spatial distribution. Typically, the majority of survey observations will have zero or very low density, with a few high-density observations representing the majority of the biomass. This research looks at model-based estimates of abundance based on the assumption that samples are drawn from a mixture of distributions representing distinct biological processes. The behavior of mixture model estimators is examined using both simulated and real data. GLMs are used to examine spatial pattern in component classification for Atka mackerel data from the Aleutian Islands, Alaska, and model-predicted classes are used as the basis for both mixture model and restratified estimators.

### Forecasts of Salmon Returns

❖ Saang-Yoon Hyun, Columbia River Inter-Tribal Fish Commission, 729 NE Oregon Street, Suite 200, Portland, OR 97232, *hyus@critfc.org*; David H. Salinger, University of Washington

***Key Words:*** preseason and in-season forecasts, salmon returns, interannual variability, probability density, optimization, Bayesian methods

Preseason and in-season forecasts of salmon returns are required for commercial fishermen and harvest managers. We make a preseason forecast with historical data on sibling returns and ocean conditions using ordinary regression and autoregressive models. Also, we make an in-season forecast with data on fish run proportions from the past years and daily fish returns collected during in-season. Interannual variability in the ecosystem and fish biotic metrics are the main obstacle to making an accurate forecast early during in-season. We develop a probability density and use a pattern-matching optimization to generate an in-season forecast. We also use Bayesian methods to express forecast uncertainty. Finally, we combine preseason and in-season forecasts to reduce forecast uncertainty. We illustrate our ideas using Columbia River Chinook salmon and Alaska Bristol Bay sockeye salmon.

### Using Multivariate Statistics To Resolve Issues of Scale with Salmon Survival and Ocean Environmental Data

❖ Rishi Sharma, University of Washington, QERM, Seattle, WA 98195, *rsharma@u.washington.edu*

***Key Words:*** SCAD, multivariate, cluster

From Oregon to Alaska, Chinook salmon stocks exhibit varying life history patterns. In addition, these salmon spend between 50--90% of their adult life in ocean environments. Recent studies indicate ocean conditions might affect survival. In order to demonstrate this hypothesis on large spatial scales, we analyzed tag recovery data used in management and estimated survival for numerous stocks. As these stocks are distributed from hundreds to thousands of kilometers apart, we studied ocean environmental data that would cover a similar geographical range. A broad-scale ocean signal is detected in ocean and survival anomalies for that time and space using cluster analysis and principal component analysis techniques. To resolve finer-scale resolution on survival, we use environmental data that correlate on stock-specific spatial scales. The implications on management are discussed.

### Reconciling Biological Realities with Statistical Requirements in Fitting Growth Curves with Emphasis on Growth Models for Sharks

❖ Nicole Vega, University of Washington, Box 355020, 1122 NE Boat Street, Seattle, WA 98105, *dryad@u.washington.edu*; Vincent Gallucci, University of Washington

***Key Words:*** asymptotic growth, growth curves, nonlinear regression, sharks

Fitting asymptotic growth curves to shark data presents opportunities and problems as compared with other fishes. At the small end of the size spectrum, accurate measurements of age at length are easy to take, but it may be difficult to obtain these samples. At the largest ages/lengths, ageing may carry high variance, and, because it is difficult to obtain information about these large animals, most predicted asymptotic sizes are projections based on data from smaller sharks. This research investigates the relationship between the distribution of the data over the range of observable lengths and the associated precision of the estimates. When the sample data omit regions of the possible length range, the effect is a loss of precision in parameter estimates. The effect is most pronounced when missing lengths are on the extremes of the length range.

## 54 Overview and Results from the 2005 National Census Test ● ✪

Section on Survey Research Methods

**Sunday, August 6, 4:00 pm–5:50 pm**

### An Overview of the 2005 National Census Test

❖ Jennifer Tancreto, U.S. Census Bureau, 5610 Paynes Endeavor Drive, Bowie, MD 20720, *jennifer.guarino.tancreto@census.gov*

***Key Words:*** census, experiment

In preparation for the 2010 Census, the U.S. Census Bureau conducted the 2005 National Census Test (NCT) in the fall of 2005. The 2005 NCT was a large, multifaceted mailout test designed to study questionnaire content and design as well as methods for improving response and the efficiency of data collection procedures. In this paper, I outline the five overarching objectives of the 2005 NCT, then describe the treatments used to test them. I explain how the treatments were intermixed into one panel design for cost-saving purposes and provide the justification behind the combinations of treatments into the experimental panels. I then describe the sample design and sample selection procedures. Finally, I report issues that occurred during the data collection period and provide global metrics for the test.

### Effect of Internet Response Mode Designs on Data Quality and Ease of Use

❖ Kelly Allmang, U.S. Census Bureau, 1022 Simsbury Court, Crofton, MD 21114, *Kelly.D.Allmang@census.gov*; Kevin Zajac, U.S. Census Bureau

***Key Words:*** topic-based, person-based, census

The U.S. Census Bureau conducted the 2005 National Census Test, a multifaceted test that explored questionnaire content and design alternatives. The overall goal of the study is to enhance the internet application to take further advantage of the technology while maintaining or improving the data quality. One objective was to test two designs of the 2005 Census internet application: a person-based approach and a topic-based approach. The person-based approach resembled a traditional paper census questionnaire, where the respondent provides data for one person at a time. In the topic-based application, the respondent provided person-level data for the entire household one question at a time. This paper will compare the design approaches and make comparisons of data quality between the internet and paper response modes in the 2005 National Census Test.

### Experimental Treatment Results of the Bilingual Census Form from the 2005 National Census Test

❖ Julie Bouffard, U.S. Census Bureau, 16030 Elegant Court, Bowie, MD 20716, *julie.a.bouffard@census.gov*; Jennifer Tancreto, U.S. Census Bureau

***Key Words:*** Spanish form, swim lane design, lane jumping, language

In preparation for the 2010 Census, the U.S. Census Bureau conducted the 2005 National Census Test, which explored variations of questionnaire design and methodology, including the use of a bilingual census form. This paper focuses on the bilingual census form experimental treatment of the test. The bilingual form consisted of a “swim lane” design that provided two response columns, one in English and one in Spanish, each containing the questions and response categories. The desire to test a bilingual form stemmed from the hope that it would increase response to the census and lower item nonresponse by appealing to a broader audience than the standard English form. This analysis will compare self response rates and item nonresponse rates between the bilingual form and a standard English form as well as examine “lane jumping” between language columns of the bilingual form.

### Analysis of Self-Response Options and Respondent-Friendly Design from the 2005 National Census Test

❖ Michael Bentley, U.S. Census Bureau, 9257 Pigeon Wing Place, Columbia, MD 21045, *michael.bentley@census.gov*

***Key Words:*** self-response options, internet, questionnaire design, census

In preparation for the 2010 Census, the 2005 National Census Test was conducted. One objective was to assess the impact of different self-response options strategies on response rates and data completeness. These methods included encouraging internet response at the replacement mailing, using messaging to distinguish the replacement and initial questionnaires, and using two methods to create the mail package. The purpose of testing the self-response options was to improve the operational feasibility of a replacement mailing, given realistic 2010 schedule constraints, and to improve the percentage and speed of returns. A second aspect involved changes to the design of the paper questionnaire in an effort to improve the respondent-friendliness of the form. This paper reports the results of these experimental treatments.

### Experimental Treatment Results for the Age, Relationship, and Tenure Items from the 2005 National Census Test

❖ Joan Hill, U.S. Census Bureau, 12938 Byefield Drive, Highland, MD 20777, *joan.marie.hill@census.gov*; Jennifer Tancreto, U.S. Census Bureau; Cynthia A. Rothhaas, U.S. Census Bureau

***Key Words:*** age, relationship, tenure, 2010 Census, 2005 National Census Test

In preparation for the 2010 Census, the U.S. Census Bureau conducted the 2005 National Census Test with Census Day on September 15, 2005. Several of the experimental panels in the 2005 test were designed to test new variations in the tenure item, changes to the relationship question and its response categories, and an added instruction for age. Specifically, the tenure portion of the test examines the removal of the word "cash" when asking about rent and adds an instruction regarding home equity loans. The relationship item alternative removes the opportunity to provide a specific written response and offers reworded response categories. The experimental age item adds an instruction to identify correctly infants less than one year old as age zero. This research is aimed at identifying the best wording and instructions for these items as part of the 2010 Census questionnaire.

## 55 Statistical Issues in Veterans' Administration (VA) Health Services Research ●

Section on Health Policy Statistics

**Sunday, August 6, 4:00 pm–5:50 pm**

### Statistical Issues in Racial/Ethnic Disparities Research

❖ Roslyn A. Stone, Veteran's Affairs Pittsburgh Healthcare System, University Drive C (151C-U), Building 28, Pittsburgh, PA 15240-1000, *roslyn.stone@med.va.gov*; Huanyu Chen, VA Pittsburgh Healthcare System; Xiangyan Xu, Veteran's Affairs Pittsburgh Healthcare System

***Key Words:*** multi-level logistic models,

Statistical issues in racial/ethnic disparities research are outlined in the context of an ongoing study of racial/ethnic disparities in 30-day mortality for 124,754 veterans who were hospitalized at least once for pneumonia during FY 1996--2002 at one of 138 sites in the VA Healthcare System. The primary data sources were the VA National Patient Treatment File discharge database and the VA BIRLS Death File. Statistical issues include defining disparities, addressing the dramatic imbalances in race/ethnicity across sites, assessing the importance of adjustments for age and comorbidity, and accounting for socioeconomic status and contextual variables. In addition, we describe a sub-study to validate the VA-based data sources for mortality against the National Death Index and assess whether mortality ascertainment varies by race/ethnicity.

### The Use of Hierarchical Linear Models To Evaluate Methods for the Delivery of Primary Care

❖ Martin Lee, University of California, Los Angeles, 3941 Eureka Drive, Studio City, CA 91604, *martin.l.lee@att.net*

***Key Words:*** hierarchical linear models, group randomization, primary care research, sample size, random coefficients

In the evaluation of techniques for improving the delivery of primary care, difficulty arises in the determination of proper designs for the measurement of process and outcome. Typically, the patient is the unit of analysis, but the intervention may be invoked at the level of the physician, or, more frequently, at the institutional level (or group of institutions). This approach is necessary to prevent inter-study group contamination. When the institution is the unit of randomization, it is usually necessary to consider matching at this level, which introduces additional design and analytical complications. We demonstrate techniques for the determination of sample size in a multiple hierarchical situation, as well as various analytical methods. An example is given on the evaluation of methods to deliver information on the cessation of tobacco intake.

### Understanding Variation in Patient Safety Measures in the VA: How Bayesian Methods Can Help

❖ Cindy Christiansen, Boston University, 200 Springs Road, Bldg. 70, Bedford, MA 01730, *cindylc@bu.edu*

***Key Words:*** Bayesian methods, profiling

One goal of quality improvement (QI) efforts is to reduce adverse events related to health care. Recent research funded by the VA HSR&D tested and implemented the Agency for Healthcare Research and Quality's Patient Safety Indicators (PSIs) in the VA. We use these administrative data measures to study variation in PSI rates among VA medical centers (VAMCs). Our premise is that Bayesian methods improve translation of the study findings when compared to traditional statistical methods involving p-values and confidence intervals. Improved translation of the statistical results can help target QI efforts, reduce costs of QI programs, and, ultimately, improve patient safety in the VA. Objectives are to interpret variation in PSI rates among VAMCs using Bayesian methods and contrast the utility of these methods with traditional ones for identifying VAMCs with extreme PSI rates.

### A Decision-Theoretic Approach to Identifying Future High-Cost Patients

❖ Kenneth Pietz, U.S. Department of Veterans Affairs, , *kcpietz@aol.com*; Margaret M. Byrne, University of Miami; Laura A. Petersen, U.S. Department of Veterans Affairs

***Key Words:*** risk adjustment, diagnostic cost groups, Bayesian, decision theory

The VA allocates a special fund to each network every year for very high cost (VHC) patients, defined as those whose care costs more than $75,000 in one year. The network allocates the sum to individual hospitals. Allocation using the distribution of VHC patients from the previous year may lead to perverse incentives. We investigated the use of a diagnosis-based decision rule to do the allocation. Diagnostic cost groups (DCGs) were used with a Bayesian decision rule to allocate VHC funds in a test network of eight hospitals. The prior probabilities are the proportion of patients who are VHC in the network. The decision rule assigns patients above a specific DCG as probable VHC patients based on costs of misclassification. The allocations generated by several values of these costs and several DCG thresholds were compared to the actual distribution in the following year.

## 56 Bayesian Student Paper Competition I ● ✪

Section on Bayesian Statistical Science

**Sunday, August 6, 4:00 pm–5:50 pm**

### Bayesian Synthesis

❖ Qingzhao Yu, The Ohio State University, 631 Cuyahoga Court, Columbus, OH 43210, *yu@stat.ohio-state.edu*; Steven N.

MacEachern, The Ohio State University; Mario Peruggia, The Ohio State University

***Key Words:*** automatic modeling, data-splitting, human intervention, model averaging, ozone, sparse model

In implementing Bayesian analysis, we face the problem of using data multiple times. We select a model and obtain the posterior distribution using the same set of data. When several analysts model the same data, we wish to combine the models efficiently, which should provide improved predictive performance. This paper tackles these problems via a novel modeling method based on data-splitting. In implementing this method, several data analysts work independently on portions of a dataset, constructing separate models that are updated eventually and combined through Bayesian model averaging. This paper provides theoretical results that characterize general conditions under which data-splitting results in improved estimation. Application of the method to a popular, real dataset shows predictive performance superior to that of many popular automatic modeling techniques.

### A Bayesian Framework To Combine Multivariate Spatial Data and Physical Models for Hurricane Surface Wind Prediction

❖ Kristen M. Foley, North Carolina State University, 8203 Campus Box, Statistics Department, Raleigh, NC 27695, *kmfoley@stat.ncsu.edu*; Montserrat Fuentes, North Carolina State University

***Key Words:*** Bayesian inference, coregionalization models, hurricane surface wind fields, non-separable multivariate models, spatial statistics, storm surge forecasts

We wish to improve the prediction of the coastal ocean response to hurricanes by combining disparate sources of observational data and output from numerical forecast models. A new Bayesian modeling framework is introduced to allow for estimation of the parameters of a multivariate spatial statistical model for wind data and parameters of a physically based deterministic model while accounting for potential bias in the observed data. For real-time storm surge prediction, we use an empirical Bayesian approach to predict hurricane surface wind fields at high spatial and temporal resolution to be used as input for a 3-D coastal ocean model. We find that this spatial model improves prediction of the wind fields when compared to the original deterministic model. These methods also are shown to improve storm surge estimates for a case study of Hurricane Charley (2004).

### A Bayesian Pooled Analysis of Doubly Censored HIV Data Using the Hierarchical Cox Model

❖ Wei Zhang, Boehringer Ingelheim, 900 Ridgebury Road, Ridgefield, CT 06810, *wzhang@rdg.boehringer-ingelheim.com*; Kathryn Chaloner, The University of Iowa; Ying Zhang, The University of Iowa; Mary K. Cowles, The University of Iowa

***Key Words:*** doubly censored data, hierarchical Cox model, MCMC methods, imputation

Two common statistical problems in pooling survival data from several studies are addressed here in the context of a real case study of HIV-infected individuals. The first problem is the data are collected from multiple studies, and it is likely that heterogeneity exists among the study populations. A random-effects hierarchical Cox proportional hazards model is therefore used to perform the pooled analysis. The second problem is the data are doubly censored in that the HIV infection time is interval-censored and the death event may be right-censored. Two approaches to incorporating the uncertainty of interval-censored HIV infection time are developed and compared to a more usual analysis using imputation. We recommend the use of the fully Bayesian approach, which adequately incorporates the uncertainty of interval-censored HIV infection times.

### An Adaptive Bayesian Approach to Jointly Modeling Response and Toxicity in Phase I Dose-Finding Trials

❖ Meihua Wang, University of Pittsburgh, Department of Biostatistics, Suite 350 Sterling Plaza, 201 N. Craig Street, Pittsburgh, PA 15213, *wang_m@nsabp.pitt.edu*; Roger Day, University of Pittsburgh

***Key Words:*** Bayesian design, adaptive design, log-bivariate normal distribution, threshold, dose-finding, phase I trials

We present an adaptive Bayesian method for dose-finding in phase I clinical trials based on both response and toxicity under the assumption that the thresholds of response and toxicity jointly follow a bivariate log-normal distribution. Responses are rare in cancer trials. But “biological responses” may be common and help decide how aggressive a phase I escalation should be. The model assumes response and toxicity events happen depending on respective thresholds and provides a framework for incorporating prior information about the population threshold distribution, as well as accumulated data. For simplicity, we restrict attention to the probability of each outcome for the next patient only. The next dose can be assigned to maximize expected utility. Simulation results show the proposed design reliably chooses the preferred doses under different scenarios and priors.

### Hierarchical State-Space Model for Microarray Short Time Course Experiments

❖ Haiyan Wu, Emory University, 3131 N. Druid Hills Road 9312, Decatur, GA 30033, *haiyanwood@msn.com*; Ming Yuan, Georgia Institute of Technology; Susan Kaech, Yale University; M. Elizabeth Halloran, Fred Hutchinson Cancer Research Center

***Key Words:*** microarray, time course, hidden Markov model, empirical Bayes, informative prior

Microarray time course experiments are important in investigating dynamic biological processes. Four common tasks in microarray time course experiments are hypothesis testing, pattern recognition, trajectory formation, and network/pathway reconstruction. NaÔve approaches, treating time series as independent, tend to ignore the correlation information across time and suffer from lower sensitivity and higher misclassification rate. Model-based approaches have been proposed to rank/identify temporally differentially expressed genes. However, most of the up-to-date models are inappropriate for temporal microarray experiments with short time series. Here, we introduce a hierarchical model integrated with Hidden Markov Model structure and auto-regression to model the temporal gene expression profile on both expression and state levels to achieve these four common tasks.

## 57 Student Paper Competition Award Presentations ●

Section on Government Statistics, Section on Survey Research Methods, Social Statistics Section

**Sunday, August 6, 4:00 pm–5:50 pm**

### Robust Model-Based Predictor of Finite Population Total

❖ Yan Li, University of Maryland, 3422 Tulane Drive, #14, Hyattsville, MD 20783-1838, *yli@survey.umd.edu*; Partha Lahiri, University of Maryland

***Key Words:*** superpopulation, empirical best prediction, Box-Cox transformation

The prediction approach has received considerable attention in finite population sampling. Under this approach, the finite population is assumed to be realization from a superpopulation model. Often, the standard linear model is assumed on the response variable. However, this assumption has been found to be inadequate in many survey applications. As a remedy, a known transformation on response variable is used frequently. Specification of a transformation is not a trivial problem in many applications. This paper introduces Box-Cox transformation in survey data analysis. The method is adaptive in that the appropriate transformation is determined automatically by the data. Thus, robustness with respect to the model misspecification is achieved. The need for selecting a model within the Box-Cox class is justified by Monte Carlo simulation and real-life data analysis.

### Causal Inference Based on Directed Acyclic Graphical Models and the Randomization Distribution: a Probability-Sampling Approach

❖ Joel E. Hanson, University of California, Berkeley, 747 Red Oak Ave., Apt. 751, Albany, CA 94706, *jhanson@stat.berkeley.edu*

***Key Words:*** Horvitz-Thompson, counterfactual distributions, causal inference, randomization distribution, model-dependent, design based methods

This paper develops a method of nonparametric causal inference from a sample to a finite population. Causal effects are defined as finite population parameters based on the causal structure of the population. Only a nonparametric model for the joint distribution of the causal structure, which can be expressed as a directed acyclic graphical model, must be assumed. The finite population causal effects are expressed in two equivalent forms: stratified sampling--type estimators with the strata determined by the confounders and Horvitz-Thompson--type estimators with the selection probabilities replaced by the propensity score. Both forms lead to sampling estimators that are subtle variations on common estimators, and we provide a delta method for computing their approximate variances. Finally, Robin’s marginal structural model approach is linked to pseudo-likelihood methods.

### Small-Area Estimation for Business Surveys

❖ Hukum Chandra, University of Southampton, Southampton Statistical Sciences Research Institute, Division of Social Statistics, Highfield, Southampton, SO1 71BJ UK, *hchandra@soton.ac.uk*

***Key Words:*** skewed data, model calibration, expected value model, MBD approach

In business surveys, data typically are skewed and the standard approach for SAE based on linear mixed models leads to inefficient estimates. In this paper, we discuss SAE techniques for skewed data that are linear following a suitable transformation. In this context, implementation of the EBLUP approach under transformation to a linear mixed model is complicated. However, this is not the case with the model-based direct (MBD) approach (Chambers and Chandra, 2005), which is based on weighted linear estimators. We extend the MBD approach to skewed data using sample weights derived via model calibration based on a lognormal model with random area effects. Our results show this estimator is both efficient and robust with respect to the distribution of these random effects. An application to real data demonstrates the satisfactory performance of the method.

### An Application of Parametric Bootstrap Method in Small-Area Estimation Problem

❖ Huilin Li, University of Maryland, 9314 Cherry Hill Road, Apt #1024, College Park, MD 20740, *huilin@math.umd.edu*

***Key Words:*** parametric bootstrap, small area, confidence intervals, Fay-Herriot model, nested error regression model

In this paper, we apply the recently developed parametric bootstrap method in constructing confidence intervals of small-area means for two well-known small-area models: Fay-Herriot model and the nested error regression model. Using a Monte Carlo simulation study, we compare our method with rival methods in terms of coverage probabilities and average lengths. We then demonstrate the utility of the parametric bootstrap method by analyzing several real-life datasets.

### Local Polynomial Regression for Small-Area Estimation

❖ Pushpal Mukhopadhyay, Iowa State University, 204 Snedecor Hall, Ames, IA 50010, *pushpal@iastate.edu*; Tapabrata Maiti, Iowa State University

***Key Words:*** nonparametric, multi-level modeling, James-Stein estimator, area level small area model

Estimation of small-area means in the presence of area-level auxiliary information is considered. A class of estimators based on local polynomial regression is proposed. The assumptions on the area-level regression are considerably weaker than standard small-area models. Both the small-area mean function and the between-area variance function are modeled as smooth functions of the area-level covariates. A composite estimator that is a convex combination of the design-weighted mean and the prediction from the nonparametric model is developed. The estimator is shown to be asymptotically consistent under mild regularity conditions. An approximation of the mean squared error (MSE), based on Taylor linearization, and an estimate of the approximate MSE are developed and their theoretical properties studied.

## 58 Going beyond the Law: Ethical Aspects of Privacy in Surveys ● ✪

Social Statistics Section

**Sunday, August 6, 4:00 pm–5:50 pm**

### Going beyond the Law: Ethical Aspects of Privacy in Surveys

❖ Gerald Gates, U.S. Census Bureau, 4700 Silver Hill Road, Suitland, MD 20233, *gerald.w.gates@census.gov*; ❖ Pamela White, Statistics Canada, 120 Parkdale Ave., Room 17087 Main Building, Ottawa, ON K1A 0T6 Canada, *pamela.white@statcan.ca*; ❖ Jeffery Rodamar, U.S. Department of Education, 3830 39th Street, NW F114, Washington, DC 20016, *jeffery.rodamar@ed.gov*

***Key Words:*** ethics, privacy, informed consent, human subjects protection

This session highlights the important role for ethics in the program decisions of statistical agencies. Legal mandates direct how information is collected and protected, but often there are gray areas subject to interpretation and discretion. In today’s technologically advanced environment, it is even more imperative that agencies consider ethics in a range of survey and census issues. Panelists will address aspects of the role of professional ethics through the survey lifecycle with a particular focus on privacy considerations. They will look at the ethical dimensions of important decisions such as deciding what information to collect, the nature of informed consent, and the rules on employee access to identified data.

## 59 How Is the TI-83 Calculator Changing How We Teach the Introductory Course in Statistics? Or Is It? ●

Section on Statistical Education

**Sunday, August 6, 4:00 pm–5:50 pm**

### How Is the TI-83 Calculator Changing How We Teach the Introductory Course in Statistics? Or Is it?

❖ Dexter Whittinghill, Rowan University, Department of Mathematics, 201 Mullica Hill Road, Glassboro, NJ 08028, *whittinghill@rowan.edu*; ❖ Christopher Mecklin, Murray State University, 1396 State Route 94 W., Murray, KY 42071, *christopher.mecklin@murraystate.edu*; ❖ Carolyn P. Dobler, Gustavus Adolphus College, 800 W. College Ave., Saint Peter, MN 56082, *dobler@gac.edu*; ❖ James Davis, Radford University, 210 Arnold Ave., Radford, VA 24141, *jimdavis@pobox.com*; ❖ Madhuri Mulekar, University of South Alabama, Department of Mathematics and Statistics, 307 University Blvd, ILB 325, Mobile, AL 36688, *mmulekar@jaguar1.usouthal.edu*

***Key Words:*** TI-83 calculator, teaching, introductory statistics

The TI-83 (or TI-84) calculator has a plethora of statistical functions and distributions. As instructors, we no longer have to spend significant time on formulas, calculations, and tables; the calculator has it all. But is this beneficial to students? Do they understand what they are doing on the calculator? After a little history on how some schools came to use the TI calculator, we will discuss how the presenter uses the TI-83 in his/her classroom, what difficulties arise, and how he/she addresses helping students understand underlying concepts. We also will consider whether we should be using the calculator at all.

## 60 The Nontechnical Side of Statistical Consulting: Reflections on Careers as Working Statisticians and Suggestions and Guidance for Those on the Way

Section on Statistical Consulting, Section on Statistical Education

**Sunday, August 6, 4:00 pm–5:50 pm**

### The Nontechnical Side of Statistical Consulting: Reflections on Careers as Working Statisticians and Suggestions and Guidance for Those on the Way

❖ Thomas Boardman, Colorado State University, Department of Statistics, 217 Statistics Bldg, Fort Collins, CO 80523-1877, *boardman@colostate.edu*; ❖ Gerald van Belle, University of Washington, Biostatistics 357232, Seattle, WA 98195-7232, *vanbelle@u.washington.edu*; ❖ John Bartko, Retired, 7 Pine Circle, Newville, PA 17241, *JJBartko@Starpower.net*; ❖ Ross Prentice, Fred Hutchinson Cancer Research Center, 1100 Fairview AVE N. / M3-A410, P.O.Box 19024, Seattle, WA 98109, *rprentic@whi.org*

This session will be of interest to statisticians looking for expert suggestions regarding career advancement and progression, development, and professional visibility. A panel of four experienced statisticians will present their views based on more than 150 years of combined statistical experience over a range of settings. Their emphasis will be placed on mostly the nontechnical issues that abound in interactions between statisticians and consulting clients and between statisticians and the media, journal editors, and the general public. The manner in which these are handled often determines whether the outcome is a success or failure. Floor discussion is anticipated to be lively and contributory.

## 61 Statistical Aspects of Pharmaceutical Industry Proof-of-Concept Studies

Biopharmaceutical Section

**Sunday, August 6, 4:00 pm–5:50 pm**

### Statististical Aspects of Pharmaceutical Industry Proof-of-Concept Studies

❖ Surya Mohanty, Johnson & Johnson Pharmaceutical R&D, 1125 Trenton-Harbourton Road, Titusville, NJ 08560, *smohanty@prdus.jnj.com*; ❖ Glen Laird, Novartis Pharmaceuticals Corporation, 1 Health Plaza, East Hanover, NJ 07936, *glen.laird@novartis.com*; ❖ Alfred Balch, Novartis Pharmaceuticals Corporation, 2 Old Waterloo Road, Budd Lake, NJ 07828, *fredbalch@juno.com*; ❖ Jens Praestgaard, Novartis Pharmaceuticals Corporation, One Health Plaza, Building 435 Office 1176, East Hanover, NJ 07936, *jens.praestgaard@novartis.com*

Development costs for a novel drug often run as high as $800 million. Further, many potential drugs are discontinued in late-stage (phase III) development, when much of the cost has been incurred. Hence, pharmaceutical companies are looking for innovative ways to conduct so-called proof-of-concept studies for early decisionmaking. Proof-of-concept (PoC) trials are small ($n < 100$ enrolled subjects) clinical trials carried out once a novel chemical entity has passed animal toxicology and been found safe and tolerable, usually in healthy volunteers. The purpose is to give an initial determination of efficacy, or other key property and guide the company's decisionmaking about whether to sponsor full development. The trials differ from registrational efficacy trials where clinical endpoints and analyses must adhere to regulatory agency standards of proof.

## 62 Measuring Gene Expression ●

Biometrics Section, WNAR, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

## Clustering of Time-Course Gene Expression Data Using Functional Data Analysis

❖ Joon Jin Song, University of Arkansas, Department of Mathematical Sciences, SCEN 301, Fayetteville, AR 72701, *jjsong@uark.edu*; Ho-Jin Lee, Schering-Plough Corporation; Jeffrey S. Morris, M. D. Anderson Cancer Center; Sanghoon Kang, Oak Ridge National Laboratory

***Key Words:*** time-course microarray, functional data analysis, clustering analysis, principle component analysis

Time-course microarray experiments are effective in studying gene expression profile levels over a period of time. Since biological processes are dynamic and complex systems, such characteristics are essential factors in understanding how the underlying mechanisms regulate cellular processes and gene functions. We propose a unified approach for gene clustering and dimension reduction based on Functional Data Analysis to group observed curves with respect to their shapes or patterns by using the sample information in time-course microarray experiments. We apply this method to a time course microarray data set on the yeast cell cycle and a synthetic data set, and demonstrate that our method is able to identify tight clusters of genes with expression profile focused on particular phases of the cell cycle.

## Dynamic Network Analysis of Time-Course Gene Expression Data

❖ Donatello Telesca, University of Washington, Department of Statistics, BOX 354322, Seattle, WA 98195-4322, *telesca@stat.washington.edu*; Lurdes Y. T. Inoue, University of Washington

***Key Words:*** time-course gene expression, dynamic networks, compound processes, dynamic time warping, functional similarity, MCMC

Time-course gene expression data consist of RNA expression from a common set of genes, collected at selected time points, usually spanning over the domain of an underlying biological process developing over time. In order to identify gene to gene interactions, we assume that a sample of genes is a realization of a compound process where gene expression profiles over time are modeled as a random functional transformation of a reference curve. We propose measures of functional similarity and time order based on the estimated warping functions. This allows for novel inferences on dynamic network which takes full account of the timing structure of functional features associated with the gene expression profiles. We discuss the application of our model to simulated and time-course microarray data arising from animal models on prostate cancer progression.

## A Bayes Approach to Virus Gene Time Course Expression Data

❖ I-shou Chang, National Health Research Institutes, 35 Keyan Road, Zhunan Miaoli, 350 Taiwan, *ischang@nhri.org.tw*

***Key Words:*** Bernstein polynomials, Markov chain Monte Carlo, time course expression, virus gene

We propose a Bayes regression model to study the time course expression profiles of a virus gene, based on data from micro-array experiments. Since the expression level of a virus gene in a cell is constantly zero initially, increasing for a while, and then decreasing, we consider a regression model in which the mean function takes this shape restriction, with the prior introduced by Bernstein polynomials, which incorporates this geometric information naturally. We implement the inference by Metropolis-Hastings reversible-jump algorithm. We will report simulation results to indicate the performance of this method and illustrate it with the analysis of a real data set for Baculavirus.

## Comparing Distance Measures for Clustering Time-Course Microarray Data

❖ Theresa Scharl, Vienna University of Technology, Wiedner Haupstrasse 7, Vienna, 1040 Austria, *theresa.scharl@ci.tuwien.ac.at*; Friedrich Leisch, University of Munich

***Key Words:*** microarray data, cluster analysis, distance measures, R

Clustering time-course microarray data is an important tool for finding co-regulated genes and groups of genes with similar temporal or spatial expression patterns. Depending on the distance measure and cluster algorithm used, different kinds of clusters will be found. In a simulation study on various datasets, the influence of cluster algorithm and distance measure used is investigated. The focus is not to find the "best" combination of cluster algorithm and distance measure, but to gain a deeper understanding for what is going on when different methods are used.

## Bayesian Markov Chain Monte Carlo and Restricted Maximum Likelihood Study of Gene Expression Patterns across Time

❖ Feng Hong, University of Illinois, 101 Illini Hall, 725 South Wright St, Champaign, IL 61820, *fenghong@uiuc.edu*; Sandra Rodriguez-Zas, University of Illinois

***Key Words:*** Bayesian, microarrays, gene expression, REML

The performance of Restricted Maximum Likelihood (REML) and Bayesian approaches to study gene expression were compared. The expression of cDNAs in bees across 6 ages was measured using microarrays, analyzed with a linear mixed effects model with dye, time and array effects, assuming a Gaussian distribution. Bayesian non-informative and informative (based on all cDNAs) priors were evaluated. A total of 437 cDNAs were differentially expressed (P< 10-4 and maximum fold change between ages > 2) in REML. Of these, 409 and 429 cDNAs had comparable Bayes Factor (BF) values with non-informative and informative priors, respectively. The correlation of maximum fold changes between approaches was 0.995. Out of 500 cDNAs non-differentially expressed in REML, 458 cDNAs had comparable BF values. The differences between REML and Bayesian approaches may be due to the prior distribution.

## Semiparametric Analysis of Gene Expression Patterns across Ages

❖ Sandra Rodriguez-Zas, University of Illinois, 1207 W. Gregory Drive, Urbana, IL 61801, *rodrgzzs@uiuc.edu*; Bruce Southey, University of Illinois; Gene Robinson, University of Illinois

***Key Words:*** semiparametric, microarray, gene expression, mixed model

A semiparametric approach was used to identify groups of genes with distinct expression profiles during behavioral maturation in honey bees. The semiparametric approach provided unambiguous criteria to detect groups of genes, trajectories and probability of gene membership to groups. The groups and trajectories identified from bees raised in one colony were cross-validated with data from bees raised in a second colony. Ten groups of genes with distinct patterns were identified using likelihood criteria and 9 groups had highly similar trajectories in both data sets. Differences in the trajectory of the remainder group were consistent with reports of accelerated maturation in one of the colonies. Gene Ontology analysis enhanced by genome annotation

confirmed the semiparametric results and revealed that most genes with similar or related function were assigned to the same or similar groups.

### Connectivity, Module-Conformity, and Significance: Understanding Gene Coexpression Network Methods

❖ Jun Dong, University of California, Los Angeles, 376 E. Wilbur Road, Apt. 103, Thousand Oaks, CA 91360, *jundong@ucla.edu*; Steve Horvath, University of California, Los Angeles; Andy Yip, National University of Singapore

***Key Words:*** connectivity, hub gene, gene network, clustering coefficient, module, module eigengene

In gene networks, modules may correspond to pathways. Within modules, highly connected `hub'- nodes have often been found to be biologically or clinically interesting targets. This paper presents theoretical and empirical results that show how intramodular connectivity, the clustering coefficient and other fundamental network concepts are related to each other. We define the concept of node conformity for general networks and use it to relate node-specific and module-specific network concepts to each other. We arrive at the notion of a module eigengene, which effectively summarizes the gene expression profiles of the entire module. We explain the meaning of fundamental network concepts inside a module in terms of the corresponding module eigengene. We provide a simple model for explaining why intramodular hub genes have often been found to be biologically significant.

## 63 Advances in Analyzing fMRI Studies ●

Biometrics Section, ENAR
**Sunday, August 6, 4:00 pm–5:50 pm**

### Intrinsic Voxel Correlation in fMRI

❖ Daniel Rowe, Medical College of Wisconsin, Department of Biophysics, 8701 Watertown Plank Road, Milwaukee, WI 53226, *dbrowe@mcw.edu*; Raymond G. Hoffmann, Medical College of Wisconsin

***Key Words:*** fMRI, voxel correlation, k-space, brain imaging, spatial frequency, connectivity

In fMRI, complex-valued spatial frequency measurements are acquired on a rectangular grid. These measurements are transformed to a complex-valued image by inverse Fourier transform. It is known that image voxels are spatially correlated. A property of the inverse Fourier transformation is that (un)correlated spatial frequency measurements yield spatially (un)correlated voxel measurements and vice versa. Spatially correlated voxels result from correlated spatial frequency measurements. This work describes the resulting correlation structure between complex-valued voxel measurements. A real-valued representation for complex-valued measurements is introduced with an associated multivariate normal distribution. An implication is that one source of voxel correlation may be attributed to temporally autocorrelated spatial frequencies. True voxel connectivity may be less than thought previously.

### Robust Independent Component Analysis in fMRI

❖ Ping Bai, The University of North Carolina at Chapel Hill, 100 Melville Loop, Apt. 15, Chapel Hill, NC 27514, *pbai@email.unc.edu*; Young Truong, The University of North Carolina at Chapel Hill

***Key Words:*** functional magnetic resonance imaging (fMRI), independent component analysis (ICA), singular value decomposition (SVD)

By generating high quality "movies" of the brain in action, functional Magnetic Resonance Imaging (fMRI) helps us to determine which parts of the human brains are activated by different task performances. Independent Component Analysis (ICA) has been successfully applied in analyzing fMRI data, to recover all the different source signals from different parts of the brain. However, due to the high sensitivity of the MR scanner, outliers are hardly evitable in acquiring the fMRI datasets, while they cause misleading effects for the analysis. In this paper, we introduce a robust ICA procedure, which uses robust singular value decomposition (rSVD) as a data-reduction step before applying ICA. We illustrate this method by both simulated data and a real fMRI dataset. Our method is proven to be powerful and advantageous for handling the outlier-situation.

### A Semiparametric Approach To Estimate the Family-Wise Error Rate in fMRI Using Resting-State Data

❖ Rajesh Nandy, University of California, Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, *nandy@psych.ucla.edu*

***Key Words:*** fMRI, multiple comparison, semi-parametric, resampling, normalized spacings, p-value

An important consideration in any hypothesis based fMRI data analysis is to choose the appropriate threshold to construct the activation maps, which is usually based on p-values. However, there are three factors which necessitate severe corrections in the process of estimating the p-values. First, the fMRI time series at an individual voxel has strong temporal autocorrelation. The second factor is the multiple comparisons problem arising from simultaneously testing tens of thousands of voxels for activation. The third problem is the effect of inherent low frequency processes in the brain that may introduce a large number of false positives without proper adjustment. A novel semi-parametric method, using resampling of normalized spacings of order statistics, is introduced to address all the three problems mentioned above. Results using the proposed method are compared with SPM2.

### Spatio-Temporal Modeling of Functional Magnetic Resonance Imaging Data

❖ Qihua Lin, Southern Methodist University, Department of Statistical Science, 3225 Daniel Ave, Dallas, TX 75275-0332, *qlin@smu.edu*; Patrick S. Carmack, The University of Texas Southwestern Medical Center at Dallas; Richard F. Gunst, Southern Methodist University; William R. Schucany, Southern Methodist University; Jeffrey S. Spence, The University of Texas Southwestern Medical Center at Dallas

***Key Words:*** spatiotemporal models, fMRI

Functional magnetic resonance imaging (fMRI) is used in fields, such as neuroscience, to study the functioning of human brains. Data from fMRI experiments are very complex. A rich spatial and temporal correlation structure is inherent in such data and the signal-to-noise ratio is generally low. A class of spatiotemporal models is introduced to model such data. Spatially varying drift and hemodynamic response functions are modeled to explain some large-scale variation. Spatially and temporally correlated small-scale variations are modeled by autoregressive moving average processes. An algorithm for model identification and parameter estimation is outlined. Comparisons to SPM analysis are made using real fMRI data. These methods offer a strong alternative to SPM for analyzing fMRI data.

### Interpreting Experience-Based Cognition from fMRI

❖ Rajan Patel, Rice University, 17 Via Colinas, Westlake Village, CA 91362, *rajan@alumni.rice.edu*; F. DuBois Bowman, Emory University; Ying Guo, Emory University; Gordana Derado, Emory University; Lance Waller, Emory University; Amita K. Manatunga, Emory University

***Key Words:*** fMRI, prediction, support vector machines, boosting, principle components

Recent advances in statistics and functional neuroimaging allow the use of distributed patterns of brain activity to predict subjective human experience. The Pittsburgh Brain Activity Competition set forth to challenge groups from multiple disciplines to use various statistical and data mining techniques to infer subjective experience from fMRI. Individuals were scanned watching three segments of videos after which each provided behavioral time vector ratings of experience coding categories ( i.e. human faces, tools, emotion). Groups were provided with functional scans from each video but time vector ratings of only the first two, with the goal to predict vector ratings for the third video. In this talk, we describe and compare several approaches, from iterative boosting and supervised principal components to support vector machines.

### Detecting Cerebral Activation from Functional Magnetic Resonance Imaging Data

❖ William Baumann, Iowa State University, 1405 Maxwell Ave., Ames, IA 50010, *wbaumann@iastate.edu*; Ranjan Maitra, Iowa State University

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive radiologic tool used to detect activity in the human brain. An usual approach is to perform hypothesis-testing on a general linear model which fits the time-course sequence at a voxel to the input sequence. Multiple comparisons are then used to identify activation regions. We propose, instead, an estimation approach which fits a mixture model to the distribution of the p-value at each voxel. Mixing proportions and the parameters of the non-central t-density are estimated in a hierarchical setting which incorporates spatial context. These are then used to identify regions of activated voxels with high probability.

## 64 Methodological Issues in Genetics Studies ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### An Importance Sampling Procedure for Obtaining Confidence Intervals of Disease Loci with General Pedigree Data

❖ Shuyan Wan, The Ohio State University, Cockins Hall, Room 404, 1958 Neil Avenue, Columbus, OH 43210, *sabrina@stat.ohio-state.edu*; Shili Lin, The Ohio State University

***Key Words:*** importance sampling, pedigree data, CSI, LOD

To localize a disease gene in a whole-genome study, we extended the confidence set approach idea of Lin et al.[2001] to general pedigree data assuming a known disease model. We test every position on the chromosomes based on the LOD score statistic. Those positions not rejected form a confidence interval for the disease locus. Three approaches were proposed to perform the tests, including a crude procedure for testing every fixed distance, a refined procedure based on inference for LOD score and an importance sampling (IS) procedure. Performance of the approaches was evaluated based on simulation studies. The IS procedure stands out in terms of the power and computing efficiency. Asymptotic properties of the null distribution was also examined. We will compare the performance of the asymptotic approximation with that of the simulation methods and apply our methods to a real data set.

### Correcting for Measurement Errors in Structured Association Tests

❖ Jasmin Divers, The University of Alabama at Birmingham, 1665 University Blvd., Suite 327, Birmingham, AL 35294, *jdivers@uab.edu*; Laura K. Vaughan, The University of Alabama at Birmingham; David Redden, The University of Alabama at Birmingham; Jose R. Fernandez, The University of Alabama at Birmingham; David B. Allison, The University of Alabama at Birmingham

***Key Words:*** structure association tests, linkage desequilibrium, measurement error, population stratification, genetic background variables, ancestry estimation

Although long blocks of linkage disequilibrium in admixed individuals offer the potential for admixture mapping, admixture can confound association testing. Confounding occurs when the allele frequencies of markers to be tested and the distribution of the phenotype of interest differ among the parental populations. Various methods have been proposed to control for such genetic heterogeneity by estimating a measure of genetic background (GB) to be used as covariate in the test of association. However, these approaches do not account for variation between estimated and true GB values. Including variables measured with error in a regression analysis reduce the ability to control for confounders resulting in a higher than expected type I error rate. We present several measurement error correction methods applicable to this problem and conduct simulation studies to compare their efficacy.

### Genomic Control for Association Studies When the Genetic Model Is Unknown

❖ Gang Zheng, National Heart, Lung, and Blood Institute, 6701 Rockledge Drive, MSC 7938, Bethesda, MD 20892, *zhengg@nhlbi.nih.gov*; Boris Freidlin, National Cancer Institute; Joseph Gastwirth, The George Washington University

***Key Words:*** genomic control, genetic association, case-control data, general association, trend test, population stratifications

Population-based case-control association studies can be affected by population stratification, which may result in a higher than nominal rate of false positive results. One approach to preserving the nominal type I error is to apply genomic control (GC), which adjusts the variance of the Cochran-Armitage trend test. When the underlying genetic model is known, GC can be applied to the corresponding optimal trend tests. In practice, however, the mode of inheritance is unknown. The genotype-based chi-square test for a general association does not depend on the underlying genetic model. This test has a chi-square distribution with two degrees of freedom. We discuss how to apply GC to this general association test.

### Pedigree Disequilibrium Test for X-Chromosome Markers

❖ Jie Ding, The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210, *ding@stat.ohio-state.edu*;

Shili Lin, The Ohio State University

***Key Words:*** association test, x-chromosome, missing genotypes, Monte Carlo

Because of the need for fine mapping of disease susceptibility loci and the availability of dense genetic markers, many forms of association tests have been developed. Some of them only apply to nuclear families and some of them can not handle missing genotypes. Here, we extended the pedigree disequilibrium test (PDT) to marker loci on X-chromosome. Furthermore, we treated families with missing genotypes by averaging over a Monte Carlo sample of those genotypes, which are generated conditioning on known genotypes. Association test based on haplotypes can also be carried out in this framework. The method was compared to existing approaches through simulation, and was applied to analyze a multiple sclerosis dataset.

### Incorporating Endophenotypes into Allelic Association Studies

❖ Chao Hsiung, National Health Research Institutes, 35 Keyan Road, Div of Biostatistics NHRI, Zhunan Miaoli, 350 Taiwan, *hsiung@nhri.org.tw*

***Key Words:*** endophenotype, concordant sibpair, discordant sibpair, association study

For a genetic study in which there are concordant and discordant sibships for a complex disease trait and there are also available the measurements of other endophenotypes for each of the individuals, we describe a test for association that utilizes nonparametrically the additional endophenotypes. Data from the Stanford Asian Pacific Program in Hypertension and Insulin Resistance (SAPPHIRe) are used to illustrate the method.

### Allowing for Etiologic Heterogeneity by Disease Subtype Increases the Power of Tests for Genetic Association

❖ Peter Kraft, Harvard University, Departments of Epidemiology and Biostatistics, 665 Huntington Avenue, Bldg II, 2nd Floor, Boston, MA 02115, *pkraft@hsph.harvard.edu*; Sholom Wacholder, National Cancer Institute; Nilanjan Chatterjee, National Cancer Institute

***Key Words:*** genetic epidemiology, categorical outcome, polytomous regression

Many complex diseases can be broken into several pathological or clinical subtypes. A putative disease susceptibility gene may influence risk of one, several, or all of these subtypes. We present two simple procedures based on polytomous logistic regression to test the global null hypothesis that a polymorphism is not associated with variation in disease outcome. The first compares the null model to a saturated alternative; the second uses a two-stage model to reduce the number of free parameters in the alternative. When there is no subtype-specific genetic effect, these procedures are only slightly less powerful than standard dichotomous logistic regression procedures; when there is a subgroup-specific effect, these procedures can be dramatically more powerful. We illustrate these procedures with an application to a case-control study of breast cancer, classified by ER and PR status.

### A Multiple Test Procedure Controling Type I Error for Genome Scan Association Studies using HapMap Data

❖ Renfang Jiang, Michigan Technological University, Department of Mathematical Sciences, Michigan Tech University, Houghton, MI 49931, *rjiang@mtu.edu*; Jianping Dong, Michigan Technological University; Shuanglin Zhang, Michigan Technological University; Qiuying Sha, Michigan Technological University

***Key Words:*** genome scan, hapmap, haplotype blocks, multiple testing, association study, complex diseases

Genome scan becomes more and more a preferable choice for studying association between markers and complex diseases. Hundreds of thousands of SNPs are tested for association with a complex disease. The problem of multiple testing must be addressed before obtaining any meaningful association without making false positive findings. There are many existing methods to deal with this problem. Some of them are too conservative or too general, and they are not designed for dealing with a specific genome scan problem. We propose to find a new method, which can be adapted to a sample from a particular population and for a specific chromosome segment. The HapMap data provided a good sample of genotype data of the whole genome from several important populations. We will build our method based on the HapMap data, and to use the haplotype blocks information provided by the data.

## 65 Sample Survey Design I ●

Section on Survey Research Methods
**Sunday, August 6, 4:00 pm–5:50 pm**

### Model-Based Sampling Designs for Optimum Estimation

❖ Sun Woong Kim, Dongguk University, JungGu Pil Dong 3 Ga 26, Seoul, 100-715 South Korea, *sunwk@dongguk.edu*; Steven G. Heeringa, University of Michigan; Peter W. Solenberger, University of Michigan

***Key Words:*** superpopulation model, probability sampling, optimization problem

Concerned more with bias than variance, many survey samplers prefer design-based inferences over model-based approaches. There have been many studies of methods for reducing variances of estimates through probability sampling. In designing samples, we may have a belief about or an experience with a population, specified as a superpopulation model. In this paper we show how to determine the optimum sampling plan to minimize sampling variance under a variety of model assumptions. These optimal sampling plans will be based on inclusion probability proportional to size (IPPS) sampling and will satisfy certain desirable properties with respect to variance estimation. The presence of the intercept term in a superpopulation model is an interesting issue in model-based design. We will show that these approaches depend on the form of the variance being considered as well as model assumptions.

### Optimum Allocation in Two-Stage and Stratified Two-Stage Sampling for Multivariate Surveys

❖ M. G. M. Khan, The University of the South Pacific, School of Computing, Information, and Mathematical Science, Suva, 1168 Fiji, *khan_mg@usp.ac.fj*; Munish A. Chand, The University of the South Pacific

***Key Words:*** multivariate two-stage sampling, multivariate stratified two-stage sampling, first-stage sampling units, second-stage sampling units, optimum allocation, nonlinear programming problem

When more than one characteristics are under study it is not possible for one reason or the other to use the individual optimum allocation of first-stage and second-stage sampling units to each stage and to various strata while using two-stage and stratified two-stage sampling designs. In such situations some criterion is needed to work out an acceptable allocation which is optimum for all characteristics in some sense. In this paper the problems of the optimum allocation in multivariate two-stage and multivariate stratified two-stage sampling are formulated as Nonlinear Programming Problems (NLPP). The NLPPs are then solved using Lagrange multiplier technique and explicit formulas are obtained for the optimum allocation of the first-stage and second-stage sampling units.

### An Application of Genetic Algorithms to Multivariate Optimal Allocation in Stratified Sample Designs

❖ Charles Day, U.S. Internal Revenue Service, P.O. Box 2608, Washington, DC 20013, *Charles.D.Day@irs.gov*

***Key Words:*** evolutionary algorithm, stochastic search

Sampling statisticians are often interested in allocating sample units to strata in a stratified probability sample in such a way that variance constraints for two or more variables of interest are satisfied with least cost. This problem falls into the class of convex mathematical programming problems. It is usually solved numerically, using a program that searches for an arbitrarily close approximation to the optimum. Genetic Algorithms (GA's) use a model of computation based on biological evolution to perform a stochastic search of the solution space of an optimization problem. This paper reports on the use of a GA to solve the multivariate optimal allocation problem. Because of the flexibility of fitness functions (objective functions) in GA's, this approach has the potential to be extended to meet additional optimality criteria involving more complicated objectives.

### The Effect of the Number Eligible and Number Selected within Households on Reported Income and Other Socioeconomic Characteristics in the 2004 NSDUH

❖ Tania Robbins, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *trobbins@rti.org*

***Key Words:*** household composition, pair selection, income questions

In the 2006 National Survey on Drug Use and Health (NSDUH), a portion of the respondents will receive a reduced set of income questions. Some of these respondents will come from the general selection algorithm that allows selection of 0, 1, or 2 persons at all households, while the rest will come from households with just 1 person selected. In a broader study, we wish to investigate the impact that the new income questions will have on reported outcomes of personal and family income, government assistance, poverty, and health insurance. But in this paper, we initially wish to investigate the impact of both the number selected and number eligible on those reported outcomes, based on data from the 2004 NSDUH. These results will be used to inform later analyses on the potential differences in reported income and related measures due to selection and eligibility into the survey.

### NASS/USDA Area Frame Sample Allocation for Estimation of Number of Farms Not on the Ag Census Mailing List

❖ Floyd Spears, Harding University, Box 10764, Searcy, AR 72149, *mspears@harding.edu*; Raj S. Chhikara, University of Houston-Clear Lake; Charles R. Perry, National Agricultural Statistics Service; Phillip S. Kott, National Agricultural Statistics Service

***Key Words:*** optimum sample allocation, area frame sample, modeling standard deviations, agriculture census, not on mailing list

USDA/NASS needs to estimate the number of farms that are not on the census mailing list (NML) for the 2007 Ag Census. Additional sample segments are planned during its 2007 annual ag survey to determine estimates for several NML items (number of farms, and subdomains corresponding to several minority and specialty farms). Stratum standard deviations from the 2002 area frame sample data are modeled in terms of certain NML farm characteristics. The NASS multivariate allocation procedure is applied to determine sample allocations for several different sets of ag items, which include NML farms and the regular annual ag survey items. Various combinations of the actual 2002 design allocation and those obtained using the NML items are investigated. The allocation that most closely meets the sample size and precision goals is identified.

### Simple Power Calculations: How Do We Know We Are Doing It the Right Way?

❖ Michael Vorburger, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *vorburger@rti.org*

***Key Words:*** power analysis, sample size, minimum detectable difference, arcsine transformation

Sample size and power determination is a crucial step in setting up efficient studies. It ensures that there are enough sampling units to enable the detection of anticipated effects. After determining the hypothesis test on a parameter, the significance level, estimates of parameters needed to compute the power function; the researcher explores several scenarios by considering different power levels and effect size (also known as minimum detectable difference). We are interested in the problem of estimating power and the effect size for specific sample sizes when assessing the difference between two proportions. We compare the accuracy of alternative methods and discuss the applicability of these results for practitioners.

### Estimation in Network Populations

❖ Mike Kwanisai, National Opinion Research Center, 1662K Carlyle Drive, Crofton, 2114, *kwanisai-mike@norc.org*

***Key Words:*** snowball sampling, link-tracing, Bayesian,

In social networks, subjects are linked to one another forming structures that are usually of interest. The subjects may have a variable of interest which, in general, can only be observed after the subjects have been interviewed. Link-tracing sampling designs are commonly used to draw samples from such network populations. Besides being convenient, these sampling designs produce biased samples that make the estimation of quantities difficult. We discuss link-tracing designs in which only a fraction of relations are traced to include subjects into the study. Using simulated and real study datasets, we demonstrate how estimation for population quantities can be done.

## 66 Sample Survey Quality I ●

Section on Survey Research Methods
**Sunday, August 6, 4:00 pm–5:50 pm**

## Nonresponse Bias Studies: 2003--2004 School and Staffing Survey

❖ Robyn Sirkis, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233, *Robyn.B.Sirkis@census.gov*; Bac Tran, U.S. Census Bureau; Phyllis Singer, U.S. Census Bureau

***Key Words:*** nonsampling, nonresponse, bias, substantive difference, SASS

Nonresponse error is one kind of nonsampling error. Nonresponse error is a function of response rates and the differences in responses between respondents and nonrespondents. A low response rate cannot cause bias alone, however it suggests the potential for bias to occur. Nonresponse error can also cause misinterpretation of the results published. The Office of Management and Budget (OMB) is considering a mandate to conduct nonresponse bias studies if the expected unit response rate is below 80 percent and the expected item response rate is below 70 percent. The National Center for Education Statistics requested that a Nonresponse Bias Study be conducted for the 2003-2004 School and Staffing Survey. Our discussion focuses on portions of the methodology applied, results, lessons learned, as well as future endeavors for nonresponse bias studies.

## Call Efforts and Relational Estimates: Preliminary Findings

❖ Chung-tung Lin, U.S. Food and Drug Administration, 505 Lamberton Drive, Silver Spring, MD 20902, *chung-tung.lin@fda.hhs.gov*

***Key Words:*** nonresponse, RDD, telephone, survey, relation

Despite the widespread concern that low response rates may produce less representative samples and contribute to larger non-response biases, several recent studies have failed to show widespread or significant discrepancies of estimated statistics in RDD surveys. Built on the literature, this study explores potential influences on relational estimates of a survey by excluding respondents who required more rigorous call effort, such as more call attempts and refusal conversion. The study will show how the exclusion would affect (1) percentage estimates and (2) estimated relationships between outcomes and covariates. This study is an important addition to the literature as RDD survey data are often used to examine and estimate how various question items are associated with each other and to understand psychological, behavioral, and socio-demographic differences.

## Estimation of Low Incidence Rates under Selection Bias

❖ Bin Wang, University of South Alabama, 9713 Concord Place, Mobile, AL 36695, *bwang@jaguar1.usouthal.edu*; Jiayang Sun, Case Western Reserve University

***Key Words:*** missing data, semi-parametric, biased sampling

This paper is motivated by a study of the cancer risks of Vietnamese Americans along the central Gulf Coast. In the study, researchers encountered two difficulties: selection bias and low incidence rate. This may cause under-estimate in estimation by a standard logistic regression model. In this paper, the authors considered using a semiparametric method to evaluate the risk of rare events from biased data with nonignorable missing values. A generalized additive model is used and a modified iterative reweighted least square estimator is developed to correct the selection bias and account for missing values. The new method will be compared with existing methods, and simulation will be performed to illustrate the performance of the new estimators.

## Assessment of Diagnostic Tests in the Presence of Verification Bias Using Multiple Imputation and Resampling Methods

❖ Michael P. McDermott, University of Rochester Medical Center, Department of Biostatistics and Computational Biology, 601 Elmwood Avenue, Box 630, Rochester, NY 14642, *mikem@bst.rochester.edu*; Hua He, University of Rochester Medical Center

***Key Words:*** bootstrap, missing data, multiple imputation, sensitivity, specificity, verification bias

Sensitivity and specificity are common measures of the accuracy of a diagnostic test. The usual estimators of these quantities are unbiased if data on the diagnostic test result and the true disease status are obtained from all subjects in an appropriate sample. In some studies, verification of the true disease status is performed only for a subset of subjects, possibly depending on the result of the diagnostic test and other characteristics of the subjects. Estimators of sensitivity and specificity based on this subset of subjects are typically biased. We view this as a missing data problem and apply commonly-used techniques in the survey sampling literature, multiple imputation and bootstrap resampling, to derive estimators and corresponding confidence intervals that are corrected for this verification bias. Comparisons are made between these and existing bias-correction methods.

## Treatment of Spatial Autocorrelation in Geocoded Crime Data

❖ Krista Collins, Statistics Canada, 16G R.H. Coats Building, Tunneys Pasture, Ottawa, ON K1A0T6 Canada, *krista.collins@statcan.ca*; Colin Babyak, Statistics Canada

***Key Words:*** spatial autocorrelation, spatial lag model, crime data

In examining the spatial distribution of crime in city neighborhoods, the measurements at one location may be influenced by the measurements at neighboring or nearby locations. This can result in spatial autocorrelation, which violates the assumption of independence among observations. We examine the basic theory and methods used to analyze spatial lattice data (data aggregated to regions as opposed to discrete points). Our focus is on spatial autoregressive models for lattice data and how to account for the spatial structure of the data. In a study funded by the National Crime Prevention Center at Public Safety Canada, we implement the spatial lag model in a study of crime rates in the neighborhoods on the Island of Montreal, Canada. We examine the differences in the models obtained with and without modeling the spatial effects.

## Assessing Population Coverage in a Health Survey

❖ Karen Davis, National Center for Health Statistics, 3311 Toledo Road, Rm 3213, Hyattsville, MD 20782, *kedavis@cdc.gov*; Chris Moriarity, National Center for Health Statistics

***Key Words:*** sample survey, oversampling, population coverage

The National Health Interview Survey (NHIS) is one of the major data collection programs of the National Center for Health Statistics (NCHS). In 2002, research was conducted to assess options for oversampling minority (Hispanic, Black, and Asian) persons age 65 years and older, to meet the goal of improving the precision of estimates while retaining the same precision for other estimates and keeping the overall sample size constant. Beginning with the 2006 NHIS, the protocol of selecting one sample adult per household is being continued, but minority persons age 65 years and older will have an increased probability of selection. We describe the research that has been conducted to assess the population coverage for sample adults, and present results from several simulations to estimate the expected increase in sample yield for minority persons 65 years and over.

### A Study of IRS Administrative Payroll as a Substitute for Missing Payroll

❖ Melvin McCullough, U.S. Census Bureau, 4700 Silver Hill Road, Building 3 Room 1176, Suitland, MD 20746, *melvin.mccullough@census.gov*

***Key Words:*** general imputation, replication, bias, single units

The Business Expenses Survey (BES) collects operating expense data of businesses in the retail, wholesale and services sectors of the United States economy. Traditionally, missing payroll expense data have been imputed using a "ratio of identicals." This method of imputing for missing payroll takes a ratio of the weighted sum of non-missing payroll to the weighted sum of operating expenses for respondents within an industry and multiplies this result by reported operating expenses to obtain an estimate for missing payroll for a given case. The availability of IRS administrative payroll data as a possible alternative to the traditional ratio imputation resulted in a study. Our objective was to determine if IRS administrative payroll could be reliably substituted for missing BES payroll for single-establishment businesses. For each industry, some of the payroll respond

## 67 Industrial Applications ● ✪

Section on Quality and Productivity, Section on Physical and Engineering Sciences

**Sunday, August 6, 4:00 pm–5:50 pm**

### Hierarchcial Modeling Using GLMs To Improve Yield

❖ Christina Mastrangelo, University of Washington, 2319 44th Ave., SW, Seattle, WA 98116, *mastr@u.washington.edu*; Naveen Kumar, University of Washington

***Key Words:*** generalized linear models, hierarchical modeling

In a complex manufacturing environment such as semiconductor manufacturing, there are hundreds of interrelated processes. In such an environment, modeling the impact of critical process parameters on final performance metrics such as defectivity or yield is a challenging task. Issues such as low number of observations compared to process variables, difficulty in formulating a high dimensional design matrix, and missing data due to failures pose serious challenges in using empirical modeling techniques. Our approach is to use generalized linear modeling in a hierarchy to understand the impact of key process and subprocess variables on the system output. Issues such as bias and variance estimation are considered. The hierarchical GLM approach helps not only in improving output metrics, but also in identifying and improving subprocess variables attributable to poor performance.

### Sequential Analysis on Misspecified Distributions

❖ Theresa Utlaut, Intel Corporation, 29430 Dutch Canyon Road, Scappoose, OR 97056, *theresa.l.utlaut@intel.com*; Kevin Anderson, Intel Corporation

***Key Words:*** sequential, SPRT

Sequential analysis is ideal in environments where resources are limited or delayed decisions result in penalties. The sequential probability ratio test (SPRT) has been shown by Wald and others to guarantee the minimum average sample number to make a decision. As with many statistical tests, the SPRT is based on an underlying assumed distribution. The presentation will investigate the impact of the errors and average sample number on the SPRT when the distributional assumptions have been both slightly and grossly violated. It will demonstrate the use of misspecified SPRT's in a semiconductor manufacturing application, and recommend some potential approaches for successful decision-making in the face of misspecification.

### On Robust Statistics

❖ Kevin Anderson, Intel Corporation, 1600 Rio Rancho Blvd., Mailstop RR5-454, Rio Rancho, NM 87124-109, *kevin.c.anderson@intel.com*

***Key Words:*** robust, inference

Classical statistical inference is performed using a set of mathematically rigorous tools that have been studied and applied through the centuries. Inference is based on observations and assumptions. These assumptions are mathematically convenient rationalizations of knowledge and beliefs. However, classically optimum procedures often behave very poorly in the presence of even slight violations of the model assumptions. Robust Statistics is the statistics of approximate parametric models. In recent decades, a number of robust tools have been proposed to deal more effectively with the problems. This presentation will discuss robust statistics, compare them with classical methods, and demonstrate their use and utility. It will also propose some ideas of why these are relatively unused among industrial statisticians, and make some specific recommendations for their use.

### Robust Analysis of Variance: Process Design and Quality Improvement

❖ Avi Giloni, Yeshiva University, 500 W. 185th Street, New York, NY 10033, *agiloni@yu.edu*; Sridhar Seshadri, New York University; Jeffrey Simonoff, New York University

***Key Words:*** ANOVA, Taguchi, robust design, quality engineering, robust statistics, outlier

We discuss the use of robust analysis of variance techniques as applied to quality engineering. Our goal is to utilize methodologies that yield similar results to standard methods when the underlying assumptions are satisfied, but also are relatively unaffected by outliers. We study several examples to illustrate how using standard techniques can lead to misleading inferences about the process being examined, which are avoided when using a robust analysis. We further demonstrate that assessments of the importance of factors for quality design can be seriously compromised when utilizing standard methods as opposed to robust methods.

### Statistical Quality Control of Loadboards for Electronic Package Testers

❖ Meihui Guo, National Sun Yat-sen University, Department of Applied Math, National Sun Yatsen Univ, Kaohsiung, 804 Taiwan, *guomh@math.nsysu.edu.tw*; Yu-Jung Huang, I-Shou University; Ming-Kun Chen, I-Shou University

***Key Words:*** statistical quality control, loadboard, lifetime, preventive maintenance

Each packaged IC undergoes a rigorous process of electrical / electronics testing to assure the quality of the finished products before shipping to the manufacturer as well as the customer. These electrical testing machines were composed of the machine tester, test interface unit, and the handler. The TIU, which is mainly composed of the loadboards, provides the electrical and mechanical connection of the device under test and the tester. In this work, we perform statistical quality control of loadboards for electronic package testers. In particular, counts of failure and lifetime of loadboards are studied. Control charts of mean time between failure are established. Optimal preventive maintenance schedule of planned maintenance actions is then proposed to aim at the prevention of breakdowns and failures. The ideal preventive maintenance program would enable more accurate and effective

### Statistical Monitoring of Multistage Processes

❖ Fugee Tsung, The Hong Kong University of Science and Technology, IELM Department HKUST, Clear Water Bay, Kowloon, 852 China, *season@ust.hk*

***Key Words:*** statistical process control, quality control, multistage operations, SPC

Most manufacturing processes comprise not a single stage but a large number of stages. Engineering modeling of multistage processes considering physical and mechanical laws in a linear state space form has been studied extensively. Such modeling describes the quality linkage among stages. However, recent research on statistical monitoring of multistage processes usually makes no use of these engineering models. Here, a statistical process control (SPC) method for a process with multiple stages is proposed based on an engineering model.

### Stochastic Models for Predicting Product Failure Rate of Parenterals Due to Particulate Matter

❖ Chi-Hse Teng, Pfizer Inc., Science Center Drive, San Diego, CA 92121, *chi-hse.teng@pfizer.com*

***Key Words:*** particulate matter, CMC, quality control, prediction, parenterals

Since the 60s, many articles have discussed the particulate issue in parenteral product, such as safety concerns, sources of particulates, and analytical chemistry methods for detecting particulates. However, the method for predicting the chance of the final product meeting the specifications limit has not been well established, which is useful for the developer and reviewer. The failure rate can be established empirically or predicted by models. In many cases, due to limited resources, it is not possible to estimate the failure rate empirically. We will discuss several stochastic models for different situations, including Poisson renewal processes model, overdispersion model for limits with single-size category and multiple-size category, and log normal model.

## 68 Risk Assessment in Business and Finance ● ✪

Section on Risk Analysis, Section on Quality and Productivity

**Sunday, August 6, 4:00 pm–5:50 pm**

### On the Application of the Latent-Variable Model To Predict Business Default

❖ K. Paul Chin, Dun & Bradstreet, Inc., 103 JFK Parkway, Short Hills, NJ 07078, *chinp@dnb.com*; Edgar Ortiz, Dun & Bradstreet, Inc.; Jianjing Ling, Dun & Bradstreet, Inc.

***Key Words:*** latent variable model, business default, bivariate logit regression model

This paper discusses the application of the Latent variable model to predict business default. In this approach business default is viewed as the result of the complex interaction of the underperformance of several key business indicators. Our findings contrast results attained leveraging latent variable model with those using traditional bivariate logit regression model.

### Mixture Models Applied to Reject Inference

❖ Billie Anderson, The University of Alabama, ISM Department, Box 870226, Tuscaloosa, AL 35487, *billiesueanderson1@yahoo.com*; J. Michael Hardin, The University of Alabama; Ana Landeros, The University of Alabama; Michael Conerly, The University of Alabama

***Key Words:*** credit scoring, reject inference, mixture models, EM algorithm

Credit Scorecards are commonly built using data available within an organizations transactional database. Such data, however, will only contain information for those applicants who were ‘accepted’ or awarded credit by the organization; data will not be available for those applicants who were ‘rejected’. The use of reject inference to adjust credit scorecard models for the missing data represented by rejected loan applications is common practice and several approaches are used in today’s financial industry. In this paper, we investigate the use of mixture models as an alternative approach for reject inference. To estimate the parameters of the mixture model, we use the EM algorithm (Dempster, Laird, and Rubin, 1977) wherein the data associated with the ‘rejected’ applicants is treated as missing completely at random (Rubin, 1976).

### How To Address Click Fraud in Pay-per-Click Programs

❖ Vincent Granville, Authenticlick, 2428 35th Ave., NE, Issaquah, WA 98029, *vgranville@authenticlick.net*

***Key Words:*** click fraud, scoring, fraud detection, risk analysis, web mining, data mining

The problem of click fraud and nonvalid clicks is nearly as old as paid-per-click advertising. Only recently has it been acknowledged by major search engine companies. It represents the most serious threat to the online advertising industry. Although there is no formal definition to click fraud, a good description can be found at datashaping.com/security.shtml. We review various state-of-the-art solutions to address click fraud in pay-per-click programs, including proprietary IP blacklists, entrapment, ad-hoc design of experiment, and detection of false positives. Our core contribution is a scoring system tied to the advertiser’s ROI and matching conversion rate distributions. New original fraud cases will be discussed, including impression and ad relevancy fraud, automated clicks from shareholders and political activists, and accidental click fraud generated by email spammers.

### Application of Kernel Methods to Fraud Detection

❖ Ravi Mallela, Equbits, 1844 Camino Pablo, Moraga, CA 94556, *ravi@equbits.com*

***Key Words:*** SVM, kernel, fraud, financial, services

Kernel Methods were introduced in the early 1990s and have become a popular method for solving classification and regression modeling problems. Over the last few years, a number of applications of kernel methods have been presented. This paper assesses the applicability of kernel methods to challenging modeling problems in the financial services industry. We will provide an overview of kernel methods and new knowledge-based approaches that allow modelers to extract meaningful business rules from SVMs. An example assessing merchant risk management will be used for illustration.

### Partial Hedging Using Malliavin Calculus

❖ Lan Nygren, Rider University, 567 Grant Street, Newtown, PA 18940, *lnygren@rider.edu*; Lakner Peter, New York University

***Key Words:*** contingent claim, partial hedging, Malliavin calculus, lookback option, convex duality, shortfall risk

We consider the problem of partial hedging of derivative risk when the investor's attitude towards the shortfall is captured by a general convex loss function. We derive the dual problem from the Legendre-Fenchel transform of the loss function and interpret the optimal strategy as the perfect hedging strategy for a modified claim. However, computation of the hedging strategy for this modified claim requires other tools than the well-known "delta" hedging technique since the associated optimal wealth process is usually too complicated to possess an analytic expression. We show how the Malliavin calculus approach can be used to derive the hedging strategy for the modified claim. The advantage of this probability based hedging approach is illustrated in a couple of explicitly worked out examples.

### An Econometric Model for Insurance Underwriting Using Bivariate Zero-Inflated Count Models

K. Paul Chin, Dun & Bradstreet, Inc.; ❖ Edgar Ortiz, Dun & Bradstreet, Inc., 103 JFK Parkway, Short Hills, NJ 07078, *ortized@dnb.com*

***Key Words:*** bivariate zero-inflated count model, overdispersion, insurance claim intensity, insurance underwriting risk

This paper introduces the application of Bivariate Zero-Inflated Count models to predict the joint event distribution of account delinquency and insurance claim intensity. Insurance underwriting decisions need to incorporate the joint probability of occurrences of these two events to get a more accurate assessment of insurance risk. This paper introduces the framework of bivariate zero-inflated count models to derive more accurate predictions of insurance underwriting risk.

### Investigating the Determinants of Financial Harm and Predatory Lending through RDD and Victim Population Surveys

❖ Danna Moore, Washington State University, SESRC, PO box 644014, Pullman, WA 99164-4014, *moored@wsu.edu*

***Key Words:*** survey, predatory lending, mortgage harm, financial harm

The increase in subprime mortgage credit has allowed many individuals that have less than prime-quality credit to borrow money. Many of these borrowers have fallen prey to predatory lenders. Little is known about these victims. Our analysis takes advantage of a unique survey of mortgage borrowers involved in a large predatory lending settlement in WA State and makes comparisons with the general population. The purpose is to examine the factors affecting the likelihood borrowers will fall victim to predatory lending. We measure the extensiveness of damage. Profiles revealed that the entire gamut of demographics were represented and harm is most associated with financial behaviors. Financial literacy is a complex aggregate analysis of factors knowledge, experience, and behaviors. Logit analyses provide the predicted odds of factors characterizing financial harm.

## 69 Modeling

Section on Statistical Computing
**Sunday, August 6, 4:00 pm–5:50 pm**

### Finite Elements Methods for Density Estimation

❖ George Terrell, Virginia Polytechnic Institute and State University, Statistics Department

Virginia Tech, Virginia Tech, Blacksburg, VA 24061, *terrell@vt.edu*

***Key Words:*** polynomial splines, b-splines, penalized least-squares, ash methods

Nonparametric density estimates may be efficiently and rapidly computed and the results compactly stored using finite elements methods. These consist of polynomial splines estimated by least squares, which may then be smoothed by a combination of roughness penalties and direct averaging of the sort used in ASH methods. The results closely approximate popular classical methods such as kernels, but are generally faster to obtain. Boundary adjustments and variable smoothing are easy to include. Multivariate extensions are straightforward, using tensor-product splines.

### On the Mixture of Multivariate Skew Normal Distributions

❖ Jack C. Lee, National Chiao Tung University, Institute of Statistics, Hsinchu, 300 Taiwan, *jclee@stat.nctu.edu.tw*; Tsung-I Lin, National Chung Hsing University

***Key Words:*** EM algorithm, Fisher information, normal mixture model, skew normal mixtures, stochastic representation, truncated normal distributions

A finite mixture of distributions, particularly the use of normal components, has received much attention and is known to be powerful for modeling an extremely wide variety of random phenomena. However, the usefulness of normal mixture models is somewhat limited, and there still exists drawbacks in various applied problems. In this paper, we introduce a flexible mixture modeling framework using the multivariate skew normal distribution. A feasible EM algorithm is developed for carrying out maximum likelihood estimation of parameters. In addition, a general information-based method for obtaining the asymptotic covariance matrix of maximum likelihood estimates is presented. We apply the procedures to a real multivariate dataset and compare the results with those from fitting Gaussian mixtures.

### Latent Regression

❖ Thaddeus Tarpey, Wright State University, 120 MM, Dayton, OH 45435, *thaddeus.tarpey@wright.edu*; Eva Petkova, Columbia University

***Key Words:*** EM algorithm, beta distribution, placebo response

The finite mixture model is cast as a regression model with a latent Bernoulli predictor. The mixture model is then generalized by allowing the latent predictor to have a continuous distribution on the interval (0,1). An EM algorithm is used to estimate parameters of the latent regression model. Examples and simulations are given to illustrate the latent regression model. In particular, the latent regression model is used to study the placebo effect and determine if there exist two distinct classes of individuals (those who do and do not exhibit a placebo effect), or if the placebo effect varies continuously among individuals with no distinct latent classes.

### Estimation for Finite Mixture Multinomial Models

❖ Nagaraj Neerchal, University of Maryland Baltimore County, Department of Math and Statistics, 1000 Hilltop Circle, Baltimore, MD 21250, *inference@gmail.com*; Minglei Liu, Medtronic, Inc.; Jorge Morel, Procter & Gamble

***Key Words:*** mixture model, Fisher's scoring algorithm, EM algorithm, multinomial

Mixture of multinomial distributions is important for both theoretical and practical reasons. Because of the specialty of this model, it is not easy to get the maximum likelihood estimates of the parameters. Several methods, including Fisher's Scoring algorithm and EM algorithm are available in the literature to get the MLE numerically. The authors consider an approximate of the information matrix of the mixture of multinomial model and proposes an Approximate Fisher's Scoring algorithm. We also investigate the properties of this algorithm and compare it with the other well known algorithms by simulation.

### Predictive Discrepancy Using Full Cross-Validation for Regression Models

❖ Mark Greenwood, Montana State University, Department of Mathematical Sciences, PO BOX 172400, Bozeman, MT 59717-2400, *greenwood@math.montana.edu*

***Key Words:*** model selection, cross validation, linear regression

Full cross-validation (FCV) based estimates of the Mean Squared Error of Prediction (MSEP) have been recommended as potential model selection criteria (Bunke, Droge and Polzehl, 1999). They have shown FCV can be used to find better estimates of MSEP than typical cross-validation (CV) and suggest its use in model selection criteria. Neath, Davies and Cavanaugh (2004) suggest a model selection criterion based on cross-validation called the Predictive Discrepancy Criterion (PDC) which is an estimate of the Kullback-Leibler discrepancy. It provides improved performance over typical CV-based criteria and the AIC. The development of a FCV-based analogue of the PDC is described. All the different criteria are then compared using simulations. Extensions to nonlinear regression model selection are also discussed.

### On the Nonnegative Garrote Estimator

❖ Ming Yuan, Georgia Institute of Technology, 427 Groseclose Building, 765 Ferst Drive, Atlanta, GA 30332, *myuan@isye.gatech.edu*; Yi Lin, University of Wisconsin-Madison

***Key Words:*** nonnegative garrote, path consistency, piecewise linear solution path, LASSO

We study the nonnegative garrote estimator from three different aspects: computation, consistency and flexibility. We show that the nonnegative garrote estimate has a piecewise linear solution path. Using this fact, we propose an efficient algorithm for computing the whole solution path for the nonnegative garrote estimate. We also show that the nonnegative garrote has the nice property that with probability tending to one, the solution path contains an estimate that correctly identifies the set of important variables and is consistent for the coefficients of the important variables. We propose a slight modification that retains the attractive properties of the original nonnegative garrote, but is more widely applicable. To demonstrate the flexibility of the proposed estimator, we consider an extension to the nonparametric regression setup.

### Latent Transition Analysis: Inference and Estimation

❖ Hwan Chung, Michigan State University, B601 W. Fee Hall, East Lansing, MI 48824, *hchung@epi.msu.edu*

***Key Words:*** latent transition analysis, maximum-likelihood method, Bayesian method

Latent transition analysis (LTA) is designed to estimate the probability of individuals' membership in hypothesized latent classes and their transition rate among these classes. LTA derives from the family of latent class analysis in which manifest items are treated as fallible indicators of unseen states. Routines for maximum-likelihood (ML) estimation are currently available in statistical software. In many examples, however, the likelihood function exhibits unusual features, causing ML estimates to behave erratically. In this talk, we explore a variety of theoretical and practical issues surrounding the use of the LTA model, including Bayesian alternatives to ML estimation. We illustrate difficulties in ML and Bayesian methods with an example from substance use behaviors in adolescent females.

## 70 Bayesian Spatial and Spatio-Temporal Models ● ✪

Section on Bayesian Statistical Science

**Sunday, August 6, 4:00 pm–5:50 pm**

### Bayesian Change Point Analysis for Local Linear Regression: a New Approach to Prior Selection

❖ Rajib Paul, The Ohio State University, 1958 Neil Ave., Department of Statistics, Columbus, OH 43210, *rajib@stat.ohio-state.edu*; L. Mark Berliner, The Ohio State University

***Key Words:*** glacial dynamics, Markov chain Monte Carlo

In the Bayesian approach, we treat both the number and locations of change points as unknown parameters. We pay special attention to the development of reasonable priors that can reflect a variety of prior information. Posterior results are obtained via Markov Chain Monte Carlo. We apply our change point analysis in modeling ice flow velocities for ice sheets in the Lambert Glacial Basin of Antarctica based on surface and basal topographies of these sheets. We rely on some simple physics-based approximations for velocity. These simplifications lead to acceptable local models. however, the need for multiple change points is very prominent in our case. Selection of priors is always an issue for Bayesian statisticians. In this paper we emphasize how we can select priors efficiently in different set-ups. Various interesting aspects and extensions are discussed.

### A Bayesian Dynamic Spatio-Temporal Interaction Model

❖ Jacob Oleson, The University of Iowa, 200 Hawkins Drive, C22GH, Department of Biostatistics, Iowa City, IA 52242-1009, *jacob-oleson@uiowa.edu*; Hoon Kim, California State Polytechnic University, Pomona

***Key Words:*** autoregressive prior, disease mapping, hierarchical Bayes, lob-linear mixed model

During the past three decades, prostate cancer incidence has changed substantially in the United States. A fully Bayesian hierarchical spatio-temporal interaction model is proposed to estimate prostate incidence rates in Iowa. We introduce random spatial effects to capture the local dependence among regions, random temporal effects to explain the nonlinearity of rates over time, and random spatio-temporal interactions. In addition, we introduce fixed-age effects, as most epidemiologic data are related strongly to age. We found prostate cancer incidence in Iowa counties increased sharply over age while incidence rates increased initially, then decreased over time. We identify hotspots of high and low rates for age groups and time periods using disease mapping.

### Multiresolution Hierarchical Dynamical Models for Spatio-Temporal Processes

❖ Ali Arab, University of Missouri-Columbia, Department of Statistics, 146 Middlebush Hall, University of Missouri Columbia, Columbia, MO 65211, *aa5vf@mizzou.edu*; Christopher K. Wikle, University of Missouri-Columbia

***Key Words:*** hierarchical models, dynamical models, Bayesian analysis, environmental sciences, spatio-temporal models

Spatio-temporal processes are ubiquitous in the environmental and physical sciences. The complexity of these processes exhibited through different scales of spatial and temporal variability necessitates the implementation of hierarchical models with computationally efficient and sparse structures applicable to very large datasets. We develop hierarchical Bayesian models incorporating efficient parameterizations to model such dynamical processes. Specifically, our multiresolution approach represents high-dimensional dynamics via mappings to low-dimensional hidden processes. This representation allows sub-processes at different scales to interact, yet keeps the number of effective parameters to a minimum.

### Spatial Bayesian Modeling of fMRI Data: a Multiple-Subject Analysis

❖ Lei Xu, University of Michigan, Ann Arbor, MI , *leix@umich.edu*; Timothy D. Johnson, University of Michigan; Thomas Nichols, University of Michigan

***Key Words:*** function brain mapping, Bayesian hierarchical model, reversible jump MCMC, mixture model, multiple-subject fMRI data analysis

The identification of a specific brain region with a specific function is a central problem in fMRI data analysis. Conventionally, a classical, mass-univariate approach is taken, and the resulting statistic image is searched for significance. Our work develops a hierarchical Bayesian model that improves the standard methods in several ways. First, we pose an explicit spatial model for activations. Second, while some authors have proposed spatial models, they have only considered single subject data; our method models multi-subject data, accounting for intersubject heterogeneity in activation location about a population location. Lastly, we use a fully Bayesian framework, so all sources of uncertainty can be considered and quantified. We estimate the quantities of interest using reversible jump MCMC methods and demonstrate the method on simulated and real data.

### Bayesian Hierarchical Spatially Correlated Functional Data Analysis with Application to Colon Carcinogenesis

❖ Veera Baladandayuthapani, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *veera@mdanderson.org*; Raymond J. Carroll, Texas A&M University; Bani K. Mallick, Texas A&M University; Mee Young Hong, Texas A&M University

***Key Words:*** functional data analysis, mixed models, spatial correlation, semiparametric methods, hierarchical models, carcinogenesis

We present new methods to analyze data from an experiment using rodent models to investigate the biological mechanisms surrounding p27, an important biomarker predictive of early colon carcinogenesis. The responses modeled are essentially functions nested within a two-stage hierarchy. Moreover, in our experiment, there is substantial biological motivation for the existence of spatial correlation among the functions, which arise from the locations of biological structures called colonic crypts: this possible functional correlation is a phenomenon we term crypt signaling. Thus, as a point of general methodology, we require an analysis that allows for functions to be correlated at the deepest level of the hierarchy. Analysis of this data set gives new insights into the structure of p27 expression in early colon carcinogenesis and suggests the existence of significant crypt signaling.

## 71 Designs for Clinical Trials and Other Studies

IMS, Biometrics Section, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### Identifiably of Placebo Responders via Potential Outcomes

❖ Eva Petkova, Columbia University, , *ep120@columbia.edu*; Thaddeus Tarpey, Wright State University; Yimeng Lu, Columbia University; Donald Klein, Columbia University

***Key Words:*** antidepressant treatment, biomarker, latent class model, randomized experiments

An important problem in the clinical practice and research on mental health and control of pain is identifying and differentiating placebo from a true drug effect. Two types for medical studies can be distinguished for their utility for addressing this problem. Acute treatment studies randomize ill subjects to either drug or placebo. Treatment discontinuation studies randomize subjects who improve after acute treatment to ether continuing treatment with drug, or switching to placebo. Using the potential outcomes framework we discuss what effects are identifiable with information from these two study types. Conditions ensuring identifiability are explicitly stated. Data from an acute treatment and a treatment discontinuation trial for depression are used to motivate and illustrate the discussion.

### Efficient Adaptive Designs for Clinical Trials

❖ Jay Bartroff, Stanford University, Department of Statistics, Sequoia Hall, Stanford, CA 94305, *bartroff@stat.stanford.edu*; Tze Leung Lai, Stanford University

***Key Words:*** adaptive design, conditional power, futility, Hoeffding's information bound, Kullback-Leibler information

Whereas previous works on adaptive design of clinical trials and mid-course sample size re-estimation have focused on conditional power to determine the second-stage sample size, we consider here a new approach that involves choosing the second-stage sample size by mimicking efficient fully-sequential tests using a flexible number of stages. Not only does this approach maintain the prescribed type I error, but it also provides an asymptotically efficient multistage test whose finite-sample performance, measured in terms of the expected sample size and power, is shown to be superior to existing conditional power and adaptive and non-adaptive group sequential designs.

### Remodified Continual Reassessment Method and the PBTC Experience

❖ Arzu Onar, St. Jude Children's Research Hospital, Memphis, TN 38105-2794, *arzu.onar@stjude.org*; Mehmet Kocak, St. Jude Children's Research Hospital; James Boyett, St. Jude Children's Research Hospital

***Key Words:*** phase 1 trials, biostatistics, oncology, simulation

The continuous reassessment method (CRM) has served as an appealing alternative to the traditional phase I design due to its model-based approach. It has, however, received harsh criticism as a result of the higher levels of toxicities associated with the original version, which led to various modifications and its subsequent successful application in adult studies. Pediatric clinical trials pose their unique challenges to the application of CRM due mainly to relatively high starting doses associated with pediatric studies. In this talk, we will outline our modifications to the approach and present various simulation results motivated by our experience gained through the Phase I trials within the Pediatric Brain Tumor Consortium.

### D-Optimal Designs for Compartmental Models

❖ Gang Li, GlaxoSmithKline, 1250 S. Collegeville Road, Collegeville, PA 19426, *gangli_stat@yahoo.com*; Dibyen Majumdar, University of Illinois at Chicago

***Key Words:*** compartmental models, Tchebyschev system, equivalence theorem

In pharmacokinetics the plasma concentration versus time profiles of a drug are often characterized by compartmental models. In this paper, we consider the optimal designs for three compartmental models. By equivalence theorem and Tchebyschev-system we show that the locally D-optimal designs for these models have the same number of support points as the number of their model parameters. The efficiencies of some practically implemented designs were evaluated. We also investigate the effect of mis-specification of initial parameter estimates on the resulted designs.

### Robust Designs for Binomial Data

❖ Adeniyi Adewale, University of Alberta, Mathematical and Statistical Sciences, Edmonton, AB T6G2G1 Canada, *aadewale@ualberta.ca*; Douglas P. Wiens, University of Alberta

***Key Words:*** robust regression design, generalized linear models, simulated annealing, model misspecification, Fisher information

We have developed criteria that generate robust designs which insure against possible misspecifications in generalized linear models (GLM) for binomial data. Modeling via GLM requires the specification of the random, systematic and link components of the model. All these model components are subject to misspecification and our design criteria accommodate these possible misspecifications. We addressed the problems of misspecification of the systematic and link components by some form of averaging of the loss - a function of sampling and bias errors - over the misspecification neighborhood. The problem of extrabinomial variation is addressed by an approach akin to quasi-likelihood estimation but adapted for design purposes. Examples for various misspecification scenarios, including some real dataset examples are presented.

### Exact D-Optimal Designs for Second-Order Response Surface Model on a Sphere and with Qualitative Factors

❖ Chuan-Pin Lee, National Sun Yat-sen University, Department of Applied Mathematics, 70 Lienhai Rd, Kaohsiung, 80424 Taiwan, *leecb@mail.math.nsysu.edu.tw*; Mong-Na Lo Huang, National Sun Yat-sen University; Ray-Bing Chen, National University of Kaohsiung

***Key Words:*** arithmetic-geometric inequality for matrices, dispersion function, optimal block designs

The exact designs for response surface model have widespread use in industry, but in many situations, it is difficult to obtain the close form of exact D-optimal designs. Here, we are interested in finding exact D-optimal designs for second-order response surface model with 2 quantitative factors on a sphere where the approximate D-optimal design provided by Kiefer (1960) is utilized. We focus on the class of uniform designs supported on the vertices of certain regular polygons and show that, for any sample size N, there is a convex combination of designs belonging to the above class with 5 to 9 vertices to be the exact D-optimal. Qualitative factors or block effects are also important factors, and the exact D-optimal designs obtained for models with only quantitative factors can be applied to models with not only quantitative but also qualitative factors or block effects.

### D-Optimal Designs for Combined Polynomial and Trigonometric Regression on a Partial Circle

❖ Fu-Chuen Chang, National Sun Yat-sen University, Department of Applied Math, 70 Lienhai Road, Kaohsiung, 804 Taiwan, *changfc@math.nsysu.edu.tw*

***Key Words:*** d-optimal designs, Taylor expansion, polynomial regression, trigonometric regression, implicit function theorem, recursive algorithm

Consider the D-optimal designs for a combined polynomial of degree d and trigonometric of order m regression on a partial circle [see Graybill (1976), p. 324]. It is shown that the structure of the optimal design depends only on the length of the design interval and that the support points are analytic functions of this parameter. Moreover, the Taylor expansion of the optimal support points can be determined efficiently by a recursive procedure.

## 72 Clinical Trial Design and Analysis ●

Biopharmaceutical Section, Biometrics Section, WNAR, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### Note on Randomization-Based Inferences for Randomized Clinical Trials

❖ Guohua Pan, Johnson & Johnson Pharmaceutical R&D, 1125 Trenton-Harbourton Road, Titusville, NJ 08560, *jpa3@prdus.jnj.com*; Yibin Wang, Novartis Pharmaceuticals Corporation

In many randomized clinical trials, study subjects represent a non-random convenience sample of available subjects from a perceived target population. However, random sampling based statistical methods are frequently used in analyzing these randomized available case studies. For this type of studies, randomization tests have been proposed in the literature as replacements of or alternatives to sampling based methods. Compared with sampling based methods, randomization based methods differ in several aspects. This paper further discusses and contrasts sampling and randomization based inferences for analyzing randomized clinical trials.

### Four Types of Sums of Squares and Estimates of Treatment Differences in Multicenter Clinical Trials

❖ Daozhi Zhang, DOV Pharmaceutical, Inc., 433 Hackensack Ave., Hackensack, NJ 07601, *daozhizhang@yahoo.com*

***Key Words:*** multicenter, clinical, trial, estimate, effect

How to analyze the data from multicenter clinical trials has been a subject of debate. The center effect is at the center of the debate. This is mainly because different approaches of handling center effect yield different statistical conclusions when patients are not equally distributed across centers. The simulation results in this presentation provide useful knowledge on how to choose estimates that result from the different approaches.

### Optimal Allocation of Units When Comparing k Treatments to Two Controls of Unequal Importance

❖ Nairanjana Dasgupta, Washington State University, Department of Statistics, Washington State University, Pullman, WA 99164, *dasgupta@wsu.edu*

***Key Words:*** MV-optimality, a-optimality, negative control, positive control

Experiments in pharmaceutical and health sciences often involve comparisons of test treatments to more than one control. However, the controls may not always be of equal importance. The question of allocating resources to the treatments and the controls depends upon the relative importance of the controls. Here we discuss this allocation from a weighted MV optimality and A optimality framework. We provide approximate as well as analytical results. One-way-layout and Block Design lay-out are studied.

### A Method for Testing a Prespecified Subgroup in Clinical Trials

❖ Yang Song, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., Raritan, NJ 08869, *ysong9@prdus.jnj.com*; George Chi, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** subgroup analysis, type I error rate, closed testing procedure, alpha allocation, adjusted p-values

In clinical trials, investigators often are interested in the effect of a given study treatment on a subgroup of patients with certain clinical or biological attributes in addition to its effect on the overall population. Such a subgroup analysis would become more important if an efficacy claim could be made for the subgroup when the test for the overall study population fails at a prespecified alpha level. In practice, such a claim is often dependent on prespecification of the subgroup and certain implicit or explicit requirements placed on the study results. By carefully considering these requirements, we propose a general statistical methodology for testing both the overall and subgroup hypotheses, which has optimal power and strongly controls the family-wise type I error rate.

### Some Issues in Fitting Clinical Count Data with Poisson Regression Model

❖ Abdul Sankoh, sanofi-aventis, 200 Crossing Blvd., Bridgewater, NJ 08807, *abdul.j.sankoh@aventis.com*

***Key Words:*** clinical count data, goodness of fit, Poisson regression, type I error rate

Poisson regression is routinely used to model response rates or clinical count data as a function of covariate levels. However, regulatory agencies are generally concerned about the goodness of fit of Poisson distribution to clinical data. Even when there is a good fit, regulators may still be concerned about the possibility of false-positive findings resulting from under-estimation of standard error. We examine in this presentation the appropriateness of the regular Poisson regression model for the analysis of clinical count data from randomized clinical trials by comparing its performance to other models including the generalized Poisson and negative binomial models regarding goodness of fit, type I error rate control, and power. Simulation results show comparable performance for the regular Poisson model regarding type I error rate control under reasonable sample size.

### Interval Estimation of Risk Ratio in the Simple Compliance Randomized Trial

❖ Kung-Jong Lui, San Diego State University, Department of Mathematics and Statistics, San Diego, CA 92182-7720, *kjl@rohan.sdsu.edu*

***Key Words:*** efficiency, simple compliance trial, efficacy, risk ratio, coverage probability

Consider the simple compliance randomized trial (SCRT) in which patients assigned to an experimental group may switch to receive a control treatment, but patients assigned to a control group are assumed to receive their assigned treatment. Five asymptotic interval estimators for the relative risk (RR) of probabilities of response among patients who would comply with the experimental treatment under the SCRT are developed. Monte Carlo simulations are employed to evaluate the performance of these interval estimators in a variety of situations. A systematic discussion on comparisons and findings is included.

## 73 Phase II Trials ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Sunday, August 6, 4:00 pm–5:50 pm**

### A Parallel Phase I/II Clinical Trial Design for Combination Therapies

❖ Xuelin Huang, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *xlhuang@mdanderson.org*; Swati Biswas, University of North Texas Health Science Center; Yasuhiro Oki, M. D. Anderson Cancer Center; Jean-Pierre Issa, M. D. Anderson Cancer Center; Donald Berry, The University of Texas

***Key Words:*** Bayesian, dose selection, logistic regression, synergy

In place of the traditional, separate phase I and II trials, we propose using a parallel phase I/II clinical trial design to evaluate simultaneously the safety and efficacy of combination therapies. The proposed design uses all data accumulated from the beginning of the trial to evaluate the toxicity and efficacy parameters and a new method to determine adaptive assignment probabilities. After an initial period of dose escalation, patients are assigned randomly to admissible dose levels. Early stopping rules are specified. Simulations show the proposed design saves sample size, has better power, and efficiently assigns more patients to doses with higher efficacy levels. For illustration, we apply the design to a combination chemotherapy trial for leukemia.

### Three-Outcome Design for Randomized Comparative Phase II Clinical Trials

❖ Shengyan Hong, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *hongsh@lilly.com*; Yanping Wang, Eli Lilly and Company

***Key Words:*** randomized phase II trial, three-outcome design, statistical comparison, two-stage design

Randomized design has been increasingly called for in oncology phase II trials to avoid patient selection bias in single arm design. However, formal statistical comparison is rarely conducted due to sample size restriction, despite its appealing feature. We offer an approach to sample size reduction by extending the three-outcome design of Sargent et al (2001) for single arm trials to randomized comparative trials. The design allows a third outcome: reject neither null nor alternative hypotheses when testing result is in "grey area". It could reduce sample size up to 30% over standard two-outcome design, making randomized comparative design an affordable option for Phase II clinical trials. The statistical formulation of three-outcome comparative design will be described and sample sizes for single-stage and Simon two-stage design will be tabulated.

### Optimal Trial Designs for Screening Cancer Therapeutic Agents

❖ Vandana Mukhi, New York University School of Medicine, 60 Sunflower Ridge Road, South Setauket, NY 11720, *vandana.mukhi@med.nyu.edu*; Yongzhao Shao, New York University; Judith D. Goldberg, New York University School of Medicine

***Key Words:*** phase II, false discovery rate, clinical trial, design

With the large number of new cancer therapeutic agents under development, investigators are seeking to design and implement clinical trials to improve outcomes for cancer patients. Improved methods for the prioritization and selection of new treatments for phase III evaluation can expedite the delivery of new agents to patients. We can accomplish this goal for improvements in study designs and procedures to screen out inefficacious treatments and select those treatments that are truly effective in a cost-effective manner. We focus on the evaluation and comparison of several optimal phase II screening trial designs based on the False Discovery Rate (FDR) and other criteria. Simulation results will be discussed.

### Optimal Adaptive Designs in Phase II Trials

❖ Anindita Banerjee, North Carolina State University, 2311 Champion Court, Raleigh, NC 27606, *abanerj2@ncsu.edu*; Anastasios A. Tsiatis, North Carolina State University

***Key Words:*** two-stage adaptive design, backward induction, Bayesian decision theory, simulated annealing

Phase-II trials provide a platform where, on the basis of the efficacy of the drugs, ineffective drugs are screened out and promising drugs move to the next phase. Simon (1989) proposed optimal fixed two-stage designs which minimize the expected sample size under the null hypothesis. We have derived optimal adaptive designs at the null that perform better than Simon's design, although the gains are modest (Banerjee and Tsiatis 2006). We further explore optimal adaptive designs that minimize the expected sample size at the alternative hypothesis, at a probability midpoint between the null and alternative hypothesis and a weighted combination of the response probabilities. We also construct an envelope function that gives the lowest expected sample size for any possible value of the response probability. The different designs are compared to each other as well as the envelope function.

### Optimal Two-Stage Designs for Phase II Clinical Trials for Continuous Endpoints

❖ Chinfu Hsiao, National Health Research Institutes, 35 Keyan Road Zhunan Town, Miaoli County, 350 Taiwan, *chinfu@nhri.org.tw*; Hsiao-Hui Tsou, National Health Research Institutes; Jen-pei Liu, National Taiwan University; Shein-Chung Chow, Duke University

***Key Words:*** phase II trial, two-stage design, continuous endpoint

Pharmaceutical development is a long and risky investment, sometimes with profitable return. More than half of development duration is spent in clinical trials. During the early phase II stage, there is an urgent need for efficient and cost-effective designs to screen these potential drug candidates using the idea of the proof of concept for efficacy in a rapid and reliable manner to minimize the total sample size and hence shorten the duration of the trials. In this talk, an optimal two-stage design for continuous efficacy endpoints is proposed. The proposed two-stage possesses the property, which minimizes the expected sample size given type I and type II error rates and specification of undesirable and targeted mean values.

### Critical Statistical Issues in the Design and Analysis of Proof-of-Concept Clinical Trials in Multiple Sclerosis

❖ Chris Assaid, Merck & Co., Inc., P.O. Box 1000, UG1C-46, North Wales, PA 19454-1099, *christopher_assaid@merck.com*

***Key Words:*** multiple sclerosis, futility, inferiority, imaging endpoints

In spite of the proliferation of experimental treatments for MS, an appropriate measure of effect in short-term proof-of-concept (POC) trials as a surrogate for long-term clinical efficacy endpoints remains elusive at best. The typical surrogate in MS POC trials has been based on imaging data and is generally highly variable and skewed. The recent withdrawal of a highly effective treatment also enhances the scrutiny of safety in these trials. Design and analysis approaches for shorter-duration trials that seek to maximize power and the objectivity of the primary endpoint and implement a comprehensive group-sequential approach for monitoring safety and efficacy over the course of the trial that is appropriate for the POC framework will be presented. Discussion will include appropriateness of futility versus inferiority monitoring in this setting.

## 74 Methods for Incomplete Data ●

Biometrics Section

**Sunday, August 6, 4:00 pm–5:50 pm**

### Semiparametric Models and Sensitivity Analysis of Longitudinal Data with Nonrandom Dropouts

❖ David Todem, Michigan State University, Department of Epidemiology, B601 W. Fee Hall, East lansing, MI 48823, *todem@msu.edu*; KyungMann Kim, University of Wisconsin-Madison; Jason P. Fine, University of Wisconsin-Madison

***Key Words:*** exponential family distribution, functional estimators, global tests and extreme statistics, incomplete longitudinal data, nonparametric mixture, uniform convergence

We propose a family of semi-parametric non-response models to adjust for informative dropouts in the analysis of longitudinal data. The approach conceptually focuses on generalized linear mixed effects models with an unspecified random effects distribution. A novel formulation of a shared latent class model is presented and shown to provide parameters that have a meaningful interpretation. We show how to use the non-identifiability of some model characteristics to construct new global tests of covariate effects over the whole support of the sensitivity parameter. Simulations demonstrate a large reduction of bias for the nonparametric model relative to the parametric model at times where the dropout rate is high or the dropout model is misspecified. The methodology's practical utility is illustrated in a psychiatric data analysis.

## Methods on Longitudinal Data with Drop-Outs and Mismeasured Covariates

❖ Grace Y. Yi, University of Waterloo, 200 University Ave., W., Waterloo, ON M3G 1P3 Canada, *yyi@uwaterloo.ca*

***Key Words:*** longitudinal data, drop-outs, measurement error

Longitudinal data commonly contain missing observations and error-contaminated covariates. There has been quite a deal of research dealing with missingness of longitudinal studies. For example, maximum likelihood, multiple imputation, and inverse probability weighted generalized estimation equations approaches have been extensively discussed to handle missing observations. Relatively little work has been available to deal with measurement error in covariates. In this talk, I will discuss marginal methods for analyzing longitudinal data when both missingness and error-prone covariates are present. Numerical studies will be conducted with the proposed methods.

## Semiparametric Analysis of Longitudinal Data with Potential Right Censoring

❖ Mengling Liu, New York University School of Medicine, 650 First Ave., ROOM 526, New York, NY 10016, *mengling.liu@med.nyu.edu*; Zhiliang Ying, Columbia University

***Key Words:*** counting process, latent variable, least square estimation, marginal model, normal transformation

A package of semiparametric models is proposed for longitudinal data under possibly irregular observation occasions and potentially informative censorship. The models are motivated by the idea of shared random effects in joint modeling of longitudinal responses and event times, and are valid under a variety of assumptions of censoring mechanism. Specifically, we assume a semiparametric normal transformation model for the informative censoring time and a semiparametric regression model for the longitudinal response variable conditional on the observed censoring information. Asymptotically unbiased estimating equations are constructed and yield least square type estimators for the finite dimensional regression parameters of the semiparametric regression model. The estimators are consistent and asymptotically normal with the variance matrix ready to be estimated by the plug-in rule.

## A Multiple Imputation Approach for Responders Analysis in Longitudinal Studies

❖ Liqiu Jiang, North Carolina State University, 1506 Ashley Downs Drive, Apex, NC 27502, *liqiu00@yahoo.com*; Kaifeng Lu, Merck & Co., Inc.; Anastasios A. Tsiatis, North Carolina State University

***Key Words:*** missing data, multiple imputation, repeated measures, logistic regression

Often a binary variable is generated by dichotomizing an underlying continuous measurement. Ordinarily, a logistic regression model is used to estimate the effects of covariates on the binary response. When the underlying continuous measurements are from a longitudinal study, the repeated measurements are often analyzed using a repeated measures model. This motivates us to use repeated measures model as an imputation approach in the presence of missing data on the responder status. We, then, apply the logistic regression model on the observed or otherwise imputed responder status. Large sample properties of the estimators are derived and simulation studies carried out to assess the performance of the estimators in situations where either the imputation model or the response model is misspecified. We show that the estimators are robust to misspecification.

## Estimation of Transition Probabilities in a Discrete-Time Markov Chain with Missing Observations

❖ Hung-Wen Yeh, The University of Texas School of Public Health, 1200 Herman Pressler, Houston, TX 77030, *hung-wen.yeh@uth.tmc.edu*; Wenyaw Chan, The University of Texas School of Public Health

***Key Words:*** discrete-time Markov chain, EM algorithm, missing observation

The discrete-time Markov chain is commonly used in describing health states for chronic diseases. Statistical inferences on comparing treatment effects or on finding determinants of disease progression usually require estimation of transition probabilities. When collecting data of health states for these studies, researchers often encounter the problem of patients' occasional unavailability. Craig and Sendi (2002) use EM algorithm to estimate the transition probabilities when one scheduled observation was possibly missing in between two observed outcomes. In this research, EM algorithm will be applied to handle the situation when various numbers of scheduled observations are not observed between two measurements. An empirical study will be performed to examine the accuracy of the procedure and to compare the results with other methods. A real data set will be used for demonstration.

## Nonparametric Comparison of Two Survival Functions with Dependent Censoring via Nonparametric Multiple Imputation

❖ Chiu-Hsieh Hsu, University of Arizona, 1515 N. Campbell, Room 2942, PO Box 245024, Tucson, AZ 85724-5024, *phsu@azcc.arizona.edu*; Jeremy M. G. Taylor, University of Michigan

***Key Words:*** dependent censoring, multiple imputation, logrank test

When the event time depends on censoring time, the conventional two-sample test method could produce an invalid test. We extend our previous work in survival estimation to propose a multiple imputation approach to using auxiliary variables to adjust for dependent censoring while comparing two survival functions. To conduct the imputation, we use two working PH models to define an imputing risk set. One is for the event times and the other for the censoring times. Based on the imputing risk set, a nonparametric multiple imputation method is used to impute a future event or censoring time for each censored observation. Simulation studies show that the sizes of the log-rank and Wilcoxon test constructed on the imputed datasets derived from the bootstrap samples are comparable to the nominal level in the presence of dependent censoring if either one of the two working models is correct.

## Multivariate One-Sided Hypotheses Testing with Complete or Incomplete Data

❖ Tao Wang, The University of British Columbia, Statistics Department, Box 592, 6335 Thunderbird Crescent, Vancouver, BC v6t 2g9 Canada, *twang@stat.ubc.ca*; Lang Wu, The University of British Columbia

***Key Words:*** multivariate one-sided hypotheses, bootstrap, missing data, likelihood ratio test, observed-data likelihood, EM

Multivariate one-sided hypotheses testing problems arise frequently in practice. Up to now some testing methods have been proposed. However, most of them require normal or large sample assumptions, which may not be reasonable. Also, the commonly-used likelihood-based tests are conservative because the null distributions often depend on nuisance parameters. Attempting to solve these problems, a bootstrap test method will be proposed. On the other hand, all of the existing tests only apply to complete data, but in practice there are often missing data. So a likelihood ratio test will be proposed for incomplete data.

## 75 Contributed Posters

Biometrics Section, Biopharmaceutical Section, General Methodology, Section on Nonparametric Statistics, Section on Physical and Engineering Sciences, Section on Survey Research Methods

**Sunday, August 6, 4:00 pm–5:50 pm**

### Examining the Effect of Biomarkers in Terms of Pathological Compartmentalization and a Continuous Variable

❖ Irene Helenowski, Northwestern University, 680 N. Lake Shore Drive, Suite 1102, Chicago, IL 60611, *i-helenowski@northwestern.edu*; Edward F. Vonesh, Baxter Healthcare Corporation; Ryan J. Deaton, University of Illinois at Chicago; Borko Jovanovic, Northwestern University; Alfred W. Rademaker, Northwestern University; Sally A. Freels, University of Illinois; Vijayalakshmi Ananthanarayanan, University of Illinois at Chicago; Peter H. Gann, University of Illinois at Chicago

***Key Words:*** biomarkers, pathological compartmentalization, continuous variables, regression

In studying biomarkers associated with carcinogenesis, we need to examine the spatial-temporal mechanisms involved. Do biomarker levels follow a gradient related to distance from the cancer region? What model and covariance structure best fit the data? This work stems from previous work concerning the effects of a continuous variable, namely distance, on biomarker reproducibility (Helenowski et al. 2005). Those studies focused on the effects of a continuous variable on reproducibility solely in the normal tissue compartment. We now examine the effects of a continuous variable, adjusting for pathological compartmentalization on associations between biomarker samples based on least square differences. Future studies will involve examining reproducibility in terms of these two factors. We apply our methods to MCM2 index values coming from the prostate gland.

### Minimum Sample Size in Control Group When Comparing Efficacy Rate with Several Treatment Groups

❖ Alan Davis, Pharmanet, 10150 Mallard Creek Road, Suite 500, Charlotte, NC 28262, *adavis@pharmanet.com*; Inder J. Sharma, Sharma Associates, Inc.

***Key Words:*** odds ratio, Dunnett's, clinical trial, power, alpha, simulation

In evaluating a treatment in a clinical trial, it is often necessary to compare the success rates in several treatment groups against a single group used as a placebo control. Sometimes it is not practical to enroll many patients in a control group, due to the ethical necessity of providing treatment. This can provide a significant loss to information that can be gathered, particularly in consideration of the desirability of evaluating several treatment groups against the control. This study will evaluate the cost in terms of power and alpha of reducing the number of patients in a control group relative to several treatment groups. The method of determining significance will use Dunnett's test extended to the evaluation of odds of success for 3-5 groups (k) where one group is a control or reference group and the outcome is a binomial used to report drug efficacy.

### Exploring the Relationship between Extended Oral Anticoagulant Therapy after a First Episode of Venous Thrombosis and Mortality Using Meta-analysis

❖ Brianna Miller, The University of Oklahoma, 500 W. Main Street, Apt 307, Oklahoma City, OK 73102, *Brianna-Miller@ouhsc.edu*

***Key Words:*** meta-analysis, venous thrombosis, anticoagulant therapy, mortality

The primary endpoint of studies examining the effect of extended oral anticoagulant therapy on recurrent venous thromboembolism after a first episode of venous thrombosis is recurrence, while the secondary endpoints are mortality and major bleeding. This study examines the relationship between extended oral anticoagulant therapy after a first episode of venous thrombosis and mortality using meta-analysis. Results from seven articles were examined for inclusion. Odds ratios and 95% CIs for the relationship between mortality and extended therapy were estimated. ORs were combined using the Mantel-Haenszel method. A dot plot displays ORs with 95% CIs for each study. A chi-square heterogeneity statistic for the ORs is shown. Combinability of studies based on independence, patient population, and study design and conclusions about the main relationship are discussed.

### Interval Estimation of Binomial Proportion in Clinical Trials with a Two-Stage Design

❖ Chen Chia Min, National Cheng Kung University, Tainan, 701, *cmchen@email.stat.ncku.edu.tw*

***Key Words:*** two-stage design, Simon's two-stage design, truncated binomial distribution, coverage probability

A two-stage design is employed in Phase II clinical trials to avoid giving patients an ineffective drug. If the number of patients with significant improvement, which is a binomial response, is greater than a pre-specified value at the first stage then another binomial response at the second stage is also observed. This paper considers interval estimation of the response probability when the second stage is allowed to continue. Then the Wald, score, Clopper and Pearson, and Sterne intervals are constructed based on the two binomial responses from this two-stage design. The mean actual coverage probability and expected interval width are employed to evaluate the performance of these interval estimators. Comparison results recommend the use of score interval since its mean coverage probability is more close to the nominal confidence level than the other interval estimates investigated.

### GLUMIP 2.0: Free SAS/IMLÆ Software for Planning Internal Pilots

❖ John Kairalla, The University of North Carolina at Chapel Hill, 123 Chaucer Court, Carrboro, NC 27510, *jkairall@bios.unc.edu*; Christopher S. Coffey, The University of Alabama at Birmingham; Keith E. Muller, The University of North Carolina at Chapel Hill

***Key Words:*** internal pilots, power, type I error, adaptive designs, sample size, re-estimation

We present the new version of our free SAS/IMLÆ software for planning internal pilots (IPs). IPs use interim power analysis (w/o data analysis) to modify final sample size. For IPs in the General Linear Univariate Model (GLUM) framework, an unadjusted test may lead to Type I error inflation. A bounding test controls Type I error while retaining most IP advantages. Our previous software was slow and unstable. GLUMIP 2.0 uses new exact forms for the test statistic making it much more fast and stable. The analytic forms in the new software solve many problems of current IP techniques for linear models. The software simplifies performing exact power analysis for IPs in the GLUM framework with Gaussian errors and is useful for a broad class of ANOVA and regression problems. Restrictions: Gaussian errors, fixed predictors, common design for all replications, no missing data.

## Stroke Clinical Trials and Response-Adaptive Randomization: an Ideal Match

❖ Yuko Palesch, Medical University of South Carolina, Department of Biostatistics, Bioinformatics, and Epidemiology, 135 Cannon Street Suite 303, Charleston, SC 29425, *paleschy@musc.edu*; Amy Bardeen, Medical University of South Carolina; Renee Martin, Medical University of South Carolina

***Key Words:*** response-adaptive randomization, acute stroke clinical trials

Generally, randomized clinical trials are conducted to benefit future patients, and response-adaptive randomization, such as the randomized-play-the-winner (RPW) method, would allow that as well as to help patients who participate in the trials. The RPW utilizes accumulating information and assigns the potentially better performing treatment to more subjects. Acute stroke trials prove ideal for this design due to their short follow-up time (24 hours) to obtain a reliable surrogate outcome measure. RPW was simulated with data from a placebo-controlled NINDS rt-PA Stroke Study in which rt-PA was shown to be effective. The RPW rule skewed the 50:50 treatment allocation to an average of 58:42 in favor of the rt-PA arm; hence 37% more subjects would receive the more successful treatment (rt-PA) under RPW. The statistical efficiency of RPW will be compared to the simple randomization method.

## Multivariate Applications in Systems Biology

❖ Amber Anderson, GlaxoSmithKline, 709 Swedeland Road, King of Prussia, PA 19406, *Amber.D.Anderson@gsk.com*; Zhu Lei, GlaxoSmithKline; Edit Kurali, GlaxoSmithKline; Amit Bhattacharyya, GlaxoSmithKIine; Kwan Lee, GlaxoSmithKline; Michael Durante, GlaxoSmithKline

***Key Words:*** systems-biology, multivariate analysis, PLSDA, multiple platforms, biomarkers

The systems-biology approach targets a comprehensive understanding of drug effect in the larger context of the whole organism or system. It can provide new understanding of drug and disease mechanisms to improve therapeutics to treat disease. Effects of different drugs were investigated in patients by assessing changes in various known and exploratory biomarkers from multiple platforms (e.g. blood chemistry, lipid, transcript, etc.). Partial Least Squares Discriminant Analysis (PLSDA) was performed on the integrated multiple platform data to extract systems-knowledge. The analysis assessed relative importance of the biomarkers across platforms; revealed cross platform marker relationship under the drug perturbation; and evaluated relative platform contribution. The advantage and strength of integrative analysis will be demonstrated through real data from the drug discovery process.

## Comparing the Performance of Three Asymptotic Methods in Estimating the Sample Size for a Therapeutic Equivalence Study Based on Difference of Proportions

❖ Xiaoning Li, The University of Oklahoma Health Sciences Center, 402 Wadsack Drive, Apt. A, Norman, OK 73072, *xiaoning-li@ouhsc.edu*; Sara K. Vesely, The University of Oklahoma Health Sciences Center

***Key Words:*** sample size, therapeutic equivalence study, asymptotic methods, difference of proportions

This study compared three asymptotic methods (Blackwelder, Dunnet and Gent, Farrington and Manning) with respect to the accuracy of sample size estimation for a therapeutic equivalence study with low dichotomous response rates (< 10%) and the impacts of misestimation of the population response rate on the empirical power (EP). EP was calculated based on the number of times the null hypothesis was rejected out of 1000 simulated sample data sets using three methods. The EP was compared with its nominal value for each method and scenario and among three methods. In most of the scenarios, the EPs were greater than their corresponding nominal powers. The effect of the misestimation on the EP decreases as delta, pis and pie increases. For pis or pie>=0.07 and delta>=3/2pis, at least one of the three tests still maintained EPs above 70% after considering all the possibilities of misestimation.

## Recent Development in Exact Inference for Parallel Group Design with Repeated Binary Measurements

❖ Dar-Shong Hwang, B.R.S.I., 11 Trinity Place, East Hanover, NJ 07936, *dshwang@optonline.net*; James Lee, Sankyo Pharma Development

***Key Words:*** exact conditional distribution, correlated binary variables, generalized constant odds ratio, generalized treatment effect, generalized interaction effect, parallel group design

Parallel groups with repeated binary measurements occur often in clinical trials. Hwang & Lee(2005) investigate an exact analysis of treatment*time interaction and treatment effect for 2 treatment groups with 2 or 3 repeated binary measurements. This paper further indicates the difference in the definition of interaction, treatment and time effects between repeated or independent binary measurements. The treatment effect assuming no treatment*time interaction is also clarified. Analogous to Zelen's test of interaction, and Birch-Cox-Gart's test of treatment effect assuming no interaction for independent binary data, parameters representing generalized treatment*time interaction, treatment or time effects are developed assuming no generalized interaction or constant odds ratio for correlated data. Existence and optimality of exact inference are demonstrated and actual distribution derived.

## Issues of Covariate Adjustments in Clinical Trials

❖ Moh-Jee Ng, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Blgd 22 Rm 5336, Silver Spring, MD 20993, *ngm@cder.fda.gov*; Tie-Hua Ng, U.S. Food and Drug Administration

***Key Words:*** covariate adjustments, clinical trials

Baseline covariate adjustments are often made in the analyses of randomized controlled trials. However, the results with adjustments for covariates that are measured not at baseline but after randomization could be misleading because such covariates may be affected by the

treatments. Therefore, such results must be viewed with caution. Examples will be presented for illustration purposes.

## Can We Recruit Additional Subjects for a Failed Study?

❖ Paul Hshieh, U.S. Food and Drug Administration, 1401 Rockville Pike, CBER, Suite 231S HFM219, Rockville, MD 20852-1448, *hshieh@cber.fda.gov*; Tie-Hua Ng, U.S. Food and Drug Administration

***Key Words:*** unplanned interim analysis, sample size, binary outcomes

When a clinical trial fails to meet the study objective, but the results almost reach a statistical significance, it is tempting to recruit additional subjects in the hope that enlarging the study will result in a statistical significance. Such an action is operationally similar to performing an unplanned interim analysis; in general, it would lead to an inflation of the type I error rate and therefore, would be unacceptable. This problem always exists when evaluating continuous endpoints because the nominal alpha level is attained by the statistical tests. However, for discrete outcomes, such as the proportion of successes, the nominal alpha level is frequently not attained by the statistical tests. In such cases, the “leftover” alpha may be “spent” by the second analysis with additional subjects and the prespecified type I error rate is maintained. Examples will be given to illustrate

## Multiple Imputation by Chained Equations: Predictive Mean Matching

❖ Gerald Kolm, Emory University, 7454 Woodruff Way, Stone Mountain, GA 30087, *paul.kolm@emory.edu*; Deborah Ehrenthal, Christiana Care Health System; Edward Ewen, Christiana Care Health System

***Key Words:*** missing data, multiple imputation, predictive mean matching, posterior distribution

Software for multiple imputation of missing data has become readily available. A number of packages have options for imputation methods and understanding the methods and their implications is crucial to obtaining reasonable imputed values. As an illustration, we compare differences in predictive mean matching with random draws from the predictive posterior distribution using multivariate imputation by chained equations (MICE) as implemented in the Stata routine, ice (Royston). The data base contains 3,594 records of current patients from an Adult Medicine Office, and includes demographic, clinical and laboratory variables. The amount of missing values range from 1% for blood pressure to over 35% for some lipid values. Results show that imputed values were not always consistent with observed values depending on imputation method and the distribution of the variables.

## Weighted Logrank-Type Tests Based on Doubly Truncated Data

❖ Su Pei Fang, National Cheng Kung University, Tainan, 701 Taiwan, *spf@stat.ncku.edu.tw*

***Key Words:*** doubly truncated data, survival function, self-consistent estimator

Recently, several algorithms, such as Turnbull (1976) and Efron and Petrosian (1999), have been derived from the likelihood function to estimate the survival function based on doubly truncated data. However, the nonparametric methods for comparing two survival functions have not been developed. Therefore, this paper proposes two test statistics to test the equality of two survival functions for doubly truncated data. One is based on the integrated weighted difference between two estimated survival functions, while the other is similar in form to the usual weighted logrank test. The comparative results from a simulation study are presented and the implementation of these methods to Taiwan’s building lifetimes data is presented.

## Kernel-Assisted EM Algorithm

❖ Suzanne Dubnicka, Kansas State University, Department of Statistics, 101 Dickens Hall, Manhattan, KS 66506, *dubnicka@stat.ksu.edu*

***Key Words:*** EM algorithm, kernel density estimate, Nadaraya-Watson estimator, missing data

The EM algorithm is a very popular tool for computing maximum likelihood estimates in the presence of incomplete data. As with any parametric method, one disadvantage of the EM algorithm is the need to specify a model for the data which includes the distribution of the errors/responses. We propose a kernel-assisted EM algorithm which uses a kernel density estimate of the likelihood function. Furthermore, the Nadaraya-Watson estimator is used the E-step of the algorithm. This methodology will be applied to the problem of estimating a treatment effect in a pretest-posttest study with the posttest measurements subject to missingness.

## The Analysis of Mixed-Effects Compartmental Systems Using Bayesian and non-Bayesian Methods

❖ Yi Wang, University of Nebraska-Lincoln, 623 S. 18th Street, Apt. 24, Lincoln, NE 68508, *wangyi9876@yahoo.com*; Kent M. Eskridge, University of Nebraska-Lincoln; Shunpu Zhang, University of Nebraska-Lincoln

***Key Words:*** compartmental analysis, mixed-effects modeling, ordinary differential equations, Bayesian and non-Bayesian methods, minimal model

Compartmental analysis is used to model dynamic biological systems and widely applied to the kinetics of drugs in the body. We use mixed-effects modeling, which quantifies between- and within-subject variability, and pharmacokinetic models with solutions from systems of ordinary differential equations (ODEs) to analyze population data. Non-Bayesian software (nlme and nlmeODE in R or NLINMIX in SAS) and Bayesian software (WinBUGS and WBDiff) enable the mixed-effect analysis of complicated systems of ODEs with and without a closed-form solution. The aim is to use several examples (particularly glucose-insulin minimal model) to illustrate the applicability of Bayesian and non-Bayesian methods for compartmental analysis of population data. Our results indicate that the two methods are numerically stable and provide accurate parameter estimates for standard population data used in this paper.

## An Application of Accelerated Lifetime Design/ Analysis for Estimating the Lifetime of CDs and DVDs

❖ James J. Filliben, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *james.filliben@nist.gov*; Adriana Hornikova, National Institute of Standards and Technology; Frederick R. Byers, National Institute of Standards and Technology

***Key Words:*** accelerated lifetime testing, simulation, uncertainty, Eyring model, design of experiment

The durability and lifetime of storage media such as CDs and DVDs was investigated utilizing an accelerated testing experiment. An optimally designed experiment was determined to minimize the bias and uncertainty of the predicted median lifetime at ambient conditions. This approach will apply for different CD and DVD manufactures and types. The ISO 18 927 and ANSI standards specify five different stress conditions in conjunction with the Eyring model, for predicting the ambient life expectancy of CDs and DVDs. For variance stabilization we used the logarithm of the failure times for the least squares regression fit to predict failure at ambient conditions. A simulation was carried out to determine the optimal accelerated testing design for this study, and to determine the lifetime estimates (and uncertainty) at ambient.

## Parametric Distance Estimators versus Maximum Likelihood Estimators in Estimating Quantiles with Misclassified Data

❖ Elliott Nebenzahl, California State University, East Bay, 1196 Stafford Drive, Cupertino, CA 95014, *elliott.nebenzahl@csueastbay.edu*; Dean Fearn, California State University, East Bay

***Key Words:*** quantile distance estimators, minimum distance estimators, survival distributions, quantile functions

We create parametric quantile estimators by minimizing distances chosen from a family of distances between a parametric quantile function and the sample quantile function. We compare these estimators to each other and the standard maximum likelihood (ML) estimators in terms of "closeness" of the fitted quantile function to the true quantile function. We are particularly interested in robustness, where the true population for the data does not agree with the parametric family we are considering. For example, we consider fitting a Weibull parametric model to generalized exponential data. For various data and different sample sizes and parametric models, we also compare the distance estimators and ML estimators in terms of unbiasedness and mean-squared error of the respective estimated pth ($0< p< 1$) quantiles.

## A Hyperbolastic Model for Survival Data

❖ Zoran Bursac, University of Arkansas for Medical Sciences, 4301 W. Markham Street, # 781, Little Rock, AR 72205, *zbursac@uams.edu*; Mohammad Tabatabai, Cameron University; David K. Williams, University of Arkansas for Medical Sciences; Karan P. Singh, University of North Texas Health Science Center

***Key Words:*** hyperbolastic survival model, Cox model, Weibull model, log-likelihood, survival probability prediction, log-logistic model

Modeling the outcomes that factor in the element of time has been extensively studied and applied in a wide range of medical and biological studies. In continuation of our work on a family of hyperbolastic growth models (Tabatabai et. al. 2005, Bursac et. al 2006) a new hyperbolastic survival model is introduced. The model is utilized for analysis of a published survival data set and the results are compared with those obtained from commonly used survival models (e.g. Cox, Weibull, log-logistic and log-normal). With the same number of parameters as in the classical models, the new model performs better in terms of values of log-likelihood and survival probability prediction. The new hyperbolastic survival model produces flexible hazards, accommodates a mix of covariates and may be a useful predictive tool in various research fields that employ time-to-event data.

## Survival Analysis on Recurrent Event Data: an Application to Alcoholism Study

❖ Jian Han, Bristol-Myers Squibb Company, 5 Research Parkway, Wallingford, CT 06492, *jian.han@bms.com*

***Key Words:*** multivariate survival, recurrent event

In alcoholism treatment clinical trials, one of conventional analyses of efficacy outcomes focuses on time to the first event ('any drinking' or 'heavy drinking'). The naturalistic course of drinking behavior during treatment often shows a gradual change in drinking before occurrence of a more stable drinking or abstinence pattern. Therefore, analyzing time to first event may not be sensitive enough to capture change in patient's drinking behavior if the treatment effect is observed in later relapse events. The multivariate survival time analysis or recurrent event analysis appears to be a better approach. Three commonly used strategies for multivariate survival time are available with emphasis of different aspects of recurrent events. These methods are explored and evaluated with respect to model assumption and validity in application in double blinded clinical trial data.

## Exact Test for an Epidemic Change in a Sequence of Exponentially Distributed Random Variables

❖ Ping Shing Chan, The Chinese University of Hong Kong, Department of Statistics, Hong Kong, 11111 China, *benchan@cuhk.edu.hk*; Kim Fung Lai, The Chinese University of Hong Kong

***Key Words:*** epidemic change, exact test, exponential variables, Wald test

Suppose we have a sequence of independent exponentially distributed observations. We would like to test the hypothesis that the sequence has the same mean versus the alternative hypothesis that there is an epidemic change in the sequence. Epidemic change refers to a change of mean after an unknown point, for an unknown duration in the sequence. We first review the asymptotic null distributions of the Wald test and likelihood ratio test. However these tests are good for large to moderate sample sizes only. In this thesis, we consider the Wald test of the epidemic change when the sample size is small. The null distribution of the statistic is a linear combination of Dirichlet random variables. A recursive formula has been derived to obtain the probabilities. The critical values are then tabulated. The powers of this exact test and its asymptotic counterpart are then compared.

## Locally Efficient Estimators for Semiparametric Models with Measurement Error

❖ Yanyuan Ma, Texas A&M University, College Station, TX , *ma@stat.tamu.edu*; Raymond J. Carroll, Texas A&M University

***Key Words:*** semiparametric, influence function, efficiency, measurement error, backfitting

We derive constructive locally efficient estimators in semiparametric measurement error models. The setting is one where the likelihood function depends on variables measured with and without error, where the variables measured without error can be modeled nonparametrically. The algorithm is based on backfitting. We show that if one adopts a parametric model for the latent variable measured with error---and if this model is correct---then the estimator is semiparametric efficient; if the latent variable model is misspecified, our methods lead to a consistent and asymptotically normal estimator. Our method further produces an estimator of the nonparametric function that achieves the standard bias and variance property.

## A Graphical Method for Testing the Equality of Regression Curves

❖ Kee-Hoon Kang, Hankuk University of Foreign Studies, Wangsan, Mohyun Myeon, Yongin, 449-791 South Korea, *khkang@hufs.ac.kr*; Cheolwoo Park, University of Georgia

**Key Words:** sizer, nonparametric regression, comparison of regression curves

One of the most important problems of statistical inference is the comparison of two or more groups, which occurs in a variety of contexts. Comparison of population curves in the nonparametric way includes densities, regression curves, survival functions and some other characteristic functions of the variable of interest. In this article we introduce a new graphical method for testing the equality of several independent regression curves. Our method is based on analyzing SiZer (SIgnificant ZERo crossing of the differences) maps, which are scale-space based visualization tools for statistical inference. We use SiZer for comparing residual distributions under the null hypothesis. The proposed method is tested under various simulation settings and real examples.

### A Permutation Test for Compound Symmetry

❖ Tracy Morris, Oklahoma State University, 3032 NW 65th Street, Oklahoma City, OK 73116, *tracy.morris@okstate.edu*; Mark Payton, Oklahoma State University

**Key Words:** covariance structure, permutation test, randomization test, compound symmetry

Many statistical procedures, such as repeated measures and factor analysis, require an assessment of the structure of the underlying covariance matrix. The classical parametric method of testing such a hypothesis involves the use of a likelihood ratio test. These tests have many limitations, including the need for very large sample sizes and the requirement of a random sample from a multivariate normal population. In situations in which some or none of these assumptions are met, permutation or randomization tests are appropriate. This research involves the development and analysis of a permutation/randomization test for the special case of compound symmetry. Samples of various sizes and number of measures on each subject were simulated from multiple distributions. In each case, the type I error rates and power were examined.

## 76 Contributed Poster Session with Opening Mixer: a Look at the Richness of Statistical Interests

ENAR, Section on Statistical Computing, Section on Statistics in Epidemiology, Biometrics Section, Biopharmaceutical Section, Business and Economics Statistics Section, General Methodology, Section on Government Statistics, Section on Physical and Engineering Sciences, Social Statistics Section, Section on Survey Research Methods, Section on Statistical Education, Section on Statistics and the Environment

**Sunday, August 6, 8:00 pm–9:50 pm**

### A Bayesian Approach to Semicontinuous Longitudinal Data

❖ Bing Han, The Pennsylvania State University, 333 Thomas Building, University Park, PA 16802, *buh113@psu.edu*; Wei Huang, Temple University

**Key Words:** semicontinuous data, panel data, Bayes hierarchical model, two-parts model

Semicontinuous data describes the mixture in response of a continuous component and a degenerate component. The classical models on this type of data consist of two-parts model (Olsen and Schafer 2001) and Heckman selection model (Heckman 1977). We adapted the classical two-parts model to a hierarchical Bayesian structure, by which longitudinal correlation is modeled by random effects. An MCMC procedure is developed and several optimizations for MCMC, including blocking parameter space and missing indicators, are discussed. We further incorporated a Bayes mixture as the top hierarchical level. The mixture identifies subjects in panel data into different classes corresponding to different responsive behavior. Finally, the methodology has been applied to the case study of R&D cost and patent in the telecommunication industry.

### Survival Instantaneous Log-Odds Ratio from Empirical Functions

❖ JungAh Jung, Novartis Pharmaceuticals Corporation, Oncology Biostatistics and Statistical Reporting, East Hanover, NJ 07936, *jungah.jung@novartis.com*; J. Wanzer Drane, University of South Carolina

**Key Words:** odds ratio, log-or, instantaneous LOR, survival ILOR, copula

The objective of this work is to introduce and derive the survival instantaneous log-odds ratio (SILOR) and its standard error and compare it to logistic regression. Hip fracture on white women, AGE and BMI from NHANES III, were used to calculate empirical survival functions (SF) for the adverse health outcome (AHO) and non-AHO. A stable copula was used to create a parametric bivariate SF that was fitted to the empirical bivariate SF. PROC NLIN in SAS was used to estimate the parameters of the copula models. Variations were observed that could not be anticipated and that were poorly, if at all, modeled by logistic regression. The bivariate SF had SILOR contours the nonconstancy of SILOR. The proposed method out-performed logistic regression. The present method is straightforward, and it captures all but random variability of the data.

### LTAS.NET: a NIOSH Life Table Analysis System for the Windows Environment

❖ Mary Schubauer-Berigan, National Institute for Occupational Safety and Health, 5555 Ridge Ave., Cincinnati, OH 45220, *zcg3@cdc.gov*; William R. Raudabaugh, Constella, Inc.; Avima Ruder, National Institute for Occupational Safety and Health; Misty Hein, National Institute for Occupational Safety and Health; Sharon R. Silver, National Institute for Occupational Safety and Health; Patricia Laber, National Institute for Occupational Safety and Health; Kathleen Waters, National Institute for Occupational Safety and Health; Jinghui Liu, Westat; Steven Spaeth, ; Kyle Steenland, Emory University

**Key Words:** epidemiology, occupational, methods, life table, cohort

Life table analysis is an essential tool of occupational epidemiology. Most life table analysis systems (LTAS) are limited by platform or analysis options. A new LTAS was created for Windows (LTAS.NET) to permit the analysis of more than one exposure variable, and to allow stratification by fixed and time-dependent covariates. LTAS.NET is written in Microsoft Visual Studio.NET using SQL Server. Options for statistical analyses include indirectly and directly standardized rate ratios, confidence intervals based on Poisson and exact methods, and trend tests for linear associations. LTAS.NET allows simultaneous stratification and analysis of multiple exposure variables. Time-dependent and fixed user-defined variables can be incorporated. The import, stratification and reporting options are highly flexible. Stratified event and person-time data are exportable for use in Poisson regression.

## Estimating Lifetime Prevalence Using Data from Disease Registries

❖ Limin X. Clegg, National Cancer Institute, Suite 504, MSC 8317, 6116 Executive Boulevard, Bethesda, MD 20892, *cleggl@mail.nih.gov*

Lifetime prevalence is the probability that an individual has ever been diagnosed with the disease at the given point in time t, whether or not the person is symptom-free from disease at t. For example, the estimated lifetime cancer prevalence gives the percent of cancer survivors in a population at time t although they may be cancer symptom-free. Prevalence may be estimated using data collected by disease registries. Two issues arise when estimating prevalence using registry data: (1) patients may be lost to follow-up, so that it is unknown if they are dead or alive at time t and (2) patients may be diagnosed before starting registration and they are not captured by the registry. This study presents an approach that takes into account these issues when estimate lifetime prevalence using registry data.

## Epidemiology of Herpes Zoster (Shingles)

❖ Peter Wollan, Olmsted Medical Center, 210 9th Street, SE, Research Department, Rochester, MN 55904, *pwollan@olmmed.org*; Patricia Saddier, Merck Research Laboratories; Lina Sy, Merck Research Laboratories; Barbara P. Yawn, Olmsted Medical Center

***Key Words:*** epidemiology, incidence, herpes zoster

All incident Herpes Zoster (shingles) cases in adults (age 22+) in Olmsted County, MN, from 1996 through 2001, were identified using the Rochester Epidemiology Project. Cases were confirmed using medical records. Results were analyzed using Poisson and logistic regression. HZ is common, averaging 3.5 cases per 1000 adults per year. The incidence rate increased over the 6 year period (p<.0001), up to 4.1 in 2000-2001. Women had a higher age-adjusted rate than men (p<.0001). The rate increased with age (p<.0001), from 1.2 per 1000 (age 22-29) to 10.6 (age 80-89), with an increase in the slope at ~ age 40. Long-lasting pain was the most common complication, and was more likely among older or immunocompromised patients and women. The distribution of the nature of non-pain complications, also more likely among older or immunocompromised patients but not in women, remained stable with age.

## What We Know about Unsuccessful and Successful High-Risk R&D Projects and What We Can Learn from Them

❖ Stephanie Shipp, National Institute of Standards and Technology, 100 Bureau Drive, MS 4710, Gaithersburg, MD 20899-4710, *sshipp@nist.gov*

***Key Words:*** advanced technology program, innovation, risk, social benefits, economic benefits

The Advanced Technology Program challenges industry to accelerate the development of technologies that, because they are risky, are unlikely to be developed at all or in time to compete in rapidly changing markets. The innovative technologies that ATP funds have the potential to generate significant commercial payoffs and widespread benefits to the U.S. economy, which is the ultimate goal of the program. ATP funds projects in all areas of science and technology. ATP tracks the progress of these funded projects during the life of the project funding (3 to 5 years) and for up six years after ATP funding ends, and even longer if the project is successful. Projects can fail for technical or commercial reasons. This poster will present the characteristics of successful and unsuccessful ATP-funded projects.

## Data Simulation Methodologies for Determining Sample Size Requirements To Test Gene-Drug Interactions in Genetically Pre-Screened Populations

❖ Kimberly Lowe, University of Arizona College of Public Health, 1940 S. Hermosa Drive, Tucson, AZ 85713, *klowe@email.arizona.edu*; James Ranger-Moore, University of Arizona College of Public Health; Patricia Thompson, Arizona Cancer Center

***Key Words:*** pharmacogenetics, gene-drug interactions, data simulation, sample size

Testing gene-drug interactions in randomized controlled trials is often unfeasible due to large sample size requirements and loss of power. We illustrate data simulation methodologies that account for variant allele frequencies under a range of effect sizes to identify sample sizes sufficiently powered to test specific gene-drug interactions. It is hypothesized that sample size requirements can be minimized if subjects are pre-screened for polymorphisms of interest and the resulting genetic information used as inclusion criteria to enrich study populations. A candidate gene-drug interaction between flavin monooxygenase 3 (FMO-3)and sulindac (trade name Clinoril) was used to model this hypothesis. There was greater than a 10% reduction in the required sample size to test the gene-drug interaction among simulated genetically-enriched study populations, assuming 80% power.

## Wavelet-Based Functional Mixed Model Data Analysis: Computational Considerations

❖ Richard Herrick, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, *rcherric@mdanderson.org*; Jeffrey S. Morris, M. D. Anderson Cancer Center

***Key Words:*** computation, wavelets, Bayesian, mixed-model, MCMC

Wavelet-based Functional Mixed Models is a new Bayesian method extending mixed models to irregular functional data (Morris and Carroll, JRSS-B, 2006). These data sets are typically very large and can quickly run into memory and time constraints unless these issues are carefully dealt with in the software. We reduce runtime by 1.) identifying and optimizing hotspots, 2.) using wavelet compression to do less computation with minimal impact on results, and 3.) dividing the code into multiple executables to be run in parallel using a grid computing resource. We discuss rules of thumb for estimating memory requirements and computation times in terms of model and data set parameters. We present examples and benchmarks demonstrating that it is practical to analyze very large data sets with readily available computing resources. This code is freely available on our website.

## Hierarchical Bayesian Calibration of Untested Devices

❖ Reid Landes, University of Arkansas for Medical Sciences, 4301 W. Markham Street, # 781, Little Rock, AR 72205, *rdlandes@uams.edu*

***Key Words:*** measurement error, MCMC, prediction

We consider the problem of calibration of mass produced measuring devices, including the issue of extending inferences from some tested devices to a similar one for which no calibration data is available. Statistical methods are well-established for cases where the reference instrument is exact and inferences are needed only for tested devices. However, when the reference instrument is subject to measurement error, and particularly when inferences for an untested device are desired, new methods are needed. We study the properties of a method for producing such predictions of quantities from an untested device based on

a Bayes hierarchical model applied in Landes (2005). We illustrate the method using a calibration experiment involving resistance temperature devices and an accurate, relatively precise thermometer. We evaluate the statistical properties of the method via simulation.

## Strip Transect Sampling To Estimate Object Abundance in Homogeneous and Nonhomogeneous Poisson Fields: a Simulation Study of the Effects of Transect Width and Number

❖ Timothy C. Coburn, Abilene Christian University, ACU Box 29315, Abilene, TX 79699, *tim.coburn@coba.acu.edu*; Sean A. McKenna, Sandia National Laboratories; Hirotaka Saito, University of California, Riverside; Orlando T. Garcia, Sandia National Laboratories

***Key Words:*** strip transect sampling, variance and bias estimation, spatial Poisson fields, unexploded ordnance (UXO)

Natural phenomena often occur as spatial point processes which can be modeled by the Poisson distribution. This paper investigates the use of strip transect sampling to estimate object intensity/abundance in such situations. Both homogeneous and non-homogeneous fields are considered. Because the number and size of transects have important operational and economic implications, the objective is to determine how many and what size transects should be selected in order to satisfy specified precision and bias constraints. The results suggest that percent of field coverage is more important than specific combinations of transect size and number, and that sampling requirements to produce high quality estimates are higher than anticipated. The sampling requirements are highest for non-homogeneous fields containing clusters or targets. A specific application to unexploded ordnance is discussed.

## The Impact of Computer Programming Languages on Statistics

❖ Morteza Marzjarani, Saginaw Valley State University, Computer Science Department, 7400 Bay Road, University Center, MI 48710, *marzjara@svsu.edu*

***Key Words:*** comp. stat, languages

The evolution of computer programming languages throughout the years has played a significant role in the way the data is analyzed using the statistical software packages. In this paper, we will introduce a short history of these languages, and their impact on statistics. Special consideration will be given on the impact of the newly introduced computer programming languages on statistics in general, and on the statistical software packages in particular. Statistical software packages are developed using one or perhaps several computer languages. Therefore, it would be of interest to know which of theses languages are more applicable to the field of statistics.

## Analyzing a Metabolomics Dataset

❖ Teresa Norris, 2620 Northchase Parkway, SE, A202, Wilmington, NC 28405, *teg8104@yahoo.com*

***Key Words:*** metabolomics, data sets, biomarkers

Metabolomics is the newest of the "-omics" sciences and shows great potential in identifying potential biomarkers for drug discovery. However, metabolomic data sets are complex and provide a number of challenges both in its analysis and interpretation. We will present some of the statistical challenges involved in analyzing these data sets as well as methods that can be used to overcome these issues.

## Inferring Quantitative Trait Loci Using a Bayesian Variable Selection Model and Markov Chain Monte Carlo Convergence Diagnostics

❖ Daniel Shriner, The University of Alabama at Birmingham, RPHB 327, 1530 3rd Ave S, Birmingham, AL 35294, *dshriner@ms.soph.uab.edu*; Nengjun Yi, The University of Alabama at Birmingham

***Key Words:*** convergence diagnostics, variable selection, quantitative trait loci

We previously developed a Bayesian composite space model for the genome-wide identification of quantitative trait loci. Since the number of quantitative trait loci is unknown, we use a variable selection method to address the unknown dimensionality of the model. In this work, we examine the behavior of our Markov chain Monte Carlo algorithm using standard convergence diagnostics. The degrees of cross-correlation and autocorrelation for the primary variables in the model were not very high. There was no significant burn-in period for 93% of the runs. The number of loci with inclusion probabilities reaching significance as assessed by Bayes factors was highly insensitive to the prior number of included loci, although effect sizes were easily overestimated. We suggest three possible criteria for determining the most parsimonious subset of variables for which the data support inclusion.

## Estimating p-Values in Small Microarray Experiments

❖ Hyuna Yang, The Jackson Lab, 600 Main Street, Box 303, Bar Harbor, ME 04609, *hyunayang@gmail.com*; Gary Churchill, The Jackson Lab

***Key Words:*** permutation method, p-value, microarray experiments

Microarray data typically have small numbers of observations per gene. This can result in low power for statistical tests and also presents challenges in assessing significance. Testing procedures that borrow information from all of the genes can improve power but these statistics have non-standard distributions and their significance must be assessed using permutation analysis. When sample sizes are small, the number of distinct permutations can be severely limited and pooling the permutation-derived test statistics across all genes has been proposed. However, this method is not appropriate because the null distribution of the test statistics under permutation is not the same. We propose a permutation based method for estimating p-values using a selected subset of data. This method is shown to have correct type I error rates and to provide accurate estimates of the FDR.

## Multivariate Simulation of Gene Expression Data

❖ Rudolph Parrish, University of Louisville, School of Public Health, 555 S Floyd St Suite 4039, Louisville, KY 40292, *rsparr01@louisville.edu*; Horace J. Spencer, University of Arkansas for Medical Sciences

***Key Words:*** microarrays, distribution modeling, simulation, multivariate, normalization

Simulation of microarray data in a realistic form is valuable for the purpose of evaluating the many different proposed techniques for preprocessing and analyzing such data and for designing experiments. For example, normalization procedures can be investigated to determine their relative statistical properties. Mathematical and statistical distribution modeling methods can be applied advantageously to help address this problem. A computational method is described in which variation among arrays can be characterized simultaneously for a large number of genes resulting in a multivariate model of gene expression which incorporates correlations. This model is used to simulate microarray expression data and probe intensity data and then to compare different normalization approaches.

### Treating Breaks in Time Series in the Current Employment Statistics State and Area Program

❖ James White, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212, *white_j@bls.gov*; Kenneth Shipp, Bureau of Labor Statistics

***Key Words:*** discontinuities, seasonal adjustment, government statistics, ARIMA, current employment statistics

Each month the U.S. Bureau of Labor Statistics publishes over 18,000 employment time series categorized by industry for states and metropolitan areas (MSAs). Through time, changes to industry and area definitions create discontinuities in these series. These discontinuities or "breaks" do not represent actual shifts in the level of employment; instead they reflect non-economic changes and diminish the integrity of the series. Finding a single method to rectify breaks in these CES time series presents multiple challenges: in addition to being numerous, the series are also quite diverse in terms of geography, industry, seasonality, and employment level. This paper describes potential remedies for series discontinuities caused by recent major administrative changes such as the North America Industrial Classification System (NAICS) and MSA redefinitions.

### On the Likelihood Ratio Test for the Numbers of Factors in Exploratory Factor Analysis

❖ Kentaro Hayashi, University of Hawaii at Manoa, Department of Psychology, 2430 Campus Road, Honolulu, HI 96822, *hayashik@hawaii.edu*; Peter M. Bentler, University of California, Los Angeles; Ke-Hai Yuan, University of Notre Dame

***Key Words:*** factor analysis, structural equation modeling, model selection, likelihood ratio test, Bartlett correction, AIC

In the exploratory factor analysis, when the number of factors exceeds the true number of factors, the likelihood-ratio test statistic no longer follows the chi-square distribution due to a problem of rank deficiency and of non-identifiability of model parameters. As a result, decisions regarding the number of factors may be incorrect. We support our argument by a simulation study.

### GEE Models for Longitudinal Analysis of Long-Term Occupational Radiation Exposures in Russian Nuclear Workers

❖ Adina Soaita, University of Pittsburgh, 130 De Soto Street, PUBLH A477, Pittsburgh, PA 15261, *ais8@pitt.edu*; Ada O. Youk, University of Pittsburgh; Richard Day, University of Pittsburgh; Tamara Azizova, Southern Ural Biophysics Institute; Niel Wald, University of Pittsburgh; Mike Kuniak, University of Pittsburgh; David M. Slaughter, University of Utah; Carol K. Redmond, University of Pittsburgh

***Key Words:*** GEE, longitudinal data, radiation exposure, Mayak Pa

This poster presents a longitudinal GEE analysis of the relationship between dosimetric (yearly gamma exposure) and hematological (lymphocyte count) data collected from a historical cohort (1948--1956) of highly exposed radiation workers at the first Russian nuclear plant. The analysis utilizes marginal models with repeated measurements and an unbalanced number of observations per subject. STATA programming tools are implemented to check correct marginal model specification with respect to the assumptions required by the GEE model (e.g., the necessity of variable transformation and the nature of missing data mechanisms). The results show a significant inverse relationship between the natural log of the lymphocyte count and the natural log of yearly gamma exposure, adjusting for baseline count, gender, work history, and lifestyle indicators (e.g., smoking and alcohol consumption).

### Antioxidant Use Predicts Transitions to Amnestic MCI and Dementia

❖ Marta Mendiondo, University of Kentucky, 207 Sanders Brown Center on Aging, 800 S. Limestone, Lexington, KY 40536-0023, *marta@email.uky.edu*; Richard J. Kryscio, University of Kentucky; Fred A. Schmitt, University of Kentucky

***Key Words:*** mild cognitive impairment, Alzheimer's disease, Markov chain, shared random effects, polytomous logistic regression, risk factors

The effect of antioxidants on transitions from normal to dementia is unknown. We examined the annual cognitive tests from the BRAINS cohort, a group of 785 initially cognitively normal subjects followed at the University of Kentucky Alzheimer's Disease Center. Subjects have an average of 8 assessments. Subjects were placed into one of three transient states at each assessment: cognitively normal, amnestic mild cognitive impairment (MCI), mixed MCI. Conversion to dementia or death are considered absorbing states. A multi-state Markov chain based on a shared random effects polytomous logistic regression model was used to model transitions among states. After controlling for age, gender, education, family history of dementia, and the presence of an APOE 4 allele, baseline antioxidant use predicts transitions from normal to amnestic MCI and from amnestic MCI to dementia.

### Structural Nested Mean Models for Assessing Time-Varying Effect Moderation: a Comparison of Two Estimation Methods

❖ Daniel Almirall, University of Michigan, 605 Hidden Valley Club Drive, Apt 205, Ann Arbor, MI 48104, *dalmiral@umich.edu*

***Key Words:*** effect modification, estimating equations, g-estimation, 2-stage estimation, time-varying treatment, time-varying covariate

We consider the problem of assessing causal effect moderation in longitudinal settings in which treatment (or exposure) is time-varying and so are the covariates said to moderate its effect. \emph{Intermediate Causal Effects} that describe time-varying causal effects of treatment conditional on past covariate history are introduced and considered as part of Robins' Structural Nested Mean Model. Two estimators of the intermediate causal effects, and their standard errors, are presented and discussed: The first is a proposed 2-Stage Regression Estimator, which can be used using standard regression software. The second is Robins' G-Estimator. The methodology is illustrated using longitudinal data from the randomized controlled trial PROSPECT. Simulation results that compare and shed light on various bias/variance properties of the two proposed estimators are presented.

### Methods for Assessing Changes in the FMRI Visual Field Map after Surgery

❖ Raymond G. Hoffmann, Medical College of Wisconsin, Division of Biostatistics, 8701 Watertown Plank Road, Milwaukee, 53226, *hoffmann@mcw.edu*; Paul Savarapian, Marquette University; Mary Jo Maciejewski, Medical College of Wisconsin; Edward A. DeYoe, Medical College of Wisconsin; Daniel Rowe, Medical College of Wisconsin

***Key Words:*** fMRI, neuroscience, spatial point process, image analysis

The visual field map is produced by mapping the active voxels of the visual cortex to a circular disk that corresponds to the retina. Activation of the voxels is determined by matching the fMRI time series to a time series based on a complex visual stimulus. Each scan (pre-surgical, post-surgical and late term recovery) produces a different set of points due to variability in location of the head between scans and due to noise in the fMRI response to the visual target. A spatial-temporal non-homogeneous Poisson process is used to test for changes in the density of the active voxels. Changes may occur both post-surgery, as well as during cortical re-organization during the healing process. Using simulations with known masks, we demonstrate the amount of change that can be identified and the robustness of the model.

## Design-Based versus Model-Based Methods: a Comparative Study Using Longitudinal Survey Data

❖ Sunita Ghosh, University of Saskatchewan, I.ARE.H., 306-105 Cumberland Ave., S., Saskatoon, SK S7N1L7 Canada, *sunita.ghosh@usask.ca*; Punam Pahwa, University of Saskatchewan; Geert Molenberghs, Limburgs Universitair Centrum

***Key Words:*** survey methods, GEE, NPHS, marginal model, multi-stage sampling

Survey data analysis using complex sampling designs ought to account for clustering, stratification and unequal probability of selection. Design-based and model-based methods are two commonly used routes taken to account for such survey designs. Several studies of cross-sectional survey designs have shown that these two approaches provide similar results when the model fits the data well. The present paper aims at comparing these two approaches using the National Population Health Survey (NPHS) dataset. The NPHS is an ongoing longitudinal study, and a stratified multi-stage sampling design was used. A marginal modeling approach proposed by Rao (1998) will be used by way of a design-based method. The Generalized Estimating Equation (GEE) method, proposed by Liang and Zeger (1986), will be used as a typical model-based approach. Results obtained from these methods will be compared.

## Hierarchical Generalized Linear Models for Data from Complex Sampling Designs

❖ Prabhu Bhagavatheeswaran, Southern Methodist University, 5937 Milton st, 124, Dallas, TX 75206, *pbhagava@smu.edu*; Ian Harris, Southern Methodist University

***Key Words:*** hierarchical generalized linear models, variance components, sampling weights, PQL, method of moments

Penalized quasi-likelihood (PQL), an approximate method of inference, is a simple estimation procedure for hierarchical generalized linear models. However, it has been noticed that PQL tends to underestimate variance components, especially when the response variable is binary. A modified estimation procedure based on method of moments, that is computationally less intensive than the original PQL method, is proposed here. The modified estimation procedure is illustrated on the simple one way random effects model. Using a simple adjustment, approximately unbiased estimators are obtained. The estimator can be adjusted to incorporate sampling weights, that arise for example in modeling complex sample survey data. This is illustrated using data from the National Assessment of Educational Progress (NAEP).

## Mediation Analysis with Multilevel Data

❖ Jungwha Lee, Institute for Health Research and Policy, 1747 W. Roosevelt Road, 555 CU9, Chicago, IL 60608, *jlee203@uic.edu*; Eisuke Segawa, Institute for Health Research and Policy; Sue Curry, University of Illinois at Chicago

***Key Words:*** mediation analysis, multilevel data

A basic mediation model has three variables, the response variable (Y), the covariate (X), and the mediator (M). How X effects Y is of our interest, by assuming that X causes M which in turn causes Y. In multilevel data, Y, X, and M can be measured in different levels of hierarchy. Krull and MacKinnon (2001) introduced multilevel modeling of mediated effects by first solving two multilevel regressions, and then combining the two results to get the direct and indirect effects. However, if the residuals of the two regressions are correlated, their approach may not be ideal. We present mediating mechanisms to incorporate the multilevel regressions. Helping Young Smokers Quit (HYSQ) study data as well as the simulated data are used to apply mediation analyses.

## The Application of the Kalman Filter to Nonstationary Time Series through Time Deformation

❖ Zhu Wang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., LE-400, Seattle, WA 98109-1024, *zhu.wang@gmail.com*; Henry L. Gray, Southern Methodist University; Wayne A. Woodward, Southern Methodist University

***Key Words:*** nonstationary, time deformation, continuous time, autoregressive, Kalman filter

An increasing valuable tool for modeling nonstationary time series data is time deformation. However, since the time transformation transforms the scale, equally spaced data become unequally spaced data. Interpolation is therefore often used to obtain regularly sampled data, which can be modeled by the classical ARMA modeling techniques. In this paper, the need for interpolation is eliminated by employing the continuous time autoregressive model. To estimate the parameters, the maximum likelihood function is decomposed by means of the Kalman filter. The resulting modeling improvements include, more accurate estimation of the spectrum, better forecasts, and the separation of the data into its smoothed time-varying components. The technique is applied to simulated and real data for illustrations.

## Is It Normal? A Simulation Study of Properties of Some Normality Tests

❖ Daniel M. Sultana, California State University, East Bay, 26970 Hayward Blvd., Apt. 102, Hayward, CA 94542, *dan_sultana@yahoo.com*; Charlyn J. Suarez, California State University, East Bay; Bruce E. Trumbo, California State University, East Bay; Eric A. Suess, California State University, East Bay

***Key Words:*** normality test, R language, Anderson-Darling, Shapiro-Wilk, Cramer-von Mises, Kolmogorov-Smirnov

Statistical packages can perform several different goodness-of-fit tests of normality. We consider the tests of Anderson-Darling, Shapiro-Wilk, Cramer- von Mises, and Kolmogorov-Smirnov. For a given dataset these tests sometimes lead to different conclusions about normality, possibly leaving students and practitioners confused about which test to believe. We use the statistical package R to simulate normal and non-normal data and to compare behaviors of these four tests. Specifically, we explore differences among the tests in several ways, focusing on reasons for their disagreement, on their relative power for several kinds of nonnormal data, and effects of using the tests in combination (for example, in terms of maximum and minimum P-values of several tests). Methods and R code are at an appropriate level for classroom use.

### Classroom Simulation: False Indications of Ouliers in Boxplots of Normal Data

❖ Bruce E. Trumbo, California State University, East Bay, 27047 Belfast Lane, Hayward, CA 94542, *bruce.trumbo@csueastbay.edu*; Eric A. Suess, California State University, East Bay; Jacob Colvin, California State University, East Bay

***Key Words:*** boxplot, outlier, simulation, R/S-Plus, pedagogy, teaching undergraduates

Computer packages often use boxplots of data to indicate "outliers": data values beyond fences located a certain multiple, often 1.5, of the interquartile range (IQR) on either side of the box bounded by the lower and upper quartiles. Simulation, presented here, shows that this definition of outlier yields surprisingly many false outlier indications in normal data of small or moderate sample size, and that the proportion of such indications is very sensitive to sample size. Simulation studies using R investigate the behavior of such outlier indications for several sample sizes, several multiples of IQR, and several parent populations. One behavior studied is the proportion of simulated samples with one or more outlier indications. Concepts and simulation programs are at a level appropriate for use in undergraduate statistics classes.

## 77 Section on Statistical Consulting Roundtable with Coffee (fee event)

Section on Statistical Consulting

**Monday, August 7, 7:00 am–8:15 am**

### Effective Collaboration via Concise Statistical Graphics

❖ Thomas G. Filloon, Procter & Gamble, 8700 Mason-Montgomery Road, Mason, OH 45040, *filloon.tg@pg.com*

***Key Words:*** statistical graphics, effective collaboration, ,

It can be helpful for both statistician and researcher (i.e., client) to use a graphical display as the common language (i.e., lingua franca) for collaboration. This mode of communication allows a less technical, common 'language' to be used as the foundation for gaining the necessary, deeper understanding of the 'problem.' Furthermore, this approach ensures subsequent statistical graphical output will meet expectations by both parties. Various types of useful statistical graphics will be discussed (e.g., categorical data displays, trellis data displays, box-percentile plots), as well as situations where this graphical approach may not work (if any).

## 78 Section on Statistical Education Roundtable with Coffee (fee event)

Section on Statistical Education

**Monday, August 7, 7:00 am–8:15 am**

### Using Japanese Lesson Study in Teaching Statistics

❖ Robert DelMas, University of Minnesota, 354 Appleby Hall, 128 Pleasant Street SE, Minneapolis, MN 55455, *delma001@umn.edu*; Joan Garfield, University of Minnesota

***Key Words:*** Japanese lesson study, research-based lessons, introductory statistics

The discussion leaders will provide a brief overview of how Japanese Lesson Study was used in an undergraduate introductory statistics course and address participants' questions about how to start a Japanese Lesson Study group; the role of leadership; necessary time commitments; needed materials and equipment; how to develop, present, and observe research-based lessons; how to conduct meetings; how observations are used to make revisions; and the products of Japanese Lesson Study.

## 79 Section on Statistics and the Environment Roundtable with Coffee (fee event)

Section on Statistics and the Environment

**Monday, August 7, 7:00 am–8:15 am**

### Comparison of Laboratory Methods for the Same Chemical

❖ Estelle Russek-Cohen, U.S. Food and Drug Administration, 1350 Piccard Drive, CDRH/OSB/DBS/DX, Rockville, MD 20850, *eyr@cdrh.fda.gov*

***Key Words:*** agreement measures, Deming regression, robust regression, measurement error models

In lab science, whether in a toxicology lab or in a clinical lab, a common task is to compare two lab methods for the same quantity. This can happen because of a modification to a lab procedure. I am hoping to compare how different disciplines have handled this issue. When I worked at the University of Maryland, the approaches taken were quite different than the ones I now see at FDA. I think a dialog about what works when and why would be useful.

## 80 Section on Statistical Graphics Roundtable with Coffee (fee event)

Section on Statistical Graphics

**Monday, August 7, 7:00 am–8:15 am**

### Graphics for Data Mining

❖ Martin Theus, University of Augsburg, Universit‰tsstr. 14, Augsburg, 86135 Germany, *martin.theus@math.uni-augsburg.de*

***Key Words:*** large datasets, data mining, interactive graphics, data cleaning, databases

This coffee session will highlight experiences with using graphics in data mining applications. Graphics are not only one of the most effective tools for checking the quality of data, they also can help discover special features in a large dataset. The design and implementation of statistical graphics should pay attention to the challenges from big datasets. For many users, this has not been an issue until now, so some statistical and graphics packages can have problems with graphics of more than 10,000 cases. However, most plots used in statistical graphics can be scaled up to be usable with large datasets.

## 81 Section on Physical and Engineering Sciences Roundtable with Coffee (fee event)

Section on Physical and Engineering Sciences
**Monday, August 7, 7:00 am–8:15 am**

### Data from Automotive Online Shoppers: Fact, Fiction, or Somewhere in between?

❖ Lynn Truss, GM Research & Development, MC 480-106 359, 30500 Mound Rd, Warren, MI 48090-9055, *lynn.t.truss@gm.com*

***Key Words:*** automotive, web data, online shoppers

Most automotive customers go online to research and shop for vehicles. Are they really providing a true signal for their wants and needs, or is it mostly noise? Maybe a combination of both? What's the best we can do as manufacturers with this kind of data? We can discuss key statistical issues, such as sampling and prediction, that arise in this context.

## 82 Section on Teaching Statistics in the Health Sciences Roundtable with Coffee (fee event)

Section on Teaching Statistics in the Health Sciences
**Monday, August 7, 7:00 am–8:15 am**

### Online-Based Approaches to Statistical Education

❖ John McGready, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205, *jmcgread@jhsph.edu*

***Key Words:*** online education, web-based courses

In this session, we will discuss online approaches to basic statistical education. I will give demonstrations of how we teach a "statistical reasoning"-centric course online at Johns Hopkins, and we will discuss the approaches and experiences of the session participants. We also will brainstorm on how to take an intensive, real-time, interactive laboratory-based course in statistical computing and make it web-ready. For years, we have taught two-week intensive courses in data analyses at Johns Hopkins and want to pursue it as an online offering. The varied and innovative perspectives of session participants guarantee to make this an exciting and informative exercise.

## 83 Introductory Overview Lectures: Statistical Consulting ● ✪

The ASA, ENAR, IMS, SSC, WNAR, Section on Statistical Consulting, Section on Statistical Education
**Monday, August 7, 8:30 am–10:20 am**

### The Business Side of Consulting

❖ Susan J. Devlin, The Artemis Group LLC, 1906 Westfield Ave., Suite B, Scotch Plains, NJ 07076, *sdevlin@theartemisgroup.com*

***Key Words:*** statistical consulting, building consulting practice

Through school, training, books, and journals, a statistical consultant acquires many valuable tools to practice effectively. However, very little training and publications are available that discuss the business aspects of running a successful consulting practice. In this talk, I focus on the impact the business side of consulting has on a practice. The following topics will be addressed: establishing your niche and business plan, structuring your fees, networking and finding clients, negotiating contracts and budgets, meeting deadlines, and managing your time. Practicing as a statistical consultant can be quite demanding and often stressful work. However, if a good a blend is achieved between the technical, business, and human side of statistical consulting, then a practice can be very rewarding, both monetarily and professionally.

### Creating Effective Encounters: the Heart of Meeting Global Challenges

❖ Douglas Zahn, 5x3 Associates, 249 Timberlane Road, Tallahassee, FL 32312, *zahn@stat.fsu.edu*

***Key Words:*** relationship, attitude, emotion, videotape, consulting, reactance

Statistical practice is a complex system, including teaching, research, consulting, and administration. At the heart of statistical practice is encounter-a purposeful meeting of a statistician with another person: colleague, student, client, supervisor, or member of staff. An encounter is a process consisting of five steps: preparing, beginning, working, ending, and reviewing. There is variation in this process as not all encounters are effective. To understand this variation and systematically reduce the number of ineffective encounters, gather primary data on the event by videotaping it. Analyze the data by using three lenses (interpersonal, intrapersonal, and technical) to identify and address barriers to effective encounters. I will give you an opportunity to learn how to apply this process to one of your current tough problems by using one of my videotapes of an actual consultation.

## 84 Statistical Models for Networks ✪

Chance, Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences, Section on Statistical Graphics
**Monday, August 7, 8:30 am–10:20 am**

### Information Processing in Cellular Networks

❖ Ravi Iyengar, Mount Sinai School of Medicine, Department of Pharmacology, Box 1215, One Gustave Levy Place, New York, NY 10029, *ravi.iyengar@mssm.edu*; Avi Ma'ayan, Mount Sinai School of Medicine

***Key Words:*** cell signaling, network dynamics, regulatory motifs, information processing

Information processing in mammalian cells occurs by interactions between cellular components that form a coupled-chemical-reaction network that has substantial information processing ability. To study this, we developed a network of 545 components and 1,259 interactions representing signaling pathways and cellular machines in a neuron. Using graph theory, we analyzed signal flow induced by receptor activation. Networking resulted in the emergence of regulatory motifs (feedback and feedforward loops and bifans) that process information. Key regulators of neuronal plasticity are highly connected nodes and required for formation of the motifs, suggesting these emergent regulatory motifs may be important for choosing between homeostasis and plasticity.

### Bayesian Solutions to the Degeneracy of Network Models

❖ Mark S. Handcock, University of Washington, Box 354322, Seattle, WA 98195-4322, *handcock@stat.washington.edu*; Martina Morris, University of Washington

***Key Words:*** random graph models, social networks, Markov chain Monte Carlo, Markov chain Monte Carlo; statistical exponential families, social sciences

Major barriers to the stochastic modeling of social networks are the specification of realistic models, the algorithmic difficulties of the inferential methods, and the assessment of the degree to which the graph structure produced by the models matches that of the data. These are related to the issue of model degeneracy for commonly proposed model classes. We proposes a Bayesian solution to this problem using a so-called nondegeneracy prior. We discuss inference under this model and an MCMC algorithm to implement it.

### Dynamical Analysis of Networks in Neural Systems

❖ Emery N. Brown, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, 77 Massachusetts Avenue 46-6079, Cambridge, MA 02139, *enbrown1@mit.edu*

***Key Words:*** state space models, adaptive filtering, Kalman filter, point processes, hidden Markov models

An important problem in neuroscience is characterizing how various brain regions represent information in their networks of neurons. These representations are studied using multielectrode arrays, electroencephalography, magnetoencephalography, and a range of imaging modalities, including functional magnetic resonance imaging. The statistical analysis of these high-dimensional and highly dynamic time series present interesting challenges for solving dynamic inverse problems. In this talk, we will summarize our recent work on the design and application of Bayesian signal processing algorithms to characterize the dynamics of brain activity. We will illustrate these methods using multielectrode recordings from neurons in the rat hippocampus while the animal executes learning tasks and EEG, MEG, and fMRI recordings from human subjects executing somatosensory and motor tasks.

## 85 Measurement Error Models ●

ENAR, Biometrics Section, WNAR

**Monday, August 7, 8:30 am–10:20 am**

### Semiparametric Methods and the Estimation of Nutrient Intakes

❖ Raymond J. Carroll, Texas A&M University, Department of Statistics, 3143 TAMU, College Station, TX 77843-3143, *carroll@stat.tamu.edu*; Arnab Maity, Texas A&M University; Yanyuan Ma, Texas A&M University

***Key Words:*** measurement error, semiparametrics, longitudinal data, repeated measures, nonparametric regression, nutrition

We consider general semiparametric models with repeated measures. The model consists of a parametric part and a nonparametric part consisting of the same function evaluated repeatedly for each subject. The problem thus includes semiparametric modeling of a very general kind for longitudinal/clustered data, for matched studies, for multivariate measurement error, etc. One example is a partially linear model with a correlation structure consisting of a person-specific effect plus random noise. In a random effects model, the person-specific effect is a random variable independent of the covariates. In a fixed effects model, the person-specific effect may be correlated in some unspecified manner with the covariates. We describe a methodology for both such problems, and apply the methods to an example involving income and caloric intake.

### A Latent Variable Model for Measurement Error Correction Using Replicate Data

❖ Sohee Park, National Cancer Center, Korea, 809 Madu1-dong, Ilsandong-gu, Goyang-si, Gyeonggi-do, 410-769 Republic of Korea, *shpark@ncc.re.kr*; Louise Ryan, Harvard School of Public Health; David Ruppert, Cornell University; John Meeker, University of Michigan; Russ Hauser, Harvard School of Public Health

***Key Words:*** measurement error, short-term and long-term variability, regression calibration, maximum likelihood, estimating equation

It is well-known that when the exposure variable is measured with error, the estimated relationship between exposure and outcome can be seriously biased unless appropriate adjustments are made. To assess within-person variability in the exposure, replicate measurements often are obtained. Using an estimating equations approach, we formulate the correctly estimated variance for the parameter of association between exposure and outcome. This method allows for various covariance structures among replicates (e.g., where exposure variable has short-term and long-term variation so the repeated measures are not simply replicates of each other and replicates are observed on only a subset of study subjects). We also apply these methods to real data from a study of pesticides exposure and male reproductive health outcome.

### Measurement Error in Population Dynamics Models

❖ John Staudenmayer, University of Massachusetts, 16 Jason Court, Amherst, MA 01002, *jstauden@math.umass.edu*; John Buonaccorsi, University of Massachusetts

***Key Words:*** error in variables, Ricker's model, random walk with drift, heteroscedasticity

Population abundances are rarely, if ever, known. Instead, they are estimated with some amount of uncertainty. The resulting measurement error has consequences on subsequent analyses that model population dynamics and estimate probabilities about abundances at future points in time. This work addresses outstanding questions about the consequences of, and ways to correct for, measurement error in several dynamic models. We provide analytical results on the biases that result from a naive analysis, which ignores measurement error under a class of models that allow for heteroscedasticity in the measurement errors. We also develop and evaluate moment- and likelihood-based methods to correct for measurement error, with and without the use of estimated measurement error variances. The methods are applied to existing population dynamics datasets. This is joint work with John Buonaccorsi.

### Generalized Measurement Error Models and Bias Reduction

❖ Leonard A. Stefanski, North Carolina State University, Department of Statistics, Room 201-B, Patterson Hall, 2501 Founders Drive, Raleigh, NC 27695-8203, *stefansk@stat.ncsu.edu*

***Key Words:*** estimating degrees of freedom, information reduction, noise addition, phony variable addition, variable selection, weighted LS

The talk will open with a review and unification of bias reduction methods in general, and for measurement error models in particular. Then I will explain how the common principles of bias reduction can be used to address standard and not-so-standard measurement error regression models; and also to address problems that are not usually regarded as measurement error problems per se, but that are, in a broad sense, generalized measurement error models (GMEM). Applications to measurement error regression modeling, variable selection (in non-measurement error regression models), and estimating degrees of freedom will be used to motivate, explain and illustrate the bias reduction strategy.

## 86 Directions for the Second Statistics Course ●

Section on Statistical Education, Section on Teaching Statistics in the Health Sciences

**Monday, August 7, 8:30 am–10:20 am**

### What Should Be the Second Statistics Course?

❖ Karla V. Ballman, Mayo Clinic College of Medicine, College of Medicine, Division of Biostatistics, Kahler 1A, 200 First Street SW, Rochester, MN 55905, *ballman@mayo.edu*

***Key Words:*** statistical education, statistics course

My response to this question is similar to my response to students who ask if it is more correct to report the mean or the median: It depends. Essentially, the choice of the second statistics course is driven by what the student plans to do beyond attainment of their undergraduate degree. Different second courses may be appropriate for different majors (i.e., statistics versus mathematics versus biology versus engineering versus economics versus something else). However, in most instances, I would advocate the second course should emphasize statistical modeling. Ideally, second courses designed for students who are not mathematics or statistics majors would emphasize modeling used in their major field of study. My discussion will focus on reasons for advocating the second statistics course be based on statistical modeling.

### Methods of Data Analysis: a Second Course in Statistics at Oregon State University

❖ Fred Ramsey, Oregon State University, Department of Statistics, Kidder 44, OSU, Corvallis, OR 97331, *ramsey@science.oregonstate.edu*

The objectives of the second course in statistics at Oregon State University primarily target graduate students in fields other than statistics. We equip them with tools that enable them to use statistical software for data analysis problems they will encounter in their graduate research. We train them to read, understand, and critically evaluate statistical arguments presented in the literature of their fields. They are schooled in the principals of experimental design, and they practice writing statistical summaries of data analysis problems, with special emphasis on the scope of inferences allowed by the study design. The course also serves to motivate and prepare undergraduates who may enter graduate studies in statistics.

### Second Statistics Course at a Liberal Arts College

❖ Michael Kahn, Wheaton College, 123 E. Main Street, Norton, MA 02766, *mkahn@wheatonma.edu*

***Key Words:*** second course, case studies, small college

We will discuss a second course in statistics at a small liberal arts college. The design is similar to courses discussed in Schafer and Ramsey (2003) and Roback (2003). The course is driven by case studies and readings from scientific literature. Students are mostly mathematics majors or minors, but also are social and life science majors with less mathematical background. The audience moderates the case studies to a large extent, and the goal is to engage in collaborative learning projects throughout the course, each student bringing their area-specific skills and interests to the scientific and statistical analyses.

### Risky Business: a Second Statistics Course at Business Schools

❖ Norean R. Sharpe, Babson College, Babson Hall, Babson Park, MA 02457, *sharpen@babson.edu*; Richard Cleary, Bentley College

***Key Words:*** teaching, modeling, business statistics, decision analysis, time series

More quantitative skills are being expected of business school graduates. Given the advances in technology and the challenges facing corporations, future managers and leaders need to have a solid grasp of randomness, modeling techniques, and data analysis strategies. Whether students aspire to work in a financial, marketing, or managerial context, the concepts of risk and uncertainty play an important role. To meet these needs and improve student learning at the undergraduate level, this presentation will outline second statistics courses that include topics in modeling, time series analysis, decisionmaking, and simulation. Our hope is that the more students are taught to identify and analyze risk, the better they will understand the impact of their decisions.

### Spanning the Parametric/Nonparametric Divide

❖ John Holcomb, Jr., Cleveland State University, 2121 Euclid Ave., Rhodes Tower 1515, Cleveland, OH 44115-2214, *j.p.holcomb@csuohio.edu*

***Key Words:*** parametric statistics, nonparametric statistics, statistics education

This paper will address the difficult issue of teaching a second course in statistics with current statistics textbooks. Currently, students learn of parametric tests early in the book, and the nonparametric tests are relegated to the last chapter. In this session, I advocate having students in a second course learn the graphical and statistical tests for normality and testing hypothesis for means with both parametric and nonparametric tests. The second course is an opportunity to reteach the parametric methods taught too quickly in a first course and compare them with their nonparametric counterparts.

## 87 Forest Fires and Fish Stocks: Statistics on Long-Term Trends from Lake Sediments and Tree Rings ● ✪

SSC, Section on Statistics and the Environment

**Monday, August 7, 8:30 am–10:20 am**

### Patterns of Biological Productivity in Saanich Inlet

❖ Frank Whitney, Institute of Ocean Sciences, 9860 W. Saanich Road, Sidney, BC V8L 4B2 Canada, *whitneyf@pac.dfo-mpo.gc.ca*; Ann Gargett, Institute of Ocean Sciences; Melissa McQuoid, Institute of Ocean Sciences

Biological activity in the surface waters of Saanich Inlet are controlled strongly by solar and lunar cycles. Primary productivity is high in spring and subsequently shows pulses on two-week tidal cycles. A simple physical model, using tidal mixing and discharge from local rivers to drive circulation, is able to reproduce the periodic supply of nutrient-rich waters to the inlet. Because Saanich Inlet has anoxic sediments, a record of the annual and fortnightly productivity cycles is preserved. Cores taken from the inlet basin show light and dark banding that reflects sedimentation alternately of phytoplankton or terrigenous materials. A long core shows there has been little change in the phytoplankton populations of the inlet over the past 7,000 years, although there is evidence that primary production and fish abundance may have decreased over the past millennium.

### Inference of Past Environmental Changes from Fossil Midges

❖ Ian Walker, The University of British Columbia, Department of Biology, 3333 University Way, Kelowna, BC V1V 1V7 Canada, *ian.walker@ubc.ca*

Midge fossils have emerged as one of the leading indicators of past environmental changes. They are being widely used to infer past changes in lakewater chemistry and summer climate. In this paper, I review the development of midge fossils as palaeoenvironmental indicators, with emphasis on current statistical inference procedures and problems.

### Forest Fires in Space and Time: Comparing Long Records of Forest Fires To Understand Their Climatic Controls

❖ Dan Gavin, University of Vermont, Botany and Agricultural Biochemistry, 120B Marsh Life Science, 109 Carrigan Dr., Burlington, VT 05405-0086, *dgavin@uvm.edu*; Kenneth Lertzman, Simon Fraser University; Feng S. Hu, University of Illinois

A century-long trend of increasing summer temperature in the Pacific Northwest, along with increasing loads of forest fuels, has raised concerns that forest fires are becoming more frequent, larger, and more difficult to control. We show how fire records are constructed from the detection of charcoal peaks in lake sediment cores. We then compare fire-event records of the past 5,000 years from two lakes that experienced the same late-Holocene climatic changes. Using such tools as the Weibull distribution and bivariate Ripley K-function, we evaluated the fire-interval distribution and fire synchroneity. Results indicate that sites with similar modern conditions may have experienced different fire intervals and asynchronous patterns in fire episodes and that the influence of climate on fire occurrence has varied with changing climatic variability over time.

## 88 Mining and Exploring Clinical Trials Data: Why, What, and How? ● ✪

Biopharmaceutical Section, Committee on Applied Statisticians, Section on Statistical Consulting, ENAR, Section on Statistical Graphics

**Monday, August 7, 8:30 am–10:20 am**

### Secondary Analysis of Clinical Trials and Claims Data for Risk Minimization and Outcomes Evaluation

❖ Michael O'Connell, Insightful Corporation, 2505 Meridian Parkway, Suite 175, Durham, NC 27713, *moconnell@insightful.com*

***Key Words:*** clinical, data mining, causal inference, propensity scoring, cross-trial analysis

Phase-three clinical trial results are typically complex, with multiple efficacy endpoints, multiple analyses for each endpoint, and multiple safety measures (i.e., adverse events). While these data typically are not analyzed after the regulatory filing, we contend that secondary trial analyses can identify populations with enhanced therapeutic effects, biomarkers for use in future studies, and safety effects for creation of risk minimization action plans. Retrospective analysis of postmarketing databases is a common approach to understanding clinical outcomes and therapy cost-benefit. Because these data are not from prospective, randomized, blinded designs, causal inference methods are required for valid inference. This presentation covers issues in the analysis of secondary clinical trial and retrospective postmarketing data by way of case studies in these settings.

### Data Mining Trees: Mining Clinical Trials Data

❖ Javier Cabrera, Rutgers University, Statistics and Biostatistics, Hill Center, Busch, Piscataway, NJ 08854, *cabrera@rci.rutgers.edu*

***Key Words:*** datamining, clinical trials, bump hunting, recursive partition

Mining clinical trial data is becoming an important tool for extracting information that may help design better clinical trials. One important objective is to identify the characteristics of a subset of cases that responds much differently than the rest of the cases. For example, what are the characteristics of placebo respondents or the highest respondents or lowest respondents to some treatment? Are secondary endpoints higher for some group of patients? The two existing methodologies that try to address these issues are "bump hunting" and "recursive partitioning." We introduce data mining trees as a method that compromises between recursive partitioning and bump hunting. We illustrate the methodology with examples that use clinical trial data. This work is a collaboration with J. Alvir; H. Nguyen; M. Lakshminarayanan, Pfizer; and D. Amaratunga, JnJPRD.

### Logistic Regression on Autopilot

❖ Trevor Hastie, Stanford University, Statistics Department, Stanford, CA 94305, *hastie@stanford.edu*

***Key Words:*** boosting, LASSO

Logistic regression is a standard modeling tool in biopharmaceutical applications. It typically is used in controlled settings with a limited number of variables and low-order interactions, if present. In this talk, I will discuss two procedures for automating logistic regression modeling with large datasets: GLMpath fits an entire lasso sequence of logistic regression models in an efficient manner and tree-based gradient boosting does the same with automatic detection of interaction effects.

## 89 Statistics and the Millennium Development Goals ● ✪

Section on Government Statistics
**Monday, August 7, 8:30 am–10:20 am**

### The Role of Monitoring and Evaluation in Development Programs

❖ Paul Wassenich, Inter-American Development Bank,
1300 New York Ave., NW, Washington, DC 20577,
*PWASSENICH@Contractual.iadb.org*

Considerable resources have been directed to "development" or "poverty reduction" in poor countries over the past 50 years. While some success stories provide encouraging exceptions, the results generally have been disappointing. As significant, albeit limited, resources continue to be invested in interventions to reduce poverty, it is imperative we understand how these interventions actually affect the welfare of the poor so they can be channeled in ways that achieve maximum impact. Yet empirical evidence of what works in development is very limited. This paper will discuss the present role and suggest the future role of monitoring and evaluation within development programs.

### Statisticians and the Millennium Development Goals

❖ Paul Fields, Brigham Young University, Department of Statistics,
230 TMCB, Provo, UT 84602, *pjfields@stat.byu.edu*

The Millennium Development Goals (MDGs) represent a significant initiative on the part of the United Nations and its member countries. Achieving these goals would have huge impact on the quality of life of millions of people around the world. However, the time frame set for achieving the MDGs makes this a highly ambitious project. Consequently, there is too much at stake to base billion-dollar decisions on anecdotal evidence. It is vitally essential to establish "best practices" based on rational and objective examination of results during and after programs. If the funding for all development programs included funds for statisticians, the probability of achieving the MDGs could increase significantly. We suggest the ASA take a proactive leadership role to bring statisticians into the development effort and work collaboratively with our statistics colleagues around the world.

### A Role for Experimental Evaluation in Efforts To Achieve Millennium Development Goals

❖ David J. Fitch, Universidad del Valle de Guatemala, Apartado
Postal 82, 01901, Guatemala, 01901 Guatemala, *dfitch@uvg.edu.gt*

***Key Words:*** education, Guatemala, LQAS

To achieve MDGs, Guatemala needs better schools. Teachers are poorly paid high-school graduates. The current Minister has well-thought-out plans for improvement that should be introduced to make possible an experimental evaluation. The country would be divided into districts with a junior high and senior high in each. Similar districts would be paired. One pair would be selected initially, but as more teachers---who would be university graduates---became available, more pairs would be selected. One district of each pair would be picked randomly to receive the program. Over the years, criterion data would be collected in the experimental and control districts, such as salary, development, academic achievement, school attendance, percent of girls graduating, age of marriage, malnutrition rates, crime, birth rate, etc. Such evaluations would strengthen an MDG grant application. Donors can't help.

## 90 Semisupervised Learning

IMS, Biometrics Section, Section on Nonparametric Statistics
**Monday, August 7, 8:30 am–10:20 am**

### Semisupervised Learning: an Overview

❖ Xiaojin Zhu, University of Wisconsin-Madison, Madison, WI
53705, *jerryzhu@cs.wisc.edu*

Semi-supervised learning learns from both labeled and unlabeled data. It is important because labeled instances are often difficult, expensive, or time-consuming to obtain. They require the efforts of experienced human annotators, while unlabeled data may be relatively easy to collect. In this talk, I will review interesting ideas in the field. Topics include mixture models and EM, self-training, co-training, transductive support vector machines, and graph-based regularization. I will then present a few issues and challenges in semi-supervised learning research.

### Statistical and Geometric Principles of Semisupervised Learning

❖ Mikhail Belkin, The Ohio State University, Computer Science
Department, 395 Dreese Laboratories, 2015 Neil Avenue, Columbus,
OH 43210-1277, *mbelkin@cse.ohio-state.edu*

***Key Words:*** unlabeled data, geometry of data, semi-supervised

While unlabeled data is ubiquitous in the real world and arguably plays a key role in natural learning and perception, relatively little is known about its theoretical and algorithmic properties. I will discuss geometric properties of probability distributions---which make semisupervised learning possible---and methods for nonparametric inference based on that point of view.

### Variable Selection for Semisupervised Learning

❖ Elizaveta Levina, University of Michigan, Department of Statistics,
1085 S. University, Ann Arbor, MI 48109, *elevina@umich.edu*; Ji Zhu,
University of Michigan

***Key Words:*** semi-supervised learning, manifold projections, variable selection, LASSO, regularization

Manifold regularization has been shown to be a useful tool in semi-supervised learning, because it provides a way of learning the local data structure using both labeled and unlabeled data. Manifold- and kernel-based classifiers, however, are impossible to interpret in terms of the original variables and cannot be used for variable selection. We combine manifold regularization with variable selection by considering linear classifiers penalized by a lasso penalty and a "local-structure" penalty. We show this approach can work well in both supervised and semi-supervised settings, and suggest an extension to nonlinear kernels.

## 91 Some Current Issues in Industrial Statistics ● ✪

Section on Physical and Engineering Sciences

**Monday, August 7, 8:30 am–10:20 am**

### Use of Computer Simulation To Plan Complicated Industrial Experiments

❖ William Q. Meeker, Jr., Iowa State University, 304C Snedecor Hall, Department of Statistics, Ames, IA 50011-1210, *wqmeeker@iastate.edu*

***Key Words:*** censoring, maximum likelihood, experimental design, reliability, nonlinear design

This talk describes methods and procedures for using Monte Carlo simulation and graphics to plan reliability and other nonstandard industrial experiments. Such experiments include accelerated life tests, repeated measures degradation tests, and accelerated destructive degradation tests. The design of such experiments typically answers questions about sample size, length of the test, and allocation of test units to levels of the accelerating variable(s). Models have complications such as random effects and censoring. Thus, standard experimental design tools need to be extended. I will describe methods that employ graphical displays for combinations of large-sample approximations for precision metrics and for the display of simulation results. Simulation will be shown to be a particularly versatile and valuable tool for providing insights into such complicated design problems.

### Forecasting Warranty Claims

Jerald F. Lawless, University of Waterloo; ❖ Marc Fredette, HEC MontrÈal, 3000 Chemin de la Cote Sainte Catherine, Montreal, PQ H3T 2A7 Canada, *marc.fredette@hec.ca*

***Key Words:*** prediction, repeated events, Poisson processes, random effects

We discuss methods for predicting the eventual total number of warranty claims in a large population of units manufactured and sold over a period of time. Prediction intervals are sought. These will initially be based on limited data concerning the units in question, and thus be rather wide. As warranty claims data accumulate, the intervals will narrow. Methods based on Poisson processes with random effects, combined with calibration by simulation, will be used to generate frequentist prediction intervals. An application to the prediction of automobile warranty claims will be considered.

### Degradation-Based Models for Analyzing Repairable Systems

❖ Vijay Nair, University of Michigan, Department of Statistics, 1085 South University 439 West Hall, Ann Arborm, MI 48109-1107, *vnn@umich.edu*; Anupap Somboonsavatdee, University of Michigan; Ananda Sen, University of Michigan

***Key Words:*** reliability

The nonhomogeneous Poisson process (NHPP) is perhaps the most commonly used model for analyzing failure data on repairable systems. In this talk, we describe an analogous model for analyzing degradation data from repairable systems based on nonhomogeneous Gaussian processes. Extensions to nonhomogeneous Levy processes will be mentioned. Just as the NHPP can handle different types of intensity functions, these processes can accommodate a variety of degradation rates. Different types of imperfect repair and aging of the system and environment also can be incorporated. Both nonparametric and parametric inference will be described.

## 92 Recent Advances in Design and Analysis of Vaccine Studies

Biometrics Section, ENAR, WNAR

**Monday, August 7, 8:30 am–10:20 am**

### On Estimation of Vaccine Efficacy Using Validation Samples with Selection Bias

❖ Michael Daniels, University of Florida, 207 Griffin Floyd Hall, Department of Statistics, Gainesville, FL 32611, *mdaniels@stat.ufl.edu*; Daniel Scharfstein, Johns Hopkins Bloomberg School of Public Health; M. Elizabeth Halloran, Fred Hutchinson Cancer Research Center; Haitao Chu, Johns Hopkins Bloomberg School of Public Health

Using validation sets for outcomes can improve greatly the estimation of vaccine efficacy (VE) in the field. Most statistical methods for using validation sets rely on the assumption that outcomes on those with no cultures are missing at random. However, often the validation sets will not be chosen at random. For example, confirmational cultures often are done on people with influenza-like illness as part of routine influenza surveillance. Vaccine efficacy estimates based on such nonrandom validation sets could be biased. Here, we propose an approach to sensitivity to selection bias when using validation sets in estimating vaccine efficacy, building on the work of Rotnitzky et al. (1998, 2001), Scharfstein et al. (1999, 2003), and Robins et al. (2000). We develop both frequentist and Bayesian approaches.

### Accounting for Variability in Sample Size Estimation with Application to a Malaria Vaccine Phase II Trial

❖ Michael P. Fay, National Institute of Allergy and Infectious Diseases, 6700B Rockledge Drive, MSC 7609, Bethesda, MD 20892-7609, *mfay@niaid.nih.gov*; M. Elizabeth Halloran, Fred Hutchinson Cancer Research Center; Dean A. Follmann, National Institute of Allergy and Infectious Diseases

***Key Words:*** sample size, binomial, compliance, Poisson, normal, variability

We use data from a longitudinal study in Mali to study several possible endpoints for designing a Phase 2 vaccine trial for a Malaria vaccine. We assume the differences in endpoints between 4-year-old control subjects and 4-year-old vaccinated subjects are similar to observed differences between nonvaccinated 4 and 8 year olds. Because there are fewer than 50 children within one year of each age group in the preliminary data, there is considerable variability in the resulting sample size estimates. We develop a general method for adjusting the sample size estimates to account for this variability so the final sample size estimates have the proper power, assuming only that the preminary data and the planned study follow the same probability model. In addition, we adjust the sample sizes to account for the case when only a proportion of vaccinated subjects respond to the vaccine.

### Statistical Methods for Assessing HIV Vaccine Efficacy in Repeated Low-Dose Challenge Experiments

❖ Michael G. Hudgens, The University of North Carolina at Chapel Hill, 3107E McGavran-Greenberg Building, Chapel Hill, NC 27599, *mhudgens@bios.unc.edu*; Peter Gilbert, Fred Hutchinson Cancer Research Center/University of Washington

***Key Words:*** causal inference, HIV, vaccines

Evaluation of HIV vaccine candidates in nonhuman primates (NHPs) is a critical step toward developing a successful vaccine to control the HIV pandemic. Historically, HIV vaccine regimens have been tested in NHPs by administering a single high dose of the challenge virus. More recently, evaluation of candidate HIV vaccines has entailed repeated low-dose challenges that more closely mimic typical exposure in natural transmission settings. In this talk, we will consider evaluation of the type and magnitude of vaccine efficacy from such experiments. Based on the principal stratification framework, we also will discuss design and analysis considerations in assessing potential immunological correlates and surrogates of protection.

## 93 Statistical and Computational Issues in Climate Research ● ✪

Section on Statistical Computing, Section on Physical and Engineering Sciences, Section on Statistical Graphics
**Monday, August 7, 8:30 am–10:20 am**

### Statistical Analysis of Spatial Patterns of Climate Variability

❖ J. Michael Wallace, University of Washington, Department of Atmospheric Sciences, 106 King Building, Box 354235, Seattle, WA 98195-4235, *wallace@atmos.washington.edu*

The role of statistical methods in identifying and defining favored patterns or "modes" of climate variability, such as ENSO and the annular modes, will be reviewed and discussed. The spatial patterns of climate trends over the past 50 years project strongly on these modes. The modes, in turn, are useful in relating global-scale climate variability to local climate impacts, such as the frequency of cold air outbreaks or heavy rain events. Among the statistical issues involved in attempting to characterize climate variability in terms of the time-dependent behavior of a limited number of "modes" are: To what extent is climate variability linear? Are the modes recovered by principal component analysis physically relevant? Does rotation of the principal components improve the robustness and physical relevance of the patterns?

### Interpreting Recent Climate Change

❖ Francis W. Zwiers, Canadian Centre for Climate Modelling and Analysis, P.O. Box 1700, STN CSC, Victoria, BC V8W 2Y2 Canada, *francis.zwiers@ec.gc.ca*

***Key Words:*** climate modeling, climate change detection and attribution, climatic extremes

The climate research community makes extensive use of computationally expensive climate models to interpret past climate change as recorded in the historical instrumental data of the past 100--150 years and paleo-climate reconstructions of the past 1--2 millennia. This research has quantified the influence of external forcing from natural and anthropogenic sources on the climate system on global and regional scales, has contributed substantially to our understanding of the sensitivity of the climate system to forcing change such as a doubling of greenhouse gas concentrations, and is beginning to broach questions concerning changes in the intensity and frequency of climate extremes. In this talk, I will describe briefly climate models, the use of those models in the interpretation of historical climate changes, and some of the ensuing statistical issues.

### Statistical Problems in Climate Change and Geophysical Fluids

❖ Carl Wunsch, Massachusetts Institute of Technology, Room 54-1524, MIT, Cambridge, MA 02140, *cwunsch@mit.edu*

The study of climate change and global-scale fluids in the climate system (ocean and atmosphere) raises many statistically challenging problems. These range from the difficulties of separating nonstationarity from non-normal behavior---and the significance of apparent trends in phenomena with multiple memory scales---to time-series analysis in the presence of unknown time-errors, to the display error fields in model results, where the dimensionality is overwhelming. One must do statistical inference with short, noisy, records; the interactions of complex physics, chemistry, and biology; and where inferences that are important to society are being drawn. Examples will be given of these various issues, including inferences about the possibility of abrupt climate change, "resonances" in the climate system, nonlinear phase entrainment, and state estimates with order 10^9 dof.

## 94 SAMSI Session on Statistical Issues in Metabolomics ● ✪

Statistical and Applied Mathematical Sciences Institute, Biometrics Section
**Monday, August 7, 8:30 am–10:20 am**

### Measurement Issues in Metabolomics Data

❖ Xiaodong Lin, University of Cincinnati, Department of Mathematics, Cincinnati, OH 45221, *linxd@math.uc.edu*

Metabolomics is a new bioinformatics area that has great potential for diagnosing certain kinds of disease, assessing drug toxicity early, and improving understanding of biochemical pathways. Progress depends on statistical characterization of the measurement process and the ability to make cross-platform comparisons. Doing this requires appropriate algorithms to baseline-correct signals in the mass spectrometry, extract abundances from apparent peaks, and describe variance components associated with each part of this complex measurement process.

### Exploring a Complex Metabolomics Dataset

❖ Susan J. Simmons, The University of North Carolina at Wilmington, Department of Mathematics and Statistics, 601 South College Road, Wilmington, NC 28403, *simmonssj@uncw.edu*; Xiaodong Lin, University of Cincinnati

***Key Words:*** metabolomics, data mining

Metabolomics, also known as biochemical profiling, has the potential for identifying possible targets for pharmaceutical interventions. The concentrations of all small molecules, known as metabolites, produce the best and most direct measure of physiological activity in an organism. By examining differences in biochemical profiles of individuals who are 'healthy' to those who are 'diseased,' we can identify potential targets for drug intervention. However, metabolomic datasets are often complex with a number of interesting statistical challenges. We examine a complex metabolomic dataset to identify differences in the biochemical profiles of diseased individuals versus those who are healthy.

### Pathway-Based Analysis of Metabolic Profiles

❖ Jacqueline Hughes-Oliver, North Carolina State University, Department of Statistics, Box 8203, Raleigh, NC 27695, *hughesol@stat.ncsu.edu*

Metabolomics is emerging as an attractive component of the extensive body of platforms for systems biology. The relatively small number of metabolites and their mostly known network of interconnectivity and relation to disease make metabolomics a prime candidate for improving both disease diagnosis and treatment. Unfortunately, analysis of metabolomic data typically ignores knowledge of disease pathways, and hence the opportunity for taking full advantage of domain knowledge is lost. Using the cholesterol portion of the biosynthesis of steroids pathway, we develop a technique for incorporating domain knowledge in the analysis of metabolites. Our focus is on diagnosing individuals at high risk for cholesterol-related disease.

## 95 Design and Analysis in Medical Devices ● ✪

Biopharmaceutical Section, Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Hypothesis Testing in Clinical Study with Mixed Patient Population

❖ Yonghong Gao, U.S. Food and Drug Administration, 1350 Piccard Drive, Rockville, MD 20850, *yonghong.gao@fda.hhs.gov*; Chul H. Ahn, U.S. Food and Drug Administration

***Key Words:*** testing, mixed patient, subgroup, OPC

In a one-arm registry study for a medical device, an objective performance criterion (OPC) is sometimes used to approve the studied device. In some medical device registry studies, the target patient population consists of two subgroups of patients with unknown proportions, and the overall OPC for the target population (usually unknown) is expressed as the weighted average of two individual OPCs defined for each subgroup. This presentation will propose the idea of random sampling to estimate the overall OPC and test statistics to implement the hypothesis testing. Simulation results will be given to compare the performance of the proposed test statistics.

### Placebo Control Group in Medical Device Clinical Trials

❖ Sarah Kogut, W. L. Gore & Associates, Inc., 4100 W. Kiltie Lane, Flagstaff, AZ 86001, *skogut@wlgore.com*

***Key Words:*** therapeutic medical device, clinical trials, placebo control group, ethics, research methodology

Technologically advanced therapeutic medical devices, such as cardiovascular pacemakers or neurostimulators, provide a valuable option for patients when standard treatments are inadequate. Demonstration to regulatory authority is based on providing valid scientific evidence that the device is safe and effective. In earlier days of the medical device industry, single-arm studies were the usual design in device clinical trials. More recently, concern has been expressed regarding placebo effect, where patient outcome may be influenced by belief that device therapy is effective. There has been discussion about the design of a placebo control arm, including ethical, practical, and scientific aspects. This talk will give a historical description of the debate and present specific examples of placebo controls. Statistical design issues and possible strategies will be discussed.

### To Pair or Not To Pair? Noninferiority Test for the Difference in Correlated Binary Outcomes

❖ Dong Li, Guidant Corporation, 4100 Hamline Ave., N., Suite F321, St. Paul, MN 55112, *dong.li@guidant.com*; David Breiter, Guidant Corporation; Shu Han, Guidant Corporation; Jia Wang, Guidant Corporation

***Key Words:*** non-inferiority test, correlated binary outcomes

Noninferiority tests are used widely in the medical device industry. Two designs are based on extreme assumptions: unpaired setting assumes independency between two responses while paired setting assumes high correlation between two responses. Nam and Blackwelder's studies show the paired design has a clear advantage over the unpaired design in terms of asymptotic relative efficiency (ARE) when two responses of the pair are highly positively correlated. However, extended power analysis for the noninferiority test is not yet available to describe how the test power changes while the correlation between two responses varies. We evaluate the noninferiority test power with correlation varying between two responses. The results are compared further with the ARE between paired and unpaired settings. Extensive Monte Carlo simulations are conducted to demonstrate our findings.

### Covariate-Adjusted Noninferiority Tests with Binary Endpoints

❖ Lei Peng, Guidant Corporation, 3200 Lakeside Drive, Santa Clara, CA 95054, *lpeng@guidant.com*

***Key Words:*** non-inferiority test, binomial endpoint, covariates adjustment, odds ratio, non-parametric analysis of covariance, binary endpoint

Given the number of approved drugs, it is more common to conduct noninferiority clinical trials. The present study focuses on hypothesis testing of binary endpoints after adjustment for covariates. In the absence of covariates, asymptotic test or confidence limit methods can be applied directly to the estimates of proportions. However, issues arise for covariate-adjusted noninferiority tests. The confidence limit cannot be compared directly with pre-defined noninferiority margin because treatment effects in covariate-adjusted analysis, such as logistic regression, is measured in odds ratio, rather than difference in proportions. Two methods have been proposed for covariate-adjusted tests. One is to use the equivalent noninferiority margin for odds ratio. The other is based on a nonparametric analysis of covariance. We will compare the merits of these methods and present simulation results.

### Noninferiority Testing with a Variable Margin

❖ Zhiwei Zhang, U.S. Food and Drug Administration, 1350 Piccard Drive, Division of Biostatistics HFZ550, Rockville, MD 20850, *zhiwei.zhang@fda.hhs.gov*

***Key Words:*** delta method, non-inferiority, likelihood ratio test, power, sample size, score test

There has been growing interest, when comparing an experimental treatment with an active control with respect to a binary outcome, in allowing the noninferiority margin to depend on the unknown success rate in the control group. It does not seem universally recognized, however, that the statistical test should appropriately adjust for the uncertainty surrounding the noninferiority margin. We inspect a naive procedure that treats an "observed margin" as if it were fixed a priori, and explain why it might not be valid. We then derive a class of tests based on the delta method, including the Wald test and the score test, for a smooth margin. We discuss the asymptotic behavior of these tests and the likelihood ratio test when applied to a piecewise smooth margin. Simulation experiments are conducted to evaluate the finite-sample performance of the asymptotic tests studied.

## 96 Modeling and Testing of Economic Time Series ● ✪

Business and Economics Statistics Section
**Monday, August 7, 8:30 am–10:20 am**

### Cost-of-Living Index Based on an Estimated Variable Elasticity of Substitution Utility Function

❖ Peter Zadrozny, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212, *zadrozny.peter@bls.gov*

***Key Words:*** estimating time series models, solving nonlinear equations

Using monthly price-quantity data for eight categories of consumer expenditures, we obtain ML parameter estimates of a variable-elasticity-of-substitution utility function, a generalization of the CES utility function. Estimating equations are a linear-in-variables VARMA process, subject to nonlinear parameter restrictions representing VES utility maximization. We also account for trend, cycle, and seasonality, which represent time-varying consumer preferences. Then, we use the multistep perturbation method---for solving nonlinear equations---to compute month-to-month changes in the cost-of-living index based on the estimated VES utility function for the price-quantity data. The resulting COLI should reflect substitution in aggregate consumer expenditures more accurately than commonly publicized, so-called superlative indexes---such as Tornqvist---due to relative price changes.

### Stochastic Volatility and Cointegration in Statistics Canada's Retail Trade Series

❖ Thierno A. Balde, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A0T6 Canada, *thierno.balde@statcan.ca*; Ioana Schiopu-Kratina, Statistics Canada; Benoit Quenneville, Statistics Canada

***Key Words:*** stochastic volatility, ARCH-GARCH models, cointegration, ARIMA transfer functions, linear regression

In this paper, we illustrate the concepts of volatility and cointegration introduced by R. F. Engle and C. W. Granger on Statistics Canada's Retail Trade Series. Although these concepts have been used extensively for economic time series, they do not seem to have been applied to seasonally adjusted time series typically produced by statistical agencies. We show that there is volatility in the New Motor Vehicle Survey series (NMVS). We fit a GARCH- and an ARIMA-type model and compare their performances. We find that the GARCH model gives better predictions than the ARIMA model, as it takes into account the existing volatility. We also find a cointegration relationship among the Monthly Retail Trade Survey series and two of its components, namely NMVS and Large Monthly Retailers series. Then, We use ARIMA transfer functions to model this cointegration relationship.

### Model-Based Formulas for Growth Rates and Their Standard Errors

❖ Tucker S. McElroy, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-9100, *tucker.s.mcelroy@census.gov*

***Key Words:*** seasonal adjustment, signal extraction, ARIMA models, lognormal distribution, unobserved components models

The growth rate of a time-series component gives the percentage change from one time to another. Growth rates, calculated from seasonally adjusted macroeconomic indicators and trends, are among the most widely used statistics in discussions of the state of the economy. Model-based methods of calculating mean square optimal estimates of seasonal adjustments, trends, and other data components can be extended to produce optimal estimates of their growth rates, as well as associated standard error measures of their uncertainty. This is useful information that non-model--based approaches cannot provide readily. In a model-based context, we present formulas for growth rates, which are minimum mean square error estimates given a finite sample of log-normal data. We also derive formulas for the standard errors and extend to the non-Gaussian case through the use of approximate formulas.

### Measuring Trend Growth Rates with an Application to Consumer Prices

❖ Peter Kenny, PBK Research, 53 Lindal Crescent, Enfield, EN2 7RP UK, *pkenny@globalnet.co.uk*

***Key Words:*** trend, growth rate, error margin

Many time series are measured as levels, but the main interest lies with the rate of change; the obvious example is price indices. There are well-known ways of measuring the trend of the level, but the trend of the growth rate is not well-defined. It is shown that the one-period growth rate of the usual level trend is not a satisfactory measure. I will investigate three questions: How do we measure the trend growth rate historically? What is the best current approximation to the historical growth rate? How can we give an error margin for the current estimate? The general considerations are illustrated by application to the growth rate of the British Consumer Price Index, which is the main target variable for the Bank of England's Monetary Policy Committee. In this context, the trend growth rate appears to be a useful additional interpretative tool.

### Adjustment of Data from Period Reporters in Estimates of Monthly Retail Trade

❖ Donald Martin, U.S. Census Bureau, Statistical Research Division, mail stop 9100, Washington, DC 20233, *donald.e.martin@census.gov*; David Findley, U.S. Census Bureau

***Key Words:*** period reporters, calendarization, benchmarking

In this paper, we discuss the adjustment of data for four- or five-week periods collected by the Service Sector Statistics Division (SSSD) of the U.S. Census Bureau for the purpose of obtaining monthly estimates of retail trade. Properties of the present adjustment procedure used by SSSD are compared with those of three others: one that arises by allowing a variable in a model for deriving the present adjustment factor to take an alternative value, the calendarization method of Cholette and Chhab (1991), and a generalization of the latter procedure. The methods are applied to real and simulated data.

# 97 Statistical Approaches to Assess Ethnic Differences: Impact on Drug Development ● ✪

ENAR
**Monday, August 7, 8:30 am–10:20 am**

### Design and Analysis for Showing the Similarity of Drug Efficacy between Two Clinical Studies

❖ Yoshiharu Horie, Nippon Boehringer Ingelheim Co., Ltd., Hyogo, Japan, *HORIEY@kaw.boehringer-ingelheim.com*; Hajime Uno, Kitasato University; Masahiro Takeuchi, Kitasato University

***Key Words:*** proportion of similar response, similarity, bridging study

One of the challenging issues regarding “bridging study,” defined in the ICH E5 guideline in new drug developments, is to evaluate the similarity of treatment effects between an original region and the other new region. The proportion of similar response (PSR) is a measure to quantify the overlap of two density functions of continuous random variables, which gives a direct interpretation of the similarity between those two random variables. Therefore, the PSR approach is appropriate for this purpose. We illustrate this applied to actual data.

### Evaluation of Treatment-Country Interaction in Global Clinical Trials

❖ Hajime Uno, Kitasato University, 60 Babcock Street, APT 22, BROOKLINE, MA 02446, *hajime@unosan.sent.com*; Yoshiharu Horie, Nippon Boehringer Ingelheim Co., Ltd.; Masahiro Takeuchi, Kitasato University

***Key Words:*** global clinical trial, prediction model

When making inferences for the treatment effect as a global measure in a global clinical trial, it is important to evaluate the presence of the treatment-country interaction. A popular way to do this is through a significance test; however, this may not be appropriate. We propose an approach to evaluate the treatment-country interaction in terms of prediction accuracy measures for future patients’ outcome. Applying a large sample theory for prediction accuracy measures, we report point and interval estimates of those measures to see how much the treatment-country interaction can improve the prediction of the future patients’ outcome in participating countries. We illustrate this approach using real data from a global clinical trial.

### Minority Differences in Cancer Survival on Cooperative Group Clinical Trials

❖ Beow Yeap, Harvard Medical School/Massachusetts General Hospital, 50 Staniford Street, Suite 560, Boston, MA 02114, *byeap@partners.org*; Marvin Zelen, Harvard School of Public Health

***Key Words:*** African-American, race, matching, gender

Outcome comparisons between racial groups generally lack the data to adjust for patient variation in disease, therapy, and institution factors. The Eastern Cooperative Oncology Group (ECOG) conducts multicenter clinical trials in adult cancers nationwide. From 1976--1995, the ECOG accrual of about 74,000 patients on therapeutic protocols was 9% African American (AA). To minimize patient variation without explicit covariate modeling, each AA proband was matched by gender to a white patient enrolled in the same protocol, assigned to the same arm, and treated at the same hospital. An overall 10% higher death rate among AA patients is distributed unevenly between genders with more of the burden on AA women. In latter years, the excess mortality has diminished so that AA patients appear to have similar benefits as whites. Follow up and enrollment data will be updated.

### Delivering Robust Outcomes from Multiregional Clinical Trials

❖ Yoko Tanaka, Eli Lilly and Company, Lilly Corporate Center, DC 6166, Indianpolis, IN 46285, *yokot@lilly.com*

***Key Words:*** heterogeneity, multi-regional, generalizability

Principles and strategies are considered for addressing heterogeneity due to geographic diversity in multiregional clinical trials. This diversity can be the result of differences in race, culture, standard of care, and/or quality of data across regions---each of which will be discussed. In light of these various sources of variability, we will consider implications of global data to specific regions and the generalizability of results. An important focus throughout will be the question of what constitutes an internally consistent global trial.

### Statistical and Related Issues of Global Drug Development: Experience in the PMDA Consultation Meetings

❖ Yuki Ando, Pharmaceuticals and Medical Devices Agency, Shin Kasumigaseki Building, 3 3 2 Kasumigaseki Chiyodaku, Tokyo, 100-0013 Japan, *ando-yuki@pmda.go.jp*

***Key Words:*** global drug development, multinational clinical trial

After the ICH-E5 guidance, “Guideline for Ethnic Factors in the Acceptability of Foreign Clinical Data,” was issued, efficiency in Japanese new drug development was improved with utilization of clinical study data from other regions. Recently, Japanese pharmaceutical companies have planned to participate in multinational clinical trials for global simultaneous drug development. In this paper, statistical and related issues that arise in the Pharmaceuticals and Medical Devices Agency consultation meetings on multinational clinical trials will be discussed.

# 98 Prediction and Detection in Defense and Homeland Security Applications ✪

Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences
**Monday, August 7, 8:30 am–10:20 am**

### Geospatial Modeling in an Information Theoretic Framework as Applied to Forecasting of Insurgent Activity

❖ Jason Dalton, Spatial Data Analytics Corporation, 1950 Old Gallows Road, Suite 300, Vienna, VA 22182, *jdalton@spadac.com*

***Key Words:*** spatial, forecast, mutual information, area reduction, resource allocation

Geospatial modeling for defense and intelligence amounts to providing an ordered set of location elements where the order conveys a difference in likelihood for the occurrence of a particular event of interest. This presentation will describe one approach to geospatial modeling that accounts for redundant and irrelevant variables through a mutual information measure applied to pair-wise comparisons of features. By using distribution divergence measures, derived spatial features can be removed when they do not contribute positively to the resulting forecast. A chief innovation of this work is its use of a credit/blame assignment to variables to adjust weights over subsequent assessments. This work builds on prior work by developing a utility model for target decision tradeoffs. Applications in IED emplacement, high-value individual tracking, and buried targets will be discussed.

### Signal Detection in Radiation Portal Monitoring Data

❖ Tom Burr, Los Alamos National Laboratory, Mail Stop F600, Los Alamos, NM 87545, *tburr@lanl.gov*; Jim Gattiker, Los Alamos National Laboratory; George Tompkins, Los Alamos National Laboratory

***Key Words:*** radiation, monitors, signal, detection, background, suppression

Data from passive radiation portal monitors (RPMs) have been collected since 2002 to detect potentially harmful radioactive cargo (i.e., special nuclear material (SNM)). Detection of illicit SNM using RPMs is complicated by several factors, including the possibility of shielding the emitted radiation, drifting background and/or sensor response, variable-length vehicle scan time, cross-talk between neighboring lanes leading to either suppression or elevation of respective lanes, and vehicle shadow-shielding, which implies that vehicles with or without radioactive material will suppress the natural background. This talk will describe nuisance and statistical alarms, background suppression, energy windowing, alarm rules, and an injection study that considers the effect of five factors on SNM detection in a full, five-factor factorial experiment.

### Cokriging with Generalized Cross-Covariances for Detecting Radioactivity

❖ Chunfeng Huang, The Ohio State University, Cockins Hall, Room 404, 1958 Neil Avenue, Columbus, OH 43210-1247, *chuang@stat.ohio-state.edu*; Noel Cressie, The Ohio State University; Yonggang Yao, The Ohio State University; Tailen Hsing, The Ohio State University

***Key Words:*** cokriging, generalized cross covariance, spatial, radioactivity, sensors

Radioactivity detection and prediction are important for assessing the effect of a so-called 'dirty bomb' on people and their environment. Suppose there are two types of sensors deployed to augment each other---one being costly but providing high-quality information at a small number of key locations and the other being inexpensive with many sensors, low-quality information, and covering a larger area. If both types of observations are intrinsic random processes, we can introduce a generalized cross-covariance to describe their spatial cross-dependencies. A nonparametric method is proposed for estimation of all spatial dependencies. We can then use cokriging to predict radioactivity at unsampled locations based on both types of data.

### Monitoring Safety of Food Supply by Analyzing Consumer Complaints

❖ Artur Dubrawski, Carnegie Mellon University, 5000 Forbes Ave., NSH 3121, Pittsburgh, PA 15213, *awd@cs.cmu.edu*; Maheshkumar Sabhnani, Carnegie Mellon University; Andrew Moore, Carnegie Mellon University

***Key Words:*** food safety, biosurveillance, Bayesian methods, causal models, machine learning, biosurveillance, Bayesian methods, causal models, machine learning

We present a new approach, called TipMon, to monitoring streams of multivariate, event-based data such as customer complaints, public safety hotline tips, or individual patient health events. Its objective is to identify groups of independently collected records that may originate from common, underlying causes. It has been applied successfully to screen consumer food complaints for emerging patterns of adverse events of natural and intentional (criminal or terrorist) origins. We use that example to explain theoretical background of our approach and its implementation details. We also discuss its performance observed so far. We believe early success in food safety domain indicates a wider applicability of TipMon. It possesses a unique ability to remain sensitive to small signals in multivariate, heterogeneous data---even when it is spotty, noisy, and in short supply.

### Logistic Joinpoint Models with Applications in Criminal Processes

❖ Ryan Gill, University of Louisville, 328 Natural Sciences Building, Louisville, KY 40292, *rsgill01@louisville.edu*

***Key Words:*** logistic regression, change point models, occurrence data

We consider a logistic joinpoint model for estimating the probability of crime occurrence based on associated features. This includes examination of backward and forward algorithms for selecting the number of joinpoints. We apply these methods to the detection of joinpoints in criminal processes.

## 99 Competing Risk Events in Cancer Epidemiology ●

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Competing Risks Models in the Monogenic Cancer Susceptibility Syndromes

❖ Philip S. Rosenberg, National Cancer Institute, 6120 Executive Blvd., EPS 8022, Rockville, MD 20852, *rosenbep@mail.nih.gov*

***Key Words:*** competing risks, spline functions, cancer, cause-specific hazards, genetic epidemiology

The monogenetic cancer susceptibility syndromes are complex diseases with multiple competing adverse events. Competing risks theory provides a framework to model the natural history. We have developed models for Fanconi anemia---associated with bone marrow failure, acute myeloid leukemia (AML), and solid tumors; severe congenital neutropenia---associated with sepsis death and AML; and hereditary breast and ovary cancer. We model the absolute cause-specific hazard functions nonparametrically using splines. It is of interest to identify covariates that modulate the hazard of various event types and to gauge the clinical impact of these covariates on the entire disease. Covariates can be modeled by stratification or by combining spline estimators of the baseline hazard functions with proportional hazards models. These models can account for genotype/phenotype associations with outcomes.

### Analyzing Survival Data with Competing Risks

❖ Leah Ben-Porat, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street 3rd Floor, New York, NY 10021, *benporal@mskcc.org*; Jaya M. Satagopan, Memorial Sloan-Kettering Cancer Center

***Key Words:*** cumulative incidence, informative censoring, survival probability, Kaplan-Meier estimate

In survival analysis, a nonparametric estimate of the survival probability can be obtained using the Kaplan-Meier method. However, a patient may experience an event other than the one of interest (i.e., competing risk event) that precludes the onset of the outcome of interest. In this setting, the cumulative incidence of the outcome of interest must be calculated by appropriately adjusting for the competing risk events. This is because a patient experiencing a competing risk event is censored in an informative manner so the Kaplan-Meier estimation procedure may not be directly applicable. The importance of such adjustments will be illustrated using published datasets. The analysis approach and the interpretation of the results will be demonstrated using the cmprisk library in R.

### Bayesian Competing Risks Analysis of Cancer Survival Data from the SEER Program

❖ Sanjib Basu, Northern Illinois University, Division of Statistics, DU 366, De Kalb, IL 60115, *basu@niu.edu*; Ram Tiwari, National Institutes of Health

***Key Words:*** Bayes factor, cause-specific hazard, mixture cure, model selection

The SEER Program of the National Cancer Institute collects and publishes cancer incidence and survival data from 14 population-based cancer registries and three supplemental registries covering approximately 26% of the U.S. population. Competing risks analysis of such cancer survival data considers the simultaneous risks of cancer and other causes and includes relative survival as well as cause-specific survival approaches. Cure rate models, on the other hand, postulate a fraction of the patients to be cured from cancer. We propose a model that incorporates competing risks and, at the same time, allows a fraction of patients to be cured. We describe Bayesian analysis of this model, discuss both conceptual and methodological issues related to model building and selection, and consider application in survival data for breast cancer patients.

### Competing Risks Analysis in Breast Cancer with Missing Cause of Death

❖ Mousumi Banerjee, University of Michigan, Department of Biostatistics, 1420 Washington Heights SPH II, Ann Arbor, MI 48109, *mousumib@umich.edu*; Ananda Sen, University of Michigan

***Key Words:*** semiparametric Bayesian, survival analysis, competing risks, SEER

Analyzing survival data with competing risks has received considerable attention in the statistical literature. Earlier work in this area focused on the scenario when the cause of death is known. A challenging twist to this problem arises when the cause of death is not known exactly, but can be narrowed down to a set of potential causes that do not necessarily act independently. In this talk, we will present an overview of issues and existing methodologies to handle such data. We propose a semiparametric Bayesian model that attempts to address some of the methodological challenges. The proposed methodology is illustrated using survival data on breast cancer patients from the SEER registry.

### Development of Cohort Life Tables for "Other Causes" for Use in Simulation Modeling

❖ Marjorie Rosenberg, University of Wisconsin-Madison, 975 University Ave., Madison, WI 53706, *mrosenberg@bus.wisc.edu*

***Key Words:*** competing risks, simulation, actuarial methods, life tables, natural history models

The development of new cancer treatments or interventions is aimed at reducing cancer mortality. Simulation modeling is one approach to quantifying the impact of new treatments or interventions at the population-level. The all-cause mortality rate is partitioned into two components: the rate of mortality for the specific type of cancer and the rate of mortality for the all-cause mortality rate with that type of cancer removed. We developed actuarial life tables that do not specify the shape of the mortality curve and are directly associated with all of the data for other cause mortality for four types of cancer: breast, colorectal, lung, and prostate for use by the Cancer Intervention and Surveillance Modeling Network, NCI-sponsored investigators using modeling techniques to measure the impact of cancer control interventions on population trends in incidence and mortality.

## 100 Bayesian Models in Finance ●

Section on Bayesian Statistical Science

**Monday, August 7, 8:30 am–10:20 am**

### Bayesian Forecasting of Prepayment Rates for Individual Pools of Mortgages

❖ Ivilina Popova, Seattle University, Department of Finance, Albers School of Business and Economics, Seattle, WA 98122, *popovai@seattleu.edu*; Elmira Popova, The University of Texas at Austin; Edward I. George, University of Pennsylvania

***Key Words:*** finance, prepayment, mixture, Bayesian

This paper proposes a novel approach for modeling prepayment rates of individual pools of mortgages. The model incorporates the empirical evidence that prepayment is past-dependent via Bayesian methodology. There are many factors that influence the prepayment behavior, and, for many, there is no available information. We implement this issue by creating a Bayesian mixture model and constructing a Markov chain Monte Carlo algorithm to estimate the parameters. We assess the model on a dataset from the Bloomberg Database. Our results show the burnout effect is a significant variable for explaining normal prepayment activities. This result does not hold when prepayment is triggered by nonpool-dependent events.

### Bayesian Inference for Derivative Prices

❖ Jonathan Stroud, University of Pennsylvania, The Wharton School, Philadelphia, PA 19104, *stroud@wharton.upenn.edu*; Nicholas Polson, The University of Chicago

***Key Words:*** stochastic volatility, jump diffusion, optimal filtering, particle filtering, option pricing, stock returns

This paper develops a methodology for parameter and state variable inference using both asset and derivative price information. We combine theoretical pricing models and asset dynamics to generate a joint posterior for parameters and state variables and provide an MCMC strategy for inference. There are several advantages of our inferential approach. First, more precise parameter estimates are obtained when both asset and derivative price information are used. Second, we provide a diag-

nostic tool for model misspecification based on agreement of the state and parameter estimates with and without derivative price information. Furthermore, the time series properties of the state variables also can be used to evaluate model fit. We illustrate our methodology using daily equity index options on the S&P 500 index from 1998--2002.

### Reliability and Survival in Financial Risk

❖ Nozer Singpurwalla, The George Washington University, Department of Statistics, 2140 Pennsylvania Avenue, Washington, DC 20052, *nozer@gwu.edu*

***Key Words:*** bonds, nonparametric Bayes, Dirichlet processes, monotone interest rates, stochastic interest rates

In this talk, we will point out connections between the asset pricing formula of fixed income investments and the exponentiation formula of reliability and survival. Once this connection is made, results from one discipline can be imported to the other and vice versa. Several results in reliability and survival have a Bayesian angle, and the relevance of such results in mathematical finance can be articulated.

### A Computational Approach to Bayesian Portfolio Selection

❖ Refik Soyer, The George Washington University, Funger Hall 415 Decision Sciences, School of Business, Washington, DC 20052, *soyer@gwu.edu*

***Key Words:*** decision analysis, stochastic volatility, MCMC

We consider multiperiod portfolio selection problems for a decision-maker with a specified utility function when the variance of security returns is described by a discrete time stochastic model. We will present the Bayesian decision theoretic setup for the problem. The solution involves a dynamic programming formulation and backward induction. We will present a simulation-based method to solve these problems, adopting an approach that replaces the preposterior analysis by a surface fitting--based optimization approach. We will provide examples to illustrate the implementation of our approach.

## 101 Inference for Dynamic Graphical Models

Section on Bayesian Statistical Science

**Monday, August 7, 8:30 am–10:20 am**

### Stochastic ARMA Models

❖ Bo Thiesson, Microsoft Research, One Microsoft Way, Redmond, WA 98052, *thiesson@microsoft.com*; Jesper Lind, Microsoft Research; David M. Chickering, Microsoft Research; David Heckerman, Microsoft Research; Christopher Meek, Microsoft Research

***Key Words:*** time series, graphical models, ARMA, Bayes net, incomplete data, SQL server

The class of stochastic ARMA models extends the classic ARMA time-series models by replacing (or smoothing) the deterministic relationship between target and regressors in these models with a conditional Gaussian distribution having a small controllable variance. As this variance approaches zero, a stochastic ARMA model approaches a classic ARMA model. We represent a stochastic ARMA model as a directed graphical model. In doing so, we benefit from the ability to apply standard graphical-model-inference algorithms during parameter estimation (including estimation in the presence of missing data), model selection, and prediction. The graphical model representation also offers a visual representation that is easy to interpret. We demonstrate how the graphical representation in this way lends itself as a conceptually easy way to handle cross-predicting time series, periodicity, and trends.

### Strategies for Online Inference with Dynamic Graphs

❖ Makram Talih, City University of New York-Hunter College, 695 Park Ave., Room 905 HE, Department of Mathematics and Statistics, New York, NY 10021, *makram.talih@hunter.cuny.edu*

***Key Words:*** dynamic graphs, graphical models, sequential Monte Carlo, posterior inference, hidden Markov model, particle filters

We will present strategies for online learning of dynamic graphs via Sequential Monte Carlo (SMC). SMC algorithms are based on maintaining in parameter space an ensemble of particles, each of which tracks a particular realization of the process under study. In our framework, the process is governed by the posterior distribution of the parameters (e.g., the precision matrix) and hidden variables (e.g., the underlying undirected graph), given the data sequence. This talk will further explore the use of geometric methods for designing an efficient importance distribution.

### Network-Based Marketing

❖ Shawndra Hill, New York University, Leonard N. Stern School of Business, 44 W 4th Street, New York, NY 10012, *shill@stern.nyu.edu*; Chris Volinsky, AT&T Labs-Research; Foster Provost, New York University

***Key Words:*** social networks, dynamic networks, marketing

We investigate the hypothesis that those consumers who have communicated with a customer of a particular service have increased likelihood of adopting the service. We survey the diverse literature on such "viral marketing," providing a categorization of the specific research questions asked, the data analyzed, and the statistical methods used. We highlight a striking gap in the literature: No prior study has had both of the key types of data necessary to provide direct support for the hypothesis---data on communications between consumers and data on product adoption. We suggest a type of service for which both types of data are available: telecommunications services. Then, for a particular telecommunication service, we show support for the hypothesis.

### Bayesian Analysis of Longitudinal Binary Data Using Markov Regression Models with Skewed Links

Seongho Song, University of Cincinnati; ❖ Younshik Chung, Pusan National University, 30 JangJeon-dong, Geumjeong-gu, Dept. of Statistics, Pusan, 609-735 Republic of Korea, *yschung@pusan.ac.kr*; Dipak Dey, University of Connecticut; Alaattin Erkanli, Duke University Medical Center

***Key Words:*** Bayesian Markov regression model, correlated Bernoulli process, skewed link, latent variables, reversible jumps MCMC

In this research, we consider nonhomogeneous Markov regression models of unknown order as a means to assess the duration of autoregressive dependence in longitudinal binary data. We describe a subject's transition probability evolving over time using logistic regression models for his/her past outcomes and covariates. Our main goal is to develop the appropriate probability model for the correlated Bernoulli process in the presence of covariate information. In this model, we consider the skewed links for the link function given by Chen, Dey, and Shao (1999). Then, the model parameters order of transitions are estimated using reversible jump Markov chain Monte Carlo (MCMC) approach (Green 1995; Green and Richardson 1997).

## 102 Finite Population Correction Factors ●

Section on Survey Research Methods

**Monday, August 7, 8:30 am–10:20 am**

### Finite Population Correction Factors

❖ Wayne Fuller, Iowa State University, 202B Snedecor Hall, Ames, IA 50011-1210, *waf@iastate.edu*; ❖ Keith Rust, Westat/ University of Maryland, 1650 Research Blvd., Rockville, MD 20850, *KeithRust@westat.com*; ❖ Phillip S. Kott, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, VA 20230-1504, *pkott@nass.usda.gov*; ❖ Barry I. Graubard, National Cancer Institute, Biostatistics Branch, 6120 Executive Blvd, Rm 8024, Bethesda, MD 20892, *graubarb@mail.nih.gov*; ❖ S. Lynne Stokes, Southern Methodist University, 3225 Daniel Ave., Department of Statistics, Dallas, TX 75275, *slstokes@mail.smu.edu*

***Key Words:*** variance, sample size, super population, complex sample designs

It is common practice to use finite population correction factors (fpc) in estimating variances when sampling from a finite population. Various approximate fpcs are used with more complex designs sometimes. When the interest is in a wider population than the specific finite sampling frame, many argue that it suffices to drop the fpc from the variance estimate, but others maintain this is appropriate only in a limited number of contexts.

## 103 Advances in Variable Selection ●

Biometrics Section

**Monday, August 7, 8:30 am–10:20 am**

### Bayesian Variable Selection in Cox Models

❖ Naijun Sha, The University of Texas at El Paso, 500 W. University Ave., Bell Hall 203, El Paso, TX 79968, *naijun@math.utep.edu*; Mahlet G. Tadesse, University of Pennsylvania; Marina Vannucci, Texas A&M University

***Key Words:*** Bayesian variable selection, Cox model, survival analysis, censored time, MCMC

In this paper, we investigate variable selection methods for Cox's proportional hazard model. We develop selection methods that allow for censored data. Our methods lead to simultaneously estimates of the survival function as well as to the identification of the factors that affect the survival outcome. We handle the problem of selecting a few predictors among the prohibitively vast number of variables through the introduction of a binary exclusion/inclusion latent vector. This vector is updated via an MCMC technique to identify promising models. We describe strategies for prosterior inference and explore the performance of the methodology with simulated and real datasets.

### FSR Methods in the Cox Proportional Hazards Model

❖ Yun Chen, North Carolina State University, 1911 Wolf Tech Lane, Apt. 304, Raleigh, NC 27603, *ychen8@ncsu.edu*; Dennis A. Boos, North Carolina State University; Leonard A. Stefanski, North Carolina State University

***Key Words:*** false selection rate (FSR), SCAD, LASSO, the Cox model, model selection

We study the performance of the False Selection Rate (FSR) method under the Cox proportional hazards model. The FSR method introduced by Wu, Boos and Stefanski (2004) controls the proportion of uninformative variables in a regression model by adding a number of pseudo variables to the original data set and monitoring the proportion of pseudo variables selected. For the Cox model, the FSR method with forward selection is compared to SCAD (Fan and Li, 2001, 2002), LASSO (Tibshirani, 1996, 1997), and forward selection with AIC and BIC. In addition, we present a new approach to estimating the tuning parameters of SCAD and LASSO based on the FSR.

### Variable Selection in Linear Mixed Models for Longitudinal Data

❖ Lan Lan, North Carolina State University, 4806 Rockport Drive, Department of Statistics, Durham, NC 27703, *llan@stat.ncsu.edu*; Daowen Zhang, sanofi-aventis; Hao Zhang, North Carolina State University

***Key Words:*** oracle property, REML, SCAD, variance components

Fan and Li (JASA 2001) proposed a family of variable selection procedures for certain parametric models via a nonconcave penalized likelihood approach, where significant variable selection and parameter estimation were done simultaneously, and the procedures were shown to have the oracle property. In this presentation, we extend the nonconcave penalized likelihood approach to linear mixed models for longitudinal data. Two new approaches are proposed to select significant covariates and estimate fixed effect parameters and variance components. In particular, we show the new approaches also possess the oracle property when the tuning parameter is chosen appropriately. We assess the performance of the proposed approaches via simulation and apply the procedures to data from the Multicenter AIDS Cohort Study.

### Variable Selection with Penalized GEE

❖ John Dziak, The Pennsylvania State University, 315 W. Beaver Ave., Apartment 2, State College, PA 16801, *dziakj1@alumni.cua.edu*

***Key Words:*** LASSO, SCAD, GEE, longitudinal, BIC

Despite the importance of variable selection criteria in linear modeling, only recently has there been work on generalizing these criteria for longitudinal and cluster-correlated data. I survey the existing literature on extending the Cp, AIC, and LASSO criteria to marginal longitudinal models with generalized estimating equations (Fu 2003 and Cantoni 2005), and propose new extensions of the BIC and SCAD (Fan and Li 2001) criteria based on penalizing a generalized least-squares loss function. The latter criteria led to a sparser solution. Theoretical and empirical results on the performance of resulting estimators will be explored.

### Adaptive-LASSO for Cox's Proportional Hazards Model

❖ Wenbin Lu, North Carolina State University, 210E Patterson Hall, 2501 Founders Drive, Raleigh, NC 27695, *lu@stat.ncsu.edu*; Hao Zhang, North Carolina State University

***Key Words:*** adaptive LASSO (ALASSO), LASSO, penalized partial likelihood, proportional hazards model, variable selection

We investigate the variable selection problem for Cox's proportional hazards model and propose a unified model selection and estimation procedure with desired theoretical properties and computational convenience. The new method is based on a penalized log partial likelihood with the adaptively weighted L_1 penalty on regression coefficients and is named adaptive-LASSO (ALASSO) estimator. Instead of applying the same penalty to all the coefficients as other shrinkage methods, the ALASSO advocates different penalties for different coefficients: Unimportant variables receive larger penalties than important variables. In this way, important variables can be protectively preserved in the model selection process, while unimportant ones are shrunk more toward zero and thus more likely to be dropped from the model. We study the consistency and oracle properties of the proposed estimator.

### The LASSO Method for Variable Selection for Right-Censored Data

❖ Lili Yu, The Ohio State University, Trumbull Court, Columbus, 43210, *yu.246@osu.edu*; Dennis K. Pearl, The Ohio State University

***Key Words:*** sieve likelihood, LASSO, model selection, right censored data

Tibshirani proposed a variation of the "lasso" method that was to minimize the log partial likelihood subject to the sum of the absolute values of the parameters being bounded by a constant in Cox's proportional hazards model. Due to the nature of this constraint, it shrinks coefficients and produces coefficients that are exactly zero. We apply this method to a class of semiparametric models (linear transformation models) in which the response variable is right-censored and the error is symmetric at zero but its distribution is unknown. We propose to use sieve-likelihood method to calculate the log likelihood and the parameters simultaneously. Simulations indicate using sieve-likelihood to calculate the lasso criteria in this setting can pick approximately the correct number of zero coefficients.

## 104 Proteomics ●

Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Design and Analysis of Experiments in Proteomics

❖ John Aleong, University of Vermont, Department of Mathematics and Statistics, 24 Hills, Burlington, VT 05405, *jaleong@uvm.edu*

***Key Words:*** protein expression, data analysis, clinical trials,

Proteomics examines the proteins in cells or tissues. An organism's proteome catalogs all the proteins expressed through the life of the organism under all conditions. The proteomics challenges are enormous, protein diversity, environmental variability, and various technologies. Comparative experiments of protein expression in normal and diseased tissues have enormous potential applications in medicine. Following an introduction of proteomics, we will review some efficient statistical designs and analyses of experiments, which will lead to more efficient clinical trials and potentially new treatments.

### Logistic and Probit Regression Modeling of Proteomic Mass Spectra in a Case Control Study on Diagnosis for Colon Cancer

❖ Bart Mertens, Leiden University Medical Center, Department of Medical Statistics, LUMC Postal Zone S5P PO Box 9600, Leiden, 2300 RC The Netherlands, *b.mertens@lumc.nl*

***Key Words:*** logistic regression, probit regression, Bayesian analysis, mass spectrometry, proteomic diagnosis, birth-death process

We adapt logistic and probit regression models for the evaluation of diagnostic potential of mass spectroscopic data in proteomics case control studies. Instead of a direct attempt to model the observed case-control data as regression on peaks, we parameterize the predictor as a linear combination of Gaussian basis functions along the mass/charge axis. The location of these basis functions is treated as a random variable and must be estimated from the data. A fully Bayesian implementation is pursued, which treats the number of functional components as a random variable. Calculations are implemented through birth-death process modeling. We evaluate the models on data from a randomized blocked case-control designed experiment, which was carried out recently at Leiden University (LUMC). The experiment compares spectra of serum samples of 63 colon cancer patients with 50 controls.

### Preprocessing Method and Nested Cross-Validation Classification of Lung Cancer Using Mass Spectrometry Proteomics

❖ Jingjing Ye, University of California, Davis, 1535 Westgate Ave., Apt 1, Los Angeles, CA 90025, *jye@ucdavis.edu*

***Key Words:*** proteomics, mass spectrometry, baseline correction, nearest-neighbor classification

Mass spectrometry is used widely to determine protein functions in living organs. Extracting useful information is a key point. There is no standard for measuring which extraction methodology gives the best result. We proposed a methodology to preprocess the spectrum data using misclassification rate as a rule to evaluate and compare preprocessing methods. In this paper, an objective function of baseline correction is proposed, subject to two constraints. After necessary adjustments of the data, our approach uses a nested nearest-neighbor classification scheme, applying three commonly used distance measures: correlation between patients, Euclidean metric, and Canberra metric. Our finding shows our preprocessing method yields a plausible rate of misclassification rates and a better result than previous studies.

### Statistical Approaches to Discovery of Biomarkers for Early Detection of Cancer Using LC-MS/MS

❖ Xiaochun Li, Dana-Farber Cancer Institute, Department of Biostatistics and Computational Biology, 44 Binney St, Boston, MA 02115, *xiaochun@jimmy.harvard.edu*; Meredith A. Goldwasser, Dana-Farber Cancer Institute

***Key Words:*** proteomics, LC-MS/MS, normalization, quality control, biomarker selection

With the advent of high-throughput proteomics, technologies for the measurement of large numbers of proteins or peptides have become available to biologists to study biological functions in cells directly. Identification of unique protein signatures would enable the development of new tools for proteomics-based, noninvasive screening of cancers. We will focus on the following issues related to the technology: quality control, normalization, methods of putative marker selection, and evaluation of those markers. As some of these issues are common to mass spectrometry data, our approach is generalizable to other MS platforms. We will illustrate our approach with an example from endometrial cancer, which uses cervical liquid PAP smear samples from women with and without endometrial cancer to identify unique changes in proteins/peptides.

### Statistical Methods for Protein Interactions Predictions

❖ Inyoung Kim, Yale University, Department of Epidemiology and Public health, School of Medicine 60 College Street, New Haven, CT 06520, *inyoung.kim@yale.edu*; Yin Liu, Yale University; Hongyu Zhao, Yale University

***Key Words:*** Bayesian method, domain-domain interactions, EM algorithm, protein-protein interactions

As protein domains are the functional units of proteins and protein-protein interactions are achieved through domain-domain interactions, the modeling and analysis of protein interactions level may be more informative and insightful at the domain level. The fact that protein domains are likely evolutionarily conserved allows us to pool information from data across multiple organisms for the inference of domain-domain and protein-protein interaction probabilities. However, the number of parameters to be estimated is very large and the amount of information for statistical inference is quite limited. We propose a full Bayesian method and a semi-Bayesian method for estimating domain-domain interaction probabilities through integrating large-scale protein interaction data from three organisms. We compare our methods with likelihood based approach (Deng et al., 2002 and Liu et al., 2005).

### Multi-Dimensional NMR Spectra Identification for Protein Structure Determination

❖ Nicoleta Serban, Georgia Institute of Technology, 755 Ferst Drive, School of ISyE, Atlanta, GA 30332-0205, *nserban@isye.gatech.edu*

***Key Words:*** NMR, protein structure, mixture regression model, wavelet decomposition, backfitting, mixture detection

Determining the three-dimensional structures for large proteins using multidimensional Nuclear Magnetic Resonance (NMR) poses a formidable undertaking because of systematic noise, local correlated noise, and a large number of protons that resonate at similar frequencies. The primary objective of the research presented in this talk is to develop a statistical technique for identification and characterization of multi-dimensional NMR spectra. Our statistical method takes a novel overall perspective: (1) It incorporates a preliminary step for separating the signal from the background using a method that adapts for sharp changes in the data and non-homogeneous signal; (2) The locations, widths and amplitudes of the NMR spectra are estimated using a computational efficient algorithm; (3) It detects mixtures of spectra to solve ambiguities due to protons with similar resonance frequencies.

### Inferring Protein Associations Using Protein Pull-Down Assays

❖ Julia Sharp, Montana State University, Department of Mathematical Sciences, PO Box 172400, Bozeman, MT 59718, *jsharp@math.montana.edu*; Kevin K. Anderson, Pacific Northwest National Laboratory; Don S. Daly, Pacific Northwest National Laboratory; Deanna L. Auberry, Pacific Northwest National Laboratory; John Borkowski, Montana State University; William R. Cannon, Pacific Northwest National Laboratory

***Key Words:*** protein association, likelihood ratio test, Monte Carlo simulation

One method to reveal protein-protein interactions is a “bait-prey pull-down” assay using a protein affinity agent and a protein detection method that is prone to various errors. A pull-down study generates a presence/absence matrix wherein each column or row corresponds to a bait protein sample or prey protein, respectively. Our method evaluates the presence/absence pattern across a row with a Likelihood Ratio Test (LRT) and simulated LRT test statistic distributions, checking the statistical assumptions with simulated binomial random variates. Each prey protein is assigned a category (specific, non-specific, or not associated) that is then appraised with respect to a pull-down study’s goal and design. The LRT p-values objectively reveal specific and ubiquitous prey, as well as potential systematic errors. The LRT screen reduces the bias introduced when applying ad hoc filtering rules.

## 105 Salient Variables for Select Research Populations

Social Statistics Section

**Monday, August 7, 8:30 am–10:20 am**

### Is There Evidence of Racial Bias for the Federal Death Sentence?

❖ Matthias Schonlau, RAND Corporation, 201 N. Craig Street, Pittsburgh, PA 15213, *matt@rand.org*

***Key Words:*** capital punishment, causality, prediction

In addition to the death penalty laws in various states, there is also a federal death penalty. Most death sentences are handed down under state law; there have been few studies of the federal death penalty. Key concepts in this line of research are “capriciousness” and “arbitrariness.” “Capriciousness” means (roughly) that the sentencing decisions occur at random and cannot be predicted from the characteristics of the legal case. “Arbitrariness” means that the sentencing decisions are influenced by a characteristic other than the severity of the case. Such a characteristic might be, for example, race of the victim. I will give results from a study that analyzed data from 1988 to 2000.

### A Measure of Intergroup Discrimination Other Than the Difference between Median Wage Incomes

❖ John Angle, Economic Research Service, 1800 M Street, NW, Room N4097, Washington, DC 20036, *jangle@ers.usda.gov*

***Key Words:*** color, discrimination, gamma PDF, gender, minority group, wage income

The effect of intergroup discrimination on minority and majority wage incomes is conventionally measured by the difference between majority and minority group wage medians. This paper shows that majority and minority wage income distributions conditioned on education and age are approximately the same distribution with a different wage scale factor. In terms of a gamma pdf model, the shape parameters in each education-age partial distribution are approximately equal but the scale parameters of the minority and majority distribution are offset. This model accounts for several statistical features of the joint distribution of wage income to the minority and majority groups besides the difference in medians and provides a new interpretation of the effect of discrimination on wage incomes. Examples are given from nonmetro wage income data, 1961-2003.

## Missing Value Imputation and Weights Adjustment for Binary Variables

❖ Mingue Park, University of Nebraska-Lincoln, 340 Hardin Hall, N., Lincoln, NE 68583, *mpark2@unl.edu*; David Johnson, University of Nebraska-Lincoln

***Key Words:*** hot-deck imputation

Missing value imputation for dichotomous variables is considered. In estimating the proportion, two methods, imputation and weight adjustment, are compared. Based on the ratio of variance increase due to imputation to variance decrease due to sample size increase, a statistic that is helpful to decide the use of imputation is introduced. A variance estimator for an estimator of proportion constructed with imputed data is suggested.

## The Analysis of Seasonal Variations in Vital Statistics in Croatia: Two Approaches

❖ Ante Rozga, University of Split, Matice Hrvatske 31, Split, 021 Croatia, *rozga@efst.hr*; Zeljko Banovic, T-Com Croatia

***Key Words:*** vital statistics, X-12-ARIMA, TRAMO/SEATS

We have analyzed the following vital statistics in Croatia: live births, deaths, marriages and divorces. Seasonal and other variations were extracted from monthly time series. We have also made forecasts for five years ahead. It would be interesting because Croatian population has been decreasing for several years and also number of marriages has dramatically fallen, due to various reasons. Two different methods for seasonal adjustment were employed: X-12-ARIMA which is empirically based (“ad hoc method”) and TRAMO/SEATS which is model based method. The first one dominates official statistics, while the second one is being used by several statistical agencies. We investigated the results and statistical diagnostics relating to both methods and have found TRAMO/SEATS to be slightly in advantage

## Mixtures of Regressions

❖ Derek Young, The Pennsylvania State University, 150 Northbrook, Apt 115, State College, PA 16803, *dsy109@stat.psu.edu*

***Key Words:*** mixture model, regression

Finite mixture models are used when sampling from a population that consists of subpopulations with similar characteristics, but with at least one separate identifying mark. Here, a mixture of regressions model is considered to describe pairs of data (X, Y). For example, the mixture of regressions is used in modeling judgments for a child’s thinking strategy when given certain tasks (Thomas and Horton 1997). The goal of this talk is to introduce allowing the mixing weights of the model to depend on the predictor. This enables the investigator, given the predictor, to determine which of the K regression equations in the mixture model best predicts the value of the response.

## Joint Modeling of Quality of Life and Disease Progression

❖ Rebecca Hubbard, University of Washington, 4510 3rd Ave., NW, Seattle, WA 98107, *rhubb@u.washington.edu*; Lurdes Y. T. Inoue, University of Washington

***Key Words:*** quality of life, disease progression, continuous time Markov model

In the study of chronic disease, quality of life and disease progression may be monitored longitudinally to describe patients’ disease experience. Understanding the temporal course of the disease and its relationship with the individual’s well being is important in choosing treatment options that would, for example, slow disease progression while producing outcomes most satisfactory to the patient. We propose a model in which quality of life depends on patient’s current disease state which in turn evolutes according to a continuous time discrete state Markov model. We examine the properties of estimators for transition rates and parameters describing the quality of life/disease relationship based on this model. We also investigate extensions to the inhomogeneous Markov model for disease progression. We apply our model to data on patients undergoing hematopoietic stem cell transplant.

## Detecting Bias in Jury Selection

❖ Bruce Barrett, The University of Alabama, Box 870226, University Blvd., 300 Alston Hall, Tuscaloosa, AL 35487-0226, *bbarrett@cba.ua.edu*

***Key Words:*** jury selection, Lexian distribution, peremptory strikes, Batson challenge

Recent U.S. Supreme Court rulings have held that attorneys in criminal trials may not exercise peremptory strikes to systematically exclude prospective jurors on the basis of race or gender. The first step in establishing a charge of improper bias requires the challenging party to show evidence that his opponent’s strikes are inconsistent with random consideration of these characteristics. Frequently, court procedure dictates that there is some alternating between Prosecution and Defense in the striking process. As a result, choices for each side impact those of the other, and simply comparing the composition of jury pool with the peremptory strikes is insufficient for establishing any inference of bias. For these situations, we present a methodology for assessing the neutrality of juror strikes, based on the Lexian or Poisson binomial distribution.

# 106 Sample Survey Quality II ●

Section on Survey Research Methods

**Monday, August 7, 8:30 am–10:20 am**

## Analysis of Inconsistency in Coverage Estimates for Children in the 2000 Census

❖ Andrew Keller, U.S. Census Bureau, 705 Monroe Street, Apt 303, Rockville, MD 20850, *andrew.d.keller@census.gov*

Using dual system methodology, the Census Bureau conducted the Accuracy and Coverage Evaluation Revision II (A.C.E.) to produce improved estimates of the net coverage error of Census 2000. Dual system estimates were created from race/Hispanic origin, tenure, relationship, household size, completion time, residence area, return rate, and age/sex post-strata. Population estimates were also created using demographic analysis. The Census Bureau expected the two estimates to be consistent. However, the A.C.E. estimated that Census 2000 had a small net overcount (not significantly different from zero) of children 0 to 9 years while demographic analysis estimated a 2.56% net undercount. The demographic analysis estimate is believed to be accurate since it depends primarily on reliable recent birth registration data. This paper probes the apparent inconsistency between the estimates.

## Difficulties and Solutions for Surveying Refugees: Bosnian Refugees in St. Louis

❖ Kevin McIntyre, Saint Louis University, Research Methodology, St Louis, MO 63108, *mcintykp@slu.edu*; Hisako Matsuo, Saint Louis University; Terry Tomazic, Saint Louis University

*Key Words:* surveying refugees, difficulties, Bosnians, sampling frame, causal model

This paper addresses some of the difficulties of surveying refugees and the methods that the authors used to solve these difficulties. Using Bosnian refugees in St. Louis as a target population, the paper also elaborates on the methods used for conceptualization, operationalization, sampling, and data collection. Some of the major difficulties of surveying refugee populations are 1) lack of English competence, 2) cultural barriers in conceptualization, 3) uncertainty about validity of scales, 4) refugees' unfamiliarity with Likert-type scales, and 5) their transient residence. There are about 50,000 Bosnian refugees in the St. Louis area, creating a critical mass of this population. The authors used a grounded theory method for conceptualization, collecting qualitative data, then developed a causal model which was assumed to explain Bosnian refugees' adaptation to the US society.

## Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters

❖ Sylvia Dohrmann, Westat, 1650 Research Blvd., Rockville, MD 20850, *SylviaDohrmann@Westat.com*; Daifeng Han, Westat; Leyla Mohadjer, Westat

*Key Words:* area sample, listing, USPS, National Health and Nutrition Examination Survey

Traditionally, area sample frames of dwelling units are created by field staff visiting the sampled area and preparing a list of all addresses within its boundaries. To reduce the high cost of area listing, many researchers are evaluating the feasibility of alternative listing procedures. Most popular among the alternatives is using purchased lists of residential delivery addresses originating from the U.S. Postal Service. However, there are issues with getting information on special units such as dorms or other group quarters. In this paper we will discuss the comparability of the lists from different vendors, their cost and coverage compared to traditional listing, their coverage of group quarters, and practical aspects of using these lists as sampling frames for area samples.

## Employment in Nonprofit Entities: Coverage, Bias, and Measurement Errors in QCEW and Public IRS Information, 1999--2003

❖ Martin David, Urban Institute, 425 8th Street, NW, Apt 1144, Washington, DC 20004-2115, *David@ssc.wisc.edu*

*Key Words:* imputation, administrative records, nonprofit sector

Employment estimates combine microdata from the BLS Quarterly Census of Employment and Wages (QCEW) with IRS information on the universe of nonprofit organizations and Form 990 filed by 501(c)(3) charities, refining Salamon and Sokolowski (2005). Systematic errors in EINs correlate with number of employees and jurisdiction. Model-based weights for QCEW establishments matched to IRS records reduce undercount of employment. Undercount varies by NAICS three-digit. Matching also reveals nonreporting of employment on Form 990. Substitution of QCEW employment again reduces bias. State differences in employer obligation to participate in the Unemployment Compensation System exclude employment in many small organizations. The differences cause significant variation in extent of matching, implying additional nonprofit employment. Matched data reduce bias in employment estimates.

## Use of Administrative Data To Explore Effect of Establishment Nonresponse Adjustment on the National Compensation Survey Estimates

❖ Chester Ponikowski, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 3160, Washington, DC 20212, *Ponikowski_C@bls.gov*; Erin E. McNulty, Bureau of Labor Statistics

*Key Words:* non-response, bias, weighting cells

Non-response is a common but undesirable feature of a survey. It may lead to biases in survey estimates and an increase in survey sampling variance. Survey practitioners use various techniques to reduce bias due to non-response. The most common technique is to adjust sampling weights of responding units to account for non-responding units within a specified set of weighting classes or cells. In the National Compensation Survey (NCS) the weighting cells are formed using available auxiliary information: ownership, industry, and establishment employment size. In this paper, we explore how effective the formed cells are in reducing potential bias in the NCS estimates. We use administrative data to estimate average wages for responding units in the NCS. We generate and compare full sample wage estimates to estimates based on responding units with weights adjusted for non-responding units.

## Does a Final Coverage Check Reduce Census Coverage Errors?

❖ Elizabeth Martin, U.S. Census Bureau, 3715 FOB 3, Washington, DC 22308, *emartin@census.gov*; Don Dillman, Washington State University

*Key Words:* questionnaire design, within-household omissions, decennial census, split-panel experiment

In 2010, automated matching of name and date of birth for every census enumeration will better identify duplicates than in past censuses, but the problem of omissions persists. Ironically, eliminating duplicates might result in a larger net census undercount. This research evaluates new questions to reduce omissions. At the end of the short form, respondents are given reminders of people who might be missed, and asked to review the form and answer 2 questions about possible errors. The questions provide a clear stopping point in the census form, which helps avoid confusion. In a March 2006 national test, experimental and control forms are mailed to random samples of 7,100 households each. Returns are analyzed to determine whether (1) respondents answer the questions, (2) a final review reduces errors, and (3) the questions flag errors, as determined by a follow up interview.

## Nonresponse Bias Analysis in the 2003 National Survey of Recent College Graduates

❖ Aref Dajani, U.S. Census Bureau, 6101 Sebring Drive, Columbia, MD 21044, *aref.n.dajani@census.gov*; Jerry J. Maples, U.S. Census Bureau ; Ronald S. Fecso, National Science Foundation

*Key Words:* nonresponse bias, logistic regression, reweighting

The National Survey of Recent College Graduates recently experienced an increase in nonresponse rates. A nonresponse bias analysis was indicated to assess any impact on key estimates. Logistic modeling was used to explore possible relationships between early and late responders and variables not currently used in the creation of nonresponse adjustment cells. To analyze estimates by level of effort, late respondents and late ineligibles served as proxies for nonrespondents. This augmented set of nonrespondents was used to reweight the sample. Key estimates were compared. After reweighting the sample, there was no pattern of differences between weighted and reweighted estimates. Further, the readjusted counts had the same distribution as the original adjusted counts.

Future analysis should test whether recommended modifications will yield favorable results.

# 107 Unit Nonresponse in Surveys II

●

Section on Survey Research Methods
**Monday, August 7, 8:30 am–10:20 am**

## Modeling the Relationship between Cell Phone Usage and RDD Contact Effort

❖ Joseph Sakshaug, University of Michigan, P.O. Box 4214, Ann Arbor, MI 48106, *joesaks@umich.edu*

***Key Words:*** cell phone, landline, RDD

It has been speculated that persons in households with both landline and cellular phones may be harder to reach in traditional random-digit dial (RDD) surveys than are those with only landline phones (Tuckel and O'Neill, 2004). "Heavy users" of cell phones may be especially unlikely to answer their landline counterpart, thus leaving RDD call centers in an awkward situation of increasing contact attempts until a respondent is reached (or not). To get at this phenomenon we utilized respondent-level data from a 2003 supplement to the Surveys of Consumers at the University of Michigan. The supplement contained questions on a variety of cell phone usage behaviors. In combination with call record data we modeled the relationship between cell phone usage and RDD contact attempts in a bid to link cell phones to the level of effort to reach a respondent.

## Web versus Email Data Collection: Experience in the Current Employment Statistics Program

❖ Richard Rosen, Bureau of Labor Statistics, Washington, DC , *rosen.richard@bls.gov*; Antonio Gomes, Bureau of Labor Statistics; Louis Harrell, Bureau of Labor Statistics; Jason Chute, Bureau of Labor Statistics; Hong Yu, Bureau of Labor Statistics

***Key Words:*** internet reporting, survey response, mixed-mode collection

The Current Employment Statistics Program of the Bureau of Labor Statistics collects employment, payroll, and hours data from a sample of about 300,000 businesses each month. For the past 10 years CES has offered a Web reporting options. Currently about 5,000 units report via the Web. Due to the complexities of registration and ongoing security requirements for monthly login, response rates have been lower than expected. This paper reports on two new initiates; A streamlined Web system and E-mail. The simplified Web systems removed most of the registration process and account password verification currently used. Under the E-mail option, respondents are sent an E-mail with a replica of the CES form embedded in the E-mail. Respondents can then fill in the form and click the "submit" button. The paper will summarize and compare conversion rates and response rates for these two tests.

## Effects of Incentives in the U.S. Consumer Expenditures Quarterly Survey

❖ David McGrath, BAE Systems/BLS, 2 Massachusettes Ave., NE, Washington, DC 20212, *mcgrath.david@bls.gov*

***Key Words:*** response rates, incentives

Response rates to the Consumer Expenditures Quarterly Survey have fallen from about 86 percent in 1990 to 76 percent by 2004. To combat the falling rates, the Bureau of Labor Statistics introduced an incentives experiment beginning in November, 2005. The goal was to increase response rates by offering respondents an unconditional, pre-paid monetary incentive. To achieve this goal, we mailed debit-card incentives along with the survey's advance letter, prior to contacting the potential survey respondent. The experimental design contrasts a control group receiving $0 with groups that receive either $20 or $40 debit cards. In this paper, we show the design of the incentive experiment, discuss some field issues that arose during implementation, and provide preliminary results that focus on the effects of the incentives on response rates.

## Indirect Monetary Incentives with a Complex Agricultural Establishment Survey

❖ Daniel Beckler, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, VA 22030, *dan_beckler@nass.usda.gov*; Kathleen Ott, National Agricultural Statistics Service

***Key Words:*** response rate, financial incentives, agricultural survey

The USDA's National Agricultural Statistics Service conducts agricultural surveys. One of the most complex is the Agricultural Resource Management Survey Phase III, which collects detailed economic data, such as assets, expenses, income, debt, and operator characteristics. Part of this survey's sample uses a 16-page questionnaire with mail-out/mail-back data collection and face-to-face nonresponse follow up. Both prepaid and promised indirect monetary incentives were used in 2004 to increase mail response rates and reduce costly face-to-face follow-up interviews. Five treatment groups, including a control group, were used for the incentive experiment. Prepaid and promised indirect cash incentives---in the form of $20 ATM cards---and priority mail were used as stimuli. Response rates, ATM card usage, and costs for the treatment groups will be presented.

## A Study of Nonrespondents in the Canadian Vehicle Survey

❖ Martin Beaulieu, Statistics Canada, R.H. Coats Building, 17P, Tunneys Pasture, Ottawa, ON K1A 0T6 Canada, *martinj.beaulieu@statcan.ca*; Francois Gagnon, Statistics Canada

***Key Words:*** nonresponse bias, respondent follow-up, travel surveys, nonrespondents' characteristics

While many efforts are put into following up respondents, nonresponse remains an important issue in the Canadian Vehicle Survey (CVS). The low response rates observed may lead to biased estimates if the response mechanism is non-ignorable. The sponsors of the survey, Transport Canada and Natural Resources Canada are interested in knowing if a nonresponse bias exists and, if so, its direction and magnitude. A study of nonrespondents was conducted in early 2006 with three main objectives: (i) estimate the nonresponse bias, (ii) determine the reasons why nonrespondents did not complete the survey questionnaire in order to improve the data collection procedures, and finally (iii) collect nonrespondents' characteristics in order to improve the nonresponse treatment. This paper will describe the methodology of the CVS study of nonrespondents and then present the analysis of the results.

## The Impact of Questionnaire Length on Economic Census Return Rates

❖ Diane K. Willimack, U.S. Census Bureau, 5001 Regency Place, Alexandria, VA 22304, *diane.k.willimack@census.gov*

***Key Words:*** establishment surveys, response rates, respondent burden

A recurring question in survey methodology is the effect of survey length on response rates. For self-administered paper forms, a common measure of length is the number of pages. The typical hypothesis is that response rates fall as the number of pages increases. This paper examines the effect of the number of questionnaire pages on response rates in U.S. economic censuses. Tailored by industry, over 500 different questionnaire versions have historically ranged from 2 to 16 pages in length. Due to extensive formatting changes, 2002 Economic Census forms ranged from 3 to 27 pages in length. As a result, survey managers were concerned that response rates would decline substantially. Considering the form as our unit of analysis, we used multivariate methods to determine if the number of questionnaire pages affected response rates differently for 2002 than it did in previous censuses.

### Respondents Reasons for Participation in Telephone Surveys

❖ Nadra Garas, American University, 3701 Connecticut Ave., NW, 820, Washington, DC 20008, *ngaras@devassoc.com*; Johnny Blair, Abt Associates Inc.

***Key Words:*** nonresponse, participation, telephone survey, introduction, refusal, reasons for participation

Telephone survey nonresponse is a major survey quality concern. While it is important to learn about reasons for refusals, it is also useful to examine reasons respondents participate. In six general population omnibus phone surveys conducted at the University of Maryland Survey Research Center from 1990 to 2001, at the end of the interview, respondents were asked: "There are different reasons people agree to begin a survey. What was the main reason you agreed to begin this survey?" The responses were coded to tabulate reasons including sponsorship, interest in topic, interviewer persistence, and "just to be helpful." We examine reasons for participation by demographic groups, whether the interview was a converted refusal, and by changes in the distribution of reasons over time. We discuss implications for designing telephone survey introductions.

## 108 Genetic Interactions/Genetic Imprinting ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Case-Control Studies of Haplotype Environment Interactions with Genetic Misclassification

❖ Christine Spinka, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211, *spinkac@missouri.edu*; Raymond J. Carroll, Texas A&M University; Nilanjan Chatterjee, National Cancer Institute

***Key Words:*** genetic epidemiology, haplotypes, misclassification, pseudolikelihood

Epidemiologic studies investigating the interaction between haplotypes and environmental factors have become increasingly common in recent years. For rare diseases, these studies are often conducted using a case-control study design. In this setting we utilize a logistic regression model to estimate the probability of disease and include covariates for the effects of the haplotypes, environmental factors and their interaction. We develop a method to estimate these parameters in the presence of genetic misclassification, focusing on the case where some individuals have their genotype measured more than once. Our estimates are asymptotically normal and unbiased with known variance. We illustrate our method by simulation and on a data set with a high level of misclassification.

### Strategy for Analyzing Multifactorial Epidemiological Data Involving Host, Genetic, and Environmental Factors

❖ John Molitor, University of Southern California, , *jmolitor@usc.edu*

***Key Words:*** Bayesian statistics, epidemiology, genetic mapping, variable selection

Classical analysis methods that model main effects and a small number of interactions encounter difficulties when applied to large epidemiological studies where extensive covariate data is collected simultaneously on risk factors. We propose various Bayesian clustering techniques that reduce dimensionality and allow for examination of interactions on a small number of cluster parameters related to lifestyle, genetic variants for key genes and environmental variables.

### A Tree-Based Regression Model for Exploring Complex Gene-Gene and Gene-Environment Interactions

❖ Jinbo Chen, University of Pennsylvania, Department of Biostatistics and Epidemiology, 612 Blockley Hall, Philadelphia, PA 19104, *jchen@cceb.upenn.edu*; Terry M. Therneau, Mayo Clinic College of Medicine

***Key Words:*** gene-gene interaction, gene-environment interaction, tree model, logistic regression model

It is now commonly understood that complex diseases are consequences of interplays between many genes and between gene and environment exposures. Owing to the completion of Human Genome Project, genetic epidemiologists are able to genotype hundreds of single nucleotide polymorphisms (SNPs) in many candidate genes. It is desirable to explore joint effects of these SNPs. Logistic regression model is convenient for screening significant main effects and possibly low-order interactions, but not for exploring complex gene-gene or gene-environment interactions. Tree-based methodology is an attractive alternative in this regard. However, tree models do not allow modeling of main effects and are clumsy for modeling continuous covariates. This work proposes a novel regression model that retains the advantages of both models for exploring gene-gene and/or gene-environment interactions.

### Genetic Interaction Networks in Association Studies of Complex Diseases

❖ Momiao Xiong, The University of Texas Health Science Center at Houston, 1200 Herman Pressler, Houston, TX 77030, *Momiao.Xiong@uth.tmc.edu*

***Key Words:*** genetic interaction networks, complex diseases, interactions, genetic epidemiology, association studies, test statistic

Recently, synthetic genetic analysis has been successfully applied to model organisms to generate genetic interaction networks. The critic question is whether the genetic interaction networks in humans can be uncovered. To address this issue, a novel definition and a new measure of interaction between two unlinked loci (or genes) are introduced. A novel statistic to test for interactions between two loci is proposed. Computational algorithms for construction of genetic interaction networks are developed. The proposed method for construction of genetic interaction networks is applied to the atherosclerosis data set that includes a total 1027 SNPs from 114 candidate genes typed in 916 samples. It is surprising to find that the interacted genes were assembled into a large genetic interaction network that connects the genes in inflammatory, antioxidant and coagulation pathways.

### Estimation of Gene by Exposure Interactions in Case-Parent Triad Studies

❖ Tracy Bergemann, University of Minnesota, 420 Delaware Street, SE, MMC 303, Minneapolis, MN 55455, *berge319@umn.edu*

***Key Words:*** genetic epidemiology, case-parent triad design, model selection, haplotype models

The case-parent triad design genotypes samples drawn from an affected offspring, manifesting a phenotype of interest, as well as from the parents. In my collaborations, we are applying this design to a genetic study of adolescent osteosarcoma patients. We will genotype 2-6 SNPs within each of 11 candidate genes, as well as exposure information for 3 different variables. We test for association, not only of single SNPs, but also any possible gene-gene interactions and gene-environment interactions. Hence, the number of potential log-linear models to fit the data is quite large. We suggest a strategy to find optimal models that incorporate both biological information about the SNPs, as well as traditional methods for model selection such as the BIC. Further, we expand upon existing methods to test for haplotype association in case-control studies, and apply them to nuclear trios.

### Linkage Analysis of Affected Sib Pairs Allowing for Parent-of-Origin Effects: Multilocus Trait Models

❖ Chih-Chieh Wu, M. D. Anderson Cancer Center, Department of Epidemiology, 1155 Pressler Street - Unit 1340, Houston, TX 77030, *ccwu@mdanderson.org*; Sanjay Shete, M. D. Anderson Cancer Center

***Key Words:*** parent-of-origin effects, genomic imprinting, linkage analysis, affected sib pairs, multi-locus models, genetics

Parent-of-origin effects, also known as genomic imprinting, differentiate a higher level of expression of genes inherited from one of the two parental chromosomes. Some genes that affect development and behavior in mammals are known to be imprinted. The statistical methods for testing linkage while allowing for parent-of-origin effects generally have greater power for imprinted loci than the usual statistical methods that ignore the parent-of-origin effects. In order to investigate genetically complex traits in the presence of parent-of-origin effects, multi-locus models of inheritance need to be specified. Here, we present extensions of our previous methods to multi-locus models of inheritance. We propose two types of multi-locus model for incorporation of parent-of-origin effects into linkage analysis.

### Testing for Genomic Imprinting Using Relative Pairs

❖ Wei Guo, The University of Hong Kong, Hong Kong, 00852 China, *guoweidora@hkusua.hku.hk*; Wing K. Fung, The University of Hong Kong

***Key Words:*** genomic imprinting, relative pair, identical-by-descent, kinship coefficient, likelihood ratio test

Genomic imprinting, also known as parent-of-origin effects, is known to be an important epigenetic factor. For imprinted genes, the expression of an allele depends on whether it has been transmitted from the father or mother. The usual statistical analysis for genetic linkage maybe not valid due to the effects of the genomic imprinting, so it is necessary to test for the genomic imprinting as the first step though many authors devoted to incorporate the genomic imprinting effects into the existing linkage analysis. A number of statistical techniques have been used for detection of the genomic imprinting, on the basis of case-parents triads (Weinberg et al.1998; Wilcox et al. 1998; Weinberg 1999). However, the parents of the proband maybe not available for some late onset diseases. In this article, we develop the imprinting tests based on the independent relative pairs.

## 109 Nonstandard Regression and Correlation Problems with Environmental Data ●

Section on Statistics and the Environment, WNAR

**Monday, August 7, 8:30 am–10:20 am**

### Over-Estimation of Trend Caused by Negative Binomial Regression Fit to Zero-Inflated Count Data

❖ Mihoko Minami, The Institute of Statistical Mathematics, 4 6 7 Minami Azabu, Minatoku Tokyo, 106 8569 Japan, *mminami@ism.ac.jp*

***Key Words:*** zero-inflated negative binomial regression model, bycatch data, size parameter, partial dependence, temporal trend

We show that applying negative binomial regression model to data with many zero-valued observation could cause severe overestimation for effects of covariates. In some situations, count data may contain many zero-valued observations, but also include large values. For example, catch (bycatch) of non-target species by a set in fisheries could be mostly zeros, but might take a large value when aggregations of animals are caught. Negative binomial regression model is a widely used regression model to overdispersed count data. For data with many zero-valued observations, negative binomial regression model might look fit adequately well. However, it might overestimate effects of covariates and result in false warning of trend. We investigate this phenomenon theoretically and show some examples in a real situation.

### Estimating Correlation with Multiply Censored Data

❖ Elizabeth Newton, Silent Spring Institute, 29 Crafts Street, Newton, MA 02458, *newton@silentspring.org*; Ruthann Rudel, Silent Spring Institute

***Key Words:*** correlation, censored, detection limit, environmental data

Environmental data frequently are left censored due to detection limits of laboratory procedures. This presents difficulties in statistical analysis of the data. Here we examine methods for estimating the correlation between two variables each of which is multiply censored. We introduce a maximum likelihood estimator (MLE) that, instead of estimating all parameters simultaneously, relies on more accurate estimates of mean and variance. We compare ML methods with Kendall's tau-b, a modification Kendall's tau adjusted for censoring, and several commonly used ad-hoc methods: correlations estimated with non-detects set to DL/2 and correlations of detects only (DET). With increasing levels of censoring most methods are highly biased. The ad-hoc methods in general tend toward zero if singly censored and one if multiply censored. Based on RMSE, DET performs the worst and MLE the best.

### Bayesian Modeling for Ordinal Substrate Size Using EPA Stream Data

❖ Megan Dailey Higgs, Colorado State University, 240 N. McKinley Ave., Fort Collins, CO 80521, *dailey@stat.colostate.edu*; Jennifer A. Hoeting, Colorado State University; Brian Bledsoe, Colorado State University

***Key Words:*** ordinal data, spatial models, Bayesian, categorical data

Substrate size can be indicative of stream health and an important determinant of habitat suitability for fish and macroinvertebrates. The U.S. Environmental Protection Agency (EPA) collected data at 485 stream sites in Oregon and Washington between 1994 and 2004. The measurement and recording protocol for substrate size resulted in ordered categorical data. Previous attempts at building successful predictive models for substrate size have treated the variable as a continuous measurement. We investigate methods to model it as an ordinal categorical variable rather than naively assuming it is continuous. Additionally, we incorporate the spatial correlation inherent in the data using Bayesian methods to build an ordinal categorical spatial model.

### Empirical Evaluation of Sufficient Similarity in Dose-Responsiveness for Environmental Risk Assessment of Chemical Mixtures

❖ LeAnna G. Stork, Monsanto Company, 800 N. Lindbergh Blvd., Mail Zone O3A, St Louis, MO 63167, *leanna.g.stork@monsanto.com*; Chris Gennings, Virginia Commonwealth University; W. Hans Carter, Jr., Virginia Commonwealth University; Linda Teuschler, U.S. Environmental Protection Agency; Edward W. Carney, The Dow Chemical Company

***Key Words:*** equivalence testing, mixed models

When toxicity data are not available for a chemical mixture of concern, U.S. EPA guidelines allow risk assessment to be based on data for a surrogate mixture considered "sufficiently similar". As a supplementary approach we develop statistical methodology to define sufficient similarity in dose-responsiveness for mixtures of many chemicals containing the same components with different ratios. Statistical equivalence testing logic is applied to determine boundary ratios for mixtures with mean dose-response relationships sufficiently similar to an observed mixture, based on a specified biologically meaningful dose-response region of similarity. The similarity region is defined by the investigator or regulator using expert biological judgment. Dose-response data from Rajapakse et al. (EHP, 2002) are used to illustrate the method. (This research is not associated with Monsanto Co.)

### Developing a Worldwide Botanical Database: Factors That Predict the Overlap of Collectors at Herbaria

❖ Cathy Furlong, FCPS/American University, 9412 Cello Court, Graduaute Student, Vienna, VA 22182, *catherine.furlong@fcps.edu*

***Key Words:*** worldwide, botanical, herbaria, policy analysis, data set

The digital availability of 400 or more years of herbarium collections would be a great benefit to both the public and scientists around the world; the development of this dataset is also a high priority to GBIF and NSF. However, to date, fewer than a dozen of the more than 3,000 herbaria around the world have had both the personnel and financial support to begin the process of digitizing their collections. Working with Rusty Russell, collections manager for the National Herbarium, we developed and tested two statistical models---an ANOVA and a linear regression---that indicated varying levels of collector overlap dependent on a combination of geographical specialty and size of an herbarium. The analyses indicated clusters of collector overlap and will be used to present a list of policy and financial support recommendations to both GBIF and NSF.

### Semiparametric Composite Likelihood Inference in Spatial Generalized Linear Mixed Models

❖ Tatiyana Apanasovich, Cornell University, 228 Rhodes Hall School of ORIE, Ithaca, NY 14853, *tanya@orie.cornell.edu*

***Key Words:*** spatial statistics, GLMM

Spatial GLMMs are flexible models for a variety of applications where we have observations of spatially dependent and non-Gaussian random variables. As in a standard GLMM given the random effects, the observations are conditionally independent and follow GLM. The mean is modeled in a general way using regression splines. In a number of applications, neither Bayesian nor maximum likelihood approaches appear practical for large sets of correlated data. To gain computational efficiency, one may approximate the objective function. Instead of the likelihood, we consider a composite likelihood, which is the product of likelihoods for subsets of data, and estimate parameters by maximizing this product. The asymptotic properties of such estimators will be outlined. The application of the methods to the Modeling Electric Power Distribution System Outages in Hurricanes will be presented.

### Search for Multivariate Structure for EMAP Fish Data Using Partition Modeling Approach

❖ Feng Gao, Virginia Polytechnic Institute and State University, 1218 University City Blvd., APT B23, Blacksburg, VA 24060, *fgao@vt.edu*; Eric P. Smith, Virginia Polytechnic Institute and State University; Samantha C. Prins, Virginia Polytechnic Institute and State University

***Key Words:*** Voronoi tessellation, partition model, fish data, RDA/CCA

A multivariate partition modeling approach - a method of clustering sites with multivariate abundance of species as response in order to find the stressor-response relationships of interest will be presented. The method uses random Voronoi tessellations assign sites to one of k clusters that subdivide a region. The BIC-like optimal criterion and hot-spot detection criterion are proposed to find the 'best' clustering or cluster of interest. The BIC-like criterion uses proportion of constrained inertia variance) over total unconstrained inertia (variance) in Canonical Correspondence Analysis (CCA)/Redundancy Analysis (RDA) as a R-square measurement. The method then is applied to EMAP fish data to search for underlying multivariate structure.

## *110* Statistical Applications in Hydrology and Geosciences ●

Section on Statistics and the Environment, Section on Physical and Engineering Sciences, WNAR

**Monday, August 7, 8:30 am–10:20 am**

## Nonparametric Transfer Function Models for Hydrological Forecasting

❖ Heung Wong, The Hong Kong Polytechnic University, Department of Applied Mathematics, Hung Hom, Kowloon, Hong Kong China, *mathwong@polyu.edu.hk*

***Key Words:*** time-series, forecasting, transfer function model, functional-coefficients, back-fitting

To perform hydrological forecasting, linear time series models are employed often. To explore the influence of the inflow on the outflow in a river system and exploit the internal interactions of the outflows, bivariate time series models are needed. The transfer function (TF) model and the semiparametric regression (SPR) model are used widely. In this paper, a new model---the nonparametric and functional-coefficient autoregressive (NFCAR)---is proposed. It consists of two parts: the nonparametric part, which explains the influences of the inflows on the outflow in a river system, and the functional-coefficient autoregressive, which reveals the interactions among the outflows in a river system. By comparing the calibration and forecasting of the models, it is found that the NFCAR model performs well.

## Testing Outliers Using a Mixture Population When Some Data Are Missing and Training Data Are Unlabeled

❖ Aruna Saram, Sam Houston State University, 2501 Lake Road, Apt 236, Huntsville, TX 77340, *stdads32@shsu.edu*; Ferry Butar Butar, Sam Houston State University

***Key Words:*** missing data, EM algorithm, outlier testing, mixture population

This paper is concerned with the problem of multivariate outlier testing for purpose of distinguishing seismic signals of underground nuclear events from training samples based on non-nuclear seismic events when some of the data are missing and unlabeled. Suppose some of the observations are missing and we assume that training data follow a multivariate normal distribution. Using generalized likelihood ratio test procedure to perform the outlier testing and Hotelling's T2 distribution for critical values always perfect when the data are not missing. We describe an EM algorithm base procedure for using the modified likelihood ratio test to test for outliers when the training data follow a mixture distribution and some of the observations are missing. We use seismic data and simulated data to describe this procedure.

## State-Space Models for within-Stream Network Dependence

❖ William Coar, Colorado State University, 3400 Stanford Road, B212, Fort Collins, CO 80525, *coar@stat.colostate.edu*; F. Jay Breidt, Colorado State University

***Key Words:*** Kalman filter, Gaussian likelihood

Because of the natural flow of water in a stream network, characteristics of a downstream reach may depend on characteristics of upstream reaches. The flow of water from reach to reach provides a natural time-like ordering throughout the stream network. We propose a state-space model to describe the spatial dependence in this tree-like structure with ordering based on flow. The model formulation is flexible, allowing for a variety of spatial and temporal covariance structures in the state and measurement equations. A variation of the Kalman filter and smoother is derived to allow recursive estimation of unobserved states and prediction of missing observations on the network, as well as computation of the Gaussian likelihood. Several forms of dependence on the network are described, such as network analogues of autoregressive-moving average models and of local linear trend models.

## Estimation of Space-Time Branching Process Models in Seismology Using an EM-Type Algorithm

❖ Alejandro Veen, IBM T.J. Watson Research Center, Statistical Analysis and Forecasting, 1101 Kitchawan Road, Route 134, Room 29-252, Yorktown Heights, NY 10598, *aveen@us.ibm.com*; Frederic P. Schoenberg, University of California, Los Angeles

***Key Words:*** point process models, branching process, earthquakes, seismology, ETAS, EM

The estimation of branching process models via a Maximum Likelihood can be unstable and computationally difficult. Viewing branching processes as incomplete data problems, however, suggests using the Expectation-Maximization algorithm as a practical estimation method. Using an application from seismology, we show that the Epidemic-type Aftershock Sequence (ETAS) model can in fact be estimated this way and we propose a particularly efficient procedure based on maximizing a partial log-likelihood function. Using a space-time ETAS model, we demonstrate that this method is extremely robust and accurate and use it to estimate declustered background seismicity rates of geologically distinct regions in Southern California.

## Robust Estimation for Periodic Autoregressive Time Series

❖ Qin Shao, The University of Toledo, Math Department, Mailstop 942, Toledo, OH 43615, *qin.shao@utoledo.edu*

***Key Words:*** periodically stationary time series, robust estimators, estimating equations, asymptotic relative efficiency, periodic autoregressive models

A robust estimation procedure for periodic autoregressive time series is introduced. The asymptotic properties and the asymptotic relative efficiency are discussed by the estimating equation approach. The performance of the robust estimators for periodic autoregressive time series models with order one is illustrated by a simulation study. The technique is applied to fitting a model for quarterly streamflow data of the Pecatonica River at Martintown, WI.

## Functional Clustering of Water Pressure Data

❖ Snehalata Huzurbazar, University of Wyoming, Department of Statistics, Department 3332, 1000 E University Avenue, Laramie, WY 82071, *lata@uwyo.edu*

***Key Words:*** functional data analysis, clustering

Functional Clustering of Water Pressure Data Water pressure data were collected from 16 boreholes drilled into the Bench Glacier in Alaska. Data were collected over a field season of about 50 days, at approximately 15 minute time intervals; however, the times of measurement were different across boreholes. The glaciological question of interest focuses on whether there is any clustering of the boreholes. We explore this question by treating the water pressure observations as functional data and performing cluster analysis. Both the functions as well as the first order derivatives are used in the analysis

## Circulant Embedded Extended CAR Models for Large Spatial Data

❖ Ernst Linder, University of New Hampshire, Kingsbury Hall, UNH, Durham, NH 03824, *elinder@unh.edu*

*Key Words:* Gaussian random fields, spatio-temporal analysis, large data

We examine an extension of the usual CAR model for spatial lattice data. The extension incorporates a second spatial parameter that governs smoothness of the underlying spatial field. We show that for a circulant embedded rectangular lattice this model has the same discrete Fourier spectrum as the MatÈrn class of covariance functions for spatial point data. This extended model can also be applied to point referenced data using appropriately defined distance-based weight functions. Thus we achieve an estimation scheme based on the fast Fourier transformation that is computationally efficient for large spatial data. We illustrate how this model and the circulant embedding is implemented within a hierarchical Bayesian estimation framework for spatial and spatio-temporal data. We give examples from earth systems science related to carbon cycling and large scale hydrological systems.

## 111 Prediction and Classification in Genetics and Analysis of Phylogenetic Trees

Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Boosting Nearest Shrunken Centroid Classifier for Microarray Data

❖ Baolin Wu, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware St SE, Minneapolis, MN 55455, *baolin@biostat.umn.edu*

*Key Words:* microarray, sample classification, boosting

Nearest shrunken centroid classifier (NSC) is a class of linear classifiers with built-in feature selections that has proven useful for analyzing microarray data. The simple linear structure of the classification boundary makes NSC easy to interpret and implement, but sometimes this simple structure fails to generalize well for some data. In this paper, we propose boosting NSC to improve its performance, which is based on the development of a novel penalized weighted linear regression model. Through application to public microarray data, we illustrate the favorable performance of the proposed boosted NSC.

### Robust-Affected Sib Pair Linkage Analysis for a Stratified Sample

❖ Guan Xing, Case Western Reserve University, 10900 Euclid Ave., Dept of EPBI, Cleveland, OH 44106, *gxx4@case.edu*; Tao Wang, Case Western Reserve University; Robert C. Elston, Case Western Reserve University; J. S. Rao, Case Western Reserve University

*Key Words:* robust, linkage, classification, Bayesian

For an affected sib pair study using model-free linkage analysis, correct classification of samples into subpopulations is very important. However, the traditional self-report definition of ethnicity is usually subjective and imprecise. The program STRUCTURE has been used to make inferences about population structure with multiple marker information on independent samples. Here we propose a new method to deal with the imperfect sample data. Our approach aims to make use of both self report and genotypic information from correlated family data. Based on the classification obtained with self report information, a new Bayesian classification approach is derived to use the available genotypic information in order to exclude subsets of misclassified families.

### Application of Bayesian Logistic Regression on Gene Expression Prediction

❖ Yuan Yuan, Harvard University, 1 Oxford Street, Statistics Department, Cambridge, MA 02138, *yyuan@stat.harvard.edu*; Lei Guo, Harvard University; Lei Shen, GlaxoSmithKline; Jun Liu, Harvard University

*Key Words:* Bayesian logistic regression, Metropolis-Hastings sampling, gene regulation, transcription factor binding sites

The past few years have seen a remarkable increase of interest in gene regulatory network research, by means of both biological and computational experiments. As for the latter, exploring the relationship between gene expression patterns and their genomic DNA sequences has special importance. We applied Bayesian logistic regression to model this relationship. A Metropolis-Hastings sampling scheme with variable selection is employed to obtain posterior samples of gene expression patterns given their sequence information, which is encoded according to predicted occurrences of over-representing motifs. This model not only provides comparable predict accuracy as other methods, but also suggests a way to improve motif discovery. Furthermore, the posterior samples can be used to evaluate measures of uncertainty of the predicted gene expression level under various conditions.

### Statistical Learning for Analyzing Functional Genomic Data

❖ Axel Benner, German Cancer Research Center, INF 280, Heidelberg, 69120 Germany, *benner@dkfz.de*; Carina Ittrich, German Cancer Research Center

*Key Words:* statistical learning, functional genomics, boosting, penalized regression, survival analysis, model selection

An important topic concerning the statistical analysis of functional genomic data is multivariable predictive modeling. Since in microarray studies the number of predictor variables is much larger than the number of observations, standard statistical model building does not work properly. Statistical learning is a new approach to develop prediction models allowing the inclusion of all available data. Selection methods like boosting and regularization methods like penalized regression have been recognized as important statistical learning methods which can control for complexity. Validation of the fitted models by using bootstrap resampling or cross validation is another important issue. The methods presented above enable for adaptive model selection. We illustrate and compare the different approaches using a data set on predicting survival for patients with acute myeloid leukemia.

### Nonnegative Matrix Factorization: a New Paradigm for Large-Scale Biological Data Analysis

❖ Karthik Devarajan, Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111, *karthik.devarajan@fccc.edu*

*Key Words:* nonnegative matrix factorization, molecular pattern discovery, high-throughput biology, class prediction and class comparison, gene expression, microarray

Nonnegative matrix factorization (NMF) was introduced as an unsupervised, parts-based learning paradigm by Lee and Seung (Nature, 1999). It involves the decomposition of a nonnegative matrix V into two nonnegative matrices W and H, via a multiplicative updates algorithm. In the context of a gene expression matrix V ~ WH, each column of W defines a metagene and each column of H represents the metagene expression pattern of the corresponding sample. NMF has been primarily applied in an unsupervised setting in facial pattern recognition and text mining; and more recently for molecular pattern discovery in high-throughput biological studies. In this paper, we discuss and review the potential applicability of NMF in a supervised learning framework, and as an exploratory tool for dimensionality reduction. We illustrate our methods with several examples using cancer microarray data.

### A Two-Stage Peeling Algorithm and Its Applications to Phylogeny

❖ Arindam RoyChoudhury, University of Washington, Box 354322, Seattle, WA 98195, *arindam@stat.washington.edu*; Joseph Felsenstein, University of Washington; Elizabeth A. Thompson, University of Washington

***Key Words:*** coalescent, divergence time, maximum likelihood, peeling, phylogeny, SNP

We developed a peeling algorithm for likelihood estimation of phylogenetic tree for populations within same species. Our method utilizes the differences accumulated from random genetic drift in allelic count data from single nucleotide polymorphisms. The likelihood computation involves two steps. First, the likelihood is maximized across branch lengths for a given tree topology; then likelihood is maximized across topologies. Our focus is the first step. The peeling algorithm translates the data at the tip of the tree to arrays of probabilities at the root; at the root, the arrays determine the likelihood. The arrays consist of probabilities related to the number of coalescences and allelic counts among the partially coalesced lineages. Tracking these probabilities requires a two-stage algorithm. Our computation is exact, and avoids time consuming Monte Carlo methods.

### Phylogeography of Modern Africa Gorillas Using MCMC

❖ Joungyoun Kim, University of Wisconsin-Madison, 4725 Sheboygan Ave. 242, madison, WI 53705, *joungkim@stat.wisc.edu*

***Key Words:*** phylogeography, gorilla, MCMC, Bayesian, mitochondria, phylogeny

The purpose of this paper is to explore the geographical distribution of genetic variation in modern gorilla populations in order to better understand the phylogeographic history of them. We base our inference on two types of DNA sequence data. One type is HV1, the first hyper-variable region in the mitochondrial control region. The other type is so-called Numts, DNA sequences in the nuclear chromosomes which are thought to be copies of HV1. We expect that the substitution rate in mitochondrial sequences is much higher than in nuclear genes. Therefore, our model uses different rates for each sequence type. We want to estimate (1) the phylogenetic relationships among modern gorilla populations; (2) times of the events when mitochondrial sequence was transferred to nuclear chromosomes. We take a Bayesian approach using MCMC methods for the computation.

## 112 Bayesian Methods in Biopharmaceuticals ●

Biopharmaceutical Section, Biometrics Section, Section on Bayesian Statistical Science, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Bayesian Adaptive Noninferiority Assessment with Safety Measure

❖ Melissa Spann, Eli Lilly and Company, 450 South Madison, Indianapolis, IN 46225, *spannme@lilly.com*; Stacy Lindborg, Eli Lilly and Company; John W. Seaman, Baylor University

***Key Words:*** Bayesian, adaptive, non-inferiority, safety measure, SAS, WinBUGS

We will present a Bayesian adaptive approach to determining if an experimental treatment is non-inferior to an active control treatment within a clinical trial that includes a placebo arm. The design uses joint posterior predictive probabilities of safety and efficacy to determine adaptive allocation probabilities. As part of a retrospective study, analyses of data from a previously FDA approved treatment were performed. We will present results of those analyses and simulation results using programming executed in SAS and WinBugs.

### Bayesian Sequential Analysis for Survival Data

❖ Lili Zhao, The University of Iowa, 513 Hawkeye Drive, Iowa CIty, IA 52246, *lili-zhao-1@uiowa.edu*; George G. Woodworth, The University of Iowa

***Key Words:*** sequential analysis, interim study, non-inferiority test, backward induction, piecewise exponential, Bayesian

Sequential designs are widely used in clinical trials. In some cases, it is more natural to construct stopping rules based on time-to-event variables. However, Bayesian decision-theoretic methods for designing group sequential monitoring with time-to-event endpoints have not been described in the literature. In this study, we propose a simulation-based solution for a single-armed Phase IIB study to decide whether a new treatment is non-inferior or inferior, relative to a standard. Once the decision boundaries are obtained, we compute Bayesian and frequentist operating characteristics including Type I error, power, and expected run length; we further explore the optimal number of interim looks and the effect of sample size on operating characteristics.

### Bayesian Approach for Predicting the Margin of Safety in Nonclinical Safety Assessment Studies: a Case Study

❖ Gheorghe Doros, Yale University, 715 Albany Street, T420E, Boston, 02118, *gheorghe.doros@aya.yale.edu*; Viencent Reynolds, Eli Lilly and Company; Eyas Aby-Raddad, Eli Lilly and Company

***Key Words:*** Bayesian analysis, nonclinical dose selection, margin of safety

For nonclinical safety assessment studies, dose selection is critical factor that influences the margin of safety (MOS) which will ultimately be defined by the study. We developed a Bayesian approach to address the problem of selecting doses and anticipating their associated MOSs in toxicology studies of compounds for which prior information is available on the safety and efficacy of other compounds in the same pharmacologic class. For the compound class under study, prior history had established that dose- and time-dependent cardiomegaly was a common finding that frequently defined clinical dosing limits. Use of a Bayesian statistical model for heart weight changes over time provided a predictive means not only to evaluate the risk of exceeding putative NOAELs, but also to maximize the probability of achieving specified MOSs with different dose selection scenarios.

### Bayesian and Composite Designs for Drug Combination Studies

❖ Yuehui Wu, GlaxoSmithKline, 1250 S. Collegeville Road, P O Box 5089, Collegeville, PA 19426, *Yuehui.2.Wu@gsk.com*; Vladimir Dragalin, GlaxoSmithKline; Vlareii Fedorov, GlaxoSmithKline

***Key Words:*** optimal design, bivariate probit model, two-stage composite design, Bayesian design, combination of drugs

The major target of our study is to implement optimal design techniques in dose ranging studies on efficacy and toxicity responses using bivariate probit model. Optimal experimental design methodology was used to construct efficient dose allocation procedures for estimation of parameters of the dose-response relationship as accurately as possible given ethical concerns and prior information. Traditional locally optimal design provides the most accurate parameter estimation per patient (or "penalty unit") if the prior knowledge provides a "good" guesses of the unknown parameters of interest. The two-stage composite design is more appealing in practice since it does not rely heavily on those guesses by taking into account the results of initial experiment. When enough patients are available (total number of patients goes to infinity), it converges to the optimal design.

### A Bayesian Simulation-Based Approach in Investigating Physiologically-Based Drug-Drug Interaction Prediction

❖ Zhiping Wang, Indiana University Purdue University Indianapolis, 1050 Wishard Blvd., RG4101, Indiana University, indianapolis, IN 46227, *zhipwang@iupui.edu*; Lang Li, Indiana University; Stephen Hall, Indiana University

***Key Words:*** drug-drug interaction, pharmacokinetic

Drug-drug interactions (DDIs) are a significant cause of adverse drug reactions that in turn results in significant morbidity and mortality. We propose to develop a suite of Bayesian tools to predict DDIs based on a general physiologically-based pharmacokinetic (PBPK) model framework, and to develop a web-based interface to implement these tools. This is a population approach designed to characterize DDIs by considering all variation sources. False negative rate (FNR) in DDI prediction is carefully defined. Our data analysis and simulation studies have shown that FNR is not only negatively associated with the DDI parameter, , but also positively associated with between subject variations of PK parameters. Its application is demonstrated with a ketoconazole-midazolam example.

### Bayesian Adaptive Dose Selection

Melissa Spann, Eli Lilly and Company; ❖ David Manner, Eli Lilly and Company, Lilly Corporate Center, DC 2233, Indianapolis, IN 46285, *mannerdh@lilly.com*; John W. Seaman, Baylor University

***Key Words:*** Bayesian, adaptive, dose selection, SAS, WinBUGS

We will present a Bayesian adaptive approach to dose selection that uses effect sizes of doses relative to placebo to select the most efficacious dose. We assume a parallel design with multiple treatment arms including a placebo arm and a continuous outcome measure. The proposed design removes treatment arms if their performance relative to placebo or other treatment arms is undesirable. A linear or quadratic function is used to determine the rate (slow or fast) at which treatment arms can be removed. This allows the investigator flexibility by presetting the criteria of acceptable performance for a treatment arm for a given trial. We will present simulation results based on programming executed in SAS and WinBugs.

### Predicting Phase III Trial

❖ Madhuja Mallick, Merck Research Laboratories, 126E Lincoln Ave., RY 34 A316, Rahway, NJ 07065, *madhujamallick@hotmail.com*; Bret Musser, Merck Research Laboratories

Phase II proof-of-concept studies are used to decide whether further investment in a compound is worthwhile. Statistical significance for the treatment difference is not sufficient evidence to enter into another lengthy and expensive future trial; instead, determining if a clinically meaningful difference did (or could) occur is of interest. Typically, results from the Phase II study are assessed qualitatively to determine the chance of future success. The purpose of this research is to quantitatively estimate the probability of success. There are various means of accomplishing this: simple frequentist methods, Bayesian posterior probabilities, and Bayesian predictive probabilities. Bayesian methods would weight prior beliefs alongside data generated from the proof of concept study in a model-based approach. Examples of such procedures will be described via simulations.

## 113 Diagnostic Tests and Cancer Screening ●

Biometrics Section, Section on Health Policy Statistics, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Bootstrap Confidence Intervals for the Area under the ROC Curve

❖ Gengsheng Qin, Georgia State University, 30 Pryor Street, Atlanta, GA 30303, *gqin@gsu.edu*; Lejla Hotilovac, Georgia State University

***Key Words:*** ROC, AUC, bootstrap, confidence interval, diagnostic test

The accuracy of a diagnostic test is of high importance in clinical medicine. It is measured often by the area under the ROC curve (AUC). To report this area properly, it is necessary to construct a confidence interval for its value. In this talk, we will present five bootstrap confidence intervals for the AUC. Then, we will conduct simulation studies to compare the performance of these confidence intervals with the existing confidence intervals for the AUC. The results will be compared in terms of coverage probability and average interval length. A real example will be used to illustrate application of the new methods.

### Analysis of Medical Diagnostic Test Data with a Test Ignorance Region

❖ Andrzej Kosinski, Duke University, Department of Biostatistics and Bioinformatics, PO Box 17969, Durham, NC 27715-7969, *andrzej.kosinski@duke.edu*

***Key Words:*** medical diagnostic test, test ignorance region, missing gold standard, sensitivity, specificity

When collecting data for a diagnostic test evaluation, one often finds that not all patients undergoing the test have their disease status verified with the gold standard. In this situation, point estimates for sensitivity and specificity of a test are only possible under more or less realistic assumptions about the missing data. The Test Ignorance Region (TIR) (Kosinski and Barnhart 2003) is an assumption free region encompassing all combinations of sensitivity and specificity values compatible with the observed data. We argue that the TIR should be a routine reporting tool for a first step in evaluation of a diagnostic test when the gold standard is partially missing. This way the information included in the actually observed data and "information" induced by assumptions can be clearly separated. More recent extensions to subregions will be discussed and real data examples analyzed.

### Skill Curves: a New Method for Evaluating Diagnostic Tests

Russell Zaretzki, University of Tennessee; ❖ William M. Briggs, Weill Medical College of Cornell University, 525 E. 68th, Box 46, New York, NY 10021, *wib2004@med.cornell.edu*

***Key Words:*** ROC curves, forecast evaluation, skill, predictive power

In this talk we introduce a method called the skill curve which is useful in evaluating the diagnostic power of a test variable. For simple diagnostics based on a single continuous measurement we demonstrate the advantages over the frequently used ROC curve. Skill curves provide information about the effectiveness of diagnostic test based on every available decision boundary. In addition, these curves provide information on the sensitivity of the test to the decision boundary chosen. We discuss parametric, non-parametric and semi-parametric methods for fitting these curves as well and report properties of skill based confidence intervals.

### Incorporation of Metabolic Insight into Analysis of High-Dimensional Structural Lipid Datasets

❖ Michelle Wiest, Lipomics Technologies, Inc., P.O. Box 593, Knights Landing, CA 95645, *mwiest@lipomics.com*; UyenThao Nguyen, Lipomics Technologies, Inc.; Aldo Bernasconi, Lipomics Technologies, Inc.

***Key Words:*** metabolomics, high-throughput, lipids

The emerging field of Metabolomics promises to diagnose and guide human health. The failure of first-generation diagnostics to reach clinical practice derives from the use of datasets where properties of metabolites were measured but the metabolites were not identified. While valuable for screening, such platforms are unable to identify metabolic pathways that cause subsequent health problems. Quantitative platforms focusing on known metabolites can resolve statistically robust and biologically meaningful differences. The statistical challenge is combining analytical precision and biological accuracy into the modeling of high density metabolite datasets. Examples of accurate, comprehensive measurements of certain known metabolites (structural lipids) show the power of biologically supervised statistical estimation to create actionable diagnostics suitable for personal health assessment.

### Challenges for Statisticians in Cervical Cancer Screening Research

❖ Jong Soo Lee, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 193, Houston, TX 77030, *jslee@stat.rice.edu*

We highlight some of the collaborative research effort between biomedical engineers and statisticians in a cervical cancer screening project. The biomedical engineers have created an optical device for use in early detection of cervical cancer, and we first present some background on this medical device. Next, we discuss the statistical techniques used. During the development of this novel technology, statisticians were involved with everything from preliminary data screening to inference on the device output to the development of supervised learning algorithms for disease classification. We show how the statistical methods have been vital to this project and why the development of new methods became necessary. Lastly, we outline further research opportunities in this joint venture.

### Quantitative Impact of Length-Biased Sampling in Cancer Screening

❖ Sonya Heltshe, University of Colorado at Denver and Health Sciences Center, 1069 Pearl Street, Apt 7, Denver, CO 80203, *sonya.heltshe@uchsc.edu*; Karen Kafadar, University of Colorado

***Key Words:*** randomized screening trial, length-bias, sojourn time, benefit time, cancer

Sometimes measured data are subject to length-biased sampling (e.g., gas lines or hospitalization stays). When the variable, which is length-biased sampled, is observable, the distribution of the observations is known (Cox and Lewis 1972). Length bias exists in screening programs where disease may be detected during the preclinical stage, as a longer sojourn time yields a greater likelihood of being screen-detected (Zelen 1976). This paper quantifies the effect of length-biased sampling on clinical duration when cases are subject to periodic screening. Underlying bivariate distribution, correlation (between sojourn and clinical time), mixing proportion of slow versus fast progressing disease, and the ratio of screening interval length to mean sojourn time all influence the magnitude of the effect length-biased sampling has on the distribution of clinical duration.

### Bayesian Inference for the Lead Time in Periodic Cancer Screening

❖ Dongfeng Wu, Mississippi State University, Department of Math and Statistics, Mississippi State, MS 39762, *dwu@math.msstate.edu*; Gary L. Rosner, M. D. Anderson Cancer Center; Lyle D. Broemeling, M. D. Anderson Cancer Center

***Key Words:*** lead time, periodic screening exams, breast cancer early detection, sensitivity, sojourn time, transition probability

This paper develops a probability distribution for the lead time in periodic cancer screening exams. The general aim is to provide statistical inferences for the lead time, the length of time the diagnosis is advanced by screening. The lead time is distributed as a mixture of a point mass and a piecewise continuous distribution. Simulations are carried out, using the HIP data, to estimate under different screening time intervals, the proportion of breast cancer patients who truly benefit from the periodic screening exams and the proportion that do not. The mean, the mode, the standard deviation and the density curve of the lead time are also provided. This provides important information to policy makers regarding the screening period and the long-term benefit for women who take part in the program. The result is also applicable to other kinds of chronic disease.

## 114 Estimation

Section on Statistical Computing

**Monday, August 7, 8:30 am–10:20 am**

### The Gentle Side of Kalman Filtering

❖ Yolanda Munoz Maldonado, The University of Texas School of Public Health, 1200 Herman Pressler, Suite RAS W834, Houston, TX 77030, *Yolanda.M.Munoz@uth.tmc.edu*

***Key Words:*** complete likelihood, ridge regression, varying coefficients, mixed models, diffuse priors, computational efficiency

Kalman Filtering is a powerful filtering algorithm that reduces the computational burden of calculating estimators, predictors, and evaluation of likelihoods of stochastic processes that can be modeled using a state-space representation. However, the Kalman filter formulas, and its onerous notations, have intimidated the more applied statistician and restricted its influence. In this talk, we present examples of Kalman filtering implementation in the settings of ridge regression, randomized block designs, varying coefficient models, and regression with correlated errors. Our objective is to demystify the Kalman filter by explicitly writing the state-space model for each of these settings and show how to obtain estimators, predictors, and nuisance parameters by implementing an efficient, O(n), Kalman filter algorithm that also deals with diffuse initial conditions.

## On Some Aspects of Estimation of a Common Mean of Two Independent Normal Populations

❖ Pranab Mitra, University of Maryland Baltimore County, 912 Hooper Ave., Apt C, Baltimore, MD 21229, *pranab1@math.umbc.edu*

***Key Words:*** common mean, first order efficient, Fisher fisher information, Jeffrey's prior, Rao-Cramer lower bound, Rao-Cramer lower bound, second order efficient.

In this paper, we study aspects of estimation of a common mean of two normal populations from an asymptotic point of view. The Bayes estimate of the common mean under Jeffrey's prior also is considered. A simulation study is carried out to compare several competing estimates in small samples.

## A Comparison Study of Procedures for Estimating the Tail Index of Heavy-Tailed Distributions

❖ Bruno C. de Sousa, Universidade do Minho, Escola de Ciencias DMCT, Campus de Azurem, Guimaraes, 4800 058 Portugal, *bruno@mct.uminho.pt*; George Michailidis, University of Michigan

***Key Words:*** heavy-tailed distributions, tail index estimators, hill estimator, qq-estimator, Zipf plot, sum plot

Many procedures, most asymptotic in nature, have been proposed for estimating the tail index in heavy-tailed distributions. This statistical problem is linked to the problem of determining the appropriate number of upper order statistics used in the estimation of the tail index. The study shows that adaptive algorithms, based on minimizing the asymptotic mean squared error of the Hill estimator, together with the Sum plot procedure and methods based on the T_k and Q_k statistics, prove to be competitive in all the scenarios examined. A good performance is also observed for the procedures based on the H^{(k)} and K^{(k)} plots, while the Zipf plot procedure did not exhibit satisfactory results. An important practical implication of this study is that the attractive graphical features of the Sum plot procedure make it extremely useful in practice when combined with other procedures.

## Estimation of the Parameter of the Skewed Double Exponential Distributions

❖ Keshav Jagannathan, Coastal Carolina University, Department of Mathematics and Statistics, Conway, SC 29528, *kjaganna@coastal.edu*

***Key Words:*** skewed double exponential distribution, asymptotic unbiasedness, consistent estimators

This talk will focus on estimating the parameter of the Skewed Double Exponential distributions using two different methodologies. Shortcomings of the estimators, as well as key properties will be discussed. Uniqueness results pertaining to the estimators as well as simulation studies will also be presented.

## A Note on the Estimation of Extreme Value Distributions Using Maximum Product of Spacings

❖ Tony Siu Tung Wong, The University of Hong Kong, Department of Statistics and Actuarial Science, Pokfulam Road, Hong Kong, 00852 China, *h0127272@hkusua.hku.hk*; Wai K. Li, The University of Hong Kong

***Key Words:*** generalized extreme value distribution, generalized Pareto distribution, maximum likelihood, maximum product of spacings, Moran's statistic

The maximum product of spacings (MPS) is employed in the estimation of the Generalized Extreme Value Distribution (GEV) and the Generalized Pareto Distribution (GPD). Efficient estimators are obtained by the MPS for all ?. This outperforms the maximum likelihood estimators (MLEs) which are only valid for ?<1. It is then shown that the MPS gave estimators closer to the true parameters compared to the MLE in a simulation study. In cases where sample sizes are small, the MPS performs stably while the MLE does not. Finally, as by-product of the MPS, a goodness of fit statistics, Moran's statistic, is available for the extreme value distributions. Empirical significance levels of Moran's statistic calculated are found to be satisfactory with the desired level.

## Inference on the Mean Parameter of the Skewed Distribution

❖ Toshinari Kamakura, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 Japan, *kamakura@indsys.chuo-u.ac.jp*

***Key Words:*** simulation, approximation, skewed distribution, survival analysis

In the field of the survival analysis the positive skewed distributions play important roles for describing lifetime, survival time or failure time. The mean value is the one of the most important parameters that specifies the underlying distributions. In case of the large sample size the central limit theory supports well normal approximation of the distributions of the sample mean and ML estimates. However, for small sample this approximation can not be adequate to test the mean parameter and make confidence intervals. In this article we investigate the behavior the several estimates for the mean parameter of the skewed distribution by simulation studies and report the properties of the estimates and represent the possibility of the new method.

## Comparing Ratio Estimators Based on Systematic Samples

❖ Hasan Hamdan, James Madison University, Department of Mathematics and Statistics, MSC 7803, Harrisonburg, VA 22801, *hamdanhx@jmu.edu*

***Key Words:*** systematic sample, ratio estimator, bootstrap

The purpose of this study is to evaluate two competitive ratio estimators, often called the mean of ratios and ratio of means, when a systematic sample of size n with a random start is used. For example, in stereology this might be the ratio by volume of mitochondria in a liver cell or the proportion of a mineral in a sample of a rock. Since there is no explicit formula for the variance of a ratio estimator in such cases, several current variance estimators are presented and compared using simulated objects. Two new approaches are also suggested. The first new method is basically a bootstrap estimate using a non-linear additive regression technique. A Monte-Carlo simulation is done using the predicted values from the fitted model to find estimates for the variances. The second method is based on finding the best linear unbiased estimator of the slope assuming a non-constant variance

# 115 Statistical Methods and Applications ●

General Methodology

**Monday, August 7, 8:30 am–10:20 am**

## On the Stability of Statistical Tests

❖ Daniele De Martini, Universit‡ del Piemonte Orientale,
Via Perrone 18, Dipartimento SEMEQ, Novara, 28100 Italy,
*demartin@eco.unipmn.it*

***Key Words:*** plug-in power estimation, observed power, rejection region thresholds, power lower bounds

Stemming from the variability of statistical tests, we define rejection regions on the basis of the estimator of the power of the test, which also measures the reproducibility probability of statistical significance. The threshold for the power estimator to define statistical tests turns out to be 1/2, at least approximately. We then introduce the notion of stability, which is satisfied when the lower bound of the power falls in the rejection region. We show that stability evaluation can be performed in parametric and nonparametric frameworks.

## A Test of Independence in Two-Way Contingency Tables Based on Maximal Correlation

❖ Deniz Yenigun, Bowling Green State University, Department
of Mathematics and Statistics, BGSU, Bowling Green, OH 43403,
*denizy@bgnet.bgsu.edu*; Gabor Szekely, Bowling Green State
University

***Key Words:*** maximal correlation, tests of independence

The maximal correlation has several attractive properties for measuring dependence between two random variables. However, except for a few special cases, it is hard to evaluate the maximal correlation explicitly. In case of two-way contingency tables, we discuss a procedure for estimating the maximal correlation which we use for constructing a test of independence. We compare this test with other existing tests of independence for several alternatives.

## Multiple Comparison Procedures

❖ Yan Li, The Pennsylvania State University, Department of
Statistics, 265 Blue Course Drive, Apt 8A, State College, PA 16803,
*yanli@stat.psu.edu*

***Key Words:*** type I error, Bonferroni procedure, power, multivariate normal probability, correlation

Hypothesis test in multiple comparison setting is an important topic in statistical literature. A number of modified procedures are available that control the Type I error by adjusting the significance level. Among these procedures, the Bonferroni procedure is the most well known and widely used for its simplicity in calculation and understanding. However, all proposed procedures are conservative compared to the exact approach. My study objective is to investigate how far the Bonferroni procedure will deviate away from the exact approach in terms of the rejection critical values and its sample size in a three-armed clinical trial scenario. My numerical results indicate that the Bonferroni procedure performs quite well when no correlation exists and the number of multiple comparisons is not large. But a positive dependency will weaken the power of the Bonferroni procedure dramatically.

## Using Permutation Tests To Study Infant Handling by Female Baboons

❖ Thomas Moore, Grinnell College, Department of Mathematics
and Statistics, Grinnell, IA 50112, *mooret@grinnell.edu*; Vicki
Bentley-Condit, Grinnell College

***Key Words:*** permutation tests, primatology, power

Female baboons, some with infants, were observed and counts made of interactions in which females handled the infants of other females (so-called infant-handling). Independent of these observations, each baboon is assigned a dominance rank of "low," "medium," or "high." Researchers hypothesized that females tend to handle infants of females ranked below them. The data form an array with row labels being infant labels and columns being female labels. Entry (i,j) counts total infant handlings of infant i by female j. Each count corresponds to one of 9 combinations of female by infant/mother ranks, which induces a 3-by-3 table of total interactions. We use a permutation test to support the research hypothesis, where ranks are permuted at random. We also discuss statistical properties of our method such as choice of test statistic, power, and sensitivity to individual observations.

## Selection Bias Due to Immigration in Pharmacoepidemiologic Studies

❖ Henrik Stovring, University of Southern Denmark, JB Winslows
Vej 9A, Research Unit of General Practice, Odense, 5000 Denmark,
*hstovring@health.sdu.dk*

***Key Words:*** pharmacoepidemiology, selection bias, waiting time distribution, prevalence

In pharmacoepidemiology, a so-called run-in period is used often to determine treatment status at an index date. Individuals immigrating during the run-in period are, however, excluded from further analysis, as their treatment status cannot be determined. This may lead to selection bias. Using the recently developed waiting time distribution (WTD) method, which avoids use of a run-in period, the size of this bias is studied with respect to prevalence of use of antidiabetic drugs in the County of Funen, Denmark, from 1993--2003. The relative bias in prevalence estimates, due to the selection created by using a half-year run-in period, is found to be approximately 1% for each calendar year (range .73%; 1.36%), but when controlled for gender and age, this bias vanishes, except for ages under 40. Immigrants generally had a prevalence half the size of nonimmigrants.

## Playing Fast and Loose with Time and Space: Statistics in Forensic Science

❖ Max Houck, West Virginia University, Forensic Science Initiative,
886 Chestnut Ridge Road, Morgantown, WV 26506-6216,
*max.houck@mail.wvu.edu*

***Key Words:*** forensic science, uniqueness, fingerprints, DNA

Many forensic sciences allege that the results of their comparisons yield uniqueness of source attribution. This uniqueness comes with certain assumptions that affect the resolution of their significance. The first is that all things are unique in space and, thus, their properties are non-overlapping. The second is that properties are constant over time. These assumptions are not provable---the population size of "all things that might be evidence" is simply too large to account. Forensic science is therefore relegated to using statistics to interpret its results, but many forensic methods explicitly deny the applicability of statistics, such as fingerprints, while others deeply embrace statistics, such as DNA. This paper lays the foundation for why statistics should be used---in the proper evidentiary context---for all forensic analyses.

### Identifying and Interpreting Regional Convergence Clusters across Europe: Asymptotic versus Boostrapped Inference

❖ Luisa Corrado, University of Cambridge, Sedgwick Site, Robinson Building, Cambridge, CB39DD UK, *lc242@econ.cam.ac.uk*; Melvyn Weeks, University of Cambridge

***Key Words:*** regional convergence, new economic geography, cluster methods, bootstrapping, stationarity tests

In this paper we test for regional convergence clusters across the European Union. We base the analysis on a cluster methodology using the following two steps: (i) generate regional clusters based upon pairwise stationarity tests; (ii) test the resulting cluster outcomes against a number of possible hypothesized cluster patterns based upon geographic, socio-demographic and politico-institutional factors as suggested by the new regional economic 'theory'. Further, our method allows regional convergence clusters to vary over time. Given that the time series are relatively short, there are potential problems in basing inference on asymptotic results for stationarity tests. To circumvent this problem we bootstrap the stationarity test and explore the robustness of the cluster outcomes.

## 116 Frontiers in Bioinformatics

IMS, Biometrics Section, ENAR

**Monday, August 7, 8:30 am–10:20 am**

### Nonparametric Functional Mapping of Quantitative Trait Loci with Incomplete Genotypic Data

❖ Jie Yang, University of Florida, Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611, *jyang81@ufl.edu*; George Casella, University of Florida

***Key Words:*** functional mapping, missing data, nonparametric statistics, mixed model

The dataset used for QTL (quantitative trait loci) mapping contains not only genotypic values like markers but also phenotypic values measured from some traits. In practice, missing data situations are usually faced due to many reasons. A popular way of handling missing genotypic data is to simply discard subjects with missing markers, which directly results in loss of power. Here we propose to use full information from all subjects, whether having complete marker information or not, to gain as much power as possible. Nonparametric functional mapping approach is used to find evidence of existing QTL, which uses nonparametric function estimation to characterize phenotypic trajectories, and then construct a likelihood ratio test throughout the linkage map. Simulation studies and application to a real dataset are provided to illustrate the power-boosting advantage of our proposed algorithms

### Context-Dependent Models for Discovery of Transcription Factor Binding Sites

❖ Chuancai Wang, The Pennsylvania State University, 600 Centerview Drive, Suite 2200, Hershey, PA 17033, *cwang@hes.hmc.psu.edu*; Jun Xie, Purdue University; Bruce A. Craig, Purdue University

***Key Words:*** conditional position dependent motif model, Gibbs sampler, Markov model, Markov order determination, transcription factor binding site

Transcription factors play a crucial role in gene regulation, and the identification of transcription factor binding sites helps gain insight into gene regulatory mechanisms. The overall goal of this work is to describe a new method of binding site detection called Motif Discovery via Context Dependent Models (MDCDM). We characterize the motif (i.e., binding sites) by a series of position-dependent first-order Markov models. In addition, a ``step-up" testing procedure is used to automatically determine the best-fitting Markov model for the background (i.e., nonsite regions). We compare our approach with the existing methods using both real and simulated data sets. The results show that the detection of binding sites can be greatly improved by accounting for dependence across positions in a motif and appropriately modeling the background dependence.

### Estimating the Variation in S Phase Duration Using Branching Processes

❖ Sara Larsson, Lund University, Box 118, Lund, 221 00 Sweden, *sara@maths.lth.se*

***Key Words:*** branching processes, cell cycle, rate of DNA replication, BrdUrd-DNA flow cytometry

It is of great interest to increase the knowledge of cell cycle kinetics parameters, such as the length of the cell cycle and its various phases. Of particular interest is the S phase, during which DNA is replicated. By measuring the DNA content with the BrdUrd-DNA flow cytometry method it is possible to follow the progression of cells through the cell cycle. Branching processes are used to model tumor cells in order to estimate the DNA synthesis time. The total rate of DNA replication is modeled as a base line rate multiplied by a random variable. To extract information about the variation in replication rate within the cell population, measurements from several time points are desirable. By assuming a distribution of the DNA synthesis time it is possible to estimate the unknown parameters using the maximum likelihood method.

### Determination of Differentially Expressed Features in a Combined LC-MS and LC-MS/MS Proteomics Work Flow

❖ Olga Vitek, Institute for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, *ovitek@systemsbiology.org*; Andrew Garbutt, Institute for Systems Biology; Ruedi Aebersold, Institute for Molecular Systems Biology

***Key Words:*** proteomics, mass spectrometry, biomarker discovery, sample size, differential expression, correlation

Liquid chromatography coupled to mass spectrometry (LC-MS and LC-MS/MS) is used increasingly for global protein profiling. However, reliable and reproducible analysis of these data is challenging. LC-MS spectra are sensitive in quantifying the abundance of features, but yield features with unknown identities and many false positives. LC-MS/MS approach is specific in sequences identification, but fails to detect many of the features. We combine the advantages of LC-MS and LC-MS/MS approaches in a statistical framework for determination of differentially expressed features. We use identities of MS/MS peaks and shared abundance profiles to account for correlation structure in the data. We also discuss aspects of experimental design---particularly choices of sample size---for mass spectrometry-based experiments. Finally, we illustrate our approach using two complex biological samples.

### Estimation of Variance in Two-Way Semilinear Models

❖ Weihua Tang, Rutgers University, 506 Dillion Court, Dept. of Statistics, North Brunswick, NJ 08902, *weihuat@stat.rutgers.edu*; Jian Huang, The University of Iowa; Cun-Hui Zhang, Rutgers University

***Key Words:*** microarray, semilinear model, variance estimation, convergence rate, Neyman-Scott, bias correction, convergence rate, Neyman-Scott, bias correction

Estimation of variance is essential in significance analysis of microarray data. In the setting of two-way semilinear linear models for microarray data, we propose to estimate noise level smoothly in intensity, moreover with bias correction which due to a Neyman-Scott type phenomenon. The proposed noise level estimator is proved with optimal convergence rates. Variance estimation for some linear combinations of gene effect estimates is also investigated. Studies on real and simulated datasets demonstrate the validity and superiority of our methods.

### Treating Expression Levels of Different Genes as a Sample: Does It Do a Good Job?

❖ Andrei Yakovlev, University of Rochester, 601 Elmwood Ave., Rochester, NY 14642, *Andrei_Yakovlev@urmc.rochester.edu*; Lev Klebanov, Charles University

***Key Words:*** microarrays, data analysis, sampling, stochastic dependence

Many modern methods of microarray data analysis resort to pooling the expression measurements across genes. Such methods tend to treat the pooled measurements as a sample drawn from some distribution. Several universal laws were proposed to analytically describe this distribution. The present paper discusses various pitfalls inherent in this approach from theoretical and practical perspectives. Our main conclusion is that pooling across genes leads to huge losses of information in defiance of the frequently encountered opinion that it allows one to "borrow strength" from other genes when making inference on a particular gene.

### Model-Based Analysis of Tiling-Arrays for ChIP-chip

❖ William Evan Johnson, Harvard University, 73 Marathon Street, Arlington, MA 02474, *wjohnson@hsph.harvard.edu*; Wei Li, Dana Farber Cancer Institute; Clifford Meyer, Dana Farber Cancer Institute; X. Shirley Liu, Dana Farber Cancer Institute

***Key Words:*** tiling microarrays, transcription factor, chip-chip, normalization, linear model, false discovery rate

We propose an analysis algorithm to reliably detect regions enriched by transcription factor Chromatin ImmunoPrecipitation (ChIP) on Affymetrix tiling arrays (chip). We model the baseline probe behavior using probe sequences and copy number on each array and standardize the probes via this probe model, resulting in a unique single chip simultaneous normalization/background adjustment. A robust scoring function is proposed to score regions for ChIP-enrichment, which allows for p-value and FDR calculations. Our method can detect ChIP-regions from single ChIP samples, multiple ChIP samples, or multiple ChIP samples with controls. The single array ChIP-region detection feature minimizes logistical constraints for labs newly adopting ChIP-chip to test protocols and antibodies or allows established ChIP-chip labs to identify samples with questionable quality that contaminate their data.

## 117 Frames

Section on Government Statistics
**Monday, August 7, 8:30 am–10:20 am**

### A Confidence Set for Estimates from Data Collected Using Double Sampling

❖ Zhanyun Zhao, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08540, *zzhao@mathematica-mpr.com*; John Hall, Mathematica Policy Research, Inc.

***Key Words:*** double sampling, confidence set, binomial, unemployment insurance

Double sampling is one of the acceptance sampling techniques used in the quality control field. Double sampling uses an initial sample from a "lot" of items, based on which, the lot may be rejected, accepted or a second sample may be inspected. After the second sample is inspected, the lot is either accepted or rejected. This paper attempts to develop a confidence set for the observed failure rate where double sampling has been used since the statistical quality control literature does not appear to have defined such a confidence set. The issue of a confidence set arose in a project for the US Department of labor where double sampling was used to check the accuracy of Unemployment Insurance (UI) records within each state. To plan for a subsequent inspection, a confidence set based on the first inspection would assist the states in judging if corrective action was needed.

### Using Sensitivity Analysis To Manage Uncertain Matching in Multiple Systems Estimation

❖ Patrick Ball, Human Rights Data Analysis Group, 480 California Ave., Suite 201, Palo Alto, CA 94306, *Patrick.B@benetech.org*; Scott Weikart, Human Rights Data Analysis Group; Rapheal Kaplan, Human Rights Data Analysis Group; Romesh Silva, Human Rights Data Analysis Group; Jeff Klingner, Human Rights Data Analysis Group

***Key Words:*** multiple systems estimation, record linkage

Multiple Systems Estimation (MSE) assumes that the identification of duplicates within a system and the record-linkage (matching) process between systems is error-free. When analysts work with data collected in unsettled situations, conflict zones and least developed countries, biographical information about individuals is often missing or incorrect. As a result, the matching decisions are subject to substantial uncertainty. Consequently, we developed a parameterized record-linkage process which generates multiple sets of matched data by varying the criteria upon which record-linkage decisions are made. We then calculate MSE estimates for all sets of the matched data and assess the distributional spread of the resulting estimates. Examples will be presented using mortality data from Timor-Leste and Sierra Leone.

### Measuring Employment and Wages in the Aftermath of Hurricanes Katrina and Rita

❖ Linda Unger, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 4840, Washington, DC 20212, *Unger.Linda@bls.gov*; Richard Clayton, Bureau of Labor Statistics

***Key Words:*** imputation, estimation, economic shocks, GIS, FEMA, hurricane

Hurricanes Katrina and Rita caused vast damage to the Gulf Coast and New Orleans posing difficult measurement issues and opportunities to understand the scope of the damage and assist those planning the recovery. BLS applied its Quarterly Census of Employment and Wages to this challenge. First, the geocoded list of all businesses was linked to FEMA GIS damage areas to profile potential damage when no other current measures existed. The utility and accuracy of the FEMA damage zones for economic determinations are explored. Also, this paper describes special treatments and changes to imputation addressing the special situations caused by large population shifts and damage to business and infrastructure. Lastly, the QCEW business register provides monthly employment and wages through December 2005, thus profiling "before and after" for this early stage of the recovery.

### How Quickly Do New Construction Addresses Appear on the Delivery Sequence File?

❖ Tina Flanagan, U.S. Census Bureau, 2312 Pinefield Road, Waldorf, MD 20601, *tina.michelle.flanagan@census.gov*; Clifford Loudermilk, U.S. Census Bureau

***Key Words:*** master address file, delivery sequence file, permit frame, new construction, lag

Currently several household surveys conducted by the Census Bureau sample from multiple frames including a unit frame which contains addresses from the decennial census and a permit frame which contains new construction addresses identified by permit offices. In the future, the Census Bureau would like to sample from a single frame consisting of the Master Address File (MAF). The MAF contains addresses from the 2000 census and is continually updated with new construction addresses by the U.S. Postal Service's Delivery Sequence File (DSF) and other sources. In this paper, we measure the time interval between the permit issue date and the DSF version date to see if the DSF can provide adequate coverage for new construction. The results show that the median lag time is 7 months, and by 24 months past the permit issue date, we have received 95% of permit frame units on the DSF.

### Evaluation of Two Subcategories of the Delivery Sequence File

❖ Robert Colosi, U.S. Census Bureau, 9130 Whitney Way, Owings, MD 20736, *robert.w.colosi@census.gov*; Aliza Kwiat, U.S. Census Bureau

***Key Words:*** frame, housing unit coverage, American Community Survey, census master address file, address list

The Master Address File, a list of all housing units in the U.S., is being updated semi-annually with a file sent from U.S. Postal Service. This report will focus on two subcategories of addresses on that file, called the Delivery Sequence File. The subcategories are: Excluded from Delivery Statistics (EDS) records (most often representing new construction units that may eventually become residential, mailable addresses) and "waffling"/unstable addresses. A waffling address is one where the value of a descriptor variable is repeatedly changing over time. The goal of the analysis is to identify addresses that are less likely to represent housing units on the ground. Both the EDS and waffling addresses are eligible for sample selection for the American Community Survey (ACS). This study tracks these addresses through the ACS process and presents the validity rates associated with them.

### Identifying and Accounting for Mergers and Acquisitions in Measuring Employment

❖ Gordon Mikkelson, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 4985, Washington, DC 20906, *Mikkelson_G@bls.gov*; Linda Unger, Bureau of Labor Statistics; Doreen LeBel, Connecticut Department of Labor

***Key Words:*** business transfers, partials, predecessors, successors, linkages, scoring

The dynamic U.S. economy causes many challenges to accurately measuring employment and wages. One of the more difficult areas is accounting for mergers and acquisitions, also called predecessors and successors. Inaccurate or incomplete measurement leads to overestimation. This paper profiles research findings and describes implementation into the BLS' Quarterly Census of Employment and Wages (QCEW) program, which is the BLS sampling frame and source of a number of economic series. QCEW data are used to identify job growth and decline. The QCEW program is about to implement an initiative to improve the linkage of transfers of businesses from one record to another. This will clearly distinguish births and deaths from expansions and contractions, recognize possible duplicate or missing reporting, and better target potential linkages using wage records and scoring techniques.

## 118 Process Monitoring ●

Section on Quality and Productivity, Section on Physical and Engineering Sciences
**Monday, August 7, 8:30 am–10:20 am**

### Monitoring in the Presence of Dynamically Changing Observations

❖ Emmanuel Yashchin, IBM Research, Box 218, Yorktown Heights, NY 10598, *yashchi@us.ibm.com*

***Key Words:*** control charts, change points, cusum, reliability, sequential analysis, warranty

We consider the situation in which a relatively large portion of data that serves as a basis for a monitoring scheme changes with time. Unlike in the case of conventional control schemes, in which new data records generally pertain to monitored items that have been introduced recently, in our case a new data record can affect information related to items introduced at earlier stages. Situations of this type are common, for example, in the area of warranty data monitoring, in which incoming repair records from the field typically provide information related to reliability of earlier product vintages. In this paper we introduce a methodology for monitoring data of this type and discuss an example related to analysis of component failure rates based on warranty data.

### Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening

❖ James D. Williams, GE Global Research, 1 Research Circle, K14C27A, Niskayuna, NY 12309, *james.williams@research.ge.com*; Jeffrey B. Birch, Virginia Polytechnic Institute and State University; William H. Woodall, Virginia Polytechnic Institute and State University; Nancy Ferry, DuPont Crop Protection

***Key Words:*** bioassay, control chart, lack-of-fit, multivariate SPC

In agricultural crop product discovery, in vivo bioassay experiments are used to identify promising compounds for further research. A replicated dose-response of commercial crop protection products is assayed and used to monitor the bioassay quality. The activity of these compounds on the test organisms is characterized by a dose-response curve measured from the bioassay. If undesirable conditions in the bioassay arise then a bioassay monitoring procedure is needed to quickly detect such issues. In this paper we illustrate a proposed nonlinear profile monitoring method to monitor the variability of multiple assays, the adequacy of the dose-response model chosen, and the estimated dose-response curves for aberrant cases in the presence of heteroscedasticity. We illustrate these methods with in vivo bioassay data collected over one year from DuPont Crop Protection.

### Profile Monitoring via Linear Mixed Models

❖ Willis Jensen, W. L. Gore & Associates, Inc., 303 Piedmont 7, Blacksburg, VA 24060, *willisjensen@yahoo.com*; Jeffrey B. Birch, Virginia Polytechnic Institute and State University; William H. Woodall, Virginia Polytechnic Institute and State University

***Key Words:*** correlated data, linear mixed model, multivariate statistical process control, phase I, profile monitoring

Profile monitoring is a relatively new set of techniques in quality control used when the product or process quality is best represented by a function (or a curve) at each time period. The idea is often to model the profile via some parametric method and then monitor the estimated parameters over time to determine if there have been changes in the profiles. Previous modeling methods have not incorporated the correlation structure within the profiles. We propose the use of linear mixed models to monitor the profiles to account for the correlation structure within a profile. When the data are balanced the simple analysis that ignores the correlation structure will perform just as well as a more complicated analysis that takes into account the correlation structure. When the data are unbalanced or when there are missing data, we find that linear mixed model approach is preferable.

### Monitoring of the Process Mean with Double Sampling EWMA Control Charts

❖ Keunpyo Kim, PRA International, 4105 Lewis and Clark Drive, Charlottesville, VA 22911, *KimKeunpyo@praintl.com*; Marion R. Reynolds, Jr., Virginia Polytechnic Institute and State University

***Key Words:*** statistical process control, sequential sampling, double sampling, EWMA

When control charts are used to monitor a process, one approach to improving efficiency is to use a sequential sampling scheme in which the sample size at a sample point depends on the chart statistics at that sampling point as well as past sampling points. A double sampling (DS) scheme with two possible stages is one simple sequential sampling scheme. Most of the DS charts in the statistical process control literature are based on Shewhart type control charts. In this paper we evaluate the DS scheme when an exponentially weighted moving average (EWMA) control chart is being used to monitor the process mean. It is shown that using the DS EWMA control chart substantially improves the ability to detect a shift in a process mean compared to the standard control charts, DS X-bar charts, and other variable sampling rate Shewhart charts.

### Multivariate SPC for Batch Processes

Benjamin M. Adams, The University of Alabama; ❖ Young Kim, The University of Alabama, ISM Department, Box 870226, Tuscaloosa, AL 35487-0226, *ykim@cba.ua.edu*

***Key Words:*** quality control, multivariate, batch processes

The performances of the Hotelling's T2 control chart and the Squared Prediction Error (SPE) control chart based on multi-way principal component analysis are evaluated for batch processes. A stochastic model for simulated batch process data is provided. The model allows for cross correlation of error terms at a given time period and serial correlation of error terms across time periods. The performance characterizations of the two monitoring schemes are provided for a variety of cross correlation and serial correlation levels. The impact of the time period at which process shifts occur is also investigated for the monitoring schemes. The T2 control chart is recommended for the cases considered.

### Estimation of Process Parameters To Determine the Optimum Diagnosis Interval for Control of Defective Items

❖ Abhyuday Mandal, University of Georgia, Department of Statistics, Athens, GA 30602-1952, *amandal@stat.uga.edu*; Tirthankar Dasgupta, Georgia Institute of Technology

***Key Words:*** quality, statistical process control, economic model, online monitoring, MCMC, EM algorithm

The on-line quality monitoring procedure for attributes proposed by Taguchi has been critically studied and extended by a few researchers. Determination of the optimum diagnosis interval requires estimation of some parameters related to the process failure mechanism. Improper estimates of these parameters may lead to incorrect choice of the diagnosis interval and consequently huge economic penalties. Here we highlight both the theoretical and practical problems associated with the estimation of these parameters, and propose a structured approach to solve them. For the Case II model, two estimation methods, one based on Bayesian procedure and the other on the EM algorithm, are developed. A Bayesian method is proposed for estimation of parameters in Case III. A systematic way to utilize available engineering knowledge in eliciting the prior for the parameters is also discussed.

### An Extension of a Change Point Problem

❖ Albert Vexler, National Institute of Child Health & Human Development, 6100 Executive Blvd., Rm 7B05L, Bethesda, MD 20892, *vexlera@mail.nih.gov*; Chengqing Wu, National Institute of Child Health & Human Development; Aiyi Liu, National Institute of Child Health & Human Development; Brian W. Whitcomb, National Institute of Child Health & Human Development; Enrique F. Schisterman, National Institute of Child Health & Human Development

***Key Words:*** Doob decomposition, change point, martingale transforms, cusum, Shiryayev-Roberts statistics

We consider a specific classification problem in the context of change point detection. We present generalized classical maximum likelihood tests for homogeneity of the observed sample in a simple form which avoids the complex direct estimation of unknown parameters. The paper proposes a martingale approach to transformation of test statistics. For sequential and retrospective testing problems, we propose adapted Shiryayev-Roberts statistics in order to obtain simple tests with asymptotic power one. An important application of the developed methods is to the analysis of exposure's measurements subject to limits of detection in occupational medicine.

## 119 Mixture Modeling ●

Section on Nonparametric Statistics
**Monday, August 7, 8:30 am–10:20 am**

### Nonparametric Mixture Regression

❖ Alex Rojas, Carnegie Mellon University, 5000 Forbes Ave., 132 Baker Hall, Pittsburgh, PA 15213, *arojas@stat.cmu.edu*; Christopher Genovese, Carnegie Mellon University; Larry Wasserman, Carnegie Mellon University

***Key Words:*** conditional density estimation, local likelihood, mixture models, EM algorithm

Conditional density estimation is a techniques that allow for a better understanding of the relationship between a response variable and a set of covariates in comparison with usual regression methods. Therefore, this technique is of great importance in many scientific fields where knowledge about conditional means, obtained by regression methods, is not enough to draw valuable conclusions of the problem at hand. In this paper we present a conditional density estimator based on finite mixture models and local likelihood estimation, which has the advantage of being easily interpretable.

## A New Approach to Modeling the PET Data and Input Function

❖ Huiping Jiang, Columbia University, 1051 Riverside Drive, Unit 48, Unit 48, New York, NY 10032, *hj2179@columbia.edu*

***Key Words:*** PET, mixture model, input function

A mixture approach has been developed (Jiang and Ogden) to model the time-course PET data as a weighted linear combination of a finite number of basic functions (components) given by a sum of exponential functions convoluted with an arterial blood input function. The disadvantage of the approach is that the input function was modeled separately based on the biomedical sampling from subjects' arterial blood. We propose a new non-invasive approach to estimate the parameters of the mixture model and input function simultaneously without arterial data. The method is illustrated with analysis of both simulated and real data.

## Local Linear Regression by Mixture

❖ Weixin Yao, The Pennsylvania State University, 1003 W. Aaron Drive, Apt 10C, State College, PA 16803, *wxy116@psu.edu*

***Key Words:*** finite mixture model, longitudinal data, local linear regression, kernel GEEs

Local linear regression by mixture is an alternative nonparametric regression method. The idea of our method is to assume that the data comes from the multivariate normal mixture model (Note any density can be well approximately by normal mixture). Then in each component, condition on the predictors, the regression function is a linear function of the predictors. Without knowing the components label, the regression function is a weighted linear function of the predictors. We also compared this new method with local linear regression method using simulation. The new method is slightly better than local linear, especially in the boundary. For longitudinal data, we have tried several methods to incorporate the correlation structure to our new method. In the simulation study, we found our new method is at least as well as the widely used kernel GEEs method (Lin and Carrol 2000, 2001).

## A Mixture Model with Dependent Observations for the Analysis of CFSE-Labeling Experiments

❖ Ollivier Hyrien, University of Rochester, Elmwood Ave., Rochester, 14642, *ollivier_hyrien@urmc.rochester.edu*; Martin S. Zand, University of Rochester

***Key Words:*** branching process, mixture model, dependent data, cell proliferation

CFSE-labeling experiment has become a widely used assay to study the proliferation of cell population, and lymphocytes in particular. The analysis of such data has however received little attention so far. In this talk, we will present statistical methods appropriate for the estimation of cell kinetic parameters from CFSE-labeling data. The proposed methods rely on a continuous-time branching process that models the growth of the cell population, and a mixture model is proposed to describe the data. Because of the way cells are labeled, the observations are not independent, and the method of maximum likelihood cannot be used for parameter estimation. We propose alternative estimators, which are proven to be consistent and asymptotically Gaussian. The performances of the proposed methods are studied in simulation studies. An application to T lymphocytes is presented.

## Prevalence Estimation from Multiple Incomplete Lists Using the Rasch Model

❖ Changxuan Mao, University of California, Riverside, Department of Statistics, Riverside, CA 92507, *cmao@stat.ucr.edu*

***Key Words:*** capture-recapture, Rasch model, mixture model, conditional likelihood, population size, species richness

In epidemiological studies, a target population of individuals with disease might appear in multiple registration lists. However, these lists may be incomplete. Estimating the total population size from such incomplete list data is important to both scientific research and management. A Rasch model can be used to model these presence indicators, in which the logit of the presence probability is decomposed into individual catchability and list effect. A nonparametric Rasch model is considered. A family of lower bounds to the population size is proposed.

## Semiparametric Mixture Approach for the Measurement Error Problem in the Presence of Additional Error-Free Covariate

❖ Byungtae Seo, The Pennsylvania State University, 326 Thomas Building, University Park, PA 16802, *seobt@stat.psu.edu*; Bruce G. Lindsay, The Pennsylvania State University

***Key Words:*** measurement error, semiparametric, mixture, kernel estimator, Kullback-Leibler, MLE*

In usual statistical problems, our interest is to find functional relationship between response variable Y and covariate X. Measurement error problem occurs when measuring true covariate X is expensive or impossible to directly measure. In this case, sometimes another predictor W is available such that W represents the true predictor X but with some measurement error. We studied this measurement error problem when there are two types of covariates, one is measured with error and another is measured without error. It is known that semi-parametric method produces inconsistent estimators. The new method proposed based on kernel smoothing. The consistency of the estimators proposed is obtained under fairly mild conditions. A small simulation shows that usual semi-parametric mixture approach produces inconsistent estimators and suggests that the proposed estimators perform well.

## Nonparametric Mixture Model

❖ Mian Huang, The Pennsylvania State University, 303 Farmstead Lane, Apt 15, State College, PA 16803, *mzh136@psu.edu*

***Key Words:*** local likelihood, mixture of regression, EM algorithm

Methods of mixture model deal with data where observations are from several homogeneous subgroups. When the data has two or more dimensions and there are relationships between the responses and predictors, mixture of regression model is proper to use. The model is extended to a general case so that the nonlinear relationship in each component can be estimated. The estimation is based on optimizing the local likelihood function using EM algorithm. Simulation studies show that the model had some advantages.

## 120 Statistical Graphics: Applications in Drug Discovery and Clinical Development ● ✪

Section on Statistical Graphics, Biometrics Section

**Monday, August 7, 10:30 am–12:20 pm**

### Graphical Analysis of Clinical Data: Exploratory and Production Environments

❖ Matthew Austin, Amgen Inc., One Amgen Center Drive, MS 24-2-C, Thousand Oaks, CA 91320, *maustin@amgen.com*

***Key Words:*** graphics, interactive, trellis, exploratory

There are two common formats in which data are presented in clinical trial research: informal review of results with executive management and formal clinical study reports. Presentation in these areas can be vastly different. In executive review of data, there are less restrictions on the graphics' format and it is possible to use interactive plots to help facilitate discussion. In clinical study, reports of the graphic files must adhere strictly to individual company style guides that may restrict the type of file produced, size, color choices, etc. This presentation includes examples from the area of chemotherapy-induced anemia and provides guidance in the use of exploratory analyses, interactive graphics, and preparation of publication-ready graphics for clinical study reports.

### Statistical Graphics: Applications in Drug Discovery and Clinical Development

❖ C. George Rochester, U.S. Food and Drug Administration, Building 22, Room 5235, 10903 New Hampshire Avenue, Silver Spring, MD 20903, *george.rochester@fda.hhs.gov*; Charles K. Cooper, U.S. Food and Drug Administration; Mat Soukup, U.S. Food and Drug Administration; Ana Szarfman, U.S. Food and Drug Administration; Robert T. O'Neill, U.S. Food and Drug Administration

Recent issues with drug safety have resulted in sensitivity to analysis and reporting of safety data in clinical studies, as well as new guidance's for trial sponsors in this regard. As a result, more information has to be comprehended by and communicated among investigators, collaborators, institutional review boards, data safety management boards, and the FDA. This talk will review modern visualization methods for analysis, reporting, and regulatory review of safety data and provide suggestions for standard graphical analyses. Interactivity in the use of modern graphical displays and its value to further data exploration and analysis also will be incorporated. Examples from pre-marketing and postmarketing data will be included.

### Statistical Graphics for Effective Scientific and Business Decisionmaking throughout Drug Discovery, Development, Postmarketing, and Portfolio Management

❖ Thomas G. Filloon, Procter & Gamble, 8700 Mason-Montgomery Road, Mason, OH 45040, *filloon.tg@pg.com*

***Key Words:*** statistical graphics, decision making

Statistical graphics are crucial to the understanding and conversion of raw data into actionable information. In drug discovery, statistical graphics are used to perform quality control and display signal patterns to facilitate the understanding of true positives. In early phase clinical studies, statistical graphics play a key role in evaluating drug metabolism and pharmacokinetics. In later phase clinical studies, statistical graphics provide a highly efficient way of communicating the key safety and efficacy message across study teams and collaborators. In postmarketing, statistical graphics enable concise comparisons between products. This talk provides an overview of statistical graphics used in the Procter and Gamble discovery and development areas and gives insight into some of the clinical and business processes enabled through the use of statistical graphics.

## 121 Machine Learning and beyond: a Session in Memory of Leo Breiman

Memorial, Section on Statistical Computing, Section on Statistical Graphics, Section on Nonparametric Statistics

**Monday, August 7, 10:30 am–12:20 pm**

### Leo's Many Lives: Probabilist, Statistician, Applied and Theoretical, Machine-Learning Guru and Gadfly

❖ Peter J. Bickel, University of California, Berkeley, Department of Statistics, 367 Evans Hall, Berkeley, CA 94710-3860, *bickel@stat.berkeley.edu*

***Key Words:*** probability, applied statistics, theoretical statistics, machine learning

Leo's contributions to applied statistics and machine learning will be amply discussed by other speakers. Despite my lack of expertise, I'll talk briefly about his contribution to probability theory and my view of him as a highly important, if sometimes deliberately extreme, critic of the current theory and practice of statistics.

### Predictive Learning via Rule Ensembles

❖ Jerome H. Friedman, Stanford University, Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford, CA 94305, *jhf@stanford.edu*

General regression and classification models are constructed as linear combinations of simple rules derived from data. These rule ensembles are shown to produce predictive accuracy comparable to the best methods. However, their principal advantage lies in interpretation. Each rule is easy to understand, as is its influence on the predictive model. Similarly, the degree of relevance of each of the input variables can be assessed globally, locally in different regions of the input space, or at individual prediction points. Techniques are presented for automatically identifying those variables involved in interactions with other variables, the strength and degree of those interactions, and the identities of the other variables with which they interact. Graphical representations are used to visualize both main and interaction effects.

### Random Forests: Variable Importance and Proximities

❖ Adele Cutler, Utah State University, Department of Mathematics and Statistics, Logan, UT 84322-3900, *adele.cutler@usu.edu*

***Key Words:*** classification, machine learning, ensemble, bagging, support vector machines

Leo Breiman and I were working together on random forests from late in 2000 to his death in 2005. Random forests have been shown to be about as accurate as support vector machines, but they are more suited to statistical applications because they are interpretable. Variable importance can be measured both locally and globally. Proximities allow us to view the data in illuminating ways and are also useful for detecting outliers, imputing missing values, and extracting clustering information. This talk presents recent work on variable importance and proximities.

## 122 Balancing the Multiple Objectives of a Good Design of Experiments ● ✪

Section on Quality and Productivity, Section on Physical and Engineering Sciences
**Monday, August 7, 10:30 am–12:20 pm**

### The Effect of Cost and Other Constraints on Optimal Split-Plot Designs

❖ Alejandro Heredia-Langner, Pacific Northwest National Laboratory, P.O. Box 999, MSIN K5-12, Richland, WA 99352, *alejandro.heredia-langner@pnl.gov*; Timothy Robinson, University of Wyoming; Christine M. Anderson-Cook, Los Alamos National Laboratory; Douglas C. Montgomery, Arizona State University

*Key Words:* split-plot designs, genetic algorithms, optimization

The need to use split-plot experimental designs arises when one or more factors in an experiment are so difficult or costly to change that complete randomization in run order is not practical. For this reason, split-plot designs need to be constructed using optimization algorithms that take into consideration the restrictions in randomization and any other conditions (such as model order) particular to the problem. We present a case study that shows how optimal split-plot designs change under a variety of practical conditions.

### Evaluation Criteria for Second-Order Split-Plot Designs

❖ Peter A. Parker, National Aeronautics and Space Administration, M.S. 238, Hampton, VA 23681, *peter.a.parker@larc.nasa.gov*; Christine M. Anderson-Cook, Los Alamos National Laboratory; Timothy Robinson, University of Wyoming; Li Liang, Duke Clinical Research Institute

*Key Words:* split-plot design, restricted randomization, d-efficiency, design evaluation, IV-efficiency, response surface

When planning an experimental investigation, we often encounter practical or economic constraints that result in a restriction on randomization accommodated with a split-plot design structure. In this talk, case studies from aerospace research are used to illustrate split-plot design considerations and construction strategies. Candidate second-order, split-plot designs are evaluated based on their estimation and prediction efficiency. In addition, the clarity of structure, ease of execution, and simplicity of analysis of the competing designs are considered. Throughout this talk, subjective compromises among design evaluation criteria are emphasized. Developing effective experimental strategies in a restricted randomization context depends on our ability to balance multiple design evaluation criteria, thereby guiding the design selection process.

### Criteria for Designing Experiments: Some Practical Considerations

❖ Douglas C. Montgomery, Arizona State University, P.O. Box 875906, Department of Industrial Engineering, Tempe, AZ 85287-5906, *doug.montgomery@asu.edu*

*Key Words:* optimal design, design criteria, design evaluation

The many reasons for conducting experiments are discussed, along with the criteria one should consider in selecting an appropriate design. Alphabetically optimal designs are discussed, and both the advantages and shortcomings of this approach are illustrated. Graphical methods for design evaluation are highlighted.

## 123 Consistency and Convergence Rates for Bayesian Methods

Section on Bayesian Statistical Science
**Monday, August 7, 10:30 am–12:20 pm**

### On Posterior Consistency in Nonparametric Regression Problems

❖ Taeryon Choi, University of Maryland Baltimore County, Department of Mathematics and Statistics, 1000 Hilltop circle, Baltimore, MD 21250, *tchoi@math.umbc.edu*

*Key Words:* almost sure consistency, Hellinger metric, in probability metric, empirical probability measure, sieve, differentiable functions

We give sufficient conditions to establish posterior consistency in nonparametric regression problems with Gaussian error when suitable prior distributions are used for the unknown regression function and noise variance. When the prior under consideration satisfies certain properties, the crucial condition for posterior consistency is to construct tests that make the true parameter separate from the outside of the suitable neighborhoods of the parameter. Under appropriate conditions on the regression function, we show there exist tests of which the type I and type II error probabilities are exponentially small for distinguishing the true parameter from the complements of the suitable neighborhoods of the parameter. We consider two examples of nonparametric regression problems.

### Rates of Convergence for Posterior Distributions

❖ Stephen Walker, University of Kent, IMSAS, University of Kent, Canterbury, CT2 7NZ UK, *S.G.Walker@kent.ac.uk*

*Key Words:* posterior consistency, rate of convergence, prior concentration, Bernstein polynomial

We provide details of rates of convergence for posterior distributions in infinite dimensional models. This is achieved without the use of sieves, based on entropy measures, which have provided the basis for classical solutions and many recent Bayesian solutions. It turns out that a simple condition required for posterior consistency, which involves a finiteness of sums of square roots of prior probabilities, is all that is required for the rate of convergence to depend solely on how the prior concentrates about the correct sampling density. An optimal rate of convergence is obtained and examples presented.

### Recent Advances in Posterior Convergence Problem

❖ Subhashis Ghosal, North Carolina State University, Department of Statistics, 2501 Founders Drive, Raleigh, NC 27695-8203, *sghosal@stat.ncsu.edu*

***Key Words:*** posterior, consistency, convergence rate, infinite dimensional model

The study of posterior consistency and convergence rates of Bayes procedures in infinite dimensional models is challenging. Counterexamples show convergence may fail fairly easily. Conditions ensuring consistency were known from the mid-60s under the i.i.d. set-up, and were generalized in the 90s. Lately, these ideas have been extended to study the posterior convergence rates in the infinite dimensional setting. Attention has been given to generalizing the results to incorporate non-i.i.d. observations. We will present results in that direction. Recently, an alternative approach to posterior convergence via a martingale property of integrated likelihood was considered in the literature. We also will discuss relations of some of our results with the ones obtained by the martingale approach.

## 124 Theory and Analysis of Complex Networks ● ✪

IMS, Section on Statisticians in Defense and National Security, Section on Statistical Graphics

**Monday, August 7, 10:30 am–12:20 pm**

### Community Detection in Large Networks

❖ Jörg Reichardt, University of Bremen, Complex Systems Lab, Otto-Hahn-Allee, 28359, *reichardt@itp.uni-bremen.de*

The problem of community detection or graph clustering has received considerable attention in recent years for its wide range of applications in exploratory data analysis and hypothesis formation for relational data sets. By using an analogy with a statistical mechanics problem, the so called Potts model, the talk will shed some light onto the nature of the problem and provide insights into the structure of its solutions. A concise definition of community, hierarchical and overlapping community structures will be given and comparisons with alternative definitions and community detection methods are made. An application to a large real world network with close to one million nodes will be shown and the statistical significance of the findings will be discussed.

### Semimetric and Scale-Free Structure in Document Networks

❖ Luis Rocha, Indiana University, Informatics Building, 901 E. 10th Street, Informatics Department, Bloomington, IN 47408-3912, *rocha@indiana.edu*

This talk will describe the nature of semimetric structures in document networks; their relation to scale-free structures; and their uses in text mining, information retrieval, etc.

### Exponential-Family Models of Random Graphs for Social Networks

❖ Martina Morris, University of Washington, Box 353340, Seattle, WA 98195-3340, *morrism@u.washington.edu*

This talk will discuss recent advances in the use of exponential-family random graphs to model social networks, including maximum-likelihood estimation via Monte Carlo, the nature of the appropriate sufficient statistics, and the ability to incorporate the effects of different kinds of generative mechanisms in these models.

### Depth Partitions and Hierarchical Structure in a Tokyo Industrial District

Tsutomu Nakano, Kwansei Gakunin University; ❖ Douglas White, University of California, Irvine, 3151 Social Science Plaza, Department of Anthropology,, Irvine, CA 92697, *drwhite@uci.edu*

We will explore implications of the Harrison White theory (Markets from Networks: Socioeconomic Models of Production) using a network statistics analysis of an industrial district of 8,500 firms in Ohta District, Tokyo. We explore also networks-affect-prices theory and predictive cohesion theory.

## 125 Real-Life Problems Advancing Missing Data Techniques ●

Biopharmaceutical Section, Biometrics Section, Section on Health Policy Statistics

**Monday, August 7, 10:30 am–12:20 pm**

### Missing Data in Smoking Cessation Trials

❖ Michael D. Larsen, Iowa State University, Department of Statistics, 220 Snedecor Hall, Ames, IA 50011, *larsen@iastate.edu*; Sunni A. Barnes, Mayo Clinic College of Medicine; Paul Decker, Mayo Clinic College of Medicine; Darrell Schroeder, Mayo Clinic College of Medicine; Robert Klesges, Mayo Clinic College of Medicine; Mark W. Vander Weg, Mayo Clinic College of Medicine

***Key Words:*** hot deck imputation, propensity score matching, multiple imputation, drop out, nonignorable missingness, nonresponse

Standard practice assumes a participant in a smoking cessation trail who does not report a smoking status six months after the end of treatment is smoking. Using data from an Air Force study of smoking cessation programs in basic training that has virtually no missing data, a simulation is conducted to study the impact of assumptions about missing data probability models and analysis procedures. Missing data indicators are generated using MCAR, MAR, and NMAR assumptions. Results using the standard assumption or imputation of smoking are compared to results using hot deck matching, propensity score matching, and multiple imputation modeling. A further analysis treating the late responders as missing cases---which they would have been had fewer contact attempts been made---also is reported. The value of carefully considering predictors of outcomes is demonstrated clearly.

### A Novel Pattern Mixture Model Applied to QOL Data

❖ Page C. Moore, University of Arkansas for Medical Sciences, Dept. of Biostatistics, 4301 West Markham #781, Little Rock, AR 72205-7199, *pagemoore@yahoo.com*; John W. Seaman, Baylor University; Sunni A. Barnes, Mayo Clinic College of Medicine

***Key Words:*** multiple imputation (MI), missing data, NMAR, pattern mixture model, discrete longitudinal data, quality-of-life (QOL)

Clinical trial endpoints are traditionally either physical or laboratory responses. However, such endpoints fail to reflect how patients feel or function in their daily activities. Quality of life issues often contribute to patient attrition in clinical trials. Many researchers often resort to ad hoc methods (e.g. case-deletion or mean imputation) when they are faced with missing data, which can lead to biased results. An alternative to these ad hoc methods is multiple imputation. Pattern mixture models, involving restriction methods and multiple imputation, are often used to handle data under the Not Missing at Random (NMAR) assumption. The purpose of this presentation is to introduce pattern mixture models, restriction methods, and to introduce a new pattern-mixture model. The new method will be illustrated using quality-of-life data.

### The Impact of Missing Data and How It Is Handled on the Rate of False Positive Results in Drug Development

❖ Stacy Lindborg, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *slindborg@lilly.com*; Craig Mallinckrodt, Eli Lilly and Company; Michael K. Carter, Eli Lilly and Company; Sunni A. Barnes, Mayo Clinic College of Medicine

***Key Words:*** missing data, regulatory risk, Bayesian, multiple imputation, MMRM

In drug development a common choice for the primary analysis is to assess mean changes via analysis of (co)variance with missing data imputed by carrying the last or baseline observations forward (LOCF, BOCF). These approaches assume data are missing completely at random (MCAR). Multiple imputation (MI) and likelihood-based repeated measures (MMRM) are less restrictive as they assume data are missing at random (MAR). We report results from a simulation study that compared the rate of false positive results from LOCF, BOCF, MI, and MMRM. These results illustrate the MAR methods provide better control of false positive rates than the MCAR methods. We discuss the need for and benefit from using MAR methods as the primary analysis in drug development.

## 126 JASA Theory and Methods Invited Paper Session

JASA, Theory and Methods
**Monday, August 7, 10:30 am–12:20 pm**

### Quantile Autoregression

❖ Roger Koenker, University of Illinois, Champaign, IL 61820, *rkoenker@uiuc.edu*

***Key Words:*** quantile regression, random coefficient, autoregressive model

We consider quantile autoregression (QAR) models in which the autoregressive coefficients can be expressed as monotone functions of a single, scalar random variable. The models can capture systematic influences of conditioning variables on the location, scale, and shape of the conditional distribution of the response, and therefore constitute a significant extension of classical constant coefficient linear time series models in which the effect of conditioning is confined to a location shift. The models may be interpreted as a special case of the general random coefficient autoregression model with strongly dependent coefficients. Statistical properties of the proposed model and associated estimators and inference methods are studied. Empirical applications of the model to the U.S. unemployment rate and U.S. gasoline prices highlight the potential of the model.

## 127 Spatial Modeling To Address Problems in Ecology ● ✪

Section on Statistics and the Environment, WNAR
**Monday, August 7, 10:30 am–12:20 pm**

### Geostatistical Modeling: Model Selection and Parameter Estimation

❖ Jennifer A. Hoeting, Colorado State University, Department of Statistics, Fort Collins, CO 80526-1877, *jah@lamar.colostate.edu*

***Key Words:*** Aikaike information criterion, model selection, spatial correlation, matern and exponential autocorrelation function

Spatial correlation often is ignored in the selection of explanatory variables and can influence model selection results. We show that using AIC for a geostatistical model is superior to the more traditional approach of ignoring spatial correlation in the selection of explanatory variables. We explore several key asymptotic assumptions required for AIC, related to properties of the parameter estimates. We provide simulations that support the theoretical results and discuss implications of these results for designing sampling plans and implementing model selection for geospatial data. Software for implementing the geostatistical model selection methods described is available at www.stat.colostate.edu/~jah.

### Markov Chain Monte Carlo for a Spatial-Temporal Autologistic Regression Model

❖ Jun Zhu, University of Wisconsin-Madison, Department of Statistics, 1300 University Avenue, Madison, WI 53706, *jzhu@stat.wisc.edu*

***Key Words:*** autologistic model, Gibbs sampler, perfect simulation, spatial-temporal model

A spatial-temporal autologistic regression model developed by Zhu et al. (2005) relates a binary response variable to potential covariates while accounting for both dependence on a spatial lattice and dependence over discrete time points, which may be useful for analyzing spatial-temporal binary data. However, the existing statistical inference is via maximum pseudolikelihood, which is statistically inefficient---especially when the spatial and temporal dependence are strong. Here, we propose a fully Bayesian approach for both model parameter inference and prediction at future time points using Markov chain Monte Carlo (MCMC). We demonstrate the methodology and compare the results with maximum pseudolikelihood and MCMC maximum likelihood approaches via a real data example concerning beetle outbreaks.

### Application and Comparison of Alternative Spatial Modeling Approaches of Zero-Inflated Data

❖ Mary C. Christman, University of Florida, Department of Statistics, IFAS, 406 McCarty C, Gainesville, FL 32611-0339, *mcxman@ufl.edu*

***Key Words:*** zero-inflated, Poisson model, negative binomial model, abundance estimation, prediction performance

Many spatially indexed datasets display zero-inflated frequencies; that is, the data arise from a mixture of structural zeroes and a probability distribution for the nonzero values. Examples are rampant in the ecological literature, especially where the data are species richness or abundance counts. This talk compares the prediction performances of several alternative spatial modeling approaches for lattice data, including the spatial zero-inflated Poisson model; two-stage fitting procedures in which a binary model is first fitted to determine presence and then, conditional on the first outcome, a model for the counts is fitted; Bayesian hierarchical approaches; and the negative binomial spatial model.

## 128 Cross-National Research on Immigration ● ✪

Social Statistics Section
**Monday, August 7, 10:30 am–12:20 pm**

### Measuring Migration between Canada and the United States: a Case Study for Best Practices

Melissa L. Therrien, U.S. Census Bureau; ❖ Margaret Michalowski, Statistics Canada, Ottawa, ON K1A 0T6 Canada, *Margaret.Michalowski@statcan.ca*

***Key Words:*** emigration, international migration, estimates, U.S., Canada, best practices

Sending countries require information about numbers and characteristics of their residents involved in international movement. This paper discusses opportunities and challenges encountered when using census and survey data from receiving countries to measure emigration. Discussion is based on an evaluation of American Community Survey data on Canadians living in the United States. Statistics on migrants from Canada to the United States are developed using data from a question on place of residence one year ago and data on year of arrival to the United States. They are then assessed against Canadian sources on emigrants to the United States from administrative data sources. The paper concludes with a discussion about whether the Canadian-U.S. collaboration has applicability for measuring emigration across other countries.

### Counting Immigrants and Expatriates: a New Perspective

❖ Jean-Christophe Dumont, Organisation for Economic Co-operation and Development, 2, rue AndrÈ Pascal, F-75775 Paris Cedex 16, France, *Jean-Christophe.dumont@oecd.org*; Georges LeMaitre, Organisation for Economic Co-operation and Development

***Key Words:*** international migration, immigrants, expatriates, emigration

The mobility of the highly skilled has been high on the policy agenda in many countries. Although shortages in information and communications technology professionals have abated, "brain drain" issues continue to attract considerable policy attention. There is considerable discussion concerning the issues mentioned here, but there seems to be a shortage of solid international statistics to support the work and debate. We have embarked a joint compilation project to collect available data on the stock of immigrants in selected countries, cross-classified by a number of relevant variables, including country of origin, educational attainment, and occupation. In this paper, we discuss how this work helps to advance the cause of international comparative statistics for use in the international debate on the migration of the highly skilled.

### Mexican Immigration to the United States: Challenges and Options

❖ Jeffrey S. Passel, Pew Hispanic Center, 1615 L. Street NW, Washington, DC 20036, *jpassel@pewhispanic.org*

***Key Words:*** Illegal immigration, demographic structure, projections

Mexico has been the largest source of both legal and unauthorized migration to the United States for four decades. As a result, the Mexican born population of the United States grew from 760,000 in 1970 to more than 11.6 million in 2005; it represents almost one-third of the immigrants living in the United States. With this astounding growth, about 10% of Mexico's population is in the United States. The paper describes principal characteristics of Mexican migrants--demographic structure, education, income, geographic distribution, and legal status--and recent changes. Migration flows have contributed to significant demographic changes within Mexico. The paper closes with an assessment of migration trends and impacts on population size and age structure in both countries over the next fifty years.

## 129 Subset Selection in Random Effects and Variance Component Models ●

ENAR, Biometrics Section, WNAR
**Monday, August 7, 10:30 am–12:20 pm**

### A New Class of Model Selection Strategies with Applications in Mixed Model Selection

❖ Jiming Jiang, University of California, Davis, 1 Shields Ave., Davis, CA 95616-5270, *jiang@wald.ucdavis.edu*

Many model search strategies involve trading off model fit with model complexity in a penalized goodness-of-fit measure. Such strategies do not always translate into good finite sample performance, as the procedure is typically overly sensitive to the setting of penalty parameters. Furthermore, these strategies do not generalize naturally to more complex models and nonconventional problems, in which the penalties and model complexity may not be defined naturally. In this talk, we introduce a new class of model selection strategies that are potentially less subjective and applicable to both conventional and nonconventional problems. We apply these strategies to linear and generalized linear mixed model selection problems.

### Variance Components and Reliability in FMRI Studies

❖ Hal Stern, University of California, Irvine, Department of Statistics, 346D Computer Science, Irvine, CA 92697-1250, *sternh@uci.edu*; Sandip Sinharay, Educational Testing Service

***Key Words:*** nested design, Bayesian inference, random effects

It is of great interest to combine functional magnetic resonance imaging study data across institutions. As part of one such multi-site study, data were collected to assess intra- and inter-site sources of variability. The analysis of these data raises important questions about the assessment and interpretation of random effects and variance components.

### BART: a Nonparametric Random Effects Model

Hugh A. Chipman, Acadia University; ❖ Edward I. George, University of Pennsylvania, The Wharton School, Statistics

Department, 3730 Walnut Street, 446 JMHH, Philadelphia, PA 19104-6340, *edgeorge@wharton.upenn.edu*; Robert E. McCulloch, The University of Chicago

***Key Words:*** Bayesian backfitting, boosting, CART, MCMC, sum-of-trees model, weak learner

We propose using BART (Bayesian Additive Regression Trees) to discover the nature of the relationship between y and a set of potential predictors before attempting to find a relevant subset and suitable parametric model. BART approximates the conditional mean of y by a sum of regression trees where each tree is constrained by a regularization prior to be a weak learner, as in boosting. BART is effectively a nonparametric random effects model with a dynamic adaptive random basis. Fitted by an iterative backfitting MCMC algorithm, BART appears to be remarkably effective at finding highly nonlinear relationships hidden within a large number of irrelevant potential predictors.

## 130 Inference under Biased Sampling ●

General Methodology

**Monday, August 7, 10:30 am–12:20 pm**

### Estimation of Treatment Effects in Observational Studies

❖ Danny Pfeffermann, Hebrew University/University of Southampton, Department of Statistics, Hebrew University, Mount Scopus, Jerusalem, 91905 Israel, *msdanny@huji.ac.il*

***Key Words:*** randomized experiment, propensity scores, instrumental variables, sample distribution, sample surveys, PISA

In observational studies the assignment of units to treatments is with unknown probabilities, and the units exposed to the various treatments could differ in important unknown characteristics. Consequently, estimation and comparison of treatment effects based on the empirical outcome distributions can be misleading. We consider an approach to deal with this problem that attempts to approximate the parametric distribution of the observed responses under a given treatment as a function of the distribution that would be obtained under strongly ignorable assignments and the assignment rule, which is modeled as a function of the observed responses and covariates. The use of this approach is founded by showing that the distribution of the observed responses is identifiable under fairly general conditions. The goodness of fit of this distribution can be tested using simple test statistics

### Maximum-Likelihood Inference on a Mixed Conditionally and Marginally Specified Regression Model in Genetic Epidemiologic Studies with Two-Phase Sampling

❖ Nilanjan Chatterjee, National Cancer Institute, 6120 Executive Blvd., EPS 8038, Rockville, MD 20852, *chattern@mail.nih.gov*; Yi-Hau Chen, Academia Sinica

***Key Words:*** case-control study, gene-environment interaction, missing data, outcome dependent sampling, semiparametric estimation

Two-phase sampling can reduce the cost of genetic epidemiologic studies by limiting expensive ascertainments of genetic and environmental exposure to an efficiently selected subsample (phase II) of the main study (phase I). Case-control status, family history, and other covariate information may be used for sampling of informative subjects at phase II. We develop a maximum-likelihood method for analysis of data from such studies using a novel regression model that permits the estimation of 'marginal' risks associated with the genetic and environmental covariates while simultaneously characterizing the 'conditional' risk of the disease associated with family history after adjusting for the covariates. The methods are developed with and without an assumption of gene-environment independence. Performances of the methods are studied using both real and simulated data.

### The Design and Analysis of Two-Phase Sampling

❖ Alastair J. Scott, The University of Auckland, 36 Princes Street, Auckland, 1 New Zealand, *scott@stat.auckland.ac.nz*; Christopher J. Wild, The University of Auckland; Yannan Jiang, The University of Auckland

***Key Words:*** two-phase sampling, semiparametric efficiency, auxiliary information, biased sampling

In many health surveys, information about some characteristics is available for all units in the population; an example is the Auckland Collaborative Birthweight Study. Birthweights and many other characteristics are collected routinely for all mothers and babies in the target population. Other characteristics can be measured for a sample only. We want to build a regression model for birthweight as a function of the other characteristics. The task is to design an efficient sampling scheme and method for fitting the model, using all the information available. We will look at methods for computing efficient semiparametric estimates and alternatives based on standard survey techniques for using auxiliary information. These alternative methods are reasonably efficient and relatively robust and simple to implement. We also look at some aspects of design.

### Weighted Likelihood for Semiparametric Models and Two-Phase Stratified Samples with Application to Cox Regression

❖ Norman E. Breslow, University of Washington, Department of Biostatistics, Mail Stop 357232, Seattle, WA 98195-7232, *norm@u.washington.edu*; Jon A. Wellner, University of Washington

***Key Words:*** case-cohort, estimated weights, failure time, inverse probability weights, missing data

Inverse probability weighted (IPW) likelihood equations lead to simple and robust estimators for two-phase stratified samples when used with semiparametric models having regular root N consistent estimators of Euclidean and nonparametric parameters. Under Bernoulli sampling, the influence function for the weighted likelihood estimator of the Euclidean parameter is the IPW version of the ordinary influence function. By proving weak convergence of the IPW empirical process and borrowing results on weighted bootstrap empirical processes, we show a parallel asymptotic expansion holds for finite population stratified sampling. Our key results were derived earlier for the special case of Cox regression with stratified case-cohort studies; other complex survey designs and missing data problems are discussed more generally. The paper interprets this prior work and paves the way for other applications.

# 131 Having an Impact in a Multidisciplinary Setting ● ✪

Section on Statistical Consulting, Section on Statistical Education, Section on Teaching Statistics in the Health Sciences

**Monday, August 7, 10:30 am–12:20 pm**

## Having an Impact in a Multidisciplinary Setting

❖ Janice Derr, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993, *Janice.Derr@fda.hhs.gov*; ❖ Janet Powell, University of Washington, Division of Occupational Therapy, Box 356490, Seattle, WA 98195-6490, *jmpowell@u.washington.edu*; ❖ Lillian Lin, Centers for Disease Control and Prevention, Atlanta, GA , *lel5@cdc.gov*; ❖ Kevin Cain, University of Washington, Seattle, WA , *Cain@u.washington.edu*; ❖ W. Scott Clark, Eli Lilly and Company, Indianapolis, IN 46284, *Clark_W_Scott@Lilly.com*

***Key Words:*** collaboration, statistical consulting, multi-disciplinary, inter-disciplinary

Statisticians who work in multidisciplinary settings have the opportunity to promote good statistical practices and statistical thinking wherever they work. A full involvement can make the most of a statistician's expertise. However, it is challenging to work effectively across disciplines. Based on input from statisticians in this situation, "consulting tales" of successful and not-so-successful experiences were developed. Panel members, four statisticians, and one nonstatistician investigator represent biomedical and epidemiological fields from government, academia, and industry. Panelists will use the consulting tales, along with their own experiences, to identify issues that are often the culprits when a statistician does not feel fully involved in a project. They also will propose more effective approaches.

# 132 Advances in Confidence Intervals on Variance Components ● ✪

Section on Health Policy Statistics

**Monday, August 7, 10:30 am–12:20 pm**

## A Comparison of Two Confidence Interval Approaches on the Dependability Coefficient in a Two-Factor Crossed Design

❖ Joseph Cappelleri, Pfizer Inc., Eastern Point Road, MS 8260 2222, Groton, CT 06340, *joseph.c.cappelleri@pfizer.com*; Naitee Ting, Pfizer Inc.

***Key Words:*** generalizability theory, variance components, confidence intervals, modified large-sample, analysis of variance, dependability coefficient

Generalizability theory employs analysis-of-variance methods and capitalizes on variance components in allowing a researcher to untangle multiple sources of error variance. For decisions based on the absolute level of performance among individuals, the dependability coefficient---a ratio of variance components---is used as a measure of reliability. A simulation study is conducted to investigate and compare the confidence interval coverage for the dependability coefficient of two modified large-sample approaches in a two-factor (one-facet) crossed design. One is proposed by Arteaga, Jeyaratnam, and Graybill (AJG) and another proposed by Cappelleri and Ting (CT). Both methods meet the nominal coverage, but the CT method tends to be more accurate and has a shorter interval width, making it preferable. Both methods are illustrated with examples.

## Confidence Intervals on General Variance Components Model: Modified Large Sample Approach

❖ Yonghee Lee, Ewha Womans University, 111 Daehyundong SedaemunGu, Seoul, 120-750 South Korea, *ylee@ewha.ac.kr*

***Key Words:*** variance components, confidence intervals, unbalanced designs

Inference on variance components is not straight forward because of the complex sampling distributions of ANOVA-type estimators. Especially in unbalanced designs, it is more difficult to make inferences on the function of variance components, as the sum of squares from ANOVA may not follow chi-square distribution or sum of squares are not independent. By using the results of Lee, Shao, and Chow (2004), I will introduce an extension of Modified Large Sample (MLS) method for construction of confidence intervals on a linear combination of variance components when ANOVA sum of squares do not follow a chi-square distribution or when estimators are dependent. The proposed method is based on an approximation to a linear combination of chi-square random variables whose degrees of freedom are one and bootstrap calibration. I will discuss applications to nested unbalanced designs as well.

## Confidence Intervals on Intermediate Precision Measures in Analytical Method Validations and Transfers

❖ Richard Burdick, Amgen Inc., 4000 Nelson Road, Mail Stop AC22A, Longmont, CO 80503, *rburdick@amgen.com*; Shea Watrin, Amgen Inc.

***Key Words:*** variance components, random models, mixed models

An important characteristic used to evaluate analytical method validations and transfers is intermediate precision (IP). IP is a measure of variation when analysts perform the method on different days or with different equipment. It is often appropriate to consider analyst as a random effect in such studies. However, it is usually not possible to sample more than two or three analysts, and resulting confidence intervals on IP are too wide to be useful. In this study, we consider three methods to model the analyst effect: random effect, fixed effect, and pooled with repeatability. We develop methods for constructing confidence intervals on IP under each condition using generalized inference and closed-form approximations. Recommendations are provided for selecting an appropriate method based on computer simulations.

## Confidence Intervals of a Common Intraclass Correlation Coefficient

❖ Lili Tian, University of Buffalo, 249 Farber Hall, 3435 Main St Bldg 26, Buffalo, NY 14214-3000, *ltian@buffalo.edu*

The intraclass correlation coefficient (ICC) has been used widely in many fields of research. For a single-sample problem, the inference theory concerning an ICC has been well-developed. In real life, researchers often collect information on familial aggregation in multiple samples. Therefore, we often need to make inference concerning ICC from several samples. This talk will present a novel approach for the confidence interval estimation of a common ICC derived from several samples under unequal family sizes. The proposed approach is developed using the concept of generalized pivots. Comparisons are made with a large sample procedure on the coverage probabilities.

## 133 Recent Advances in Small-Area Issues ●

Section on Survey Research Methods
**Monday, August 7, 10:30 am–12:20 pm**

### Small-Area Modeling for Survey Data with Smoothed Error Covariance Structure via Generalized Design Effects

❖ Ralph Folsom, RTI International, 3040 Cornwallis Road, Post Office Box 12194, Research Triangle Park, NC 27709-2194, *ref@rti.org*; Avinash C. Singh, Statistics Canada; Akhil Vaish, RTI International

***Key Words:*** estimating functions, generalized deffs, ignorable and nonignorable designs, unstable estimated error covariance

We consider specifying the design-based error covariance structure in small-area modeling with survey data. While it is customary to treat the estimated covariance as known, it is often unstable. To alleviate this problem, one can either model the design-based covariance matrix or smooth the estimated covariance by specifying only its mean function. We prefer smoothing over modeling because of the strong assumptions needed to model the error covariance structure of SAEs. To smooth nondiagonal error covariance matrices, we make use of the g-deff (generalized design effect), defined earlier by Rao and Scott (1981) in the context of categorical data analysis. Simulation results for SAEs based on a linear mixed model show that the author's EFGL method (FCSM 2005) with the proposed smoothing provides improved coverage of confidence intervals.

### Hierarchical and Empirical Bayesian Analysis of Bivariate Binary Data: an Application to Small-Area Estimation

❖ Ananya Roy, University of Florida, Department of Statistics, 103 Griffin Floyd Hall PO Box 118545, Gainesville, FL 32611, *aroy2@stat.ufl.edu*; Malay Ghosh, University of Florida

***Key Words:*** estimating functions, infant mortality rates, mean squared error, second order correct

This paper addresses small-area estimation problems when the response is bivariate binary. Empirical Bayes estimators of small-area means are provided and the associated asymptotic mean squared errors and estimated mean squared errors found. The results are applied to the analysis of a real dataset, and the superiority of the empirical Bayes estimators over the direct estimators is established. We also provide a hierarchical Bayesian analysis and compare our results.

### Robust Estimation of the Mean Square Error of an EBLUP of a Small-Area Mean

❖ Shijie Chen, RTI International, 701 13th Street, NW, Suite 750, Washington, DC 20005-3967, *schen@rti.org*; Partha Lahiri, University of Maryland; Jon N. K. Rao, Carleton University

In this paper, we present a general method for estimating the mean square error (MSE) of EBLUP for the well-known Fay-Herriot small-area model. Unlike the normality-based MSE approximations, our result involves the kurtosis (but not the skewness) of both the sampling and model errors and depends on the method of estimating variance component. For the method of moments estimator of the variance component, estimation of the kurtosis of the model error is not required. Estimation of the kurtosis is necessary for other methods of estimating variance components (e.g., the method proposed by Fay and Herriot (1979)). We propose a method for estimating the kurtosis that provides a method of MSE estimation for a general class of variance component estimators with the non-normal sampling and model errors.

### Small-Area Estimation Using Nonparametric Regression

❖ Tathagata Bandyopadhyay, Indian Institute of Management Ahmedabad, Wing14 Vastrapur, Ahmedabad, 380015 India, *tathagata@iimahd.ernet.in*

***Key Words:*** small area estimate, nonparametric regression

Small-area estimates are being used increasingly by policymakers for making important decisions. However, small sample sizes---often available for small areas---make the estimates unreliable. An often used approach for producing reliable small-area estimates is to employ a mixed linear model. It borrows strength from survey data collected in other small areas for getting an estimate of a small area. The success of this model-based approach depends largely on the choice of the model and, of course, the availability of good auxiliary data. We address the question of choosing a correct regression model. More specifically, we consider finding small-area estimates using a nonparametric model.

### Bayesian Estimation of the Percentiles of Overweight

❖ Jai Choi, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, *jwc7@cdc.gov*; Balgobin Nandram, Worcester Polytechnic Institute

***Key Words:*** finite population percentile, logistic spline regression, selection bias, nonresponse bias

We analyze body mass index (BMI) data from the third National Health and Nutrition Examination Survey (NHANES III), which assesses the health status of the U.S. population. We studied two important features in these data. First, because there are no BMI values for a considerable number of the children and adolescents under study, there was serious chance for nonresponse bias. Second, there are differential probabilities of election. To assess the status of overweight and obesity in children and adolescents, we predicted the finite population percentile BMI for domains (small areas) formed by age, race, and sex in 35 large counties and showed how to include the selection probabilities into a nonignorable nonresponse model, which is difficult because there is no simple functional relation between the BMI values and the selection probabilities.

## 134 Meta-analysis in Survey Research: Analysis of Multiple Response Rates and Other Applications ●

Section on Survey Research Methods
**Monday, August 7, 10:30 am–12:20 pm**

### Differences in Survey Nonresponse Components over Time in Various Domains

❖ Brad Edwards, Westat, 1650 Research Blvd., Rockville, MD 20850, *bradedwards@westat.com*; David Cantor, Westat; Lisa Moses, Westat

***Key Words:*** survey participation, minorities, refusal rates, unlocatable rates

This paper examines response rate trends by race/ethnicity on several Westat studies conducted since the early 1990s. Analysis is restricted to in-person studies that provide some means of determining response rates by minorities independent of coverage issues (i.e., where minority status is known even for the noncontacts). In many cases, minority participation rates are higher than the national mean, but there is considerable variation in nonresponse components among minorities (refusal rates are higher for some; unlocatable rates are higher for others), and we examine these differences and determine whether the general decline observed in study participation over the past decade is mirrored by overall declines in minority group participation.

### Monitoring Survey Response Rates and Measuring Nonresponse Bias in a Federal Statistical Agency: National Center for Education Statistics (NCES)

❖ Marilyn Seastrom, National Center for Education Statistics, 1990 K Street, NW, Washington, DC 20006, *Marilyn.Seastrom@ed.gov*

***Key Words:*** nonresponse bias analysis, survey response rates, data quality, statistical standards, meta-analysis

In 2005, OMB posted statistical standards in the Federal Register. Included is a requirement to conduct nonresponse bias analysis when response rates fall below established cut points. Related to this, the 2002 NCES Statistical Standards includes four standards on survey nonresponse: computation, parameters, achieving acceptable rates, and nonresponse bias analysis. NCES monitors survey response rates and nonresponse bias analyses. The focus is on the unit level response rates, as the nonresponse bias analysis requirements are tied to specific unit-level target parameters. This analysis will summarize the 3.5 years of experience at NCES monitoring response rates and related nonresponse bias analyses. Included will be a description of the approaches used in the nonresponse bias analyses and a summary of the response rates and findings from the different analyses.

### Sources of Variation in Response Rates to the Behavioral Risk Factor Surveillance System

❖ David W. Smith, The University of Texas School of Public Health, 13138 Queens Forest Street, San Antonio, TX 78230, *smithd2@uthscsa.edu*

***Key Words:*** survey methods, response rates, BRFSS, data quality

The BRFSS is an RDD telephone survey conducted by state and territorial health departments and sponsored by the Centers for Disease Control. Sample and questionnaire design are similar in all states in any year. In recent years, the number of respondents has been about 2,000--6,000, totaling 200,000--300,000 from all states. The BRFSS reports several indicators of the quality of survey execution for each state, including cooperation rates, response rates, and call disposition codes. A public-use data file includes item-specific missing data codes for refused and don't know/not sure. Respondents who complete partial surveys may be identified from the data. Variation among the indicators of survey quality exceeds chance across states and years and is identified with both state and personal characteristics that relate to survey quality.

### Meta-analysis in the Human Sciences

❖ Mack Shelley, Iowa State University, 323 Snedecor Hall, Ames, IA 50011-1210, *mshelley@iastate.edu*

***Key Words:*** human sciences, meta-analysis, health research, education research, nonnormality

The use of meta-analysis in research in the human sciences spans topics in education, the social sciences, health, physical activity, nutrition, and related disciplines. This paper offers an overview of the use of meta-analysis methods in the human sciences, focusing on some of the uses---and problems with---the use of meta-analysis in education and human health research. We address the results from meta-analysis of 334 findings from 37 studies of the effect of physical fitness interventions on strength, stamina, and endurance outcomes, controlling for placebo/treatment/control statistical main effects and with covariates for age, gender, randomization, double-blind procedures, study duration, training load, training frequency, and training status. We address issues related to weighting and the use of ranked or transformed effect sizes to deal with lack of normality.

## 135 Recent Advances in Bayesian Computation and Bioinformatics ● ✪

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### Model-Based Correlations: a Tool for Revealing Interactions in Microarray Data

❖ Beatrix Jones, Massey University, Institute of Information and Mathematical Sciences, Private Bag 102 904 N Shore Mail Centre, Auckland, 0000 New Zealand, *m.b.jones@massey.ac.nz*; Marie Fitch, Massey University

***Key Words:*** covariance selection, factor analysis, graphical models, microarray data

Correlation is a basic and popular tool to look for potential interactions among many continuous variables, as occur in a microarray experiment. As well as thresholded raw correlations, one can use model-based correlations (e.g. derived from Bayesian covariance selection or factor analysis models). Interactions also can be examined using the inverse covariance, which represents the conditional independence structure. Do these more computationally intensive methods offer advantages---better or complementary information about interactions? Using simulated data and microarray data from yeast, we examine the abilities of these methods to reveal and represent the known interactions present.

### Recent Developments in Population Monte Carlo

❖ David Stephens, Imperial College London, Department of Mathematics, Huxley Building, 180 Queens Gate, London, SW7 2AZ UK, *d.stephens@imperial.ac.uk*

***Key Words:*** Markov chain Monte Carlo, sequential Monte Carlo, population algorithms, mixture models, gene expression profiles

We review population approaches to Markov chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) with particular application to mixture modeling and cluster analysis. The Bayesian inference problem for mixtures when the data being modeled are large in number and potentially high-dimensional is especially challenging, but this is precisely the context that coincides with the analysis of gene expression profiles. Assuming a regression model for the underlying profiles, we demonstrate that population MC methods provide a computationally feasible method of solution for the resulting mixtures of regressions problem. We illustrate the implementation of population MC methods on simulated data and several real datasets from microarray experiments investigating the functional genomics of different organisms.

### Multiple Testing Using the Posterior Probability of Half-Space: Application to Microarray Data

❖ Aurelie Labbe, Universite Laval, Pavillon Vachon, Dpt de mathematiques et de statistique, Quebec, G1K7P4 Canada, *alabbe@mat.ulaval.ca*; Mary Thompson, University of Waterloo

***Key Words:*** multiple testing, posterior probability, microarray data

The aim of this talk is to connect Bayesian analysis and frequentist theory in the context of multiple comparisons. Considering the problem of testing the equality of two sample means, we show that the posterior probability of the one-sided alternative hypothesis---defined as a half-space---shares with the frequentist p-value the property of uniformity under the null hypothesis. Ultimately, the posterior probability regarded is used in the same spirit as a p-value in the Benjamini-Hochberg procedure, or in any of its extensions. By doing so, we are taking advantage of both the control of the False Discovery Rate and the flexibility brought by the Bayesian framework.

### Markov Chain Density Estimation with Applications in Bioinformatics

❖ Mark Briers, University of Cambridge, Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ UK, *mb511@cam.ac.uk*; Arnaud Doucet, The University of British Columbia; Raphael Gottardo, The University of British Columbia

***Key Words:*** Markov chain Monte Carlo, density estimation, marginal likelihood, parallel computation, bioinformatics

The use of efficient density estimation techniques from the output of a Markov chain simulation (in a general state-space) is a key component in many numerical estimation procedures. In this paper, we extend the recent importance sampling-based density estimation technique of Chen (1994, 2005) and show how it can be used in bioinformatics. Specifically, we apply this density estimator to perform model selection in SNP-genotyping microarrays and introduce a novel parallel Markov chain Monte Carlo processing technique, which can be applied to problems where large datasets can render such techniques impractical.

### Bayesian Analysis of ChIP-chip Experiments

❖ Raphael Gottardo, The University of British Columbia, 333 6356 Agricultural Road, Vancouver, BC V6T1Z2 Canada, *raph@stat.ubc.ca*

***Key Words:*** bioinformatics, hierarchical model, Markov random fields, mixture models, tiling arrays

The advent of microarray technology has enabled biomedical researchers to monitor changes in the expression levels of thousands of genes. Until recently, however, the mechanisms driving these changes were harder to study in a similarly high-throughput level. A recent technological innovation, chromatin immuno-precipitation coupled with microarray (chip) analysis (hence the name ChIP-chip), now makes it possible for researchers to identify regions of a given genome bound by specific DNA binding proteins. During this talk, I will present a flexible and robust Bayesian hierarchical model for the analysis of ChIP-chip data that incorporates the spatial structure (probe dependence) while borrowing strength from all the probes. Detection of bound regions will be based on the posterior distribution of the parameters. The model will be illustrated using Affymetrix tiling arrays.

## 136 Coalescent/Evolution Theory

● ✪

Section on Statistics in Epidemiology

**Monday, August 7, 10:30 am–12:20 pm**

### The McDonald-Kreitman Test in Studies of Molecular Evolution

❖ Jason Gardenier, The College of William & Mary, 115 Saint Andrews Drive, NE, Vienna, VA 22180, *drgarden@verizon.net*

***Key Words:*** molecular evolution, Mcdonald-Kreitman, MK-test, DNA sequencing

The purpose of this presentation is to demonstrate how the McDonald-Kreitman (MK) test is being used to clarify the driving force of molecular evolution. By comparing the ratios of salient versus replacement substitutions between two species and within species, one can use the G-test of statistical significance to test the Neutral Theory of Molecular Evolution. As the speed with which DNA can be sequenced has been increasing rapidly, the need to have exact readings of nucleotides has become critical. The MK test is being used more often to elucidate the relative importance of Darwinian natural selection versus neutral genetic drift in driving overall evolution.

### Beyond the TDT

❖ Knut Wittkowski, The Rockefeller University, General Clinical Research Center, 1230 York Ave Box 322, New York, NY 10021, *kmw@rockefeller.edu*; Ephraim Sehayek, The Rockefeller University

***Key Words:*** McNemar test, u scores, chlolesterin, heart disease, personalized medicine, pharmacogenetics

Proposed in 1993, the Transmission Disequilibrium Test (TDT) by Spielman et al. has been used widely to analyze genetic information from trios (affected child and parents). In 2002 (discussion: 2004), we identified a flaw in the TDT (and many of its extensions). In short, genetical "independence" of alleles transmitted to the same child does not imply statistical "independence" of observations. Stratification by parental mating types increases power and reduces other shortcomings of the TDT. More importantly, new insights open the door for new tests targeting more specific questions. Using data on risk factors for cardiovascular diseases, we present versions particularly suited for diseases with low and high dominance and extension to uni- and multivariate ordinal and semiquantitative phenotypes. Asympt/exact versions of the tests are available from mustat.rockefeller.edu.

### Coalescent Theory for a Completely Random Mating Monoecious Population

❖ Edward Pollak, Iowa State University, Department of Statistics, Ames, IA 50011-1210, *pllk@iastate.edu*

***Key Words:*** monoecious population, random mating, coalescent theory

Consider a random mating monoecious diploid population that has N individuals in each generation.I assume that at time 0 a random sample of n << N copies of a gene are taken from this population. It is also assumed that G1,...,GN, the numbers of successful gametes produced by parents 1,...,N, are exchangeable random variables. It is shown that if time is measured backward in units of 8N/ E[ G1( G1-1)] = 2Ne generations, where Ne is the effective population size, the separate copies of a gene ancestral to those at time 0 are almost certain to come from separate individuals as Ne tends to infinity. It is then possible to obtain a generalization of coalescent theory for haploid populations if the distribution of G1 has a finite second moment and its third moment divided by N approaches 0 as N tends to infinity.

### Poisson Stars and Phylogenetic Trees: Limits of Inferences on Population History

❖ James Dunyak, The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730, *jdunyak@mitre.org*; Marc Colosimo, The MITRE Corporation; Lynette Hirschman, The MITRE Corporation

***Key Words:*** phylogenetic tree, SEIR, coalescence

The distribution of DNA sequences in a population and the resulting phylogenetic trees involve a complex interaction between the biological clock, the coalescence process, and the population history. In the most challenging case, all mutations occur after final coalescence, and the resulting tree is star-shaped with all branches emanating from a single node and Poisson-distributed genetic distances. We explore population inferences associated with rapidly spreading influenza virus in a single outbreak season in a single geographic area. Both exponential and susceptible-exposed-infectious-removed models are considered for population growth models. Estimates are then developed for the probability of obtaining a Poisson star and inferences are made on the initial seed size, growth rate, and other parameters.

### Strategies for Optimally Identifying Significant SNPs in Case-Control and Trio Association Data

❖ Christina Bromley, BioStat Solutions Inc., 5 Ridgeside Court, Suite 202, Mount Airy, MD 21771, *enabromley@biostat-solutions.com*; Ronald L. Bromley, BioStat Solutions Inc.; Sandy Close Kirkwood, Eli Lilly and Company; Richard D. Hockett, Eli Lilly and Company; Nitai Mukhopadhyay, Eli Lilly and Company

***Key Words:*** pharmacogenomic, SNP, identification, association, case control, trio

Large genetic association studies often include hundreds of thousands of genetic polymorphisms. The use of traditional multiple comparison adjustments alone could decrease the chance of detecting meaningful polymorphisms. The correlation between closely spaced SNPs further compounds the issue. Ascertainment bias, LD structure, and population stratification also impact significant SNP recognition. A genome-wide genetic association study was completed on a case-control cohort using >1 million SNPs on the initial screen, followed by individual genotyping for ~30,000 SNPs and replication in a cohort of parent-child trios. We propose a combination of statistical, genetic, epidemiologic, and clinical SNP selection strategies to confirm potentially important findings. Identification of SNPs between two designs needs careful consideration but can be used to validate positive findings.

## 137 Bayesian Bioinformatics ● ✪

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### Bayesian Analysis of EST Data with Multiple Libraries and Multiple Types of Tissues

❖ Fang Yu, University of Connecticut, Department of Statistics, 215 Glenbrook Road Unit 4120, Storrs, CT 06269, *fangyu@stat.uconn.edu*; Ming-Hui Chen, University of Connecticut; Lynn Kuo, University of Connecticut; Peng Huang, Medical University of South Carolina; Wanling Yang, The University of Hong Kong

***Key Words:*** Dirichlet distribution, gene expression, mixture distributions, multinomial distribution, shrinkage estimators

ESTs (Expressed Sequence Tags) are usually a one-pass sequencing reading of cloned cDNAs derived from a certain tissue. The frequency of unique tags among different unbiased cDNA libraries is used to infer the relative expression level of each tag. In this paper, we consider a multinomial model with novel priors of nonlinear Dirichlet distributions for EST data with multiple libraries and/or multiple types of tissues. The properties of the priors and the implied posteriors are examined in detail. Gene selection algorithms are developed to detect the co-expression within the same type of tissue and the differential expression between different types of tissues. A real EST dataset is used to illustrate the proposed model.

### Protein Structure Prediction: Statistical and Machine-Learning Approaches

❖ Sujay Datta, Texas A&M University, Department of Statistics, Blocker Building, College Station, TX 77843--3143, *sujayd5247@yahoo.com*

***Key Words:*** amino acid sequencing, secondary structure, tertiary structure, protein folding, Bayesian methodology, machine learning techniques

Although a protein can be characterized by its amino acid sequence first, most proteins fold into 3-D structures that determine their functions. After many successful attempts of sequencing, the focus is on structural bioinformatics where techniques are to be developed for predicting the secondary, tertiary, and quaternary structures of protein molecules from the sequence data. Several decades of research failed to produce any reliable prediction method, as experimental methods such as X-ray crystallography and NMR spectroscopy are slow and expensive and computational searches in the space of all possible 3-D configurations is impractical. Recently, breakthroughs have been achieved using statistical and machine learning approaches. After a quick review of the background needed, we will discuss the challenges involved and give a brief tour of these new approaches.

### Identifying Activated Molecular Pathways Using Bayesian Methods

❖ Yifang Zhao, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, *yifang@stat.uconn.edu*; Lynn Kuo, University of Connecticut; Dong-Guk Shin, University of Connecticut; Fang Yu, University of Connecticut

***Key Words:*** pathway, Bayes factor, MCMC

Identifying molecular pathways most activated in a defined stage of cell differentiation or when cells are exposed to environmental stimuli provides more insights on the functional information about genes. We propose novel methods to evaluate a set of possible pathways obtained from KEGG or BioCarta databases on their activations from the microarray and proteomic data. These high-throughput data are further supplemented by prior information constructed from literature search on the gene-to-gene promotion or inhibition knowledge. The Bayes factor approach is used to evaluate the evidence for each activated pathway. Essentially, we develop Markov chain Monte Carlo methods and Bayesian model selection methods to identify a set of pathways most activated by observing the high-throughput data.

### Normalization of Microrarrays in Transcription Inhibition Experiments

❖ Yan Zheng, University of Minnesota, 1023 29th Ave., SE, Apt D, Minneapolis, MN 55414, *yanzheng@biostat.umn.edu*; Cavan Reilly, University of Minnesota

***Key Words:*** Bayesian inference, gene expression, Gibbs sampler, microarray normalization

Almost all of the existing methods for normalization assume not too many of the genes differ in expression levels across arrays. Hence when the level of expression for many genes is not roughly constant across arrays, these standard methods are inappropriate. Here, we develop a model for normalization in the context of an experiment that attempts to measure mRNA half-lifes by stopping transcription and measuring gene expression at certain later times. This model does not assume most genes are constant across arrays, but rather that some genes have long half-lifes. By supposing there are genes with long half-lifes relative to the duration of the experiment, the model allows estimation of normalizing terms. Certain weaknesses of the basic model are noted and a more sophisticated model is developed that addresses these shortcomings.

### Bayesian Models for Pooling Microarray Studies with Multiple Sources of Replications

❖ Erin Conlon, University of Massachusetts, Department of Mathematics and Statistics, 710 North Pleasant Street, Amherst, MA 01003, *econlon@mathstat.umass.edu*

***Key Words:*** microarray, Bayesian models, meta-analysis, Bayesian false discovery rate

We introduce a Bayesian hierarchical model to pool cDNA microarray data across multiple independent studies to identify highly expressed genes. Each study has multiple sources of variation (i.e., replicate slides within repeated identical experiments). Our model produces the gene-specific posterior probability of differential expression---which provides a direct method for ranking genes---and provides Bayesian estimates of false discovery rates (FDR). In simulations combining two and five independent studies, with fixed FDR levels, we observed large increases in the number of discovered genes in pooled versus individual analyses. We also observed that, in our simulation studies, the Bayesian FDR estimates tracked the true FDRs well. We will illustrate our approach using two independent studies in Bacillus subtilis. This is joint work with Joon Jin Song and Jun S. Liu.

## 138 Statistical Issues in Cardiovascular Medical Device Clinical Trials ● ✪

Biopharmaceutical Section, Biometrics Section, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### The Use of a Correlated Binary Data Model for Sample Size Calculation and Analysis of Accessions Data

❖ John C. Evans, Boston Scientific Corporation, 100 Boston Scientific Way, Marlborough, MA 01752-1242, *evansj1@bsci.com*; Corina M. Sirbu, Boston Scientific Corporation; Kay M. Larholt, Boston Scientific Corporation; Brian Johnson, Boston Scientific Corporation

***Key Words:*** implantable port venous accessions, complication rate, temporally correlated binary data, sample size, interim analysis

Sample size calculations and analyses in studies of port comparisons have assumed independence of accessions across time. However, accessions usually are correlated, thus requiring more study patients. If this correlation is ignored when counting complications for the individual accessions, and the proportions of complications are compared by methods assuming independence of accessions, then incorrect conclusions may be drawn. We will present results for a new study, comparing sample size calculations and corresponding interim analyses under the usual assumption of independence of accessions with those obtained when a positive correlation is assumed among accessions. In the latter case, we use a correlated binary data model to estimate the number of patients required with an average of 13 accessions over a six months and do a corresponding interim analysis.

### Statistical and Clinical Significance: a Practical Example Utilizing ST-Segment Monitor Endpoints in Acute ST Elevation MI (STEMI)

❖ Cynthia Green, Duke Clinical Research Institute, 2400 Pratt Street, Durham, NC 27705, *green054@mc.duke.edu*

***Key Words:*** clinical trials, medical devices, biomarkers, clinical meaning, statistical significance

While mortality is the best measure of treatment benefit, thousands of patients often are required to show a statistically measurable reduction. Nonmortality biomarkers, capturing data at different times using different data collection methods and representing different physiological information, can show a significant treatment effect in smaller populations. A biomarker should be statistically and clinically significant. Statistical significance provides >95% likelihood that the observed biomarker change is the result of treatment, not chance or confounding. Clinical meaning denotes the biomarker change is associated with improved clinical outcome. Statistically significant benefits are not always clinically meaningful. This study shows a statistical model defining a threshold for clinical meaning for ST-segment endpoints used to measure therapeutic benefit in acute MI patients.

### An Application of Propensity Score Method in Medical Device Trials

❖ Minglei Liu, Medtronic, Inc., Medtronic Vascular, 3576 Unocal Place, Santa Rosa, CA 95403, *minglei.liu@medtronic.com*; Jianming Wang, Medtronic, Inc.; Jane Gau, Medtronic, Inc.

***Key Words:*** propensity score, imputation, treatment effect

Propensity score method is a useful tool to adjust the selection bias and accurately estimate treatment effects. While it is used widely in all kinds of clinical trials, there is no general rule about how the subgroup (quintiles) should be combined. We will discuss how this is applied in a case study in medical device clinical trials, including missing value imputation, score calculation, subgroup analysis, and the combination method.

### Patient-Prosthesis Mismatch and Operative Mortality after Aortic Valve Replacement Surgery

❖ Sean M. O'Brien, Duke University, DUMC 3850, Durham, NC 27710, *obrie027@mc.duke.edu*; Elizabeth R. DeLong, Duke

University; Lawrence H. Muhlbaier, Duke University; Charles R. Bridges, University of Pennsylvania; Eric D. Peterson, Duke University

***Key Words:*** observational studies, measurement error, hidden confounders, sensitivity analysis, hierarchical models

Patient-prosthesis mismatch (PPM) following aortic valve replacement surgery is said to be present when the blood flow area of the prosthetic valve is too small relative to the patient's body size, leading to high pressure gradients. Guidelines now recommend not implanting a valve if the ratio of its effective orifice area relative to body surface area is less than 0.85. Yet, the clinical relevance and effects of PPM remain controversial. This paper describes numerous methodological challenges we encountered when analyzing the association between PPM and operative mortality in a large cardiac surgery registry. Issues we addressed include measurement error and sensitivity to hidden confounders. Our data suggest that while PPM may be causally related to increased operative mortality, the relationship is highly complex and may be confounded by numerous unmeasured variables.

### Receiver Operating Characteristic Curve Method for Evaluating Surrogate Endpoints for Predicting Target Lesion Revascularization in a Second-Generation Drug Eluting Stent Trial

❖ Hong Wang, Boston Scientific Corporation, CP4, One Boston Scientific Place, Natick, MA 01760, *hong.wang@bsci.com*; Sarah Bass, Boston Scientific Corporation ; Peter S. Lam, Boston Scientific Corporation

***Key Words:*** receiver operating characteristic, c-statistics, diagnostic test, drug eluting stent, surrogate endpoint

In this paper, we used the receiver operating characteristic (ROC) curve and area under the ROC curve method to access the correlations of the quantitative coronary angiography (QCA) parameters with the clinical outcome of target lesion revascularization (TLR) in the second-generation drug eluting stent trial. The examined QCA surrogate endpoints included the analysis segment and stented region of late loss, percent diameter stenosis, and minimum lumen diameter. The c-statistics were computed with its 95% confidence interval for each QCA parameter. A threshold was determined by using the probability cutoff method, and a diagnostic test was performed to test the difference of the c-statistics of the paired QCA parameters for their correlation with TLR. The in-segment percent diameter stenosis was identified as the QCA surrogate endpoint with the highest correlation with TLR.

## 139 Methods in Cancer Research ●

Biometrics Section, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### A Primer on Quantitative RT-PCR (q-RT-PCR) with Applications in Colorectal Cancer Biomarker Studies

❖ Terry Hyslop, Thomas Jefferson University, 211 S. 9th Street, Suite 602, Walnut Towers, Philadelphia, PA 19107, *thyslop@mail.jci.tju.edu*; Yanyan Li, Thomas Jefferson University ; Inna Chervoneva, Thomas Jefferson University

***Key Words:*** quantitative RT-PCR, simulations, colorectal cancer

Quantitative RT-PCR, while used in research laboratories for ~10 years, is a relatively recent technique in clinical investigations. With data from a prospective multi-center trial testing the prognostic ability of a biomarker for colorectal cancer, we present technical and statistical challenges these data present. Using probabilistic approaches, we investigated several methods to simulate data in order to investigate the estimation properties of existing approaches as well as new alternative methods of quantification. Additionally, we include an error structure to incorporate the serial dilution process, which is common in these experiments. In this presentation, we will demonstrate the simulation approach that most closely models the real data we see in our clinical studies.

### Estimation of RT-PCR Amplification Efficiency with Application to Relative Quantification

❖ Inna Chervoneva, Thomas Jefferson University, 211 S. 9th Street, Suite 602, Walnut Towers, Philadelphia, OR 19107, *i_chervoneva@mail.jci.tju.edu*; Yanyan Li, Thomas Jefferson University ; Terry Hyslop, Thomas Jefferson University; Boris Iglewicz, Temple University

***Key Words:*** RT-PCT efficiency, amplification rate, RT-PCR kinetic, relative quantification, amplification efficiency, branching process

Amplification efficiency is the key issue in real time (RT-PCR) quantification because many PCRs do not have ideal amplification efficiency and quantification should be adjusted for the actually achieved efficiency. Efficiency of a single reaction may be estimated by considering kinetic data from this reaction. Most proposed methods for estimating amplification efficiency fit an exponential growth curve to the limited subset of kinetic data in early exponential and loglinear phases. In this work, we use a four-parameter logistic model for all kinetic RT-PCR data. We show that both exponential and logistic models for RT-PCR kinetic data may be obtained as approximations of generalized theoretical model for PCR kinetics of Schnell and Mendoza (1997) and that amplification efficiency estimates from logistic models are more precise than those from the exponential models.

### Sample Size for FDR-Control in Microarray Data Analysis

❖ Sin-Ho Jung, Duke University, Hock Plaza, Suite 802, 2424 Erwin Road, Durham, NC 27705, *jung0005@mc.duke.edu*

***Key Words:*** block compound symmetry, family-wise, prognostic gene, true rejection, two-sample t-test

We consider identifying differentially expressing genes between two patient groups using a microarray experiment. We propose a sample-size calculation method for a specified number of true rejections while controlling the false discovery rate at a desired level. Input parameters for the sample-size calculation include the allocation proportion in each group, the number of genes in each array, the number of differentially expressing genes, and the effect sizes among the differentially expressing genes. We have a closed-form sample size formula if the projected effect sizes are equal among differentially expressing genes. Otherwise, our method requires a numerical method to solve an equation. Simulation studies are conducted to show that the calculated sample sizes are accurate in practical settings. The proposed method is demonstrated with a real study.

### An Experimental Design for Clinical Trials Evaluating Combination Agents

❖ Christine McLaren, University of California, Irvine, Epidemiology Division, 224 Irvine Hall, Irvine, CA 92697-7550, *cmclaren@uci.edu*; Vernon M. Chinchilli, The Pennsylvania State University; Wen-Pin Chen, Chao Family Comprehensive Cancer Center; Frank L. Meyskens, Chao Family Comprehensive Cancer Center

***Key Words:*** 2x2 factorial design, Cochran-Armitage trend test, cancer prevention, simple loop alternative

For clinical trials that intend to test combination agents, new designs are needed to achieve results with fewer patients and less expense. We propose a 2x2 balanced factorial design to test whether a combination of agents is better than the individual treatments or placebo with regard to reducing the proportion of recurrent adenomas, precursors to colorectal cancer. The Cochran-Armitage test of trend can be used to compare the proportion of patients developing at least one adenoma between placebo and treatment groups. We derive a pair of two-sided tests for the ordered alternative of increasing or decreasing proportions across treatment groups and a method for power and sample-size calculations. Simulations demonstrate type I and II error rates. The design achieves a reduction in sample size of more than 30% and promises economy of patients and expense.

### Adaptive Phase I Clinical Trial Designs for Biologic Agents and Molecularly Targeted Therapies

❖ Sumithra Mandrekar, Mayo Clinic College of Medicine, Division of Biostatistics, 200 1st St SW, Rochester, MN 55905, *mandrekar.sumithra@mayo.edu*; Daniel Sargent, Mayo Clinic College of Medicine

***Key Words:*** phase I, CRM, targeted therapy

Agents with specific biologic targets with unknown dose efficacy curves and limited toxicity profile require novel designs to establish biologically optimal doses. We present two designs for identifying an optimal dose of a single or two drug combinations utilizing both toxicity and efficacy data. A continual reassessment method with straight forward dose selection criteria using accumulated data from all patients treated thus far is employed based on the flexible continuation-ratio model. In a two-agent setting, a generalization of the single agent model allowing for separate toxicity and efficacy curves for each agent to generate a dose outcome surface for the combination is used. Our simulations demonstrated that the proposed designs have favorable operating characteristics in terms of experimentation and recommendation rates and the average sample size under a variety of scenarios.

## 140 Expanding Access to Research Data: Reconciling Risks and Opportunity

Section on Government Statistics, Section on Statistical Education, Section on Statistical Consulting

**Monday, August 7, 10:30 am–12:20 pm**

### Expanding Access to Research Data: Reconciling Risks and Opportunities

❖ Eleanor Singer, University of Michigan, Institute for Social Research, Box 1248, Ann Arbor, MI 48105, *esinger@isr.umich.edu*; ❖ Margo Anderson, University of Wisconsin-Milwaukee, History Department, Milwaukee, WI 53201, *margo@uwm.edu*; ❖ John Haltiwanger, University of Maryland, 3105 Tydings Hall, College Park, MD 20742, *haltiwan@econ.umd.edu*; ❖ Daniel Weinberg, U.S. Census Bureau, 4700 Silver Hill Road, MS 6300, Washington, DC 20233, *daniel.h.weinberg@census.gov*; ❖ Donald Rubin, Harvard University, 1 Oxford Street, Cambridge, MA 02138, *rubin@stat.harvard.edu*

***Key Words:*** data access, confidentiality, restricted access, restricted data

The recently released National Research Council report, "Expanding Access to Research Data: Reconciling Risks and Opportunities," examines societal and technological conditions that have complicated the release of confidential statistical data to users. The report recommends a variety of approaches to reconcile the conflicting demands of making data widely available while protecting their confidentiality, such as providing data through multiple modes, tailored to the needs of different types of users; improving the utility and confidentiality protection of newer access modes; improving the efficiency of agency review processes; and measuring the level of research data use and the frequency with which confidentiality breaches occur. This session will bring together stakeholders who have not participated in the NRC panel's deliberations to offer their perspectives.

## 141 Causal Inference and Noncompliance

Biometrics Section, Section on Health Policy Statistics, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### On Estimating Treatment Effects under Noncompliance in Randomized Clinical Trials

❖ Heejung Bang, Cornell University, 411 E. 69th Street, New York NY, NY 10021, *heb2013@med.cornell.edu*; Clarence E. Davis, The University of North Carolina at Chapel Hill

***Key Words:*** as treated, instrumental variables, intention-to-treat, noncompliance, per-protocol

In this report, we compared four estimators---intention-to-treat (ITT), as treated (AT), per-protocol (PP), and instrumental variables (IV) estimators---considered conventionally for treatment effect estimation by simulation under different noncompliance scenarios in typical clinical trial settings. We found that ITT and IV estimators are not perfect and can be problematic, although they carry desirable properties.

### Efficient Nonparametric Estimation of Causal Effects in Randomized Trials with Noncompliance

❖ Jing Cheng, University of Pennsylvania, 423 Guardian Drive, 503 Blockley Hall, Biostatistics, Center for Clinical Epidemiology and Bio, Philadelphia, PA 19104, *jcheng@cceb.upenn.edu*; Dylan S. Small, University of Pennsylvania; Thomas R. Ten Have, University of Pennsylvania; Zhiqiang Tan, Johns Hopkins University

***Key Words:*** causal effects, randomized trials, noncompliance, efficient nonparametric estimation, compliance class

Causal approaches based on the potential outcome framework provide a useful tool for addressing the noncompliance problems in randomized trials. Various estimators (e.g., instrumental variable (IV) estimator) have been proposed for causal effects of treatment. We propose a new estimator by applying the empirical likelihood with moment restrictions to the mixture outcome distributions. Simulation studies show this estimator is robust to parametric distribution assumptions and more efficient than the standard IV estimator. The method is applied to data from a randomized trial of an encouragement intervention to improve adherence to prescribed depression treatments among depressed elderly patients in primary care practices.

## Nested Markov Compliance Class Model in the Presence of Time-Varying Noncompliance

❖ Julia Lin, University of Pennsylvania, *jlin@cceb.upenn.edu*; Thomas R. Ten Have, University of Pennsylvania; Michael R. Elliott, University of Michigan

***Key Words:*** compliance class, principal strata, Markov, potential outcomes, randomized, longitudinal

We consider a Markov structure for partially unobserved, time-varying compliance classes in the Imbens-Rubin (1997) compliance model framework. The context is a longitudinal randomized intervention study where subjects are randomized once at baseline, outcomes and patient adherence are measured at multiple follow-ups, and patient adherence to the randomized treatment could vary over time. We propose a nested latent compliance class model where we use time-invariant, subject-specific compliance principal strata to summarize longitudinal trends of subject-specific time-varying compliance patterns. The principal strata are formed using Markov models that related current compliance behavior to compliance history. Causal treatment effects are estimated as intent-to-treat effects within the compliance principal strata.

## Use of a Marginal Structural Model To Estimate the Causal Effect of Epoetin Use on Hemactocrit Value among Hemodialysis Patients

❖ Yi Zhang, Medical Technology and Practice Patterns Institute, 4733 Bethesda Ave., #510, Bethesda, MD 20814, *yz@mtppi.org*; Mae Thamer, Medical Technology and Practice Patterns Institute; Dennis J. Cotter, Medical Technology and Practice Patterns Institute; James Kaufman, U.S. Department of Veterans Affairs VA Boston Healthcare System; Miguel A. Hern·n, Harvard School of Public Health

***Key Words:*** causal inference, marginal structural model, inverse probability of treatment weighted estimation, erythropoietion, hematocrit

Standard methods to adjust for time-varying confounders that are affected by previous treatment may lead to biased estimates. Marginal structural models (MSMs), introduced by Robins, can appropriately control for this type of time-dependent confounding. The parameters of MSMs are estimated by inverse probability weighting. We present an application of MSMs to estimate the causal effect of erythropoietin (EPO) on hematocrit (Hct) response among renal failure patients with anemia. The research question is motivated by the discrepancies between observational and clinical trial results examining EPO treatment to ameliorate anemia. Specifically, in contrast to clinical trials, observational studies have reported a significantly inverse association between EPO and Hct.

## The Causal Effect of Finasteride on the Severity of Prostate Cancer

❖ Bryan Shepherd, Vanderbilt University, 1161 21st Ave., S., S2323 MCN, Nashville, TN 37232-2158, *bryan.shepherd@vanderbilt.edu*; Mary W. Redman, Southwest Oncology Group; Donna P. Ankerst, University of Munich

***Key Words:*** causal inference, sensitivity analysis, counterfactuals, principal stratification

The Prostate Cancer Prevention Trial found that finasteride lowered the risk of prostate cancer. However, those diagnosed with cancer in the finasteride arm tended to have worse cases of cancer than those with cancer in placebo. Did finasteride cause more severe cancer or was this due to post-randomization selection bias? We answer this question by estimating the mean treatment difference in prostate cancer severity for those who would have had cancer regardless of treatment assignment. We consider assumptions that allow identification of this estimand: conditional on diagnosis of cancer, independence of cancer severity and incidence of cancer if randomized to other arm; monotonicity; and outcomes missing completely at random. We perform sensitivity analyses by systematically relaxing these assumptions using plausible ranges of sensitivity parameters and relevant covariates.

## Location Design in Location-Controlled Follow-Up Studies

❖ Fan Li, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, E3036, Baltimore, MD 21205, *fli@jhsph.edu*

***Key Words:*** design, causal effect, principal stratification, partially controlled studies, needle exchange program

Good designs are essential to the quality of data analysis. Also, new (valid) evaluation methods imply existence of novel good designs for similar studies in future. We focus on a general class of location-controlled designs where the location of services affects utilization of those services and measurement of data. Design strategies for the location-controlled programs are rarely investigated in public health. Based on the novel evaluation approach of Frangakis et al. (2004) for partially controlled studies using principal stratification, we propose to develop optimal designs in the following sense: (1) yielding largest benefit to community; (2) yielding the best accuracy in periodical evaluation; (3) balancing a large benefit based on current information, with ability of accurate evaluation. We illustrate our methods in the design of a Needle Exchange Program.

## How To Make Inference from Experiments Conducted on Populations with Different Disease Determinants

❖ Steven D. Mark, University of Colorado Health Sciences Center, 4200 E. 9th Ave., Campus Box 119, Denver, CO 80262, *steven.mark@uchsc.edu*

***Key Words:*** causal inference, fundamental disease probability, drug development, informative contrasts

Frequently in the study of human disease one wishes to make inferences from two experiments where the subjects in experiment A are not exchangeable with the subjects in experiment B. For instance, in developing new agents to treat human cancers one first studies the effect of a drug on cancers in mice (experiment A), and then predicts what the outcome will be in a randomized trial of the drug on cancers in humans (experiment B). I have proposed a formal structure for such inference, the Fundamental Disease Probability (FDP) paradigm. A fundamental disease probability is the probability of disease conditional on both measured and unmeasured disease risk factors. In the FDP paradigm the goals of inference are defined in terms of contrasts of FDPs within and/or between experiments. In this talk I illustrate the structure of FDP inference by describing its application in two simple examples.

# 142 Nonparametric Regression Methods I

Section on Nonparametric Statistics

**Monday, August 7, 10:30 am–12:20 pm**

## Two-Sample Comparison with Long Memory Errors

❖ Fang Li, Indiana University Purdue University Indianapolis, Department of Mathematical Sciences, 402 N Blackford Street LD270, Indianapolis, IN 46202, *fli@math.iupui.edu*

***Key Words:*** partial sum process, fractional Brownian motion, Monte Carlo simulation

This paper discusses the problem of testing the equality of two nonparametric regression functions against two-sided alternatives for uniform design on [0,1] with long memory moving average errors. The standard deviations and the long memory parameters are possibly different for the two errors. The paper adapts the partial sum process idea used in the independent observations settings to construct the tests and derives their asymptotic null distributions. The paper also shows that these tests are consistent for general alternatives and obtains their limiting distributions under a sequence of local alternatives. Monte Carlo simulations are then conducted to study the finite sample level and power behavior of these tests at some alternatives.

## Some Diagnostic Methods for Choosing the Degree of Smoothing in Nonparametric Regression

❖ Craig Andres, Kettering University, , *craig_andres@yahoo.com*; Robert Kushler, Oakland University

***Key Words:*** nonparametric regression, smoothing parameter selection, diagnostics

Automated methods for selection of the smoothing parameter in nonparametric regression modeling, such as (generalized) cross validation, occasionally result in overfitting the data. Some diagnostic techniques for detecting when this is happening, and some alternative methods for smoothing parameter selection, will be discussed.

## Spline-Backfitted Kernel Smoothing of Additive Regression Models

❖ Jing Wang, Michigan State University, Department of Statistics and Probability, East Lansing, MI 48824, *wangjin3@msu.edu*; Lijian Yang, Michigan State University

***Key Words:*** bandwidths, b spline, knots, local linear estimator, Nadaraya-Watson estimator

A great deal of efforts has been devoted to the inference of additive model in the last decade. Among the many existing procedures, the kernel type are too costly to implement for large number of variables or for large sample sizes, while the spline type provides no asymptotic distribution or any measure of uniform accuracy. We propose a synthetic estimator of the component function in an additive regression model, using a one step backfitting, with spline estimators in the first stage and kernel estimators for the second stage. It is established that under very weak conditions, the proposed estimator's pointwise distribution is asymptotically equivalent to an ordinary univariate Nadaraya-Watson estimator, hence the dimension is effectively reduced to one at any point. This dimension reduction holds uniformly over an interval under stronger assumptions of normal errors.

## Boundary Kernel Method in Nonparametric Deconvolution

❖ Shunpu Zhang, University of Nebraska-Lincoln, 340 Hardin Hall, N., Lincoln, NE 68583, *szhang3@unl.edu*

***Key Words:*** deconvolution, boundary kernel estimator, density estimation, boundary effect

This paper considers nonparametric deconvolution problem when the true density function is truncated. We propose to use the deconvolution boundary kernel method to remove the boundary effect of the conventional deconvolution density estimator. Methods of constructing deconvolution boundary kernels are provided. The mean square error properties, including the rates of convergence, are investigated for supersmooth and ordinary smooth errors. It is shown that the deconvolution boundary kernel estimator successfully removes the boundary effects of the conventional deconvolution density estimator. Simulations are carried out to compare the performance of the deconvolution boundary kernel estimator and the conventional deconvolution estimator.

## Double-Smoothing for Bias Reduction in Local Linear Regression

❖ Hua He, University of Rochester Medical Center, Department of Biostatistics and Computational Biology, 601 Elmwood Avenue, Box 630, Rochester, NY 14623, *huahe@bst.rochester.edu*; Li-Shan Huang, University of Rochester

***Key Words:*** local linear regression, edge effect, mean squared error, asymptotic bias, asymptotic variance

Nonparametric local linear regression is commonly used in practice because of its excellent numerical and theoretical properties. It involves fitting a straight line segment over a small region whose midpoint is the abscissa, x, and the predicted value at x is the estimated intercept of that straight line segment.In this paper, we propose a new estimator, the double-smoothing local linear estimator, which combines all fitted local lines in the neighborhood of x with another round of smoothing. The new estimator makes uses of all local intercepts and slopes, in contrast of using only the intercepts in local linear regression. The new estimator reduces the bias to an order of h^4 comparing with h^2 for local linear regression without affecting the order of magnitude of variance. This paper also investigates the edge effects of the proposed new estimator, and simulation results are also given

## A Comparative Study of Stage-1 and Stage-2 Semiparametric Estimation of Hemodynamic Response Function for fMRI Data

❖ Chunming Zhang, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, *cmzhang@stat.wisc.edu*; Yuan Jiang, University of Wisconsin-Madison; Tao Yu, University of Wisconsin-Madison

***Key Words:*** false discovery rate, multiple comparison, semiparametric test, stimuli, time resolution

The standard tool for analyzing functional magnetic resonance imaging (fMRI) data is some variant of the linear model, which is restrictive in modeling assumptions. We develop a semiparametric approach, based on the cubic smoothing splines, to obtain statistically more efficient estimates of the hemodynamic response function (HRF) associated with fMRI experiments. The hypothesis testing of HRF is conducted to identify regions of the brain which are activated when a subject performs a particular task. Furthermore, we compare stage-1 and stage-2 semiparametric estimates of HRF in significance tests for detecting the activated brain regions. Our simulation studies demonstrate that the stage-1 estimates combined with a bias correction procedure perform best in detecting the activated brain regions. We illustrate this method using a real fMRI dataset and compare with popular methods.

# 143 Contributions in National Security ✪

Section on Statisticians in Defense and National Security

**Monday, August 7, 10:30 am–12:20 pm**

## Albert Einstein: Centennial Appreciation of a Master Statistician

❖ Boris Iglewicz, Temple University, Department of Statistics, 00600, 1810 N 13th Street, Philadelphia, PA 19122, *borisi@temple.edu*

***Key Words:*** Bose-Einstein, photoelectric effect, Brownian motion, scientific method

Fifty years have passed since Einstein's death, and one century since the appearance of his 1905 revolutionary papers. This talk serves as a modest tribute to this anniversary with an emphasis on the statistical component of Einstein's work. Also considered are his early years that led up to 1905. There will be discussion of the Bose-Einstein, Brownian motion, and photoelectric papers. In addition, an alternative interpretation of his E = mc2 formula will be given. Evidence, including statistical, is provided for why Einstein is viewed as the greatest scientist of the 20th century. An explanation is provided for why the title's phrase, "master statistician," is reasonable. This exploratory journey provides insight into the scientific method and the role of statistics in science.

## Toward Effective Anomaly Detection in Public Health Surveillance

❖ Colin Goodall, AT&T Labs-Research, 200 S. Laurel Ave., Middletown, NJ 07748, *cgoodall@att.com*; Sylvia Halasz, AT&T Labs-Research; Arnold Lent, AT&T Labs - Research; Simon Tse, AT&T Labs - Research; John Allegra, Emergency Medical Associates Research Foundation; Dennis Cochrane, Emergency Medical Associates Research Foundation

***Key Words:*** anomaly detection, public health, biosurveillance, emergency department data, alert

In biosurveillance' use of statistical anomaly detection, multiple streams of data---from hospital emergency departments, testing laboratories, pharmacies, etc. - are continually examined for abnormal patterns. The detected anomalies, with attached measures of significance and severity, are queued for investigation, leading to possible action. A partially-adaptive anomaly detection method has been applied to daily feeds of emergency department data using a statistically-principled system of generated outbreaks. Alerts were generated for a variety of parameter settings, giving state public health officials a tradeoff among number of false alerts, the smallest outbreak to be detected, the outbreak profile, and the timeliness of detection.

## A Bayesian Approach to Radiographic Surveillance in Children

❖ Namhee Kim, The Ohio State University, 320 W. 10th Ave., Center for Biostatistics, Columbus, OH 43210, *kim.1862@osu.edu*; Prem K. Goel, The Ohio State University; Bema Bonsu, Columbus Children's Hospital; M. Marvin, Children's Hospital

***Key Words:*** bivariate dynamic model, biosurveillance, time series, outbreak detection

Recently, there has been a growing demand on timely detection of abnormal pattern in the time-series data with epidemiological interest. The spread of SARS virus in 2003 provides an example to motivate this study. We focus on detection of respiratory diseases with abnormal daily count of children who have to undergo chest radiograph at an emergency department (ED). We use bivariate dynamic time-series model for the daily counts of children coming to ED (census) and those given chest radiographs. To detect abnormal departure from realistic underlying situation, it is crucial to learn the regular pattern from historical data and possible variants that already exist in historical data, but not involving outbreaks. These issues are addressed and a decision rule to suggest the existence of outbreak in the current data is discussed and implemented with artificially attempted outbreaks.

## Statistical Analysis of Department of Energy Safety Data

❖ Rama Sastry, U.S. Department of Energy, EH31, Washington, DC 20585, *Rama.sastry@eh.doe.gov*

***Key Words:*** VPP, safety, clustering, box plots

The Department of Energy operates many nuclear and non-nuclear facilities and National laboratories located at various places in the United States. The Department collects safety data to monitor health and safety of workers at these facilities. The objective of this paperis to present some statistical methods applied to analyze such data and offer suggestions to improve safety. The statistical techniques used include Hierarchical Clustering analysis, Box plots and Trend analysis with Kalman filtering. The results of the analysis indicate that the facilities adopting a safety program called Voluntary Protection Program (VPP) showed better safety performance than others in the Department.

## Time Transformation Methods for Analyzing Time Series with Time-Varying Frequencies

❖ James R. Haney, Southern Methodist University, Department of Statistical Science, Dallas, TX 75275-0332, *jhaney@smu.edu*; Wayne A. Woodward, Southern Methodist University; Henry L. Gray, Southern Methodist University

***Key Words:*** time series, time transformation, time-varying frequency

A time series with time-varying frequencies can often be transformed into a stationary time series by transforming the time axis. After this time transformation, traditional methods for analyzing stationary time series can then be used. This paper discusses these time-transformation methods, which can be usefully applied to many seismological and biomedical time series where potential applications include classification and smoothing. Examples are discussed in detail.

## Minimum Hellinger Distance Estimation Using Synthetic Data

❖ Bret Hanlon, Cornell University, DSS 301 Mallot Hall, Ithaca, NY 14853-3801, *bmh35@cornell.edu*; Anand Vidyashankar, Cornell University

*Key Words:* data confidentiality, privacy, synthetic data, minimum Hellinger distance, statistical efficiency

Data Confidentiality is emerging to be an important issue in several areas of science and technology, especially within the healthcare industry. Synthetic data has been proposed as a tool to mitigate some problems of confidentiality. In this talk, we describe a new "robust" methodology to construct and analyze synthetic data. Our construction of synthetic data is nonparametric in nature while the modeling and analyses are of parametric type. We introduce minimum Hellinger distance methodology in this context and establish the classical asymptotic properties. Furthermore, we describe the privacy properties of our methodology and evaluate the trade-off between privacy and efficiency. Finally, we establish precise connections between our methods to the other methods in the area of data confidentiality. We apply our methodology to the analysis of total earnings in the 2000 census data.

### Error Modeling in Vector-Based GIS Data

❖ Kimberly Love, Virginia Polytechnic Institute and State University, 1307 University City Blvd., Apt 3, Blacksburg, VA 24060, *krlove@vt.edu*; Keying Ye, The University of Texas at San Antonio; Eric P. Smith, Virginia Polytechnic Institute and State University; Stephen Prisley, Virginia Polytechnic Institute and State University

*Key Words:* GIS, vector, coordinates

Geographic Information Systems (GIS) provide researchers with flexibility, information and methods for displaying the information visually. Researchers typically view the results as fixed although uncertainty and error are possible. This error affects many aspects of GIS; in vector data it affects coordinate features, line features, polygon features, and the relationships between them. Thus the error is significant in establishing existence of important features such as roads, buildings, streams, and so on. We provide a quick overview of existing work on use of stochastic models in GIS and detail extensions and new innovations in this area. We will also explore the applications of statistical methodologies in modeling the error structure in vector-based GIS data.

## 144 Epidemiologic Modeling ● ✪

Section on Statistics in Epidemiology, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### A Minimum-Distance Approach to Logistic Regression via the Case-Control Formulation

❖ Howard Bondell, North Carolina State University, 220D Patterson Hall, Raleigh, NC 27695, *bondell@stat.ncsu.edu*

*Key Words:* biased sampling, logistic regression, robust, minimum distance

Maximum likelihood fitting of the logistic regression parameters can be affected greatly by atypical observations. Several robust alternatives have been proposed and implemented in statistical software packages. However, upon considering the model via the case-control viewpoint, it is clear that current techniques can exhibit poor behavior in many common situations. A new robust class of procedures is introduced. The estimates are constructed via a minimum-distance approach after identifying the model with a semiparametric biased sampling model. A particularly useful choice of distance measure is described via a semiparametric empirical characteristic function. These new approaches can be highly efficient, while remaining robust to small deviations in the model. Thus, they can be used to fit the logistic regression model, even in the presence of atypical observations.

### A Wavelet-Based Method for the Prospective Monitoring of Disease Incidence Counts in Space and Time

❖ J. Brooke Marshall, Virginia Polytechnic Institute and State University, 403 Progress Street, NE, Apt 8, Blacksburg, VA 24060, *jemarsh2@vt.edu*; Dan Spitzner, Virginia Polytechnic Institute and State University; William H. Woodall, Virginia Polytechnic Institute and State University

*Key Words:* public health surveillance, wavelets, Poisson regression, disease clusters, control charts, ARL performance

In epidemiology it is useful to monitor disease occurrences prospectively to allow for quicker detection of disease clusters. Here we present a prospective method for monitoring disease occurrences in a geographical region. In this method, a surface of incidence counts is modeled over time in the region of interest. This surface is modeled using Poisson regression where the regressors are functions from the Haar wavelet basis. The surface is estimated each time incidence data is obtained using both past and current observations, weighing current observations more heavily. The flexibility of this method allows for the detection of changes in the incidence surface, increases in the overall mean incidence count, and clusters of disease within smaller areas of the region, by using control charts. The control limits for these charts are determined by average run length performance.

### Estimating the Causal Effect of LTVV on ALI with G-Computation

❖ Weiwei Wang, Johns Hopkins University, 615 N. Wolfe Street, E3040, Baltimore, MD 21205, *wewang@jhsph.edu*; Daniel Scharfstein, Johns Hopkins Bloomberg School of Public Health; Dale Needham, Johns Hopkins Medical Institutions; Roy Brower, Johns Hopkins Medical Institutions; Peter Pronovost, Johns Hopkins Medical Institutions

*Key Words:* causal inference, g-computation, dynamic treatment regime, potential outcomes

The Lower Tidal Volume Ventilation (LTVV) approach is a dynamic treatment regime using volume- and pressure-limited mechanical ventilation for patients with Acute Lung Injury (ALI). We used G-Computation (Lok et al., 2004) to estimate the potential survival function of the LTVV approach among ALI patients in an ARDSNetwork clinical trial (ARDSNetwork, 2000). Our results demonstrate that strict adherence to LTVV is efficacious at reducing short-term mortality.

### Joint Modeling of Birthweight and Gestational Age

❖ Betsy Enstrom, Duke University, 5500 Bradford Pear Court, Institue of Statistics and Decision Sciences, Raleigh, NC 27606, *betsy1181@yahoo.com*; Alan E. Gelfand, Duke University; Geeta K. Swamy, Duke University Medical Center; Marie L. Miranda, Duke University

*Key Words:* mixture model, MCMC algorithm, EM algorithm, cumulative logit

Modeling birthweight (BW) and gestational age (GA) have received much discussion in the literature and it's linkage to social, environmental, and genetic risk factors. A clear definition of these two birth cohorts is beneficial to many fields of science. We propose joint modeling with the objective of better characterization of adverse birth outcomes. The data set consists of 107045 live births as recorded in the North Carolina Detailed Birth Record for 2003, we develop the model specifications in a hierarchical Bayesian framework as well as a likelihoodist setting. A

MCMC and EM algorithm are illustrated for the Bayesian and frequentist approach, respectively. Practical issues that arise when fitting these types of models are discussed including the number of mixture components, convergence issues. GA given BW is fit with logistic regression with a cumulative logit link function.

### A Cross-Cultural Investigation of College Student Alcohol Consumption: a Classification Trees Analysis

❖ Panagiota Kitsantas, East Carolina University, Austin 124, Greenville, NC 27513, *kitsantap@ecu.edu*; Anastasia Kitsantas, George Mason University; Tanya Anagnostopoulou, Hellenic Institute of Psychology and Health

***Key Words:*** alcohol consumption, college students, classification trees

The present cross-cultural study attempted to identify high-risk subgroups for alcohol consumption among college students using classification trees. Sixty-six American and 66 Greek college students responded to a number of questions regarding alcohol consumption, religious beliefs, attitudes toward drinking, advertisement influences, parental monitoring, and drinking consequences. Heavy drinkers in the U.S. group were characterized by their younger age and lack of religiosity. In contrast, infrequent drinkers were older and more religious. In the Greek group, heavy drinkers denied the negative results of alcohol and used a permissive attitude to justify their drinking, whereas infrequent drinkers were more likely to be monitored by their parents. Classification trees may provide a useful analytical tool in identifying high-risk populations in the field of substance abuse.

### Smoothing U.S. State Life Tables for Years 1999--2001

❖ Rong Wei, National Center for Health Statistics, 3311 Toledo Road, Room 3114, Hyattsville, MD 20782, *rrw5@cdc.gov*; Lester R. Curtin, Centers for Disease Control and Prevention; Robert Anderson, National Center for Health Statistics; Elizabeth Arias, National Center for Health Statistics

***Key Words:*** US mortalities, life tables, reliability criteria, smoothing methods, variances of LT functions

The NCHS publishes race-gender specific life tables for individual States decennially. Many of those tables were not published due to concern over the statistical reliability of the table functions. In constructing State life tables for 1999-2001, two different statistical models were considered: a Bayesian method that borrows strength from the national data and a two-dimensional local weighted regression to use 30 years of State data. Both methods eliminate the arbitrary nature in the past smoothing technique and allow a greater number of life tables to be published. To evaluate the new life table functions, minimum total deaths, number of adjusted death and changes in life expectancy at birth being used for criteria of reliability are calculated. Restrictions on relative standard errors for probabilities of dying and life expectancy at birth are also justified for the new methods.

### Clustered Odds Ratio

❖ Wanjie Sun, The George Washington University, 6110 Executive Blvd., #750, Biostatistics Center, Rockville, MD 20852, *wsun@biostat.bsc.gwu.edu*; Patricia Cleary, The George Washington University; John M. Lachin, The George Washington University

***Key Words:*** clustered odds ratio, ALR, GEE, bias, efficiency

Clustered binary data occur frequently in many fields, especially in genetics where we try to estimate sib-sib Odds Ratio (OR) in presence of multiple siblings. Standard methods that ignore the within-cluster correlation underestimate the variance of OR, therefore, inflate the type-1 error rate. In this paper we compared the bias and efficiency of four methods: 1) ordinary OR, 2) clustered OR from Rao & Scott (1992), 3) clustered OR from first-order GEE (Liang & Zegar, 1986), and 4) clustered OR from second-order GEE, Alternative Logistic Regression (ALR) (Carey etc.1993). Simulation shows that ALR is the most consistent, and most efficient of the four. And the efficiency improves with growing cluster size n. So besides the fact that ALR models the main effect (y/x relationship) and association effect (OR/x relationship) simultaneously, it is also the most robust method of the four.

## 145 Survey-Based Estimation I ●

Section on Survey Research Methods
**Monday, August 7, 10:30 am–12:20 pm**

### Estimating Dynamic Price Indexes

❖ Alan Dorfman, Bureau of Labor Statistics, 7807 Custer Road, Bethesda, MD 20814, *dorfman_a@bls.gov*; Mary Kokoski, Bureau of Labor Statistics

***Key Words:*** hedonic price index, sampling

Standard price indexes fix on a set of goods, each very precisely specified as to what it is and where sold. Measurements of the change in price of each good are made, and these combined by formula to get an overall measure of change. In a “dynamic universe”, goods disappear and new ones appear, making a fixed product index difficult to construct. In cross area indexes, in which we compare the overall prices of goods in one area to those in another, there may be no common specific goods available. In both cases, indexes are constructed using hedonic regression. In hedonic regression, prices are modeled on the properties of the goods, and the properties serve as the basis of comparison from one time period to the next or one area to another. We here investigate the behavior of dynamic price indexes compared to standard indexes, and of estimators of dynamic indexes based on samples.

### Categorical Time Series Modeling Using Data from the Current Population Survey

❖ Stephen Miller, Bureau of Labor Statistics, 2 Massachusetes Ave., NE, room 4985, Washington, DC 20212, *miller_s@bls.gov*

***Key Words:*** labor force estimates, gross flows, model assisted sampling, measurement error

The Current Population Survey (CPS) collects labor force information from individuals in approximately 50,000 households each month. The 4-8-4 rotation sampling design of the CPS results in approximately a 75% overlap of respondents month-to-month, and approximately 50% over a twelve month span. This overlap allows for the estimation of Gross Flow (month-to-month) changes in Labor Force status, as well as improved estimation of monthly Stock estimates via Composite Estimation. We propose a unified modeling approach which combines both cross-sectional, time series, and (potentially) auxiliary information to estimate labor force classifications over time. We examine several micro-level modeling approaches, and combine the model-based approach along with the design-based approach through the use of estimating equations. Some additional modeling of measurement error is also examined.

### An Examination of Poststratification Techniques for the Behavioral Risk Factor Surveillance System

❖ Michael P. Battaglia, Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138, *mike_Battaglia@abtassoc.com*; Martin R. Frankel, Abt Associates Inc.; Michael W. Link, Centers for Disease Control and Prevention

***Key Words:*** raking, weighting methodology, current population survey, mean squared error

Random-digit-dialing surveys such as the Behavioral Risk Factor Surveillance System typically poststratify on age by gender using control totals from a source such as the 2000 Census, the Current Population Survey, or the American Community Survey. Using logistic regression and CHAID we identified key main effect demographic variables and important two-factor interactions associated with several health risk factor outcomes measured in the BRFSS. A procedure was developed to construct control totals, which were consistent with estimates of age and gender obtained from a commercial source and distributions of other variables from the CPS. Raking was used to incorporate main effects and two-factor interaction margins into the weighting of the BRFSS data. The risk factor estimates were compared with those based on the current BRFSS weighting methodology and MSE estimates were developed.

### Power Analysis of the Rao-Scott First-Order Adjustment to the Pearson Test for Homogeneity

❖ Sunyeong Heo, Changwon National University, Department of Statistics, 9 Sarimdong, Changwon, 641-773 South Korea, *syheo@sarim.changwon.ac.kr*

***Key Words:*** complex sample, homogeneity test, Rao-Scott first order adjustment, Wald test, distribution of weighted sum of noncentral chisquare random variables, 2003 NHIS (National Health Interview Survey)

In a secondary categorical data analysis of complex sample data, situations often arise when the full variance-covariance matrices of cell proportion estimates are not available but only their variances. In this case, researchers often use the Rao-Scott first order adjustments for inferences. In some circumstances, the type I error rate is greater than the nominal level when using the adjusted test. In addition, there is a concern about the loss of power. I have searched for a way to evaluate powers of the Rao-Scott first order adjustments to Pearson-type test statistics for the homogeneity test, and compared them with the powers of the Wald test. To evaluate power, I used Solomon and Stephens' three-moments approximation. The proposed methods are applied to 2003 NHIS (National Health Interview Survey) public use data.

### Estimating Counterfactual Temporal Trends

❖ Andrea Piesse, Westat, 1650 Research Blvd., RE476, Rockville, MD 20850, *andreapiesse@westat.com*; David R. Judkins, Westat

***Key Words:*** temporal trends, dose-response analysis, propensity scoring, program evaluation

Recent evaluations of advertising campaigns addressed at changing health behaviors have used both uncontrolled trends in outcomes and confounder-controlled exposure-outcome relationships to measure campaign effects. Interpretation of these effects can be problematic when, for example, there is a significant exposure-outcome association but the trend is flat. We have developed methodology for decomposing a trend into separate "counterfactual temporal trends" for exposed and unexposed individuals. The counterfactual trends are controlled for confounders through propensity scoring and weighted to represent the populations in which either all individuals are exposed or all are unexposed. A graphical display of the overall and counterfactual trends helps to isolate effects of the campaign from those of other societal forces and is a useful communication tool when dealing with a lay audience.

### Testing the Robustness of HLM and SUDAAN

❖ Jiaquan Fan, Westat, 1650 Research Blvd., Rockville, MD 20850, *jiaquanfan@westat.com*; David R. Judkins, Westat

***Key Words:*** multi-level modeling, semiparametric ANOVA, clustered data analysis

Two statistical software packages, HLM and SUDAAN, can be used to analyze clustered data. It has been suggested that SUDAAN is more robust than HLM to departures from normal homoscedastic errors. We investigated this question directly with a series of simulations. Specifically, we examined robustness in the context of a randomized complete block design, where each "plot" is a small group of children at the same nursery school and a series of measurements of each child are made. We constructed a series of superpopulations in which the standard assumptions of hierarchical (mixed effects) linear models were violated. In addition to testing HLM and SUDAAN, we tested SAS PROC MIXED, and a semi-parametric analysis of variance along the lines discussed by Rosenbaum (2002).

### Hierarchical Linear Modeling Using Complex Survey Data Based on the Pre-Elementary Education Longitudinal Study (PEELS)

❖ Hyunshik Lee, Westat, 1650 Research Blvd., Rockville, MD 20854, *LeeH1@westat.com*; Frank Jenkins, Westat; Sunyeong Heo, Changwon National University

***Key Words:*** multi-level modeling, HLM 6.0, growth curve analysis, simulation study

PEELS studies children with disabilities using a multistage survey design. Children are followed for 5 waves at which numerous background and cognitive measures are gathered. The goal is to identify correlates of cognitive development for the population. The study will gauge development by estimating cognitive growth curves for individual children and the average growth for key subgroups. The HLM 6 program will be used for the analysis. It is not known how well the program estimates growth and variance parameters for data with complex sampling, missing data points, and imputed values for missing variables. This study evaluates the performance of HLM by analyzing simulated data with sample and missing data characteristics varying over a plausible range while on average similar to the actual PEELS data. Bias and precision of estimates as functions of sample factors will be determined.

## 146 Disease Surveillance Methods

● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

## Disease Outbreak Surveillance: Using Prescription Data as a Proxy Source for Detection

❖ A. Elizabeth Allen, IMS Health, 660 W. Germantown Pike, Plymouth Meeting, PA 19462, *eallen@us.imshealth.com*; Kennon Copeland, IMS Health

***Key Words:*** disease, outbreak, prescription, data

Geographic patterns of diseases are of critical interest for public health monitoring and policy. However, information used to estimate disease patterns is generally limited, either in terms of geographic or population coverage. By contrast, availability of information on pharmaceutical drug dispensing is very comprehensive. Pharmaceutical drug information has been found to allow a broader sense of potential disease outbreaks as it is more timely and provides higher coverage than physician reported data. In a prior study, Copeland and Allen (2005) showed Tamiflu prescription data evidenced similar outbreak trends as influenza-like-illness reported by the Centers for Disease Control (CDC) at both national and regional levels. Further analysis here looks at more sensitive local outbreak detection methods, as well as incorporating daily prescription estimates for more timely detection.

## A Comparison of Two Methods for Disease Surveillance

❖ Gerald Shoultz, Grand Valley State University, Department of Statistics, 1133 Makinac Hall, Allendale, MI 49401, *shoultzg@gvsu.edu*; Paul Stephenson, Grand Valley State University; J. Wanzer Drane, University of South Carolina

***Key Words:*** cusum, disease surveillance, Texas, Shewhart

This paper compares two methods for testing hypothesis usable in disease surveillance and process control to determine which is more powerful: TEXAS (Hardy et al 1980) and CUSUM (Hawkins and Olwell 1998). TEXAS, a modification of the procedures of Shewhart (1931), uses a two-step decision rule to determine when a process is under control. CUSUM finds a process to be out-of-control when the sum of a set of measurements exceeds a given threshold. First, the authors will discuss how these process control procedures can be used to monitor disease surveillance. Then the authors will present a simulation that compares the performance of the TEXAS and CUSUM methods to examine which method is more powerful for a variety of hypotheses.

## Prospective Surveillance of Influenza Data Using Hidden Markov Models

❖ Al Ozonoff, Boston University, 715 Albany Street, T4E, School of Public Health, Boston, MA 02118, *aozonoff@bu.edu*; Paola Sebastiani, Boston University

***Key Words:*** influenza, surveillance, hidden Markov models

We propose a Hidden Markov model approach to investigating spatio-temporal patterns in influenza mortality data. Motivated by prospective surveillance methods, as used in syndromic surveillance, we consider various levels of spatial aggregation and calculate deviations from expected patterns of disease in order to better understand the dynamics in both space and time of influenza transmission, morbidity, and mortality. We demonstrate the improved fit of HMMs to influenza data and consider the implications for prospective surveillance efforts at the national, regional, and local levels.

## Modeling and Prediction of Influenza Outbreaks Using Chest Radiograph Data

Peter F. Craigmile, The Ohio State University; Namhee Kim, The Ohio State University; ❖ Soledad Fernandez, The Ohio State University, 320 W. 10th Ave., Center for Biostatistics, Columbus, OH 43221, *fernandez.71@osu.edu*; Bema Bonsu, Columbus Children's Hospital

***Key Words:*** one step prediction errors, filtering, alarm processes, regression models, seasonal autoregressive moving average models

We apply time-series methods to the modeling and prediction of influenza outbreaks using chest radiograph data in children's emergency departments. These models incorporate available covariate information, such as (possibly smoothed) visit counts and temperature series, as well as time-series dependencies on daily and weekly seasonal scales. Outbreaks detection is based on filtering the one-step-ahead prediction errors obtained from these time-series models. Using simulation experiments, we illustrate the effect of the choice of filter and the statistical models on outbreak detection.

## Detection of Outbreaks in Syndromic Surveillance Data Using Monotonic Regression

❖ Jared Burdin, The MITRE Corporation, 202 Burlington Road, Bedford, MA 01730, *jburdin@mitre.org*; James Dunyak, The MITRE Corporation; Mojdeh Mohtashemi, The MITRE Corporation/MIT/AI Lab; Martin Kulldorff, Harvard Medical School/Harvard Pilgrim Health Care

***Key Words:*** syndromic surveillance, outbreak detection, generalized maximum likelihood ratio test, monotonic regression, pool-adjacent-violators algorithm, Poisson regression

Due to nonstationarity and substantial variability in outbreak profiles, early detection of disease outbreaks is challenging. In this paper we propose a method to detect outbreaks in syndromic surveillance data using a generalized maximum likelihood ratio test in which both the null and alternative hypotheses are Poisson distributed. The data is daily counts of interactions between patients and the Harvard Pilgrim Healthcare System in the Boston area. Using Poisson regression, we estimate the daily means and variances of the data as well as day of the week variations. The estimated means serve as the means under the null hypothesis. To determine the means under the alternative hypothesis we use a generalized form of the Pool-Adjacent-Violators algorithm on seven day windows of data. For each window a test statistic is computed and an outbreak is indicated if it exceeds a threshold.

## Surveillance of Occupational Drivers Using k Nearest Neighbor Methods on the Line

❖ Svetla Slavova, University of Kentucky, KY Injury Prevention and Research Center and Department of Statistics, 333 Waller Ave, Suite 202, Lexington, KY 40504, *ssslav2@email.uky.edu*; Terry Bunn, University of Kentucky; Dmitri Pavlov, Pfizer Inc.; Richard J. Kryscio, University of Kentucky

***Key Words:*** case-control, density ratio, k nearest neighbor, parallel implementation, occupational driving, surveillance

In the surveillance of non rare events on the line the ratio of the densities of cases to controls is used to define a local cluster whenever this ratio exceeds a predetermined constant. A nonparametric k Nearest Neighbor method (kNNM) is used to estimate this ratio since with large samples it yields simple to compute confidence intervals for the ratio and straightforward edge corrections. Numerical simulations confirm these asymptotic properties. A parallel implementation of the kNNM is proposed which does not compromise these properties. Applications to two large Kentucky Occupational Safety and Health Surveillance databases show that out of state drivers have significantly higher accident risk past age 50 with opposite results for the youngest drivers and that occupational drivers have a higher accident risk during the early morning hours.

### A Mathematical Model for the Influence of the Pneumococcal Vaccine on S. pneumoniae Vaccine and Nonvaccine Serotypes

❖ Robertino Mera, GlaxoSmithKline, 5 Moore Drive, MAI B209 2D, Research Triangle Park, NC 27709, *robertino.mera@gsk.com*

***Key Words:*** Streptococcus pneumoniae, mathematical models, mixed effect models

Coverage of a conjugate vaccine that provides protection against the most common serotypes was 73% in 2004. Concern exists over an increase in carriage and disease by nonvaccine serotypes. A deterministic transmission model was used to study how vaccination and herd immunity (direct and indirect effects) influence the coexistence of vaccine and nonvaccine serotypes. Insights from the mathematical model were utilized to fit a generalized linear mixed-effects model to surveillance data from before (1998--1999) and after (2003--2004) the introduction of the vaccine. Direct and indirect effects act solely through the transmission coefficient on the reproductive number equation. The model correctly predicts complete serotype switch before coverage reaches 90% due to population herd immunity. Serotype replacement was complete by 2004, with a lower prevalence of invasive disease.

## 147 Reliability and Life Testing

Section on Physical and Engineering Sciences, Section on Quality and Productivity

**Monday, August 7, 10:30 am–12:20 pm**

### Estimating Failure Probability: Exploring Methods

❖ Jason Overstreet, Honeywell Inc., 13350 U.S. Highway, MS 7321, Clearwater, FL 33764, *jason.overstreet@honeywell.com*; Eric B. Howington, Coastal Carolina University; Alexia M. Athienitis-Makris, Applied Health Outcomes

***Key Words:*** reliability, Weibull, logistic regression, failure probability, censored

A common problem in the field of engineering and biology is estimating the probability of a “failure” based on an associated time variable. There are a number of methods available that allow the analyst to estimate the probability of failure with a prescribed level of certainty. Two popular techniques that are widely used to estimate failure probability based on a continuous variable (usually time) are Weibull analysis and binary logistic regression. This paper will present some unique results on the application of these and other methods that we hope will be of practical value to the practitioner.

### Graphical Estimators of Location and Scale from Probability Plots with Censored Data

❖ Anupap Somboonsavatdee, University of Michigan, Department of Statistics, 1085 South University 439 West Hall, Ann Arbor, MI 48109-1107, *asomboon@umich.edu*; Vijay Nair, University of Michigan; Ananda Sen, University of Michigan

***Key Words:*** asymptotic distribution, graphical methods, location scale, least squares, censored data

Probability plots are popular graphical tools for assessing distributional assumptions in reliability applications. These are particularly well-suited for the location-scale family of distributions. When the plot indicates a conformity to the assumed distribution, it is common practice to make an initial guess about the distributional parameters using the least-squares line through the plot. Indeed, current versions of software---such as Minitab---use this as the default estimation method. In this talk, we investigate this quick and dirty method of estimation with censored data and compare it to the maximum likelihood. Specifically, weighted least squares estimators based on the probability plot are considered and their asymptotic properties contrasted with those of the MLEs. Small-sample behavior of the estimators is investigated through an extensive simulation study.

### Design of Sudden Death Tests for Estimation of a Weibull Percentile

❖ John McCool, The Pennsylvania State University, 30 E. Swedesford Road, Malvern, PA 19355, *mpt@psu.edu*

***Key Words:*** sudden death tests, maximum likelihood, precision, type II censoring, Weibull distribution

Sudden Death Tests are a way of conducting life tests in which n specimens are divided into g groups each of size m (n=gm). Testing continues on each group until the first of the m failures occurs in each. The testing thus results in g failures among the n specimens. It is shown how to determine the group size m so that the p-th percentile of the Weibull distribution of life may be estimated with greater precision than in a conventional life test wherein n specimens are tested until the occurrence of the g-th failure. Comparisons of the expected duration of the life tests are given in both cases for several combinations of g, m and p.

### Accelerate Life Test Planning with Independent Weibull Competing Risks with Known Shape Parameter

❖ Francis Pascual, Washington State University, Mathematics Department, PO Box 643113, Pullman, WA 99164-3113, *jpascual@math.wsu.edu*

***Key Words:*** Arrhenius relationship, d-optimality, Fisher information, maximum likelihood, inverse-power relationship, time failure censoring

We present methodology for accelerated life test (ALT) planning when there are two or more failure modes or competing risks which are dependent on one accelerating factor. We assume that the failure modes have respective latent failure times, and the minimum of these times corresponds to the product lifetime. The latent failure times are assumed to be independently distributed Weibull with known common shape parameter. We present expressions for the Fisher information matrix and test plan criteria. We apply the methodology to ALT of Class-H insulation for motorettes where temperature is the accelerating factor. We also present two-level and 4:2:1 allocation test plans based on determinants and on estimating quantiles or hazard functions.

### Proportional Odds Families of Lifetime Distributions

❖ James Gleaton, University of North Florida, 4567 St. Johns Bluff Road, S., Jacksonville, FL 32224-2645, *jgleaton@unf.edu*; James Lynch, University of South Carolina

***Key Words:*** proportional odds, equivalence relation, stochastic ordering, Kullback-Leibler information

Given a lifetime distribution function W(x), another lifetime distribution may be generated through a proportional odds transformation

$W(x; t) = W(x)/(W(x) + t(1 - W(x)))$, where $t > 0$. It is shown that the p.o. transformation partitions the set of all lifetime distributions. In each equivalence class, all members are related through p.o. transformations. Within each class, there is a stochastic ordering of distributions according to t, with dispersion of the transformed distribution being a strictly increasing function of t. Conditions for the existence of moments and moment-generating functions are established; it is shown that either every member of a class has a m.g.f., or none does. It is also shown that the Kullback-Leibler information function for two distributions related by a proportional odds transformation is a strictly increasing function of t, for $t > 1$.

### Robust Prediction and Extrapolation Designs for Censored Data

❖ Xiaojian Xu, University of Alberta, 109 Lees Ave., Apt. 619, Ottawa, ON K1S 5L5 Canada, *xiaojian@ualberta.ca*

***Key Words:*** regression design, extrapolation, censoring, nonsmooth optimization, accelerated life testing, maximum likelihood estimation

We present the construction of robust designs for a possibly misspecified generalized linear regression model when the data are censored. The minimax designs are found for maximum likelihood estimation in the context of both prediction and extrapolation problems with or without restraint of design unbiasedness. This paper extends preceding work of robust designs for complete data by incorporating censoring and maximum likelihood estimation. It also broadens former work of robust designs for censored data by considering both nonlinearity and much more arbitrary uncertainty in fitted regression response and by dropping all restrictions on the structure of regressors. Solutions are derived by a nonsmooth optimization technique. A typical example in accelerated life testing also is demonstrated.

### Comparison Sequential Testing for Reliability: Optimal Test Truncation

❖ Genady Grabarnik, IBM T. J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, *genady@us.ibm.com*; Haim Michlin, Technion-Israel Institute of Technology

***Key Words:*** quality and reliability, sequential testing, comparison testing, MTBF ratio

The study deals with simultaneous testing of two systems, one “basic” (subscript b) and the other “new” (n) - with an exponential distribution of their times between failures. It is checked whether the mean TBFn/MTBFb ratio equals a given value, versus whether it is smaller than the latter. These tests yield a binomial pattern. A recursive algorithm was implemented for calculating the probability of a given combination of failure numbers in the systems, permitting rapid and accurate determination of the test characteristics. The influence is analyzed of truncation of Wald’s SPRT on the characteristics of the sample number up to termination at a given operating characteristic. These are the SN distribution up to acceptance/rejection of the null hypothesis, and it’s mean. The results were applied to the comparison estimate of naval electronic equipment quality and reliability.

## 148 Assessment of Student Performance ●

Section on Statistical Education

**Monday, August 7, 10:30 am–12:20 pm**

### Pedagogical Utilization and Assessment of the Statistic Online Computational Resource in Introductory Probability and Statistics Courses

❖ Juana Sanchez, University of California, Los Angeles, 8125 Math Sciences Building, Box 951554, Los Angeles, CA 90095-1554, *jsanchez@stat.ucla.edu*; Ivo Dinov, University of California, Los Angeles; Nicolas Christou, University of California, Los Angeles

***Key Words:*** Java applets, online course material, education research, teaching with technology, probability and statistics

The NSF-funded Statistics Online Computational Resource (SOCR) provides a number of interactive tools for enhancing instruction in various undergraduate and graduate courses in probability and statistics. Quasi-experimental studies to assess the effectiveness of technology in student learning outcomes and other outcomes are rare in Statistics Education. In this paper we present the results of a study with quasi-experimental components that assesses the effectiveness of using SOCR in Introduction to Probability classes and Introduction to Statistics for the Life Sciences. We use three different designs each with a treatment (SOCR enhanced instruction) and a control (traditional instruction) groups. We compare the SOCR and control groups’ demographics and learning styles at baseline. And then, at the end of the study, we compared the outcomes of the study.

### Assessing College Students’ Success in the Elementary Statistics Course

❖ Jen-Ting Wang, SUNY at Oneonta, 322 Fitz, SUNY Oneonta, Oneonta, NY 13820, *wangj@oneonta.edu*; Shu-Yi Tu, University of Michigan; Yann-Yann Shieh,

***Key Words:*** correlation, coefficient of determination, multiple regression analysis

Elementary Statistics is a required course for most college students. Research has been conducted for determinants of achievement in college mathematics courses; however, there has been little investigation for statistics courses. In this exploratory study, data concerning students’ grades received in this course, the academic performance in high school and in college, as well as numbers of collegiate credits earned were collected from a public four-year college. This study aims to identify the most significant factors of students’ grades in this course. In addition, comparisons between performances of male and female students as well as those of freshmen and non-freshmen were examined. Class size effect was also discussed. In addition to searching for the most important factors, a prediction model for the course grade was established from multiple linear regression

### Be Realistic! Analysis and Pedagogical Benefits of Soliciting Students’ Exam Score Estimates

❖ Douglas M. Andrews, Wittenberg University, Department of Math and Computer Science, PO Box 720, Springfield, OH 45501, *dandrews@wittenberg.edu*

***Key Words:*** exam estimation, pedagogy, regression

Students were invited to estimate their test scores as they submitted each exam in twelve sections of two introductory statistics courses spanning four years. This simple practice takes essentially no class time but has several pedagogical benefits. Most significantly, students become much more realistic throughout the semester about how much they do and don’t understand. The resulting dataset can also be used to illustrate many principles of data analysis and regression modeling in introductory and intermediate courses, all in a context of inherent interest to students. Curiously, females tend to overestimate by less than males - though this sex difference can be explained almost entirely by the more pervasive tendency of higher-scoring students to overestimate by less than lower-scoring students, coupled with the tendency of females to score higher than males in these courses.

### Predictors of Group Performance on Projects

❖ William L. Harkness, The Pennsylvania State University, 318
Thomas Building, Department of Statistics, University Park, PA
16802, *wlh@stat.psu.edu*

***Key Words:*** group projects, group characteristics, predictors of performance

We examine the relationship between characteristics of groups--gender composition, attitude towards statistics, experience with group work, test anxiety, motivation, grade point average--based on survey questions, and various assessment measures, including performance on midterm and final exams, group scores on projects, lab quizzes, and homework. The study was conducted in an elementary statistics course in the Fall Semester, 2004 and replicated one year later. The course met three times per week, with one large group meeting and two meetings per week in computer labs in sections of about 75-80 students each. Records on 280-300 students in 75+ groups were obtained. Groups were self-selected of sizes 3-5 based on seating area. The basic research questions were concerned with ascertaining factors that affect student performance on group projects.

### Are the Students Ready for the Challenge?

❖ Mammo Woldie, Texas Southern University, 3100 Cleburne,
Houston, TX 77004, *woldie_mx@tsu.edu*

***Key Words:*** competency test, business statistics, remedial measures

A mathematics competency test was administered at the beginning of each semester to undergraduate and graduate business students. These students were enrolled in three different business statistics courses. The test scores of these groups were used to determine their level of preparedness for the courses in which they were enrolled. An attempt will be made to point out areas of weaknesses and then suggest remedial measures.

### A Comparison of Gender Performance on Parallel Mathematics Questions

❖ Kenn Pendleton, GED Testing Service, 13537 Walnutwood Lane,
Germantown, MD 20874, *kenn_pendleton@ace.nche.edu*

***Key Words:*** mathematics, education, gender, constructed-response

The new Series 2002 GED Mathematics tests contain both constructed-response (CR) and traditional multiple-choice (MC) items. A study was conducted to investigate whether response format on parallel mathematics items affected gender performances differently. Four pairs of parallel items, one requiring a constructed response and the other a multiple-choice response, were created. Each pair was spaced widely within one of four 25-item tryout booklets. Two booklets presented the CR item first, and two presented the MC item first. Each of the booklets was administered to approximately 140--170 graduating high school seniors divided by gender fairly evenly. The performances of the males and females on each pair of items were analyzed using both chi-square tests and t-tests of correlations. No significant difference in gender performance was revealed.

### Changing Answers in Exams: for the Better or for the Worse?

❖ Juergen Symanzik, Utah State University, Department of
Mathematics and Statistics, Logan, UT 84322-3900, *juergen.symanzik@usu.edu*; Natascha Vukasinovic, Utah State University

***Key Words:*** teaching, electronic textbook, cyberstats, multiple answers

Throughout the year 2004, Utah State University offered three introductory long-distance statistics courses (Stat 2000) for students in its International Program in Hong Kong. The courses were based on the electronic textbook CyberStats. The exams were given electronically and all student answers were stored in the CyberStats data base. `All answers' means that if a student changed the answer to a multiple-choice question or corrected a typo in a text-based answer, all previous answers were still accessible and not only the final answer. In this talk, we will address the question whether students gained more points when changing their answers or when keeping their first answer.

## 149 Insights for Advanced Undergraduate Statistics Courses ●

Section on Statistical Education, Section on Statistical Consulting

**Monday, August 7, 10:30 am–12:20 pm**

### Writing Experiences in a Second Statistics Class

❖ Terry King, Northwest Missouri State University, Department
of Mathematics and Statistics, 800 University Drive, Maryville, MO
64468, *tlking@nwmissouri.edu*

***Key Words:*** education, writing

In a General Statistics II class, the students submitted 3-4 assignments that were written in the format of a journal article. The process of choosing a format, and experiences working with the chosen format, will be discussed.

### Datasets for Teaching Statistics and Design

❖ Charles Stegman, University of Arkansas, NORMES 302
West Annex, University of Arkansas, Fayetteville, AR 72701,
*cstegman@uark.edu*; Calli Holaway-Johnson, University of Arkansas

***Key Words:*** data sets, teaching statistics, teaching design

A major focus in teaching courses in statistics, measurement, and design should be the analysis of data. Results can be used to illustrate key concepts underlying the procedures discussed, help students learn how to analyze their own data, aid in interpreting and presenting research results, and contribute to preparing future researchers. This paper presents information on a multitude of data sets applicable for teaching courses at multiple levels, from introductory statistical analysis through regression, ANOVA, multivariate, and advanced measurement. The accompanying CD contains over 100 datasets from a variety of fields and will be distributed free of charge. The datasets on the CD are either public use datasets or Monte Carlo computer generated datasets and are in a format ready for use in statistics courses.

### Attracting the Brightest Students into Statistics

❖ Greg Taylor, Winston-Salem State University, 132 Reynolds
Center, Dept. of Accouting and MIS, Winston-Salem, NC 27110,
*taylorg@wssu.edu*

***Key Words:*** education, k-12, co-curricular, problem solving

The most capable problem solvers are highly sought after by many scientific, engineering, and mathematics-related disciplines, including statistics. Approaches to attracting young minds taken by these disciplines and the on-going efforts of the statistics community are rapidly changing. Many fields, including theoretical mathematics are creating awareness and interest through problem solving communities and co-curricular activities. Many of these activities provide young participants with realistic notions of the types of problems faced by professionals within specific disciplines. Adopting the best of these practices within the statistics community will attract talented youth.

### Intermediate Statistics with SAS: Interactive

❖ Phyllis Curtiss, Grand Valley State University, Department of Statistics, 1 Campus Drive, Allendale, MI 49401, *curtissp@gvsu.edu*

***Key Words:*** interactive, labs, SAS, programming

The best way to learn SAS programming is by doing it. Programming skills can only be developed through practice. With this in mind, I am writing a statistics book for an intermediate applied statistics class. The book contains SAS labs. The students follow directions to learn various SAS commands and practice statistical inference procedures, while working at their own pace. There are two types of labs. The first type of lab deals with basic SAS programming. Here the emphasis is on options and commands for the data step, and using SAS procedures. The second type of lab deals with using SAS to obtain output for statistical inference. In this part of the book, I introduce the inference procedure in class and do an example. The students then do the lab to get practice, while learning the SAS coding. I will be showing some examples of the content and ideas behind this book.

### Computer Activities To Support Learning Traditional Math Stat Topics

❖ Mary Parker, The University of Texas at Austin/Austin Community College, 1622 Waterston Ave., Austin, TX 78703, *parker@math.utexas.edu*

***Key Words:*** math stat, computer simulations, activities, estimation

Various computer activities and computer simulations will be presented which support upper-division and master's level students in learning mathematical statistics topics. Some of the topics are (1) compare and evaluate different estimators, (2) choose a test statistic for a hypothesis test, (3) the applicability of asymptotic techniques for various finite sample sizes, and (4) the robustness of some statistical results against deviations from "necessary" assumptions.

### Teaching an Undergraduate Capstone Course in Statistical Consulting

❖ Heather Smith, California Polytechnic State University, San Luis Obispo, Statistics Department, 1 Grand Av, San Luis Obispo, CA 93407, *hsmith@calpoly.edu*; John Walker, California Polytechnic State University, San Luis Obispo

***Key Words:*** statistics education, consulting

Cal Poly's statistics curriculum requires a capstone course in statistical communication and consulting for all undergraduate statistics majors. We have developed a course that teaches effective consulting techniques and strategies through a series of group projects. At the same time, these projects provide a rigorous technical review of statistical methods taught throughout the curriculum. Each project involves introductory and follow-up consulting sessions, which are videotaped and reviewed in class. Initially, the instructors act as clients. Later, the students interact with real clients that are screened by the instructors. We present an overview of the course and a sample project with video of consulting sessions. We also discuss the challenges in teaching a statistical consulting course to undergraduate students.

### Integrating Statistical Knowledge through an Undergraduate Capstone Course

❖ John Walker, California Polytechnic State University, San Luis Obispo, Statistics Department, 1 Grand Av, San Luis Obispo, CA 93407, *jwalker@calpoly.edu*; Heather Smith, California Polytechnic State University, San Luis Obispo

***Key Words:*** statistics education, consulting, learning objectives

Cal Poly's statistics curriculum requires a capstone course in statistical communication and consulting for all undergraduate statistics majors. One objective of the course is to provide students with a capstone experience that helps them to integrate the knowledge and skills from their previous statistics courses. This is difficult because many students think of courses as self-contained units rather than seeing how they fit together in the big picture. Our curriculum itself is partly at fault since little time is spent in class describing how each particular course fits into the whole. We discuss specific difficulties that students have in integrating key skills and understandings from different statistics courses to address consulting problems. We also describe ideas to change major courses to encourage more integration throughout the curriculum.

## 150 Mixture Models and Misspecified Models

Section on Bayesian Statistical Science

**Monday, August 7, 10:30 am–12:20 pm**

### Bayes Methodology Accounting for Uncertainty of Commonality in 'Random Effects' in a Linear Mixed Model

❖ Guofen Yan, University of Virginia, 1942 Powell Creek Court, Charlottesville, VA 22911, *gy4g@virginia.edu*; Joseph Sedransk, Case Western Reserve University

***Key Words:*** hierarchical model, partition

We describe and evaluate Bayesian methodology to improve inference for linear mixed effects models by embedding in a more general model the usual specification that the random effects come from a single normal distribution. If the usual specification is incorrect, standard inferences for the variances and random effects will be unsatisfactory. We demonstrate this analytically through a set of examples. The new methodology may be of special value when the random effects correspond to "small areas."

### How Well Does a Logistic Regression Model Estimated with Complex, Multistage Survey Data Fit Data from a New Sample?

❖ Tyson Rogers, University of Minnesota, Hematology MMC 480 Mayo 8480, 420 Delaware, Minneapolis, MN 55455, *roge0285@umn.edu*

***Key Words:*** small area estimation, NHANES III, hierarchical model, observational study

In the absence of a control group, a single arm observational study often compares health outcomes in its subjects to those in the gen-

eral population. Estimates for the general population may be based on complex, multi-stage national health surveys. While such surveys have large sample sizes, estimates of the prevalence of health outcomes and estimates of relationships between health outcomes and other measured covariates have non-negligible standard errors. Small area estimates for the subset of the population with covariates similar to those of subjects in the observational study can be more uncertain. We develop and demonstrate methodology using posterior samples from a Bayesian hierarchical model to evaluate whether and how a logistic regression model of proteinuria differs in an observational study of kidney donors compared to NHANES III.

### Bayesian Model for Misclassified Binary Response with Covariate Subject to Measurement Error

❖ Anna McGlothlin, Baylor University, 1624 S. 5th Street, Apt. 416, Waco, TX 76706, *Anna_McGlothlin@baylor.edu*

***Key Words:*** misclassification, measurement error, random effects, Bayesian inference

We will consider a Bayesian analysis for modeling a binary outcome that is subject to misclassification. Additionally, an explanatory variable is assumed to be unobservable, but measurements are available on its surrogate. A binary regression model is developed to incorporate the measurement error in the covariate as well as the misclassification in the response. We will investigate the inclusion of a random effect to accommodate additional variation. A Bayesian model selection criterion will be utilized. Markov chain Monte Carlo methods are employed to perform the necessary computations. The methods developed will be illustrated through an example.

### Label Switching in Finite Mixture Models

❖ Tong Wang, University of Southern California, 1138 W. 29th Street, Apt. 12, Los Angeles, CA 90007, *wangtong@usc.edu*; Steven L. Scott, University of Southern California

***Key Words:*** finite mixture model, identifiability, label switching

We examine a label switching phenomenon that can occur in the Bayesian analysis of finite mixture models. Label switching occurs because finite mixture models are only identified up to a permutation of the state labels. Thus the finite mixture of $k$ distributions contains $k!$ symmetric modes. We propose a method of removing label switching by running a Markov chain that is transient with respect to all but one of the modes. The practical implementation of our method is similar to a decision theoretic relabeling scheme of Stephens (2000). Our method offers computational savings over Stephens' method, is easier to apply, and avoids imposing artificial boundaries in the multivariate parameter space.

### Bayesian Inference of Population Structure from Dominant Markers Using Mixture of Betas

❖ Rongwei Fu, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, CB669, Portland, OR 97239, *fur@ohsu.edu*; Dipak Dey, University of Connecticut; Kent E. Holsinger, University of Connecticut

***Key Words:*** mixture of betas, dominant markers, population structure, FST, reversible jump

Hierarchical Bayesian methods provide a natural way to incorporate the hierarchical structure inherent in genetic data and have been increasingly applied to genetic samples. By assuming a common FST across all loci, Bayesian methods have been used to estimate FST directly from dominant markers. However, a common FST across all loci may not be the best representation of the data. So we assume that there are groups of loci with similar mutation rates exhibiting similar degrees of genetic differentiation, thus there is a FST for each group. We develop a model of mixture of betas to model the allele frequency and use reversible jump method to identify and estimate different FST's across loci. Uncertainty about the magnitude of inbreeding is incorporated into the estimate of FST. The model is illustrated with RAPD data from 14 populations of a North American orchid, Platanthera leucophaea.

### Flexible Bayesian Variable Selection in Multivariate Linear Regression

❖ Nan Lin, Washington University in St. Louis, Campus Box 1146, 1 Brookings Dr, Saint louis, MO 63130, *nlin@math.wustl.edu*

***Key Words:*** Bayesian variable selection, multivariate regression, Gibbs sampling, posterior probability, mixture model, latent indicator variable

We extend the stochastic search variable selection method in George and McCulloch (1993) to the linear regression model with multivariate responses. Typically, the predictors are either selected or excluded for all the response variables. Our method allows the selection of predictors different for each response variable by transforming the model representation. In the proposed procedure, we describe multivariate linear regression as a Bayesian hierarchical mixture model, in which latent indicator variables are used to select predictors. Predictors with significant effects can be identified as those with higher posterior probability to be included in the model. A Gibbs sampling scheme is used to generate samples from the posterior distribution. The performance of our method is illustrated using both simulation studies and real data examples.

### The Average Effects of Misspecified Models and Diffuse Interaction Models

❖ Juxin Liu, The University of British Columbia, Room 333, 6365 Agricultural Road, Vancouver, BC V6T 1Z2 Canada, *jliu@stat.ubc.ca*; Paul Gustafson, The University of British Columbia

***Key Words:*** average effect, interaction, diffuse interaction model

In an observational study exploring the association between a health outcome and numerous risk factors, how to measure and interpret interactions amongst the risk factors has been the subject of a lively debate. We consider two aspects of this issue. First, what happens if we fit the data, actually generated from a pairwise interaction model, by an additive model? Here we pay special attention to the average effect of an individual risk factor, i.e., averaged over the joint distribution of the other risk factors. Second, we consider diffuse interaction models, as proposed by Gustafson, Kazi and Levy (2005). We compare this interaction model and a pairwise interaction model in terms of their abilities to detect interaction effects at a given sample size.

## 151 Model Selection and Diagnostics

IMS

**Monday, August 7, 10:30 am–12:20 pm**

### AIC for Change-Point Models

❖ Yoshiyuki Ninomiya, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Graduate School of Mathematics, Fukuoka, 812-8581 Japan, *nino@math.kyushu-u.ac.jp*

***Key Words:*** structural change, model selection, information criterion, random walk, Brownian motion

AIC-type information criterion is generally estimated by the bias-corrected maximum log-likelihood. In regular models, the bias can be estimated by $p$, where $p$ is the number of regular parameters. This presentation considers the AIC-type information criterion for change-point models, which are not regular, the bias of which will not be the same as for regular models. Under some contiguity condition, it is shown that the bias can be evaluated by $3m+p$, where $m$ is the number of change-points, which differs from regular models.

## Iterative Bias Correction on Cross-Validation

❖ Hirokazu Yanagihara, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, 305-8573 Japan, *yanagi@sk.tsukuba.ac.jp*; Hironori Fujisawa, The Institute of Statistical Mathematics

***Key Words:*** bias correction, cross-validation, leave-k-out cross-validation, model selection, asymptotic expansion, divergence

It is known that the cross-validation (CV) criterion is a second-order unbiased estimator of the risk between the candidate model and the true model. In this talk, we show that a 2k-th higher-order unbiased estimator can be proposed by using a liner combination of the leave-one-out,...,leave-k-out CV criteria. The distinguishing point is that the proposed idea can give a smaller bias than a jackknife method without any analytic calculation. We verify by numerical experiments that the proposed estimator has a smaller bias than other criteria including the generalized information criterion (GIC), the extended information criterion (EIC), the ordinary CV criterion, and so on.

## Model Assessment Tools for a Model False World

❖ Jiawei Liu, Georgia State University, 718 COE Building, 30 Pryor Street, Atlanta, GA 30303, *jliu@mathstat.gsu.edu*; Bruce G. Lindsay, The Pennsylvania State University

***Key Words:*** model selection, statistical distance, model credibility index, normality, bootstrap

A standard goal of model evaluation and selection is to find a model that approximates the truth well while at the same time is as parsimonious as possible. In this paper we emphasize the point of view that the models under consideration are almost always false, if viewed realistically, and so we should analyze model adequacy from that point of view. We investigate this issue in large samples by looking at two types of model credibility indices, which are designed to serve as one-number summary measures of model adequacy. We define these indices to be the maximum sample size at which samples from the model and those from the true data generating mechanism are nearly indistinguishable. We show that these definitions lead us to some new ways of viewing models as flawed but useful. These concepts are an extension of some important work of Davies (1995).

## Context Tree Estimation for Not Necessarily Finite Memory Processes, via BIC and MDL

❖ Zsolt Talata, Georgia Institute of Technology, 686 Cherry Street, NW, School of Mathematics Georgia Tech, Atlanta, GA 30332-0160, *talata@math.gatech.edu*; Imre Csiszar, AlfrÈd RÈnyi Institute of Mathematics

***Key Words:*** model selection, context tree, consistent estimation, Bayesian information criterion (BIC), minimum description length (MDL), infinite memory

The concept of context tree, usually defined for finite memory processes, is extended to arbitrary stationary ergodic processes (with finite alphabet). These context trees are not necessarily complete, and may be of infinite depth. The familiar Bayesian Information Criterion (BIC) and Minimum Description Length (MDL) principles are shown to provide strongly consistent estimators of the context tree, via optimization of a criterion for hypothetical context trees of finite depth, allowed to grow with the sample size n as o(log n). Algorithms are provided to compute these estimators in O(n) time, and to compute them on-line for all i< =n in o(n log n) time.

## Variable Selection via Penalized Likehood in Semiparametric Regression

❖ Xiao Ni, North Carolina State University, Department of Statistics, 2501 founders Drive, Raleigh, NC 27695-8203, *xni@stat.ncsu.edu*; Daowen Zhang, sanofi-aventis; Hao Zhang, North Carolina State University

***Key Words:*** penalized likelihood, smoothly clipped absolute deviation, smoothing spline, linear mixed models, partial linear models

We introduce an estimation and model-selection procedure for semiparametric regression models via doubly penalized likelihood. The smoothly clipped absolute deviation (SCAD) penalty is used for variable selection. The proposed method not only selects important variables and estimates their coefficients simultaneously, but estimates the unknown nonparametric function with smoothing splines. We propose a linear mixed model framework in which the unknown parameters are computed iteratively. This framework also enables us to estimate the smoothing parameter as a variance component directly. Theoretical properties of the semiparametric estimators are explored. Simulation results are presented to provide empirical support. Further extension of our proposed method to correlated data will be discussed.

## Highest Posterior Model Selection

❖ Tanujit Dey, Case Western Reserve University, 323 Yost Hall Department of Statistics, 10900 Euclid Avenue, Cleveland, OH 44106-7054, *txd30@case.edu*; Hemant Ishwaran, The Cleveland Clinic; J. S. Rao, Case Western Reserve University

We consider the properties of the highest posterior probability model in a linear regression setting. Under a spike and slab hierarchy we find that the highest posterior model is total risk consistent for model selection, but that it also possesses some curious properties. Most important of these is a marked underfitting in finite samples, a phenomenon well noted in the literature for Bayesian Information Criterion (BIC) related procedures, but not often associated with highest posterior model selection. We employ a rescaling of the hierarchy and show the resulting rescaled spike and slab models mitigate the effects of underfitting.

## Residuals and Diagnostics in Dirichlet Regression

❖ Rafiq Hijazi, United Arab Emirates University, CBE Statistics Department, UAE University Jimi Campus, Al Ain, 17555 United Arab Emirates, *rhijazi@uaeu.ac.ae*

***Key Words:*** compositional data, Dirichlet regression, quantile residuals, Aitchison's distance

Compositional data are rarely analyzed with the usual multivariate statistical methods. One approach to model such data is Dirichlet regression. We present various diagnostic methods for Dirichlet regression models. We discuss the use of quantile residuals to check the distributional assumptions. Measures of total variability and goodness of fit are proposed to assess the adequacy of the suggested models. An R-square measure based on Aitchison's distance is introduced. The likelihood distance is employed to identify the influential compositions. Finally, an example with real data is presented and discussed.

# 152 Missing Covariates, Covariate Measurement Error, and Misclassification ●

Biometrics Section, WNAR, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

## A Pseudolikelihood Approach for Analyzing Nutritional Epidemiologic Data in the Presence of Dietary Measurement Error

❖ Samiran Sinha, Texas A&M University, Department of Statistics, 3143 TAMU, College Station, TX 77843-3143, *sinha@stat.tamu.edu*; Raymond J. Carroll, Texas A&M University; Bani K. Mallick, Texas A&M University

***Key Words:*** calibration study, EM algorithm, estimated likelihood, p-splines, sandwich method

Motivated by a real nutritional epidemiologic data, we develop a semi-parametric method for estimating disease-diet association in the presence of measurement error in the diet variable. Usually in these problems, the primary data consist of a binary disease variable (Y) and a response to Food frequency questionary (FFQ) which is considered as a surrogate variable (Q) for the true diet variable (X), and a smaller study which is called validation data, contain the surrogate variable (Q) and repeated measures of an unbiased instrumental variable of the true diet. At the first stage, using the validation data we estimate how Q depends on the X, and at the second stage we estimate how the disease risk is associated with X. The method is semiparametric in the sense that at both the stages we use P-splines technique to estimate the non-parametric smooth functions.

## Instrumental Variable Estimation in Logistic Regression Models with Measurement Error

❖ Kimberly Weems, North Carolina State University, Department of Statistics, 220 D Patterson Hall, Raleigh, NC 27695, *ksweems@stat.ncsu.edu*; Leonard A. Stefanski, North Carolina State University

***Key Words:*** logistic regression, measurement error, instrumental variables

We consider parameter estimation in logistic regression models with measurement error using a binary instrumental variable. The conditional-score method of Stefanski and Carroll (1987) is used to obtain unbiased estimating equations. We obtain sufficient statistics for the unobserved predictors and the conditional distribution of the observed data given these sufficient statistics. Unbiased score functions that are free of the unknown predictors are then used to derive unbiased estimating equations for the model parameters. Our work generalizes that of Buzas and Stefanski (1996) to non-normal instrumental variables. Simulation results and an application to real data are presented to illustrate this method.

## Simultaneous Inference for Semiparametric Nonlinear Mixed-effects Models with Covariate Measurement Errors and Missing Responses

❖ Wei Liu, The University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC V6T 1Z2 Canada, *wei@stat.ubc.ca*; Lang Wu, The University of British Columbia

***Key Words:*** cubic spline basis, longitudinal data, Monte Carlo EM algorithm, random effects model

Semiparametric nonlinear mixed-effects (NLME) models are flexible for modeling complex longitudinal data. Covariates are usually introduced in the models to partially explain inter-individual variations. Some covariates, however, may be measured with substantial errors. Moreover, the responses may be missing and the missingness may be nonignorable. We propose two approximate likelihood methods for semiparametric NLME models with covariate measurement errors and nonignorable missing responses. The methods are illustrated in a real data example. Simulation results show that both methods perform well and are much better than the commonly used naive method.

## Sieve Maximum Likelihood Estimation for Missing Covariates in Regression Models

❖ Qingxia Chen, Vanderbilt University, 1161 21st Ave., S., S2323 Medical Center North, Nashville, TN 37232, *cindy.chen@vanderbilt.edu*; Donglin Zeng, The University of North Carolina at Chapel Hill; Joseph G. Ibrahim, The University of North Carolina at Chapel Hill

***Key Words:*** missing covariates, generalized linear model, model misspecification, semiparametric efficient, b-spline, profile likelihood

Misspecification of the covariate distribution in missing data problems is studied for regression models with several missing covariates. We propose a new semiparametric method which specifies a fully nonparametric model for the conditional distribution of the missing covariates given the completely observed covariates, assuming the missing covariates are missing at random (MAR). For ease of exposition, we first deal with the problem of one missing continuous covariate. To obtain the estimates, the log of the fully unspecified covariate joint density is approximated by B-spline functions and the estimates of the regression coefficients are obtained by maximizing a pseudo-likelihood function over a sieve space. Such estimators are shown to be consistent and asymptotically normal with their asymptotic covariance matrix achieving the semiparametric efficiency bound. Profile likelihood metho

## Composite Tests Using Matrix Pooling: Minimizing Tests, Maximizing Results

❖ Bethany Hedt, Harvard School of Public Health, 63 Allston Street, No. 2, Cambridge, MA 02139, *bhedt@hsph.harvard.edu*; Marcello Pagano, Harvard School of Public Health

***Key Words:*** pooled testing, surveillance

Using imperfect biological tests to detect infected individuals leads to misclassification, an undesirable outcome. In order to lower misclassification errors, one may perform repeated tests, which increases the cost. Under certain conditions, one may increase the accuracy of the testing procedure at the same time as decreasing the costs of testing by intelligently pooling samples from various individuals. We introduce Matrix Pooling with imperfect tests and show that this algorithm under certain conditions does indeed increase the accuracy of testing whilst decreasing the costs. At the same time, Matrix Pooling decreases the time required for obtaining results when compared to individual tests and other pooling methods.

## The Effect of Differential Misclassification in the Chuuk's Mudslide Study

❖ Tzesan Lee, National Center for Environmental Health, 1600 Clifton Road, CDC, Atlanta, GA 30333, *tjl3@cdc.gov*; Josephine Malilay, National Center for Environmental Health

***Key Words:*** adjusted odds ratio, case-control study, misclassification error, proxy respondent, sensitivity and specificity

On July 2, 2002, Chuuk received almost 20 inches of rain within 24 hours from tropical storm Chata'an. Twelve mudslides on six islands on resulted in 43 deaths, hundreds of injuries, and more than 1,000 displaced residents. Interviews were conducted to assess the association between demographic factors, knowledge of warning systems, behavior response, access to health care, and deaths. Because proxies were used for both cases and controls, a misclassification error was likely to occur when classifying the exposure status of subjects. The aim of this paper is to study the effect of differential misclassification on the estimation of true unknown odds ratio for the death-exposure association in Chuuk's mudslide study.

## 153 Inference

Section on Statistical Computing
**Monday, August 7, 10:30 am–12:20 pm**

### Inference on Reliability in Two-Parameter Exponential Stress-Strength Model

❖ Shubhabrata Mukherjee, University of Louisiana at Lafayette, 110 Cherry Street, Apt 7, Lafayette, LA 70506-3669, *s_mukherjee09@yahoo.com*; Kalimuthu Krishnamoorthy, University of Louisiana at Lafayette; Huizhen Guo, Xavier University

***Key Words:*** coverage probability, generalized confidence limit, generalized p-value, location-scale invariance, power distribution, Pareto distribution

The problem of hypothesis testing and interval estimation of the reliability parameter in a stress-strength model involving two-parameter exponential distributions is considered. Test and interval estimation procedures based on the generalized variable approach are given. Statistical properties of the generalized variable approach and an asymptotic method are evaluated by Monte Carlo simulation. Simulation studies show that the proposed generalized variable approach is satisfactory for practical applications while the asymptotic approach is not satisfactory even for large samples. The results are illustrated using simulated data.

### The Distribution of the S-Statistic for Samples of Size 4 Drawn from Uniform and Exponential Populations

❖ Winston Richards, The Pennsylvania State University, 777 W. Harrisburg Pike, Middletown, PA 17057, *ugu@psu.edu*

***Key Words:*** S-statistic, method of sections, common content

The method of sections is used to determine the common content of a sphere and regular tetrahedron. The dimensions of the regular tetrahedron are held fixed, and the sphere of common center is allowed an increasing radius. As the sphere expands, it will first intersect the faces of the regular tetrahedron, then its edges, then it will circumscribe the regular tetrahedron. Common contents for all the cases mentioned will be determined. These results will then be used to determine the exact distribution of the S-statistic for samples of size 4 drawn from uniform and exponential populations, where S is the biased estimate of the standard deviations of these populations.

### Modified Normal Approximations to the Binomial Distribution

❖ David Vlieger, Northwest Missouri State University, Department of Mathematics and Statistics, 1663 GarrettStrong, Maryville, MO 64468, *vlieger@nwmissouri.edu*

Traditional Normal approximations to the binomial distribution tend to be quite inaccurate for small sample sizes, particularly in the tails. For this reason, approximate confidence intervals for the parameter p of a binomial distribution based on these approximations yield mixed results, and a large body of literature exists addressing these problems. In this study two new distributions (derived from the Normal distribution) are introduced, along with their use as approximations to the binomial distribution. Some of the statistical properties of these distributions are investigated, and the accuracy of the approximations to the binomial distribution is evaluated. The confidence intervals for the binomial parameter p are given and the coverage probabilities and sharpness properties investigated.

### On the Simultaneous Lower Confidence Bounds for Order Restricted Inference

❖ Chu-In C. Lee, Memorial University of Newfoundland, St. John's, NF A1C 5S7 Canada, *clee@math.mun.ca*; Jianan Peng, Acadia University; Lin Liu, University of California, San Diego

***Key Words:*** Dunnett's procedure, Gupta's subset selection, isotonic regression, Kuhn-Tucker theorem, lower confidence bounds

In many experiments researchers are interested in comparing several treatment means with a control mean. When there are some treatments significantly better than the control, it is often of interest to evaluate the difference between the best treatment mean and the control mean and to identify the best treatment. We derive simultaneous lower confidence bounds for the aforementioned difference for the case that treatments are at least as effective as the control and for the case that no restriction is placed on the treatment means and the control mean. The evaluation of the simultaneous lower confidence bound for the difference between the best treatment mean and the control mean is a concave programming problem subject to homogeneous linear constraints. Two efficient computation algorithms are proposed. Expected lower confidence bounds of the two procedures are compared with Dunnett's.

### A Partially Exchangeable Model and Its Applications in Correlated Data

❖ Latonya Garner, University of Mississippi, P.O. Box 3958, University, MS 38677, *lcgarne1@olemiss.edu*; Hanxiang Peng, University of Mississippi

In this talk, we generalize the binomial model by relaxing independence to partial exchangeability. We derive the distribution when the sequence of Bernoulli trials forms a Markov chain. We show that the generalized binomial distribution by George and Bowman (1995) hence the classical binomial distribution is a special case. A simulation study will be conducted to illustrate that the proposed distribution can be used to study correlated data, which is our motivation and our future research.

### A Continuing Study on a New Resampling Method To Reduce Small-Sample Bias: an Extension to Nonnormal Distributions

❖ Haiyan Bai, University of Cincinnati, 522 Riddle Crest Lane, Apt 1, Cincinnati, OH 45220, *baih@email.uc.edu*; Wei Pan, University of Cincinnati

*Key Words:* resampling, small-sample bias, simulation, non-normal distributions

The generalizability of using small samples to make statistical inference has long been a concern in the literature. There exist four commonly used resampling methods-bootstrap, jackknife, randomization exact test, and cross-validation-to deal with the small sample issues, but they have limitations, such as little robust to outliers and dependent observations. Recently, a new resampling method was developed, using a union of randomly generated samples to obtain an enlarged, less biased sample which has larger statistical power, independent observations, and robustness to outliers. However, the enlarged sample from this new method was based on normal distribution. This current study is a continuing research by extending the new method to non-normal distributions, such as exponential, chi-square, and Poisson distributions, which increases the applicability of the new resampling method.

### Shape-Restricted Regression Splines and Applications

❖ Mary Meyer, University of Georgia, Statistics Building, Athens, GA 30605, *mmeyer@stat.uga.edu*

*Key Words:* smoothing, nonparametric, splines, regression

Regression splines are smooth and parsimonious nonparametric regression estimators, but are known to be sensitive to the number and placement of the knots. When shape restrictions such as monotonicity or convexity can be imposed, the fits are robust to knot choices. Consequently, inference about the regression function is feasible. Several examples of practical applications are given.

## 154 Adaptive Methods ●

Biopharmaceutical Section

**Monday, August 7, 10:30 am–12:20 pm**

### A Two-Stage Adaptive Design for Phase III Trials To Establish Noninferiority and Superiority

❖ Yulan Li, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, *yulan.li@novartis.com*; Qing Liu, Johnson & Johnson; Jeffrey Maca, Novartis Pharmaceuticals Corporation

*Key Words:* two-stage adaptive design, adaptive closed-testing procedure, adaptation guidelines, interim analysis, operational bias, conditional power

We consider a phase III trial in which two doses of an experimental drug are compared to an active control to establish a noninferiority claim in efficacy. Based on phase II results, the high dose also is expected to be superior to the active control in efficacy. But its safety profiles are uncertain over the course of treatment in the phase III trial. Due to practical constraints, a sponsor for drug development may not be willing to commit up front to power the trial for demonstrating superiority of the high dose. We propose an exact two-stage adaptive design, by which the sponsor can decide to pursue the superiority objective after an interim analysis only if the high dose still holds up the efficacy expectation and its safety data do not raise major concerns. The procedure strongly controls type I error rates. The issue of operational bias will be discussed for real-life applications.

### Confidence Intervals Following an Adaptive Group Sequential Design

❖ Cyrus Mehta, Cytel Inc., 675 Massachusetts Ave., Cambridge, MA 02139, *mehta@cytel.com*

*Key Words:* adaptive clinical trials, group sequential, repeated confidence intervals, median unbiased estimates, Muller and Schafer

This paper proposes a method for computing confidence intervals with exact or conservative coverage following a group sequential test in which an adaptive design change is made one or more times over the course of the trial. The key idea, due to Muller and Schafer (2001), is that by preserving the null conditional rejection probability of the remainder of the trial at the time of each adaptive change, the overall type I error, taken unconditionally over all possible design modifications, also is preserved. This idea is further extended by considering the dual tests of repeated confidence intervals (Jennison and Turnbull 1989) and of stage-wise adjusted confidence intervals (Tsiatis, Rosner, and Mehta 1984). The method extends to the computation of median unbiased point estimates.

### Implementing Adaptive Designs in Clinical Trials: Risks and Benefits

❖ Christopher Khedouri, U.S. Food and Drug Administration, 4652 Tall Maple Court, Ellicott City, MD 21043, *christopher.khedouri@fda.hhs.gov*; Thamban Valappil, U.S. Food and Drug Administration; Mohammad Huque, U.S. Food and Drug Administration

*Key Words:* adaptive design, non-inferiority, clinical trial, anti-infective drug products, sample size, operational bias

Adaptive clinical trial designs allow for changes in the study design, sample size, or analysis methods based on information observed during a trial. Risks of adaptive designs include statistical, operational and misclassification biases, both real and perceived. Benefits include possibilities for more efficient and ethical trial designs. We explore three types of adaptive sample size designs from the published literature and compare them in both superiority and non-inferiority settings to traditional, fixed sample size designs. We also provide some results based on simulations to address the effects of various misspecifications of assumptions used in the initial sample size calculation. Finally, we summarize the risks and benefits of implementing each of the four approaches. Examples pertaining to clinical trial designs for anti-infective drug products are also discussed.

### Dynamic Treatment Allocation and Randomization Tests in Clinical Trials

❖ Lee-Lian Kim, Centocor R&D, Inc., 200 Great Valley Parkway, Malvern, PA 19355, *lilianne_kim@yahoo.com*; Yaung-Kaung Shen, Centocor R&D, Inc.; Jewel Johanns, Centocor R&D, Inc.; Aparna Raychaudhuri, Centocor R&D, Inc.

*Key Words:* randomization tests, dynamic treatment allocation, SAS

There are different methods to randomize subjects to treatment arms in a clinical trial. One method used is dynamic treatment allocation. Here, the treatment allocated to a subject depends on the distribution of subjects that had been randomized previously. It has been noted by the European Agency for the Evaluation of Medicinal Drugs (EMEA) that if a clinical trial utilizes dynamic treatment allocation to randomize subjects to the treatment arms of the clinical study, randomization tests would be the proper tool to account for the dynamic treatment allocation. In this presentation, we describe the method to perform randomization tests in SAS for categorical data. This can also be expanded to continuous data.

### Resampling Methods for Adaptive Designs

❖ Hui Zhang, Bristol-Myers Squibb Company, F209B Global Biometric Sciences, 5 research parkway, Wallingford, CT 06492, *hui.zhang@bms.com*

***Key Words:*** resampling methods, adaptive designs

In clinical trials, the estimation of effects difference is often of primary importance. Proper resampling methods will provide second-order correct estimates, which will outperform the traditional normal approximation. Bootstrapping has been known for i.i.d. random variables. Unfortunately, traditional bootstrap is not applicable because the observations of adaptive designs are dependent due to adaptive allocation. We address this problem by developing and studying resampling methods for dependent data in adaptive designs, including theoretic results and simulations.

### Evaluating Exploratory and Confirmatory Evidence Collectively

❖ Qian Li, U.S. Food and Drug Administration, 10802 Whiterim Drive, Potomac, MD 20854, *qian.li@fda.hhs.gov*

***Key Words:*** clinical trials, collective evidence, confirmatory evidence, exploratory evidence, flexibly designed

The concept of collective evidence is used to reason that certain valid exploratory information can be used, in conjunction with confirmatory information from at least one phase III clinical trial, to support a claim in a new drug approval. The standard of the collective evidence is equivalent to the requirement of two positive independent phase 3 clinical trials, the common interpretation of the “substantial evidence” required by the regulation. Two cases from investigational new drug submissions are explored to illustrate the potential use of combining exploratory and confirmatory information due to an adaptive design or adaptation in the middle of a trial.

### Adjusted Two-Sided Combination Tests for Adaptive Clinical Trials

❖ Zhilong Yuan, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., PO Box 300, Raritan, NJ 08869, *zyuan1@prdus.jnj.com*; Yang Song, Johnson & Johnson Pharmaceutical R&D; Xiaolong Luo, Johnson & Johnson Pharmaceutical R&D; George Chi, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** adaptive design, combination test, two-sided tests, rejection region, futility boundary

We consider a multi-stage adaptive clinical trial studying a two-sided null hypothesis of no treatment effect, where the evidence from the different stages is combined by a combination test. A two-sided combination test, however, is known to have significant less power than the corresponding one-sided test. We investigate this phenomenon by a geometric analysis of the rejection region, and propose to restrict it to that targeting overall positive and negative effect. A simulation study shows that the loss of power is recovered by the adjusted test. The effect of a futility boundary is also assessed by comparing the adjusted test to a doubly one-sided test which requires treatment effects in all stages to have the same sign. The former is shown to have comparable power, and is also more robust to various factors.

## 155 Analysis of Microarrays ●

Biometrics Section, ENAR

**Monday, August 7, 10:30 am–12:20 pm**

### Biweight Correlation as a Measure of Distance between Genes on a Microarray

❖ Aya Mitani, Pitzer College, 1050 North Mills Ave., Claremont, CA 91711, *aya.mitani@pitzer.edu*

***Key Words:*** biostatistics, microarray

The underlying goal of microarray experiments is to identify genetic patterns across different experimental conditions. Genes contained in a particular pathway or that respond similarly to experimental conditions should be coregulated and show similar patterns of expression on a microarray. Using any of a variety of clustering methods or gene network analyses, we can partition genes of interest into groups, clusters, or modules based on measures of similarity. Typically, Pearson correlation is used to measure distance (or similarity) before implementing a clustering algorithm. Pearson correlation is susceptible to outliers, however, an unfortunate characteristic when dealing with microarray data (well-known to be typically noisy.) We propose a robust similarity metric based on Tukey’s biweight estimate of multivariate scale and location.

### Modified Wilcoxon Mann-Whitney Methods for Identifying Functional Gene Categories in Microarray Experiments

❖ Liu Hua, University of Kentucky, KY 40515, *hualiu@ms.uky.edu*; Constance Wood, University of Kentucky; Arnold J. Stromberg, University of Kentucky

***Key Words:*** microarray, functional gene category, Wilcoxon Mann-Whitney, partial area under the ROC curve, asymptotic

Microarray technology allows investigators to study huge numbers of genes simultaneously. To obtain insight into the biological phenomena based on the differential gene expressions, the over-representation of the functional gene categories is commonly tested. Existing statistical methods for identifying over-represented functional gene categories include Fisher’s exact tests, two-sample Kolmogorov-Smirnov tests, and Wilcoxon Mann-Whitney tests. These tests are not sensitive to differences in distributions over specific subsets of the real line. To improve the power, we propose a truncated Wilcoxon Mann-Whitney statistic that evaluates the partial area under the receiver operating characteristic curve and a one-sided version of this statistics. The asymptotic properties of the two statistic and applications to a microarray experiment will be discussed.

### MicroRNA Regulation of mRNA Expression in Neuronal Development

❖ Diane Richardson, Rutgers University, 501 Hill Center, Department of Statistics, Piscataway, NJ 08854, *drichard@stat.rutgers.edu*; Rebecka Jornsten, Rutgers University

***Key Words:*** microarray, microRNA, regulatory motif, gene expression, clustering, hierarchical mixture

We jointly analyze microarray data of mRNA and microRNA expression in developing cells and neuronal tissue. A hierarchical mixture model is derived to cluster expression patterns that are defined by differentiating patterns over time and/or cell-lines. We discuss how the clustering outcome can be used to discover regulatory motifs of mRNAs and regulatory pathways between mRNA targets and microRNAs. We demonstrate how data bases of transcription factor binding sites can be mined using the joint clustering outcome.

### Point and Interval Predictions of Protein Concentrations in ELISA Microarray Assays

❖ Don S. Daly, Pacific Northwest National Laboratory, P.O. Box 999, MS K6 08, Richland, WA 99354, *ds.daly@pnl.gov*; Kevin K. Anderson, Pacific Northwest National Laboratory; Amanda M. White, Pacific Northwest National Laboratory; Susan S. Varnum, Pacific Northwest National Laboratory; Richard C. Zangar, Pacific Northwest National Laboratory

***Key Words:*** ELISA microarray immunoassay, monotonic spline, propagation of error, Monte Carlo simulation

An ELISA microarray assay predicts simultaneously the concentrations of numerous proteins in a small sample. These predictions are uncertain due to processing error so that prediction error estimates are critical to inferring biological significance and improving the assay. Using real and simulated data, we investigate blends of classic and new methods to make point and interval ELISA concentration predictions: logistic and spline models, NLS and PCLS fits, delta method and Monte Carlo predictors. All blends give acceptable results in most cases. The classic blends are more understandable to biologists but do not always converge and do generate unrealistically divergent intervals. The new blends, though less comprehensible to a layman, are robust and true to the underlying physics. These methods are available as a modular Java/R open source application.

### A Nonparametric Likelihood Ratio Test To Identify Differentially Expressed Genes from Microarray Data

❖ Sunil Mathur, University of Mississippi, Hume 325, Department of Mathematics, University of Mississippi, University, MS 38677, *skmathur@olemiss.edu*; Sankar Bokka, University of Mississippi

***Key Words:*** genes, nonparametric, likelihood ratio, power, efficient

Motivation: Microarray experiments contribute significantly to the progress in disease treatment by enabling a precise and early diagnosis. One of the major objectives of microarray experiments is to identify differentially expressed genes under various conditions. The statistical methods, currently available in literature to analyze microarray data are not up to the mark, mainly due to the lack of understanding of the distribution of microarray data. Results: We present a nonparametric likelihood ratio (NPLR) test to identify differentially expressed genes using microarray data. The NPLR test is highly robust against extreme values and does not assume the distribution of parent population. Simulation studies show that the NPLR test is more powerful than some of the commonly used methods. Asymptotic results are discussed.

### A Statistical Framework To Infer Functional Gene Associations from Multiple Biologically Dependent Microarray Experiments

❖ Siew-Leng Teng, University of California, Berkeley, Department of Statistics, 367 Evans Hall, Berkeley, CA 94720-3860, *slteng@stat.berkeley.edu*

***Key Words:*** microarray, functional gene associations, experimental dependencies, gene dependencies, functional gene relationships

Microarray data from multiple biologically related experiments now allow for a more complete portrayal of gene dynamics in a biological process. However, a critical issue has been widely ignored in current integrative analyses: existence of dependencies among gene expressions across related experiments. These dependencies can arise either due to similar conditions or relevant external perturbations among the experiments, and can result in inaccurate estimates of gene associations and hence incorrect biological conclusions. To address this issue, we propose a statistical framework and Knorm correlation to quantify functional gene associations in presence of such experimental dependencies. Our model underlines a unique dependency structure that maintains the same experiment (gene) correlations across genes (experiments). Our measure yielded promising results in experimental datasets.

### Cluster Analysis for Gene Expression Data with Liquid Association Structure

❖ Yijing Shen, University of California, Los Angeles, 11140 Rose Ave., 114, Los Angeles, CA 90034, *yshen@stat.ucla.edu*; Ker-Chau Li, University of California, Los Angeles; Shinsheng Yuan, University of California, Los Angeles

***Key Words:*** genes, clustering, liquid association, k-means, EM algorithm

An important goal in gene clustering analysis is to identify functionally related gene groups that could be used in functional annotation for new or uncharacterized genes. We analyze small gene sets selected from the Yeast genome by the method of Liquid Association (Li, 2002)-the correlation between the expression levels of such gene pairs depends on the change of a relevant external variable. Traditional approaches may not always provide the best results because functionally related genes may only be co-expressed under a subset of conditions. For such structured datasets, expression profile of the external variable is an important factor in guiding the clustering. By employing the external information we develop a new clustering algorithm that optimally selects two subsets of conditions for generating best clustering; then we refine the clustering result through the use of Gene Ontology.

## 156 Models for Data Collected over Time ●

Biometrics Section

**Monday, August 7, 10:30 am–12:20 pm**

### Analysis of Mixture Random Effects Models for Longitudinal Data

❖ Yimeng Lu, Columbia University, 100 Haven Ave., 26G, New York, NY 10032, *yl2058@columbia.edu*; Hongtu Zhu, Columbia University and New York State Psychiatric Institute; Thaddeus Tarpey, Wright State University; Eva Petkova, Columbia University

***Key Words:*** mixture random effects modeling, Bayesian inference, MCMC, antidepressant study

We propose a mixture random effects model to classify subjects based on repeated measurements over time. The hierarchical model accounts for random subject-specific regression coefficients, component-specific distribution of the random effects, and the covariate effect on classification (i.e., on the multinomial probabilities of subjects belonging to given components of the mixture). A Bayesian procedure is developed to estimate component allocation and mean regression coefficients for each component simultaneously. A Markov chain Monte Carlo method is implemented to produce the solution. The procedure is tested through simulations and applied to longitudinal data originated from an antidepressant study with the goal of identifying different types of responders, such as placebo responders or true drug responders.

### Nonparametric Inference for High-Dimensional Longitudinal Data

❖ Ke Zhang, Kansas State University, 2212 Prairie Glen, Manhattan, KS 66502, *kezhang@ksu.edu*; Haiyan Wang, Kansas State University

***Key Words:*** high dimension, longitudinal, rank statistics, lipidomics

Analysis of high dimensional data has received increasing attention in recent years as large amount of data become available due to fast development of modern technologies. In this paper, we consider testing the main effect of factors and the temporal change of the factor effects when there are a large number of levels for some factors and each subject is measured repeatedly over time. Test statistics are proposed based on both the original observations and their (mid-) ranks. The asymptotic distributions of the test statistics are obtained under corresponding null hypotheses. Simulation results will demonstrate the property of the tests. A lipidomics dataset are analyzed to illustrate the application of the proposed tests.

### Violating the Assumption of Independence of the Error Components in the Linear Mixed Model for Longitudinal Data

❖ Matthew Gurka, University of Virginia, P.O. Box 800717, Charlottesville, VA 22908-0717, *mgurka@virginia.edu*; Lloyd Edwards, The University of North Carolina at Chapel Hill; Keith E. Muller, The University of North Carolina at Chapel Hill

***Key Words:*** variance components, correlation, random effects, model diagnostics, obesity

The linear mixed model is commonly used in the analysis of longitudinal data, where one can simultaneously model the among-unit error (the random effects) and the within-unit error. The effect of violating the normality assumption of these two sources of deviation has been studied. However, the impact of violating the assumption of independence between the random effects and the within-unit error has not received any attention. An examination of data from an intervention trial targeting obesity reveals evidence of a possible violation of the independence assumption. We argue that this violation may occur often in practice, particularly in areas of research like obesity. If the two sources of deviation are dependent, estimation and inference about the variance components of the model, and consequently the random effects estimators, are adversely affected.

### A Marginal Model for Multistate Panel Data under Heterogeneity

❖ Wei-Ting Hwang, University of Pennsylvania School of Medicine, 628 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, *whwang@cceb.upenn.edu*

***Key Words:*** panel data, frailty, Markov process, estimating equation

The goal of this work is to estimate disease progression rates from panel data with multiple stages when the population is heterogeneous because of unobserved variables. Frailty is introduced to describe the heterogeneity in the population. We developed a marginal model based on generalized estimating equations to obtain the parameter estimates. Comparison to conditional model based on the likelihood function will be discussed. The results and methods are applied to data on Alzheimer's Disease (AD) research where repeated measurements on subjects' cognitive functions were collected over time through a Clinical Dementia Rating (CDR) scale. This scale has five ordinal stages which measure the functional impairment in areas such as memory, orientation and language.

### Nonparametric Inference for Panel Count Data

❖ Ying Zhang, The University of Iowa, Department of Biostatistics, Iowa City, IA 52242, *ying-j-zhang@uiowa.edu*

***Key Words:*** panel count data, interval censored data, nonparametric test, asymptotic distribution

We study a simple nonparametric inference procedure for panel count data, a type of complicated data often appearing in clinical trials. We propose an easy-to-implement nonparametric estimation method for the mean function of counting process by maximizing a pseudo-likelihood function established from a nonhomogeneous Poisson process. We derive the asymptotic normality of a smooth functional of the estimator. This smooth function is estimated easily, hence warranting a useful inference procedure for panel count data. We further propose a simple nonparametric test for the comparison of the mean functions among k independent samples. The test is validated through simulation studies and demonstrated by the two real-life examples.

### Estimation of the Mean Function of Panel Count Data Using Monotone Polynomial Splines

❖ Minggen Lu, The University of Iowa, 598 Hawkeye Court, Iowa City, IA 52246, *minggen-lu@uiowa.edu*; Ying Zhang, The University of Iowa; Jian Huang, The University of Iowa

We study the nonparametric pseudo-likelihood and full likelihood estimators of the mean function of a counting process based on panel count data using monotone polynomial splines. The setting for panel count data is one in which n independent subjects, each with a counting process with common mean function, are observed at several possibly different times during a study. Generalized Rosen algorithm was used to compute the estimators. We show the proposed spline estimators are asymptotically consistent and the rate of convergence is higher than 1/3. The simulation study show the spline-based estimators have smaller variances and mean square errors than nonparametric pseudo and full likelihood estimators proposed in Wellner and Zhang (2000). A real Bladder cancer trial example is used to illustrate the method.

## 157 Topic-Contributed Poster Session: Data Exposition

Section on Statistical Graphics

**Monday, August 7, 10:30 am–12:20 pm**

### Using Data Mining Tools in the Study of NASA Ozone Data

❖ Wei-hong Wang, The College of New Jersey, Math Department, Ewing, NJ 08628-4700, *wang@tcnj.edu*; Pin-Shuo Liu, William Paterson University

***Key Words:*** data displays, spatial data & GIS, temporal data, neural network, predictive modeling, genetic algorithm

This paper investigates the ozone data as presented in the 2006 ASA Data Exposition (www.amstat-online.org/sections/graphics/dataexpo/2006.php). The data contains eight variables in 506 files and was collected on 24x24 geographical grids over 72 months. Our investigation treats ozone as the target variable and focuses on the spatial and temporal behaviors of the target and predictors. The study also compares the predictive powers of various data mining tools in the following two areas: regression, neural network, boosted trees, random forest, support vector machines, genetic algorithm, and MARSplines for

cross-sectional data and neural network and temporal system-equation model for time series data. The paper will discuss a variety of ways in the comparison of the models and in the dynamic displays of spatial and temporal data.

## Identifying Outliers in Multivariate Spatial Data

❖ Anthony Franklin, Coastal Carolina University, 2110 Hawksmoor Drive, Conway, SC 29526, *amfrankl@coastal.edu*; Eric B. Howington, Coastal Carolina University; Keshav Jagannathan, Coastal Carolina University

***Key Words:*** spatial analysis, multivariate analysis, outlier detection, data expo

Our entry into the 2006 Data Expo consists of an analysis focusing on detecting unusual data observations. Our approach will incorporate both multivariate and spatial data analysis techniques. An emphasis will be placed on graphical methods for detecting outliers.

## Exploratory Data Analysis of Meteorological Data Using SAS Stat Studio

❖ Frederick Wicklin, SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513, *Rick.Wicklin@sas.com*; Yun Chen, North Carolina State University

***Key Words:*** meteorological data, exploratory data analysis, statistical software, dynamically-linked graphics, SAS STAT studio, semivariogram

We explored the geographic and atmospheric variables in NASA meteorological data. We discovered inconsistent pressure readings, indicating a possible systematic measurement error. We discovered geographic categories for which the atmospheric variables are similar. We adjusted the atmospheric variables to account for the geographic effects, then formulated models that explain the relationship between ground-based temperature measurement and satellite-based measurements. Similarly, we formulated models to explore the seasonal and long-term variation of spatial averages of the atmospheric variables. Lastly, we constructed empirical semivariograms to visualize the spatial dependence of temporal averages of the atmospheric variables. We analyzed this data using Stat Studio, a new statistical software product from SAS that provides dynamically-linked graphics for exploring multivariate data.

## SparkMats: a Graphical Method of Exploring Spatially Distributed Time Series

❖ John Emerson, Yale University, 438 Humphrey Street, New Haven, CT 06511, *john.emerson@yale.edu*; Walton Green, Yale University

***Key Words:*** sparclines, graphical, exploratory, time, space, NASA

The proliferation of data in the computer age has made the search for methods of compressing higher dimensions onto a flat, two-dimensional graphical display more difficult and important. Here, we illustrate a graphical, exploratory approach using a generalization of Edward Tufte's 'sparklines,' or compact time series, that can be used to represent data of continuous variables distributed in space and time. Using this and other graphical tools, we identify certain peculiarities of seven four-year monthly time series of climate data on a 24 by 24 raster covering an area of South and Central America. Our methods of graphical display reveal variable truncation and multiple covariation, which must be addressed before applying predicative autoregressive models.

## Novel Two-Step Process for Graphically Summarizing Multivariate Spatial Temporal Data in Two Dimensions

❖ Svetlana K. Eden, Vanderbilt University, 1161 21st Ave., S., S-2323 Medical Center North, Nashville, TN 37232-2158, *svetlana.eden@vanderbilt.edu*; Theresa A. Scott, Vanderbilt University; Angel An, Vanderbilt University; Jeffrey Horner, Vanderbilt University; Cathy Jenkins, Vanderbilt University

***Key Words:*** seasonality, variability, polar coordinate, spatial, temporal, multivariate

It is difficult to graphically summarize important features of multivariate spatial temporal data in two dimensions. To explore spatial and temporal clustering, a large matrix-like geographic graph was generated for each atmospheric variable. Polar coordinate systems were plotted at each longitude/latitude intersection in the spatial grid; radius designated deviation from a global minimum and angle designated time. Using the data from a chosen set of grid points, a matrix of scatter plots was used to investigate any possible pairwise relationships between the atmospheric variables. Using this method, features such as seasonal, yearly and spatial variability, points of interest, and associations were discerned easily. By incorporating the mentioned sources of variation, our process expedited revealing key details of the data.

## Using Kriging and 3-D Graphics To Explore Trends of Total Column Ozone Amount and Tropospheric Weather Systems in Central America from 1995 to 2000

❖ Kening Wang, University of Arkansas, 346 N. West Ave., 302 WAAX, Fayetteville, AR 72701, *kxw06@uark.edu*; Charles Stegman, University of Arkansas; Sean W. Mulvenon, University of Arkansas; Yanling Xia, University of Arkansas

***Key Words:*** graphics, map, ozone, Central America

In this study, prediction maps of ozone and surface temperature were generated using kriging. Bar charts of air pressure, cloud coverage, and elevation also were plotted on the maps. 3-D graphics were created using PowerPoint. All the graphics were able to be visualized through a web-based interactive system. Ozone exhibits strong seasonality and latitude dependence. Interannual ozone variability is also present. Ozone amounts in the tropics are rather low, with the lowest in winter and the highest in summer. In the northern region, ozone amounts are high in spring, decrease in summer and fall, and rise again in winter. Like ozone, temperature also varies seasonally with latitude, but no direct relationship with ozone is observed. Cloud coverage variability depends on region and season. Air pressure is generally consistent, and it mainly changes with elevation.

## Dynamic Data Visualization of Meteorological Data

❖ Bruce Peterson, Terastat, 22904 NE 51st Street, Redmond, WA 98053, *bapeters@terastat.com*

***Key Words:*** dynamic, animation

Data visualization technologies are applied to complex datasets to leverage the innate human pattern recognition capability. By accentuating selected types of graphical patterns, the analyst can develop hypotheses about the data that may be explored with other techniques or used to develop a statistical model. Advances in computer technology have enabled the use of animation in data visualization, aiding the recognition of dynamical patterns. This poster will illustrate dynami-

cal patterns found in the meteorological dataset identified using a tool built specifically for visualization of dynamic data. A summary of the use of the visualization tool's major components on the meteorological dataset will be presented.

## A Web-Centric Graphical Approach to Gain Insight into NASA's NUMB3RS

❖ Robert Allison, SAS Institute, Inc., 318 Indigo Drive, Cary, NC 27513, *Robert.Allison@sas.com*

***Key Words:*** geographical, temporal, visualization, animation, web-based, SAS

We all use numbers every day, especially NASA. And in lieu of a math genius, graphics can help analyze those numbers and answer questions such as Where is the hole in the Ozone Layer? Are there anomalies in the NASA data? Are there trends and correlations in the NASA data? Are those trends and correlations the same in all geographical areas? What's the hottest or sunniest location for my next vacation? This poster describes new techniques (developed especially for this data) for visualizing and comparing large amounts of data on a single screen. Innovative combinations of maps, small multiples, animations, drill-downs, interactive graphics, graphical bread-crumbs, and thumbnail glyphs are utilized--- all via a simple web browser. With these innovative graphics, the important NUMB3RS jump right out at you---no math genius required!

## Another View at Central America

Hadley Wickham, Iowa State University; ❖ Jonathan Hobbs, Iowa State University, 2804 Stange Road, Apt. 8, Ames, IA 50010, *jonhobbs@iastate.edu*; Dianne Cook, Iowa State University; Heike Hofmann, Iowa State University

***Key Words:*** visualization, graphics, climate data, satellite data

In the legacy of John Tukey, we have been trying to come up with good pictures to "force the unexpected upon us" as the example of this year's ASA Data Exposition. The data are geographic and atmospheric measures on a coarse, 24x24 grid covering Central America. Using conventional graphics and less-conventional interactive graphics, we were able to find familiar features in the data, such as the El NiÒo and La NiÒa events, as well as more surprising results. Do you want to know about the relationship between high temperatures in Bogota in May and a white Christmas in Florida? Come and see for yourself!

## Graphical Display of Model-Based Temperature Data

❖ Jeff Slezak, Mayo Clinic College of Medicine, 200 1st Street, SW, Rochester, MN 55905, *slezak@mayo.edu*

***Key Words:*** graphics, multiple regression, data visualization

Multiple linear regression was used to model temperature, looking especially for time trends and factors with strong time interactions. Graphical methods include map-based display of actual and model-predicted temperature at each time point and a display of model-predicted temporal trends in temperature. The final model (R2=0.89) showed strong spatial and temporal trends, including seasonal temperature variations between northern and southern latitudes. Factors with strong time interactions included latitude, longitude, elevation, and cloud cover. Predicted time trends generally showed small annual gains, although some areas---mostly over oceans---showed slight declines and locations in Mexico and South America showed increases of up to three degrees per year.

## Multiple Lagged Differences in Spatial Time Series

Rafe Donahue, Vanderbilt University Medical Center; ❖ Jeffrey Horner, Vanderbilt University, 1161 21st Ave., S., S-2323 Medical Center North, Nashville, TN 37232-2158, *jeff.horner@vanderbilt.edu*

***Key Words:*** data display, spatial time series

Typical time series graphs plot a measured variable over time to visually demonstrate the features in the data. For strongly periodic atmospheric time series, simply graphing all differencing orders can convey visual periodicity. With the assumption that the atmospheric data contain annual periodicity, graphing the variance of the differencing orders from the mean of each differencing order can further display aberrations and interesting characteristics in the data. We introduce a novel graph to expose spatial time series data by plotting all differencing orders in an array. Our resulting plots clearly show the seasonality and interesting time periods for the air temperature and cloud data. Other weather measures show similar and distinct features. A strong aberration in air pressure data also is shown.

## Exploring Spatial and Temporal Characteristics of Atmospheric Ozone Concentration Using Visualization

❖ Sudeshna Paul, Purdue University, 150 N. University Street, Dept Of Statistics, West Lafayette, IN 47907, *paul0@purdue.edu*; Souleymane Fall, Purdue University; Devdutta Niyogi, Purdue University; Bruce A. Craig, Purdue University

***Key Words:*** global warming, ozone, stratosphere, correlation, depletion

In this work, we investigate NASA Langley Research Center datasets containing primarily monthly geographic and atmospheric measures on a coarse grid covering Central America. Time series analysis was performed to check and adjust for serial correlation and identify important trends in atmospheric ozone concentration. Relationships between ozone and higher atmospheric temperature, pressure, and cloud cover were analyzed. Visualization techniques using ArcGIS and R showed there might be a relationship between ozone depletion and warming of the near-surface atmosphere, also known as "global warming." Additional data obtained from NASA showed indication of cooling of the upper stratosphere, which results in further ozone depletion.

## Visualizing Several Abnormal Climate Changes in Central America from January 1995--December 2000

❖ Sang-Hoon Cho, University of Wisconsin-Madison, 5639 Longford Terrace 104, Fitchburg, WI 53711, *cho@stat.wisc.edu*; Hyonho Chun, University of Wisconsin-Madison; Deepayan Sarkar, State University of Wisconsin

***Key Words:*** El Nino, global warming, ozone, surface temperature, graphics, visualization

We aim to effectively visualize several uncovered abnormal patterns of the dataset. As an initial step, we analyzed the changes of the variables of the dataset at each location on a 24x24 grid over time and found abnormal patterns, including an overall increasing pattern of the surface temperature at each location once adjusted for its elevation, the surge of the surface temperature in the ocean near the coast of Peru during summer 1997, and the surge of the surface temperature at a particular location in Mexico. The first two findings correspond to the natural phenomena that are global warming and El Nino, whereas the last seems to be related to the rapid decrease of the ozone level. Dynamic displays are being considered to show the surface temperature changing over time and plots to emphasize the abnormal surges of the surface temperature.

### Data Display Principles Revealed in the NASA Data

❖ Rafe Donahue, Vanderbilt University Medical Center, 1716 Danforth Park Close, Brentwood, TN 37027, *rafe.donahue@vanderbilt.edu*

***Key Words:*** data display principles, data visualization

The complex nature of the NASA weather data certainly makes it difficult to understand. The data are a multivariate spatial time series with nonstationary components that are both easily identifiable (annual periodicities) and difficult to view (trends and data anomalies). Since typical plots of single variables across time fail to show anything but the most fundamental components of this dataset, further efforts are needed to appreciate the more subtle elements in the data. Working with a fundamental principle that "the data display is the model" for the data, plots that show pertinent sources of variability in the data will be developed in an effort to better understand the nature of the dataset as a whole.

## 158 Contributed Posters

General Methodology, Section on Statistical Computing, Biometrics Section, Section on Statistics and the Environment, Section on Statistical Graphics

**Monday, August 7, 10:30 am–12:20 pm**

### Intervention Models To Avoid

❖ Bradley Huitema, Western Michigan University, Department of Psychology, Kalamazoo, MI 49008, *brad.huitema@wmich.edu*

***Key Words:*** intervention analysis, time series experiment

Two methods of interrupted time-series analysis known as ITSA and ITSACORR have become increasingly visible in published applied social science research during the past five years. Unfortunately, these methods contain fatal flaws that lead to greatly distorted intervention effect parameter estimates. The problems with the ITSE method (described in Huitema, 2004)are multiplied in the more recent ITSACORR approach. We demonstrate that neither the descriptive nor inferential properties of ITSACORR are satisfactory. Indeed, each component of the ITSACORR framework, including the structural model, the design matrix, the autocorrelation estimator, the ultimate parameter estimation scheme, and the inferential method contains fatal flaws. We recommend that the method not be used. Certain ARIMA and regression based methods provide far more satisfactory analysis options.

### JSL Scripts for Extending Available Statistical Tests in JMP Version 6

❖ Andy Mauromoustakos, University of Arkansas, 101 Agricultural Annex, AGRX 101, Fayetteville, AR 72701, *amauro@uark.edu*; Kevin Thompson, University of Arkansas

***Key Words:*** JSL, JMP

The new version of JMP (Version 6) provides data analysts with and easy and powerful sets of statistical and graphical platforms to handle most common data analysis situations. Fortunately for JMP users it includes a powerful scripting language to extend its capabilities. In this paper we will use JSL to program additional statistical tests that exists in the literature but are not usually included in the standard statistical methods texts and do not exists in most statistical software including JMP. We will limit this paper on comparisons of two statistics (involving means, medians, variances, counts, correlations, proportions) under a variety of conditions that include parametric assumptions, distribution free tests and when the two might be correlated. The JSL code for the test will be made available to the JMP community so that users will have more tests available to them.

### Interactive Web Site for Data Analysis

❖ Roger Lamb, Saginaw Valley State University, 7400 Bay Road, University Center, MI 48710, *marzjara@svsu.edu*; Morteza Marzjarani, Saginaw Valley State University; Josh Urbain, Saginaw Valley State University

***Key Words:*** website, data analysis, data mining

An interactive Web page is developed to gather information from the user for analysis. The analysis would involve the collection of data, modify the data (where needed) and then incorporate a statistical software such as SAS, SPSS, or Minitab into the Web page to perform the analysis of data. The ultimate goal is to develop an "intelligent" Web page which accepts data from the user, cleans the data, and prepares them for analysis. The software then prompts the user to select a statistical software (SAS, SPSS, or Minitab) and suggests one or more statistical model for analyzing the data.

### Novel Methods in the Visualization of Transitional Phenomena

❖ Bruce Swihart, University of Colorado at Denver and Health Sciences Center, 622 Panorama Drive, Grand Junction, CO 81503, *BruceSwihart@yahoo.com*; Brian Caffo, Johns Hopkins University; Matthew Strand, University of Colorado at Denver and Health Sciences Center; Naresh Punjabi, Johns Hopkins University

***Key Words:*** history matrix, history matrix visualization, longitudinal categorical data, Markov process, multi-state survival analysis, sleep apnea

When studying a group of individuals over time there exists an inclination to explore the data and report results with composite summary measures that collapse the time component. An example of this would be "Those with sleep apnea spent 18% of their total sleep time in REM, whereas those without sleep apnea spent 21%." Although these composite summary measures are convenient, they do obscure any time-related effects. History Matrix Visualization (HMV) is a technique that can allow the exploration and guide the analysis of longitudinal categorical data, and may provide insights into transitional processes that may have been otherwise obscured when implementing traditional composite summary measures. The HMV method is demonstrated using data of 59 matched apneic and non-apneic individuals from the Sleep Heart Health Study.

### Using Multivariate Statistical Techniques To Analyze Environmental Data

❖ Kyle Bradford, James Madison University, 7803 Dr. Hasan Hamdan, Burruss Hall, Harrisonburg, VA 22807, *bradfokb@jmu.edu*; Steaphanie Pearson, James Madison University

***Key Words:*** MANOVA, time series regression, factor analysis, canonical correlation

In this project, multiple regression, multivariate time series regression, MANOVA, and/or canonical correlation techniques will be used to predict some existing variables. Some of the variables are used as they are and others are manipulated to create new variables, these new variables may be dependent or independent. For example, the tentative plan is to consider the difference between surface to air temperature, elevation, latitude and longitude, as independent variables and the rest are dependent variables.

## 159 Statistics in Sports Speaker with Lunch (fee event)

Section on Statistics in Sports

**Monday, August 7, 12:30 pm–1:50 pm**

### Every Play, Every Day: a Success Story for Statistics in Sports

❖ Gilbert Fellingham, Brigham Young University, 212 TMCB, Provo, UT 84602, *gwf@byu.edu*

***Key Words:*** Bayesian hierarchical model, sports

In an effort to help the USA Men's National Volleyball Team focus practice sessions in preparation for the 2004 Olympic Games in Athens, Greece, we analyzed data from the 2002 World Championships and 2003 World Cup matches. These data include a grade for each skill performed every time a ball was touched (excluding setting). Using a hierarchical Bayesian model, we estimated an importance score for each skill-rating combination that accounted for variability and frequency of occurrence. Using the importance scores and information on the performance of the three top-ranked teams in the world at that time, a set of coaching guidelines was developed outlining changes to implement in the practice schedule. In this talk, I will discuss the important modeling aspects and data requirements that led the USA team to improve from 14th in the world to 4th in the Olympic games.

## 160 Section on Bayesian Statistical Science Roundtable with Lunch (fee event)

Section on Bayesian Statistical Science

**Monday, August 7, 12:30 pm–1:50 pm**

### Bayesian Bioinformatics

❖ Jeffrey S. Morris, M. D. Anderson Cancer Center, Department of Biostatistics and Applied Mathematics, 1515 Holcombe Blvd Box 447, Houston, TX 77030, *jeffmo@odin.mdacc.tmc.edu*

***Key Words:*** bioinformatics, Bayesian analysis, functional data analysis, microarrays, proteomics, metabonomics

Advances in science and technology have led to a variety of new devices for obtaining measurements from biological samples. Some of these produce complex, extremely high-dimensional data (e.g., microarrays, array cGH, 2d gels, MALDI-MS, LC-MS, NMR). These data present significant analytic challenges that require careful thought and care to address. Bioinformatics, the quantitative field dealing with issues with these high-throughput biological data, has exploded in recent years as researchers have worked to address these issues. The coherent and flexible modeling framework provided by the Bayesian paradigm has proven useful for attacking problems in this area and shows considerable promise as an area of future research. In this roundtable luncheon, we will discuss Bayesian bioinformatics and bioinformatics in general, what it has accomplished, and what its future holds.

## 161 Biopharmaceutical Section Roundtables with Lunch (fee event)

Biopharmaceutical Section

**Monday, August 7, 12:30 pm–1:50 pm**

### New Development and Challenges in Phase I and Phase I/II Dose-Finding Studies

❖ Weili He, Merck & Co., Inc., 1129 Blvd., Westfield, NJ 07090, *weili_he@merck.com*

The primary goal of a phase I cancer clinical trial is to determine the maximum tolerated dose (MTD) to be used for subsequent phase II and III trials evaluating efficacy. There are two schools of thought in designing phase I cancer clinical trials: one treats the MTD as being observed from the data and the other treats the MTD as a parameter to be estimated from a model. Diverse opinions have been expressed in strong support of one design over the other, and there has been no clear guidance on the choice of a design. The intention of this roundtable is for people who are interested in this area to discuss their ideas of design choices and share new development and challenges in designing a phase I and phase I/II dose-finding study.

### Analysis and Evaluation of Safety Information from Clinical Trial Data

❖ H. Amy Xia, Amgen Inc., 1 Amgen Center Drive, MS 242A, Thousand Oaks, CA 91320, *hxia@amgen.com*

***Key Words:*** drug safety, clinical trials, graphical tools, Bayesian hierarchical modeling, signal detection

There has been a growing awareness of importance relative to statistical evaluation of drug safety data recently. The use of statistics in assessing clinical trial safety data is important and often needs additional consideration. Some key problems in safety analyses will be highlighted. Some of the emerging statistical methodology developments in this area will be discussed, such as the application of the Bayesian hierarchical modeling of clinical trial safety data to address multiplicity issues, metaanalysis techniques for analyzing rare adverse events data, and some graphical advances in mining safety data. In addition, issues with evaluation of aggregate data for understanding the evolving safety profile, and especially for detecting potential safety signals during the course of clinical trials in drug development, will be discussed.

### Time-to-Event Analysis with Uncertain Endpoints

❖ Li Chen, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, *lic@amgen.com*

Time-to-event analysis of nonfatal endpoints or fatal events due to certain causes often involves classification of the endpoints as true or false according to decision rules. If these endpoints are not obvious to decipher whether true or false, they exhibit a level of uncertainty. Typically, only the endpoints confirmed as true events by an endpoint committee are used in the analysis. Information contained in the level of certainty is lost. Snapinn (1998) proposed a weighted Cox regression model that allows all potential endpoints to be included in the analysis, along with the level of certainty of each. This session will discuss the benefit of the method, recent development of the method, and its application in clinical trials.

## Analyses of Stratified Trials: Tips for Improving Power

❖ Devan V. Mehrotra, Merck Research Laboratories, UN-A102, 785 Jolly Road, Bldg C, Blue Bell, PA 19422, *devan_mehrotra@merck.com*

***Key Words:*** interaction, Mantel-Haenszel test, minimum risk weights, null variance, stratified Wilcoxon rank sum test, stratum-invariant weights

We will discuss simple ways to improve power when comparing two treatments in a stratified clinical trial with either a binary or a continuous endpoint using either a parametric or a rank-based test. The stratification could be due to one or more factors---such as study center, gender, and age group---and the objective of the trial could be to either show noninferiority or superiority of the new treatment versus a control treatment. Real examples and simulation results will be used to illustrate the notable power advantages of the relatively new approaches over 'standard' methods of analysis. Examples of new approaches include using minimum-risk weights (Mehrotra and Railkar, 2000), instead of Cochran-Mantel-Haenszel weights for binary data, and using stratum-invariant ranks, instead of stratum-specific ranks for the stratified Wilcoxon rank sum test (Li and Mehrotra, 2005).

## Using Computer Simulation To Aid in Dose Selection in Clinical Trials

❖ Kenneth Liu, Merck & Co., Inc., BL 3 2, PO Box 4, West Point, PA 19486, *Kenneth_Liu@merck.com*

***Key Words:*** phase 1-3 clinical trials

Phase I and II studies provide limited dosing information. Even phase II dose-ranging studies can provide unclear dose selection for phase III trials. Using computer simulations can aid in dose selection in phase III clinical trials.

## Assessing the Concordance of Two-Measurement Methods

❖ Jason Liao, Merck Research Laboratories, P.O. Box 4, WP37 C305, West Point, PA 19486, *jason_liao@merck.com*

***Key Words:*** reproducibility, concordance, agreement interval, bias

It is often necessary in medicine and other experimental sciences to assess the agreement between two measurement methods. It can happen in all phases of a drug's development. The two sets of measurements can be generated from two gene sequences in disease mapping, two analysts, two laboratories, two instruments, two treatments, two formulations (bioequivalence), two clinical endpoints, two models, two assays, two batches, two raters, test vs. retest, among others. The agreement problem covers a range of data, and the applications can come from different fields. In this roundtable luncheon, approaches will be discussed with real-life examples.

## Pharmacogenomics for Drug Development and Personalized Medicine

❖ Xuejun Peng, Takeda Global Research and Development Center, 475 Half Day Road, Lincolnshire, IL 60069, *xpeng@tgrd.com*

***Key Words:*** pharmacogenomics, clinical trials, personalized medicine, RX/DX dual development

The advance of pharmacogenomics makes it promising for genomics/proteomics tests to be used to support drug development and personalized medicine. For example, efficacy and safety in subgroups of patients may be improved by identifying the genomics profiles of individuals; clinical trials may be conducted for the dual purposes of developing drugs and diagnostic tools. However, many challenges still exist to fulfill such promises. In this luncheon, we will discuss technical and statistical issues with respect to design and analysis of clinical trials with pharmacogenomics.

## What To Do with Interaction Effects at Interim Analysis?

❖ Yuko Palesch, Medical University of South Carolina, Department of Biostatistics, Bioinformatics, and Epidemiology, 135 Cannon Street Suite 303, Charleston, SC 29425, *paleschy@musc.edu*

***Key Words:*** clinical trial, interim analysis, interaction effect

It has become rare to conduct a large clinical trial without prespecified interim analyses. Each interim analysis should be identical to the primary analysis at the end of the trial. A problem arises when a model-based primary analysis includes a treatment x covariate interaction term. Does one include it in the interim analysis? What if it is highly significant? The possible courses of action are to ignore the interaction effect and base the decision to stop or continue solely on the main treatment effect or to do subgroup analyses. For the latter, the test is unlikely to be powered properly for each subgroup, and if inconsistent results are obtained among the subgroups, the decision regarding the study conduct becomes complex. Roundtable luncheon participants will discuss the pros and cons of each course and share their experiences.

## Preparation of Interim Reports for Independent Data Monitoring Committee Review

❖ KyungMann Kim, University of Wisconsin-Madison, 600 Highland Ave., Bx 4675, Madison, WI 53792, *kmkim@biostat.wisc.edu*

***Key Words:*** clinical trials, interim analysis, IDMC

An independent data monitoring committee (IDMC) is charged to review the ongoing safety of trial participants and the scientific validity and merit of the trial and to make recommendations to the sponsor whether the trial should continue as planned, modify, or terminate because of safety, futility, or efficacy. Because a statistical analysis plan (SAP) serves as a benchmark for the final report, it needs to be explicit about the contents and format. An interim analysis plan, however, serves as a benchmark for the ongoing review of the safety and efficacy by the IDMC and cannot be codified as in SAP. Instead, it needs to be flexible to address any concerns emerging during the trial and to evolve with the trial. A dialogue will be held among statisticians representing sponsor, independent statistical center, and IDMC to improve the contents and format of the interim reports.

## Statistical Design and Analysis Issues Associated with the Establishment of the Safety and Effectiveness of Medical Devices

❖ Gary Kamer, U.S. Food and Drug Administration, 1350 Piccard Drive, Suite 100, CDRH/OSB/DBS, HFZ 550, Rockville, MD 20850, *gary.kamer@fda.hhs.gov*

***Key Words:*** FDA, clinical, ethical, compare

The process of establishing a new medical device as safe and effective enough to be approved by FDA for marketing purposes may require somewhat different design and analysis approaches than would a simple head-to-head comparison of two already-approved treatments. From sample size considerations to interim analysis to interpretation of results, clinical and ethical concerns influence the study design, implementation, and analysis. This roundtable discussion is intended to provide examples and compare competing approaches.

### Issues in Planning Two-Arm Clinical Trials of Active Drugs

❖ Sheela Talwalker, T'Walker Consulting, 11276 Caminito Rodar, San Diego, CA 92126, *sheela_talwalker@hotmail.com*

***Key Words:*** two - arm trial, superiority, equivalence/non-inferiority

In a trial of superiority over an active drug, do we need to re-establish assay sensitivity or effectiveness of the active comparator? Can we assume it? Can we use historical information collected in the recent past? If not, do we need to establish it by including placebo arm in the trial? Can we test equivalence or noninferiority when superiority cannot be established? How does one resolve the issues regarding adequate sample size, appropriate analysis population, and clinically meaningful difference delta for declaring equivalence or noninferiority? Gray areas in decisionmaking while testing equivalence or noninferiority lead to the issue of an ambiguity in the present practice of testing superiority.

### Analysis of Multiple Failure Outcomes

❖ Guowen Sun, sanofi-aventis, 3 Ash Court, Clinton, NJ 08809, *gordon.sun@sanofi-aventis.com*

***Key Words:*** failure-time data, failure outcomes, survival models, multiple events, clinical trial

Multiple failure outcomes or multiple failure-time data are encountered frequently in biomedical and other investigations. These data arise from repeated occurrences of the same type of event or multiple events of different types. In this luncheon, we will review/discuss appropriate procedures and survival models for analyzing different multiple failure outcomes.

### What Are the Statistical Issues in Subgroup Analysis - Design, Analysis, and Interpretation?

❖ Chul H. Ahn, U.S. Food and Drug Administration, 1350 Piccard Drive, CDHR, Rockville, MD 20850, *chul.ahn@fda.hhs.gov*

***Key Words:*** multiplicity, familywise error, subgroup p-value, interaction test, data dredging, regression to the mean

Subgroup analysis appears in almost every device or drug clinical trial. When the sponsor reports the results by demographic variables, they immediately run into subgroup analysis problems. In this discussion, we will consider two major problems in subgroup analysis---multiple testing and statistical power---among others. Then, subgroup analysis will be discussed in more detail in design, analysis, and interpretation.

### Conducting Multiple Event Analysis in Clinical Trials

❖ Xiang Zhang, Amgen Inc., One Amgen Center Drive, MS B24 2C, Thousand Oaks, CA 91320, *xiangz@amgen.com*

***Key Words:*** multiple-event analysis, recurrent events, multiple-type events, AG model, PWP model, WLW model

Many clinical trials contain complex failure-time data. For example: 1) Multiple-type events, 2) Recurrent events, 3) Mixture of multiple-type events and recurrent events, 4) Single failure times on different subjects may be correlated, arising, for example, from genetic effects. Several models have been proposed such as the Anderson-Gill model, the PWP model (Prentice, William, and Peterson), the WLW model (Wei, Lin, and Weissfeld), the LWA model (Lee, Wei, and Amato), and a frailty model. However, multiple-event analysis is not often primarily considered in the design of clinical trial. During this group discussion, I would like to focus on the following types of data: a) multiple-type events; b) recurrent events; c) mixture of recurrent and multiple-type events and share experience with respect to analysis methods, study design consideration, interactions with regulatory agencies, and labeling negotiations.

## 162 Business and Economics Statistics Section Roundtable with Lunch (fee event)

Business and Economics Statistics Section

**Monday, August 7, 12:30 pm–1:50 pm**

### MBA Statistics Courses Should Start with Regression

❖ J. Keith Ord, Georgetown University, McDonough School of Business, 204 Old North, Washington, DC 20057, *ordk@georgetown.edu*

***Key Words:*** introductory course, MBA program, regression

Introductory statistics courses for MBA students have tended to shrink in length over the years, yet the topical coverage often remains unchanged. Indeed, some courses barely touch upon regression analysis---even though the evidence suggests regression modeling is used far more widely in later MBA courses than any aspect of traditional inference. In this discussion, we will explore the notion of starting out with regression and then weaving in the other elements as needed, an approach inspired by Harry Roberts' book on data analysis for managers.

## 163 Section on Statisticians in Defense and National Security Roundtable with Lunch (fee event)

Section on Statisticians in Defense and National Security

**Monday, August 7, 12:30 pm–1:50 pm**

### Metrics for National Defense: What Metrics Would You Use To Measure Success or Make Decisions If You Were the Secretary of Defense or a Senator or Congressman?

❖ Nancy Spruill, Office of the Secretary of Defense, 300 Defense Pentagon, Washington, DC 20301-3000, *Nancy.Spruill@osd.mil*

***Key Words:*** defense, security, indicators of success

One of the thrusts of the latest Quadrennial Defense Review of the Department of Defense---released in February 2006---is to develop more meaningful, useful metrics for defense decisionmakers. If you were Secretary of Defense or a senator or congressman with oversight in the defense area, what metrics would you use to indicate the health of the Department of Defense? What metrics does the department currently provide? This roundtable will talk about dashboards, balanced scorecards, readiness, and business metrics that are used and others that might be proposed for future use. Attendees' ideas are welcome.

## 164 Section on Statistical Education Roundtables with Lunch (fee event)

Section on Statistical Education

**Monday, August 7, 12:30 pm–1:50 pm**

### Service Learning throughout the Statistics Curriculum

❖ Craig A. Johnson, Brigham Young University, Idaho, 232w Ricks Building, Mathematics Department, Rexburg, ID 83440, *johnsonc@byui.edu*

***Key Words:*** service learning, statistical education, pedagogy, curriculum

Service learning can enrich the traditional statistics curriculum greatly. Data-rich experiences provided by service learning motivate the course content and increase interest. In addition, exposure to real (and often messy) data provides valuable preparation for future applications. Service learning can be rewarding, and it does not need to take a lot of time and effort. We will discuss ways to incorporate service learning into students' educational experiences successfully, from the introductory statistics course through the doctoral level.

### Finding Internet Resources for Teaching Statistics Using CAUSEweb

❖ Ginger Rowell, Middle Tennessee State University, Department of Mathematical Sciences, PO box 34, Murfreesboro, TN 37132, *rowell@mtsu.edu*

***Key Words:*** statistics education, technology, digital library, CAUSEweb

Are you looking for one-stop shopping for resources to teach undergraduate statistics? This roundtable highlights CAUSEweb, a statistics education digital library being developed by the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE). Participants will leave with an overview of this digital library, searching tips to make finding statistics internet sites more efficient, and ways to integrate materials into lessons. Discussion will also include how to use CAUSEweb for finding professional development opportunities and research in statistics education.

### What Can We Do To Implement the GAISE Guidelines in Our Classrooms?

❖ Mary Parker, The University of Texas at Austin/Austin Community College, 1622 Waterston Ave., Austin, TX 78703, *parker@math.utexas.edu*

***Key Words:*** education, GAISE

The GAISE Guidelines recommend emphasizing conceptual understanding and improving statistical thinking by using active learning, real data, technology, and varied assessments to both improve and evaluate student learning. We'll talk about what we have done to implement these in our classrooms and what we'd like to do next. The GAISE Guidelines can be found at www.amstat.org/education/gaise.

## 165 Section on Statistics in Epidemiology Roundtable with Lunch (fee event)

Section on Statistics in Epidemiology

**Monday, August 7, 12:30 pm–1:50 pm**

### Sampling from Large Cohorts When Covariate Ascertainment Is Expensive

❖ William Barlow, Cancer Research and Biostatistics, 1730 Minor Ave., Suite 1900, Seattle, WA 98101, *williamb@crab.org*

***Key Words:*** sampling, cohort designs, case-control, case-cohort, screening

We will discuss sampling strategies when we have a large cohort available with known outcomes. We wish to sample efficiently to reduce costs of covariate ascertainment. The covariates could be biomarkers or risk factors obtainable only by medical record review. Possible choices for sampling include nested case-control, case-cohort, and countermatching designs. We will concentrate on analyses that use failure time outcomes. The possible choices will be illustrated with reference to evaluating the efficacy of mammography screening on breast cancer mortality when chart review of screening history is required. Participants are welcome to contribute other possible applications.

## 166 Section on Government Statistics Roundtable with Lunch (fee event)

Section on Government Statistics

**Monday, August 7, 12:30 pm–1:50 pm**

### Katrina: Unanticipated Data Needs

❖ Christa Jones, U.S. Census Bureau, 3231 Valley Drive, Alexandria, VA 22302, *christa.d.jones@census.gov*

***Key Words:*** Katrina, population estimates, disaster

Each year, natural disasters occur, impacting the lives of people in the United States. Hurricane Katrina, however, caused unprecedented devastation, presenting statistical agencies with new and unanticipated data needs. The massive migration and destruction of housing in the wake of Hurricane Katrina also disrupted ongoing monthly surveys and impacted the assumptions underlying population estimates. These programs historically have provided federal agencies with the information to evaluate and adapt programs to meet public needs. This roundtable discussion will focus on impacts the hurricane and other disasters can have on current data and possible approaches the statistical system can take to adjust for them.

## 167 Section on Health Policy Statistics Roundtable with Lunch (fee event)

Section on Health Policy Statistics

**Monday, August 7, 12:30 pm–1:50 pm**

### Measuring Health Disparities

❖ James Scanlan, James P. Scanlan, Attorney at Law, 1529 Wisconsin Ave., NW, Suite 300, Washington, DC 20007, *jps@jpscanlan.com*

***Key Words:*** health disparities, health inequalities, mortality differentials, measuring, relative differences, absolute differences

When two groups differ in their susceptibility to an outcome, a decline in the prevalence of an outcome tends to increase relative differences in experiencing the outcome and reduce relative differences in avoiding the outcome. Such declines also cause various changes to other indexes used to measure the size of differences in experiencing an outcome. These tendencies call into question almost all research to date on changes in the size of health disparities and raise questions as to whether such changes can, in fact, be evaluated.

## 168 Section on Physical and Engineering Sciences Roundtable with Lunch (fee event)

Section on Physical and Engineering Sciences

**Monday, August 7, 12:30 pm–1:50 pm**

### Using All Them Machines: Grid Computing for Statistical Applications

❖ Randall Tobias, SAS Institute, Inc., SAS Campus Drive, Cary, NC 27513, *randy.tobias@sas.com*; Peter Westfall, Texas Tech University

***Key Words:*** grid computing, statistical computing, simulation

Compute-intensive statistical tasks often can be factored into repeating many small, independent tasks. Examples include repetitive analyses on partitions of large databases, bootstrapping, simulation, and Bayesian sampling. Such tasks are potentially amenable to grid computing, where the small tasks are farmed to multiple “server nodes.” Issues to be discussed: What sorts of realistic compute-intensive problems do people run across? Can grid computing be expected to help? What hardware and software are required? How do you go about turning a one-machine task into a multiple-machine task? The leaders have experience with statistical grid computing, mainly using SAS software. They will be prepared to introduce the methodology and talk about their successes (and otherwise!). Participants should come with real statistical problems in mind that may benefit from grids.

## 169 Section on Quality and Productivity Roundtables with Lunch (fee event)

Section on Quality and Productivity

**Monday, August 7, 12:30 pm–1:50 pm**

### Use of Genetic Algorithms in Experimental Design

❖ John Borkowski, Montana State University, Department of Mathematical Sciences, Bozeman, MT 59715, *jobo@math.montana.edu*

***Key Words:*** genetic algorithm, experiments, design, response surface, DOE

A genetic algorithm (GA) is an evolutionary search strategy used to find an acceptable solution to a problem based on the optimization of a fitness function. A successful GA will ‘evolve’ more fit solutions until an acceptable solution has been generated. This roundtable discussion will involve a brief review of existing applications of GAs to the generation of experimental designs (e.g., optimal designs). This will be followed by a discussion of the potential of GAs to address other existing experimental design problems (e.g., constrained design regions, mixture designs, restrictions on randomization). Participants are encouraged to share their experiences with GAs, as well as any design generation problems they are encountering for which GAs may be applied.

### Bayesian Methods in Reliability

❖ Alyson Wilson, Los Alamos National Laboratory, P.O. Box 1663, MS F600, Los Alamos, NM 87544, *agw@lanl.gov*

***Key Words:*** reliability, Bayesian

Bayesian methods are used across reliability. Although once thought of only as methods to introduce “expert judgment” into an analysis, Bayesian methods also provide important tools for systems analysis, experimental design, and information integration.

## 170 Section on Risk Analysis Roundtable with Lunch (fee event)

Section on Risk Analysis

**Monday, August 7, 12:30 pm–1:50 pm**

### Risks of Computerized Voting Systems

❖ Michael Orkin, Exponent, Inc., 149 Commonwealth Drive, Menlo Park, CA 94025, *morkin@exponent.com*

***Key Words:*** voting machines, touch screen, computerized voting, recount, vote, DRE

Touch screen and other computerized voting systems have advantages over traditional, mechanical voting machines. They also introduce new risks of corrupting an election and subverting the democratic process. We will discuss these risks (e.g., computer glitches, malicious code, network security breaches) along with safeguards to help alleviate them. We also will talk about recounts, audits, and voting machine reliability standards and discuss efforts to implement these safeguards.

## 171 Section on Survey Research Methods Roundtables with Lunch (fee event)

Section on Survey Research Methods

**Monday, August 7, 12:30 pm–1:50 pm**

### The Importance of Nonresponse for Survey Design

❖ Roger Tourangeau, University of Maryland, 1218 LeFrak Hall, College Park, MD 20742, *rtourangeau@survey.umd.edu*

***Key Words:*** survey research, response rates, nonresponse bias

This roundtable will discuss the recent decline in survey response rates and its impact on survey costs and design. Has the decline in response rates biased survey estimates? When should we be concerned about the decline in response rates? What are the options for limiting the problems due to falling response rates? Is "responsive design" the answer to the problem?

### Optimization of Survey Procedures in the Presence of Limited Cost Information

❖ John L. Eltinge, Bureau of Labor Statistics, Office of Survey Methods Research, PSB 1950, 2 Massachusetts Avenue NE, Washington, DC 20212, *Eltinge.John@bls.gov*

*Key Words:* frame development and refinement, global and local optimization, interviewer assignments, mode effects, sample allocation, total survey error

Practical survey design work often involves trade-offs between prospective costs and prospective effects on data quality. With some notable exceptions, methodological research has tended to place primary emphasis on evaluation of data quality, and has devoted less attention to evaluation of survey cost structures. This roundtable reviews the information generally available on survey cost structures, outlines some of the reasons for limitations on availability of information on some cost components, and explores methods for optimization of survey procedures in the presence of limited cost information.

## 172 Social Statistics Section Roundtable with Lunch (fee event)

Social Statistics Section

**Monday, August 7, 12:30 pm–1:50 pm**

### The Future of U.S. Income Statistics

❖ Connie Citro, Committee on National Statistics, Keck Center, Room 1139, 500 5th St NW, Washington, DC 20001, *ccitro@nas.edu*

*Key Words:* income, poverty, surveys, administrative records, wealth, indirect estimators

Public concern is growing about income adequacy, inequality, and, generally, how Americans are faring economically compared with the past and other countries. Informed debate on these issues requires good statistics on income and wealth. The luncheon organizer, who directed studies of poverty measurement at the Committee on National Statistics, will share her views and seek to stimulate discussion on needed improvements to the U.S. income statistics system. Issues will include the role of surveys---including SIPP, CPS, and ACS---the possible use of administrative records to correct survey responses for reporting errors, and the role of indirect estimation techniques.

## 173 Late-Breaking Session #1: Statistical/Mathematical Challenges in Biodefense Immune Modeling

The ASA, ENAR, IMS, SSC, WNAR

**Monday, August 7, 2:00 pm–3:50 pm**

### A Kernel Method for Subpopulation Discovery and Analysis in Polychromatic Flow Cytometry

David Foster, Duke University; Cliburn Chan, Duke University;
❖ Thomas B. Kepler, Duke University, 101A North Building, Research Drive, Box 90090, Durham, NC 27708-0090, *keple003@mc.duke.edu*

*Key Words:* cell biology, clustering, smoothing, density estimation

Technological developments in flow cytometry allow increasing numbers of cell-surface markers to be assessed simultaneously and promise to enhance phenotypic analysis. The analysis of such high-dimensional data presents significant obstacles; improved statistical methods will be required to realize its potential. To this end, we developed a novel method based on kernel density estimation to identify and analyze subpopulations in polychromatic flow-cytometric data. We implemented these methods, along with visualization tools, in software to automate subpopulation discovery and comparison between treatment datasets. Our system automatically optimizes the kernel parameters, identifies potential subpopulations, validates them using independent statistical methods, and plots their modes on optimized projections onto two-dimensional contour plots. It also estimates the proportion of total events falling into the subpopulations and compares the distribution of events between pairs of datasets. We will describe the method and illustrate it with 14-color data.

### Multidimensional Scaling Analysis To Study Temporal Transcriptome Fingerprint Clustering in Human Dendritic Cells Infected with Wild-Type and Chimeric Viruses

❖ Yongchao Ge, Mount Sinai School of Medicine, Department of Neurology and Microbiology and Center for Translational Systems Biology, One Gustave L. Levy Place, Box 1023, New York, NY 10029, *Yongchao.Ge@mssm.edu*; Ana Fernandez-Sesma, Mount Sinai School of Medicine; Thomas M. Moran, Mount Sinai School of Medicine; Stuart C. Sealfon, Mount Sinai School of Medicine

*Key Words:* multidimensional scaling, clustering, QPCR, dentritic cells, viruses

Our lab has measured temporal transcriptome fingerprints across many genes from different buffy coats of human dendritic cells (DC) from different donors infected with wild-type and chimeric viruses. Due to changes in interest in specific targets over time, many genes have missing values in qPCR measurements, which can be problematic for statistical analysis. To address the missing value problem, we construct a distance that only uses the common genes between the two buffy coats and that automatically normalizes qPCR measurements. We then apply multidimensional scaling analysis and find that the buffy coats are clearly clustered into three groups according to the type of virus infected. Thus, dimensionality reduction of complex transcriptional changes shows high correlation for the same DC exposures across different donors.

### Data Analysis for Multiplex Assays

❖ Shlomo Ta'asan, Carnegie Mellon University, Department of Mathematical Sciences, 5000 Forbes Avenue, Pittsburgh, PA 15213, *shlomo@andrew.cmu.edu*

*Key Words:* multiplex assays, influenza a, mathematical modeling, logistic regression

Multiplex assays are becoming standard in immunology research due to their reduced cost and their ability to produce abundant information regarding multiple cytokines in a single run. Common statistical approach for interpreting the data use standard curves based on logistic fitting. In this talk we present an alternative approach which is based on a mathematical model for the binding process of antibodies to the beads. It uses just a few physical parameters, including binding affinity and fluorescence to concentration conversion parameter. We show the model's ability to reproduce experimental results. We then describe how to use the model in conjunction with an inverse problem formulation in order to define standard curves. The advantage of the method over purely statistical models is that it identifies physical parameters which are invariant over a wide range of conditions. Results regarding Influenza A infection will be presented.

### Identifiability and Statistical Inverse Problems for Biomedical Dynamic Systems

❖ Hulin Wu, University of Rochester, School of Medicine and Dentistry, 601 Elmwood Avenue, Box 630, Rochester, NY 14642, *hwu@bst.rochester.edu*

***Key Words:*** dynamic models, HIV/AIDS, identifiability, differential equations

A vast literature on mathematical modeling of biomedical processes has appeared in the past. In many cases, a set of differential equations are proposed to describe the dynamics of biological agents or species. Although identifiability and inverse problems of dynamic models have been discussed, they mostly focus on theoretical perspective and mathematical identifiability. The Literature that actually investigates the identifiability and inverse problems with practical application perspective is sparse. In this talk, I will discuss these issues from practical application perspective. In particular, statistical identifiability and inverse problems for dynamic models will be addressed in detail. Statistical estimation methods and implementations will be illustrated carefully via simulation examples and HIV dynamic models.

## 174 Bayesian Finance

Business and Economics Statistics Section, Section on Bayesian Statistical Science

**Monday, August 7, 2:00 pm–3:50 pm**

### Random Field and Affine Models for Interest Rates: an Empirical Comparison

❖ Alan Bester, The University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, *cbester@ChicagoGSB.edu*

***Key Words:*** finance, Bayes, Markov chain Monte Carlo

Traditional affine models of the term structure are eminently tractable, but suffer from empirical difficulties. Random field models offer great flexibility in fitting the data, but are widely considered nonimplementable unless they are approximated by a low-dimensional system. I develop a state-space estimation framework where both random field and affine models can be estimated by MCMC using the same panel of forward rate data. I find random field models are much better able to fit the patterns of volatility and correlation in a long historical sample of U.S. Treasury forward rates.

### Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices

Michael Johannes, Columbia University; ❖ Nicholas Polson, The University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, *ngp@gsb.uchicago.edu*; Jonathan Stroud, University of Pennsylvania

***Key Words:*** finance, Bayes, Monte Carlo

This paper provides a methodology for optimally filtering latent state variables in discretely observed jump-diffusion models. When prices are observed continuously, state variables such as volatility, jump times, and jump sizes also are observed. But with discrete observations, these variables are unobserved. We combine discretization schemes with Monte Carlo methods to compute the optimal filtering distribution: the distribution of the latent states conditional on the observed prices. Our approach is general, applying in models with nonlinear characteristics and even nonanalytic observation equations. We use simulations to investigate how sampling frequency affects jump and volatility estimates and then extract information about volatility jointly from equity index options and index returns.

### Macroeconomic Filtering from the Yield Curve

❖ Satadru Hore, The University of Chicago, 5807 S. Woodlawn Ave., Grad School of Business, Chicago, IL 60637, *pshore@gsbphd.uchicago.edu*

***Key Words:*** finance, Bayes, Markov chain Monte Carlo

An integrated utility-based model is presented for bond prices, where CPI and real consumption follow a jointly lognormal process with latent growth rates. The volatilities of the growth rates are also stochastic, giving rise to time-varying risk premia. The model is fitted to the time-series of economic fundamentals and the yield curve jointly under different prior specifications. The prior that fits the yield curve strongly does a poor job in predicting yearly bond prices over the one that tries to find a balance between different series. The volatility in expected inflation has a large explanatory power in explaining bond risk premia, while volatility in expected consumption growth does not. A high risk-aversion parameter is computed for the tight prior specification, but it is small for other specifications, confirming existing puzzles in asset pricing.

## 175 Section on Statistics in Sports Invited Session

Section on Statistics in Sports, Section on Statistical Education

**Monday, August 7, 2:00 pm–3:50 pm**

### An Objective Scoring Method for Graded Count Variables with Applications to Olympics and Baseball

❖ John Daniels, Central Michigan University, Pearce Hall 206, Mount Pleasant, MI 48859, *john.daniels@cmich.edu*

***Key Words:*** u-statistics, multivariate, rank test, hierarchical data

In many sports competitions, participants are evaluated based on the counts of their successes; often, a single-count variable does not suffice. Thus, the assumptions of traditional approaches based on linear combinations of counts by category often are not justified on theoretical grounds. Establishing concept validity through empirical validation, as often is suggested, is problematic, in part because the choice of the criterion to validate against can shift the focus from assessing an athlete's ability to assessing context-dependent performance. This presentation proposes the use of u-statistics for scoring objects (e.g., players, teams, countries) based on several count variables. In particular, these variables are allowed to be graded by importance or relevance.

### Evaluation and Analysis of the Impact of Recruiting on College Football

❖ Justin W. Davis, University of Missouri-Columbia, 1004 Manhattan Drive, Columbia, MO 65201, *davisjwa@health.missouri.edu*; David Annis, Naval Postgraduate School

***Key Words:*** college football, recruiting, sports, rankings

Recently, there has been a dramatic increase in the scope and availability of information about the recruitment of collegiate athletes due to the advent of subscription recruiting web sites. These services publish nationwide and state-level rankings of recruits, in addition to team and position rankings. Increased recruiting coverage has allowed the college football fan to follow the recruiting process in more detail than ever before. With this comes increased scrutiny of team performance in this regard. In this talk, we will present our findings regarding the impact of recruiting on team win-loss record and the ability of recruiting services to identify accurately top-flight college football talent. We also evaluate the recruiting effectiveness of coaching staffs. Finally, we present a retrospective model for evaluating career contribution made by an athlete post-recruitment.

### The Combination of Subjective Judgment with Statistical Projections in the Evaluation of a Baseball Player

❖ Sig Mejdal, St. Louis Cardinals, 346 Dunsmuir Terrace, #3, Sunnyvale, CA 94085, *smejdal@stlcardinals.com*

***Key Words:*** baseball, player evaluation, applied statistics

A few baseball teams, including the St. Louis Cardinals, have brought in quantitative analysts (i.e., scouts) to compliment their existing evaluation system. I will present processes I use to evaluate players' hitting, pitching, fielding, and base-running skills, along with a description of the limitations and difficulties associated with combining these objective measures with the long-standing subjective evaluations.

## 176 Forensic Statistics ✪

Section on Statisticians in Defense and National Security
**Monday, August 7, 2:00 pm–3:50 pm**

### Another Look at the Kennedy Assassination

❖ Clifford Spiegelman, Texas A&M University, Room 405C, Blocker-3143, College Station, TX 77845-3143, *cliff@stat.tamu.edu*

***Key Words:*** bullet-lead, forensics

One of the first uses of the compositional analysis of bullet lead was in testimony to the House Select Committee on the Assassination of President Kennedy (HSCA). A distinguished scientist, the late Dr. Guinn, testified to the HSCA that compositional analysis of assassination bullet fragments strongly supported only two bullets striking President Kennedy and his motorcade. We will take a look at this claim and include new experimental findings.

### Forensic Statistics: Intelligence, Evidence, and Law

❖ David Kaye, Arizona State University, College of Law, Tempe, AZ 85287-7906, *k@asu.edu*

***Key Words:*** scientific evidence, law, forensic science, identification, pattern matching

This presentation discusses techniques for criminal investigation or intelligence gathering that rest on statistical inferences or raise statistical issues. Examples include DNA databases and "familial searching," DNA samples and "racial profiling," speaker identification from voice spectrographs, fingerprint identification (automated and otherwise), and the polygraph as a lie detector. It also describes the legal standards and case law on using scientific and statistical evidence obtained with these techniques as proof in criminal cases in the United States.

### The Probative Value of Trace Evidence: What Sources of Error Are Really Important?

❖ Alicia Carriquiry, Iowa State University, 102D Snedecor Hall, Department of Statistics, Ames, IA 50011-1210, *alicia@iastate.edu*

***Key Words:*** coincidental match, trace evidence, probative value, reference databases, sampling

Forensic scientists collect and analyze trace evidence from crime scenes. Trace evidence may include fibers, cartridge casings, or bullet fragments, and the data are typically multivariate measurements such as the composition of the lead alloy in bullet fragments. Suppose traces also are found on a suspect. The question of interest is whether the crime scene and suspect evidence have a common origin. This often is answered in two steps. First, a match between the two sources of evidence is established. Then, the probative value of the match is assessed. A recent NRC report discussed the errors that may arise when establishing matches, but a more difficult problem is assessing the probability that the match occurred by chance alone. We focus on the second question and discuss the information and approaches that might be needed to decide on the probative value of evidence.

## 177 Variance Estimation in the Presence of Nonresponse and Outliers ●

Section on Survey Research Methods
**Monday, August 7, 2:00 pm–3:50 pm**

### Estimation of the Total Variance of Survey Statistics under Unweighted Imputation

❖ Santanu Pramanik, University of Maryland, 1218 Lefrak Hall, Joint Program in Survey Methodology, College Park, MD 20742, *spramanik@survey.umd.edu*; Partha Lahiri, University of Maryland

***Key Words:*** missing value, imputation, response propensity

In many complex surveys, unweighted imputation methods are employed because of the unavailability of survey weights at the time of imputing missing survey data. In such situations, it is well known that certain customary design-based estimators with imputed data generally are biased under the usual uniform response mechanism within each imputation cell and the sampling design. In this paper, we present the expression of the bias of a customary design-based estimator under ignorable response mechanism and then use this expression to propose a bias-corrected estimator. The second part of the paper deals with a variance estimator that captures different sources of uncertainties. Both theory and results from a Monte Carlo simulation study are presented to justify our approach.

### Adjusted Jackknife for Imputation under Unequal Probability Sampling without Replacement

❖ Yves G. Berger, The University of Reading, Applied Statistics, Whiteknights Road, Reading, RG6 6FN UK, *yvesberger@yahoo.co.uk*; Jon N. K. Rao, Carleton University

***Key Words:*** deterministic imputation, hotdeck, inclusion probabilities, linearization, pseudo-values, random imputation

Imputation is used commonly to compensate for item nonresponse in sample surveys. If we treat the imputed values as if they are true values, and then compute the variance estimates using standard methods---such as the jackknife---we can underestimate seriously the true variances. We propose a modified jackknife variance estimator, which is defined for any without replacement of unequal probability sampling design in the presence of imputation and non-negligible sampling fraction. Mean, ratio, and random imputation methods will be considered. The practical advantage of the proposed method is its breadth of applicability.

### Variance Estimation for Complex Surveys in the Presence of Outliers

❖ Ralf T. Münnich, University of Trier, Economics and Social Statistics Department, Universitätsring, Faculty IV, Trier, 54286 Germany, *muennich@uni-trier.de*; Beat Hulliger, Swiss Federal Statistical Office

***Key Words:*** variance estimation, outlier, imputation, robust estimation

Statistical production often faces the difficulty of finding adequate methods for detecting and treating outliers. This generally is solved by either applying robust estimation methods or identifying outliers and replacing them or adjusting their weight. While, for robust estimators, approximate variance estimations exist for other methods, often one has to resort to resampling methods. Based on the EU projects EUREDIT and DACSEIS, the accuracy and sensitivity of selected variance estimation strategies with respect to outliers will be investigated in a Monte Carlo simulation. The study will be conducted on data from the Swiss Household and Budget Survey and the Swiss Environment Protection Expenditure Survey.

## 178 Statistical Issues in Disaster Response ● ✪

Section on Risk Analysis, Section on Statisticians in Defense and National Security, Section on Statistical Graphics

**Monday, August 7, 2:00 pm–3:50 pm**

### Statistical GeoInformatics of Hotspot Detection and Prioritization for Early Warning and Disaster Management

❖ Ganapati P. Patil, The Pennsylvania State University, Department of Statistics, 421 Thomas Bldg., University Park, PA 16802, *gpp@stat.psu.edu*; Luiz Duczmal, Universidade Federal de Minas Gerais; Reza Modarres, The George Washington University; Stephen L. Rathbun, University of Georgia

***Key Words:*** poset prioritization system, upper level set scan statistic system, crimes, wild fires, disease outbreaks, persistent poverty trajectories

Geoinformatic surveillance for spatial and spatiotemporal hotspot detection and prioritization, early warning, and sustainable management is a critical need for the 21st century. Our detection innovation employs the notion of an upper-level set, and accordingly is called the upper-level set scan statistic system (Patil and Taillie 2004, Environmental and Ecological Statistics, 11, 183-197). The prioritization innovation employs the notion of a partially ordered set, and accordingly is called the poset prioritization system (Patil and Taillie, 2004, Environmental and Ecological Statistics, 11, 198-228). In this presentation, we will discuss these innovations and the investigations in progress, relating to a variety of case studies involving natural disaster management and early warning (e.g., crimes, wildfires, disease outbreaks, persistent poverty trajectories, etc.).

### Changing Perspectives in the Analysis of Natural Disaster Data

❖ Maria J. Sirois, Tulane University, Department of Biostatistics, SL-18, 1440 Canal Street, Ste. 2001, New Orleans, LA 70112, *mjsstat@yahoo.com*; David Banks, Duke University

***Key Words:*** statistics, disaster, analysis, vulnerability, hazard, data

Some emergency and disaster planners now consider classical methods of analyzing disaster data to be inadequate. New approaches include using vulnerability data and analyzing data from all phases of a disaster. The inclusion of these data may be difficult due to the lack of common definitions and complete, reliable data sources. Failure to include these data may lead to results that fail to distinguish characteristic differences in disaster planning stages and certain entities in the same disaster area. A compromise must be made between efficiency and the collection and preservation of data related to a disaster. An interdisciplinary approach to disaster analysis must be flexible enough to incorporate multiple hazards and include variables related to the hazard, human behavior, and sustainable development.

### Preparing for a Disaster

❖ Vicki M. Bier, University of Wisconsin-Madison, Department of Industrial Engineering, 1550 Engineering Drive, Madison, WI 53706, *bier@engr.wisc.edu*; Lee Clarke, Rutgers University

***Key Words:*** risk, preparedness, emergency, security, cost-effectiveness

Investments in emergency preparedness and response have yet to live up to their promise as a cost-effective way to enhance public safety and security. We will discuss some of the reasons for this problem, lay out strategies for evaluation of competing policies, and point out directions for improvement.

## 179 Statistical Methods in Climate Modeling and Seismology ● ✪

WNAR, Section on Physical and Engineering Sciences, Section on Bayesian Statistical Science, Section on Statistics and the Environment

**Monday, August 7, 2:00 pm–3:50 pm**

### Probabilistic Projections of Climate Change: Bayesian Models for Analyzing Ensembles of Global Climate Models

❖ Claudia Tebaldi, National Center for Atmospheric Research, Boulder, CO , *tebaldi@ucar.edu*; Richard L. Smith, The University of North Carolina at Chapel Hill; Douglas W. Nychka, National Center for Atmospheric Research; Linda O. Mearns, National Center for Atmospheric Research

***Key Words:*** climate models, climate change, probabilistic projections, regional scale, Bayesian approach, cross-validation

General Circulation Models, or Global Climate Models (GCMs), perform complex numerical simulations of the whole climate system under alternative future scenarios of anthropogenic emissions. Their output is critical in determining projections of future climate change. However, different GCMs may differ significantly in their projections, raising the question of how best to combine ensembles of GCMs into a probability distribution of future climate change. Using a Bayesian model, we derive posterior distributions of temperature change at regional scales by treating regions jointly. A cross-validation approach is proposed to test the reasonableness of our Bayesian predictive distributions. Our analysis allows for quantification of the uncertainty in climate change projections as a Bayesian posterior PDF, a substantial extension of previous approaches to uncertainty in climate models.

### Uncertainty Estimation in Geophysics

❖ Mrinal K. Sen, The University of Texas at Austin, Institute for Geophysics, 4412 Spicewood Springs Road Bldg 600, Austin, TX 78759, *mrinal@ig.utexas.edu*

***Key Words:*** geophysics, uncertainty

Geophysical inverse problems involve estimating subsurface earth model parameters from surface measurements of noisy and inadequate geophysical data. Nonlinearity adds another degree of complexity. We employ a Bayesian formulation to estimate a posterior probability density function in model space. For practical applications, we seek a trade-off between accuracy and computational speed. A Monte Carlo importance sampling is theoretically more accurate, but the convergence rate is very slow. Alternately, we use an approximate greedy sampling algorithm based on multiple very fast simulated annealing algorithm to characterize the marginal PPD. Our numerical experiments with the inversion of resistivity and seismic waveform data reveal that the MVFSA is a useful and practical tool that provides a general framework to combine disparate data types, such as well logs and seismic and core data.

### Estimating Parametric Uncertainties of the Community Atmospheric Model (CAM3) and Processes Controlling Global Climate Change

❖ Charles S. Jackson, The University of Texas at Austin, Institute for Geophysics, 4412 Spicewood Springs Rd., Austin, TX 78759, *charles@ig.utexas.edu*

***Key Words:*** Bayesian inference, uncertainty, climate, stochastic inversion

Computational models of global climate differ widely in their sensitivity to projected increases in greenhouse gases. We do not yet have a clear picture of what processes dominate this uncertainty or how we may make better use of observations to reduce parametric uncertainties in climate models. The climate problem is one example of a broad class of problems in the geosciences currently limited by the lack of sufficient computational resources and/or the dimensionality and size of observational constraints. This talk will emphasize the characterization of the climate problem and present results of a calculation using Bayesian inference and techniques of geophysical stochastic inversion to constrain values of six nonlinearly related parameters important to convection, precipitation, and radiation in the Community Atmospheric Model (CAM3).

## 180 Sensitivity Analysis for Missing Data and Causal Inference: Principles and Practice

Biometrics Section

**Monday, August 7, 2:00 pm–3:50 pm**

### A Sensitivity Analysis Paradigm for Randomized Trials with Potentially Informative Censored Data

❖ Daniel Scharfstein, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 201205, *dscharf@jhsph.edu*

In this talk, we describe a sensitivity analysis paradigm for reporting results of randomized trials with time-to-event outcomes, which may be censored informatively. We illustrate the methodology using data from a randomized trial of a treatment designed to reduce epileptic seizures. We also show how to use auxiliary information from previous trials to reduce the range of the sensitivity analysis.

### Mixture Models and Informative Priors for Analyzing Incomplete Longitudinal Data

❖ Joseph W. Hogan, Brown University, Center for Statistical Sciences, Box G-H2, Department of Community Health, Providence, RI 02912, *jhogan@stat.brown.edu*

***Key Words:*** repeated binary data, semiparametric models, varying coefficient models, nonidentifiability, nonignorable nonresponse, HIV/AIDS

Mixture models provide an attractive framework for analyzing incomplete data because the parameter space frequently can be divided into one set of parameters indexing Y(obs) and another indexing Y(mis), given Y(obs). A subset of the parameters indexing [Y(mis)|Y(obs)] are completely nonidentifiable. We give examples of principled use of informative priors for capturing assumptions about the missing data mechanisms, including uncertainty about the assumptions. Examples use data from a longitudinal HIV cohort study and a longitudinal clinical trial of smoking cessation.

### Sensitivity Analysis for Instrumental Variables Regression with Overidentifying Restrictions

❖ Dylan S. Small, University of Pennsylvania, 400 Huntsman Hall, 3730 Walnut Street, Wharton School, Department of Statistics, Philadelphia, PA 19104, *dsmall@wharton.upenn.edu*

***Key Words:*** instrumental variables, causal inference, structural equations models

Instrumental variables (IV) regression is a method for making causal inferences about the effect of a treatment when there are unmeasured confounders. The method requires one or more valid IV; a valid IV is a variable that is independent of unmeasured confounders and has no direct effect on the outcome. Often there is uncertainty about the validity of proposed IVs. When more than one IV is proposed, the validity of the IVs can be tested via the ``overidentifying restrictions test.’ However, the test has no power versus certain alternatives. To fully address uncertainty about the validity of the proposed IVs, we develop a method of sensitivity analysis that makes full use of the information provided by the overidentifying restrictions test, but provides more information by exploring sensitivity to invalidity of the proposed IVs in directions for which the test has low power.

## 181 Management of Statistical Decision-making in a Large Organization

Section on Physical and Engineering Sciences, Section on Statistical Education, Section on Statistical Consulting
**Monday, August 7, 2:00 pm–3:50 pm**

### Maximizing the Effectiveness of Statistical Resources in Industry

❖ Henry T. Davis, Becton, Dickinson, and Company, 1 Becton Drive, Franklin Lakes, NJ 07417-1880, *henry_davis@bd.com*

***Key Words:*** statistics, para-statistician, Six Sigma, statistical decision making

Many problems faced by modern companies require appropriate statistical input to ensure the proper decision is made. However, the number of statisticians in most companies is too small and/or the statisticians are not properly placed organizationally to allow their direct participation in all decisionmaking processes that would benefit from statistical input. This talk will focus on methods the organization and/or statisticians can use to better leverage limited statistical resources and thus increase the decisionmaking processes that will benefit from statistical input. Methods discussed will include use of in-house training programs to teach employees basic statistical methods, moving from a “do” to a “review” model for project teams that have resources trained in basic statistical methods, and the creation of parastatisticans.

### The Future of Corporate Statistics Organizations

❖ Roger W. Hoerl, GE Global Research, 1 Research Circle, Schenectady, NY 12309, *hoerl@crd.ge.com*

***Key Words:*** consulting, leadership, collaborative research

The environment within which statisticians in the private sector work today is radically different than the environment of only 10--20 years ago. This situation has been brought about by numerous societal and organizational changes, such as the rise of the internet, globalization, changes in secondary education, and corporate improvement initiatives---such as Six Sigma. In short, we have gone from an environment in which a consulting paradigm for statistics organizations made sense and worked to one in which such a paradigm is inadequate and not viable long-term. This session will review briefly the changes in our environment and how our paradigm needs to change in order for corporate statistics groups to thrive in the current environment. This paradigm also requires new skills and behaviors in order for statisticians to be successful.

### Teaching Engineers To Think Statistically

❖ Scott A. Pardo, Purdue Frederick Labs, 3 Garret Mountain Plaza, Suite 303, West Paterson, NJ 07424, *scott.pardo@pharma.com*

***Key Words:*** physics, engineering, emic anthropology

Engineers schooled in physics are not taught that 90% of the work they do when they graduate involves data analysis. Those who are introduced to inference in school have little idea how it applies to problems of system design. Most statisticians have limited training in engineering sciences and may have difficulty in communicating with engineers who often are taught to think deterministically. Due to lack of knowledge in physics, statisticians often are frustrated when explaining inference to engineers. In order to teach engineers how to think statistically, statisticians can borrow ideas from “emic” anthropologists, who study a culture through the eyes of its members rather than by external observation. By understanding physics, statisticians often can find sources of uncertainty in problems and tie notions of inferential thinking to the engineer’s understanding.

### Collaboration with Nonstatisticians

❖ Joanne Wendelberger, Los Alamos National Laboratory, Statistical Sciences Group, Mailstop F600, Los Alamos, NM 87545, *joanne@lanl.gov*

***Key Words:*** collaboration, nonstatistician, research, methodology, application

Methodological advances in statistics often are motivated by the need for innovative approaches to challenging application problems. Typically, these problems arise when a nonstatistician is faced with a problem that goes beyond the capabilities of standard statistical tools available in commonly used statistical software. Working together, a statistician and subject matter expert have collective knowledge that neither possesses alone. Working with a statistician, the subject matter expert has access to an extensive collection of approaches and techniques for developing custom solutions. Working with a subject matter expert, the statistician is exposed to insights that can be key to identifying the true problem and developing a solution that addresses the issue of interest. This partnering can lead to new methods of collecting or analyzing data and making decisions.

## 182 Genome-Wide Association Studies ● ✪

Section on Statistical Computing, Biometrics Section, ENAR, WNAR
**Monday, August 7, 2:00 pm–3:50 pm**

### Genome-Wide Disease Gene Mapping by Association Analysis

❖ Jurg Ott, The Rockefeller University, Laboratory of Statistical Genetics, 1230 York Avenue, New York, NY 10021-6399, *ott@rockefeller.edu*

***Key Words:*** gene mapping, linkage disequilibrium, genetic association

After discussing the rationale for genetic association analysis, I will outline the steps currently taken for genome-wide association analysis. Problem areas to be discussed are multiple testing, false discovery rate, probability of a real finding when a result is statistically significant, different forms of heterogeneity, and multi-locus approaches. I will mention a recent success story (age-related macular degeneration) and conclude with plans for a new project in Beijing.

### The Genetics of Insulin Resistance: Clusters and SNPs

❖ Richard A. Olshen, Stanford University, Division of Biostatistics, HRP Redwood Building, Stanford, CA 94305-5405, *olshen@stat.stanford.edu*

***Key Words:*** clustering, quantization, bootstrapping, SNPS, insulin, resistance

This talk is in part about Gauss mixture vector quantization as it applies to clustering data from two-hour oral glucose tolerance tests, triglycerides, total cholesterol, HDL, and waist/hip ratio after removing the effects of age and body mass index. Subjects are 557 Chinese women in 285 sibships from the SAPPHIRe network of the NHLBI's Family Blood Pressure Program. Clustering enables precise definition of insulin resistance. There are concerns regarding how to eliminate the joint effect of age and BMI and of how to bootstrap appropriately from the dependent data. Results from SNP genotyping in genes that are candidates for hypertension are applied with other measures to classify individuals of unknown cluster membership into clusters. The work, which is in progress as this abstract is being written, is joint with Sangho Yoon, Alfred Lin, and the SAPPHIRe team.

### Identifying Interactions in Genome-Wide Association Studies

❖ Charles Kooperberg, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M3-A410, Seattle, WA 98109-1024, *clk@fhcrc.org*

***Key Words:*** SNPS, epistasis, haplotypes, gene-environment

Over the next few years, it will become increasingly common to conduct genome-wide association studies involving hundreds to thousands of Single Nucleotide Polymorphisms (SNPs) because of breakthroughs in high-throughput technologies. For many of these studies, interest will not be limited to just the characterization of individual SNPs or haplotypes that are associated with a disease outcome, but will include the identification of interactions between SNPs within a gene (as in haplotype effect), between genes (epistasis), or between gene and environment (e.g., drugs, smoking, and alcohol consumption). We will discuss situations, including study designs, where it is possible to identify particular types of interactions.

### Choices and Consequences of Genetic Marker Selection on Whole-Genome Association Scans

❖ Lon Cardon, University of Oxford, Wellcome Trust Centre for Human Genetics, Welcome Trust Principle Research Fellow, Oxford, OX3 7BN UK, *lon.cardon@well.ox.ac.uk*

***Key Words:*** genetic, association, SNP, whole-genome

Whole-genome association scans involve hundreds of thousands of genetic markers, selected from more than 4 million known variants. The scale of the problem is unprecedented in genetic research. A number of methods have been developed for marker selection, most of which are aimed at exploiting known patterns of correlations in reference samples to minimize redundancy. These design choices have a number of potential consequences on subsequent analysis strategies. Missing data in experimental conditions, which can occur at high rates and generally not at random, exacerbate the problem. We describe methods for marker selection and examine the trade-off between efficiency and restrictions in subsequent analysis and computational challenges facing this novel area of applied genetic research.

## 183 Regression Models with Functional Predictors

Section on Nonparametric Statistics

**Monday, August 7, 2:00 pm–3:50 pm**

### Functional Variance Processes and Volatility Modeling

❖ Hans-Georg Mueller, University of California, Davis, Department of Statistics, One Shields Avenue, Davis, CA 95616, *mueller@wald.ucdavis.edu*

***Key Words:*** functional data analysis, principal component analysis, random effect, stochastic process, variance function, seismology

Functional variance processes are introduced as a tool for functional data analysis, aiming at a stochastic model for quantifying variation in the local error variance in the situation where one has a sample of functions. These processes may be viewed as an extension of the concept of a variance function known from nonparametric regression. Applications that will be discussed include a seismic discrimination problem from geophysics, and the modeling of volatility from repeatedly observed series of financial market data. It is shown that the concept of a functional variance process proves useful in these contexts. This talk is based on joint work with Rituparna Sen, Ulrich Stadtmueller, and Fang Yao.

### Interpretable Functional Regression Models

❖ Gareth James, University of Southern California, Bridge Hall 401P, Information and Op Mgt Department, Los Angeles, CA 90089-0809, *gareth@usc.edu*

***Key Words:*** functional regression, interpretable models, Dantzig selector, variable selection

Regression models to relate a scalar Y to a functional predictor X(t) are becoming increasingly common. Work in this area has concentrated on estimating a coefficient function B(t) with Y related to X(t) through the integral of B(t) multiplied by X(t). Points where B(t)< >0 correspond to points where there is a positive/negative relationship between X(t) and Y. Alternatively, points where B(t)=0 indicate no relationship. Hence, identification of these regions has important implications. Unfortunately, most fitting procedures result in a B(t) with unnatural wiggles, making the curve hard to interpret. In this talk, we introduce a new approach that borrows ideas from a new class of variable selection procedures to produce coefficient curves that have simple shapes. This approach greatly simplifies the interpretation of the relationship between X(t) and Y over different regions of t

### Aspects of Feature Selection in Functional Data

❖ Philip J. Brown, University of Kent, IMSAS, Cornwallis Building., Canterbury, CT2 7NF UK, *pjb8@kent.ac.uk*

***Key Words:*** Bayesian methods, hyper-LASSO, wavelet functional modelling, variable selection, p>n problem

We look at functional data as arising from infrared applications in chemometrics and mass spectroscopy data used in proteomics. The data may contain experimental factors and covariates, but there is a desire to discriminate between two or more groups. Modeling often is facilitated by the use of wavelets. We review a variety of approaches to modeling the functional data as response and modeling directly the discriminatory categories conditional on functional data and experimental factor/covariates. Our ultimate focus will be on Bayesian models that allow regularization. To this end, we look at a variety of forms of scale mixture of normal prior distributions, including forms of hyper-lasso and approaches to robustness and stability of discrimination. We are particularly interested in fast algorithms capable of scaling up to many variables and which are flexible.

# 184 Threshold Regression Models and Applications

Section on Statistics in Epidemiology
**Monday, August 7, 2:00 pm–3:50 pm**

## Threshold Regression for Survival Analysis: Modeling Event Occurrence When Latent Health Status Decreases to a Threshold

❖ George A. Whitmore, McGill University, Faculty of Management, 1001 Sherbrooke St West, Montreal, PQ H3A 1G5 Canada, *george.whitmore@mcgill.ca*

***Key Words:*** first hitting time, threshold, non-proportional hazards, lung cancer, occupational risk, latent models

Considerable research has investigated first hitting times as models for survival and other event times. A first hitting time is the earliest time a stochastic process reaches a fixed threshold or boundary state. In the medical context, the process represents the latent health status of a subject and the threshold represents a critical level of health that triggers an adverse medical event (relapse, disease onset, death). The time scale can be calendar time or another operational measure of disease progression. Threshold regression refers to first hitting time models with regression structures that accommodate covariate data. The process, threshold parameters, and time scale all may depend on the covariates. Threshold regression methodology already has demonstrated its value in studies of infectious disease, cancer, and occupational risk.

## Application of Threshold Regression in Analyzing Lung Cancer Rates in Women

❖ Mei-Ling T. Lee, The Ohio State University, Biostatistics Division, School of Public Health, B-122 Starling-Loving Hall, 320 West 10th Ave, Columbus, OH 43210, *meilinglee@sph.osu.edu*; Bernard Rosner, Harvard Medical School

***Key Words:*** lung cancer, threshold regression, non-proportional hazard models, proportional hazard models, hazard ratios, longitudinal survey data

On the basis of longitudinal survey data from the Nurses' Health Study, we investigate women's lung cancer rates with respect to smoking exposure using threshold regression methodology. Smoking history, smoking status, and pack-years were used to measure smoking exposure. We calculate age and exposure-specific lung cancer rates and obtain hazard ratios for smokers and nonsmokers. We compare our results with those provided by a proportional hazards approach.

## Latent Process Models with Multiple Types of Observations

❖ Daniel Commenges, UniversitÈ Bordeaux, INSERM EMI 0338, Bordeaux, 33076 France, *daniel.commenges@bordeaux.inserm.fr*

***Key Words:*** latent process, joint model, coarsening, threshold, dementia

Latent process models are useful for modeling survival data but are even more useful for joint modeling several types of observations. Joint modeling appears to be a promising approach, for instance in the context of ageing, where it is possible to relate scores to psychometric tests and diagnosis of dementia to the same latent process representing cognitive ability. An observed event (for instance, dementia) may be determined by the latent process hitting a barrier; alternatively, the latent process may appear as a time-dependent variable in a proportional hazard model. We exhibit a general framework, relating observations (of events or of markers) to the latent process and making clear the difference between random transformations and coarsening.

# 185 Applications of Machine Learning Techniques in Bioinformatics and Other Areas ●

Biometrics Section, Section on Nonparametric Statistics, ENAR
**Monday, August 7, 2:00 pm–3:50 pm**

## Graph-Based Classifiers in Semi-Supervised Learning

❖ George Michailidis, University of Michigan, Department of Statistics, 439 West Hall 1085 South University, Ann Arbor, MI 48109-1107, *gmichail@umich.edu*

***Key Words:*** semi-supervised learning, sequential predictions, smoother matrices

A graph-based, nearest-neighbor classifier is proposed for semi-supervised learning, which utilizes the labeled data and topology of the graph for training and produces proper probability class estimates. In addition, a sequential procedure is developed for classifying unlabeled nodes. The procedure is iterative in nature and uses the topology of the graph and the cotraining information available in labeled and unlabeled data. The performance of the proposed classifier is assessed on several synthetic and real data. Extensions to settings where the available information comes from several domains modeled by graphs also are considered.

## Estimating Differential Equation Models of Gene Expression Dynamics

❖ Theodore Perkins, McGill University, 3775 University St, Montreal, QC H3A 2B4 Canada, *perkins@mcb.mcgill.ca*

***Key Words:*** differential equations, gene expression, genetic networks, dynamics, Drosophila melanogaster

Mathematical/computational biologists often use differential equations to describe the dynamics of genetic networks---the interacting systems of DNA, mRNA, proteins, and various molecules that regulate life at the cellular level. Quantitative molecular abundance data is increasingly available, creating a strong demand for sound techniques for fitting differential equation models. Yet, little attention has been paid to this problem in statistics, probabilistic modeling, or machine learning communities. I will discuss statistical and computational challenges in fitting differential equations to noisy, incomplete biological data and my work on combining optimization and regression approaches that address these issues. I will present results on modeling spatio-temporal expression data from the gap gene network of Drosophila melanogaster.

### Statistical Models on Protein Complex Alignments

❖ Tony Chiang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M2-B876, Seattle, WA 98109, *tchiang@fhcrc.org*

***Key Words:*** alignment, protein, interaction, complex

Modeling protein interactions is one important aspect to biological experiments, but problems arise due to high error rates incurred by the various assay technologies. These errors include both false positives and false negatives and can be either stochastic or systematic in nature. We address the general problem of estimating all multiprotein complexes in a given organism with the presence of these errors. Currently, there is no sound mechanism to determine the accuracy of protein complex estimation (clustering) algorithms. We will discuss ways to compare and align protein complexes, but focus on determining an optimal alignment that yields a highest weight based on a similarity scoring criterion. We will discuss the complexity of this alignment and give examples of deterministic and approximation algorithms.

### Neural Network Imputation: an Experience with the National Resources Inventory Survey

❖ Tapabrata Maiti, Iowa State University, Department of Statistics, 221 Snedecor, Ames, IA 50011-0001, *taps@iastate.edu*

Imputation is needed in almost all major surveys. In large-scale national surveys, different groups of people work on different stages of the surveys. Often, the statistical estimation group has little or insufficient communication with the other groups. In such situations, hotdeck type of imputation is difficult to apply. On the other hand, because of the complex nature of the survey, finding a suitable model may not be easy and thus a nonparametric method---such as neural network imputation---seems reasonable. One such national survey is the U.S. Department of Agriculture's National Resources Inventory Survey (NRI). By design, the survey has a lot of missing values, and the missing values are imputed using a donor-based method. This article develops a neural network imputation model and compares it with the results of the existing imputation method. The end result looks promising.

### Analyzing Gene Expression Data Using Tree-Based Models

❖ Bret Musser, Merck Research Laboratories, P.O. BOX 2000, RY34-A316, Rahway, NJ 07065, *bret_musser@merck.com*

***Key Words:*** gene expression, tree-based models

In any model-building process, relationships between predictor variables cause complications. In "feature selection" approaches such as recursive partitioning (RP), correlated predictors may be mutually substitutable, make additive contributions, or show synergy. These features are rarely discovered in recursive partitioning models because most RP methods generate a single tree as "the" answer. The solution is to produce not a single tree, but a "forest" of models that fit the data. Harvesting knowledge from the forest is much harder than generating it though. Not only may visually different trees be identical in the sense of generating identical rules, but given correlated predictors, different rules may lead to identical sets of terminal nodes. Different methods of building a "forest" of tree-based models will be discussed and applied to gene expression data.

## 186 Postmarket Issues in Medical Devices ●

Biopharmaceutical Section

**Monday, August 7, 2:00 pm–3:50 pm**

### Limiting Distributions of Resistances for Specific Radiation Sterilization Doses

❖ Harry Bushar, U.S. Food and Drug Administration, DHHS/PHS/FDA/CDRH/OSB/DBS, 1350 Piccard Drive, HFZ550, Rockville, MD 20850, *harry.bushar@fda.hhs.gov*

***Key Words:*** distributions of resistances, D10 value, bioburden, verification dose, substantiation for sterilization, sterilization dose

The Association for the Advancement of Medical Instrumentation (AAMI) is developing a Technical Information Report (TIR) to allow substantiation for sterilization of health care products at certain specific radiation doses (15, 17.5, 20, 22.5, 25, 27.5, 30, 32.5, and 35 kGy). Using simulations, non-unique limiting distributions of resistances are found, which consist of a relatively high D10 value (dose required to achieve inactivation of 90% of the average Bioburden [population of viable microorganisms]) (occurring with 5% frequency) and a relatively low D10 (occurring with 95% frequency), where both high D10 and low D10 decrease with increasing Bioburden, while verification dose (radiation dose predicted to give a sterility assurance level (SAL) = 0.1 [probability of a single viable microorganism]) first increases and then decreases with increasing Bioburden and evolving distributions.

### Issues Encountered in Statistical Analyses of Complex Experimental Medical Device Data

❖ Hollington Lu, Center for Devices and Radiological Health, 1350 Piccard Drive, Room 100Q, HFZ550, Rockville, MD 20850, *hollington.lu@fda.hhs.gov*; Barbara Krasnicka, U.S. Food and Drug Administration

***Key Words:*** experimental design, mixed models, longitudinal data, statistical quality control, hemodialyzer reuse

New medical device approvals are to be based on the results of in vitro and in vivo studies through research. However, a design of an in vitro experiment and the subsequent statistical analysis may be complicated. For such cases, applying simple summary statistics to characterize variability is insufficient; more sophisticated techniques are required. A short comparison of applicable approaches (e.g., quality control summary, GLM, mixed model) will be presented with special emphasis paid to the underlying assumptions, correct interpretation of results, and evaluation of methods. Also, examples will be discussed in the context of hemodialyzer reuse studies using data collected by device manufactures. Factors of interest will be type of dialyzer, lot of production, blood flow, and number of reuse.

### Frailty Model for Assessing Treatment Effect between Bare-Metal and Drug-Eluting Stents with Multiple Cardiac Events: Experience from Post-Marketing Registries

❖ So Jung Imm, Boston Scientific Corporation, 2710 Orchard Parkway, San Jose, CA 95134, *sojung.imm@bsci.com*; Scott Wehrenberg, Boston Scientific Corporation; Aijun Song, Boston Scientific Corporation; Zheng Zhou, Boston Scientific Corporation

***Key Words:*** frailty model, repeated events, event dependency, treatment effect

In clinical studies of treatment effect, it is common for patients to experience multiple events over time, and event times are correlated due to event dependency at the patient level and heterogeneity across individual patients (randomness). In coronary stent treatment assessment, cardiac events traditionally are studied by modeling only the time to first such event in the survival analysis. This may not provide a full picture of the treatment efficacy. In this exercise, repeated events were studied using a variance-corrected model (the Andersen-Gill model) and a conditional frailty model. The latter has the ability to consider both event dependency and heterogeneity across individuals. Using data from post-marketing stent registries, we implemented these models and empirically compared their results on treatment effect regarding drug-eluting versus bare-metal stents.

### Multiple Imputation for Missing Data in Propensity Score Generation: Application in Comparing Two Stenting Techniques Using Post-Marketing Registry Data

❖ Aijun Song, Boston Scientific Corporation, 1 Scimed Place, Mail Stop A242, Maple Grove, MN 55311, *Shannon.Song@bsci.com*; Scott Wehrenberg, Boston Scientific Corporation; Zheng Zhou, Boston Scientific Corporation

***Key Words:*** propensity score, multiple imputation, Markov chain Monte Carlo, regression

Propensity score has been used routinely in observational studies, comparing treatment effect and adjusting for the baseline characteristics imbalance due to the treatment not being assigned at random. Because of a large number of variables involved in the propensity score generation, missing data in some variables are common. Single imputation method, such as imputing missing value with the mean of the complete cases, is biased because the uncertainty about the predictions of missing values is not considered. In this research exercise, several multiple imputation methods will be investigated to compute the missing data. The adjusted treatment effects will be compared based on the resulting propensity score using multiple imputation. We will apply these methods to patient data comparing direct stenting and pre-dilation techniques in a post-marketing registry of drug-eluting stent.

## 187 Bayesian Applications in Medical Devices ● ✪

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Monday, August 7, 2:00 pm–3:50 pm**

### Using Auxiliary Information in Clinical Trials

❖ Shu Han, Guidant Corporation, 1150 Cushing Circle, APT 244, Saint Paul, MN 55108, *shu.han@guidant.com*; Donald Berry, The University of Texas

***Key Words:*** auxiliary information, Bayesian, clinical trials, nonparametric, kernel density

We develop parametric and nonparametric models that utilize early endpoints in making treatment comparisons concerning the primary endpoint. Using all available information in this way enables earlier decisions about treatment benefits and futility. It also improves the efficiency of adaptive designs. Our parametric model is Bayesian and assumes the data are normally distributed. Our nonparametric model uses kernel density estimation. Our simulation results demonstrate that both parametric and nonparametric models have advantages over conventional methods. The parametric model performs slightly better than the nonparametric model when the distributions of the early and primary endpoints are bivariate normal. The nonparametric model is robust in that it performs better than the parametric model when these distributions deviate from normality sufficiently.

### The Role of Accrual Rate and Follow-up Time in a Bayesian Adaptive Design

❖ Xuefeng Li, U.S. Food and Drug Administration, 1350 Piccard Drive, HFZ550, Rockville, MD 20850, *xuefeng.li@fda.hhs.gov*

***Key Words:*** adaptive design, interim analysis, Bayesian

Bayesian designs with adaptive sample size and interim looks are used more and more in clinical trials for therapeutic orthopedic devices. Clinical parameters (e.g., accrual rate and follow-up time) play an important role in this kind of design. In this study, how accrual rate and follow-up time will affect a typical Bayesian adaptive design is studied through simulations. How the sample size and operating characteristics change with the accrual rate and follow-up time are explored. Some suggestions are given for choosing appropriate Bayesian adaptive designs under different situations. Further topics and developments are discussed.

### A Tiered Treatment Design for a Historically Controlled Medical Device Clinical Trial

❖ Alistair O'Malley, Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115-5899, *omalley@hcp.med.harvard.edu*

***Key Words:*** Bayesian analysis, coronary-artery stents, historical control, hierarchical modeling, medical devices, tiered treatment design

I will describe a novel design for a single-armed trial of a new drug-coated coronary-artery stent. In order to make comparisons against coronary artery bypass graft (CABG) surgery, two historical trials involving a third treatment (bare metal stenting) are used. One of the historical trials compares the drug-coated stent to bare metal stenting, while the other compares bare metal stenting to CABG. I will discuss the advantages and disadvantages of the "tiered treatment design" and present the results from the analysis of the trial. In particular, I will demonstrate that, in this case, the tiered treatment design leads to a different conclusion than the standard historical-control analysis.

### How Bayesian Hierarchical Models Handle Multiplicity Issues Automatically: a Case Study in a Medical Device Trial

❖ Feng Tang, Medtronic, Inc., 1015 Gramsie Road, Mail Stop Z240, Saint Paul, MN 55126, *feng.tang@medtronic.com*; Lou Sherfesee, Medtronic, Inc.; Andrew Mugglin, University of Minnesota

***Key Words:*** Bayesian hierarchical model, multiplicity, stratification, type I error

We present how Bayesian hierarchical models handle multiplicity issues automatically using an example in a medical device trial. Initially, a Bayesian design was used to study the efficacy of a therapy for patients with indication type A. However, due to the slow enrollment of this group of patients into the trial and the potential similar benefit of the therapy on another group of patients with indication type B, the decision was made to include type B patients in the current trial. As a result, the initial model and design were revised. Rather than making separate inferences for the two hazard ratios, which requires adjustment of existing decision criteria to control overall simulated type I error, we added a layer of hierarchy to the existing model to allow us to keep the same decision criteria, yet with desirable frequentist properties.

## 188 Statistical Methods for Evaluating Racial-Ethnic Disparities in Health

Section on Health Policy Statistics

**Monday, August 7, 2:00 pm–3:50 pm**

### A New Method for Estimating Racial/Ethnic Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity

❖ Marc Elliott, RAND Corporation, 1776 Main Street, P O Box 2138, Santa Monica, CA 90407-2138, *elliott@rand.org*; Allen Fremont, RAND Corporation; Nicole Lurie, RAND Corporation; Peter A. Morrison, RAND Corporation; Philip Pantoja, RAND Corporation; Allan Abrahamse, RAND Corporation

***Key Words:*** Bayes theorem, surname lists, geocoding, managed care, health plan, administrative data

Racial/ethnic disparities in health care are at the forefront of health policy. We consider health plans' need to monitor racial/ethnic disparities among their enrollees. Few plans acquire self-reported racial/ethnic data from their entire membership. When these data are missing, surnames and neighborhood contextual measures can provide useful surrogate information for comparing subgroups. We describe and evaluate a hybrid method applicable where researchers must rely on administrative records lacking racial/ethnic detail. Our Bayesian algorithm takes block group race/ethnic composition as a prior multinomial distribution and updates these into posterior probabilities on the basis of the joint outcomes of Hispanic and Asian surname list 'tests.' This approach substantially outperforms other approaches in terms of classification accuracy and the MSE of health disparity estimates.

### Power of Tests for a Dichotomous Independent Variable Measured with Error

❖ Daniel McCaffrey, RAND Corporation, 201 N. Craig Street, Suite 202, Pittsburgh, PA 15213, *danielm@rand.org*; Marc Elliott, RAND Corporation

***Key Words:*** maximum likelihood, measurement error

We consider the power of three methods of testing differences in the means of groups defined by an unobserved dichotomous variable with known conditional probability p, given other predictors. The first method classifies observations into two groups by whether p exceeds a threshold. Method two tests the coefficient in a regression of the outcome on p. Method three estimates the difference in means by maximizing the outcomes' likelihood, given p. The efficiency of method one roughly scales with the square of one less the classification error. The efficiency of method two roughly scales with the R-square for predicting the unobserved dichotomous variable and is usually more powerful than method one. Method three is most powerful, but simulations show that for differences in means of 0.2 - 0.5 standard deviations, method two is about 90% as efficient as the MLE.

### Improving the Accuracy of Health Estimates for Small Racial-Ethnic Groups by Weighted Pooling over Time

❖ Brian Finch, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, *bfinch@mail.sdsu.edu*; Marc Elliott, RAND Corporation; Daniel McCaffrey, RAND Corporation; David Klein, RAND Corporation; Daniela Golinelli, RAND Corporation

While most health surveys collect substantial data on large race-ethnic groups, little is known about the extent of disparities for smaller groups, such as Asian subgroups and American Indians/Alaska Natives. In fact, major national health surveys have limited or no ability to assess adequately the health of major race-ethnic subgroups smaller than that of Mexican-American. One solution to improving the precision of estimates is to pool data over time. The simplest solution, using the average of several recent years to estimate the present year, is biased in the presence of a linear trend and inefficient in the presence of autocorrelation. We propose and evaluate a Kalman filter approach in which we detrend repeated cross-sections and exponentially weight on the basis of an autocorrelation estimated by pooling across larger race-ethnic groups.

### Differential Use of 0-10 Rating Scales by Racial-Ethnic Minorities in CAHPSÆ

❖ Robert Weech-Maldonado, University of Florida, Department of Health Services and Research Mangement Policy, P O Box 100195, Gainesville, FL 32610, *rweech@phhp.ufl.edu*; Marc Elliott, RAND Corporation; K. Cameron Schiller, University of Florida; Ron D. Hays, University of California, Los Angeles

***Key Words:*** race/ethnicity, consumer assessments, Medicaid, rating scales

Understanding cultural differences in the use of rating scales is critical to evaluating racial/ethnic disparities in consumer assessments of care. The Consumer Assessments of Healthcare Providers and Systems (CAHPSÆ) survey uses four 0-10 global ratings (i.e., personal doctor, specialist, plan, and health care), among other measures. Using multinomial logistic regression on more than 31,622 Medicaid and 229,790 commercial managed care enrollees in 2000, we find evidence that Hispanic ethnicity and Medicaid status are both associated with increased endorsement of "10" relative to an omitted category of "5"-"8" after case-mix adjusting for age, gender, education, and self-rated health status. Because of the negative skew, what actually may be a difference in dispersion or use of scale extremes may be misinterpreted as a difference in central tendency.

## 189 Ranked Sampling I ●

Section on Nonparametric Statistics

**Monday, August 7, 2:00 pm–3:50 pm**

### New Imperfect Rankings Models for Ranked-Set Sampling

❖ Jesse Frey, Villanova University, Department of Math Sciences, Villanova, PA 19085-1699, *jesse.frey@villanova.edu*

***Key Words:*** ranking models, judgment rankings, Bernstein polynomials, Birkhoff's theorem

Ranked-set sampling is an approach that leads to improved statistical inference in situations where the units to be sampled can be ranked

relative to each other before formal measurement. This ranking need not be completely accurate. In fact, results in the literature have shown that no matter how poor the ranking, procedures based on ranked-set sampling tend to be at least as efficient as procedures based on simple random sampling. However, efforts to quantify the gains in efficiency for ranked-set sampling procedures have been hampered by a shortage of models for imperfect rankings. In this paper, we introduce a new class of models for imperfect rankings and prove that essentially any reasonable model for imperfect rankings is a limit of models in this class. We then describe a specific, easily applied method for selecting an appropriate model from the class.

### Ranked Set Sample Inference under the Constraint of Stochastic Ordering of Judgment Ranking Classes

❖ Omer Ozturk, The Ohio State University, 1958 Neil Ave., Columbus, OH 43210, *omer@stat.ohio-state.edu*

***Key Words:*** isotonic regression, imperfect ranking, perfect ranking, consistent estimator, ranking model

In ranked set sampling procedure, ranking error is almost an unavoidable reality. Thus, a reasonable statistical inference should address the validity of the procedures under imperfect ranking. In this talk, we will develop statistical inference for the estimation of the cumulative distribution function of the underlying distribution under the stochastic order constraint of the judgment class distributions. We will show that the CDF estimator of judgment classes are consistent for the CDF of the ranking class distributions. We next will use these estimators to develop statistical inference for ranked set samples that may contain ranking errors. Examples will be provided where these estimators can be used to draw inference for a characteristic of the underlying distribution.

### Concomitant of Multivariate Order Statistics with Application to Judgment Post-Stratification

❖ Xinlei Wang, Southern Methodist University, 2121 Loch Haven Drive, Plano, TX 75023, *swang@smu.edu*; S. Lynne Stokes, Southern Methodist University; Johan Lim, Texas A&M University; Min Chen, The University of Texas at Austin

***Key Words:*** ranked set sampling, Gaussian quadrature, least squares estimator

In this paper, we develop general expressions for the density of the concomitant of a random vector and establish its related properties. The concomitant of a normal random vector is studied in detail and methods for calculating its moments are discussed. Finally, we apply the theory to develop new estimators of the mean from a judgment post-stratified sample, where post-strata are formed by rank classes of auxiliary variables. Our estimators are shown to be more efficient than existing ones and robust against violations of the normality assumption. They also are well-suited to applications requiring cost efficiency.

### Approximate Confidence Intervals from a Ranked Set Sample

❖ Christopher Sroka, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210, *csroka@stat.ohio-state.edu*; Elizabeth Stasny, The Ohio State University; Douglas Wolfe, The Ohio State University

***Key Words:*** judgment order statistic, variance estimation, stratified sampling, medical expenditure panel survey

Prior research has shown that ranked set sampling (RSS) results in more precise estimation than simple random sampling (SRS). However, it has not been shown how to capitalize on this increased precision to construct a shorter confidence interval for the mean. We provide a formula for a large sample confidence interval based on the standard normal distribution. This technique relies on an estimator of the variance of the judgment order statistics. This estimator has the same properties as the sample variance used in SRS. Computer simulations show our method results in confidence intervals that are over 40% shorter than confidence intervals based on an SRS of equivalent sample size. RSS and SRS perform similarly with respect to how quickly their respective confidence intervals converge to the nominal level.

## 190 Synthetic Data or No Data: Choices in Disclosure Avoidance Methods

Section on Government Statistics

**Monday, August 7, 2:00 pm–3:50 pm**

### U.S. Census Bureau Disclosure Avoidance Practices and Research: an Update for JSM 2006

❖ Laura Zayatz, U.S. Census Bureau, 4700 Silver Hill Road, Commerce Census SRD 3209 4, Washington, DC 20233-0001, *laura.zayatz@census.gov*

***Key Words:*** confidentiality, data access

The U.S. Census Bureau collects its survey and census data under Title 13 of the U.S. Code, which promises confidentiality to its respondents. The agency also has the responsibility of releasing data for the purpose of statistical analysis. The goal is to release as much high-quality data as possible without violating the pledge of confidentiality. We apply disclosure avoidance techniques prior to publicly releasing data products to protect the confidentiality of our respondents and their data. This paper discusses the types of data we release, our disclosure review process, restricted access procedures, disclosure avoidance techniques being used, and current disclosure avoidance research.

### Part 2: Myth & Reality - Complementary Cell Suppression

❖ Ramesh Dandekar, Energy Information Administration, 1000 Independence Ave., Washington, DC 20585, *ramesh.dandekar@eia.doe.gov*

***Key Words:*** tabular data, disclosure, estimation, linear programming, audit

Currently, complementary cell suppression procedures are used mostly by statistical agencies to protect sensitive tabular data from disclosure. In recent years, LP-based automated audit procedures have been advocated and are being used to ensure the adequacy of protection offered by cell suppression patterns. LP-based lower and upper bounds for suppressed tabular cells typically are used to determine the adequacy of disclosure control measures. A previous "Myth & Reality" paper (Dandekar 2005) showed how interior point method-based LP tools can be used to cause the statistical disclosure of sensitive tabular cells. In Part 2, statistical experimental design principles are used to demonstrate the relative ease with which the statistical disclosure of sensitive tabular data can occur by using readily available personal computer (PC) software tools.

## Partial Synthesis for Disclosure Avoidance

❖ Sam Hawala, U.S. Census Bureau, 4700 Silver Hill Road, room 3211A bldg 4, Washington, DC 20233, *sam.hawala@census.gov*

***Key Words:*** confidentiality protection, disclosure avoidance, partially synthetic data, microdata, group quarters

We investigate the approach of releasing partially synthetic microdata. This form of data release, in theory, has the advantage of reflecting the loss of information necessary to avoid disclosure in the inferences users may want to make. The approach relies on an optimal variable selection technique to identify records and variables for synthesis, the production of synthetic records using appropriate models, and an assessment of data utility and disclosure risk after reintegrating the synthetic records into the original dataset.

## The Disclosure Limitation Protocol for the Census Bureau's "On the Map" Origin-Destination Transportation Application

❖ Fredrik Andersson, Cornell University, CISER, 391 Pine Tree Road, Ithaca, NY 14850, *fredrik.andersson@census.gov*; John Abowd, Cornell University; Marc Roemer, U.S. Census Bureau

***Key Words:*** synthetic data, noise infusion, commuting patterns, geospatial data, posterior predictive distribution

This paper documents the disclosure limitation protocol for the Census Bureau's "On the Map" application, which manipulates and displays origin-destination patterns and associated worker and business characteristics. The version discussed in this paper consists of a block-to-block commute pattern matrix of jobs and workers in at least one state and associated data on the characteristics of the residence and workplace areas at the block-group level. The disclosure limitation protocol creates synthetic origin/residence data using the posterior predictive distribution of the workers' place of residence conditional on estimates of employed workers that have been protected using the noise-infusion disclosure limitation protocol from the Quarterly Workforce Indicators.

## Inferences on Two-Stage, Multiply-Imputed Data

❖ Satkartar Kinney, Duke University, 211 Old Chemistry Bldg., Box 90251, Durham, NC 27708, *saki@stat.duke.edu*; Jerome Reiter, Duke University

***Key Words:*** synthetic data, missing data, data confidentiality, multiple imputation

Partially synthetic datasets, where some values are replaced with multiple imputations, usually are generated for disclosure limitation. It is often the case that missing values also must be multiply imputed. We will discuss the generation and analysis of two-stage, multiply-imputed datasets, particularly significance testing for estimands from such data.

# 191 Census Coverage Measurement

●

Section on Survey Research Methods

**Monday, August 7, 2:00 pm–3:50 pm**

## 2010 Census Coverage Measurement: Initial Results of Net Error Empirical Research Using Logistic Regression

❖ Richard Griffin, U.S. Census Bureau, 104 Arrowhead Circle, Lansdale, PA 19446, *richard.a.griffin@census.gov*; Thomas Mule, U.S. Census Bureau; Doug Olson, U.S. Census Bureau

***Key Words:*** dual system estimates, logistic regression models, cross-validation

For the dual-system estimates in past censuses, a post-stratification approach has been used. The post-stratification approach has significant limitations as it limits the number of factors that can be included due to sample size constraints. Statistical modeling techniques, such as logistic regression, offer more flexibility and possibilities for reducing sampling error and bias. The March 2001 post-stratification variables are used. The focus is on investigating the trade-off between bias and variance of estimates obtained by the elimination of higher-order interaction terms in the models and examining measures to evaluate and compare the fit of alternative models. Alternative estimates and their standard errors for race/origin/tenure groups are presented and contrasted. A logarithmic penalty function is suggested for model selection, and cross validation is used.

## 2010 Census Coverage Measurement Research on Person Coverage Estimates by Housing Unit Enumeration Status

❖ Vincent Mule, U.S. Census Bureau, 15007 Nebraska Lane, Bowie, MD 20716, *vincent.t.mule.jr@census.gov*

For the 2010 Census, the goals and objectives of the Census Coverage Measurement (CCM) program have been expanded to include components of coverage errors that include erroneous enumerations and omissions. Because one of the goals and objectives is to estimate the omissions by whether the housing unit was included in the census, we are researching ways to estimate the true population by whether the housing unit was enumerated. We present our research findings on methods for estimating the population in enumerated housing units. For this population, we examined whether we can use the housing unit matching to allow census outcome covariates to be used in the modeling of the match status.

## A Nonparametric Approach to Census Population Size Estimation

❖ Song X. Chen, Iowa State University, Department of Statistics, Ames, IA 50011-1210, *songchen@iastate.edu*; Chengyong Tang, Iowa State University; Jean D. Opsomer, Iowa State University; Sarah M. Nusser, Iowa State University

***Key Words:*** capture-recapture surveys, discrete covariates, kernel smoothing method

We consider applying an approach based on the nonparametric kernel smoothing methods for inference issues surrounding the U.S. census population size estimation. The issues considered include checking the validity of commonly assumed models for the probability selection function and automatic combinations of post-stratums to gain more information from the capture-recapture type accuracy and coverage evaluation surveys.

### 2010 Census Coverage Measurement: the Hunt for the Magic Variables

❖ Eric Schindler, U.S. Census Bureau, FOB 2, Room 2404, Washington, DC 20233, *eric.l.schindler@census.gov*

***Key Words:*** post-stratification, dual system estimation

The Census 2000 Accuracy and Coverage Evaluation (A.C.E.) employed a post-stratified, dual system estimation design. The key post-stratification variables were race, Hispanic origin, age, sex, and tenure, familiarly known as ROAST. Additional variables used in or considered for A.C.E. were census region; type of enumeration area; tract mail return rate; family type; return type/return date/proxy status, state, local census office; and tract percent minority. The 2010 Census Coverage Measurement program will employ a logistic regression-based, dual-system estimator with the ROAST variables and additional variables. This paper examines the differences several of these additional variables make in the dual system estimates, especially at a higher level of aggregation, such as for states or for race/origin groups, once ROAST is included in the post-stratification model.

### Framework for Census Coverage Error Components

❖ Mary Mulry, U.S. Census Bureau, 2884 Manorwood Trail, Fort Worth, TX 76109, *mary_mulry@yahoo.com*; Donna Kostanich, U.S. Census Bureau

***Key Words:*** erroneous enumerations, omissions, undercount, dual system estimation, 2010 Census

A major goal and challenge for coverage measurement in 2010 is to design a survey that measures the components of coverage error, namely erroneous enumerations and omissions. The U.S. Census Bureau's previous coverage measurement surveys were designed primarily to estimate net census error using dual system estimation (DSE). To improve the accuracy of estimates of net error, the U.S. Census Bureau's DSE has relied on balancing some of the components of error, meaning some census omissions offset some erroneous inclusions in a manner that preserved the net error. As a result, the process produced inflated estimates of omissions and erroneous inclusions. This paper provides a framework for overcoming these inflated estimates of component errors. It also explicitly defines the components of error and discusses how they relate to traditional net error concepts.

## 192 Hiring a Consulting Statistician: What We Look for ●

Section on Statistical Consulting, Section on Statistical Education

**Monday, August 7, 2:00 pm–3:50 pm**

### Hiring a Consulting Statistician: What We Look for

❖ Christina M. Gullion, Kaiser Permanente Center for Health Research, 3800 N. Interstate Ave., Portland, OR 97227-1110, *christina.gullion@kpchr.org*; ❖ K. B. Boomer, The Pennsylvania State University, 323 Thomas Building, University Park, PA 16802-211, *boomer@stat.psu.edu*; ❖ Fred Hulting, General Mills, Inc., 9000 Plymouth Ave., N., Minneapolis, MN 55427, *fred.hulting@genmills.com*; ❖ Don Harder, Eli Lilly and Company, Lilly Corporate Center, DC 2233, Indianapolis, IN 46285, *dharder@lilly.com*

***Key Words:*** consulting, careers, employment

Recruiting and hiring a consulting statistician can be a challenge. Positions often require a particular mix of technical and consulting skills, content area knowledge, and experience with applied problems. Panelists will discuss the process and criteria they typically use in identifying and evaluating candidates for a consulting statistician position. Each panelist works in a distinctive market segment of the organizations that employ consulting statisticians. We will focus on the particular problem of hiring experienced consultants (i.e., more than three years out of training).

## 193 The Promise and Potential of the American Community Survey ● ✪

Social Statistics Section

**Monday, August 7, 2:00 pm–3:50 pm**

### The Promise and Potential of the American Community Survey

❖ Charles Hirschman, University of Washington, Sociology Department, Box 353340, Seattle, WA 98195-3340, *charles@u.washington.edu*; ❖ Linda Gage, California Department of Finance, 915 L Street, Sacramento, CA 95616, *Linda.Gage@dof.ca.gov*; ❖ Linda Jacobsen, Population Reference Bureau, 1875 Connecticut Ave., Suite 520, Washington, DC 20009-5728, *ljacobsen@prb.org*

The idea for the American Community Survey (ACS)---and many of the practical aspects of implementation---was germinated by Leslie Kish as "rolling samples" of the population. Kish and many of his fellow visionaries in the U.S. Census Bureau saw the ACS not just as a substitute for the census long form, but as a new method for collecting data with more periodicity and flexibility than the traditional model of decennial censuses. Participants in this session will discuss the promise and potential of the ACS to implement Kish's vision as a new frontier of data collection for monitoring society, research, and informing social policy.

## 194 Learning from and Applying Statistics Education Research to Our Own Teaching ●

Section on Statistical Education

**Monday, August 7, 2:00 pm–3:50 pm**

### Learning from and Applying Statistics Education Research to Our Own Teaching

❖ Christine Franklin, University of Georgia, 3111 Mars Hill Road, Watkinsville, GA 30677-1588, *chris@stat.uga.edu*; ❖ Roxy Peck, California Polytechnic State University, San Luis Obispo, 1 Grand Ave., College of Science & Mathematics, San Luis Obispo, CA 93407-9000, *rpeck@calpoly.edu*; ❖ Robert Gould, University of California, Los Angeles, 1946 Overland Ave., 104, Dept. of Statistics, Los Angeles, CA 90025, *rgould@stat.ucla.edu*; ❖ Joy Jordan, Lawrence University, P.O. Box 599, Appleton, WI 54912-0599, *joy.jordan@lawrence.edu*

***Key Words:*** statistics education, research, teaching

Each panelist selected a research study from the statistics education literature and will discuss it briefly as well as talk about how s/he has (or will) use it to inform her/his own teaching of statistics. The main idea of this session is to help us think about using existing research to inform our teaching in addition to doing research in our classrooms. The discussant will talk about the implications of the statistics education research for our teaching in light of what the three panelists present.

# 195 Sample Survey Design II ●

Section on Survey Research Methods

**Monday, August 7, 2:00 pm–3:50 pm**

## Properties and Modifications of a Probability Proportional to Size Sampling Procedure

❖ Lawrence R. Ernst, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 1950, Washington, DC 20212-0001, *Ernst.Lawrence@bls.gov*

***Key Words:*** PPS sampling, unbiased estimators of variance, sample expansion, Essen's method

There are numerous procedures for selecting a sample PPS without replacement. Brewer and Hanif (1983) discuss the properties of 50 such procedures. One of the more interesting of these procedures originally appeared in Jessen (1969). This procedure possesses several desirable properties including: relative simplicity of sample selection, fixed sample size of any desired size, selection strictly PPS without replacement, unbiased estimators of variance, and simple calculation of joint inclusion probabilities. In this paper we study possible modifications of Jessen's procedure to obtain other desirable properties such as allowing for sample expansion with the expanded sample remaining PPS without replacement and adjustment of joint selection probabilities in order to improve the stability of the Yates-Grundy variance estimator and to exclude the possibility of negative variance estimates.

## Supplementing RDD Surveys with Web-Based Survey Data

❖ Karol Krotki, RTI International, 1615 M Street, Suite 740, Washington, DC 20036, *kkrotki@rti.org*

***Key Words:*** survey mode, selection probabilities, selection bias, weighting

With declines in response rates it becomes important to seek cost-effective alternatives to the telephone survey mode. One candidate is the web-based survey which, even though it is not based on a probability sample, comes from a frame that in many respects resembles the entire population. Furthermore, the web-based survey is much less expensive than RDD. We study the differences between the RDD and web-based survey data, we develop methodologies for combining the two sets of weights in a way that accurately reflects the underlying populations, and we develop design- and model-based strategies for the estimation process including variance calculations and significance testing. Specifically, these concepts will be developed using data from the 2004 Ohio Adult Tobacco Survey which consists of approximately 2,000 completed RDD interviews and 6,000 web-based interviews.

## Comparison of Mixed-Mode and Address Frame Designs to Random Digit Dialing for General Population Surveys

❖ Michael W. Link, Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, MS:K-66, Atlanta, GA 30341-3717, *MLink@cdc.gov*; Michael P. Battaglia, Abt Associates Inc.; Martin R. Frankel, Abt Associates Inc.; Larry Osborn, Abt Associates Inc.

***Key Words:*** random-digit dialing (RDD), telephone surveys, response rates, mixed modes, survey sampling

Multiple modes and address-based sampling may be viable alternatives to random digit dialing (RDD) for general population surveys. A 6 state pilot conducted as part of the Behavioral Risk Factor Surveillance System (BRFSS) compared two designs to RDD: (1) telephone numbers selected using RDD and reverse-matched for addresses, (2) addresses sampled from a postal database and matched to telephone numbers. For both, surveys were mailed to all identified addresses, with telephone follow-up of nonrespondents. Those without an identifiable address were contacted by telephone. The pilots were conducted in parallel with the RDD-based BRFSS survey, providing a baseline for comparison. The pilot designs and RDD survey are compared on response rates; respondent representativeness, including reaching cell phone-only households and those without telephones; and, comparability of survey estimates.

## Probability Sample Designs that Impose Superpopulation Models on Survey Data

❖ Stephen Woodruff, 800 Westview Terrace, Alexandria, VA 22301, *stephenwalx@aol.com*

***Key Words:*** ratio estimator, best linear unbiased estimator, Horvitz-Thompson estimator

For a substantial class of sampling problems, the sample design not only provides the sampling distribution but also imposes a superpopulation model on the sample data. For many of these sampling problems this duality is an academic curiosity but there are some cases where inference based on probabilities of selection would be statistical malpractice. In such situations, the model provides a vital backup and a Best Linear Unbiased Estimator (BLUE) that avoids Horvitz-Thompson (HT) based methodologies. Since the design imposes the model, model failure is a peripheral issue. An example is presented for mail sampling and other situations that encounter similar problems are mentioned. For the examples presented here, an adequate frame is not available for sample design but is only available after the sample has been selected and the data collected.

## A Review of the Sample Design for the 2001, 2003, and 2005 California Health Interview Survey

❖ Ismael Flores Cervantes, Westat, 1650 Research Blvd., Rockville, MD 20850, *ismaelflores-cervantes@westat.com*; Michael E. Jones, Westat; Laura Alvarez-Rojas, Westat; J. Michael Brick, Westat; John H. Kurata, University of California, Los Angeles; David Grant, University of California, Los Angeles

***Key Words:*** RDD, health survey, geographic stratification, surname list samples, within-household sampling

Sampling procedures used for selecting samples for the California Health Interview Survey (CHIS) are described in this paper. The CHIS telephone survey started in 2001 and continued in 2003 and 2005. The objectives of CHIS were to examine issues in public health and health care and to monitor changes over time for Californians. Each round presented many challenges addressed at the sample design stage. Changes of data user needs and their impact on the sample design are discussed. Methods such as geographic stratification, oversampling of small areas, disproportionate stratified sampling for minority areas, and the use of surname list frames were incorporated into the design. Techniques for increasing telephone dialing efficiency such as purging of nonresidential numbers, oversampling of cases with a matched address, and subsampling of cases for refusal conversion are also discussed.

### Comparing Alternate Designs for a Multidomain Cluster Sample

Pedro Saavedra, ORC Macro; ❖ Mareena McKinley Wright, ORC Macro, 11785 Beltsville Drive, Calverton, MD 20705, *Mareena.McKinley.Wright@orcmacro.com*; Joseph P. Riley, U.S. Department of Housing and Urban Development

***Key Words:*** simulations, probabilities proportional to size, two-phase sample, two-stage sample, assisted housing programs

The Assisted Housing Income (Re)certification Study calls for a national sample of housing projects and tenants from three assisted housing programs. Unfortunately, different programs predominate in different localities. In the past a composite measure of size was used to select clusters, and allocations were made for each PSU so as to insure sampling the same total number of tenants for each program. For the current cycle, a two-phase sample was used to select projects from each program so that probabilities within programs remain proportional to size. While this method leads to a lower design effect due to weighting, the work load is slightly uneven across PSUs. Options, such as a two-phase sample for two of the programs and a two-stage sample equalizing the total samples by PSU, are examined through simulations.

### The Impact of Unequal Minority Distribution across Schools on PPS Oversampling Methods

❖ William Robb, ORC Macro, 126 College Street, Burlington, Vermont, *william.h.robb@burlington.orcmacro.com*; Joshua Brown, ORC Macro; James Ross, ORC Macro; Ronaldo Iachan, ORC Macro

***Key Words:*** PPS, minority, oversampling, school-based, GINI, weighted MOS

Probability Proportional to Size sampling methods are widely used in multistage cluster samples. The measure of size used for primary sampling units can be weighted so as to concentrate the sample among units that are relatively rare - a method that is used extensively to oversample minority respondents. The success of oversampling depends on the magnitude of the weights used and the degree of differential concentration of minorities among sampling units. We present recent work quantifying the degree to which minority yields can be increased in school-based surveys using these methods given the degree of unequal distribution of minorities across schools, the impact of these methods on design effects, and the implications of the degree of minority concentration in the resulting sample, relative to that in the population.

## 196 Nonresponse, Imputation, and Estimation ●

Section on Survey Research Methods
**Monday, August 7, 2:00 pm–3:50 pm**

### Multiple Imputation for Incomplete Multivariate Data under a Latent-Class Selection Model

❖ Hyekyung Jung, The Pennsylvania State University, 331A Thomas Building, Penn State University, University Park, PA 16802, *hjung@stat.psu.edu*; Joseph L. Schafer, The Pennsylvania State University

***Key Words:*** incomplete multivariate data, multiple imputation, latent class model, nonignorable missingness, missingness mechanism, selection model

Most missing data procedures assume that the missing values are missing at random (MAR). This assumption is convenient but sometimes questionable if the probabilities of missingness depend on unobserved quantities. With multivariate data, the number of distinct missingness patterns becomes large, making traditional outcome-based selection modeling unattractive, which ties probabilities to the partially observed values directly. Information in the missing-data indicators is sometimes be well summarized by a simple latent-class structure, however, suggesting that the partially observed variables could be tied to probabilities of class membership. This paper describes methods for imputing nonignorably missing response using a latent-class selection model and examines the properties of multiple imputations created under the proposed model.

### Inferences on Missing Information and the Number of Imputation

❖ Ofer Harel, University of Connecticut, Department of Statistics, 215 Glenbrook Road Unit 4120, Storrs, CT 06269-4120, *oharel@stat.uconn.edu*

***Key Words:*** missing data, missing information, multiple imputation

In the presence of missing values, researchers may be interested in the rates of missing information. The rates of missing information are important for assessing how the missing information contributes to inferential uncertainty about, Q, the population quantity of interest; are an important component in the decision of the number of imputations; and can be used to test model uncertainty and model fitting. I will present the asymptotic distribution of the rates of missing information. Numerically, I will show that the proposed asymptotic distribution agrees with the simulated one. I also will suggest the number of imputations needed to obtain reliable missing information rate estimates based on the asymptotic distribution.

### Enhancements to the 2006 Canadian Census Edit and Imputation System

❖ Wesley Benjamin, Statistics Canada, 120 Parkdale Ave., RHCoats 15-C, Ottawa, ON K1A 0T6 Canada, *wesley.benjamin@statcan.ca*

***Key Words:*** edit, imputation, census, Canada, donor, deterministic

The CANadian Census Edit and Imputation System (CANCEIS) will do deterministic imputation plus perform minimum change donor imputation for all variables (both numeric and categorical) in the 2006 Canadian Census. Significant enhancements have been made to CANCEIS for the 2006 Census, including the ability to perform deterministic imputation, process alphanumeric variables, do outlier detection of numeric variables and use failed records as donors. In addition, it is now easier to write compact decision logic tables and improvements have been made in the handling of numeric variables. These changes were implemented by methodologists, subject matter experts and systems developers working in a collaborative fashion. CANCEIS, or an earlier version of the software, has been used in the 1996 and 2001 Canadian Censuses, as well as the 2000 Swiss, 2000 Brazilian and 2005 Peruvian Censuses.

### Separating the Wheat from the Chaff: the Search for the Best Imputation Methodology

❖ Paula Weir, Energy Information Administration, 1000 Independence Ave., SW, Washington, DC 20585, *Paula.Weir@eia.doe.gov*; Pedro Saavedra, ORC Macro

***Key Words:*** imputation, nonresponse, regression, exponential smoothing

Developing an imputation methodology is often wrought with basic data issues. Yet, the interpretation and treatment of the data have bearing on the methods to be considered and the performance of competing estimators. In this study, the data suffer from a recent change in both the data elements collected and the processing system, confusion over truly missing versus zero values, and reliability of edit failed data elements. In addition, the implemented imputation must be performed without access to concurrent reports, as the system requires immediate imputation before data from other respondents are available. Exogenous data from a related survey are considered, and a number of different estimators are compared through an exploratory approach to determining an imputation model that is compatible with both processing requirements and data characteristics.

### Analyses of Measures of Respondent Burden at the National Agricultural Statistics Service

❖ Fatu Wesley, U.S. Department of Agriculture, 1400 Independence Ave., SW, Room 6344A South Building, Washington, DC 20250-0023, *fatu_wesley@nass.usda.gov*

***Key Words:*** burden, respondent, indicator, measure

The National Agricultural Statistics (NASS) uses joint burden indicators (JBIs) to measure the burden on respondents (especially those in certainty strata). The JBIs are computed annually, by using information for each survey. The JBIs are 1) the total number of surveys a respondent is asked to complete during the year, 2) the total number of times a respondent is contacted and 3) the total number of minutes the respondent spends completing a survey. Analyses were performed to compare the average JBIs computed for seven surveys in order to 1) compare surveys in regard to the burden imposed on respondents, 2) determine the level of burden on respondents in the certainty strata versus respondents in other strata, 3) determine if the indicators measure distinct aspects of respondent burden and 4) examine if any results can be used to decrease respondent burden.

### Multiple Imputation for Response Biases in NLAAS Due to Survey Instruments

❖ Jingchen Liu, Harvard University, 1 Oxford Street, Science Center 7th Floor, Cambridge, MA 02138, *jcliu@stat.harvard.edu*; Xiao-Li Meng, Harvard University; Chihnan Chen, Boston University; Margarita Alegria, Cambridge Health Alliance

Survey responses can be affected by the interviewee's previous experiences in the same or other similar surveys. This response bias is observed in the National Latino and Asian American Study (NLAAS), where there was a strong question ordering effect on the respondents' self-reported psychiatric service uses. Because NLAAS has a built-in randomization to two different orderings, one of which is considered to be (approximately) bias-free, we are able to use a Bayesian modeling approach to predict the unbiased responses. Multiple imputations are created from the corresponding posterior predictive distribution to facilitate common analyses to correct this bias. The modeling task, including the computation, turns out to be rather complicated due to the presence of high order interactions in the data. In this paper, we report our current solutions and directions for further improvements.

### Multiple Imputation Strategy for Alameda County Study

❖ Irina Bondarenko, University of Michigan, M4017 SPH II, 1420 Washington Hgts, Ann Arbor, MI 48109, *ibond@umich.edu*; Trivellore E. Raghunathan, University of Michigan

***Key Words:*** multiple imputation, longitudinal, imputations, imputation diagnostic, missing values

The Alameda County Study is a prospective study of a stratified random sample of members of Alameda County, California households, with data first collected in 1965, and five subsequent waves of data collection carried in 1974, 1983, 1995, and 1999. It is the longest ongoing, population-based study in the US that contains substantial amounts of information on health status and behavioral, social, socioeconomic, neighborhood, and socio-environmental risk factors. The study is subjected to numerous sources of missing data, including partial nonresponse, lost to follow-up, changes in questionnaires, and collection of data on a sub-sample in 1983, due to reduced funding. This paper explores multiple imputation strategy for a variety of longitudinal analyses. A number of diagnostics tests are used to check validity of the models used in imputations, and plausibility of completed datasets.

## 197 Challenges and Innovative Methods for Teaching Biostatistics in the Health Sciences ● ✪

Section on Teaching Statistics in the Health Sciences
**Monday, August 7, 2:00 pm–3:50 pm**

### The Implications of the Increasing Sophistication of Statistical Methods in The New England Journal of Medicine

❖ Suzanne Switzer, Smith College, Box 8416, 1 Chapin Way Smith College, Northampton, MA 01063, *sswitzer@email.smith.edu*; Nicholas J. Horton, Smith College

***Key Words:*** statistical education, medical training

A recent survey of Original Articles published in The New England Journal of Medicine revealed increasing use of statistical methods over time, compared with surveys conducted in 1979 and 1989 (Emerson and Colditz). In 311 articles published in 2004-2005, a substantial fraction of articles utilized relatively sophisticated statistical methodologies such as survival analysis (61%), multiple regression (51%) or power calculations (39%). Only 13% of articles used just simple descriptive statistics (e.g. percentages, means, confidence intervals). Knowledge of material typically included in an introductory statistics course increased this percentage to only 21%. We will describe the implications of this increased use of sophisticated statistical methods for medical and statistical education.

### How To Reduce the Risk of Nervous Breakdown in Physicians Taking an Introductory Biostatistics Course

❖ Patrick Arbogast, Vanderbilt University, S2323 Medical Center North, Nashville, TN 37232-2158, *patrick.arbogast@vanderbilt.edu*

***Key Words:*** education, teaching

The Master of Public Health (MPH) program at Vanderbilt University is a two-year program for physicians and other doctoral-level health care professionals. The objective of the program is to provide training for patient-oriented researchers who will conduct studies with large sample sizes. The MPH includes didactic course work and mentored research. Teaching biostatistics to non-statisticians with expectations of conducting their own data analyses poses challenges unique from other teaching settings. I teach the introductory biostatistics course in this program, and in this talk will present a course description, teaching techniques implemented, what was successful, what was not, and lessons learned. In addition, a retention tool of biostatistical knowledge developed in collaboration with a colleague and designed to aid us in future offerings of our courses, will be presented.

### Basic Biostats: Online Learning versus Onsite Learning

❖ John McGready, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205, *jmcgread@jhsph.edu*

***Key Words:*** statistical education, online learning

The technology explosion has allowed the internet to play an increasingly large role in the education of public health and medical professionals. For example, at Johns Hopkins it is now possible to receive a MPH degree with 80% of the credits for the degree done via the internet. For the past six years a two-quarter sequence, “Statistical Reasoning in Public Health” has been taught concurrently online and on campus. The same instructor has taught both sections for the past 3 years. This situation has provided an opportunity to study how the two sections compare both in terms of student demographics and class performance, and to try to help assess whether online learning is comparable to on-campus learning for basic biostatistics. We will present the results from a study designed for this purpose and conducted at Johns Hopkins in the period from September 2005 - January 2006.

### The Design and Evaluation of Computer-Assisted Instruction on Biostatistics: an Example of Central Limit Theorem

❖ Lai-Chu See, Chang Gung University, 259 Wen Hwa 1st Road, Kweisan, Taoyuan, 333 Taiwan, *lichu@mail.cgu.edu.tw*; Yu Hsuan Huang, Chang Gung Memorial Hospital; Yi Hua Chang, Chang Gung University; Pei-I Peng, Chang Gung University

***Key Words:*** central limit theorem, Flash animation, computer-assisted instruction, learning effect

The aims of this study were to design a computer-assisted instruction (CAI) of the topic of Central Limit Theorem and evaluate its learning effect. Class A students were given CAI first and followed by lecture; class B students were given lecture first and followed by CAI. Three quizzes were given before and after teaching, respectively. Two classes of graduate nursing students, 17 in class A and 32 in class B, participated in September 2005. On post-test1, class A (61.8% correction) did slightly worse than class B (65.0%) (p=.6301). Both classes showed a significant gain in the second score (18.2% for class A, 7.5% for class B). For within group comparison, only class B students showed significantly higher scores in CAI than those in lecture. 43.8% of class B students preferred lecture followed by CAI, and two-thirds of class A preferred CAI followed by lecture.

### Enhancing Medical Students’ Understanding of Risk Information Using a Large Group Interactive Audience Response System

❖ Fredric Wolf, University of Washington, E312 Health Sciences, Box 357240, Medical Educ Biomedical Informatics, Seattle, WA 98005-7240, *wolf@u.washington.edu*; David Masuda, University of Washington; Linda Pinsky, University of Washington

***Key Words:*** risk communication, decision making, teaching, framing, risk assessment

100 medical students used an Audience (or Group) Response System to indicate whether they would recommend (yes/no) a patient be treated based on the results from 5 clinical trials. The effectiveness of each trial in reducing death was presented in 5 mathematically equivalent formats, as relative risk reduction (RRR), absolute risk reduction (ARR), reduction in odds (RO), survival rate (SR), and number of patients needed to be treated (NNT) to prevent 1 additional bad outcome. Immediately graphically displaying the results indicated over 90% would recommend treating when results were presented as RRR or RO compared to < 50% for NNT, SR, and ARR. Students were surprised that these are 5 different ways to represent (frame) identical information. Providing immediate feedback helps learners appreciate risk information and the impact language and communication have in influencing decisions.

### Learner-Centered Approach to Biostatistics in Health Sciences Settings

❖ Ralph M. Turner, University of the Sciences in Philadelphia, Department of Social Sciences, 600 South 43rd Street, Philadelphia, PA 19104-4495, *r.turner@usip.edu*

***Key Words:*** teaching, biostatistics, health science, learner-centered

A learner-centered method was developed that engages students in statistics through journal articles on hot-topic research. Before class, students read reference materials on a targeted statistical analysis. In class, students read a research article that uses the targeted statistical analysis. Then students work in small groups using the assigned references to examine the analysis used in the study. The small groups work toward a consensus to answer questions about purpose, components, assumptions, critical tables, significance, and interpretation of the targeted analysis. Finally, the instructor mediates a large group discussion correcting misunderstandings and solidifying learning. Outcome data from four cohorts show the learner-centered procedure significantly increases final exam scores by 27% versus a traditional lecture delivered method after controlling for SAT scores.

## 198 Advances in Graphical Methods ●

Section on Statistical Graphics

**Monday, August 7, 2:00 pm–3:50 pm**

### Letter Value Box Plots: Box Plots for Large Datasets

❖ Karen Kafadar, University of Colorado, Department of Mathematics, P.O. Box 173364, CB 170, Denver, CO 80217-3364, *kk@math.cudenver.edu*; Heike Hofmann, Iowa State University; Hadley Wickham, Iowa State University

***Key Words:*** boxplot, distribution, shape

Conventional boxplots are useful displays for conveying rough information about the data distributional shape. Tail information beyond the quartiles (whiskers) can be unreliable in small samples so whiskers show only the extent of the outer quartiles. Thus, boxplots present two shortcomings for large data sets: more “outliers” are labeled and information about tail behavior is sparse. Letter value box plots address both shortcomings: a fixed, reasonably small, number of outliers can be labeled, and more detailed estimates of tail behavior based on letter values beyond the quartiles can be shown. We describe their construction and illustrate their usefulness on real data sets.

### Censored Q-Q Plot: Diagnostic Tool for Checking Population Heteroscedasticity

❖ Jong Kim, Portland State University, Department of Mathematics and Statistics, Portland, OR 97207, *jong@pdx.edu*

***Key Words:*** censored regression quantile, lifetime distribution, population heteroscedasticity, q-q plot, accelerated failure time model, proportional hazards model

Most textbooks in survival analysis discuss how to construct Q-Q plot. However, they do not explain in detail how to construct Q-Q plot for censored data. In this paper, I show how to construct a censored Q-Q plot in detail. Next, I display fitted lines on the Q-Q plot to show the effect of censoring as improperly dealing with censored data results in biased location and scale parameter estimates in log-location and scale family in lifetime distributions. Through a simulation study, I compare my method with others in terms of the ability to determine goodness-of-fit. Next, this method is extended to censored regression settings to detect population heteroscedasticity. Finally, examples are shown to illustrate the algorithm and compare the results with those in the literature.

### Variations on the Histogram

❖ Lorraine Denby, Avaya Labs Research, 233 Mt Airy Road, Basking Ridge, NJ 07920, *ld@research.avayalabs.com*; Colin Mallows, Avaya Labs Research

***Key Words:*** density estimation, bi-modal, outliers

It is usual to choose to make the bins in a histogram all have the same width. One could also choose to make them all have the same area. These two options have complementary strengths and weaknesses; the equal-width histogram oversmoothes in regions of high density, and is poor at identifying sharp peaks; the equal-area histogram oversmoothes in regions of low density, and so does not identify outliers. We describe a compromise approach which avoids both of these defects.

### Ideas about Forestry Data Visualization

❖ Lutong Zhou, University of Western Ontario, Department of Statistics and Actuarial Sciences, London, ON N6A 5B7 Canada, *lzhou8@uwo.ca*; Willard J. Braun, University of Western Ontario

***Key Words:*** graphics, forest fires, R

The grid package (P. Murrell, 2005) is a good tool for visualizing forestry data. The viewport, a main feature of grid, makes it possible to visually compare groups of plots arranged spatially. For instance, several viewports can be used to construct multiple caveplots (Becker et al,1994) for visualizing lightning and fire ignition data at different locations. They can also be used to modify pairwise scatterplots for exploring relationships among fire weather variables and danger ratings. The editing of grid objects (grobs) makes animating graphics in R easier. As an example, we introduce the Fengche plot for visualizing wind data. In this talk we argue that grid is easy enough to use that it should replace some traditional plotting tools.

### Of Forests, Trees, and Logs: Application and Visualization of Tree-Based Methods for Large Data

❖ Simon Urbanek, AT&T Labs-Research, Shannon Laboratory, Room D281, 180 Park Ave, Florham Park, NJ 07932, *urbanek@research.att.com*

***Key Words:*** large data, tree-based models, classification and regression trees

Tree-based methods became mainly popular due to their versatility and interpretability. However, some tree-based approaches are computationally intensive, which raises the question of their applicability to large data. Moreover, visualization and interpretability of tree models can be impaired by model complexity. In this paper we discuss the performance of tree models on large data, their inherent properties in such an application and present visualization techniques that help to enhance interpretability and extract information from both complex tree models and tree models for large data.

### Simple Visualizations of Paired Comparisons

❖ Spencer Graves, PDF Solutions, Inc., 333 W. San Carlos Street, Suite 700, San Jose, CA 95110, *spencer.graves@pdf.com*; Hans-Peter Piepho, University of Hohenheim

***Key Words:*** multiple comparisons

This presentation describes and compares two alternative methods for converting a table of k*(k-1)/2 p-values comparing k different items (e.g., levels of a factor) into a simple visual summary of the significant differences. One of these methods is a “letter-based representation” previously described by Piepho (2004). The other is a simple graphic identifying the “neighbors” that are not significantly different from each of the k items.

## 199 Longitudinal Models ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Monday, August 7, 2:00 pm–3:50 pm**

### Modeling Individual Addition Behavior Using a Mixed-Effect Model with Three States

❖ Sheng Luo, Johns Hopkins University, 6120 Executive Blvd., Room 8105, Bethesda, MD 20892, *sluo@jhsph.edu*; Ciprian M. Crainiceanu, Johns Hopkins University; Thomas A. Louis, Johns Hopkins University; Nilanjan Chatterjee, National Cancer Institute

***Key Words:*** three-state process, cure model, mixture model, tri-variate random effect

Similar to other behaviors of substance abuse and dependence, personal smoking behavior alternates between smoking and non-smoking periods. A successful quit attempt involves a “cured state” from which a subject never relapses to smoking. To model smoking cessation patterns longitudinally, we develop a mixed-effect Markov models with three states: smoking, transient quitting and cured (absorbing state). Subject level random effects are used to capture between subject heterogeneity. We illustrate this methodology using a large (29,133 subjects) longitudinal cohort lung cancer prevention study and discuss our findings.

## Quantile Regression Methods for Modeling CD4 T-Cell Trajectory among HIV-Infected Men and Women on Long Term, Highly Active Antiretroviral Therapy

❖ Haitao Chu, Johns Hopkins Bloomberg School of Public Health, N 615 Wolfe Street, RoomE7636, Baltimore, MD 21205, *hchu@jhsph.edu*; Ying Wei, Columbia University; Alvaro Munoz, Johns Hopkins Bloomberg School of Public Health; Stephen J. Gange, Johns Hopkins Bloomberg School of Public Health

***Key Words:*** conditional quantile regression methods, longitudinal data analysis, multicenter aids cohort study, Women's Interagency HIV Study, CD4 cell count

Estimating CD4 cell count trajectories for HIV-infected subjects has traditionally relied on linear/nonlinear random effect models or generalized estimating equation approaches. The construction of predictive reference curves for an individual, which has important clinical significance, has often relied on normal theory. In the analysis of longitudinal data with unequally spaced measurements, non-parametric conditional quantile regression methods (QR) offer a complementary strategy for estimating conditional quantile functions based on prior history and other covariates. The authors provided a case study using QR to estimate CD4 cell count trajectories over a 7-year period following HAART initiation for 404 HIV-infected men in the Multicenter AIDS Cohort Study and 609 HIV-infected women in the Women's Interagency HIV Study. The advantages of QR are illustrated.

## On the Equivalence of Case-Crossover and Time-Series Methods

❖ Yun Lu, Johns Hopkins University, 734 Concord Point Drive, Perryville, MD 21903, *ylu@jhsph.edu*; Scott Zeger, Johns Hopkins University

***Key Words:*** time series, case-crossover design, log-linear model, Poisson regression, environmental epidemiology, air pollution

Time series and case-crossover methods are often viewed as competing alternatives in environmental epidemiologic studies. Several recent studies have compared the time series and case-crossover methods. In this paper, we show that case-crossover using conditional logistic regression is a special case of time series analysis when there is a common exposure such as in air pollution studies. This equivalence provides computational convenience for case-crossover analyses and a better understanding of time series models. Time series log-linear regression accounts for over-dispersion of the Poisson variance, while case-crossover analyses typically do not. This equivalence also permits model checking for case-crossover data using standard log-linear model diagnostics.

## Analysis of Longitudinal Trinomial Outcome through a Surrogate Variable

❖ Wenyaw Chan, The University of Texas School of Public Health, 1200 Herman Pressler, Houston, TX 77030, *wenyaw.chan@uth.tmc.edu*; Yen-Peng Li, The University of Texas Health Science Center at Houston; Hung-Wen Yeh, The University of Texas School of Public Health

***Key Words:*** continuous-time Markov chain, EM algorithm, hidden Markov model, longitudinal studies, surrogate variable

Researchers in public health studies often encounter the situation that the outcome variable can not be measured properly due to confidentiality or limitation. In this circumstance, a surrogate variable that is accessible and can gauge the character of the targeted outcome is often adopted to carry on the study. In this research, a longitudinal surrogate outcome will be used for statistical analysis in which the targeted outcome is assumed to follow a trinomial distribution. We propose to use the EM algorithm and a continuous-time hidden Markov model to analyze a longitudinal multinomial outcome that has three categories and is accessible only through a surrogate variable. The accuracy of the point estimation and the sensitivity of the computational algorithm will be examined by an empirical study. An example in public health will be demonstrated.

## The Log Multinomial Regression Model for Nominal Outcomes with More Than Two Attributes

❖ Leigh Blizzard, Menzies Research Institute, University of Tasmania, Private Bag 23, Hobart Tasmania, 7001 Australia, *Leigh.Blizzard@utas.edu.au*; David W. Hosmer, University of Massachusetts

***Key Words:*** multinomial regression, log link, relative risk, logistic regression, Poisson regression

In principle, the log binomial model can be modified to obtain relative risk estimates for nominal outcomes with more than two attributes. We examine some technical issues that arise when one attempts to fit to multinomial data a model with log link (the "log multinomial model"). In extensive simulations, its performance is compared with that of an alternative method based on the approach proposed by Schouten et al. (1993) for binary data, and with separate fits of a Poisson regression model. Log multinomial regression resulted in "failure" rates (nonconvergence, out-of-bounds probabilities) exceeding 50% in some data settings. The other methods produced out-of-bounds probabilities for data successfully fitted by a log multinomial model. Log multinomial coefficient estimates had least relative bias and mean squared error, but with greater-than-nominal confidence interval coverage.

## Bayesian Hierarchical Models for Racial and Socioeconomic Predictors of Mortality in a Sample of the U.S. Medicare Population

❖ Yijie Zhou, Johns Hopkins University, 615 N. Wolfe Street, Room E3038, Baltimore, MD 21205, *yijzhou@jhsph.edu*; Francesca Dominici, Johns Hopkins University; Thomas A. Louis, Johns Hopkins University

***Key Words:*** race, socioeconomic status, mortality, multi-level model, multiple imputation

In this project, we investigate the association between individual race and risk of death, as well as whether this association is affected by both individual-level and zip code-level socioeconomic status (SES). We assemble a data set that links at individual level and zip code level three government databases: Medicare, US census and Medicare Current Beneficiary Survey (MCBS), for 2095 zip codes in the Northeast region of the United States. We develop and apply multi-level statistical models to this data set, and estimate population-level attributable risk, relative attributable risk and odds ratio of death comparing the black versus the white population. In addition, we develop multiple imputation methods for combining information across the Medicare, U.S. census and MCBS databases to impute missing individual-level SES data.

## Bayesian Estimation for Epidemic Models on a Social Network

❖ Crystal Linkletter, Simon Fraser University, Department Statistics and Actuarial Science, 8888 University Dr, Burnaby, BC V5A

1S6 Canada, *cdlinkle@stat.sfu.ca*; Randy R. Sitter, Simon Fraser University; Nicolas Hengartner, Los Alamos National Laboratory

***Key Words:*** Bayesian hierarchical models, social networks, epidemics

Two well-known models for the spread of epidemics are the SIS (susceptible-infected-susceptible) and SIR (susceptible-infected-removed) models. We consider the problem of fitting these models to epidemic data under the assumption that there is an underlying social network that governs contact between individuals. Modeling transitions in the health-status of such a population over time is challenging due to the large number of states and the interaction between individuals. If observations can be made on individuals over time, early explorations suggest that a Bayesian framework seems promising for estimating transmission parameters, as well as features of the social network.

# 200 Choice Experiments in Marketing

Section on Statistics and Marketing

**Monday, August 7, 2:00 pm–3:50 pm**

## Managing Large Conjoint Studies

❖ Ulderico Santarelli, Consultant, Via Bellinzona 32, MILAN, 20155 Italy, *u.santarelli@tin.it*

***Key Words:*** conjoint, full profile, nested designs, respondent's effort

It is seldom feasible in marketing practice to constrain real problems into standard conjoint tools. Real choice experiments often lead to the intermix of product concepts. Problems become large, design difficult, and respondent's workload so huge that doubt remains about response validity. The author presents a new methodology based on a revision of random utility theory, which gains much in respondent's effort. Large nested designs can be managed with an affordable effort, keeping full orthogonality of the design within nesting branches and full profile data collection mode.

## Designing a Stated Choice Survey To Study Food Product Eco-Labels

❖ Iain Pardoe, University of Oregon, 1208 University of Oregon, Eugene, OR 97405, *ipardoe@lcbmail.uoregon.edu*

***Key Words:*** agricultural economics, discrete choice, environment, marketing, multinomial logit, sustainability

Eco-labels were first introduced to the U.S. in the early 1970s to certify organic fruits and vegetables. These labels have evolved since then to represent various aspects of sustainable agricultural practices, environmental health, wildlife preservation, etc. This article describes the design of a large-scale stated choice survey to investigate consumer reactions to eco-label characteristics such as price premiums, food quality, organic certification, and production location, for a variety of common food products. This survey is part of a multi-year, multi-site project to study demand impacts and implementation of eco-labels funded by the U.S. Department of Agriculture.

## Fusing Best/Worst Choices and Ratings Data for Comparisons on a Common Scale

❖ Lynd D. Bacon, Sighthound Solutions, Inc., 1840 Gateway Drive, Suite 200, San Mateo, CA 99404-4029, *lbacon@sighthoundsolutions.com*; Peter J. Lenk, University of Michigan; Katya Seryakova, Knowledge Networks, Inc.

***Key Words:*** best/worst analysis, hierarchical Bayes, preference/importance measurement, discrete choice, data fusion, consumer behavior

Marketing researchers are using best/worst scaling (BW), a choice-based procedure, instead of rating scales, for measuring item importance or appeal. Ratings often do not discriminate between items well. BW produces differences because it forces choices. When BW choice data are analyzed using hierarchical models, respondent-level scale values can be obtained. The problem with these results is that overall differences between respondents are lost. Two respondents can have identical estimated scores for items when one likes all the items and the other does not. We describe a model that combines ratings and choice data that preserves scale location while retaining the discriminatory power of choice data. As a result, respondent score differences can be interpreted as location differences on a common scale.

## The Impact of Choice Set Complexity on Respondent Screening Behavior

❖ Joseph Retzer, Maritz Research, 2019 E. River Road, Grafton, WI 53024, *jjrr@wi.rr.com*

***Key Words:*** non-compensatory, choice based conjoint, hierarchical Bayes, marketing research

Increasing the number of alternatives in a choice-based conjoint (CBC) exercise provides more information per choice set. However, it also increases the risk of respondent attribute--level screening. The screening behavior of respondents performing a CBC exercise with a varying number of alternatives is examined using a noncompensatory hierarchal Bayes model as detailed in Gilbride and Allenby's A Choice Model with Conjunctive, Disjunctive, and Compensatory Sreening Rules. While compensatory analysis assumes alternatives are compared based on total utility, this model allows for consideration set formation using a set of alternatives that survive a screening process (noncompensatory analysis) when process measures or other data (e.g., acceptable alternatives) are not available. This approach offers superior predictive results and useful behavioral insights.

## Misspecification and Decision Strategies in Choice Design Parameter Assumptions

❖ Jennifer Golek, DuPont, 3401 N. Rockfield Drive, Wilmingon, DE 19810, *jgolek@gmail.com*; Robert W. Mee, University of Tennessee

***Key Words:*** stated preference experiment, discrete choice experiment, efficiency, entropy, non-compensatory decision strategy

We study the impact of misspecification of the assumed parameter values used in creation of optimal choice designs. We find that the impact of misspecification varies widely based on the discrepancy between the true and assumed parameter values. Further, we find that entropy becomes a more feasible secondary measure of design optimality if one considers the potential of misspecification of the values. Current design and analysis strategies for stated preference experiments assume that compensatory decisions are made. We consider how different decision strategies may be represented through manipulating the assumed parameter values used in creating the choice designs. In this context, the consequences of misspecification of the decision strategy are also evaluated.

### Thurstone Scaling via Maximum Likelihood in Order Statistics

❖ Stan Lipovetsky, GfK-CRI, 8401 Golden Valley Road, Minneapolis, MN 55427, *lipovetsky@Gfkcustomresearch.com*

***Key Words:*** Thurstone scale, order statistics, paired comparisons, ranks, maximum likelihood, marketing research

Thurstone scaling is widely used for finding priorities among the compared items. This work considers maximum likelihood objective constructed by the probabilities of the order statistics applied to the ranked or paired comparison data. Probit, logit, and multinomial links are tried to obtain the Thurstonian scale and to estimate probabilities of the items' choice. This approach is convenient and can substantially enrich both theoretical interpretation and practical application of Thurstone modeling, particularly, in marketing and advertising research.

### Partial Profile Choice Experiments: an Assessment

❖ Terry Elrod, University of Alberta, 323 School of Business, University of Alberta, Edmonton, AB T6G2R6 Canada, *terry.elrod@ualberta.ca*

***Key Words:*** consumer research, choice experiments, conjoint analysis, partial profile

Partial profile choice experiments are emerging as an important means for estimating conjoint part worth utilities. These experiments constrain choices in each choice set to differ on a subset of all. This paper compares the efficiency of partial profile and full profile designs in four different studies. Partial profile designs have been avoided in the past because they appear to be less efficient. However, conventional application of efficiency criteria such D-efficiency ignores that consumers can make more choices and make those choices more accurately when alternatives differ on only a few attributes in each choice set. A more proper application of D-efficiency shows that partial profile choice experiments provide much more efficient estimates of part worths.

## 201 Statistical Methods and Ecological Applications ●

Section on Statistics and the Environment

**Monday, August 7, 2:00 pm–3:50 pm**

### Choice of Weights in Meta-Analysis of Mark-Recapture Survival Studies

❖ James R. Faulkner, U.S. Department of Commerce, NOAA Fisheries, 2725 Montlake Blvd E., Seattle, WA 98112, *jim.faulkner@noaa.gov*; Steven G. Smith, U.S. Department of Commerce

***Key Words:*** weighting, mark-recapture, survival, meta-analysis

Meta-analyses often use weighting to account for unequal precision of parameter estimates. The method of weighting estimates by the inverse of the estimated sampling variance is often used in meta-analyses of survival estimates from capture-recapture studies (e.g., Jolly-Seber estimates). Weighting by inverse variance may give biased results, because Jolly-Seber survival estimates are correlated with their estimated sampling variances. We propose alternative weighting methods, including the inverse of the squared coefficient of variation. Using Monte Carlo simulation, we investigate the effect of weighting method on the performance of resulting estimators under scenarios involving stationary and non-stationary processes. The choice of most effective method depends on multiple factors, including analysis objective, degree of stationarity of the survival estimates, and sampling effort.

### Trend Estimation in a Periodic Survey of North-American Waterfowl

❖ Mark Otto, U.S. Fish and Wildlife Service, 11510 American Holly Drive, Merriam Building, Laurel, MD 20708-4002, *mark_otto@fws.gov*

***Key Words:*** REGCMPNT models, ARIMA, time series, signal extraction

We estimate trends in North American duck and goose populations from May Breeding Waterfowl and Habitat Survey estimates. The MBWHS is an aerial transect survey systematically placed transects in each of 70 strata encompassing the U.S. Dakotas and Montana, Canada, up to the Northwest Territories and tundra in Alaska. Transects are sampled each year. The trends in the ``true values” are estimated using the level of the first differenced series or the linear slope in the series itself. The “true values’” are the observed survey estimates apart from the survey sampling error. The survey sampling error is defined from sampling error covariances. Usually periodic surveys only account for sample variance at each time point. The change between time period estimates can be improved by accounting the sampling error covariances among time points.

### Bayesian Spatio-Temporal Models for Radio-Telemetry Contacts

❖ Albert N. Hendrix, R2 Resource Consultants, Inc., 15250 NE 95th Street, Redmond, WA 98052, *nhendrix@r2usa.com*; Rip Shively, U.S. Geological Survey; Barbara Adams, U.S. Geological Survey

***Key Words:*** MCMC, WinBUGS, spatio-temporal, fish

Radio-tagged animals are contacted in different spatial locations over time to evaluate their spatial distribution. If the study area is divided into spatial cells, the number of contacts within a cell may be modeled as a Poisson random variable. The mean of the Poisson process can be modeled as a function of environmental covariates, a temporal component, and a spatial component. The spatial component can be specified as a conditional autoregressive process, and the temporal component can be modeled in several forms: fixed time effects, time varying coefficients, or an autoregressive process. These models were evaluated in a Bayesian framework using Markov chain Monte Carlo implemented in winBUGS with simulated data. Subsequently, these models will be applied to the contacts of radio-tagged endangered suckers (fish) in Upper Klamath Lake, Oregon.

### Evaluating Sampling Approaches for Monitoring Chronic Wasting Disease (CWD) in Deer Populations

❖ Ling Huang, Iowa State University, 104 Snedecor Hall, Department of Statistics, Ames, IA 50011-1210, *lingh@iastate.edu*; Sarah M. Nusser, Iowa State University; William R. Clark, Iowa State University; David L. Otis, Iowa State University

***Key Words:*** wildlife populations, disease monitoring, sampling design

Chronic Wasting Disease (CWD) is a disease of considerable concern to wildlife scientists. Estimating the probability of detecting the presence of CWD and estimating the prevalence rate of CWD are important for conducting surveillance research. Our goal is to evaluate the properties of sample designs for monitoring CWD in deer using a realistic landscape setting and to explore the properties of estimates generated from convenience samples in relation to probability sample designs. Deer are simulated on the landscape using an algorithm that distributes deer proportional to the expected density in each habitat stratum. A model is built to predict detection probability in a year and waiting time to first detection, to evaluate properties of CWD prevalence estimates, to compare bias and precision of estimates obtained from convenience sampling designs to the probability sampling designs.

### The Federally Threatened Plant Species, Missouri Bladderpod: Model Selection and Validation

❖ William Leeds, Truman State University, 100 Normal, Kirksville, MO 63501, *hjkim@truman.edu*; Hyun-Joo Kim, Truman State University; Corey Elledge, Truman State University; Michael Kelrick, Truman State University; Elizabeth Bobzien, Truman State University; James Franklin, Truman State University

***Key Words:*** Missouri bladderpod, logistic model, model selection criteria, model selection uncertainty, model validation, environmental statistics

Missouri bladderpod (Lesquerella filiformis) is a federally threatened plant species found in southwest Missouri and northern Arkansas. Understanding how Missouri bladderpod's abundance is related to its habitat is crucial for its conservation. Abundance of Missouri bladderpod and percent cover of several habitat attributes were recorded in Springfield, MO in 1997, 1998, 2003 and 2005. Logistic regression models were created to describe the relationship between presence of L. filiformis and percent cover of the habitat attributes. We are developing optimal models using both conventional model-selection criteria and Akaike weights based on model-selection uncertainty. Once these models are created, we will assess their performances using a data set from an independent site.

### Effects of Sample Survey Design on the Accuracy of Classification Tree Models in Ecology

❖ Thomas C. Edwards, U.S. Geological Survey, 5230 Old Main Hill, Utah State University, Logan, UT 84322-5230, *tce@nr.usu.edu*; D. Richard Cutler, Utah State University; Gretchen Moisen, U.S. Forest Service; Niklaus E. Zimmermann, Swiss Federal Research Institute WSL; Linda Geiser, U.S. Department of Agriculture

***Key Words:*** sample survey, classification trees, accuracy assessment, probability samples, non-probability sample

We evaluated the effects of probabilistic (DESIGN) and non-probabilistic (PURPOSIVE) sample surveys on classification trees models for predicting the presence of four lichen species in the Pacific Northwest, USA. Models derived from both survey forms were assessed using an independent data set (EVALUATION). Accuracies as gauged by resubstitution were similar for each lichen species irrespective of the underlying sample survey form. Cross-validation accuracies were lower than resubstitution for all species and both designs, and in all cases were closer to the true prediction accuracies based on the EVALUATION data. The classification tree structures also differed both among and within the species, depending on the sample survey form. We argue that greater emphasis should be placed on reporting cross-validation accuracies rather than resubstitution accuracies.

### Spatio-Temporal Analysis Incorporating a Spatial Correlation Structure on a Long-Term Forestry Field Research Dataset

❖ Bronson Bullock, North Carolina State University, 3102 Jordan Hall, Department of Forestry & Environmental Research, Campus Box 8008, Raleigh, NC 27695-8008, *bronson_bullock@ncsu.edu*; Edward Boone, The University of North Carolina at Wilmington

***Key Words:*** loblolly pine, Pinus taeda, repeated measures, dynamic linear model

The use of spatio-temporal data sets is growing in environmental research fields. These data sets are often plagued with both spatial and temporal correlation. Often this data is collected at regular intervals and at predefined locations. A spatio-temporal model is presented using a Dynamic Linear Model with a parametric spatial correlation structure. In this paradigm the spatial correlation structure can be placed either on the mean or on the observation errors. Examples of data generated by each type of process are considered. This paradigm is applied to a loblolly pine tree (Pinus taeda) data set from a long-term forestry field research site.

## 202 Biostatistical Modeling ●

Biometrics Section, Section on Bayesian Statistical Science, ENAR

**Monday, August 7, 2:00 pm–3:50 pm**

### Inter-Rater Reliability of Pressure Ulcer Staging: Ordinal Probit Bayesian Hierarchical Model That Allows for Uncertain Rater Response

❖ Byron Gajewski, The University of Kansas Medical Center, 3901 Rainbow Blvd., MS 4043, Kansas City, KS 66160, *bgajewski@kumc.edu*; Sara Hart, The University of Kansas Medical Center; Sandra Bergquist-Beringer, The University of Kansas Medical Center; Nancy Dunton, The University of Kansas Medical Center

***Key Words:*** Markov chain Monte Carlo, random effects model, intraclass correlation coefficient, NDNQI

This article describes a method for estimating the inter-rater reliability of pressure ulcer staging from raters in National Database of Nursing Quality Indicators participating hospitals. The method models ordinal spanning data using an ordinal probit Bayesian hierarchical model across several hospitals in which raters monitor patient pressure ulcers. This novel approach allows for an "unstageable" pressure ulcer rating. We compare the ordinal probit Bayesian hierarchical model to an approximate random effects model that assumes the raw ordinal data are continuous.

### A Random-Effects Four-Part Model for Longitudinal Medical Costs

❖ Lei Liu, University of Virginia, Division of Biostatistics, 3181 Hospital, W., PO Box 800717, Charlottesville, VA 22908, *liulei@virginia.edu*; Mark R. Conaway, University of Virginia

***Key Words:*** generalized linear mixed model, hierarchical linear model, four-part model, Laplace approximation, random effects, longitudinal

We extend the four-part model (Duan et al. 1983) in the cross-sectional medical cost setting to longitudinal medical cost setting (e.g., monthly medical costs) with correlated random effects. Our model characterizes four major features of longitudinal medical cost data: a fair amount of monthly medical costs are zero; there is a remarkable difference between inpatient and outpatient monthly costs; monthly medical costs are correlated for the same subject; and the responses in the four-part model also are correlated. We introduce the random-effects four-part model and propose an estimation method by the high-order Laplace approximation technique in Raudenbush et al. (1998) and Olsen and Schafer (2001). Our method is applied to monthly medical costs for heart failure patients from clinical data repository at the University of Virginia Health System.

## Bayesian Analysis of Repeated Data with Many Zeros: Application to the Longitudinal Adolescent Substance Abuse Study

❖ Hyonggin An, The University of Iowa, 200 Hawkins Drive, C22B GH, Iowa City, IA 52242, *hyonggin-an@uiowa.edu*

***Key Words:*** zero inflated Poisson, Bayesian analysis, random effects, two-parts model

Repeated data with too many zeros are quite common in many fields of study. For these data, Tooze et. al. (2002) proposed a two-parts model with random effect for continuous outcomes while Hall (2000) proposed a zero-inflated Poisson regression with random effects for counting outcomes. The parameters in these models were estimated by maximum likelihood. In this paper, as alternatives to maximum likelihood approaches, we propose Bayesian estimation methods for both the two-part model and the zero-inflated Poisson regression with random effects. The Bayesian methods are more flexible and can be easily implemented in the existing software, R and WinBUGS. The results from real data analysis comparing the standard methods and the proposed methods in the longitudinal adolescent substance abuse study are also provided.

## Repeated Measures Mixture Modeling with Applications to Postmortem Tissue Studies in Schizophrenia

❖ Zhuoxin Sun, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, *zhuoxin@jimmy.harvard.edu*; Ori Rosen, The University of Texas at El Paso; Allan R. Sampson, University of Pittsburgh

***Key Words:*** MCMC, mixture models, repeated measures, schizophrenia

In some neurological postmortem brain tissue studies, repeated measures are observed. These observations are taken on the same subject and are correlated within the subject. Furthermore, each observation can be viewed as coming from one of a pre-specified number of populations where each population corresponds to a possible type of neuron. A mixture model to model such repeated data will be presented. In this model, the mixture components are linear regressions, and the component-indicator variables are logits with random effects. The within-subject observations are taken to be correlated through the component indicator random variables. Inference is facilitated by sampling from the posterior distribution of the parameters via MCMC methods. The model is applied to a postmortem brain tissue study to examine the differences in neuron volumes between schizophrenic and control subjects.

## A Hidden Markov Model To Describe Responses to Alcoholism Treatment

❖ Kenneth Shirley, University of Pennsylvania, The Wharton School, , *kshirley@wharton.upenn.edu*; Dylan S. Small, University of Pennsylvania

***Key Words:*** hidden Markov models, alcoholism treatment clinical trial, longitudinal data, relapsing-remitting diseases

To determine whether a drug is effective in treating alcoholism, a subject's drinking behavior must be monitored over a period of time. To capture complex features of a subject's drinking behavior, like the difference between a steady, moderate drinker and one who alternates between abstinence and binges, Markov chain models are useful. We extend Albert's Markov model for chronic diseases (Albert, 1994) to a hidden Markov model (HMM), whose hidden states represent "in-control" vs. "out-of-control" drinking. The HMM allows us to describe drinking behavior flexibly; instead of defining a subject's state at a given time strictly based on his or her observed data, the HMM essentially smoothes the observed data and lets us estimate a probability of being in each hidden state at a given time. We present simulation results, as well as an application to real data.

## Mixture Gaussian Model-Based Bayesian Clustering

❖ Wei Zhang, Harvard University, 1 Oxford Street, Statistics Dept, Cambridge, MA 02138, *weizhang@fas.harvard.edu*

***Key Words:*** mixture Gaussian model, Bayesian clustering, generalized Gibbs sampler

Identifying patterns of co-expression in DNA microarray by cluster analysis is an important task in understanding the underlying biological processes. In this presentation, we will introduce a mixture Gaussian model-based Bayesian clustering method. The model gives freedoms of drift and scale effects and allows different S/N for different clusters. Due to the high dependency structure between the parameters, some advanced techniques of the gibbs sampler were used to decrease the autocorrelation of the sampled parameters. Some simulation studies show that this model-based Bayesian clustering method is more robust and flexible than the traditional k-means clustering.

## Modeling Distortion Product Otoacoustic Emissions Using Noncentral-F Mixed Effects Models

❖ Lai Wei, The Ohio State University, 602 Stark Court, Columbus, OH 43210, *weil@stat.ohio-state.edu*; Peter F. Craigmile, The Ohio State University; Wayne M. King, The Ohio State University; Stephanie Jones, The Ohio State University

***Key Words:*** tests for periodicity, maximum likelihood, quasi-likelihood, hearing tests

Distortion product otoacoustic emissions (DPOAEs) are sinusoids produced by the nonlinearities in the healthy cochlea. In a clinical environment, they are an important nonbehavioral measure of cochlear function because they are a close analog of the behavioral pure-tone audiogram. The standard method of the detection of DPOAEs is via the spectral analysis of sound signals recorded from the ear. In this talk, the noncentral-F mixed effects regression models are explored to capture the important features of DPOAE data. In particular, these models capture within-subject-variability of DPOAEs of healthy hearing subjects. This is key to understanding the underlying processes inherent in DPOAE-based hearing tests. The penalized quasi-likelihood method is used to estimate the model parameters, and we demonstrate how to do the model selection and diagnostic.

# 203 Methodology for Survival and Censored Data ●

Biometrics Section, WNAR, ENAR

**Monday, August 7, 2:00 pm–3:50 pm**

## Constructing Multivariate Prognostic Expression Profiles for Survival Endpoints

❖ Derick R. Peterson, University of Rochester, 601 Elmwood Ave., Box 630, Dept Biostatistics and Comp Bio, Rochester, NY 14642, *peterson@bst.rochester.edu*; Alexander Pearson, University of Rochester

*Key Words:* model selection, shrinkage estimation, survival analysis, high-dimensional data, microarray

Modern high-throughput technologies allow us to simultaneously measure the expressions of a huge number of candidate predictors, some of which are likely to be associated with survival. One difficult task is to search among an enormous number of potential predictors and to correctly identify most of the important ones, without mistakenly identifying too many spurious associations. Mere variable selection is insufficient, however, for the information from the multiple predictors must be intelligently combined and calibrated to form the final composite predictor. Although it is impossible to simultaneously adjust for a huge number of predictors in an unconstrained way, we propose a method that offers a middle ground where some partial adjustments can be made in an adaptive fashion, regardless of the number of candidate predictors. We demonstrate the performance in a simulation study.

## Predict Survival using Gene Expression Data under Cox PH Models

❖ Wenqing He, University of Western Ontario, 1151 Richmond Street, N., London, ON N6A5B7 Canada, *whe@stats.uwo.ca*; Grace Y. Yi, University of Waterloo

*Key Words:* gene expression, Cox ph models, ROC

There is extensive research on prediction of various clinical phenotypes using gene expression profiles. Relatively less attention has been paid to study the connection of gene expressions to time to event of patients. One reason is that the number of genes is much larger than the number of subjects. We develop methods for selecting informative genes that are event related, and based on those genes, to build prediction models for the time to event probability. The time to event model is postulated using Cox proportional hazards models, and the accuracy of prediction is assessed by means of the ROC method. The proposed methods differ from the existing methods in the sense that our selected gene profiles have clear biological meaning. Simulation studies show that the proposed methods perform well under various conditions. A real microarray data set is analyzed using the proposed methods.

## Model Combining in Survival Analysis

❖ Lihua Chen, The University of Toledo, 2051 Brookdale Road, Toledo, OH 43606, *Lihua.Chen@utoledo.edu*

*Key Words:* model combining, model uncertainty, survival analysis

Model uncertainty is inherent in survival analysis. We present a model combining method, Adaptive Regression by Mixing (ARM) to convexly combine survival models to account for model uncertainty. The advantage of ARM compared to other model selection and combining methods is illustrated through simulation and a study of chemotherapy for lung cancers. A theoretical result on risk bound of ARM is also presented.

## Median Regression Analysis from Doubly Censored Data

❖ Sundar Subramanian, University of Maine, 5752 Neville Hall, Department of Mathematics and Statistics, Orono, ME 04469-5752, *subraman@math.umaine.edu*

*Key Words:* curse of dimensionality, censoring distribution, minimum dispersion statistic, product integral, Volterra integral equation

Median regression models provide a robust alternative to regression based on the mean. We propose a methodology for fitting a median regression model from data with both left- and right-censored observations in which the left-censoring variable is always observed. First, we set up an adjusted least-absolute-deviation-estimating function using the inverse-censoring weighted approach, whose solution specifies the estimator. We then describe the inference procedure for the regression parameter and determine the efficacy of the proposed procedure through simulation.

## Smoothed Bootstrap-Based Bandwidth Estimation

❖ Derek Bean, University of Maine, Department of Mathematics and Statistics, 5952 Neville Hall, Orono, ME 04469-5752, *derek.bean@umit.maine.edu*; Sundar Subramanian, University of Maine

*Key Words:* cross validation, Epanechnikov kernel, integrated mean squared error, missing censoring indicators, numerical integration

Survival function estimators in the so-called missing censoring model of random right censorship require (kernel) estimates of certain conditional quantities. Therefore, for computing the survival function estimators, the user would need to supply the bandwidth, a key parameter for kernel estimators. We employ the smoothed bootstrap approach from kernel density estimation, proposed by Faraway and Jhun, to produce a data-based bandwidth for computing the kernel estimators. We present simulation results.

## Equivalences of Nonparametric Estimators and Noninformative Censoring Conditions

❖ Yingfu Li, University of Houston-Clear Lake, 2700 Bay Area Blvd., Houston, TX 77058, *li@uhcl.edu*; Jiantian Wang, Kean University

*Key Words:* equivalence, nonparametric estimators, noninformative censoring, survival function

In this talk, we will show the equivalence of several nonparametric estimators of survival function as well as the equivalence of several noninformative censoring conditions proposed in the literature. Furthermore, we prove such noninformative censoring condition can be expressed as a simple structure of independent variables.

# 204 Mixed Models and Data Mining

Section on Statistical Computing

**Monday, August 7, 2:00 pm–3:50 pm**

## A Comparison of the Reporting of Problems Encountered in the Estimation of Covariance Parameters in Linear Mixed Models Using SAS, SPSS, R, Stata, and HLM

❖ Brady West, University of Michigan, 3554 Rackham Building, Ann Arbor, MI 48109, *bwest@umich.edu*; Kathy Welch, University of Michigan; Andrzej Galecki, University of Michigan

***Key Words:*** linear mixed models, statistical software, covariance parameters, estimation

Computational problems with maximum likelihood estimation of covariance parameters in linear mixed models frequently arise in practice. In some cases, estimates of covariance parameters converge to a solution close to the boundary of the parameter space, or fall outside it. As a result, the variance-covariance matrix for the random effects in the model violates (or nearly violates) positive definiteness constraints. We fit a linear mixed model to a real-world longitudinal dataset in which this problem is encountered and compare the way it is reported when using the mixed modeling procedures in SAS, SPSS, R, Stata, and HLM. We compare the estimates derived using these packages and present practical alternatives for dealing with the problem. This work can be found in Linear Mixed Models: a Practical Guide Using Statistical Software.

## A Note on Testing of Hypothesis of Kroneckar Product Covariance Structure in Doubly Multivariate Data

❖ Anuradha Roy, The University of Texas at San Antonio, Department of Management Science and Statistics, 6900 N Loop 1604 West, San Antonio, TX 78249, *aroy@utsa.edu*

***Key Words:*** covariance structure, Kroneckar product, maximum likelihood estimates, mixed effects model, PROC mixed

The decision of whether to use Kronecker product variance covariance matrix over an unstructured variance covariance matrix is an important issue in doubly multivariate data in small sample set up. In this paper, we talk about statistical tests pertaining to Kronecker product variance covariance matrices. The tests can be tested by the MIXED procedure of SAS, in two steps. As PROC MIXED has the potential problem of convergence of the maximum likelihood estimates, algorithms are discussed to calculate the test statistics using the IML procedure of SAS. We demonstrate the tests with four real datasets and compare the results with that of PROC MIXED. Simulation studies are conducted to examine the needed sample sizes. A comparison study is presented when one of the matrices in the Kronecker product is either an unstructured or a compound symmetric or an autoregressive of order one.

## On Hierarchical Linear Mixed Modeling Using the Multivariate t Distribution with Missing Information

❖ Tsung-I Lin, National Chung Hsing University, Department of Applied Mathematics, Taichung, 402 Taiwan, *tilin@amath.nchu.edu.tw*; Jack C. Lee, National Chiao Tung University

***Key Words:*** autoregressive process, Bayesian prediction, Markov chain Monte Carlo, missing values, random effects, outliers

We present a fully Bayesian approach to modeling incomplete longitudinal data using the t linear mixed model with AR(p) dependence. Markov chain Monte Carlo techniques are implemented for computing posterior distributions of parameters. To facilitate the computation, two types of auxiliary indicator matrices are incorporated into the model. Meanwhile, the constraints on the parameter space arising from the stationarity conditions for the autoregressive parameters are handled by a reparametrization scheme. Bayesian predictive inferences for the future vector also are investigated. The procedures are applied to data from two longitudinal clinical trials.

## Automatic Approximation of the Marginal Likelihood in Non-Gaussian Hierarchical Models

❖ Hans J. Skaug, University of Bergen, Department of Mathematics, Johannes Brunsgate 12, Bergen, 5008 Norway, *skaug@mi.uib.no*; David A. Fournier, Otter Research Ltd.

***Key Words:*** ad model builder, importance sampling, Laplace approximation

We show the fitting of non-Gaussian hierarchical random effects models by approximate maximum likelihood can be made automatic to the same extent that Bayesian model fitting can be automated by the program BUGS. The word “automatic” here means the technical details of computation are made transparent to the user. We achieve this by combining a technique from computer science known as “automatic differentiation” with the Laplace approximation for calculating the marginal likelihood. Automatic differentiation, which should not be confused with symbolic differentiation, is mostly unknown to statisticians, and hence we review basic ideas and results. The computational performance of our approach is compared to that of existing mixed-model software on a suite of datasets selected from the mixed-model literature.

## Adversarial Learning

❖ Bowei Xi, Purdue University, 150 N. University Street, Dept of Statistics, West Lafayette, IN 47907, *xbw@stat.purdue.edu*; Murat Kantarcioglu, The University of Texas at Dallas; Chris Clifton, Purdue University

***Key Words:*** data mining, adversarial learning, game theory

Many data mining applications, both current and proposed, are faced with an active adversary. Problems range from the annoyance of spam to the damage of computer hackers to the destruction of terrorists. These problems pose a significant new challenge: The behavior of a class (the adversary) may adapt to avoid detection. In all of these cases, data mining has been proposed as a solution: from training spam filter to using data mining to identify terrorists. We use a game theoretic approach to identify a steady-state: What happens when both parties are doing the best they can to achieve their conflicting goals? We will demonstrate that in a spam email setting.

## Maximum Entropy Data Camouflaging

❖ Kurt Pflughoeft, Market Probe, 2655 N. Mayfair Road, Milwaukee, WI 53226, *k.pflughoeft@marketprobe.com*; Ehsan S. Soofi, University of Wisconsin-Milwaukee; Refik Soyer, The George Washington University

***Key Words:*** maximum entropy distribution, moment, percentile, statistical disclosure limitation, intrusion

In the problem of data security the main concern is to avoid misuse by the intruders who seek information on individuals while providing useful information for statistical analysis to legitimate users. We propose a maximum entropy (ME) approach for masking the data. In the proposed procedure the released data maintains all the essential features of the actual data required for statistical analysis, but all the individual entries are completely camouflaged. Our ME approach is complementary to paradigm proposed by Keller-McNulty, Nakhleh, and Singpurwalla

(2005) where the information content of the data, not the actual data, is masked through increasing the entropy of data-generating distribution. The proposed ME procedure will be illustrated and compared with the existing methods.

## 205 A Changing World: Katrina, Children, Judges, and More ● ✪

Business and Economics Statistics Section
**Monday, August 7, 2:00 pm–3:50 pm**

### Statistics - Sociophysics - Mediaphysics (Statistical Physics of Social Mass Media Phenomena)

❖ Dmitri V. Kuznetsov, Media Planning Group, 101 Huntington Ave., Boston, MA 02199, *dmitri.kuznetsov@mpg.com*; Igor Mandel, Media Planning Group

***Key Words:*** sociophysics, mediaphysics, word of mouth, opinion propagation, influence, mass media

The invasion of physics to traditional for statistics fields such as economics and sociology is gaining its momentum. Its success originates from the deep similarities in rising from single particle properties to complex object features in physics and from single person characteristics to social mass behavior. In general, if statistics usually formulates hypotheses about population in terms of population itself, statistical physics does it in terms of distributions of individual objects and then generalizes to the population. In 2005 we introduced Mediaphysics as a part of sociophysics, studying processes of mass communications in social systems and demonstrated its potential for applications in different areas. In this article we discuss theoretical problems of relations between statistics and mediaphysics, and some applications of the mediaphysical technique to marketing modeling.

### The Impact of Hurricane Katrina on Business Establishments

❖ Ron S. Jarmin, U.S. Census Bureau, 4700 Silver Hill Road, Center for Economic Studies, Stop 6300, Washington, DC 20233-6300, *ron.s.jarmin@census.gov*

***Key Words:*** establishments, hurricane, Katrina, GIS, employment, damage

In this paper we provide estimates of the number of business establishments affected by Hurricane Katrina. This is done by using GIS (Geographic Information Systems) tools to merge information from FEMA disaster area maps to establishment level latitude-longitude coded microdata from the Census Bureau Business Register (BR). We then link to the Longitudinal Business Database and Economic Censuses to construct estimates of the employment, payroll, sales, capital stock and so on at affected establishments. Our estimates are useful for two broad purposes. First, the estimates themselves are of value to measure the impact of the hurricane on businesses and their employees. Second, knowing precisely which business units are affected and to what extent helps statistical organizations such as the Census Bureau assess the impact of disasters on their ongoing statistical programs.

### Estimating the Change in the Gender Wage Gap When Employment Composition Changes: Evidence for Japan, 1987–2002

❖ Daiji Kawaguchi, Hitotsubashi University, Naka 2 1, Kunitachi Tokyo, 186-8601 Japan, *kawaguch@econ.hit-u.ac.jp*; Hisahiro Naito, Tsukuba University

***Key Words:*** gender wage gap, Roy model, self-selection, trimming estimator, Japan

Gender wage gap has narrowed significantly in the last 20 years in Japan. The employment population ratio of women also has increased in the same period. The change of compositions of workers could result in the observed gender wage convergence without any change in the structure of offered wages. This paper develops a method to estimate the convergence of gender wage gap when the compositions of workers of both genders change. The newly developed estimators are trimming estimators that trim the wage distribution of certain education and age groups so the composition of workers in terms of these observable characteristics is stable over time. Applying two extreme trimming rules, the trimming estimators identify the upper and lower bounds of true gender wage convergence.

### When Do Judges Explain Themselves?

❖ Alan Izenman, Temple University, Department of Statistics, Speakman Hall 1810 N 13th St, Philadelphia, PA 19122-6083, *alan@temple.edu*; David Hoffman, Temple University

***Key Words:*** statistics and the law, judicial decisions, case complexity index, censored data, predictors, opinion or order

An important topic of interest to legal scholars is under what conditions will a judge decide to write an opinion in a case, instead of merely an order. Such an "opinion or order" decision tends to be reflected in whether the decision is published by either Westlaw or Lexis. To investigate the conditions under which a judge decides to write an opinion or an order in each case, we collected a large number of predictor variables related to the status of each judicial decision in a large random sample of civil cases from each of four jurisdictions, Maryland, California (ND), Pennsylvania (ED), and New York (SD), during 2003. Certain of these cases had not yet ended at the time of data collection, and so get treated as censored data. We describe a model that helps us deal with the complicated aspects of this question, and we present a "complexity index" for each case.

### Do Regular Cycles Occur in American Politics?

❖ Samuel Merrill, Wilkes University, 3024 43rd Court, NW, Olympia, WA 98502, *smerrill@zhonka.net*; Bernard Grofman, University of California, Irvine

***Key Words:*** spectral analysis, time series, negative feedback

Are there cycles in American politics (i.e., does the proportion of the Democratic/Republican vote share for president or seat share in Congress rise and fall over extended periods of time and, if so, are the cycles regular?) If so, what is the cycling period? If there are regular cycles, can we construct an integrated model - such as a negative feedback loop - that identifies forces that could generate the observed patterns? We use spectral analysis to test for the presence and length of cycles and develop a voter-party interaction model that depends on the tensions between parties' policy and office motivations and between voters' tendency to sustain incumbents while reacting against extreme policies. We find a plausible fit between the regular cycling that this model projects and the time series of two-party politics in America over the past century and a half.

### Tolls, Exchange Rates, and International Bridge Traffic

❖ Thomas Fullerton, The University of Texas at El Paso, CBA 236, 500 West University Avenue, El Paso, TX 79968-0543, *tomf@utep.edu*

***Key Words:*** tolls, bridges, Mexico, applied, econometrics

Tolls are being considered increasingly as means for financing roads and bridges. At present, there is relatively little research regarding the impacts of tariffs on infrastructure usage. To partially address this gap in the literature, monthly international bridge traffic from Ciudad Juarez, Mexico, to El Paso, Texas, are analyzed using ARIMA transfer functions. Toll rate, currency market, and business cycle fluctuations are included in the econometric analysis.

### High-Frequency Returns, Jumps, and the Mixture of Normals Hypothesis

❖ Jeff Fleming, Rice University, Jones Graduate School, 6100 Main Street, Houston, TX 77005, *jfleming@rice.edu*; Brad Paye, Rice University

***Key Words:*** asset prices, bipower variation, integrated volatility, jumps, realized volatility, stock returns

Previous empirical studies find both evidence of jumps in asset prices and that returns standardized by realized volatility are approximately standard normal. This paper resolves the apparent contradiction. We show, using high-frequency returns data for 20 heavily-traded US stocks, that microstructure noise may bias kurtosis estimates for standardized returns. Applying a bias-corrected realized variance estimator, we find that standardized returns are platykurtotic and that the standard normal distribution is easily rejected. When daily returns are standardized using realized bipower variation, an estimator for integrated volatility that is robust to the presence of jumps, the resulting series appears to be approximately standard normal. These results support the view that jumps should be included in models of stock prices.

## 206 Stochastic Processes with Applications ●

Biometrics Section

**Monday, August 7, 2:00 pm–3:50 pm**

### Intensity Estimates for Spike Train Data Observed under Multiple Behavioral States

❖ Matt Gregas, Harvard School of Public Health, Harvard School of Public Health, Department of Biostatistics, Boston, MA 02115, *mgregas@hsph.harvard.edu*

***Key Words:*** intensity function, spike trains, Poisson process

Spike trains are the fundamental experimental unit of many neurophysiological experiments. Intensity function estimates are usually obtained from a collection of spike trains that are modeled as independent realizations of a Poisson process with common intensity function. In practice the intensity function often depends upon the behavioral state of the subject. Thus the intensity function that generates the spike trains may be a concatinization of several separate intensity functions that are joined at behavioral markers. That is, the intensity function is dependent upon the behavioral epoch of the experiment. The lengths of these epochs vary across the experimental replications. We propose a local likelihood methodology that allow us to estimate the intensity function using all available spike train information while correctly accounting for the different lengths.

### Probabilistic Model To Evaluate Biological Process

❖ Hrishikesh Chakraborty, RTI International, Statistics Research Division, RTI International, RTP, NC 27709, *hchakraborty@rti.org*; Pranab K. Sen, The University of North Carolina at Chapel Hill

***Key Words:*** probabilistic model, HIV transmission, male-to-female, bootstrap, successive approximation, underlying biological process

To model biological relationship it is impossible to test theoretical model due to availability of data. Relevant partial information can be obtained from different independent trials or from some published results. In this paper we demonstrated how to use conditional and unconditional probability theories to combine independent study data and published results to develop models to explain biological relationship. We used bootstrap re-sampling to account for repeated observations and successive approximation to estimate model parameters. This technique was used to develop two probabilistic models to evaluate the male-to-female penile-vaginal per-sexual-act HIV-1 transmission probability as a function of changes in transmitter cell and receptor cell counts. This type of modeling is very important when it is impractical to design one such study to evaluate underlying biological process.

### Parametric Inference from Window-Censored Renewal Processes and Applications

❖ Yanxing Zhao, The Ohio State University, 632 Trumbull Court, Columbus, OH 43210, *yanxingzhao@hotmail.com*; H. N. Nagaraja, The Ohio State University

***Key Words:*** renewal process, window censoring, maximum likelihood estimate, Fisher information, simulation

Suppose we have m independent, identical renewal processes observed over a fixed length of time. Only the renewals that occur within an observation window are recorded. Assuming a parametric model for the renewal time distribution, we obtain the maximum likelihood estimates (MLE) and study their properties. We consider the exponential, gamma distributions in our study. We compute the Fisher information matrix using simulation. We use it to study the large sample properties of the MLE and to determine optimal window length when the total time of observation is fixed. Our results are applied to a longitudinal dataset of lupus relapses.

### Stochastic Models for MRI Lesion Count Data from Patients with Relapsing Remitting Multiple Sclerosis

❖ Xiaobai Li, The Ohio State University, 4490 Westborough Drive, W., Columbus, OH 43220, *li.478@osu.edu*; H. N. Nagaraja, The Ohio State University

***Key Words:*** RRMS, MRI, queueing theory, hidden Markov

The longitudinal T1-weighted Gadolinium-enhancing MRI lesion count sequences provide information on the onset and sojourn time of the lesion enhancement for patients with relapsing remitting multiple sclerosis (RRMS). The infinite-server queue with Poisson arrival process and exponential service is proposed for this type of data. The rate of the Poisson arrival process can also be allowed to be governed by a two-state hidden Markov chain. We describe the likelihood function for each model based on appropriate assumptions and fit these models to a dataset from 9 RRMS patients. We obtain the maximum likelihood estimators of the parameters of interest arising from these models and study their asymptotic properties through simulation. We discuss the

validation of the assumptions for the proposed models and examine the robustness of these estimators. We suggest the application of the models.

### Tracking of Multiple Merging and Splitting Targets with Application to Convective Systems

❖ Curtis Storlie, North Carolina State University, 2501 Founders Drive, Campus Box 8203, Raleigh, NC 27695, *storlie@stat.ncsu.edu*

***Key Words:*** target tracking, hidden Markov, convective systems, Gaussian processes, Brownian motion

A statistical approach to multiple target tracking is presented which allows for birth, death, splitting and merging of targets. Targets are also allowed to go undetected for several frames. The splitting and merging of targets is a novel addition for a statistically based tracking algorithm. This addition is essential for the tracking of storms, which is the motivation for this work. The method assumes that the location of a target behaves like a Gaussian Process when it is observable. A Markov Chain model decides when the birth, death, splitting, or merging of targets takes place. The tracking estimate is achieved by an algorithm that finds the tracks that maximize the conditional density of the unknown variables given the data. An approach to quantify the confidence in a tracking estimate and some theoretical properties of the estimate are presented as well.

### An Efficient Algorithm for Exact Distribution of Discrete Scan Statistics

❖ Morteza Ebneshahrashoob, California State University, Long Beach, Department of Mathematics and Statistics, 1250 Bellflower Boulevard, Long Beach, CA 90840, *mortezae@csulb.edu*; Tangan Gao, California State University, Long Beach; Mengnien Wu, Tamkang University

***Key Words:*** two-state Markov dependent trials, i.i.d. Bernoulli trials, probability generating functions, scan statistics

Waiting time random variables and related scan statistics have a variety of interesting and useful applications. In this paper, exact distribution of discrete scan statistics for the cases of homogeneous two-state Markov dependent trials and i.i.d. Bernoulli trials are discussed by utilizing probability-generating functions. A simple algorithm was developed to calculate the distributions. Numerical results show the algorithm is efficient and capable of handling large problems.

## 207 Estimating Functions, Goodness-of-Fit, and Smoothing for Nonparametric and Semiparametric Models

IMS, Biometrics Section, Section on Nonparametric Statistics, ENAR

**Monday, August 7, 2:00 pm–3:50 pm**

### Minimax Estimation Using Higher-Order Estimating Functions

❖ Lingling Li, Harvard School of Public Health, Department of Biostatistics, 677 Huntington Ave, Boston, MA 02115, *lili@hsph.harvard.edu*; Eric Tchetgen, Harvard School of Public Health; James Robins, Harvard School of Public Health; Aad van der Vaart, Vrije Universiteit Amsterdam

***Key Words:*** minimax, higher order u-statistic, nonparametric, semiparametric, influence function

Suppose we obtain n i.i.d random vectors O with unknown distribution F. Our goal is to construct minimax estimators for a functional b(F) in a model that places no restriction on F, other than bounds on the roughness of certain density and conditional expectation functions. We consider a general class of functionals that includes: (1) The weighted average treatment effect (2) The marginal mean of a response Y, when Y is missing at random. The functionals in our class have a non-zero semiparametric information bound. However, with the high dimensional data collected in many applications, roughness bounds that are substantively meaningful may yield non-root-n minimax rates of estimation, that are no longer attainable using standard asymptotically linear estimators. We present novel minimax estimators for functionals in our class that are based on higher order U-statistics.

### Higher Order Influence Functions for Inference in Monotone Missing Data Models

❖ Eric Tchetgen, Harvard School of Public Health, Department of Biostatistics, 677 Huntington Ave, Boston, MA 02115, *etchetge@hsph.harvard.edu*; Lingling Li, Harvard School of Public Health; James Robins, Harvard School of Public Health; Aad van der Vaart, Vrije Universiteit Amsterdam

***Key Words:*** monotone missing, higher order u-statistic, CAR, doubly-robust, nonparametric

Suppose we collect n iid high dimensional random vectors of time varying covariates {L(t),t=1,...,T} where t indexes time and T denotes the end of the study. Suppose that the main interest is in the mean of Y, a component of L(T), but some patients are missing Y due to dropping out of the study prior to completion. Under the assumption of Coarsening at Random (CAR), recent advances in semiparametric theory for monotone missing data models have led to doubly-robust (DR) estimators that are n1/2 consistent if either (i) a working model for the missingness mechanism , or (ii) a working model for the full data distribution are correctly specified. We introduce novel estimators for the mean of Y that remain consistent under weaker assumptions than (i) and (ii); however we must sacrifice n1/2 consistency for slower convergence rates comparable to those encountered in nonparametric inference.

### Minimax Interval Estimation of Optimal Treatment Stategies

❖ James Robins, Harvard School of Public Health, Kresge, 677 huntington ave, Boston, MA 02115, *robins@hsph.harvard.edu*; Eric Tchetgen, Harvard School of Public Health; Lingling Li, Harvard School of Public Health; Aad Van der Vaart, Amsterdam

***Key Words:*** higher order u-statistics, causal inference, nonparametric statistics, semiparametric statistics, optimal treatment strategy

James Robins (joint with Aad van der Vaart, Lingling Li, Eric Tchetgen) We use our unified theory of inference in non-,semi-, and fully parametric problems based on higher order U-statistics to construct optimal confidence intervals for the conditional treatment effect at each level of a high dimensional vector X of pretreatment continuous covariates based on data from either randomized or observational study data. We then use this interval to construct optimal confidence intervals for the optimal treatment strategy function d*(X) that determines the optimal treatment as a function of a patient's covariates X. This methodology helps solve the vexing problem of how to determine which apparent qualitative interactions observed in a randomized experiment are real rather than do to data snooping and sampling variability.

### A Semi-Adaptive Smoothing Algorithm in Bispectrum Estimation

❖ Wei Yang, University at Albany, 08093, *wy4515@albany.edu*; Igor Zurbenko, University at Albany

***Key Words:*** KZ statistics, KZFT, bispectrum, smoothing, squared variation, third-order periodogram

Kolmogorov-Zurbenko Fourier Transform (KZFT) is an iteration of the Fourier transform. Fourier transform has an energy leakage at the adjacent frequency. KZFT may overcome it by k degrees less where k is the number of iterations. Third-order periodogram is not a consistent estimator of the bispectrum. Averaging the periodogram in blocks of the two-dimensional frequency domain may solve this problem. One way is to fix the length of the spectral window, which may lead to biased estimation when the bispectrum exhibits a variable order of smoothness. In our method, KZFT was used to calculate the third-order periodogram, which we smoothed using a varying square window until the squared variation within the window reached a prespecified percentage of the total squared variation. The algorithm has been applied in simulated and EEG data with good results.

### Estimation in Constrained Models

❖ Hanxiang Peng, University of Mississippi, Department of Mathematics, University of Mississippi, University, MS 38677, *mmpeng@olemiss.edu*

***Key Words:*** efficiency, empirical likelihood, series estimator, influence function

In this talk, we discuss estimation in models defined through constraints. We present several procedures of estimation which may or may not yield efficient estimators in the sense of least dispersed regular estimators. Examples are given to demonstrate how much efficiency gain can be obtained with efficient estimators compared with other inefficient estimators such as empirical estimators. We also report some simulation results illustrating the behaviors of the proposed estimators.

### Goodness-of-Fit Tests via Phi-Divergences

❖ Leah R. Jager, University of Washington, 5705 Phinney Ave. N., Apt. 206, Seattle, WA 98103, *leah@stat.washington.edu*; Jon A. Wellner, University of Washington

***Key Words:*** goodness-of-fit, phi-divergence, Poisson boundaries, multiple comparisons

We introduce a unified family of goodness-of-fit statistics that includes both the supremum version of the Anderson-Darling statistic and the test statistic of Berk and Jones (1979) as special cases. The family is based on phi-divergences somewhat analogously to the tests for multinomial families introduced by Cressie and Read (1984). We show that the asymptotic null distribution theory for the Anderson-Darling and Berk-Jones statistics extends to the entire family. We describe necessary and sufficient conditions for the statistics to converge to their natural parameters under fixed alternatives, and show that the family exhibits Poisson boundary behavior for certain extreme alternatives. We prove that this family achieves the same optimal detection boundary as Tukey's higher criticism statistic for normal shift mixture alternatives, as discussed in Donoho and Jin (2004).

### Extensions of the Penalized Spline Propensity Prediction Method for Monotone Missing Data

❖ Guangyu Zhang, University of Michigan, 3566 Green Brier Blvd., Apt 412, Ann Arbor, MI 48105, *guangyuz@umich.edu*; Roderick J. Little, University of Michigan

***Key Words:*** monotone missing data, penalized spline propensity prediction

Little and An (2004) proposed a Penalized Spline Propensity Prediction (PSPP) method, which yields an estimate of the marginal mean of the missing variable with a double robustness property. We study extensions of the PSPP method to monotone missing data, where the variables can be ranged so that all subsequent variables, say Yj+1,.,Yp , are missing for cases where Yj is missing, for all j = 1,.,p-1. This pattern arises frequently in longitudinal studies due to attrition. Let (X1,.,Xq,Y1,.,Yp) denote a p+q-dimensional vector of variables with Y1,.,Yp represent the monotone missing data part and X1,.,Xq fully observed covariates. Our research objective is to estimate the marginal and conditional means of Y1,.,Yp . Results of a simulation study illustrate the efficiency and robustness property of the stepwise PSPP method. We also apply the method to an online weigh loss study.

## 208 Advances in Bayesian Computation

Section on Bayesian Statistical Science

**Monday, August 7, 2:00 pm–3:50 pm**

### Likelihood Subgradient Densities

❖ Kjell Nygren, IMS Health, 567 Grant Street, Newtown, PA 18940, *knygren@us.imshealth.com*; Lan Nygren, Rider University

***Key Words:*** Bayesian statistics, log-concave likelihood functions, generalized linear model, Poisson regression

We introduce likelihood subgradient densities and explore their basic properties. Using mixtures of likelihood subgradient densities, we propose an approach for the construction of tight enveloping functions in the Bayesian context. In the case of normal priors with normal data, the area underneath the resulting enveloping function is bounded above by 2/sqrt(pi) (which is approximately equal to 1.128). The approach is extended to k-dimensional models where the corresponding bound is (2/sqrt(pi))^(k). More generally, our approach should also yield tight enveloping functions for other models in which the data is close to normal. Such models include generalized linear models (e.g., Bayesian Poisson regression and the Bayesian logit model). Simulations based on the approach are performed for two separate models using accept-reject methods.

### Likelihood Approximations in Bayesian Multiple Curve Fitting

❖ Carsten Botts, Williams College, Bronfman Science Center, Williams College, Williamstown, MA 01267, *cbotts@williams.edu*; Michael Daniels, University of Florida

***Key Words:*** b-splines, Laplace approximation, reversible jump MCMC, unit-information prior

We model functional data from multiple subjects with a regression spline linear mixed model. In this model, the expected values for any subject (conditioned on the random effects) can be written as the sum of a population curve and a subject-specific deviate from this population curve. The population curve and subject-specific deviates are modeled as b-splines with k and k’ knots located at t_k and t_k’, respectively. We sample from the posterior p( k, t_k, k’, t_k’ | y), where y is the observed data, using reversible jump MCMC methods. Sampling from this posterior distribution is complicated by the fact that no analytical form for p(y | k, t_k, k’, t_k’) exists. We explore two approximations to this likelihood and study how each penalizes linear mixed models with too many random-effect knots.

### Video Segmentation Using a Bayesian Online EM Algorithm

❖ Johan Lindström, Lund University, Box 118, Lund, SE-22100 Sweden, *johanl@maths.lth.se*; Finn Lindgren, Lund University; Kalle Åström, Lund University; Jan Holst, Lund University; Ulla Holst, Lund University

***Key Words:*** recursive estimation, mixture models, video segmentation, Bayesian models

A method for video segmentation using Bayesian models and an online EM algorithm is presented. The method models each layer as a Gaussian mixture, with local, per pixel, parameters for the background and global parameters for the foreground. The online EM algorithm also uses a progressive learning rate allowing the relative update speed of each Gaussian component to depend on how long that component has been observed and similar foreground components are merged using a Kullback-Leibler distance. Performance of the algorithm for gray-scale and RGB videos as well as on output from a Prewitt edge detector is compared to that of another algorithm. Especially for the edge detector, performance increases dramatically.

### Two-Stage EM Algorithm on the Random Transfer Function Model

❖ Hyunyoung Choi, University of California, Santa Barbara, Statistics and Applied Probability, Santa Barbara, CA 93106-3110, *choih@pstat.ucsb.edu*; Bonnie K. Ray, IBM T. J. Watson Research Center

***Key Words:*** random transfer function model, EM algorithm, MCMC

A random transfer function model can be used to characterize a panel of time series whose individual responses to an input series are potentially different, yet follow a common distribution (Choi 2005). Previous research has used MCMC techniques to estimate the parameters of this random transfer function model. In this paper, we propose a two-stage EM algorithm for estimating model parameters as a more computationally efficient alternative and investigate its properties compared to the MCMC approach. For illustration, we apply the methods to model the impact of company-level management changes on the financial health of publicly traded companies, based on a random sample of companies that experienced such a change over the last two years.

### Bayesian Variable Selection in Clustering High Dimensional Data with Substructure

❖ Michael D. Swartz, M. D. Anderson Cancer Center, 1155 Pressler Street, Unit 1340, Department of Epidemiology, Houston, TX 77030-3721, *mdswartz@mdanderson.org*; Marina Vannucci, Texas A&M University

***Key Words:*** Bayesian variable selection, Bayesian mixture models, reversible jump Markov chain Monte Carlo, microarray analyses

Biological experimenters have incorporated microarray data into designed experiments, where the goal of inference is to uncover genes that are differentially expressed across the experimental groups. Commonly, researchers perform multiple ANOVA analyses with a multiple comparison correction. We cast the problem into one of Bayesian variable selection in clustering high dimensional data with substructure. The experimental groups naturally define substructure in the data, and clustering looks for similarities and differences among the subgroups, while the variable selection searches for those genes that best differentiate the subgroups. We have modified the reversible jump MCMC algorithm of Tadesse et al. (JASA, 2005) to account for experimental design substructure while exploring the cluster and feature space. We successfully employed the new algorithm to analyze preliminary data sets.

## 209 Design and Analysis of Response Surface Experiments

Section on Physical and Engineering Sciences, Section on Quality and Productivity

**Monday, August 7, 2:00 pm–3:50 pm**

### Semiparametric Techniques for Response Surface Methodology

❖ Stephanie Pickle, Virginia Polytechnic Institute and State University, 8534 Willow Creek Drive, Roanoke, VA 24019, *spickle@vt.edu*; Jeffrey B. Birch, Virginia Polytechnic Institute and State University; Timothy Robinson, University of Wyoming

***Key Words:*** semi-parametric regression, model robust regression, robust parameter design

Many industrial statisticians employ the techniques of Response Surface Methodology (RSM) to study and optimize products and processes. A second-order Taylor series approximation is commonly utilized to model the data; however, parametric models are not always adequate. In these situations, any degree of model misspecification may result in serious bias of the estimated response. Nonparametric methods have been suggested as an alternative as they can capture structure in the data that a misspecified parametric model cannot. Yet nonparametric fits may be highly variable especially in small sample settings which are common in RSM. Therefore, semi-parametric regression techniques are proposed for use in the RSM setting. These methods will be applied to an elementary RSM problem as well as the robust parameter design problem.

### Rethinking Steepest Ascent

❖ Robert W. Mee, University of Tennessee, 333 Stokely Management Center, Knoxville, TN 37996, *rmee@utk.edu*; Jihua Xiao, University of Tennessee

***Key Words:*** multiple responses, response surface methodology

The path of steepest ascent consists of points that maximize the predicted response for a fitted first-order model among all points with the same standard error of prediction. When there are multiple responses or additional constraints (e.g., on cost or throughput), current methods are of limited usefulness. We address the problem by constructing confidence bounds for improvement of each response relative to its performance at the center of the design. Overlaying these confidence regions provides the user with the best insight regarding the desirability of moving to new locations. While the methods are simplest when fitting first-order models, they are easily extended to more complicated models using standard statistical software.

### Sequential Methodology for Detecting Jumps in Complex Surfaces

❖ Yan Lan, University of Michigan, Department of Statistics, 439 West Hall, 1085 South University, Ann Arbor, MI 48109, *lany@umich.edu*; George Michailidis, University of Michigan

***Key Words:*** boundary curve, sequential, regression trees

In many scientific and engineering settings, the response surface of a physical phenomenon or engineering system has complex characteristics. The goal is to identify input settings that explore interesting features of the underlying surface. We present new sequential design methodology and are interested in developing computationally efficient methodology that detects and fits the underlying boundary curve, which usually has a complex shape. The proposed methodology utilizes a variety of techniques, such as regression trees, structural change models, various tests such as Anderson-Darling and likelihood ratio tests and QQ-plots. We start with a small initial design, which is subsequently refined, as more information about the boundary curve is obtained. The methodology proves fairly robust to noise and is illustrated on a number of simulated data sets.

### Minimal-Point Optimal Designs for Second-Order Response Surfaces

❖ Ray-Bing Chen, National University of Kaohsiung, No 700 Kaohsiung University Road, Institute of Statistics, Kaohsiung, 811 Taiwan, *rbchen@nuk.edu.tw*; Yu-Jen Tsai, National University of Kaohsiung; Dennis K. J. Lin, The Pennsylvania State University

***Key Words:*** best angle sampler, central composite design, d-optimality, point efficiency, simulated annealing algorithm, small composite design

In this work, we are interested in sequentially finding the minimal-point optimal designs for the second-order response surfaces when a specific first-order design is given at the first stage. A modified simulated annealing algorithm is proposed for finding these minimal-point designs according to the D-optimal criterion. Based on the different first-order designs, the corresponding D-optimal composite designs with minimal supports were found. These minimal-point designs were then compared with other small composite designs and minimal-point designs. It is shown that the proposed minimal-point designs perform well in general. In the cases they are not to be D-optimal, they have reasonably high D-efficiencies.

### Alphabet-Optimal Central Composite Designs

❖ Trevor A. Craney, Sikorsky, 6900 Main Street, Mail Stop: S317A, Stratford, CT 06615-9129, *trevor.craney@sbcglobal.net*

***Key Words:*** CCD, experimental design, response surface

Many criteria exist for which design matrices can be compared for efficiency to optimality. Most standard design matrices for first- or second-order models are limited from improvement due to the fixed values associated with their construction. Central composite designs permit users to specify the axial values, thus allowing restricted optimization to be made of the design matrix. A method will be discussed for determining G-optimal central composite designs based on their efficiencies. These efficiencies will be compared to central composite designs incorporating face-centered, spherical, and rotatable axial values for a varying number of factors and replication of center runs. This same information will be reported for D-optimality, A-optimality, and E-optimality, as well as a new class of VIF-optimal designs, which measure efficiency relative to orthogonal design matrices.

### Orthogonal Blocking in Response Surface Designs with Split-Plot Structure

❖ Li Wang, Virginia Polytechnic Institute and State University, 401 Fairfax Road, Apt. 1424, Blacksburg, VA 24060, *wangleelee2001@hotmail.com*; Scott Kowalski, Minitab Inc.; Geoff Vining, Virginia Polytechnic Institute and State University

***Key Words:*** central composite designs, response surface, orthogonal blocking, design of experiment, split-plot experiment

When all experimental runs can not be performed under homogeneous conditions, people often construct blocks to increase the power for testing the treatment effect. Orthogonal blocking provides the same estimator of the polynomial effects as the one that would be obtained by ignoring the blocks. Khuri (1992)provides the matrix notation of the orthogonal blocking condition for a completely randomized design. In response surface designs with a split-plot structure, orthogonal blocking has received little attention. Khuri (1992) and Goos (2002) show that for a balanced fixed block design, OLS estimate has to be the same as GLS estimate. In this paper, we show that for a balanced OLS-GLS equivalent splitplot design, orthogonal blocking can be achieved. The conditions for an orthogonal blocked split-plot central composite design are given.

## 210 Contributed Posters

Business and Economics Statistics Section, Section on Statistical Education, Biopharmaceutical Section, Section on Statistics in Epidemiology, Section on Bayesian Statistical Science, Section on Teaching Statistics in the Health Sciences, Section on Health Policy Statistics, Social Statistics Section, WNAR, Section on Statistical Consulting, Section on Survey Research Methods, Section on Statistics and Marketing

**Monday, August 7, 2:00 pm–3:50 pm**

### Measuring Financial Data Quality Assessment and Improvement

❖ George Sirbu, Bentley College, 175 Forest Street, Morison 351, Waltham, 02452, *gsirbu@bentley.edu*; Mary Ann Robbert, Bentley College; Donna Fletcher, Bentley College

***Key Words:*** risk management, data quality, sequential allocation

The Bank of International Settlements requires banks to have a robust system in place to validate the accuracy and consistency of rating systems, processes and all relevant risk components. Such a system relies heavily on high quality financial data. In the presentation we establish statistical measurements of the data quality of the entire data stream used in calculating pre-settlement exposure (PSE) of all of its product positions. Due to the high volume of information that will have to be processed, we recommend alternative techniques in form of sequential allocation for a more efficient analysis. Although the size of the data sampled is unknown in such case, simulations will offer an insight of what resources can be expected to be allocated for the estimation of the overall measure of the PSE.

### Perceptions of Men versus Women in a Business Organization in 2005

❖ Kris Moore, Baylor University, One Bear Place, 98005, Waco, TX 76798-8005, *Kris_Moore@baylor.edu*; Dawn Carlson, Baylor

University; Amber St. Anant, Baylor University; Dwayne Whiten, Texas A&M University

This survey of men and women executives listed in 2004 Standard & Poors Register of Corporations, Directors and Executives and the 2005 Dun & Bradstreet Reference Book of Corporate Management sought answers to questions of differences in perceptions versus gender. There are comparisons of gender versus work expectations, obstacles women face, comfort of work, perception of contribution and work life balances.

## Inference for a Hazard Rate Change Point under Dependent Censoring

❖ Nan Zhang, Rice University, 7900 Cambridge Road, APT Building28 Room 1E, Houston, TX 77054, *jennyz@rice.edu*; Xuelin Huang, M. D. Anderson Cancer Center

***Key Words:*** change point, likelihood ratio test, dependent censoring, copula

This research is motivated by the study on how the 1999 ACT affected the survival rate of insurance companies. While usually the lifetime of an insurance company can be reasonably assumed to have constant hazard rate, the hazard rate may experience a jump due to government regulations. The likelihood ratio test for constant hazard rate against a step change alternative has received considerable attention in recent years. In the presence of covariates, we propose an extended likelihood ratio test for a constant failure rate against the alternative of a failure rate involving a single change-point. We also use an assumed copula to do sensitivity analysis for the test under different degrees of dependent censoring. An application to insurance company data is presented.

## Model Identification and Forecasting of Stationary Models with GARCH(P,Q) Errors

❖ Melody Ghahramani, University of Manitoba, Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2 Canada, *umghahra@cc.umanitoba.ca*

***Key Words:*** ARMA-GARCH, kurtosis, forecasting, autocorrelation, RCA

Financial returns are often modeled as autoregressive time series with innovations having conditional heteroscedastic variances, especially with GARCH processes. We have extended the results of Thavaneswaran et al. (2005) on the kurtosis of a zero-mean GARCH process for a class of stationary ARMA (p, q) models with GARCH errors. The autocorrelation of the squared process will be useful in identifying the order of the GARCH processes. Moment properties, including kurtosis of Random Coefficient Autoregressive (RCA) models with GARCH errors are derived. We also study the forecasting problem for stationary ARMA (p, q) financial time series with errors having GARCH errors. Closed form expressions for the l-steps ahead forecasts are also given in terms of weights for stationary ARMA-GARCH models.

## The Magnet Effect of Price Limits: Evidence from Transactions Data

❖ Ping-Hung Hsieh, Oregon State University, 200 Bexell, College of Business, Corvallis, OR 97331, *hsiehph@bus.orst.edu*; Yong H. Kim, University of Cincinnati; J. Jimmy Yang, Oregon State University

***Key Words:*** magnet effect, transactions data, Taiwan stock exchange, logistic regression, price limits

We investigate the magnet effect of price limits using transactions data from the Taiwan Stock Exchange. Using a logistic regression model that incorporates explanatory variables documented in the microstructure literature to capture transaction price changes, we found that the conditional probability of price going up (down) increases significantly when the price is approaching the upper (lower) price limit, supporting the magnet effect. We found that the magnet effect starts to emerge when the price is within 10 ticks from the upper price limits and about 5 ticks from the lower price limits. We also examined firm characteristics to identify possible determinants of magnet effect and found that beta, trading volume, market capitalization, and book-to-market are positively related to the degree of magnet effect. Our overall findings generate important policy implications.

## Robust Granger Causality Tests in the VARX Framework

❖ Alex Maynard, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7 Canada, *amaynard@chass.utoronto.ca*; Dietmar Bauer, Arsenal Research

***Key Words:*** Granger causality, surplus lag, nonstationary VAR, local-to-unity, long-memory, structural breaks

Predictability and Granger causality tests often rely heavily on the choice of model for the low frequency behavior of the time series. Dolado (1996) and (1995) introduced tests in I(1) VAR models based on an additional lag in the specification of the estimated equation which is not included in the tests. This leads to conventional test distributions for stationary and (co)integrated data generating processes. By extending this approach to the VARX framework, we show that it can provide Granger causality tests that accommodate stationary, nonstationary, near-nonstationary, long-memory, and unmodeled structural break processes within the context of a single $\chi^2$ null limiting distribution. Since the distribution under the null hypothesis is the same in all cases, no prior knowledge, first-stage testing, or estimation is required.

## Uplift Modeling in Direct Marketing

❖ John Lin, Epsilon, 601 Edgewater Drive, Wakefiled, MA 01880, *jlin@epsilon.com*; Qizhi Wei, Epsilon

***Key Words:*** uplift, incremental, direct marketing (DM), ROI, predictive model

In the past, predictive models were evaluated based only on the ability to generate additional responses/sales compared to randomly audiences. With growing demands for marketing accountability, statisticians are increasingly being asked to build uplift models that identify consumers who are most positively influenced by DM campaigns and show the incremental impact of DM programs so that ROI can be more accurately measured. This in turn leads to better decisions about how marketing dollars are spread across channels as well as better decisions about overall marketing budgets. This article discusses the pros/cons of different approaches to build and validate the uplift models. In-market results will be used to illustrate how models can be enhanced. We will also propose a new method to implement the uplift models, which has been applied to a recent marketing campaign with great success.

## How Bad Could Your Data Be? Variance Maximization

❖ Jeffrey Stuart, Pacific Lutheran University, Mathematics Department, Tacoma, WA 98447, *jeffrey.stuart@plu.edu*

***Key Words:*** variance maximization

For a finite, simple random sample from a closed, bounded subset S of real n-space, how large can the variance be? We show that the answer is largely determined by the convex hull of S, and that each maximum variance subset of S is comprised of certain extreme points of the convex hull. An alternative formulation in terms of sets of end points of diameters of S is presented and used in conjunction with circumscribing spheres to bound the maximum variance. Results depend on whether the sample size is even or odd. Variance maximization has application to information theory. This work was motivated by a question originally posed to the AP Statistics newsgroup. The mathematical content is accessible to those with a semester of advanced calculus and several lectures on convex analysis.

## Characterizations of Factor Analytic Covariance Structure

❖ Timothy Costigan, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *costigan_timothy_m@lilly.com*

***Key Words:*** probability inequalities, multiple testing, multivariate normal

We show that trivariate normal distributions exhibit the factor analytic covariance structure if and only if all correlations are non-negative and all partial correlations are positive. These conditions, which are also necessary and sufficient for MTP2,are used to obtain factor analytic upper and lower bounds to higher order orthant and rectangular probabilities. A simple expression is also presented for trivariate normal distributions to be MTP2 in absolute value. These characterizations are used to derive second/third order hybrid probability inequalities that can be evaluated in SAS. Examples are presented to illustrate the accuracy of the new bounds.

## Assessment of Small-Area Estimates from a Complex Survey Cancer Surveillance Project

❖ Van Parsons, National Center for Health Statistics, 3311 Toledo Road, Room 3219, Hyattsville, MD 20782, *vparsons@cdc.gov*; Nathaniel Schenker, National Center for Health Statistics; Trivellore E. Raghunathan, University of Michigan; Dawei Xie, University of Pennsylvania; William Davis, National Cancer Institute

***Key Words:*** Bayesian, NHIS, BRFSS, cancer risk factors

Cancer surveillance research requires that estimators of prevalence for risk factors over small areas be accurate and precise. Many of these prevalences can be estimated from the National Health Interview Survey (NHIS) and from the Behavioral Risk Factors Surveillance System (BRFSS). While the NHIS is a nationally representative face-to-face survey, and the BRFSS is a system of RDD state surveys, direct estimation for small areas, e.g., counties, becomes problematic for each survey because of small or no samples over the small areas and coverage biases. We have developed a Bayesian hierarchical model on the combined NHIS+BRFSS surveys to produce county level estimates. To assess the model-based accuracy and precision, we evaluate both the model-based estimates and direct survey estimates, compare the results, and describe the merits and drawbacks of the model- based approach.

## Methodology for Estimating Age of Onset of Overweight Using Self-Reported Historical Height and Weight Data

❖ Henry Xia, Centers for Disease Control and Prevention, 3311 Toledo Road, Room 6227, Hyattsville, MD 20782, *hxia@cdc.gov*; Virginia Freid, Centers for Disease Control and Prevention; Patricia Pastor, Centers for Disease Control and Prevention

***Key Words:*** statistics, overweight, public health

Between 1988--94 and 2001--04 the percentage of Americans who were overweight increased from 56% to 66%. Understanding how the AOOW varies for different subgroups in the population is useful for health education programs. This study used up to six self-reported weights and two self-reported heights over respondent's lifetimes to determine AOOW and persistence of overweight. A sex-specific model, based on a logistic growth curve and second order polynomial curve, used measured body height to impute height for ages for which only data about weight were collected. Males, Mexicans, and people below twice poverty were more likely to experience onset of overweight before the age of 30. Females, non-Hispanic whites, and nonpoor persons were more likely to become overweight at 30 years and over. People with early onset of overweight were more likely to become obese later in life.

## A Multivariate Statistical Analysis of Substance Abuse in the United States

❖ Monique Owens, SUMSRI at Miami University, 3490 Addison Place N., Grove City, OH 43123, *msmo06@hotmail.com*; Joshua Svenson, SUMSRI at Miami University

***Key Words:*** multivariate analysis, substance abuse

Where do the major drug problems occur in this country among the states? How are social and economic factors related to substance abuse in the states? We approach these questions with multivariate statistics. By using factor analysis, we distinguish the underlying factors of a collection of variables related to substance abuse. With discriminant analysis, we design a rule for classifying states as either having a major drug problem or a minor drug problem.

## Adaptive Poisson Modeling of Medication Adherence in HIV-Positive Methadone Patients

❖ Kevin Delucchi, University of California, San Francisco, 401 Parnassus Ave., 0984 TRC, San Francisco, CA 94143-0984, *KDelucchi@lppi.ucsf.edu*; George Knafl, Oregon Health & Science University; Nancy Haug, University of California, San Francisco; James Sorensen, University of California, San Francisco

***Key Words:*** Poisson, medication adherence, longitudinal, small sample, clustering, mixture modeling

The use of adaptive Poisson regression models to capture medication adherence levels from electronic pill-bottle monitoring devices is studied. Data are from a randomized trial designed to improve medication adherence in HIV+ methadone patients. Plots display changes over time and treatment effects consistent with the more standard analysis. Predicted values were clustered via an extensive search among methods and results compared to finite mixture modeling. Clustering found the more extreme cases while mixture modeling reflected levels of adherence. The mixture-modeling grouped subjects into low, intermediate, and high adherence classes while the clustering-method generated compatible, finer adherence groups. Results replicate the primary analytic finding. Improvements in the methodology are shown and factors associated with medication adherence presented. Funding: DA009253 AI057043.

## Assessing Publication Bias in Meta-analysis

❖ Xin Li, The University of Texas at Austin, 3368D Lake Austin Blvd., Austin, TX 78703, *xinli@math.utexas.edu*; Tasha Beretvas, The University of Texas at Austin

***Key Words:*** meta-analysis, publication bias, nonparametric methods

Meta-analysis (M-A) is used to summarize results of empirical studies. It involves aggregation of results from identical studies. One major criticism of M-A is the bias that could result from selective publication. Rosenthal's fail-safe N (FSN) provides one estimate of the number of studies (k0) missing as a result of publication bias. Duval and Tweedie's trim & fill method provides a non-parametric method to estimate k0 with three ways suggested: L0, Q0, and R0. The purpose of the current study is twofold. First, trim & fill method is extended to provide L0, Q0, and R0 values for correlation coefficients (instead of for delta). Second, results of using L0, Q0, and R0 were compared with k0 estimates resulting from FSN. Data from a real meta-analysis (Heller, Watson & Ilies, 2004) were used to investigate and demonstrate differences. The full paper will detail the derivations and results.

## University Graduate Mentoring

❖ Terry Tomazic, Saint Louis University, 3750 Lindell Blvd., McGannon Hall 230, St Louis, MO 63108, *tomazictj@slu.edu*; Michael Donovan, Saint Louis University; John Hicks, Saint Louis University; Eric Watterson, Saint Louis University; Barry Katz, Saint Louis University

***Key Words:*** graduate students, mentoring

The study evaluates the prevalence and quality of mentoring relationships between graduate students and faculty at the university level. In particular, the quality of mentoring is assessed between students on an assistantship and those not. Students were asked of the benefits to their academic performance and whether there were differences by department or program. Finally, mentors were assessed for knowledge, availability, professionalism, emotional and financial support, and career development and networking on behalf of students. The results are based on a total sample of 379 graduate students at a mid-western university with major doctoral and masters level programs. The survey was conducted via the Internet. The major result found is that students on assistantship were more likely to be mentored, had higher satisfaction and were more productive in scholarly work.

## Gender Differences and Factors Affecting the Interface Performance Level

❖ Dennis Kira, Concordia University, 1455 de Maisonneuve Blvd., Montreal, QC H3G1M8 Canada, *kira@vax2.concordia.ca*; Fassil Nebebe, Concordia University; Raafat G. Saade, Concordia University

***Key Words:*** performance measure, user interfaces, cognition

The software interface is the point of interaction between the user and the application. The users' level of computer skills may influence perceived ease of use and perceived usefulness of the software directly as well as indirectly manifested in performance. The interface plays an important role in amplifying these relationships. These phenomena have been studied mainly within the command based and direct manipulation interfaces.

## Statistical Analysis of Aphid Data: a Case Study for AP Statistics

❖ James Matis, Texas A&M University, Department of Statistics, College Station, TX 77843-3143, *matis@stat.tamu.edu*; Thomas Kiffe, Texas A&M University; Timothy Matis, Texas Tech University; Douglass Stevenson, Texas A&M University

***Key Words:*** introductory statistics

Aphids cause more worldwide economic damage than any other insect pest. This poster illustrates the use of some standard statistical tools for the analysis of an exceptional data set on the pecan aphid. The four parts of the AP syllabus, here called Regular AP Statistics (RAPS), namely exploring data, planning a study, producing models, and confirming models, are demonstrated. Five concepts from Beyond AP Statistics (BAPS), specifically randomized block design, multiple linear regression, nonlinear regression, differential equations models, and stochastic processes, are also illustrated. The aphid is parthenogenetic, which is a unique reproductive system in the insect world. We suggest that the data analysis of this fascinating and economically important insect could form a very interesting unifying problem for the whole AP statistics course.

## Favorite Datasets from Early Phases of Drug Research: Part 6

Thomas E. Bradstreet, Merck Research Laboratories; ❖ Thomas H. Short, Indiana University of Pennsylvania, IUP Mathematics Department, 210 S 10th St, Indiana, PA 15705, *tshort@iup.edu*

***Key Words:*** statistics education, standard data sets, data analysis

We present small to moderate sized data sets to be used for classroom demonstrations, workshops, homework problems, projects, and exam questions. The data sets can also serve as standards in research comparisons of new and existing statistical methodologies. These data sets illustrate topics and issues including bioequivalence, dose proportionality, dose response, special target populations (such as the elderly and gender groups), blocking factors, interaction, non-normality, heteroscedasticity, concordant and discordant outliers, repeated measures for longitudinal data, unequal sequence effects in crossover trials, multiplicity, and residual analysis. The data sets originate from Preclinical (animal) and Phase I Clinical Pharmacology (human) studies in drug research. These and other pedagogically useful data sets are located at www.math.iup.edu/~tshort/Bradstreet/.

## Statistics Jeopardy!Æ

❖ Neal Rogness, Grand Valley State University, 1133 MAK, 1 Campus Drive, Allendale, MI 49401, *rognessn@gvsu.edu*; Adam Weimer, The Ohio State University

***Key Words:*** teaching, education, game, activity

The presenters will share examples of boards, based on the popular television game show Jeopardy!Æ, that were created as an interactive and entertaining way to assess student understanding of statistical concepts. The boards can be used as review mechanisms for exams and/or to remind students of information they should have retained from prerequisite courses.

## Expectations for Statistical Literacy: a Comparison among Psychology, Business, and Public Health Professions

❖ S. David Kriska, Restat Systems, 2570 Sherwood Road, Bexley, OH 43209, *davidkriska@sbcglobal.net*; Mark C. Fulcomer, Richard Stockton College of New Jersey; Marcia M. Sass, University of Medicine & Dentistry of New Jersey

***Key Words:*** statistical literacy, statistics curricula, survey

The Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) and the International Statistics Project emphasize the need for statistical literacy. Definitions of literacy tend to be general statements that need to be expanded to address specific skills. This poster will present a draft survey for identifying key specific elements of literacy. The intent of the surveys will be to make comparisons of expectations held by faculty and by employers of recent graduates. The academic areas planned to be surveyed include: business, psychology, and public health. While Fulcomer, Lyke, Sass, and Kriska (2005) noted similarities in various statistics courses that cross content lines, the current effort is in identifying similarities, differences, and potentially missing elements in statistics curricula. The session will be used to encourage comments on the draft instrument.

## A Bayesian Model for Predicting the Probability of Additional Positive Axillary Nodes in Breast Cancer Patients with Positive Sentinel Lymph Node Biopsy

❖ Sunni A. Barnes, Mayo Clinic College of Medicine, Division of Biostatistics, 200 First Ave SW, Rochester, MN 55905, *barnes.sunni@mayo.edu*; Tanya Hoskin, Mayo Clinic College of Medicine; Cody Hamilton, Baylor Health Care System

***Key Words:*** prediction, logistic regression, breast cancer, WinBUGS

Sentinel lymph node biopsy (SLNB) is the preferred technique for axillary staging in breast cancer. The current standard of care for patients with positive SLNB remains complete axillary lymph node dissection (CALND). However, 40-60% of patients have no further disease. Since CALND increases the risk of acute and late morbidity a model for predicting the presence of additional axillary disease to aid clinicians in treatment decisions and possibly spare some patients from an unnecessary surgical procedure is desirable. We fit a Bayesian hierarchical generalized linear model with a logit link to our data and compare our results to those found using a frequentist modeling approach. This comparison is based on the accuracy of the predicted probabilities using the clinical cutoff pf less than 10% as justification to not perform CALND.

## Introducing Data Quality in the Classroom

❖ Mark C. Fulcomer, Richard Stockton College of New Jersey, 48 Trainor Circle, Bordentown, NJ 08505, *mcfulcomer@aol.com*

***Key Words:*** data quality, teaching, health statistics, health policy

With a long history that dates back to the early efforts of Deming, Shewhart, and others, statistical quality control and related topics have increasingly been incorporated into introductory and intermediate courses. Unfortunately, companion issues of data quality (e.g., missingness, digit preference) have not received similar attention, particularly with respect to the collection of information in the health sciences. Beginning with a seminal paper by Hotelling and Hotelling (1932), this presentation provides several examples of how some important data quality issues can be introduced in applied statistics courses. Using large files of historical data and other approaches, these examples can both enhance and enliven the educational experiences of students with more “applied” interests. The timely relevance of such issues to recent work in “health disparities” is emphasized.

## A Menu System for Stats/List on the TI Voyage 200

❖ John Turner, U.S. Naval Academy, 225 Autumn Lane, Centreville, MD 21617, *jct@usna.edu*

***Key Words:*** calculator

The TI Voyage 200, while a bit pricey, is a quite capable machine for intro stats. I will present a menu system I developed for the Voyage 200 and consider other issues, such as data entry.

## Program Assessment in Statistics at the Master’s Level

❖ Julia Norton, California State University, East Bay, 28022 El Portal Drive, Hayward, CA 94542-2512, *julia.norton@csueastbay.edu*; Lynn Eudey, California State University, East Bay

***Key Words:*** departments of statistics, statistics masters degree, statistics curriculum, statistics programs

At California State University East Bay (formerly Hayward), assessment is mandated by the Chancellor’s Office and the State legislature. To maintain as much control over our programs as possible, we have included a first year assessment covering beginning statistical ideas from our general education course assessment, modeling questions from our undergraduate program assessment, and information from the theoretical (closed-book) and applied (open-book, computer assisted) two-day culminating examination for our MS degree program. This approach allows us to compare level appropriate materials and programs. One discovery caused us to rethink portions of our degree and proposed changes to our master’s curriculum. The analysis includes repeated measures studies and longitudinal information about our degree program.

## Discussing Factor Analysis in a 50-Minute Class Period

❖ J. Burdeane Orris, Butler University, 4600 Sunset Blvd., Indianapolis, IN 46208, *orris@butler.edu*; Bruce Bowerman, Miami University of Ohio

***Key Words:*** factor analysis, applied education

Factor analysis is an important multivariate technique used in business research. This paper will present an applied discussion of factor analysis that allows the topic to be covered by an instructor in one fifty minute class session. Discussion of principal components, factor rotation, and the final factor interpretation, along with SAS and SPSS outputs will be included. This approach allows instructors teaching a variety of statistics courses to briefly introduce factor analysis without requiring students to have an extensive background in matrix algebra. It focuses on terminology and interpretation of computer output.

## Strategies for Making Your Curriculum Vita Numerical and Graphical for Promotion, Tenure, and Career Awards

❖ Charlie Goldsmith, McMaster University, Centre for Evaluation of Medicines, 105 Main Street East Level P1, Hamilton, ON L8N 1G6 Canada, *goldsmit@mcmaster.ca*

***Key Words:*** cv evaluation, tenure, promotion, productivity patterns, criteria for success, improvement and leadership

Academic statisticians want tenure, promotion and awards. Training and productivity are found in a Curriculum Vita (CV). This paper shows how to convert productivity to tables and graphs. Tables are sorted into those that are and are not peer reviewed. Presentation tables are sorted into local, national, and international. Each row has signposts in a career: date of first contribution, PhD, etc. Contribution date (y) subtracted from current year gets contribution interval (dy). Table entries are: c,w,m, where c is count, w is weighted sum and m is mean. For each entry, w/dy is productivity per year. If the second of a pair is = the first, indicate with a + showing increased productivity; or else negative. A Leadership Index, Improvement Index and frequency table are computed from these tables. Smoothed productivity data show career patterns. Examples are used to show key points.

### Development of an Introductory Biostatistics Course for Graduate Students in Biomedical Sciences

❖ John Rutledge, Weill Medical College of Cornell University, Division of Biostatistics and Epidemiology, Department of Public Health, 411 E 69th St, New York, NY 10021, *jrr2002@med.cornell.edu*; Kylie Bryant, Weill Medical College of Cornell University; Kathy Zhou, Weill Medical College of Cornell University; Yolanda Barron, Weill Medical College of Cornell University; Anita Mesi, Weill Medical College of Cornell University; Heejung Bang, Cornell University; Eduardo Martinez-Ceballos, Weill Medical College of Cornell University; Lorraine Gudas, Weill Medical College of Cornell University; Selina Chen-Kiang, Weill Medical College of Cornell University; Madhu Mazumdar, Weill Medical College of Cornell University

***Key Words:*** biomedical sciences, introductory statistics course, experimental medicine, online course

The goals of this course included introducing the concepts of basic statistical methods used in laboratory research, distinguishing between statistical design and data analysis, teaching proper ways of reporting methods and results in papers, demonstrating how to critique statistical results in published literature, teaching use of codes written in the software program R, and introducing infrequently needed higher level methods for which statistical collaborators are needed. Course material was prepared using few books and published papers. Evaluation methods included journal clubs, a take home data analysis project, student presentations based on analysis of data generated from their laboratories, and written examination. Response has been positive. Byproducts include an online course for post-doctoral fellows and change in 'Instruction to Authors' for Journal of Experimental Medicine

## 211 President's Invited Session

The ASA, ENAR, WNAR, SSC, IMS

**Monday, August 7, 4:00 pm–5:50 pm**

### A Data-Driven World: Why Now, and What Do We Do about It?

❖ William R. Pulleyblank, IBM Business Consulting Services, Center for Business Optimization, Route 100, Somers, NY 10589, *wp@us.ibm.com*

Business and industry are experiencing a transformation in the way they handle their operations and planning due to technology advances, globalization and economic pressures, and the development of enabling methods and tools. For this transformation to succeed, we must address five fundamental questions: How do we deal with massive amounts of noisy data, both repository and streaming? Can we develop adequate methods for handling risk and uncertainty? How should we deal with issues of distributed data and computation? What is required for operational systems to take advantage of data in the same way as planning systems? How do we incorporate these changes into the emerging networked business world? I will discuss these questions and ideas we are developing at IBM to find answers.

## 212 Section on Bayesian Statistical Science Roundtable with Coffee (fee event)

Section on Bayesian Statistical Science

**Tuesday, August 8, 7:00 am–8:15 am**

### Can and Should We Teach Bayesian Inference in Stat 101?

❖ Jerome Reiter, Duke University, Box 90251, Durham, NC 27708, *jerry@stat.duke.edu*

***Key Words:*** Bayesian statistics, education, teaching

Bayesian methods are becoming more popular in statistical analysis. On one hand, Bayesian intervals and tests are easier for students to interpret than classical ones. On the other hand, they require students to understand conditional probability, which is generally challenging. Should we teach Bayesian methods in (algebra- or calculus-based) introductory statistics courses? During this roundtable, we will share experiences and opinions on this issue. The leader of the roundtable teaches both Bayesian and classical methods to students in algebra-based courses.

## 213 Section on Statistical Consulting Roundtable with Coffee (fee event)

Section on Statistical Consulting

**Tuesday, August 8, 7:00 am–8:15 am**

### Bioinformatics Consulting: Keeping up to Date

❖ Ann Hess, Colorado State University, Department of Statistics, Fort Collins, CO 80523, *hess@stat.colostate.edu*

***Key Words:*** bioinformatics, microarrays, consultant

The field of bioinformatics is experiencing rapid growth in both industry and academia. Statistical consultants need to stay current on methods and software related to bioinformatics and microarray data analysis. In addition, consultants often are asked to advise clients on the 'best way' to process the resulting data. This discussion will be an opportunity for both new and experienced consultants working in bioinformatics to share their experiences and ideas for staying up-to-date in this fast-moving field.

## 214 Section on Statistical Education Roundtable with Coffee (fee event)

Section on Statistical Education

**Tuesday, August 8, 7:00 am–8:15 am**

### Motivating the Math Major To Consider a Career in Statistics

❖ Carolyn Cuff, Westminster College, Hoyt Science Center, New Wilmington, PA 16172-0001, *ccuff@westminster.edu*

***Key Words:*** math major, careers in statistics, graduate school

We will discuss ideas for motivating the mathematics major to consider graduate school in statistics. What are factors in assuring their success in graduate school? How can we help them prepare for the transition? What are graduate schools looking for in matriculating students?

## 215 Section on Statistics and the Environment Roundtable with Coffee (fee event)

Section on Statistics and the Environment
**Tuesday, August 8, 7:00 am–8:15 am**

### Bayesian Applications in Environmental Science

❖ Eric P. Smith, Virginia Polytechnic Institute and State University, Department of Statistics, Blacksburg, VA , *epsmith@vt.edu*; Keying Ye, The University of Texas at San Antonio

***Key Words:*** environmetrics, standards assessment, Bayesian, pollution

Advances in Bayesian statistics have created opportunities for environmental applications. We will discuss successful applications of Bayesian statistics in environmental application, why we should use Bayesian statistics, and why the Bayesian approach might not meet the approval of federal agencies.

## 216 Section on Physical and Engineering Sciences Roundtable with Coffee (fee event)

Section on Physical and Engineering Sciences
**Tuesday, August 8, 7:00 am–8:15 am**

### Weibull Analysis

❖ Fritz Scholz, The Boeing Company, P.O. Box 3707, MC 7L21, Seattle, WA 98124-2207, *fritz.scholz@boeing.com*

***Key Words:*** Weibull analysis

This roundtable will discuss various aspects of Weibull analysis, especially from an industrial perspective. It will cover maximum likelihood estimation under various forms of censoring, no failures, confidence bound procedures, and software.

## 217 Section on Risk Analysis Roundtable with Coffee (fee event)

Section on Risk Analysis
**Tuesday, August 8, 7:00 am–8:15 am**

### Estimation of Health Risk from Fish Consumption Associated with Fishing in an Urban Industrial Setting

❖ Rose Ray, Exponent, Inc., 149 Commonwealth Drive, Menlo Park, CA 94025, *Rray@exponent.com*

***Key Words:*** angler survey, risk assessment, avidity, estimation, probability sample, health risk

A unique method of data collection and statistical analysis was devised to capture the site-specific information necessary to calculate the exposure factors needed to characterize the fish consumption pathway for recreational anglers in a human health risk assessment for the Passaic River. The survey used two methods to address the challenges of conducting a creel/angler survey in an urban and industrial setting with limited river access. While unique, the analytical method discussed here is based on accepted methods of interpreting survey data and basic laws of probability. The site-specific human exposure factors calculated and presented include size of angler and fish-consuming populations, annual fish consumption rate, duration of anglers' fishing careers, cooking methods for the fish consumed, and demographic information.

## 218 Section on Teaching Statistics in the Health Sciences Roundtable with Coffee (fee event)

Section on Teaching Statistics in the Health Sciences
**Tuesday, August 8, 7:00 am–8:15 am**

### Productivity versus Professionalism in Biostatistics

❖ Lemuel Moye, The University of Texas Health Science Center at Houston, School of Public Health, RAS Building E815, 1200 Herman Pressler, Houston, TX 77025, *moyelaptop@msn.com*

***Key Words:*** professionalism, ethics, coping skills

While productivity is and will be a fundamental attribute of the competent statistician, there are other core themes that must be allowed to develop, appear, and exert their influences. The presence of self-control and patience, of moral excellence and compassion, and of discipline and flexibility are as critical to the development of the junior scientist as is the acquisition of technical skills. The presence of these traits engenders collegiality, persuasive strength, responsibility, administrative diligence, influence, and vision (i.e., the qualities of charitable leadership).

## 219 Introductory Overview Lectures: Computer Experiments

The ASA, ENAR, IMS, SSC, WNAR, Section on Physical and Engineering Sciences
**Tuesday, August 8, 8:30 am–10:20 am**

### What Are Computer Experiments, and How Do We Design Them?

❖ William Notz, The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210-1247, *win@stat.ohio-state.edu*

*Key Words:* computer experiments, experimental design, space filling designs

This is the first of a series of two talks that will discuss some statistical approaches to analyzing data from computer experiments. We will begin by introducing the basic components of a computer experiment, emphasizing how computer experiments differ from physical experiments. We will then describe some common goals of computer experiments. Following this overview, we will discuss strategies for selecting an experimental design. We will present several popular space-filling designs including Latin hypercube designs, uniform designs, minimax designs, maximin designs, Sobol sequences, and designs based on multiple criteria. We will discuss software that is available for generating designs, with examples. We will also briefly describe other strategies based on statistical criteria, including sequential strategies.

### The Modeling and Analysis of Data from Computer Experiments

❖ Thomas Santner, The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210-1247, *tjs@stat.ohio-state.edu*

*Key Words:* computer experiments, gasp models

In this second talk we will consider statistical models that one might use to analyze data from computer experiments and that form the basis for sequential design strategies. We discuss the advantages and disadvantages of response surface modeling and then introduce the flexible class of Gaussian Stochastic Process (GASP) models. Because GASP models are popular in practice but unfamiliar to many statisticians we will provide an introduction to these models and give some intuition about the parameters in these models. We will discuss prediction and other inference and indicate how such models can be fit using statistical software. Examples will be used throughout.

## 220 JBES Invited Session

JBES-Journal of Business and Economic Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### On the Fit and Forecasting Performance of New Keynesian Models

❖ Frank Schorfheide, University of Pennsylvania, Department of Economics, McNeil Building, Room 525, 3718 Locust Walk, Philadelphia, PA 19104-6297, *schorf@ssc.upenn.edu*; Marco Del Negro, Federal Reserve Bank of Atlanta; Frank Smets, European Central Bank; Raf Wouters, National Bank of Belgium

This paper provides new tools for the evaluation of DSGE models and applies them to a large-scale New Keynesian DSGE model with price and wage stickiness and capital accumulation. Specifically, we approximate the DSGE model by a vector autoregression (VAR), then systematically relax the implied cross-equation restrictions. Let lambda denote the extent to which the restrictions are being relaxed. We document how the in- and out-of-sample fit of the resulting specification (DSGE-VAR) changes as a function of lambda. Furthermore, we learn about the precise nature of the misspecification by comparing the DSGE model's impulse responses to structural shocks with those of the best-fitting DSGE-VAR. We find the degree of misspecification in large-scale DSGE models is no longer so large that it prevents their use in day-to-day policy analysis.

## 221 Six Sigma: What's Missing ● ✪

Section on Quality and Productivity, Section on Statistical Education

**Tuesday, August 8, 8:30 am–10:20 am**

### Improving the Six Sigma Toolkit

❖ Gerald J. Hahn, GE Global Research/RPI, 1404 Orlyn Drive, Schenectady, NY 12309, *gerryhahn@yahoo.com*; ❖ Necip Doganaksoy, GE Global Research, 1 Research Circle, K14C35, Schenectady, NY 12309, *doganaksoy@crd.ge.com*

*Key Words:* continuous improvement, getting good data, planning investigations

Six Sigma has been in the spotlight for more than a decade. Its evolution has been a dynamic process. Although most of the initial emphasis of Six Sigma was on quality improvement in manufacturing, it is now being applied in key areas throughout the business and beyond what would traditionally be considered "quality." In this talk, we identify---based on our own experiences---statistical tools and concepts that did not receive the attention they deserve when Six Sigma was introduced originally but are now gaining recognition. Foremost among these is focusing on getting good data up front.

### Six Sigma: What Is Missing?

❖ William C. Parr, University of Tennessee, 335 Stokely Management Center, Department of Statistics, Operations and MS, Knoxville, TN 37996-0532, *wparr@utk.edu*

*Key Words:* Six Sigma, continuous improvement, black belts, change management, soft skills, non-statistical tools

The purpose of this presentation is to share my thoughts about the most important items missing from Six Sigma. I begin with a discussion of the problem of definition of Six Sigma, then given a provisional resolution to that problem and provide my of what needs to be added to the Six Sigma philosophy and toolbox to extend its lifetime. I will not try to give my view of a complete list of what is missing from Six Sigma, but instead will discuss those items I believe would have the most positive impact if added to the toolbox of Black Belts and other Six Sigma practitioners.

### Making Six Sigma Work: Statistics, Semi-Statistics, and Soft Stuff

❖ Blanton Godfrey, North Carolina State University, College of Textiles 3424, Box 8301, Raleigh, NC 27695, *blanton_godfrey@ncsu.edu*

Six Sigma has had a major impact on the manufacturing and service sectors in the United States in the past 20 years. Many of the proponents and leading companies have emphasized statistical methods, often with great effect. New software, the abundance of data, and the need to make fact-based decisions created an incredible opportunity for significant improvements using these methods. Most Six Sigma efforts soon incorporated the semi-statistical methods: FMEA, QFD, error-proofing, value-stream analysis, balanced scorecards, cause-and-effect diagrams. But the soft stuff remains a challenge, although, in many cases, it is what limits the impact of Six Sigma. Leadership, change management, and teamwork are critical parts of successful improvement initiatives of any kind. They are essential if we are to make Six Sigma achieve its potential.

## 222 Bayesian Methods in Computational Biology ● ✪

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Evolutionary Sparse Factor Modeling for Subpathway Identification and Characterization

❖ Mike West, Duke University, Durham, NC 27708-0251, *mw@isds.duke.edu*; Carlos Carvalho, Duke University; Quanli Wang, Duke University; Joseph Lucas, Duke University; Joseph Nevins, Duke University; Jeffrey Chang, Duke University

***Key Words:*** Bayesian factor analysis, gene pathways, cancer genomics, Bayesian computation, sparsity priors, pattern profiling

I will discuss high-dimensional, sparse, latent-factor models in exploratory and predictive analysis in genomics. The talk will cover methods of evolutionary stochastic computation developed for analysis of complex, large-scale, and very sparse latent factor models---models we have developed and use in a range of studies of biological pathway elucidation; general modeling questions, including the nature, mathematical representation, and relevance of sparsity; computational/simulation-based innovations; and examples in studies of subpathways of the Rb/E2F complex "pathway" that plays key roles in most human cancers. This work is collaborative with colleagues and students involved in the NCI Integrated Cancer Biology Program at Duke University.

### Bayesian Inference for Biochemical Network Dynamics

❖ Darren J. Wilkinson, University of Newcastle upon Tyne, School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU UK, *d.j.wilkinson@ncl.ac.uk*

***Key Words:*** Markov, systems biology, diffusion, modelling, Langevin, MCMC

This talk will give an overview of one of the key problems in the new science of systems biology: inference for the rate parameters underlying complex stochastic kinetic biochemical network models, using partial, discrete, and noisy time-course measurements of the system state. The basic problem will be introduced, highlighting the importance of stochastic modeling for effective estimation. Then, a range of approaches to Bayesian inference will be reviewed and compared. Some approaches recognize the discrete nature of the underlying molecular dynamics. Other techniques use a diffusion approximation to the true underlying process in order to give a nonlinear, multivariate stochastic differential equation (SDE) representation. Inference for this multivariate SDE can be carried out using MCMC techniques similar to those adopted for similar models in econometrics.

### Bayesian One-Way and Two-Way Clustering Methods for Genomic Data

❖ Jun Liu, Harvard University, 1 Oxford Street, Cambridge, MA 02138, *junliu@fas.harvard.edu*

***Key Words:*** Gibbs sampler, microarray, clustering, smoothing spline, two-way clustering

We will present Bayesian methods for time-course microarray data based on smoothing spline, nontime-course data based on Gaussian profiles, and two-way clustering based on the plaid model introduced by Owen and Lazzeroni.

### Statistical Methods for Motif Detection Incorporating Structural Features of DNA

❖ Mayetri Gupta, The University of North Carolina at Chapel Hill, Department of Biostatistics, McGavran-Greenberg Hall, Chapel Hill, NC 27599, *gupta@bios.unc.edu*

***Key Words:*** gene regulation, motif discovery, generalized hidden Markov models

Gene regulation is controlled by transcription factors (TFs) binding to specific sites (motifs) on the genome. Motif patterns tend to be short and degenerate, and similar patterns often occur outside protein-bound regions, leading genomic sequence-based motif prediction methods to have high false positive rates. Recently, new experimental methods studying genome-wide protein-DNA interactions suggest that packaging of chromatin is strongly associated with binding of TFs. We propose a Bayesian statistical model and a Monte Carlo method to infer nucleosome positioning from high density genome tiling arrays, and demonstrate that using the chromatin structure information significantly improves TF binding site predictions in an yeast genome example.

## 223 Alternative Approaches for Sample Size Planning ●

Section on Statistical Consulting, Section on Statistical Education

**Tuesday, August 8, 8:30 am–10:20 am**

### Estimating the Minimally Significant Difference for a Clinical Study

❖ Robert A. Parker, Harvard School of Public Health, Center for Biostatistics in AIDS Research, 651 Huntington Avenue/FXB-609, Boston, MA 02115, *rparker@sdac.harvard.edu*

***Key Words:*** sample size estimation, alternative hypothesis, clinically important difference, detectable difference

Although a number of factors are involved in calculating a sample size for a clinical study, perhaps the most critical issue is the difference to be detected, sometimes termed the "minimally significant difference" (MSD). This is the alternative hypothesis used for the study design. In practice, sample sizes often are calculated iteratively and assumptions modified---particularly the MSD---until the sample size is acceptable. I present a somewhat more objective method for determining an appropriate MSD for a study. I argue the MSD should be related to the overall potential of the study to impact clinical practice, both in terms of the potential number of future individuals impacted by the results of the study and the practical implications of such a change. Several studies in HIV/AIDS are used to illustrate the underlying ideas.

### A Completely Different Approach to Sample Size Planning

❖ Peter Bacchetti, University of California, San Francisco, 185 Berry Street, Lobby 4, Suite 5700, San Francisco, CA 94107-1762, *peter@biostat.ucsf.edu*; Charles E. McCulloch, University of California, San Francisco; Mark R. Segal, University of California, San Francisco

***Key Words:*** sample size, power, cost efficiency, innovation, study design, research funding

We propose a new approach for choosing sample size based on cost efficiency, the ratio of a study's scientific and/or practical value to its total cost. We show that choosing sample size to minimize cost per subject (which does not require projecting the study's value) never falls short of the most cost-efficient sample size and can give smaller sample sizes than traditional methods. An alternative, minimizing total cost divided by the square root of sample size, never falls short for innovative studies, often performs well, and is justifiable in other cases. Notably, for study value assumed to be proportional to power at a given alternative and total cost a linear function of sample size, this choice is more cost-efficient than the sample size producing 90% power. In many situations, these methods are easier, more reliable, and better justified than current conventional approaches.

### Sample Size and the Value of Information

❖ Kimberly M. Thompson, Harvard School of Public Health, 677
Huntington Ave., 3rd Floor, Boston, MA 02115, *kimt@hsph.harvard.edu*

***Key Words:*** value of information, Bayes theorem, decision analysis, clinical trials

Selecting an appropriate sample size represents a choice that leads to real trade-offs. Smaller sample sizes mean lower costs and quicker results, but they run the risk of not producing a significant result. In contrast, larger sample sizes mean longer studies, higher costs, and potentially costly delays in getting lifesaving interventions into mainstream use. In spite of an illusion of objectivity in most statistical analyses of clinical trials, subjectivity exists in any trial. This talk presents a value-of-information approach to data collection and emphasizes the need for statisticians to recognize the values that exist under seemingly value-free calculations. The talk discusses the challenges to broader implementation of value-of-information approaches and provides insights from prior studies.

## 224 Statistical Methods for Integrative Genomics ● ✪

Biometrics Section, WNAR, ENAR
**Tuesday, August 8, 8:30 am–10:20 am**

### Integrative Correlation: a Gene-Level Measure of Cross-Study Reproducibility

❖ Leslie Cope, Johns Hopkins University, Sidney Kimmel
Comprehensive Cancer Center, MD , *cope@jhu.edu*; Liz Garrett-Mayer, Johns Hopkins University; Edward Gabrielson, Johns Hopkins University School of Medicine; Giovanni Parmigiani, Johns Hopkins University

***Key Words:*** microarray, reproducibility, correlation, cross-study

Results from a combined analysis of two or more microarray experiments can be improved by concentrating the analysis on those genes, which are well measured in all experiments. The integrative correlation coefficient is a gene-specific measure of cross-study reproducibility. Integrative correlation has proved itself in many applications, facilitating the joint analysis of diverse gene expression datasets. Unsupervised and fully data-driven, it is related closely to canonical correlation analysis, and many of the statistical properties of the measure can be derived by exploiting that relationship. We describe and interpret the integrative correlation coefficient, illustrating its use on expression microarray data. We present key statistical properties, including an appropriate null distribution and tight upper-bounds on the coefficients.

### A Bayesian Model for Cross-Study Differential Expression

❖ Andrew Nobel, The University of North Carolina at Chapel
Hill, NC 27599-3260, *nobel@email.unc.edu*; Haakon Tjelmeland, Norwegian University of Science and Technology; Rob Scharpf, Johns Hopkins University; Giovanni Parmigiani, Johns Hopkins University

***Key Words:*** microarray, cross-study, Bayesian, differential expression, gene expression

We propose and analyze a Bayesian method for cross-study analysis of differential expression in two or more gene expression studies. The method is based on a 'joint' hierarchical model of the data in each of the available studies. For every gene, the model includes a binary parameter indicating whether the gene is differentially expressed. Given data, the probability that a gene is differentially expressed is evaluated through the posterior expectation of its associated indicator, which is estimated via MCMC. We apply the method to multiple lung cancer datasets and simulated data derived from the subset of adenocarcinomas. For the simulated data, we compare the posterior mean estimates of differential expression to other cross-study methods using ROC curves and a related measure based on false and missed discovery rates

### Statistical Methods for Analysis of Copy Number and Expression Transcript Data

❖ Debashis Ghosh, University of Michigan, School of Public Health,
Department of Biostatistics, 1420 Washington Heights, Ann Arbor, MI 48109/2029, *ghoshd@umich.edu*

***Key Words:*** chromosome, DNA microarray, machine learning, cancer

Recently, there have been oncologic studies in which both gene expression and copy number have been assessed on the same samples. Based on these data, we can integrate the two types of data to determine candidate events for chromosomal aberrations. We propose procedures for inferring candidate gene fusion events, determining the presence of cis- and trans-dosage effects and testing the hypothesis that genomic instability leads to evasion of apoptotic pathway in cancer. A variety of cancer studies will be considered to illustrate the various methodologies.

### Integrative Homologous Regulation Analysis between Mouse and Human

❖ Jae K. Lee, University of Virginia, Department of Public Health
Sciences, Hospital West Complex, Room 3181, P.O.Box 800717, Charlottesville, VA 22908-0717, *jaeklee@virginia.edu*

***Key Words:*** homology analysis, integrative correlation analysis, local pooled error test, melanoma metastasis, microarray gene expression

We have analyzed microarray datasets from a cross-species melanoma study in which progressively metastatic melanomas were induced experimentally from a nonmetastatic human melanoma cell line and a mouse melanoma line by repeatedly selecting cells with the greatest metastatic potential. We performed homology search, LPE testing, and integrative correlation analyses to discover cross-species co-regulation patterns. We found these analyses resulted in many co-regulated genes between the two microarray datasets, implying that highly concordant gene regulation networks of melanoma metastasis are present between human and mouse. This cross-species investigation provided a number of interesting gene sub-clusters, including ones reported in the original study.

## 225 Network Visualization

Section on Statistical Graphics, Section on Statisticians in Defense and National Security

**Tuesday, August 8, 8:30 am–10:20 am**

### Scaling up Graph Visualization

❖ Stephen North, AT&T Labs-Research, Room D239, 180 Park Ave., Bldg 103, Florham Park, NJ 07932, *north@research.att.com*

***Key Words:*** graph visualization, large graphs

Visualization is essential for exploring and analyzing network data in many fields, and scale is one of the greatest challenges of the field. We will show real world examples, introducing visualization algorithms and tools that have been effective on networks of various sizes. In particular, we will show results from recent work with Gansner and Koren on a topological (structural) zooming viewer for very large graphs, graph stream filters, and improved optimization methods that help to cope with scale.

### Scalable Drawing of Trees and Graphs

❖ Tamara Munzner, The University of British Columbia, 2366 Main Mall, Vancouver, BC V6T 1Z4 Canada, *tmm@cs.ubc.ca*

***Key Words:*** information visualization, graph drawing

We present several scalable, interactive visualization systems for large real-world tree and graph datasets. The multilevel TopoLayout framework provides high-quality layouts of a large class of graphs, decomposing the graph into pieces based on topological structure and using layout algorithms appropriate for each piece. The H3 system uses 3D hyperbolic geometry to show a large area of interest around a changeable focus point, handling quasi-hierarchical graphs of more than 300,000 nodes. The TreeJuxtaposer system provides side-by-side visual comparison of phylogenetic trees, handling more than 4 million nodes. It uses the Accordion Drawing visualization technique, which combines a stretch-and-shrink navigation metaphor with the guaranteed visibility of marked areas.

### Visualizing Evolving Graphs by Simultaneous Embeddings

❖ Stephen G. Kobourov, University of Arizona, Department of Computer Science, 1040 E 4th St, Tucson, AZ 85721, *kobourov@cs.arizona.edu*

***Key Words:*** graph drawing, information visualization, evolving graphs

Problems in simultaneous graph visualization involve the layout of multiple related graphs. A series of related graphs may arise from one relation between a set of objects as it evolves through time or from several relationships defined on the same set of objects. In simultaneous embedding, nodes are placed in the exact same locations in all the graphs, and a series of graphs is simultaneously embeddable if it is possible to find node locations that yield straight-line, crossing-free drawings for each of the graphs. We present polynomial time algorithms for simultaneous embedding of various classes of planar graphs and prove some classes of graphs cannot be embedded simultaneously. Further, we present a near-linear time algorithm for visualizing graphs that evolve through time and demonstrate its application to problems in software engineering and databases.

## 226 Teaching Biostatistics to Health Care Professionals without Equations ●

Section on Teaching Statistics in the Health Sciences, Section on Statistical Education, Section on Statistical Graphics

**Tuesday, August 8, 8:30 am–10:20 am**

### Teaching Research Design to Health Sciences Professionals

❖ Stephen W. Looney, LSU Health Sciences Center, 11888 Longridge Ave., Apt. 1007, Baton Rouge, LA 70816-8929, *sloon1@lsuhsc.edu*

***Key Words:*** clinical research, epidemiology, observational studies

In this presentation, we describe a course in research design originally developed for faculty members in the health sciences. The course was taught under the title “Cohort Studies”; however, most of the course content applies to research design in general (especially the design of observational studies) and could be adapted for almost any curriculum intended for health sciences faculty members or other health professionals. We will describe several of the modules contained in the course and discuss how basic concepts of research design can be communicated effectively to health sciences professionals with minimal use of formulas. The modules that will be described include Selecting the Population and Sample to Study, Measurement Issues, The Use of Questionnaires, Preparation of Data for Analysis, Confounding, Bias, and Ethical Issues.

### Teaching Statistical Concepts through Visualization

❖ Edward H. Ip, Wake Forest University School of Medicine, Room 319, MRI Building, Medical Center Blvd, Winston-Salem, NC 27157, *eip@wfubmc.edu*; Gary Wolgast, Wake Forest University School of Medicine

***Key Words:*** statistics education, visualization, concept learning, heat map, longitudinal data, Java

Teaching statistical concepts to students who do not have a strong quantitative background can be a challenging task. This research explores the potential of using interactive visualization tools to help students in learning statistical concepts. The following examples will demonstrate the use of a visualization tool in illustrating specific concepts commonly encountered in biostatistics courses: a java-based interactive heat map program developed by the authors illustrates the concept of “overlay” in teaching ANOVA; a graphical interface, Xcelsius, illustrates the teaching of linear models through a dashboard of interactive dials; a commercial visualization tool, Tableau, illustrates the analytic issues involved in the analysis of longitudinal data through the use of multiple panels, colors, shapes, and sizes.

### Effective Teaching of Applied Biostatistics for Clinicians Enrolled in a Research Training Program

❖ Jodi Lapidus, Oregon Health & Science University, Division of Biostatistics, Public Health, and Preventive Medicine, 3181 SW Sam Jackson Park Rd, CB-669, Portland, OR 97239, *lapidusj@ohsu.edu*

***Key Words:*** introductory biostatistics, health professionals, health science training programs

The human investigations program and the master's program in clinical research are funded at Oregon Health & Science University by a K--30 training grant from NIH. Students primarily are clinicians (residents, fellows, and junior faculty) interested in launching research careers. Increasing enthusiasm, appreciation, and competency in biostatistics among health sciences professionals requires biostatistics faculty members to examine, and potentially alter, their teaching styles and course content. The 14-week introductory biostatistics course for these programs includes case-based examples from recent literature as motivators and a combination of examples, graphs, web applets, and laboratory data analysis sessions. Use of formulae and equations are kept to a minimum. We will discuss lessons learned and outline successful methods that can be adopted by others.

## 227 Statistical Issues in Diagnostic Medicine ●

Section on Statistics in Epidemiology, Biometrics Section, Section on Health Policy Statistics, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Evaluating the Predictiveness of a Continuous Marker

❖ Margaret S. Pepe, Fred Hutchinson Cancer Research Center/University of Washington, 1100 Fairview Ave., N., M2-B500, Seattle, WA 98109-1024, *mspepe@u.washington.edu*; Ying Huang, University of Washington; Ziding Feng, Fred Hutchinson Cancer Research Center

***Key Words:*** predictiveness curve, classification, sensitivity, receiver operating characteristic, explained variation, r squared

Various measures quantifying the predictiveness of a continuous marker for a binary outcome have been proposed, including the area under the ROC curve, proportion of variation explained, correlation of outcome with risk, Brier score, and concordance index. We argue that the predictive capacity of a marker has to do with the population distribution of risk, given the marker, and suggest a graphical tool---the predictiveness curve---to display this distribution. This provides a common scale to compare markers that may not be comparable on their original scales. Some existing measures of predictiveness are shown to be summary indexes derived from the predictiveness curve. Other measures are concerned with the performance of decision rules based on the marker, which we argue is different from its predictiveness. Applications to cancer biomarkers will be discussed.

### Double-Semiparametric ROC Regression Analysis

❖ Xiao-Hua Andrew Zhou, University of Washington, Box 357232, 1705 Pacific Street NE, Seattle, WA 98195-7232, *azhou@u.washington.edu*; Huazhen Lin, University of Washington

***Key Words:*** non-parametric, ROC curves, diagnostic accuracy, transformation models

ROC regression methodology offers an opportunity to investigate how factors affect test accuracy. For directly modeling ROC curves, both parametric methods---where the link and baseline functions are specified---and semiparametric methods---where the link function is specified but the baseline function is unspecified---have been developed. However, the misspecification of either the link or the baseline function can lead to substantial bias for the ROC curve estimates. In this talk, we extend the existing direct ROC regression models to allow arbitrary nonparametric link and baseline functions. We show the proposed estimators for the regression parameters and ROC curves are asymptotically normal and consistent with the parametric convergent rate, n^{-1/2}. We illustrate our approach with a real dataset.

### Evaluation of Diagnostic Tests in Studies with Verification Bias

❖ Marina Kondratovich, U.S. Food and Drug Administration, HFZ-550, 1350 Piccard Drive, Rockville, MD 20850, *Marina.Kondratovich@fda.hhs.gov*

***Key Words:*** diagnostic accuracy, sensitivity and specificity, verification bias, multiple imputation, types of missingness

It is well-known that in studies with verification bias, ratios of sensitivities and false positive rates are unbiased estimates. We will discuss whether the information about these ratios is enough to make a conclusion about effectiveness of the test under investigation. We also will discuss multiple imputation techniques for obtaining unbiased estimates of the tests' diagnostic accuracies and different mechanisms of missing disease status, missing completely at random, and missing at random. These types of missingness will be demonstrated by examples from the studies of comparison of the prostate-specific antigen assay and digital rectal exam for detection of prostate cancer and others. If different types of missingness is not addressed properly, the adjustment for the verification bias can produce biased estimates of the diagnostic accuracies of the tests.

## 228 Statistical Responses to Legislation: the Privacy Rule ● ✪

Committee on Privacy and Confidentiality, Section on Health Policy Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### The Statistician's Role in Developing the HIPAA De-identification Standard

❖ Alvan O. Zarate, National Center for Health Statistics, 3311 Toledo Road, Room 7116, Hyattsville, MD 20782, *aoz1@cdc.gov*

***Key Words:*** disclosure, legislation, confidentiality

During the late 1990s, a number of laws were proposed to protect the privacy of health records. All included a definition of "identifiability" that was very general. However, when the U.S. Department of Health and Human Services assumed responsibility for writing a Privacy Rule in accordance with provisions of the Health Insurance Portability and Accountability Act (HIPAA), two approaches were taken---one a very detailed statistical approach and the other a relatively simple, nonstatistical procedure. This paper describes these standards, the process by which they were adopted, and the role of statisticians in their development. As the Privacy Rule is cited widely now as a standard for protecting electronic and other information from re-disclosure, it is important they be understood and their implications discussed.

### The Federal Statisticians' Response to the HIPAA Privacy Rule

❖ Jacob Bournazian, Energy Information Administration, 1000 Independence Ave., SW, EI-70, Washington, DC 20585, *jacob.bournazian@eia.doe.gov*

***Key Words:*** confidentiality, disclosure avoidance, disclosure risk

The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule requires most "covered entities" who provide data to take reasonable steps to protect the confidentiality of health care information they possess. Statistical Policy Working Paper 22 (SPWP 22) is cited in the preamble of the Privacy Rule for guidance on what constitutes "reasonable steps" to protect the confidentiality of the data. SPWP 22 was released originally in 1994 and revised in 2005 by the Federal Committee on Statistical Methodology to reflect current practices by the federal agencies for protecting confidential data. This paper discusses the specific areas in which the HIPAA privacy rule impacted the revisions to SWSP 22 and the issues raised in considering reasonable steps to protect the confidentiality of health data.

### Statistical Deidentification and the HIPAA Rule

❖ Patrick Baier, National Opinion Research Center, Washington, DC 20011, *districtline@gmx.de*

***Key Words:*** HIPAA, de-identification

In order to protect the privacy of patients, the Health Insurance Portability and Accountability Act places restrictions on the use and sharing of patient data by members of the health care industry (e.g., hospitals, pharma companies, or marketing firms). In many situations, only "de-identified" data can be used. We present methods that have been used in practice to estimate the risk of patient re-identification and discuss solutions for keeping this risk below an acceptable minimum. We will focus on statistical safeguards that ensure patient privacy while retaining the ability for companies to conduct useful data analyses. Other safeguards (e.g., procedural, legal, etc.) may be touched briefly.

## 229 Recent Developments in Nonparametric Survival Analysis Methods

ENAR, Biometrics Section, WNAR, Section on Nonparametric Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### Partially Linear Hazard Regression for Multivariate Survival Data

❖ Jianwen Cai, The University of North Carolina at Chapel Hill, Chapel Hill, NC , *cai@bios.unc.edu*; Jianqing Fan, Princeton University; Jiancheng Jiang, Princeton University; Haibo Zhou, The University of North Carolina at Chapel Hill

***Key Words:*** local pseudo-partial likelihood, marginal hazard model, martingale, multivariate failure time, partially linear, profile pseudo-partial likelihood

Partially linear hazard regression models for multivariate survival data are studied. A profile pseudo-partial likelihood estimation method is proposed under the marginal hazard model framework. The estimation on the parameters for the linear part is accomplished via maximization of a profile pseudo-partial likelihood. This enables one to get root-n consistent estimators of the parametric component. Asymptotic normality is obtained for the estimates of both the linear and the nonlinear parts. The new technical challenge is that the nonparametric component is indirectly estimated via its integrated derivative function from a local polynomial fit. An algorithm of fast implementation of our proposed method is presented. Consistent standard error estimates are proposed. Simulations are conducted to examine the performance of the proposed method. A real dataset is analyzed for illustration.

### A General Imputation Methodology for Nonparametric Regression with Censored Data

❖ Daniel Rubin, University of California, Berkeley, Graduate Group in Biostatistics, Berkeley, CA 94709, *drubin@stat.berkeley.edu*; Mark van der Laan, University of California, Berkeley

***Key Words:*** nonparametric regression, survival analysis, double robustness, imputation, least squares estimators, locally weighted average estimators

In this talk, we consider the random design nonparametric regression problem when the response variable is subject to a general mode of missingness or censoring. In such problems, the response variable often must be transformed or truncated in some way for the regression function to remain identifiable. We present a general methodology for imputing responses with the property of double robustness, in that the method works well if either a parameter of the full data distribution (covariate and response distribution) or a parameter of the censoring mechanism is well approximated. We give general results for the method when the imputed responses are entered into commonly used nonparametric regression procedures, including least squares estimators, complexity-regularized least squares estimators, penalized least squares estimators, locally weighted average estimators, and cross-validated estimators.

### Kernel-Smoothed Profile Likelihood Function in Accelerated Failure Time Model

❖ Donglin Zeng, The University of North Carolina at Chapel Hill, Department of Biostatistics, CB7420, Chapel Hill, NC 27599-7420, *dzeng@bios.unc.edu*; Danyu Lin, The University of North Carolina at Chapel Hill

***Key Words:*** accelerated failure time, kernel smoothing, profile likelihood, asymptotic efficiency, empirical process

This paper proposes an approximate maximum-likelihood estimation for estimating parameters in an accelerate failure time model. Specifically, the estimate of the regression coefficient is obtained by maximizing a kernel-smoothed profile likelihood function. The consistency, asymptotic normality, and efficiency of the estimates are proved. Simulation studies are conducted to examine the small-sample performance of the proposed estimate. Real data are analyzed using the proposed method.

### Nonparametric Analysis of Multivariate Competing Risks Data

❖ Jason P. Fine, University of Wisconsin-Madison, Departments of Statistics and Biostatistics, K6-420, CSC, 600 Highland Avenue, Madison, WI 53792, *fine@biostat.wisc.edu*

***Key Words:*** bivariate association measures, empirical processes, uniform convergence properties, cause-specific incidence and hazard functions, nonparametric estimation

While nonparametric analyses of bivariate failure times have been studied widely, nonparametric analyses of bivariate competing risks data have not been investigated. Such analyses are important in familial association studies, where multiple interacting failure types invalidate nonparametric analyses for independently censored clustered data. We develop nonparametric estimators for the bivariate cause-specific hazards function and the bivariate cumulative incidence function, natural extensions of their univariate counterparts, which make no assumptions about the dependence of the risks. The estimators are shown to be uniformly consistent and converge weakly to Gaussian processes. Summary association measures are proposed and yield formal tests of independence in relative pairs. An illustrative analysis of dementia onset associations in the Cache County Study on Aging will be presented.

### An Additive Modulated Renewal Process

❖ Dorota Dabrowska, University of California, Los Angeles, Department of Biostatistics, Los Angeles, CA 91405, *dorota@ucla.edu*

The additive hazard regression model provides a common alternative to the proportional hazard regression model. I will discuss a generalization to additive modulated renewal processes and estimation in the presence of censoring.

## 230 Client-Oriented Perspectives on Small-Area Estimation ●

Section on Survey Research Methods, Section on Physical and Engineering Sciences

**Tuesday, August 8, 8:30 am–10:20 am**

### Characterization of Cost Structures, Perceived Value, and Optimization Issues in Small-Domain Estimation

❖ John L. Eltinge, Bureau of Labor Statistics, Office of Survey Methods Research, PSB 1950, 2 Massachusetts Avenue NE, Washington, DC 20212, *Eltinge.John@bls.gov*

***Key Words:*** adoption and diffusion of technology, components of risk, constraints, data quality, multiple comparisons, total survey error

This paper characterizes tradeoffs among cost structures, data quality, perceived value, and optimization issues in small-domain estimation. Four topics receive principal attention. First, we highlight several classes of direct and indirect costs incurred by the producers and users of small-domain estimates. Second, we use the Brackstone framework of six components of data quality to review statistical properties of direct design-based and model-based estimators for small domains and to link these properties with related components of risk. Third, we explore data users' perceptions of the value of published small-domain estimates and costs incurred through decisions not to publish estimates for some subpopulations. Fourth, we emphasize the importance of constraints in the administrative development and implementation of small-domain estimation programs.

### A Small-Area Estimation Approach in Reducing Survey Costs

❖ Partha Lahiri, University of Maryland, 1218 Lefrak Hall, JPSM, College Park, MD 20742, *plahiri@survey.umd.edu*; Paul D. Williams, National Center for Health Statistics

***Key Words:*** multi-level model, Bayesian method, EBLUP, mean squared error

The demand for producing statistics at subnational levels is increasing steadily at a time when survey agencies are looking for ways to reduce total survey costs to meet fixed budgetary requirements. We shall discuss various potential ways to reduce survey costs by using state of the art small area estimation techniques.

### Some Problems and Possible Solutions in Developing a Small-Area Estimation Product for Clients

❖ Avinash C. Singh, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A 0T6 Canada, *avi.singh@statcan.ca*

***Key Words:*** client input, data gaps for small areas, statistical tools and products, best unbiased prediction via filtering

It is argued that taking on the challenge of developing a product (a menu-driven software system with user interface for client self-use) unlike a tool (requiring statistical skills to operate) is mutually beneficial for both researchers (for a disciplined R&D, setting priorities, and client input and validation at early stages) and clients (for understanding and confidence about the product and to plan for data gaps and customer needs in advance). We describe some problems and possible solutions for a SAE-product that enhances the existing methods by allowing for smoothing of the error covariance, collapsing of areas with little or no samples, multivariate modeling with combined longitudinal and cross-sectional data, approximate self-benchmarking to estimates of subgroups of areas or target variables, and Kalman filtering for computational efficiency and innovations for diagnostics.

## 231 Missing Data in Clinical Trials: Can We Do Better Than LOCF? ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Is It Better To Use Random Effects Models in Analysis of Repeated Binary Responses with Missing Data?

❖ Guanghan Liu, Merck Research Laboratories, 785 Jolly Road, Blue Bell, PA 19422, *guanghan_frank_liu@merck.com*

***Key Words:*** repeated binary response, dropouts, random effects, GEE, LOCF

In analysis of repeated binary responses with missing data, conventional methods such as the last observation carried forward (LOCF) approach can be biased in both parameter estimates and hypothesis tests. Generalized estimating equation (GEE) method is valid when missing data are missing complete at random. However, when data are missing at random, analyses based on GEE can be biased. Several random effect--based likelihood or quasi-likelihood methods have been proposed in the literature to overcome the drawbacks. These methods are available in many statistical software packages, including SAS version 9. In this presentation, we will evaluate the random effects models with full- or quasi-likelihood methods in the analysis of repeated binary response data. Simulations are used to compare the results and performance among these methods under different assumptions.

### An Analytic Road Map for Incomplete Longitudinal Clinical Trial Data

❖ Craig Mallinckrodt, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *cmallinc@lilly.com*; Shuyi Shen, Eli Lilly and Company; Caroline Beucnkens, Limburgs Universitair Centrum; Geert Molenberghs, Limburgs Universitair Centrum

***Key Words:*** missing data, longitudinal data, maximum likelihood, mixed effects models, multiple imputation

Practice has shifted away from ad hoc methods such as LOCF and complete case analyses to more principled likelihood- and Bayesian-based approaches for analyses of incomplete longitudinal clinical trial data. Actual clinical trial data will be used to illustrate how the attributes of likelihood- and Bayesian-based methods implemented under the missing-at-random framework can be exploited to choose a primary analysis and to develop complete analytic plans. Specifically, multiple imputation and likelihood-based mixed-effect model analyses in both restrictive and inclusive modeling approaches will be discussed. In addition, selection models and pattern mixture models will be used to assess sensitivity of results to departures from the MAR assumption. Local influence will be used as another sensitivity tool to assess the impact of influential subjects.

### Mixed-Effects Latent Curve Modeling of Incomplete Longitudinal Cognitive Data

❖ John J. McArdle, University of Southern California, Department of Psychology, Los Angeles, CA 90089, *jmcardle@usc.edu*

***Key Words:*** longitudinal data, dynamic structural equation modeling, cognitive aging, mixed-effects, latent growth-decline curve models, active trial

This study examines intra-individual differences in cognition by age using contemporary techniques of mixed-effects latent curve analysis with incomplete data. The research uses a representative national sample of N>10,000 individuals based on more than 25,000 interviews from the HRS-AHEAD (1993--2000). After fitting alternative forms of latent curve models, we find comparative advantages in using incomplete age-at-testing---instead of time or date-of-testing for group and individual changes---and nonlinear age changes in cognitive data with clear cognitive declines between ages 60--75, but twice as much decline between 75--90. These basic methods are then used in a reanalysis of the ACTIVE randomized trial for the retention of cognitive functioning in older ages. The incomplete data in ACTIVE are related to self-selection by cognitive status, and the results impact training.

### Scope of the Problem, Options Available, and Empirical Performance of Techniques for Handling Missing Data in Randomized Clinical Trials: Focus on Weight Loss Trials

❖ David B. Allison, The University of Alabama at Birmingham, 317E RPHB, 1665 University Boulevard, Birmingham, AL 35294-0022, *dallison@uab.edu*

I will provide a brief overview of the theoretical challenge posed by missing data in randomized clinical trials (RCTs), the methods available for analyzing RCTs with missing data, and their a priori specifiable advantages and disadvantages. Using obesity weight loss RCTs as an example, I will present empirical evidence that quantifies the magnitude of the missing data problem, the approaches typically used for addressing it, and differences in results obtained with different methods. Based on this, I will offer inferences about the relative performance of methods in practice and interpretive heuristics.

## 232 Applications in Seasonal Adjustment

Business and Economics Statistics Section

**Tuesday, August 8, 8:30 am–10:20 am**

### Balancing SA Series as a Complement to the Direct and Indirect SA of Series

❖ Benoit Quenneville, Statistics Canada, 120 Parkdale Ave., 17 RHC BSMD, Ottawa, ON K1A 0T6 Canada, *benoit.quenneville@statcan.ca*

***Key Words:*** X-12-ARIMA, seasonal adjustment, benchmarking, reconciliation, raking

It is possible with X-12-ARIMA to seasonally adjust (SA) an aggregate with either the direct or indirect approach. For example, a national total often is obtained indirectly from a first classification, such as an industry breakdown. However, when looking at a second classification, such as a regional breakdown, the indirectly obtained national total becomes fixed and the sum of the seasonally adjusted estimates from the second classification does not necessarily equal the national total. Unfortunately, nothing is available in X-12-ARIMA to impose a fixed aggregate on its components. We propose and illustrate simple methods based on constrained weighted least squares regression to restore the additivity. We also consider the case where the SA series must be benchmarked to annual control totals. The methods are designed to be easily programmable with the SAS statistical software.

### Computing Tasks Associated with Seasonal Adjustment at Statistics Canada: Methods and Software

❖ Susie Fortier, Statistics Canada, 120 Parkdale Ave., RH Coats Building, Ottawa, ON K1A 0Y6 Canada, *Susie.Fortier@statcan.ca*; Benoit Quenneville, Statistics Canada

***Key Words:*** X-12-ARIMA, raking, benchmarking, analytical tools

This paper discusses computing tasks associated with a typical seasonal adjustment process at Statistics Canada. The tasks include importing the data series into the process, preprocessing the series before seasonal adjustment, seasonal adjustment with X-12-ARIMA, post-processing the seasonally adjusted series, benchmarking individual series to annual control totals, raking/reconciliation of a set of series to restore aggregation constraints, trend-cycle estimation, graphs and analytical tools for the release of the seasonally adjusted series, and exporting the data series out of this process. We discuss how these computing tasks are related and illustrate this with the case of Statistics Canada International Travel Survey. We also highlight recent development in the statistical methodology of this overall process.

### Exploring Model-Based Seasonal Adjustment for Industry Employment Statistics

❖ Christopher Manning, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 4860, Washington, DC 20212, *manning.chris@bls.gov*; Jeffrey A. Smith, Bureau of Labor Statistics

***Key Words:*** seasonal adjustment, X-12-ARIMA, TRAMO/SEATS, X-12-SEATS

This study explores the use of ARIMA model-based seasonal adjustment in the Current Employment Statistics (CES) program and the implications it might have on seasonally adjusted CES series. Theoretically, ARIMA model-based seasonal adjustment has the advantage over the X-11 method by shaping its seasonal decomposition to the individual series while the X-11 method must choose from a finite number of built-in filters. We attempt to determine whether this theoretical advantage leads to actual benefits by examining CES employment estimates seasonally adjusted under the different methodologies. Our test will use X-12-ARIMA, TRAMO/SEATS, and X-12-SEATS. TRAMO/SEATS and X-12-SEATS are expected to differ mainly in model selection. The evaluation of the seasonally adjusted series will check for residual seasonality and compare stability diagnostics and smoothness statistics.

### Variance Estimation for Noise Components in Time Series from a Survey

❖ Daniell Toth, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 1950, Washington, DC 20212-0001, *toth.daniell@bls.gov*; Stuart Scott, Bureau of Labor Statistics

***Key Words:*** seasonal adjustment, sampling error, structural models, X-11

Models for economic time series of the form y=trend + seasonal + irregular typically assume each term is stochastic with a noise component. A fourth noise component enters the picture when the series is observed from a survey. Chen, Wong, Morry, and Fung (2003) compared method of moments and spectral estimates of "combined error" autocovariances in X-11 seasonal adjustment. This paper revisits the topic both with and without the use of external sampling error information. For comparison, we use simulated data generated from structural models---as done by Chen et al.---and sampling error models---suggested by the Bureau of Labor Statistics employment and unemployment series. We investigate whether prior smoothing in this system adds stability to the estimation. We also address selecting a "cutoff" value for the number of autocovariance terms needed.

## 233 New Methods for Testing Sensitive Questions Using the National Health Interview Survey ✪

Section on Survey Research Methods

**Tuesday, August 8, 8:30 am–10:20 am**

### Developing New Methods and Questions for Improving Response and Measurement on Sensitive Questions on the National Health Interview Survey

❖ Peter Meyer, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, *Pmeyer1@cdc.gov*; James Dahlhamer, National Center for Health Statistics; John Pleis, National Center for Health Statistics

***Key Words:*** survey methods, NHIS, health, economic measures, NCHS

This presentation is an overview of a study conducted by the National Center for Health Statistics. An experiment was fielded that evaluated alternative methods for asking income questions, response to proposed new wealth questions, and respondent willingness to provide partial social security numbers. The field test took advantage of oversampling on the National Health Interview Survey (NHIS). These "screener" households do not have at least one black, Hispanic, or Asian member and are usually screened out of the sample. For this study, the usually excluded households were used as a separate sample and asked the test items along with standard NHIS questions. The purpose of this study is to test methods for asking sensitive questions and determine the feasibility of using "screened out" samples in future research.

### Sampling and Oversampling in the National Health Interview Survey

❖ Chris Moriarity, National Center for Health Statistics, 200 Spring Ave., Takoma Park, MD 20912, *cdm7@cdc.gov*

***Key Words:*** sample survey, screening

The National Health Interview Survey (NHIS) is a continuous survey that collects health data using personal interviews. Changes have been made in the sample design for the NHIS (a "redesign") about every 10 years, using information from the previous decennial census. The 2006 NHIS is the first year of the most recent sample redesign. Beginning with the previous design---implemented in 1995---and continuing with the 2006 design, minority persons are oversampled using two mechanisms: variation in sampling rates in substrata and screening. Primary sampling units are partitioned into substrata based on population proportions of minority persons, and substrata with higher minority population proportions are sampled at a higher rate. A proportion of the sample cases are "screened" (i.e., the household interview stops if no eligible minority persons are on the household roster).

### Unfolding the Answers? Income Brackets and Income Nonresponse in the National Health Interview Survey (NHIS)

❖ John Pleis, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, *gzp4@cdc.gov*

***Key Words:*** nonresponse, income, brackets

Nonresponse to income-related survey questions is problematic and may lead to biased estimates. In the NHIS, respondents are first asked to provide the exact dollar amount of the family's income in the previous calendar year (nonresponse = 30%). Current follow-up questions based on income intervals have had minimal effect on lowering nonresponse. Using the NHIS screened-out sample from April-June of 2006, we present the results of alternative income questions using unfolding brackets. Respondents were randomly assigned to the existing or alternative method. Alternative methods for asking about the sources of income were used to assess whether item nonresponse for income could be reduced. Instead of asking about each source separately, a flashcard approach was used where families were asked about only the income sources of which they initially indicated receipt.

### Questions People Don't Like To Answer: Wealth and Social Security Numbers

❖ James Dahlhamer, National Center for Health Statistics, 3311 Toledo Road, Room 2117, Hyattsville, MD 20782, *fzd2@cdc.gov*; Peter Meyer, National Center for Health Statistics; John Pleis, National Center for Health Statistics

***Key Words:*** sensitive questions, data quality, item nonresponse

The National Health Interview Survey (NHIS) collects data on family income, but not wealth. Yet research has identified a strong, separate association between wealth and health status. Will people answer these sensitive questions? In addition, social security numbers are needed for linkage to databases such as the National Death Index, but suffer from high item nonresponse. Can we improve reporting? This paper presents early results from a field experiment designed to address these and other research questions. Quality indicators, such as item nonresponse and interview break-off rates, are used to evaluate a set of proposed new wealth questions, while the feasibility of collecting the last four digits of social security numbers (purported to be almost as useful and less sensitive than the nine-digit approach) is assessed. Implications for future surveys are discussed.

## 234 Frontiers in Demographic Coverage Measurement ● ✪

Social Statistics Section

**Tuesday, August 8, 8:30 am–10:20 am**

### Demographic Coverage Measurement: Can Information Integration Theory Help?

❖ Dean H. Judson, U.S. Census Bureau, 15000 Fort Trail, Accokeek, MD 20607, *dean.h.judson@census.gov*

***Key Words:*** information integration, demographic analysis, coverage measurement, combining information

Statisticians have developed many ways to assess census (or other registration) coverage. The standard, of course, is dual-system estimation. Related techniques, such as reverse record check and the "megalist" method, also have been proposed. These techniques generally use microdata, rather than aggregates; use actual records, rather than estimates of aggregates; and have a well-established statistical theory behind them that allows statistical inferences. In both Census 2000 in the United States and Census 2001 in the UK, however, it was demographic methods of assessing coverage that ultimately provided important clues that the census enumeration process had over- and under-coverage in particular areas. This paper discusses whether information integration theory can help link demographic and statistical methods in an overarching framework.

### Demographic Analysis in the UK Census: a Look Back to 2001 and Looking Forward to 2011

Owen Abbott, Office for National Statistics; ❖ Garnett Compton, Office for National Statistics, Room 4300W, Segensworth Road, Fareham, PO15 5RR UK, *garnett.compton@ons.gov.uk*

***Key Words:*** census, demographic analysis, coverage, quality assurance, administrative sources

The 2001 United Kingdom Census used dual-system estimates based on a large census coverage survey to assess and adjust for census under-enumeration. A number of demographic analysis techniques, based on administrative data sources, were then used to quality assure the 2001 Census results and to make final coverage adjustments in particular areas or subpopulations. This paper will outline the methods, data sources, and operational processes used in 2001 and look forward to 2011, specifically the demographic analysis techniques that will be considered for both coverage assessment and quality assurance. Furthermore, it will look at the blurring between coverage assessment/adjustment and quality assurance resulting from wider use of demographic techniques and the potential challenges this raises.

### Coverage Assessment and Adjustment in the 2011 UK Census

❖ Owen Abbott, Office for National Statistics, Room 4300W, Segensworth Road, Fareham, PO155RR UK, *owen.abbott@ons.gov.uk*

***Key Words:*** census, coverage, post-enumeration survey, administrative data

It is traditional that census coverage be measured and the outcome disseminated to users. Most census-taking countries undertake this, usually using some form of post-enumeration survey (PES). For the 2001 UK Census, the ONS instigated the one number census (ONC) project and embarked on a program to research coverage measurement and adjustment. For the 2011 Census, it is expected that ONS will improve upon the success of the ONC significantly and use it as a platform to develop improved coverage measurement and adjustment. This paper lays out the aims of the 2011 coverage assessment and adjustment project and the areas of research that will be addressed. Early ideas for key innovations and improvements, such as the potential for using external information (e.g., administrative data) and information collected in the census on visitors away from their usual residences, will be discussed.

### Using the Audit Trail Data To Evaluate the Quality of Collection of the Canadian National Longitudinal Survey of Children and Youth

❖ Bruno Lapierre, Statistics Canada, 120 Parkdale Ave., Tunney's Pasture, Main Bldg. 2500-A, Ottawa, ON K1A 0T6 Canada, *Bruno.Lapierre@statcan.ca*; Scott Meyer, Statistics Canada

***Key Words:*** audit trail, quality, collection

The Audit Trail is a Blaise programming option available for computer-assisted interviews. It provides a detailed picture of each interview by recording the sequence of screens visited and entries made by the interviewer, each with a time stamp. This information allows for a better understanding of the collection process. In this paper, we will present a description of the Audit Trail files and steps needed to convert the raw text files into a coherent usable SAS dataset. We will present details on how Audit Trail provides a window into the collection process (e.g., time spent on each question and module, edit trigger rates and interviewer's reactions, flow problems, and special events---help, change language, remarks). Finally, we will provide examples of what we have learned from the Audit Trail data for Cycle 6 of the Canadian National Longitudinal Survey of Children and Youth.

## 235 Climate, Weather, and Spatial-Temporal Models

Section on Statistics and the Environment, Section on Statisticians in Defense and National Security, Section on Bayesian Statistical Science, WNAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Models for Multivariate Spatial Lattice Data and Assessing Climate Change

❖ Stephan Sain, University of Colorado at Denver and Health Sciences Center, P.O. Box 173364, Campus Box 170, Denver, CO 80217-3364, *ssain@math.cudenver.edu*

***Key Words:*** Markov random field, conditional autoregressive model, hierarchical model

The U.S. Climate Change Science Program Strategic Plan has recognized the need for regional climate modeling to assess climate impacts. This is the focus of the newly formed North American Regional Climate Change Assessment Program that seeks to study a number of regional climate models. In this talk, we will discuss the details of the specification of a spatial hierarchical model based on a multivariate Markov random field (also referred to as a conditional autoregressive, or CAR, model) and take a close look at the output of a particular regional climate model that uses a "business as usual" scenario to model climate over the western United States. In particular, we seek to explore how the modeled climate---seasonal measures of temperature and precipitation in particular---changes over the next 50 years.

### Spatial Patterns of Global Climate Change Fields

❖ Reinhard Furrer, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, *rfurrer@mines.edu*; Reto Knutti, National Center for Atmospheric Research

***Key Words:*** climate change, spatial processes, hierarchical Bayes, large data sets

We present probabilistic projections for spatial patterns of future climate change using a multivariate Bayesian analysis. The methodology is applied to the output from 21 global coupled climate models used for the 4th Assessment Report of the Intergovernmental Panel on Climate Change. The statistical technique is based on the assumption that spatial patterns of climate change can be separated into a large-scale signal related to the true forced climate change and a small-scale signal from model bias and variability. The scales are represented via dimension reduction techniques in a hierarchical Bayes model. Posterior probabilities are obtained with a MCMC simulation. The posterior fields can be analyzed as such or down-scaled or weighted arbitrarily. For example, we show that 74% of the land areas are likely to warm by more than 2K by the end of the century (SRES A1B).

### Modeling Precipitation Network Data When Station Reporting Times Are Misaligned

❖ Jarrett Barber, Montana State University, Department of Mathematical Sciences, PO Box 172400, Bozeman, MT 59717-2400, *jarrett@math.montana.edu*; Alan E. Gelfand, Duke University; Douglas W. Nychka, National Center for Atmospheric Research

***Key Words:*** dynamic model, Gaussian spatial process, precipitation, temporally misaligned data

We present a two-stage stochastic model for precipitation events and amounts using data from a network of monitoring stations recording hourly or daily precipitation. Daily stations report accumulations for nominal 24-hour periods, but these periods differ from station to station. That is, the reporting periods are misaligned. In addition to temporal correlation, we expect spatial correlation among events and/or amounts at nearby stations, but temporal misalignment masks spatial correlation to some degree. Our model uses a pair of conditionally independent latent Gaussian spatial processes within a dynamic framework to align data to common reporting times across stations effectively and to recover information masked by misalignment. We use the model to reconstruct an aligned daily precipitation network data product.

### Spatial and Temporal Models for Evaluating IPCC Climate Model Outputs

❖ Mikyoung Jun, Texas A&M University, Department of Statistics, 3143 TAMU, College station, TX 77843, *mjun@stat.tamu.edu*; Douglas W. Nychka, National Center for Atmospheric Research; Reto Knutti, National Center for Atmospheric Research

***Key Words:*** climate models, model bias, spatio-temporal process, spherical process, multivariate processes

There are extensive efforts to develop climate models to study climate change. We have about 20 climate models from the Intergovernmental Panel on Climate Change. The previous works with climate model outputs commonly assume climate model outputs are random samples from a symmetric distribution centered around the true climate. One of the most interesting problems to climate scientists and modelers is verifying the bias of climate models and how the biases of different models are correlated. We propose modeling the climate model outputs as spatio-temporal processes on sphere x time and focusing on the spatial and temporal covariance structure. We propose cov models as well as cross-cov models for pairs of climate model outputs. Then, we quantify model biases and classify climate models with common biases.

### A Hierarchical Bayesian Spatio-Temporal Model for Tropospheric Carbon Monoxide

❖ Anders Malmberg, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, *andersm@ucar.edu*

***Key Words:*** hierarchical, Bayesian, space-time, remote sensing

Tropospheric Carbon Monoxide (CO) is an important trace gas, air quality indicator, and one of the few atmospheric species that can be remotely sensed from space. The Measurement of Pollution in the Troposphere (MOPITT) instrument, on board the Terra satellite, provides CO mixing ratios at seven pressure levels on a global scale and can be used to better understand the transport and transformation of CO. To sensibly validate MOPITT data using in situ observations and for use in assimilation contexts, we propose a hierarchical Bayesian spatio-temporal model based on an advection/diffusion idea. These models and techniques for estimating parameters also have a general use for a variety of environmental problems that include remotely sensed measurements. We present results and discuss the properties of our model and what impact sparse data have on our model estimates.

## 236 Spatial Modeling

Section on Bayesian Statistical Science, Section on Nonparametric Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### Gaussian Processes and Limiting Linear Models

Robert Gramacy, University of Cambridge; ❖ Herbert Lee, University of California, Santa Cruz, School of Engineering, 1156 High St, Santa Cruz, CA 95064, *herbie@ams.ucsc.edu*

***Key Words:*** Bayesian statistics, semi-parametric model, spatial statistics, nonparametric regression

Gaussian processes (GPs) retain the linear model (LM) as either a special case or in the limit. We show how this relationship can be exploited when the data are at least partially linear. However, from the prospective of the Bayesian posterior, the GPs that encode the LM have either probability of nearly zero or are otherwise unattainable without the explicit construction of a prior with the limiting linear model in mind. We develop such a prior and show its practical benefits extend well beyond the computational and conceptual simplicity of the LM. For example, linearity can be extracted on a per-dimension basis or combined with treed partition models to yield a highly efficient nonstationary model. Our approach is demonstrated on synthetic and real datasets of varying linearity and dimensionality.

### Bayesian Model for Dependent Nonparametric Regressions

❖ Abel Rodriguez, Duke University, ISDS Box 90251, 214 Old Chemistry Building, Durham, NC 27708, *abel@stat.duke.edu*; Alan E. Gelfand, Duke University; David B. Dunson, National Institute of Environmental Health Sciences

***Key Words:*** Dirichlet process, non-parametric regression, dependent Dirichlet processes, metaanalysis

We explore Bayesian models incorporating dependent generalizations of the Dirichlet process (DP) that allow borrowing of information

among multiple nonparametric regressions. We first consider models in which different curves are assumed to be exchangeable using a hierarchical DP. Then, we generalize the model to describe spatially correlated functions using spatially dependent DPs. Applications to the analysis of multicenter studies and the estimation of temperature profiles in the ocean are presented.

### Sampling Importance Resampling for Computer Model Inverse Problems

❖ Matt Taddy, University of California, Santa Cruz, 231 Blackburn Street, apt 2, Santa Cruz, CA 95060, *taddy@ams.ucsc.edu*; Bruno Sanso, University of California, Santa Cruz; Herbert Lee, University of California, Santa Cruz

*Key Words:* inverse problems, computer models, sampling importance resampling

The classic approach to statistical inverse problems is to develop a surrogate model for the expensive computer simulator and to use this to find the likelihood for sampling from the posterior distribution of inputs. However, in many cases, there is a large bank of simulated values and the need for a faster algorithm that does not require predicted output at new locations. We develop a sampling importance resampling (SIR) algorithm that works in conjunction with kernel density estimation to resample from the original computer output according to the posterior distribution of input values. We will examine and compare the performance of our algorithm in examples that include multivariate output data from a nondeterministic climate model simulator.

### Bayesian Computational Methods for Models in Geosciences

❖ Alejandro Villagran, University of New Mexico, 415 Humanities Building, The University of New Mexico, Albuquerque, NM 87131-0001, *avhstat@unm.edu*; Gabriel Huerta, University of New Mexico

*Key Words:* multiple very fast simulated annealing, adaptive Metropolis-Hastings, geophysical inversion problem, Bayesian inference

In this talk, we will describe the development and application of different intensive computational techniques that arise from Bayesian approaches to geosciences models. These methods include algorithms such as multiple, very fast, simulated annealing; adaptive Metropolis-Hastings (M-H); parallel tempering; and the multiple try method. We compare the performance of each algorithm and show both advantages and drawbacks in the context of estimating an earthquake epicenter location and in climate model parameter estimation. Mainly, we focus on two aspects: optimization of the posterior density and optimal uncertainty estimation.

### The Effect of Collinearity on Parameter Estimation in Bayesian Spatially Varying Coefficient Models

❖ David Wheeler, The Ohio State University, 100 W. 2nd Ave., Apt 4, Columbus, OH 43201, *wheeler.173@osu.edu*

*Key Words:* MCMC, spatial statistics, regression, simulation study

The recognition in the sciences that relationships between explanatory variables and a response variable in a regression model are not always constant across a study area has led statisticians to develop Bayesian regression models that allow for spatially varying coefficients. The interest in the application of these models is the interpretation of the coefficients in addition to overall model fit. There has been little focus, however, on the accuracy of the estimated coefficients, particularly in the presence of explanatory variable collinearity. This presentation will investigate the effect of collinearity on the regression coefficients through use of simulation and MCMC estimation. The results show the Bayesian regression model is overall fairly robust to moderate levels of collinearity but degrades substantially with strong collinearity.

## 237 Fostering Active Learning in Statistics ●

Section on Statistical Education, Section on Teaching Statistics in the Health Sciences

**Tuesday, August 8, 8:30 am–10:20 am**

### How Do You Come up with These Things?

❖ AndrÈ Michelle Lubecke, Lander University, 320 Stanley Ave., Greenwood, SC 29649-2099, *alubecke@lander.edu*

*Key Words:* active learning, statistics education, teaching, classroom activities

Creating an active learning environment does not require thinking outside the box as much as it requires moving around comfortably inside the box and looking at things from multiple angles. It requires moving from What do I want them to know? to What do I want them to think about? What do I want them to experience? What do I want them to see? What do I want them to realize? My goal is to share the thought processes, ponderings, and inspirations behind some of my creations. My hope is this will enable others to find a way to access their own ideas and, perhaps, provide the encouragement some may need to act bravely upon those ideas.

### Using Simulation To Introduce Inference

❖ Sharon Lane-Getaz, University of Minnesota/Cal Poly, 2829 See Canyon Road, San Luis Obispo, CA 93405, *slanegetaz@msn.com*; Andrew Zieffler, University of Minnesota

*Key Words:* inference, simulation

This paper discusses how dynamic interactive technologies, including web applets and software designed for teaching statistics, were employed in an introductory statistics course to introduce inference. As computers become more readily available to educators, there is wide speculation that teaching inference via dynamic, visual simulations may make some statistics topics more accessible to introductory students (Moore, 1997). Some educators are introducing simulations early and repeatedly in the introductory course, reinforcing a three-tiered schema of the distributions encountered in various simulation activities: population, samples, and the distribution of sample statistics. The objective is to help students develop an informal understanding of the inferential logic that interconnects these tiers of distributions as a foundation for understanding formal inference.

### Active Learning in an Online Introductory Statistics Course

❖ Michelle Everson, University of Minnesota, Department of Educational Psychology, 206 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN 55455, *gaddy001@umn.edu*

*Key Words:* active learning, distance education, introductory statistics, statistics education, teaching

Researchers define active learning in different ways, but most definitions include providing students with opportunities to apply what is learned and to hone critical thinking skills by analyzing, synthesizing, and evaluating content. This paper describes an effort to incorporate active learning in an online, graduate-level introductory course. Students in this course work independently to complete activities, homework assignments, quizzes, and a project that involves data collection, exploration, and analysis. Students also collaborate in assigned discussion groups to complete small-group assignments. The way in which active learning is fostered in this course will be described, and student feedback will be presented to demonstrate the impact course assignments have on student learning and the ways in which assignments might be improved in future offerings.

### Designing Classroom Activities for Student Learning

❖ Allan Rossman, California Polytechnic State University, San Luis Obispo, Department of Statistics, Cal Poly, San Luis Obispo, CA 93407, *arossman@calpoly.edu*; Beth Chance, California Polytechnic State University, San Luis Obispo

***Key Words:*** active learning, constructivism, statistics education

We describe principles that have guided our development of classroom activities to facilitate student learning in a constructivist learning environment. We illustrate these principles with examples of activities intended to lead students to discover and explore fundamental statistical concepts, such as confidence, significance, randomization, and causation. We provide guidelines for adapting activities and developing new ones.

### Evaluation of Students' Learning in Real-Time Online Activities Environment

❖ Felix Famoye, Central Michigan University, Department of Mathematics, Mt. Pleasant, MI 48859, *felix.famoye@cmich.edu*; Carl Lee, Central Michigan University

***Key Words:*** statistical concept, real-time online activity, attitude survey, assessment

New course materials were developed for an introductory statistics course in which a real-time online database created by students was used to teach statistical concepts. This is part of an NSF-CCLI--funded project. The course materials and the real-time, online, hands-on activities were designed to create an active, learning-centered environment for the students. These activities have been incorporated into several sections of our introductory statistics class. Two instructors chose to teach two sections each. One section used the new course materials and the other section used the traditional lecture approach. Preliminary results on students' performance will be presented. Also, results from pre-post surveys on students' attitudes will be presented.

## 238 Health Policy Student Paper Awards

Section on Health Policy Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### Understanding Self-Reported Disability among the Elderly: a Novel Use of Anchoring Vignettes

❖ Kate Stewart, Harvard University, Department of Health Care Policy, 180 Longwood Ave, Boston, MA 02115, *stewart@hcp.med.harvard.edu*; Mary Beth Landrum, Harvard Medical School; David M. Cutler, Harvard University

***Key Words:*** disability, anchoring vignettes, survey, elderly

Recent studies report declining disability among the elderly. While these data often are interpreted to signify health improvements, it is unclear whether they reflect improvements in health, changes in environmental factors, or both. The goals of the study are to evaluate the effect of assistance and environmental factors on survey responses and to explore the use of anchoring vignettes---a new survey method---to factor out these effects on self-reported disability. We are conducting a survey of 400 community-dwelling elderly people in Boston. The study includes existing survey questions on disability, anchoring vignettes, and objective physical assessments. We plan to compare vignette-adjusted disability estimates with physical functioning scores to evaluate whether vignettes may be a useful tool for developing comparable disability estimates.

### Hierarchical and Joint Site-Edge Methods for Medicare Hospice Service Region Boundary Analysis

❖ Haijun Ma, University of Minnesota, 1188 Gibbs Ave., N1, Falcon Heights, MN 55108, *haijunma@biostat.umn.edu*; Bradley P. Carlin, University of Minnesota; Sudipto Banerjee, University of Minnesota

***Key Words:*** health services research, wombling, areal data, conditionally autoregressive (CAR) model, Ising model, boundary analysis

Hospice service, a convenient and ethically preferable health care option for the terminally ill, is unavailable to patients in remote areas not served by any hospice system. In this paper, we seek to determine the service area of two hospice systems in northeastern Minnesota based on only death counts abstracted from Medicare billing records. We suggest a variety of hierarchical models for areal boundary analysis that hierarchically or jointly parameterize both the areas and the edge segments. This leads to conceptually appealing solutions for our data that remain computationally feasible. Our results successfully delineate service areas for our two Minnesota hospice systems that sometimes conflict with the hospices' own service area self-reports. We also obtain boundaries for the relative rate of hospice use, an important secondary endpoint in our investigation.

### Extending the Capture-Recapture Methodology To Estimate Disease Subpopulation Sizes in the Presence of Cross-Classification

❖ Ulysses Diva, University of Connecticut, Department of Statistics, 215 Glenbrook Rd U4120, Storrs, CT 06269, *ulysses@stat.uconn.edu*; Dipak Dey, University of Connecticut; Timothy Morse, University of Connecticut Health Center

***Key Words:*** Bayesian modeling, capture-recapture, maximum likelihood approaches, MSD in Connecticut, population and subpopulation sizes

This paper attempts to extend the capture-recapture methodology to answer questions about the sizes of subpopulations, particularly in situations where there is cross-classification. Three methods of estimating the sizes of subpopulations based on the multinomial likelihood are presented. The first is a conditional maximum likelihood estimator, while the second is an integrated maximum likelihood estimator. The third is a hierarchical Bayesian implementation, which parameterized the table probabilities to implement the independence assumption. The methods were evaluated using simulations and applied to the data on work-related musculoskeletal disorders in Connecticut. We estimated about 149,900 (se=6226) work-related MSD cases in CT from 1995 to 2001, with about 41,600 (se=1781) Carpal Tunnel Syndrome and 109,900 (se=4554) cases having other types of MSD.

### Methods for Profiling the Value of Hospital Care Following Acute Myocardial Infarction

❖ Justin Timbie, Harvard University, 180 Longwood Ave., Suite 301, Boston, MA 02115, *timbie@fas.harvard.edu*

***Key Words:*** profiling, Bayesian methods, cost-effectiveness, hierarchical models

This paper compares three methods to assess the relationship between in-hospital survival and costs for hospital care following acute myocardial infarction (AMI). The first approach models the two outcomes assuming independence. The second method uses a Bayesian hierarchical cost-effectiveness model in which survival and spending for each hospital are modeled jointly. Incremental costs and survival probabilities are estimated by comparing the performance of each hospital to a standard---a hospital with matching case mix. In the third approach, survival is modeled as a function of spending per episode using a hierarchical linear model with random intercepts and slopes. The top decile of hospitals according to each method are selected for comparison. The advantages and disadvantages of each modeling approach are discussed in an attempt to identify a preferred method.

### Nonparametric Statistical Methods for a Cost-Effectiveness Analysis

❖ Phillip Dinh, University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, *ispvd@u.washington.edu*; Xiao-Hua Andrew Zhou, University of Washington

***Key Words:*** confidence intervals, cost-effectiveness analysis, edgeworth expansion, incremental cost-effectiveness ratio, net health benefit, skewness

Two measures often used in a cost-effectiveness analysis are the incremental cost-effectiveness ratio (ICER) and the net health benefit (NHB). Inferences on these quantities often are hindered by highly skewed cost data. In this paper, we derived the Edgeworth expansions for the studentized t-statistics for the two measures and showed how they could be used to guide inferences. We used the expansions to study the theoretical performance of existing confidence intervals based on normal theory and to derive new confidence intervals for the ICER and NHB. We conducted a simulation study to compare our new intervals with several existing methods. We found that our new intervals give good coverage accuracy and are narrower.

## 239 Responding to Emergency Needs for Data: Developing Population Estimates in the Wake of Hurricane Katrina

Section on Government Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### Responding to Emergency Needs for Data: Developing Population Estimates in the Wake of Hurricane Katrina

❖ Lisa Blumerman, U.S. Census Bureau, FOB 3, Room 2430, Washington, DC , *lisa.m.blumerman@census.gov*; ❖ Sally Obenski, U.S. Census Bureau, 4700 Silver Hill Road, Suitland, MD 20746, *sally.m.obenski@census.gov*; ❖ Christa Jones, U.S. Census Bureau, 3231 Valley Drive, Alexandria, VA 22302, *christa.d.jones@census.gov*; ❖ Marc Roemer, U.S. Census Bureau, 4700 Silver Hill Road, Building#2 Room 1002, Washington, DC 20233-9200, *m.i.roemer@gmail.com*; ❖ Thomas Nardone, Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212, *Nardone_T@BLS.GOV*

***Key Words:*** administrative records, record linkage, population estimates

The Census Bureau's mission is to collect and produce quality data about the nation's people and economy. This is accomplished through direct collection of survey and census data, use of administrative records, and development of data products. In the wake of Hurricane Katrina and Rita, it was clear that the traditional approach to the development of population estimates for the areas affected by the hurricanes needed to be revised. An inter-divisional effort was established to understand the effect of the hurricanes and develop methodological enhancements to produce more accurate estimates of the affected areas. A new approach, with new data sources, was needed. This panel discusses the Census Bureau's efforts to work collaboratively on the development of new estimates of the affected population. It further describes collaboration across the federal government, in understanding the issues and acquiring data to assist in measuring the changes.

## 240 Methods for High Dimensional Data

IMS

**Tuesday, August 8, 8:30 am–10:20 am**

### Constrained Dimension Reduction Based on Canonical Correlation

❖ Jianhui Zhou, University of Virginia, Department of Statistics, PO Box 400135 Kerchof hall, Charlottesville, VA 22904, *jz9p@virginia.edu*

***Key Words:*** dimension reduction, canonical correlation, LASSO

The "curse of dimensionality" makes high-dimensional data analysis usually challenging. The sliced inverse regression (SIR) and canonical correlation (CANCOR) methods reduce the dimensionality of data by replacing the explanatory variables with a small number of composite directions without losing much information. However, the composite directions estimated by SIR or CANCOR often involve all of the variables, which makes them difficult to interpret and sometimes meaningless in the following up analyses. To simplify the direction estimates, Ni, Cook and Tsai (2005) proposed the Shrinkage SIR (SSIR) method based on SIR. In this paper, we propose the Constrained CANCOR (CCANCOR) method based on CANCOR. Each direction estimated by CCANCOR consists of only a subset of the variables and is easier to interpret. The direction estimates by CCANCOR are shown to be consistent.

### Sliced Inverse Regression under Data Contamination

❖ Ulrike Genschel, Iowa State University, 312 Snedecor Hall, Ames, IA 50011, *ulrike@iastate.edu*

***Key Words:*** dimension reduction, outliers, SIR, subspace estimation, subspace metric

SIR (Li, 1991) is a dimension reduction procedure for estimating an appropriate regressor subspace. Because fitting a regression curve relies on the correct identification of such a subspace, the sensitivity of SIR to outlying observations becomes crucial to understand and we examine the influence of data contamination on the subspace estimate from SIR. Not only is it possible to overestimate the dimension of the subspace under contamination as stated in the literature, underestimation is possible as well to the extent that none of the true dimension reduction directions of the subspace are recoverable. We also demonstrate that data contamination scenarios producing erroneous subspace estimates in SIR depend on the of the covariance structure of the regressor variables as well as on the knowledge of the dimension of the true subspace. Our findings are supported by a simulation study.

### Significance and Recovery of Block Structures in Binary and Real Matrices with Noise

❖ Xing Sun, The University of North Carolina at Chapel Hill, 601 Jones Ferry Road, apt f5, carrboro, NC 27510, *xingsun@email.unc.edu*; Andrew Nobel, The University of North Carolina at Chapel Hill

***Key Words:*** statistical significance, bipartite, bi-clustering, recoverability

Biclustering/subspace clustering algorithms are recently well studied in the field of Data Mining. One equivalent form of these algorithms can be stated as follows: given a rectangular data matrix, find every submatrix satisfying certain criteria. This paper presents a theoretical analysis of several statistical questions related to some popular biclustering algorithms in a noisy environment. We begin by establishing some simple significance bounds. We then consider the noise sensitivity of a particular biclustering algorithms (Frequent Item Mining) under a simple binary additive model, and show that, even at small noise levels, it fails to directly recovered the underlying matrix with a simple block structures. On the positive side, we proposed an alternative criterion which can recover the simple underlying structures under binary additive model.

### Some Extensions of a Two-Sample Test under the Effect of High Dimension

❖ Yingli Qin, Iowa State University, 315c Snedecor Hall, Iowa State University, Ames, IA 50010, *qinyl@iastate.edu*; Song X. Chen, Iowa State University; Liang Peng, Iowa State University

***Key Words:*** high dimension, power function, significance test, hypothesis test

We proposed a two sample test for population means with high dimensional data when the dimensionality is of the same order of the sample size. The properties of the test are evaluated under a general set of conditions. The test can be accommodated to(1)different covariance matrices for the two samples (2)a wider range of covariance structures regards to concentration on the eigenvalues of the covariance matrices.

### Detecting Sparse Signals with a Large Scale of SPRTs

❖ Yanhong Wu, University of the Pacific, Department of Mathematics, Stockton, CA 95211, *ywu@pacific.edu*

***Key Words:*** sparse signal, SPRT, conditional random walk, mixture distribution

Suppose a large scale of independent and identical sequential probability ratio tests are conducted under normal observations with small same first-type and second-type error probabilities. An indifference zone [\theta_0, \theta_1=-\theta_0 ] is assumed for the design of the SPRT (\theta_0< 0). We consider the testing of existence and strength of sparse signals. First, it is shown that given that a null hypothesis is rejected, the mean estimate behaves the same asymptotically under both P_{\theta_0}(.) and P_{\theta_1}(.). Thus, the test for the existence of sparse signal depends on mainly the number of rejections. Second, given the number of rejections, the test for the strength of signal, or \theta>\theta_1 becomes a problem of testing for homogeneity under a mixture model. A combined local score test is thus proposed.

### On the Accuracy of Data Squashing

❖ Atsuyuki Kogure, Keio University, 5322 Endoh, Fujisawa, 252-8520 Japan, *kogure@sfc.keio.ac.jp*; Masahiko Sagae, Gifu University

***Key Words:*** data squashing, massive data sets, MLE, binning, local moments, kernel method

The concept of "data squashing" was introduced by DuMouchel, Volinsky, Johnson, Cortes and Pregibon (1999) to alleviate the computational burden of the statistical analysis of massive data-sets. The key idea is to squash the original massive data set into a smaller representative sample and then apply the statistical procedure such as MLE to the squashed data. While many exemplifications have been reported to show the practicability of the method, not much research has been made into the accuracy of the method. In this talk we give theoretical arguments to evaluate the accuracy of the data squashing method. To construct the squashed data, DVJCP (1999) partitions the whole data range into bins and matches the local moments of the squashed data to that of the original data on each bin. We show this as a process of density estimation and investigate it in the context of the kernel method.

### New Tools for Sparse Inference: the Regime of Moderate Significances

❖ Jiashun Jin, Purdue University, Department of Statistics, West Lafayette, IN 47907-2067, *jinj@stat.purdue.edu*

***Key Words:*** multiple comparisons, proportion of non-null effect, Kuiper belt object, sparsity, higher criticism, asymptotic optimality

Sparse and moderately strong signals are naturally found in many applications, e.g. covert communications, Kuiper Belt Objects. Here, sparsity refers to the situation that out of a massive data set, only a small fraction of data points contain relevant information or signal, while others are irrelevant or noise; moderately strong signals are those stronger than a typical noise, but not stronger than all of them. Sparse inference in the regime of moderate significances poses interesting new phenomenon as well as new challenges. In this talk, we introduce three recent inference tools for the proportion of signals: the Higher Criticism , the Higher Criticism , Meinshausen and Rice's confidence lower bound, and Cai, Jin, and Low's confidence lower bound. We will discuss the asymptotic behavior of these tools in detail, and also compare their strength and weakness.

## 241 Construction of Prior Distributions

Section on Bayesian Statistical Science

**Tuesday, August 8, 8:30 am–10:20 am**

### An Asymptotic Viewpoint on High-Dimensional Bayesian Testing

❖ Dan Spitzner, Virginia Polytechnic Institute and State University, Department of Statistics, Mail Code 0439, Blacksburg, VA 24061, *dan.spitzner@vt.edu*

***Key Words:*** Bayesian testing, smooth goodness-of-fit tests, high-dimensional testing

Bayesian testing is studied asymptotically on a high-dimensional normal means model, in which the null hypothesis of zero means in all dimensions is tested against general alternatives. This is known to serve as a canonical model for smooth goodness-of-fit testing. The asymptotic setup is such that prior mass placed on the null hypothesis is allowed to decrease as dimensionality increases, while at the same time the dispersion of the prior placed on the alternative is allowed to increase, thereby tending toward a noninformative specification. The interest is to deduce appropriate rates of change for the prior parameters in order to assure the test's reasonable behavior in large dimensions. When geometric constraints representing smoothness are imposed, it becomes sensible to weight the prior, and an expanded objective is to deduce suitable configurations for those weights.

## Novel Bayesian Variable-Selection Priors for "Large p Small n" Data Analysis

❖ Joseph Lucas, Duke University, Durham, NC 27708-0251, *joe@stat.duke.edu*

***Key Words:*** variable selection, sparcity, microarray, low sample, model selection, high dimensional

Standard Bayesian variable-selection priors for regression coefficients involve mixing a "point mass" at zero with a normal (or other) distribution assuming an unknown mixing proportion, q. We show that the use of the traditional prior for q can lead to over-estimation and significant false-positive bias. The problem is particularly apparent in highly multivariate regression and ANOVA modeling such as arises in the analysis of gene expression experiments. We describe a novel hierarchical extension of the traditional approach involving observation-variable specific indicators of inclusion and which alleviates these issues. Resulting posterior distributions, computed with custom MCMC methods, induce conservative inferences of "significant" effects consistent with the expectation of variable-selection priors. Examples in analysis of microarray data demonstrate these issues and advances.

## New Priors for Complex System Reliabilities and Trends

❖ Earl Lawrence, Los Alamos National Laboratory, Statistical Sciences Group, D1, P.O. Box 1663, MS F600, Los Alamos, NM 87545, *earl@lanl.gov*; Scott Vander Wiel, Los Alamos National Laboratory

***Key Words:*** prior, component, system, reliability

When modeling reliability degradation in complex systems it is tempting to use diffuse priors on parameters of the system's components and subcomponents. This can lead to a nonsensical system prior and have a dramatic effect on the posterior when data are sparse. We explore negative log-gamma component (NLG) priors that induce a more appropriate a priori distribution of system reliability and its evolution over time. NLG distributions are similar to the beta, but closed under products. This closure makes NLGs ideal for modeling complex series systems in which the system-level reliability is a product of the component-level reliabilities. The NLG model is easily extended to include a hierarchical prior that anticipates situations where only a few of the components will contribute substantially to the level and trend of system reliability over time.

## Reference Priors in Linear Models with General Covariance Structures

❖ Xin Zhao, Merck & Co., Inc., 806 Lincoln Woods, Lafayette Hill, PA 19444, *xin_zhao@merck.com*; Martin T. Wells, Cornell University

***Key Words:*** covariance estimation, general linear models, graphical models, linear mixed models, p-splines, reference prior

A new class of reference priors for linear models with general covariance structures are developed. A general Markov chain Monte Carlo (MCMC) algorithm is also proposed for implementing the computation. We present several examples to demonstrate the results: Bayesian penalized spline smoothing, a Bayesian approach to bivariate smoothing for a spatial model, and prior specification for a Gaussian graphical model.

## Bayesian Interval Estimation of Attributable Risk Using Probability Matching Prior

❖ Bin Huang, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, OH 45229, *bin.huang@cchmc.org*; Siva Sivaganisan, University of Cincinnati

***Key Words:*** attributable risk, probability matching prior, interval estimate, cross-sectional study, Jeffrey's prior, non-informative prior

Attributable risk (AR) is one of the most commonly used risk measures in epidemiology and public health research. We derived probability matching prior when the parameter of interest was to estimate AR. Different study designs were considered. Simulation studies examined the frequentist performance of Bayesian interval estimates for AR (i.e., rate of coverage, length of interval, and root mean square error). These results were contrasted with the Bayesian interval estimates using standard noninformative priors, or Jefferey's priors, and the commonly used asymptotic estimates.

## Bayesian Inference with Matching Priors

❖ Ana-Maria Staicu, University of Toronto, 100 St. George Street, TORONTO, ON M5S 3G3 Canada, *anamaria@utstat.utoronto.ca*

***Key Words:*** matching priors, orthogonal parameters, higher order approximations, uninformative priors, approximate Bayesian inference, logistic model

Matching priors are a natural choice in the absence of specific prior information, since they lead to inferences equivalent to frequentist inferences to the second order. We use the Tibshirani (1989) class of uninformative priors for orthogonal component parameters in exponential family models. A numerical study compares posterior tail probabilities, using this class of priors, with the frequentist first and higher-order counterpart p-values using conditional inference. The extension of this to more general models is explored and some directions are indicated.

## A General Approach for Constructing Normalizing Constants for Parameter Selection Priors

❖ Christopher Carter, University of New South Wales, Australian Graduate School of Management, Sydney, 2052 Australia, *chriscar@agsm.edu.au*

***Key Words:*** constrained Wishart distribution, Gaussian graphical models, model averaging, variable selection

A parameter selection prior is a mixture of discrete and continuous components which allows some or all of the parameters in a statistical model to be exactly zero. Two examples of such priors are a variable selection prior in a Gaussian linear regression model and a covariance selection prior, where the off-diagonal elements of the inverse of the covariance matrix are allowed to be exactly zero. Our article makes the following contributions. First, it sets out a general framework for parameter selection priors; second, it presents a simulation method for estimating normalizing constants for parameter selection priors; third, it applies the methodology to obtain a general class of covariance selection priors; and fourth, it provides an extensive empirical check on the simulation methodology for obtaining the normalizing constants for covariance selection priors.

# 242 Classification and Other Correlated Data Problems

Section on Physical and Engineering Sciences

**Tuesday, August 8, 8:30 am–10:20 am**

## InfoEvolve(TM): a Data Mining Tool Combining Information Theory and Genetic Algorithms

❖ Aaron J. Owens, DuPont, P.O. Box 80249, Wilmington, DE 19880-0249, *aaron.j.owens@usa.dupont.com*; Karen M. Bloch, DuPont

***Key Words:*** information theory, genetic algorithms, data mining, dimensionality reduction, classification, empirical modeling

We introduce the InfoEvolve(TM) empirical modeling technology, which is based on the combination of Information Theory and Genetic Algorithms. The InfoEvolve(TM) Discovery module is used for dimensionality reduction in large data sets, and its Predict module uses a committee of experts approach for accurate prediction of classifications. We present two example applications: using dimensionality reduction to improve model accuracy for a grain classification instrument, and modeling to predict movements of the S&P 500 stock index on a day-to-day basis.

## Analysis of Handwritten ZIP Code Digits Using OBSTree

❖ Atina Dunlap Brooks, North Carolina State University, Department of Statistics, Raleigh, NC 27695, *adbrook2@stat.ncsu.edu*; Jacqueline Hughes-Oliver, North Carolina State University

***Key Words:*** data mining, machine learning, handwriting analysis, prediction, classification, simulated annealing

The classification method OBSTree has been used to successfully analyze drug discovery datasets. One significant feature of these datasets is that they contain a small percentage of observations which have meaningful classes and a large number of uninteresting observations. This paper explores the algorithmic modifications necessary to perform an OBSTree analysis of a dataset in which there is no irrelevant class. The dataset used is a well known US postal zip code dataset which contains individual handwritten digits. Our results are compared to a variety of data mining methods, and the interpretability benefits of OBSTree are examined.

## A Nonparametric Approach Based on a like Markov Property for Classification

❖ Eun Sug Park, Texas Transportation Institute, 3135 TAMU, College Station, TX 77843-3135, *e-park@tamu.edu*; Clifford Spiegelman, Texas A&M University

***Key Words:*** classification, high dimensional data

We suggest a new approach for classification based on nonparametricly estimated likelihoods. Due to the scarcity of data in high dimensions, nonparametric estimation of the likelihood functions for each population is impractical. Instead, we propose to build a class of candidate likelihood models based on a Markov property and to provide multiple likelihood estimates that are useful for guiding a classification algorithm. Our density estimates require only estimates of one and two-dimensional marginal distributions, which can effectively get around the curse of dimensionality problem. A classification algorithm based on those estimated likelihoods is presented. A modification to it utilizing variable selection of differences in log of estimated marginal densities is also suggested to handle specifically high dimensional data.

## Sensitivity Analysis with Correlated Inputs: Application to Kinetic Models

❖ Sebastien Da Veiga, Institut FranÁais du PÈtrole, BP3, Vernaison, 69390 France, *sebastien.da-veiga@ifp.fr*; Francois Wahl, Institut FranÁais du PÈtrole; Fabrice Gamboa, UniversitÈ Paul Sabatier

***Key Words:*** nonparametric regression, global sensitivity indices, conditional moments estimation

We propose two original methods to estimate sensitivity indices when the inputs of a model are correlated. These methods are based on local polynomial smoothers for the estimation of conditional moments. Both have good theoretical properties which are exhibited and also illustrated through analytical examples. We use these estimators to carry out a sensitivity analysis on two real cases of kinetic models with correlated parameters: a simplified model of hydrodesulfuration and a model of isomerization of normal butane.

## A Structural Equation Method for Modeling Data Center Thermal Distribution

❖ Zhiguang Qian, Georgia Institute of Technology, 7 Burdick Street, Apt 3A, Stamford, CT 06905, *zqian@isye.gatech.edu*; Yasuo Amemiya, IBM T. J. Watson Research Center

***Key Words:*** latent variable, structural equation system, computer experiment, data center, multivariate statistical analysis

Temperature management is key in designing and running a reliable data center. Physics-based mathematical models, implemented in computer code, are often used to study air movement and temperature distribution mechanisms. A statistical method based on latent variables is introduced for analyzing the multivariate temperature readings produced by the computer experiment, and for building a surrogate model to be used for prediction. A two-stage estimation procedure is developed for the proposed latent variable model by making use of sufficient statistics and ordinary least square estimation. Also discussed is a method for determining practical configurations to meet some physical and usage requirements.

## Improved Tolerance Factors for Multivariate Normal

❖ Sumona Mondal, University of Louisiana at Lafayette, 327 General Mouton Ave., Apartment 13, Lafayette, LA 70501, *sumona@louisiana.edu*; Kalimuthu Krishnamoorthy, University of Louisiana at Lafayette

***Key Words:*** confidence, content, coverage probability, noncentral chi-square approx, Wishart distribution

In this article, an improved method of computing tolerance factors for a multivariate normal distribution is proposed. The method involves an approximation and simulation, and is more accurate than the several approximate methods considered in Krishnamoorthy and Mathew (1999,Technometrics, 41, 234-249). The accuracies of the tolerance regions are evaluated using Monte Carlo simulation. Simulation study shows that the new approach is very satisfactory even for small samples. Tolerance factors based on the proposed approach are tabulated for the dimension of the normal distribution p = 2(1)10, and various sample sizes ranging from 2p + 1 to 1000.

### K-Means Clustering: a Novel Probabilistic Formulation, with Some Applications

❖ Samiran Ghosh, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269, *samiran@stat.uconn.edu*; Dipak Dey, University of Connecticut

***Key Words:*** Bayesian computation, k-means clustering, Mahalanobis distance, Markov chain Monte Carlo, multivariate exponential power family

One of the simplest partition based clustering algorithm is K-means algorithm. It can be shown that the computational complexity of K-means does not suffer from exponential growth with dimensionality. The only crucial requirements are the knowledge of cluster number and computation of some suitably chosen similarity measure. For this simplicity and scalability, K-means remains an attractive alternative when compared to other competing clustering philosophy. However being a deterministic algorithm, traditional K-means have several drawbacks. It only offers hard decision rule, with no probabilistic interpretation. In this paper we have developed a decision theoretic framework by which traditional K-means can be given a probabilistic footstep. This will not only enable us to do a soft clustering rather whole optimization problem could be recasted into Bayesian modeling framework.

## 243 Missing or Censored Data

Section on Statistical Computing, Biometrics Section, ENAR
**Tuesday, August 8, 8:30 am–10:20 am**

### Counting the Homeless in Los Angeles County

❖ Brian Kriegler, University of California, Los Angeles, Department of Statistics, 5455 Sylmar Ave., Unit 2501, Sherman Oaks, CA 91401, *bk@stat.ucla.edu*; Richard Berk, University of California, Los Angeles; Don Ylvisaker, University of California, Los Angeles

***Key Words:*** random forests, imputing, aggregation, confidence intervals, homeless

Over the past two decades, a variety of methods have been used to count the homeless in large metropolitan areas. In this paper, we report on a recent effort to count the homeless in Los Angeles County, one that employed the sampling of census tracts. A number of complications are discussed, including the need to impute homeless counts to areas of the County not sampled. We conclude that, despite their imperfections, estimated counts provided useful and credible information to the stakeholders involved.

### An Investigation of Missing Data Methods for Decision Trees

❖ Yufeng Ding, New York University, 44 W. 4th Street, IOMS Dept, New York, NY 10012, *yding@stern.nyu.edu*; Jeffrey Simonoff, New York University

***Key Words:*** missing data, decision tree, probabilistic split, surrogate split, CART, c4.5

There are many different missing data methods used by decision tree algorithms, but not many studies have been done on their appropriateness and performance. This paper provides both analytic and Monte Carlo evidence regarding the effectiveness of six popular missing data methods. We show that in the context of decision trees, the relationship between the missingness and the dependent variable, rather than the standard missingness classification approach of Rubin (missing completely at random, missing at random, and not missing at random), is the most helpful criterion to distinguish different missing data methods. We are also able to make recommendations as to the best method to use in various situations.

### Nonlinear Neural Network Imputation

❖ Safaa Amer, National Opinion Research Center, 946 NW Circle Blvd., Apt 164, Corvallis, OR 97330, *amer-safaa@norc.org*

***Key Words:*** imputation, missing data, neural networks, non-linear

Non-linear models offer a flexible realistic way of imputing missing data. An evaluation of the performance of a class of feed-forward non-linear neural networks in imputation is presented compared to other statistical techniques. Results show that neural network imputation technique offers a tremendous speed advantage and similar or better performance relative to some other statistical imputation techniques.

### Using Objective Measures in Combination with Self-Report To Estimate Adherence

❖ Richard Thompson, Johns Hopkins Bloomberg School of Public Health, 3133 Guilford Ave., Baltimore, MD 21218, *rthompso@jhsph.edu*; Michael Griswold, Johns Hopkins Bloomberg School of Public Health; Arlene Butz, Johns Hopkins University; Michele Donithan, Johns Hopkins Bloomberg School of Public Health

***Key Words:*** sensitivity, missing data, self-report data, study design, electronic monitoring

Asthma is a major health risk for those under the age of 18. Adherence to prescribed asthma medication is often assessed purely through self-report, which is known to give biased estimates. We present methods to account for biases in self-report data through objective measures which are also surrogates for the outcome of adherence. We address this by: 1) relating the objective measure to the self-report measure, 2) adjusting the self-reported results, and 3) performing sensitivity analyses around assumptions on informative missingness mechanisms. Motivating data come from self-reported diary data of nebulizer use for rescue medication in conjunction with electronic monitoring. Our goal is to improve estimation and guide study design using self-report data. Specifically, given limit resources, what percent of the sample should be monitored in order to improve self-reported outcomes?

### Test of Independence with Incomplete Contingency Table

❖ Shin-Soo Kang, Iowa State University, 3308 Valley View Road, Ames, IA 50014, *sskang@iastate.edu*; Kenneth Koehler, Iowa State University; Michael D. Larsen, Iowa State University

***Key Words:*** complete case analysis, multiple imputation, Wald statistic

MLE is obtained from the partial log-likelihood function for the cell probabilities of two-way incomplete contingency tables proposed by Chen and Fienberg (1974). Constraints can be incorporated with the lagrangian multiplier. Variances of MLE estimators of population proportions are derived from the matrix of second derivatives of the log-likelihood with respect to cell probabilities, and Wald and likelihood ratio tests of independence are examined. Simulation results, when data are missing at random, reveal that MLE provides more efficient estimates of population proportions than either multiple imputation (MI), based on data augmentation, or complete case analysis, but neither MLE nor MI provides an improvement over complete-case analysis with respect to power of tests for independence.

### A Comparison of Parametric and Semiparametric Methods for Predicting Limited Dependent Variables

❖ Maya Sternberg, Centers for Disease Control and Prevention, 1600 Clifton Road, MS E63, Atlanta, GA 30333, *mrs7@cdc.gov*; Alula Hadgu, Centers for Disease Control and Prevention; Suephy Chen, Emory University School of Medicine

***Key Words:*** censoring, Tobit, CLAD, regression, prediction, error

This paper uses simulation to compare parametric and semi-parametric regression methods to predict a limited dependent variable. We compare the results from both parametric and semi-parametric misspecified models which ignore any censoring such as ordinary least squares regression, least absolute deviation regression, and least median squares regression. Followed by comparing both parametric and semi-parametric models which incorporate censoring such as Tobit regression, and censored least absolute deviation regression. We not only compare the unbiasedness and efficiency of the various regression estimators, but also compare the out of sample predictive performance of the proposed estimator and model specifications. W e vary the amount of censoring and basic model assumptions. We compare the performance of a variety of different forms of prediction error.

### Estimation of Correlation Coefficient in Bivariate Normal Population Based on Concomitants of Order Statistics

❖ Qinying He, The Ohio State University, 1958 Neil Ave., Columbus, OH 43210, *he@stat.ohio-state.edu*; H. N. Nagaraja, The Ohio State University

***Key Words:*** concomitants of order statistics, bivariate normal, correlation coefficient, Fisher's z transformation, censored sample

Consider a random sample from a bivariate normal population (X,Y), where only the Y's values and the ranks of associated X values are observed. Our goal is to study the estimators of the population correlation coefficient that are based on this data. We examine three such estimators and study their small and large sample properties. Using simulation, we obtain their bias, mean square error, and asymptotic distribution. One of the estimators, Gini Correlation, has an efficiency of 80--96%.

## 244 Risk Assessment: Disease, Health, and Harm ● ✪

Section on Risk Analysis, Biometrics Section, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Characterization of Occupant Anthropometry and Clearance Measures in Passenger Cars

❖ Duane Steffey, Exponent, Inc., 149 Commonwealth Drive, Menlo Park, CA 94025, *dsteffey@exponent.com*; Robert Bove, Jr., Exponent, Inc.; Jacob Fisher, Exponent, Inc.; Lauren Ciccarelli, Exponent, Inc.; Robert Cargill, Exponent, Inc.; Tara Moore, Exponent, Inc.

***Key Words:*** injury risk, motor vehicles, principal components, hierarchical regression

Recent studies have suggested a link between body size and injury potential in motor vehicle accidents. Differences in outcomes may reflect varying positions of people within the passenger compartment and clearance measures between the occupant and vehicle structures. This paper presents analyses of data on adult subjects of varying heights and weights who were seated in the driver's seat of passenger cars and asked to adjust the seat and restraint to a comfortable driving position. Principal component methods are applied to obtain interpretable lower-dimensional reductions of anthropometric and seating geometry variables. Hierarchical regression models are developed to assess vehicle-specific effects while incorporating subject-level covariates. Study findings are essential in approaching the investigation of relationships between anthropometric measures and injury potential.

### Smallpox and Pandemic Influenza: Modeling Potential Challenges for the U.S. Blood Supply

❖ Steven Anderson, U.S. Food and Drug Administration, 1401 Rockville Pike, HFM210, Rockville, MD 20852, *andersonst@cber.fda.gov*

***Key Words:*** computer modeling, smallpox, pandemic influenza, whole blood, safety, transfusion risks

Each year there are approximately four million life-saving blood transfusions in the U.S. Blood shortages or perturbations in supply can occur on a local or regional basis. Blood safety is ensured by measures including deferral of at-risk donors, testing, and others. Infectious diseases and measures to prevent disease spread can challenge the safety and supply of blood. We developed a computer model of the U.S. blood supply. Using several nationwide smallpox vaccination scenarios, we show potential impacts on the blood supply. Scenarios include vaccination programs of different durations and illustrate use of the model to evaluate possible measures that may maintain supply. We will illustrate model applications to evaluate various possible "what if" scenarios for pandemic influenza and possible measures that could preserve supply during a challenge.

### A Comparison of Bayesian Networks and MCMC Techniques for Quantitative Risk Assessment

❖ Paola Berchialla, University of Torino, Department of Statistics, Piazza Arbarello 8, Torino, 10122 Italy, *berchialla@econ.unito.it*; Silvia Snidero, University of Torino; Alexandru Stancu, University of Torino; Cecilia Scarinzi, University of Torino; Roberto Corradetti, University of Torino; Dario Gregori, University of Torino

***Key Words:*** Bayesian networks, Markov chain Monte Carlo, quantitative risk assessment

Bayesian Networks (BNs) have been successfully applied to a great number of classification problems and have been used to develop quantitative risk assessment models. The BN responds immediately to changes, such as entering evidence, because it does not use simulation and can propagate information from any point in the network to all others by Bayes's theorem. However conventional Bayesian networks require discretization of continuous prior to learning which introduces errors if continuous variables have to be discretized. Markov chain Monte Carlo approach does not require discrete variables while retaining some of the properties of the BN model, such as the ability to draw inferences from evidence. A comparison of these modeling approaches for a quantitative risk assessment model is proposed using the case study of foreign body injuries in children.

### Risk Assessment of Listeria Monocytogenes on Frankfurters with Organic Acid Salts Surface Treatments

❖ Zheng Lu, Iowa State University, 46 Schilletter Village, Apt. D, Ames, IA 50010, *zhl@iastate.edu*; Chunwang Gao, Iowa State University

***Key Words:*** risk assessment, growth curve model, food safety, l. monocytogenes, organic acid salts

Risk assessment of microbiological hazards in foods has gained increased attention in food microbiology in recent years, and has become a priority area of work for the Codex Alimentarius Commission (CAC) since 1999. Organic acid salts such as sodium lactate (SL), sodium diacetate (SD), potassium benzoate (PB), potassium sorbate (PS) and their combinations are recognized as potential inhibitions of L. monocytogenes growth on frankfurters. The objective of this study was to select a proper L. monocytogenes growth curve model for frankfurters treated with organic acid salts and to provide a means of risk assessment for L. monocytogenes in ready-to-eat (RTE) meats, treated with organic acid salts. This information is expected to improve the practical guidelines and methodology for hazard characterization of microbial pathogens on RTE meat products.

### Generalized Residual Sojourns under Truncation and Censoring

❖ John Hsieh, University of Nevada, Reno, School of Public Health, Lombardi Bldg MS274, Reno, NV 89557, *jhsieh@unr.edu*

***Key Words:*** lifetime, failure time, censoring/truncation, generalized survival function, generalized distribution function, survival analysis

This article introduces the generalized survival and distribution functions and use them to construct generalized residual lifetime and failure time transformations under censoring and truncation. Both cohort and stationary population interpretations of the expectations of these residual sojourns are described in terms of the steady-state distribution of a population renewal process. Probability distributions of these transformed random variables and their moments are derived in terms of the original life distribution of the underlying lifelength random variable. Lorenz curve and scaled total time on test transform are generalized to allow for censoring as well as truncation and to cover lifetimes as well as failure times. The probabilistic results obtained in this article provide the basis for further study on statistical inference of censored/truncated lifetimes and failure times.

### General Indexes for Spatial-Temporal Population Risk Assessment

❖ Chu-Chih Chen, National Health Research Institutes, 35 Keyan Road, Division of Biostatistics and Bioinformatics, Zhunan Town Miaoli County, 350 Taiwan, *ccchen@nhri.org.tw*

***Key Words:*** aggregate cumulative risk, kriging, lifetime cancer risk, hazard index, dioxin, incinerator

We develop formulations for risk assessment of regional residential population exposed to a hazardous chemical (carcinogen or non-carcinogen) within a certain amount of time. Exposure assessment of unsampled area is predicted using kriging based on available sample measurements at different sites. Spatial-temporal aggregate cumulative risk is then obtained by integrating the risk function over population age distribution and exposure time interval, with average risk divided by area of the region and length of time. The formulation reduces to classical risk assessment indexes of lifetime cancer risk for carcinogens and hazard index (HI) for non-cancer toxic chemicals when spatial exposure and population is homogeneous. Example for risk assessment of residential dioxin exposure generated by incinerators in local area is illustrated.

### Public Health Interventions and Transition Probabilities

❖ Paula Diehr, University of Washington, Box 357232, Seattle, WA 98195-7232, *pdiehr@u.washington.edu*

***Key Words:*** multi-state life table methods, Markov, public health

Public health interventions improve population health by improving health at baseline or by changing the probabilities of transition from one health state to another. We estimated the impact of eight types of interventions in a cohort of older adults (age 65 at baseline). Making everyone healthy at baseline would increase life expectancy by 0.23 years and increase years of healthy life (YHL) by 0.54 years. Equivalent improvement could be obtained with a 12% decrease in P(Sick|Healthy), a 16% increase in P(Healthy|Sick), a 15% decrease in P(Die|Sick), or a 14% decrease in P(Die|Healthy). All interventions increased YHL and survival, but those maximizing survival also increased years of morbidity and medical expenditures. The choice of intervention depends on the relative value society assigns to an additional year of sick life.

## 245 Misspecifications, Data Quality, and Improvements ●

Business and Economics Statistics Section

**Tuesday, August 8, 8:30 am–10:20 am**

### Sons, Daughters, and Parents' Labor Supply: New Evidence from Matched CPS Data

❖ James Vere, The University of Hong Kong, School of Economics and Finance, Pokfulam Road, Hong Kong, 00852 China, *jpvere@hku.hk*

***Key Words:*** panel data, labor supply, household production, fertility

In a 2002 article, Lundberg and Rose (LR) claimed that men's wages and labor supply increase more in response to births of sons than to births of daughters. However, underreporting of daughters in the PSID introduces nonclassical measurement error in men's fertility histories, which biases LR's fixed-effects estimates. This paper uses matched CPS data to replicate LR's analysis. Like the PSID data, the CPS data have a panel structure where respondents can be matched across years. Unlike the PSID, however, the fertility data in the CPS do not exhibit any sex-related biases. The results yield no evidence that fathers increase wages or hours worked by greater amounts in response to sons than to daughters. More generally, the results show that nonclassical forms of measurement error in fertility have potentially serious implications for microeconomic analyses of household behavior.

### Statistical Modeling: Science versus Business and Domain Expertise

❖ Igor Mandel, Media Planning Group, 9-15 Berdan Ave., Fair Lawn, NJ 07410, *igor.mandel@mpg.com*; David Hauser, Media Planning Group

***Key Words:*** statistical modeling, advertising, ROI, optimization, budgeting, yield analysis

The dichotomy between results derived from statistical modeling and business expectation is addressed: how to balance unconstrained statistical models that may contradict domain knowledge or expectation. If differences exist, the results need to be defended by the statistician to change that domain knowledge/expectation, or the application of constraints is needed to "correct" the results. This process of alignment is generally informal and usually quite subjective. All statisticians face this dilemma almost every time they present results to an audience. The combinations of "subjective" and "objective", "derived" and "desirable", "free" and "forced" aspects of the modeling could be complicated, and often pose difficult scientific and moral challenges. Different approaches are considered; recommendations follow from extensive advertising ROI modeling created for many product categories.

## Estimating Signatures on a Petition with Cis

❖ Mary M. Whiteside, The University of Texas at Arlington, Box 19437, Arlington, TX 76019, *whiteside@uta.edu*; Mark E. Eakin, The University of Texas

***Key Words:*** Goodman, exact distributions, replicated signatures

This paper presents a method for constructing a confidence interval estimate of the number of valid, un-replicated, signatures on a petition. The estimation of the number of unique individuals represented by signatures on a petition is one of many important applications of the more general problem of estimating the number of classes in a finite population. The Goodman estimator is widely used for this purpose. We derive critical values from the exact distribution and an unbiased estimator of the variance of the Goodman estimator. Thus, exact confidence intervals can be constructed for the number of unique, valid signatures on a petition.

## Estimation of Industry Distribution of Statistical Discreancy in National Income and Product Accounts

❖ Baoline Chen, Bureau of Economic Analysis, 1441 L Street, NW, Washington, DC 20230, *baoline.chen@bea.gov*

***Key Words:*** national accounts reconciliation, constrained optimization, generalized least square estimation, reliability measure of data

Gross domestic product (GDP) and gross domestic income (GDI), though conceptually equivalent, differ by statistical discrepancy (SD). Currently, there are no estimates of SD by industry. Lack of such information hinders a proper understanding of the sources of inconsistency in the national account. This paper describes and illustrates an method that can correctly estimate industry distribution of SD based on reliability of the initial estimates, and it can accurately reconcile GDI by industry account with Input-Output (I-O) account, which measures GDP as value-added. The reconciliation model is described by a constrained optimization model which minimizes the weighted sum of squares of deviation from the initial estimates in all components of I-O and GDI by industry accounts. The optimal solutions are equivalent to estimates from GLS estimation. Data used are from BEA 1997 accounts.

## Quantifying the Quality of Macroeconomic Variables

❖ Alex Teterukovsky, Statistics Sweden, Box 24300, Stockholm 10451, Sweden, *alex.t@scb.se*; Lars-Erik Öller, Statistics Sweden/ Stockholm University

***Key Words:*** forecast errors, revisions, entropy, statistical quality

Methods to quantify the quality of a macroeconomic statistical time series are presented. The measures are based on a combination of how predictable the series is and how much its statistics needs to be revised. An information window is introduced which is based on signal-to-noise ratios (SNR) and provides a snapshot of the quality. A formulation of information in terms of entropy is considered. A nonparametric approach is taken which allows for testing whether a forecast or a preliminary value is informative. Concavity and monotonous convergence of information accrual are discussed. It is shown how the suggested measures signal that either given forecasts or macroeconomic variables (GDP components) themselves, are of dubious quality.

## Incorporating EITC Participation into the SAIPE Program's County Poverty Model

❖ Wesley Basel, U.S. Census Bureau, 4700 Silver Hill Road, Room 1451 Bldg 3, Suitland, MD 20746, *wesley.w.basel@census.gov*; Brian J. O'Hara, U.S. Census Bureau

***Key Words:*** small-area estimation, small-domain areas, SAIPE, poverty, EITC

The U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces state and county income and poverty estimates using decennial census data, household survey data, administrative records and population estimates. The IRS has provided the Census Bureau new data on earned income tax credit (EITC) participation. Incorporation and evaluation of this new data in the SAIPE program models could lead to improvements in small area estimates. The EITC is a uniformly administered and widely advertised program. Regional variation in participation rates might be lower than other programs that have been utilized in the SAIPE program models. Due to the wide coverage of the EITC, a particular point of this research is improving modeling performance for counties with weak participation in other programs, thus reducing errors under the current model.

## Misspecification of Cointegrating Ranks in Seasonal Models

❖ Byeongchan Seong, Pohang University of Science and Technology, Department of Mathematics, San 31 Hyojadong, Namgu, Pohang Gyeongbuk, 790-784 South Korea, *sibici@snu.ac.kr*; Sinsup Cho, Seoul National University; Sung K. Ahn, Washington State University; S. Y. Hwang, Sookmyung Women's University

***Key Words:*** Gaussian reduced rank estimation, seasonal cointegration

We investigate the effects of mis-specification of cointegrating (CI) ranks at other frequencies on inference of seasonal cointegration at the frequency of interest such as testing of CI rank and estimation of CI vector. Earlier studies focused mostly on a single frequency corresponding to a seasonal root at a time, ignoring possible cointegration at the remaining frequencies. We investigate the effects of the mis-specification, especially with finite samples adopting Gaussian reduced rank estimation by Ahn and Reinsel (1994) that considered cointegration at all frequencies of seasonal unit roots simultaneously. It is observed that the identification of the seasonal CI rank at the frequency of interest is sensitive to the mis-prespecification of the CI ranks at other frequencies even for larger samples, especially, when the CI rank at frequency 0 is under-specified.

# 246 Nonclinical, Nonbiological Drug Development ●

Biopharmaceutical Section
**Tuesday, August 8, 8:30 am–10:20 am**

## Statistical Issues of Two-Stage Delivery Dose Uniformity Test of Orally Inhaled and Nasal Drug Products

❖ Meiyu Shen, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20993, *shenm@cder.fda.gov*; Yi Tsong, U.S. Food and Drug Administration

***Key Words:*** dose content uniformity, inhaler product, parametric tolerance intervals, quality assurance

In response to the 1998 FDA Draft Guidance for Chemistry, Manufacturing and Controls (CMC) documentation for metered dose inhalers (MDIs), dry powder inhalers (DPIs) and nasal spray drug products, pharmaceutical companies and industry groups have submitted numerous comments on the target specifications of the delivery dose and the validity of the delivered dose uniformity (DDU) test. These comments mainly focused on the unnecessary degree of restrictions and the statistical bias of the test procedure. In the last five years, both FDA and the International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS) have actively worked on developing an improved statistical testing procedure. In this presentation, we discuss the deficiencies of the 1998 FDA Draft Guidance test and a group sequential test procedure based on parametric tolerance intervals (PTT).

## Practical Considerations on Lot Acceptance Sampling Procedures

❖ Boguang Zhen, U.S. Food and Drug Administration, 11413 Duryea Drive, CBER, Potomac, MD 20854, *Boguang.zhen@fda.hhs.gov*

***Key Words:*** sampling procedure, military standard, operating characteristic curves, lot release, binomial distribution

To release a biologics product into the market, sampling inspection is required in order to assure the quality of the product. Many organizations adopted the Military Standard Tables for the inspection. Military Standard 105D is a sampling scheme and combines several individual sampling plans in a system designed to motivate the producer to sustain acceptable quality levels. This talk introduces the use of Military Standard and shows its limited use, especially when there is a higher defective rate than expected in lots to be released. A new multiple sampling procedure using the operating characteristic curves and joint binomial distributions is developed as an alternative approach to deal with the limited use of the Military Standard. Considerations of clinical impacts are also taken into accounts when using the new procedure. A real example is presented.

## Assessing the Reproducibility of an Analytical Method

❖ Jason Liao, Merck Research Laboratories, P.O. Box 4, WP37 C305, West Point, PA 19486, *jason_liao@merck.com*

***Key Words:*** validation, reproducibility, concordance, agreement interval, bias

A validated analytical method may need to be revalidated because of changes such as transfers between laboratories or analysts, changes in instruments and/or software platforms, or modifications. A fundamental parameter for this revalidation is the reproducibility of the method. In this paper, a method is proposed to define the agreement interval for each paired observation in assessing the overall concordance. The proposed approach is simple, but informative, and offers many advantages over existing approaches. It is demonstrated through an assay bridging study.

## Imputing Nonignorable Missing Data on Clinical Laboratory Assessments

❖ Kapildeb Sen, Bristol-Myers Squibb Company, 311 Pennington-Rocky Hill Road, Pennington, NJ 08543, *kapildeb.sen@bms.com*; Chen-Sheng Lin, Bristol-Myers Squibb Company; Kannan Natarajan, Novartis Pharmaceuticals Corporation; Jun Xing, Bristol-Myers Squibb Company

***Key Words:*** non-ignorable missingness, longitudinal analysis, laboratory assessments, single imputation, multiple imputation

A common problem in clinical trials is the missing data on the direct measures of clinical laboratory assessments (e.g. Glomerular Filtration Rate, low-density lipoprotein cholesterol) at various time points due to improvement or decline in the system function parameter itself that is being measured. In such cases, an approximate calculation of this measure can be made indirectly. We will investigate an imputation method that models the missing data mechanism based on information from the indirect measurements. Using simulations, this method will be compared to single imputation methods such as linear regression and hot-deck imputation and to multiple imputation methods. These methods will be evaluated with respect to the accuracy and precision of the trend estimates in a longitudinal analysis model.

## Web-Based Interactive Bayesian Multiple-Objective Optimal Designs for Dose Response Studies with Constraints

❖ Xiang-feng Wu, Stony Brook University, 8 Bowen Place, Stonybrook, NY 11790, *sterlingwu@yahoo.com*; Wei Zhu, Stony Brook University; Holger Dette, Ruhr-Universit‰t Bochum; Weng Kee Wong, University of California, Los Angeles

***Key Words:*** Bayesian multiple-objective optimal design, dose response study, logit model

We present the numerical algorithms for finding Bayesian multiple-objective optimal designs for dose-response studies subject to certain constraints. Dose-response studies are routinely conducted in drug design process. Due to safety, efficacy, and experimental design considerations, practical constraints are often imposed on dose range, dose levels, dose numbers, dose proportions and potential missing trials. The resulting controlled optimal designs satisfying these constraints can be readily adopted for optimal estimation of the parameters of interest such as the median effective dose level or the threshold dose level. In addition, we describe the methodologies and the implementation of a web based interactive optimal design platform for the practitioners. We demonstrate our results and methodology through the logit dose response model.

## Teaching Statistics to Clinical Research Staff in a Pharmaceutical Company

❖ Sunil Dhar, New Jersey Institute of Technology, 323 Martin Luther King Blvd., University Heights, Newark, NJ 07102, *dhar@njit.edu*; Farid Kianifard, Novartis Pharmaceuticals Corporation

***Key Words:*** statistics education, nonstatisticians

Education of clinical research staff in understanding statistical concepts is an area of importance for pharmaceutical companies. This understanding is needed to help them communicate with statisticians using a common language when designing clinical trials and interpreting clinical trial results. Such staff have little time for a one-semester, or even one-week, continuing education course in statistics. Faced with this reality, we developed a three-module course, for a total of 1.5 days taught over one month, that addresses the needs of this audience. We describe the format and content of the course and provide references that can serve as a resource for teaching such a course.

# 247 Methods for Agreement and Paired Data ●

Biometrics Section, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

## Concordance Correlation Coefficient for Ordinal Data

❖ Jingyun Yang, The Pennsylvania State University, 1003 W. Aaron Drive, 10C, State College, PA 16803, *mjyyang@stat.psu.edu*

***Key Words:*** kappa

Measurements of agreement are used to assess the reproducibility of a new assay or instrument, the acceptability of a new or generic process, methodology or method comparison. Kappa and weighted kappa are used when the responses are nomical and ordinal, respectively. We plan to develop a new method called multivariate kappa. Its asymptotic distribution will be given as well as its small sample performance compared with kappa and weighted kappa.

## Paradoxes Revisited: Comparison of B-Statistic with Kappa

❖ Shankar Viswanathan, The University of North Carolina at Chapel Hill, CB 7420, McGavran-Greenberg Hall, Chapel Hill, NC 27514, *shankar@bios.unc.edu*; Shrikant I. Bangdiwala, The University of North Carolina at Chapel Hill

***Key Words:*** Cohen's kappa, Bangdiwala's b, paradox, prevalence, bias

Kappa is the most commonly used reliability index to study agreement between two raters. Bangdiwala proposed the B-statistic that also corrects for chance agreement as kappa. Though kappa has been widely used, its interpretability and behavior has been questioned and debated. We revisit the paradoxes where kappa has been shown to ill-behave and compare how the B-statistic behaves under similar settings. Paradoxes: (1) Under highly symmetrically unbalanced marginals, a low kappa value observed for high observed diagonal proportion, while B-statistic was more consistently closer to Po. (2) Kappa values are higher for asymmetrical compared to symmetrical marginal imbalances, but this is untrue for B-statistic, thus eliminating the paradox. Under various bias & prevalence settings examined, B-statistics is more stable than kappa, thus we recommend using B-statistic as an alternative to kappa.

## Interobserver Variability and Intraobserver Variability

❖ Jing Han, St. Francis Hospital, P.O. Box 631, New York, 10156, *hanjlst@gmail.com*

***Key Words:*** observer variability, agreement studies, measurement studies, observer agreement, random effects models, ANOVA

The most elegant design of a clinical study will not overcome the damage caused by unreliable or imprecise measurement. The requirement that one's data be of high quality is at least as important a component of proper study design as the requirement for randomization, double blinding, controlling when necessary for prognostic factors, and so on. Existing indices of observer agreement for continuous data, such as the intraclass correlation coefficient or the concordance correlation coefficient, measure the total observer-related variability, which includes the variabilities between and within observers. This work introduces a new index that measures the observer agreement.

## Assessing Rater Exchangeability and Identifying an Atypical Rater Using a Log-linear Modeling Approach

❖ Kari Kastango, University of Pittsburgh, Graduate School of Public Health, 303 Parran Hall, 130 DeSoto Street, Pittsburgh, PA 15261, *kari.kastango@bms.com*; Roslyn A. Stone, Veteran's Affairs Pittsburgh Healthcare System

***Key Words:*** inter-rater agreement, nominal outcome

We propose a formal inferential approach to identify an atypical rater in an inter-rater agreement study involving six raters and a binary outcome. A global and heterogeneous partial agreement log linear model is fit and pair-wise comparisons of the six partial agreement parameters are made, using p-values adjusted for the number of comparisons made. Type I error and the power to correctly identify an atypical rater are assessed via simulation. The heterogeneous partial agreement parameters generally do highlight the most atypical rater in the scenarios considered, although the power is low to detect such a rater as statistically significantly different from the other raters. The pair-wise comparisons of the heterogeneous partial agreement parameters are quite likely to identify the correct rater as atypical when any rater is identified.

## The Test-Based Exact Confidence Intervals of the Difference between Two Proportions for Small-Sample Paired Binary Data

❖ Hong-Long Wang, National Taipei University, Department of Statistics, 151 University Road, San Hsia, Taipei, 237 Taiwan, *hlw@mail.ntpu.edu.tw*; Shiu-Ying Lin, National Taipei University; Xiao-Hua Andrew Zhou, University of Washington

***Key Words:*** small sample paired binary data, test based confidence intervals, Wald test, likelihood ratio test, studentized range statistics

To compare positive rates of two diagnostic tests in a paired design, Wald interval has poor coverage performance for the small sample. Zhou and Qin (2003) derived a better confidence based on the Edgeworth expansion for the studentized difference with the coverage probability coverges to the nominal level at a rate of O(n-1/2). We developed a confidence interval--based testing procedure. We showed that this confidence interval has exact conservative coverage probability. A computing program will be provided to compute confidence intervals for all combinations with sample size 5--30 using Wald test, likelihood ratio test, and studentized range statistics.

## Problems Related to Effect Size Measurement and Analysis

❖ Sibabrata Banerjee, New Jersey Institute of Technology, Department of Mathematical Sciences, University Heights, Newark, NJ 07102-1982, *sb95@njit.edu*

***Key Words:*** effect size, ROC, AUC

Comparing two treatments on the basis of their observed response is a situation which is commonly encountered in clinical research. More specifically, it is the quantity P(Y>X), denoted as Effect Size(ES). X and Y follows two unknown distributions F, G, respectively. Our interest lies in the estimation of ES and the inferences based on the amount of the shift D (=Y-X) of the vectors of observations X and Y; D itself is a random variable. We try to preserve the underlying uncertainty of the random shift and use the obtained information for our inference on ES. Receiver Operating Characteristics (ROC) Curve is a method originated in the field of signal processing and soon gained popularity in radiology. There is a close relation between ES and ROC and the area under the ROC curve(AUROC). Estimation of the ROC curve and the AUROC is also explored.

### Confidence Interval of Effect Size in Paired and Parallel-Group Study

❖ Jianrong Wu, St. Jude Children's Research Hospital, 332 N. Laucerdale Street, Memphis, TN 38105, *jianrong.wu@stjude.org*; Guoyong Jiang, Cephalon, Inc.

***Key Words:*** confidence interval, effect size, r^*-formula, signed log-likelihood ratio, small sample

Recent years have seen a heightened interest in estimating effect size, a common measure of effect magnitude, in biomedical research, because of its direct clinical relevance. In this article, we propose three methods for constructing confidence intervals of effect size that are applicable to randomized comparative parallel-group studies and paired study. We compare these methods using simulation to determine those that have good performance in terms of coverage probability. In particular, simulations show that the confidence interval generated by the modified signed log-likelihood ratio method possesses essentially exact coverage probabilities even for small samples. We provide two examples to illustrate the proposed methods.

## 248 Inference for Categorical Data ●

Biometrics Section

**Tuesday, August 8, 8:30 am–10:20 am**

### Exact Inference for Contingency Tables with Correlated Data

❖ Chris Corcoran, Utah State University, 3900 Old Main Hill, Logan, UT 84322-3900, *corcoran@math.usu.edu*

***Key Words:*** exact tests, contingency tables, correlated data, categorical data

Correlated categorical outcomes are common in biomedical applications such as genetics, ophthalmology, and teratology, where repeated measures are made over time, or clusters of subjects are sampled who share common genetic or environmental exposures. While a large body of work has yielded models for analyzing correlated categorical data, these methods generally rely on asymptotics to justify their inferences. We demonstrate how an exponential family model derived by Gourieroux et al (1984) and Zhao and Prentice (1990) allows us to formulate exact correlated-data analogues to conventional exact tests for two-way contingency tables, paying particular attention to tables of several ordered binomials or of two ordered multinomials. We discuss an efficient algorithm to obtain the exact conditional distribution, and illustrate the tests using examples from actual research.

### Exact Unconditional Inference for Multinomial Likelihoods with an Example Using 2x3 Contingency Tables

❖ Gerald Crans, Eli Lilly and Company, 389 Arbor Drive, Carmel, IN 46032, *crans889@msn.com*; Jonathan J. Shuster, University of Florida

***Key Words:*** exact unconditional inference, hypothesis testing, nuisance parameters

Exact unconditional hypothesis testing methods for 2x2 binomial data, 2x2 multinomial data, and correlated proportions exhibits a superior power advantage when compared to exact conditional approaches. Comparisons beyond the 2x2 testing scenario have not been explored due to the computational complexity of maximizing null power functions that depend on several nuisance parameters. A technique to compute the maximum of null power functions that result from the comparison of independent multinomial distributions is presented. As a special case, the 2x3 contingency table comparing independent trinomial distributions is considered. Exact critical values of the Pearson chi-square test are computed for n=10(1)70 and alpha = 0.05. Sample-size comparisons between Pearsons chi-square test statistic and the conditional p-value suggest the conditional p-value tends to be superior.

### Measuring Association in Contingency Tables

❖ Shailendra Banerjee, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, GA 30333, *snb1@cdc.gov*

***Key Words:*** association, symmetrical, contingency table

Objectives: This study aims at developing measures of association and their sampling errors for cross classification of nominal and ordinal variables. Methods: Lambda, the symmetric measure of association is obtained by taking arithmetic means of numerators and denominators of asymmetric measures. Three new measures for symmetrical situation were obtained by taking arithmetic and geometric means alternatively. For ordinal variables, Somers' measures were modified to combine the ties for both C and R to obtain a new symmetric measure d. Delta method was used to estimate the standard errors for the new measures. Methods were used on survey data obtained from 180 U.S. hospitals. Results: It is observed that although, these measures do not differ much from the known measures of association, their standard errors are lower for most of them.

### Data Information with the Cochran-Mantel-Haenszel Test

❖ Philip E. Cheng, Academia Sinica, Institute of Statistical Science, Taipei, 115 Taiwan, *pcheng@stat.sinica.edu.tw*

***Key Words:*** common odds ratio, likelihood ratio test, mutual information, relative entropy, three-way interaction

For combined 2x2 tables, the Bartlett test for no interaction and the Cochran-Mantel-Haenszel test for independence are examined. An information identity is formulated to characterize the division of data information shared between the two tests. This yields a two-step likelihood ratio test procedure that is more powerful than other tests for independence. The test scheme directly yields the Bartlett maximum likelihood estimate, and provides efficient interval estimation for the common odds ratio without using asymptotic variance estimates. As a byproduct, it evidences the drawback that the CMH test can be insensitive when the odds ratios are spread around unity.

### Modified Maximum Likelihood Estimation of the Proportional Odds Model

❖ Evrim Oral, Middle East Technical University, Department of Statistics, Middle East Technical University, Ankara, 06531 Turkey, *eoral@metu.edu.tr*

***Key Words:*** proportional odds model, modified maximum likelihood, order statistics, robustness

The proportional odds model is widely used for ordinal categorical responses. Since the likelihood equations are intractable and do not have explicit solutions, maximum likelihood estimation of the model parameters requires iterative procedures. We propose using the modified maximum likelihood (MML) method to lessen the difficulties encountered in maximum likelihood estimation of the proportional odds model. We show that the MML estimators, which are explicit functions of sample observations, are asymptotically fully efficient under some very general regularity conditions and highly efficient for small sample sizes. We study the robustness properties of these estimators via simulations. We also illustrate the method using real life examples.

### Using Variation Reduction Point To Remove Extraneous Effect in Modeling

❖ Chong Yau Fu, National Yang-Ming University, The Institute of Public Health, 155 sec 2 Li Long Street, Shih Pai, Taipei, 112 Taiwan, *chong@ym.edu.tw*; Shih-Hua Liu, National Yunlin University of Science & Technology

***Key Words:*** variation reduction, extraneous effect, age effect, stratification

In modeling, measuring the association between outcome variable and main covariate is the key issue. It may be biased from the effect of extraneous variable---age, for instance, which relates to the main covariate only---and with an indirect effect in the outcome variable. From point-of-data analysis, especially in several included covariates, it is not necessary to keep this extraneous variable in model. This study proposes a mean corrected method to remove this extraneous effect prior to modeling. This extraneous variable explains part of the total variation of the main covariate. The method removes this variation (among groups) through mean corrected technique. A Down syndrome study, relating to gestational age, is illustrated. The proposed method benefits from an easy approach to comparing with conditional logistic model using likelihood probability.

## 249 Design of Power and Sample Size in Clinical Trials ●

Biometrics Section, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Blinded Sample-Size Re-estimation in Randomized Block Trials with Continuous Endpoint

❖ Biao Xing, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, *xing.biao@gene.com*; Jitendra Ganju, BiostatWorks

***Key Words:*** sample size reestimation, adaptive design, internal pilot study, blinded estimation

Blinded sample-size re-estimation allows for modifying the sample size without compromising the integrity of the trial. One challenge in this is the blinded estimation of the variance while the trial is ongoing. Existing methods are often biased, based on untenable assumptions or applicable to only two-treatment trials. We propose a simple unbiased method, which uses subject enrollment order and randomization block size to estimate the variance. It can be applied to normal or non-normal data, trials with two or more treatments, equal or unequal allocations, and fixed or random block sizes. The proposed variance estimator performs best when the randomization block size is the smallest. Simulations and data analyses show that for many commonly used randomization block sizes, the proposed method performs well in re-estimating the variances and sample sizes in a blinded manner.

### Sample Size Re-estimation for Time-to-Event Studies

❖ Erinn Hade, The Ohio State University, 320 W. 10th Ave., M200 Starling Loving Hall, Columbus, OH 43210, *hade.2@osu.edu*; Soledad Fernandez, The Ohio State University; David Jarjoura, The Ohio State University

***Key Words:*** sample size re-estimation, survival, failure rate, time to event

Sample size re-estimation after an initial start-up period in clinical trials typically requires a planned interim analysis and re-estimation of the variance from the initial collected data. Authors have examined issues related to variance re-estimation and the timing of interim analyses, including options of blinded or unblinded data and the effect of unblinding on the bias of final estimates. Published work on sample size re-estimation has been closely examined for Gaussian and binomial outcomes. Some have explored time to event outcomes through a sequential testing design. We discuss issues related to sample size re-estimation for time to event analyses without the use of a sequential design and focus on studies of diseases that have low failure rates in both treatment and control arms. We present examples and simulations that show re-estimation can be critical with low failure rates.

### Sample Size and Power Calculation with Generalized Mixed Effects Model

❖ Qianyu Dang, University of Pittsburgh, 200 Meyran Ave., Suite 200, Pittsburgh, PA 15213, *dangq@upmc.edu*; Sati Mazumdar, University of Pittsburgh

***Key Words:*** power, sample size, GLIMMIX, penalized quasi-likelihood, mixed model

Generalized mixed effects model (GLIMMIX) has been widely used on correlated outcomes with all types of distributions. The model can now be easily implemented with SAS PROC GLIMMIX in its newest version 9. One way to fit this model is linearization by penalized quasi-likelihood (PQL) or marginal quasi-likelihood (MQL). Even though both of the methods have been shown to provide biased point estimates, the estimates for standard errors are found to be quite accurate. We derived formulas for power and sample size calculations using GLIMMIX based on PQL and MQL. A sensitivity study by simulation is carried out to show the impact of a mis-specified random effects structure. Examples are provided to illustrate use of the derived formulas for power and sample size calculations and results are compared with GEE approach.

### Design of Clinical Trials with Flexible Sample Size

❖ Lu Cui, sanofi-aventis, 902 Taggert Drive, Belle Mead, NJ 08502, *lu.cui@aventis.com*; Fang Liu, Merck Research Laboratories; Ray Zhu, sanofi-aventis

***Key Words:*** clinical trial, sample size, power, interval, regret score

Sample size determination is critical in design of clinical trials. An adequately sized trial can have sufficient power to detect the treatment difference of interest yet without over-sizing the study. However, the traditional sample size approach based on a single point initial projection of the true treatment difference may fail to achieve this goal if this initial projection misses its target for various likely reasons. In this study, a new approach to the sample size problem via minimizing regret score is proposed, assuming that the true treatment difference, though fixed, is in a likely interval rather than a single value. The regret score is used to select a better design with uniformly sufficient power and a near optimal sample size with respect to all possible values of the true treatment difference in the chosen interval.

### Approximate Confidence Intervals for Power in UNIREP Analyses

❖ Matthew Gribbin, The University of North Carolina at Chapel Hill, 101 Aberdeen Street, Biostatistics, Apt D, Chapel Hill, NC 27516, *mgribbin@gmail.com*; Keith E. Muller, The University of North Carolina at Chapel Hill; Jacqueline Johnson, The University of North Carolina at Chapel Hill

***Key Words:*** power, confidence intervals, UNIREP, Geisser-greenhouse, Huynh-Feldt

Unlike standard mixed model tests which often have inflated test size, particularly in small samples, and have few methods developed for computing power, tests using the univariate approach to repeated measures (UNIREP) accurately control test size even in small samples and have available accurate power approximations. Methods presented in Muller, et al (2005) give new power approximations for the Geisser-Greenhouse, Huynh-Feldt, Box Conservative and Uncorrected UNIREP tests. We present new theoretical methods for computing accurate confidence interval approximations for power in UNIREP tests based on the methods developed in Muller, et al (2005). We perform simulations of these confidence intervals for a variety of examples. We illustrate the application of the methods using Diffusion Tensor Imaging data.

### Designing Standardly Tailored Multicomponent Intervention Trials in Medicine To Yield Unbiased Effect Estimates

❖ Heather G. Allore, Yale University, School of Medicine, Program on Aging, 1 Church Street 7th Floor, New Haven, CT 06510, *heather.allore@yale.edu*

***Key Words:*** experimental design, medicine, multicomponent interventions, standardly-tailored

Many clinical trials test a single therapeutic agent. However, many health problems are multifactorial in etiology. A type of design (standardly-tailored multicomponent interventions) modifies risk factors, thus, attempts to reduce adverse health outcomes. Modifiable risk factors are delineated and intervention components specific to the risk factors are assigned per protocol. While standardly-tailored designs are clinically relevant and yield an estimate for the overall intervention effect, previous applications have not provided unbiased estimates of individual component effects because components often target several risk factors that may be confounded. Fractional factorial designs can estimate main effects at the sacrifice of higher-order interactions. We propose guidelines for the design of standardly-tailored multicomponent intervention trials that yield unbiased effect estimates.

### Evaluation of Several Multiple Comparison Procedures for Noninferiority Trials with Two Doses of a Treatment and a Control with a Binary Success Rate Endpoint

❖ Hongling Zhou, U.S. Food and Drug Administration, 2251 Mount Hebron Drive, Ellicott City, MD 21042, *hongling.zhou@fda.hhs.gov*; Mohammad Huque, U.S. Food and Drug Administration

***Key Words:*** non-inferiority, multiplicity, clinical trial, hypothesis testing

Some non-inferiority trials use 2 doses of the test drug for claim of non-inferiority for each dose to the active control, which should be adjusted for multiplicity. As the traditional Bonferroni and Sidak adjustments are conservative, several other approaches could be taken using the closed testing principle to increase the power of the test. These procedures first test for the intersection null hypothesis H0 (intersection of H01 and H02) for passing the gate that clears the assertion that at least one of the two doses is non-inferior to the active control without being specific. After H0 is rejected at level a, then H01 and H02, which are inferiority null hypotheses for dose 1 and 2 respectively, are then each tested at the same level a. We will address several tests in this regard and will show their asymptotic properties and power performance.

## 250 Censoring and Modeling with Censoring ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Tuesday, August 8, 8:30 am–10:20 am**

### Interval Estimation for a Measure of Stochastic Superiority

❖ Euijung Ryu, University of Florida, 300 Diamond Village, Apt. 6, Gainesville, FL 32603, *eryu@stat.ufl.edu*

***Key Words:*** measure of stochastic superiority, Mann-Whitney statistic, multinomial distribution, confidence interval

When we are interested in comparing two independent groups that have ordered categorical measurements, we can use a measure of stochastic superiority to describe the effect size. This measure embodies the relative size of outcomes, rather than actual magnitude. We investigate three confidence interval methods: the Wald-type, score-type, and likelihood ratio-based, which are based on the Mann-Whitney statistic allowing ties. These confidence intervals are obtained by considering data from two independent multinomial distributions. We consider an unrestricted and a parametric model, depending on whether there exists a relationship between cell probabilities. Simulation studies show it is better to use an unrestricted model under several conditions.

### Estimating Equation Approach for Regression Analysis of Interval-Censored Failure Time Data

❖ HeeJeong Lim, California State University, Fullerton, Department of Mathematics, 800 North State College Blvd, Fullerton, CA 92834, *heejeonglim@fullerton.edu*; Xingqiu Zhao, McMaster University; Jianguo Sun, University of Missouri-Columbia

***Key Words:*** estimating equations, interval-censoring, proportional hazards models, regression analysis

This article discusses statistical inference for the proportional hazards model when there exists interval-censoring on both survival time of interest and covariates (J. Roy. Statist. Soc. B 34, (1972) 187; Encyclopedia of Biostatistics. Wiley, New York, 1998, pp.2090-2095). In particular, we consider situations where observations on the survival time are doubly censored and covariates are interval-censored. For inference about regression parameters, a general estimating equation approach is proposed. The proposed estimate of the parameter is a generalization of the maximum partial-likelihood estimate for right-censored failure time data with known or exactly observed covariates (Statistical Analysis of Failure Time Data. Wiley, New York, 1980). The asymptotic properties of the proposed estimate are established and its finite sample properties are investigated through a simulation study.

## Sensitivity of the Kaplan-Meier Estimate to Nonignorable Censoring

❖ Tao Liu, University of Pennsylvania School of Medicine, Department of Biostatistics and Epidemiology, 501 Blockley Hall, Philadelphia, PA 19104, *tliu@cceb.upenn.edu*; Daniel F. Heitjan, University of Pennsylvania

***Key Words:*** coarse-data model, ignorability, informative censoring, ISNI, Kaplan-Meier curve, sensitivity analysis

Untestable assumptions about association between survival and censoring times can affect the validity of estimates of the survival distribution, including the Kaplan-Meier (KM) nonparametric MLE. This article explores the sensitivity of KM to nonignorable censoring by extending the index of local sensitivity to nonignorability (ISNI) (Troxel, Ma, Heitjan 2004, Zhang 2004) to the case of a nonparametric survival model. The method involves first specifying a coarse-data selection model to describe the association between the failure and censoring processes, then evaluating the slope of the nonparametric survival MLE with respect to a nonignorability parameter in the neighborhood of the ignorable model. We show in a simulation that ISNI analysis captures local sensitivity to nonignorability well. We apply the method to the Stanford Heart Transplant Data.

## Censoring Point in Logistic ELISA Standard Curves

❖ Ryan E. Wiegand, Medical University of South Carolina, P.O. Box 250853, 135 Cannon St, Suite 303, Charleston, SC 29425, *wiegand@musc.edu*; Elizabeth H. Slate, Medical University of South Carolina; Elizabeth G. Hill, Medical University of South Carolina; Jyotika K. Fernandes, Medical University of South Carolina

***Key Words:*** limit of detection, censoring, bias, clustering

We address the analysis of concentrations of various cytokines in gingival crevicular fluid (GCF) arising from a multiplex ELISA study where GCF was collected at multiple sites from each patient. Since a sample's cytokine concentration is determined by back-solving from an optical density via the standard curve, any sample with an optical density within the curve's asymptotes will map to a reliable concentration. Values close to the low asymptote, however, are unreliable because they may be indistinguishable from zero and, similarly, values at the high asymptote represent saturated signals. Hence, these samples must be censored. Using an analysis method that simultaneously accounts for left- and right- censoring and clustering, we explore various methods of determining the censoring points and evaluate the benefits of each method. This work was funded by NIH grant P20 RR-017696.

## Biases Due to Left Censoring and Left Truncation in Estimating Menopausal Transition Stages from Menstrual Cycle Data

Kevin Cain, University of Washington; ❖ Bin Nan, University of Michigan, *bnan@umich.edu*

***Key Words:*** menstrual cycle, menopause, left censoring, left truncation

Restage is a collaborative study to analyze data from four studies of menstrual cycle calendar data. Among the purposes of restage are to estimate the median age at entry to early and late menopausal transition stages as marked by the first occurrence of cycle length irregularity and a long cycle length, respectively, and to estimate the median age at menopause. In this presentation we compare results from the four data sets and also perform simulations based on the Tremin data to illustrate the bias caused by ignoring left censoring and left truncation. We also explore possible solutions to this bias. We conclude that biases due to women entering a study at older ages can be large, especially for analyses of the age at entry to early stage. Results are presented based on simplistic methods for dealing with the truncation and censoring, but more sophisticated methods need to be developed.

## Predicting Exposure at a Specified Time Based on an Extended Randomized Regression Model for Interval-Censored Data

❖ Robert Lyles, Emory University, Department of Biostatistics, 1518 Clifton Rd NE, Atlanta, GA 30322, *rlyles@sph.emory.edu*; Amita K. Manatunga, Emory University; Renee Moore, Emory University; Michele Marcus, Emory University

***Key Words:*** coarse data, environmental epidemiology, random effects, reproductive health

Motivated by a study aiming to link decaying maternal serum levels of an environmental exposure with reproductive outcomes among offspring, we use a random regression model to predict unmeasured in-utero exposures. The model is complicated due to coarse exposure data, most of which is reported to the nearest integer. We treat these data as interval-censored realizations of underlying lognormal exposure levels and maximize the resulting integrated likelihood. We derive empirical Bayes and constrained Bayes predictions of in-utero exposures accounting for the censoring and use simulations to compare their performances with those of estimation approaches proposed in the reproductive health literature. In addition to simulation results, we provide an analysis of data consisting of longitudinal serum measurements of polybrominated biphenyls (PBBs) from the Michigan Female Health Study.

## Inference for Partially Observed Systems with an Application to Cholera

Edward L. Ionides, University of Michigan; ❖ Carles Breto, University of Michigan, 439 West Hall, Ann Arbor, MI 48109, *cbreto@umich.edu*; Aaron A. King, University of Michigan

***Key Words:*** nonlinear state space model, maximum likelihood estimator, cholera

Inference for nonlinear state space models can be a challenging step in developing models for stochastic dynamical systems. The system investigated as a case study is a population model for cholera in Bangladesh. This example motivates the introduction of a new method for likelihood based inference, which we call MAPLE (Maximum A Posteriori Limit Estimation). The MAPLE algorithm computes a maximum likelihood estimator as the limit of an average of Bayesian posterior mean estimators. The new methodology has computational advantages and a theoretical justification.

# 251 Design and Analysis Issues in Epidemiological Modeling ● ✪

Section on Statistics in Epidemiology

**Tuesday, August 8, 8:30 am–10:20 am**

## Data Analysis under Case-Cohort Designs with Clustered Binary Outcome Data

❖ Shou-En Lu, University of Medicine & Dentistry of New Jersey, 683 Hoes Lane, W., Rm 220, Piscataway, NJ 08854, *lus2@umdnj.edu*; Yong Lin, University of Medicine & Dentistry of New Jersey; Joanna H. Shih, National Cancer Institute

***Key Words:*** case-cohort design, estimating function, intra-cluster association, binary outcomes

Cluster case-cohort design (Lu and Shih 2006) is a cost effective design that incorporates cluster structure in a cohort into the sampling frame. In contrast to assuming all cohort members to be independent, it accounts for the dependency between cluster members and extends the well-known case-cohort design of Prentice (1986) which only applies for cohorts of independent individuals. This paper aims to develop a statistical methodology for analyzing binary outcomes under the cluster case-cohort design. Statistical inference is developed to estimate the regression parameters in the marginal logistic regression model and the intra-cluster association. Statistical properties of the proposed estimators are developed. The performance and statistical efficiencies of the proposed estimators are investigated with simulation studies. A data example is used to illustrate the proposed methodology.

## Early, Cost-Effective Identification of High-Risk/Priority Control Areas in Foot-and-Mouth Disease Epidemics

❖ Steven Schwager, Cornell University, Department of BSCB, 434 Warren Hall, Ithaca, NY 14853, *sjs5@cornell.edu*; Ariel Rivas, Cornell University; Stephen Smith, Cornell University; Antoni Magri, Cornell University

***Key Words:*** foot-and-mouth disease, epidemic, multivariate methods

Data from the 2001 Uruguayan Foot-and-Mouth Disease epidemic were examined to seek improved cost-benefit based policies. Variables analyzed were location and size of 4,022 individual land parcels (574 infected over 60 days); animal density; percentage of dairy farms per county; and road density. Each variable was categorized and the cases per class at epidemic days 1-3 and 4-6 were compared. More cases were found at days 4-6 than 1-3 in areas with small parcels, high animal density, >20% dairy farms, and high road density (each p< 0.03). These classes had greater proportions of cases at days 7-60 than their proportions of total area. The region constructed by intersecting the classes with more cases at days 4-6 included 50.4% of all cases at days 7-60 in only 30.6% of the area. The area per case in this region was =33% lower and covered =45% less area than any one-variable approach.

## Capture-Recapture Revisited

❖ Lawrence Lessner, SUNY University at Albany, One University Place, Room 150, Rensselaer, NY 12144, *LLessner@nycap.rr.com*

***Key Words:*** capture recapture, Poisson processes, marked process

The objective of this work is to present a probability model using nonhomogeneous marked Poisson processes to model the data in a two-source capture-recapture situation. The model begins by assuming the incidence of membership in a study population is a nonhomogeneous process. A sequence of constructions is obtained that result in four stochastically independent process:< Z(A,j)> that models incidence of persons using only source A, < Z(B,j)> incidence for B only, < Z(AB,j)> incidence for both A and B, and < Z(0,j)> is the incidence of subjects not selecting A or B. The probability of capture in the three cells is no longer required to be a constant. Intensities for all processes have close-form expressions.

## Heterogeneity on Estimation Size of Hard-to-Reach Population Using Capture-Recapture Data

❖ Shenghai Zhang, Public Health Agency of Canada, 16 Oldfield Street, Ottawa, ON K2G 6V8 Canada, *s.zhang@rogers.com*

***Key Words:*** capture-recapture, hard-to-reach people, epidemiologic methods, generalized estimating equations

Capture-recapture estimation methods for closed wildlife population has been adapted by epidemiologists to estimate the size of a hidden or hard-to-reach population. The heterogeneity of capture probabilities on the estimation of population size using capture-recapture data is considered in this presentation. A generalized estimating equation approach to the estimating capturing probabilities is presented by considering the heterogeneity of the study population. Estimated probabilities, then serve as denominators for calculating the size of the population.

## Analysis of a Disease and Probability of Exposure Association Using a Replicated Error-Prone Exposure Assessment

❖ Chengxing Lu, Emory University, 3448 N. Druid Hills Road, Apt N, Decatur, GA 30033, *clu@emory.edu*; Robert Lyles, Emory University

***Key Words:*** beta-binomial, bias, logistic regression, misclassification, replicates

In environmental epidemiologic studies, it is common for a binary exposure to be assessed multiple times in a manner subject to misclassification. In a case-control setting, we focus on exploring the association between a disease and the probability of exposure given such replicates and in the absence of a gold standard for exposure. Assuming a beta distribution for the exposure probability, we obtain the estimated association by maximizing the marginal likelihood of the observed exposure replicates and the disease status. In simulation studies, we compare the performance of the proposed method with that of a logical but biased approach that replaces the unknown true exposure probability by the sample mean of the replicates. The proposed method is shown to be superior in terms of bias and confidence interval coverage. A real data example will also be presented for demonstration.

## A Two-Phase Analysis of Air Pollution and Adverse Birth Outcomes

❖ Katherine Hoggatt, University of Michigan, 109 Observatory Street, 3010 SPH 1, Ann Arbor, MI 48109, *khoggatt@umich.edu*; Sander Greenland, University of California, Los Angeles; Beate Ritz, University of California, Los Angeles

***Key Words:*** epidemiology, two-phase, two-stage, case-control

We show how a two-phase design and analysis can be used to collect detailed covariate data on a subsample and correct for possible selection bias and improve the efficiency of the estimates relative to what would be expected if only the subsample were analyzed. Our Phase 1 group was year 2003 births in select Los Angeles ZIP codes. Our Phase 2 group was births selected for the nested case-control study where the mother responded to the study questionnaire. To assess possible selection bias, we estimated the association between ambient carbon monoxide exposure and low weight birth (< 2500g) in the Phase 1 and 2 groups with only first-phase variables included in the models. We then used three two-phase estimators, which allowed us to use second-phase covariates to better control for confounding, reduce the apparent selection bias, and increase the efficiency of our estimates.

## 252 Survey-Based Variance Estimation I ●

Section on Survey Research Methods

**Tuesday, August 8, 8:30 am–10:20 am**

### Confidence Intervals for Quantile Estimation from Complex Survey Data

❖ Babubhai V. Shah, SAFAL Institute Inc., 22 Autumn Woods Drive, Durham, NC 27713, *babushah@earthlink.net*; Akhil Vaish, RTI International

***Key Words:*** quantile estimates, confidence intervals, complex survey data

A method for estimating quantiles and their confidence intervals based on the paper by Francisco and Fuller (1991) has been implemented in SUDAANÆ software. There have been problems encountered in practical application of the Francisco-Fuller method: It requires evaluation of bounds at many points, the resulting limits are, at times, not monotonic or internally consistent, and the accuracy is not very good. The objective of this paper is to develop an improved method that overcomes these problems. The revised method is based on the confidence intervals based on the estimating equations (Godambe 1999). An adjustment to the empirical distribution is proposed to achieve internal consistency and reduce bias.

### Estimating Prevalence When a Subset of the Sample Requires Further Evaluation

❖ David C. Hoaglin, Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138, *dave_hoaglin@abtassoc.com*; K. P. Srinath, Abt Associates Inc.

***Key Words:*** survey, variance estimation, Chronic Fatigue Syndrome

Population-based estimates of the prevalence of some health conditions require that a subset of the survey sample have a clinical evaluation, in addition to an initial personal interview. Chronic fatigue syndrome (CFS) provides a motivating example: an interview can elicit whether respondents have the symptoms that are part of the definition of CFS, as well as whether they have exclusionary conditions, but the actual diagnosis of CFS depends on laboratory data and the results of a physical examination and other tests. Thus, all subjects with sufficient symptoms (and no exclusions) are eligible for clinical evaluation, and other subsets of subjects may be selected for comparison. For types of fatiguing illness whose definition involves clinical data, the paper examines two approaches for estimating standard errors of prevalence estimates.

### Variance Estimation for Ordered Categories

❖ Steven Pedlow, National Opinion Research Center, 55 E. Monroe Street, Suite 4800, Chicago, IL 60603, *pedlow-steven@norc.uchicago.edu*; Yongyi Wang, National Opinion Research Center; Kirk Wolter, National Opinion Research Center

***Key Words:*** Taylor series, jackknife, balanced half samples, balanced repeated replication, BRR, BHS

The National Longitudinal Survey of Youth (NLSY97), sponsored by the U.S. Department of Labor, examines youth transitions from school to work. Approximately 9,000 youths born in 1980-1984 were interviewed in 1997 and are still being interviewed annually. Job classification information has been collected and the sum of the top five occupations of youths for various subgroups has been reported. We consider this top five variable to be a measure of concentration. Since the ranking of occupations is subject to sampling error, calculating variances for the sum of the top five is an interesting problem. The variance in the measure of concentration is not equivalent to the variance of being in the actual top five observed occupations. This paper compares several different methods of variance estimation for this measure of concentration.

### Variance Estimation in a Survey for Iowa's State Board of Education

Michael D. Larsen, Iowa State University; ❖ Lu Lu, Iowa State University, 202D Snedecor Hall, Dept. of Statistics, Ames, IA 50014, *icyemma@iastate.edu*

***Key Words:*** stratified multistage sampling, one-psu-per-stratum design, non-invariant design, collapsed stratum estimator, generalized variance function, ratio estimator

Iowa's State Board of Education carried out a stratified multistage sample survey to study the availability of employment preparation courses and the degree to which students enroll in those courses in Iowa's public schools. Three survey sampling issues arose in the design that was implemented. First, the design actually is not invariant. Usual variance estimators are compared to simulated sampling distributions. Second, some PSUs in the stratified multistage sampling design have only one PSU per stratum. The collapsed variance estimator, model-based estimators based on generalized variance functions, and a ratio estimator based on auxiliary variables are proposed and compared. Results are based on a simulation study. Third, as part of ongoing work, the relative costs of adding schools versus adding students to the sample will be evaluated.

### Variance Estimation in Complex Survey Sampling for Generalized Linear Models

❖ Sundar Natarajan, New York University Medical Center, 423 E. 23rd Street, Room 11101S, New York, NY 10010, *sundar.natarajan@med.nyu.edu*; Stuart Lipsitz, Brigham and Women's Hospital; Garrett Fitzmaurice, Harvard University; Charity G. Moore, University of South Carolina

***Key Words:*** population survey, quasi-likelihood, weighted estimating equations, medical expenditure panel survey

In complex surveys (sample a fraction of a large finite population) each subject has a different probability of selection. For generalizing from the sample to the population, the analysis must incorporate both the design and the probability of selection. We focus on non-standard regression models for complex surveys; the outcome variable is the subject's total yearly health care costs. In prior analyses of medical cost data the variance was approximately equal to the mean raised to the power of 1.5, a non-standard variance function. Currently, regression parameters for this model cannot be estimated in standard packages for analyzing sample surveys. We propose a simple, two-step method to obtain consistent regression parameter and variance estimates; the method can be implemented within any standard sample survey package and is applicable to complex surveys with any number of stages.

### Variance Estimation of the Survey-Weighted Kappa Measure of Agreement

❖ Moshe Feder, RTI International, Cox Building 317, POBox 12194, Research Triangle Park, NC 27709, *mfeder@rti.org*

***Key Words:*** complex sample, Taylor approximation, kappa, variance estimation, re-interview

The National Survey of Drug Use and Health (NSDUH) is currently conducting a study to assess the reliability of responses. An interview/re-interview method is employed where individuals are interviewed on two occasions. Reliability is assessed by comparing the responses from the two interviews. To measure the reliability of categorical responses, the kappa (?) index of inter-rater reliability is used. This measure is the statistic most-often used to assess reliability of categorical variables. The common asymptotic variance estimation assumes simple random sample. The NSDUH sample design is complex, affecting the point estimates of ? and its variance. While correcting the point estimate of ? for the design is straightforward, the variance estimate is more involved. We present a Taylor linearization derivation, along with numerical results and comparisons with the asymptotic formula

### Improved Confidence Intervals for the Bernoulli Parameter

Wheyming Tina Song, National Tsing Hua University; ❖ Chia-Jung Chang, National Tsing Hua University, Lab 720, Engineering Building 1, NTHU, 101 Section 2 Kuang Fu Road,, Hsinchu, 30013 Taiwan, *g943808@oz.nthu.edu.tw*

***Key Words:*** Bernoulli process, confidence interval, Monte Carlo experiments

Despite the simplicity of the Bernoulli process, developing good confidence interval procedures for its parameter-the probability of success p-is deceptively difficult. The binary data yield a discrete number of successes from a discrete number of trials, n. This discreteness results in actual coverage probabilities that oscillate with the n for fixed values of p (and with p for fixed n). We suggest ideas designed to reduce the effect of discreteness on coverage p probability while having minimal effect on properties of the interval width. The ideas are of four types: ways to improve choice of the Student-t distribution's degrees of freedom, ways to improve the standard-error estimator, ways to apply permutation, and ways to combine intervals. We discuss four specific new procedures and illustrate improved performance with Monte Carlo experiments.

## 253 Survey-Based Estimation II ●

Section on Survey Research Methods
**Tuesday, August 8, 8:30 am–10:20 am**

### Methods for Birth Cohort Analysis for the National Immunization Survey

❖ Phil Smith, Centers for Disease Control and Prevention, 2816 Summercrest Lane, Duluth, GA 30096, *pzs6@cdc.gov*; Kirk Wolter, National Opinion Research Center; James Singleton, Centers for Disease Control and Prevention

***Key Words:*** birth cohort, cross-sectional survey, survey weight

The National Immunization Survey (NIS) is a cross-sectional survey conducted annually to estimate vaccination coverage in a target population of children 19-35 months of age. Annual estimates of vaccination coverage have traditionally been obtained using the sample of children obtained from a calendar-year sample. Each annual calendar-year sample includes several different birth cohorts. Also, each birth cohort may have different vaccination coverage at a specific age because they can experience different temporal events before that age that affects vaccination coverage, such as a vaccine shortage. In our paper we describe methods for combining data from several NIS calendar-year samples to estimate vaccination uptake for birth cohorts.

### Estimation of Regression Coefficients with Unequal Probability Samples

❖ Yu Wu, Iowa State University, 208A Snedecor, Iowa State University, Ames, IA 50011, *yuwu@iastate.edu*; Wayne Fuller, Iowa State University

We compare alternative estimators for regression coefficients estimated with data from a complex survey. The ordinary least squares estimator is a common choice for researchers, but, under an informative design, the ordinary least squares estimator is biased. The design-weighted estimator is consistent but may have a large variance. Design-consistent estimators based on generalized least squares and instrumental variable procedures are compared theoretically and in a Monte Carlo study.

### Indirect Sampling and the Problem of Identification of Links

❖ Pierre Lavallée, Statistics Canada, R.H. Coats Building, 15J, 120 Parkdale Ave, Ottawa, ON K1A 0T6 Canada, *pierre.lavallee@statcan.ca*; Xiaojian Xu, University of Alberta

***Key Words:*** weighting, over-estimation, record linkage, logistic regression, calibration

With indirect sampling, we have two populations---UA and UB---linked where we want to produce estimates for UB. Unfortunately, we have only a sampling frame for UA. We then draw a sample from UA to produce estimates for UB using the existing correspondence (or the existing links) between the two populations. To calculate the weights relative to the units surveyed within UB, we use the generalized weight share method (GWSM) that uses the links existing between UA and UB. Unfortunately, it happens in several applications that the links between UA and UB are difficult to identify. In general, we underevaluate the number of links, which causes overestimation of totals. After an overview of the GWSM, we will present in detail the problem of links identification along with proposed solutions. We then will study the performance of each solution.

### A Comparison between Ratio Estimation and Post-Stratification

❖ Chang-Tai Chao, National Cheng Kung University, No 1 University Road, Department of Statistics, Tainan City, 701 Taiwan, *ctchao@stat.ncku.edu.tw*; Tzu-Ching Chiang, National Cheng Kung University

***Key Words:*** ratio estimation, post-stratification, auxiliary variable, relative efficiency, design-based, model-based

To improve the estimation in a survey sampling problem, ratio estimator is recommended often when auxiliary variable is available. By taking advantage of the correlation between the variable of primary interest and the auxiliary variable, though biased, ratio estimator often can provide more efficient estimation result than sample mean. On the other hand, post-stratification is used when the population cannot be stratified before the sample is selected. It also is able to provide more efficient estimate than the overall sample mean of the variable of interest. In this research, these two estimation methods are compared in terms of their relative efficiency. We also discuss whether ratio estimator or post-stratification should be used in practice.

### Using Income as an Auxiliary Variable To Improve the Design of Household Expenditure Surveys

❖ Charles Mitchell, Statistics Canada, 890 Cahill Drive W., Unit 48, Ottawa, ON K1V 9A4 Canada, *charles.mitchell@statcan.ca*; Christian Nadeau, Statistics Canada

***Key Words:*** household surveys, stratification, allocation, spending

When conducting household surveys, household spending habits often are linked to household income. However, for some expenditure variables, the relationship with household income is not linear. For example, the probability of reporting a nonzero value seems to increase with household income for spending on furniture and household renovations. This variety of relationships provides a challenge in constructing an efficient sampling design. The aim of this paper is to compare alternate sampling designs with the goal of reducing the variance of estimates that show different types of relationships with household income. We will consider different two-stage sampling designs using an area frame. The variance for different estimates will be compared for alternative allocation methods that use income as an auxiliary variable.

### Model Averaging in Survey Estimation

❖ Xiaoxi Li, Iowa State University, Department of Statistics, 204 Snedecor, Ames, IA 5001, *lixiaoxi@iastate.edu*; Jean D. Opsomer, Iowa State University

***Key Words:*** local polynomial regression, cross-validation, regression estimation

Model averaging is a widely used method as it accounts for uncertainties in model selection. However, its applications in survey estimation are much to be explored. We propose a model-averaging (MA) regression estimator for the population total. Different ways to obtain this estimator are explored through simulation studies.

### Integrating Person and Housing Unit Weighting for the Current Population Survey

❖ Andrew Zbikowski, U.S. Census Bureau, DSMD, Washington, DC 20233-0001, *andrew.a.zbikowski@census.gov*; Phawn Letourneau, U.S. Census Bureau

***Key Words:*** surveys, estimation, weighting

The Current Population Survey (CPS) produces national, state and substate demographic and economic estimates of people, householders and housing units. Although the estimated numbers of householders and housing units should agree by definition, there has long been a discrepancy in these estimates from the CPS and from other demographic surveys sponsored by the Census Bureau. The Census Bureau formed a group of statisticians to study the weighting methods used by the demographic surveys, and to make recommendations. This group recommended research on weighting methods to integrate population and housing controls that would produce equal estimates of householders and housing units. For the CPS, we are researching possible new adjustments for housing unit weighting. This paper will discuss ongoing research on weighting methods for the CPS, and present some preliminary results.

## 254 Nonparametric Regression Methods II ●

Section on Nonparametric Statistics

**Tuesday, August 8, 8:30 am–10:20 am**

### Spline Single-Index Prediction Model

❖ Li Wang, Michigan State University, A400 Wells Hall, East Lansing, MI 48824, *wangli4@stt.msu.edu*; Lijian Yang, Michigan State University

***Key Words:*** b spline, geometric mixing, knots, nonparametric regression, root-n rate, strong consistency

For the past two decades, single-index model has proven to be an efficient way of coping with the high-dimensional problem in nonparametric regression. In this paper, we investigate the single-index prediction based on weakly dependent sample. The single index is identified by the best approximation to the multivariate prediction function of the response variable, regardless of whether the prediction function is a genuine single-index function. A polynomial spline estimator is proposed for the single-index coefficients and shown to be root-n consistent and asymptotically normal. An iterative program based on Newton-Raphson algorithm is developed. Simulation experiments have provided strong evidence that corroborates with the asymptotic theory. The algorithm is sufficiently fast for the user to analyze large data of high dimension within seconds.

### Testing Lack-of-Fit of Heteroscedastic Nonlinear Regression Models with Local Linear Smoothers

❖ Chin-Shang Li, St. Jude Children's Research Hospital, 332 N. Lauderdale Street, Memphis, TN 38105, *chinshang.li@stjude.org*

***Key Words:*** bandwidth selection, boundary effects, fit comparison, local linear kernel, quasi-likelihood estimator

A data-driven test is proposed for assessing the appropriateness of heteroscedastic nonlinear regression models by using local linear regression smoothers in which no boundary-corrected kernels are needed to resolve boundary effects. The bandwidth is selected based on the asymptotically optimal bandwidth under the parametric null model. This selection method leads to the data-driven test. It is shown that the test statistic is normally distributed under the null hypothesis and the test is consistent against any fixed alternative. The resulting test can be applied for the lack-of-fit of a postulated generalized linear model and is compared to existing tests. A real-life dataset is used to demonstrate the practical use of the proposed test.

### Local Analytic Curve Estimation

❖ Richard Charnigo, University of Kentucky, 851 Patterson Office Tower, Lexington, KY 40506-0027, *richc@ms.uky.edu*; Cidambi Srinivasan, University of Kentucky

***Key Words:*** nonparametric regression, global modeling, local modeling, compound estimator, calculus of variations

Consider a nonparametric regression problem in which the mean curve $\mu(x)$ is analytic. Seeking an analytic estimator with a tractable closed-form expression and without prior specification of an orthonormal basis, we formulate an integrated weighted least squares functional. Minimization of this functional reduces to solution of a second-order ordinary differential equation. The raw estimator is not consistent, but some modifications permit consistent estimation of Taylor approximants and the construction of a consistent ``compound estimator''. The compound estimator is analytic and can recover $\mu(x)$ at a uniform rate of nearly $n^{-1/2}$ on a compact set.

### Nonparametric Regression with Coarsened Predictors

❖ Aurore Delaigle, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, *delaigle@math.ucsd.edu*

***Key Words:*** deconvolution, errors-in-variables

We consider nonparametric estimation of a regression function in an errors-in-variables (EIV) problem. In the classical EIV problem, the goal is to estimate a function m, where Y=m(V)+\epsilon and a sample of (U,Y) is available, where U=V+\delta. See Fan and Truong (1993). In this talk, we assume that a sample of (W,Y) is observed, where Y=g(W)+\varepsilon, but instead of estimating g, the goal is to estimate m(x)=E(Y|X=x), where X=W+\delta and \delta is a measurement error. The motivating idea is that a training sample of accurate observations is often available, and future values of Y can be predicted from easier-to-obtain contaminated observations X of W. We propose a nonparametric estimator of m, discuss its theoretical properties and illustrate its performance via a simulation study and on a real data example. This is joint work with Peter Hall and Hans Muller.

### Asymptotic Approximation to a Nonparametric Regression Experiment with Unknown Variance

❖ Andrew Carter, University of California, Santa Barbara, Department of Statistics, UCSB, Santa Barbara, CA 93106-3110, *carter@pstat.ucsb.edu*

***Key Words:*** nonparametric regression, asymptotic equivalence of experiments

Brown and Low (1996) showed nonparametric regression experiments are asymptotically equivalent (in the sense of Le Cam's deficiency distance) to a Brownian motion process with an unknown drift function. Their result requires that the variance of the observations is known, but these variances generally are considered unknown nuisance parameters. Including the variance as an extra parameter in the definition of the regression experiment changes the form of the limiting experiment because a Brownian motion is completely informative about its variance. The appropriate approximation is a mixture of continuous Gaussian processes with different variances. The connection between the nonparametric regression and the mixed Gaussian process will be demonstrated for a homogeneous variance and a smooth variance function.

## 255 Biomarkers and Joint Models

●

Biometrics Section, WNAR, ENAR
**Tuesday, August 8, 8:30 am–10:20 am**

### Mixtures-of-Experts Joint Model for Longitudinal Biomarkers and Recurrent Events

❖ Jun Han, Georgia State University, Mathematics and Statistics Department, PO Box 4110, Atlanta, GA 30302-4110, *jhan@mathstat.gsu.edu*; Elizabeth H. Slate, Medical University of South Carolina; Edsel Peña, University of South Carolina

***Key Words:*** latent class, joint model, heterogeneous population, recurrent event, longitudinal biomarker

We propose a mixtures-of-experts joint model for a longitudinal biomarker and recurrent event, which accommodates the effects of covariates on the biomarker and event processes, the effects of accumulating event occurrences, effects caused by interventions after each event occurrence, and an underlying heterogeneous population. The EM algorithm for MLE and a penalized likelihood measure are employed to estimate the model parameters and the number of latent classes. A combined distance for biomarker and recurrent event data is suggested for initial clustering of units to derive starting values for the EM algorithm, and the performance of multiple clustering methods are compared. Properties of the parameter estimates, including four criteria for determining the number of latent classes, are examined by simulation.

### Joint Analysis of Longitudinal Measurements and Competing Risks Failure Time Data

❖ Ning Li, University of California, Los Angeles, 3281 S. Sepulveda Blvd., Apt 307, Los Angeles, CA 90034, *ningli@ucla.edu*; Robert Elashoff, University of California, Los Angeles; Gang Li, University of California, Los Angeles

***Key Words:*** competing risks model, EM algorithm, joint modeling, longitudinal data, mixed effects model, multiple outcomes

We study a joint model for longitudinal measurements and competing risks survival data. Our model consists of a linear mixed effects submodel for the longitudinal outcome and a proportional cause-specific hazards frailty sub-model (Prentice et al, 1978) for the competing risks survival data, linked together by some latent random effects. We propose to obtain the maximum likelihood estimates of the parameters by an EM-based algorithm and estimate their standard errors using a profile likelihood method. Our joint model offers a flexible approach to handle non-ignorable missing data in the longitudinal measurements after event times. It is also an extension of previous joint models with a single failure type, providing a possible way to model informative censoring as a competing risk. Our method is evaluated using both simulated data and a clinical trial for the scleroderma lung disease.

### Joint Modeling of Survival and Longitudinal Data

***Key Words:*** joint modelling, survival analysis, longitudinal data, nonparametric, MCEM

In clinical studies, longitudinal covariates are often used to monitor the progression of the disease as well as survival time. Relationship between a failure time process and some longitudinal covariates is of key interest and so is understanding the pattern of longitudinal process to learn more about health status of patients, or to get some insight into the progression of disease. Joint modeling of the longitudinal and survival data has certain advantages and has emerged as an effective way for each to gain information from the other. Typically, a parametric longitudinal model is assumed to facilitate the likelihood approach. In the talk, I will propose several nonparametric longitudinal models in the joint modeling setting. Longitudinal process is represented by some basis functions and a proportional hazard model is then used to link them with the event-time.

### Type I and II Error of Joint Multimarkers Cox Models To Predict Chronic Disease Outcomes

❖ Philimon Gona, Boston University, Department of Mathematics and Statistics, 111 Cummington St, Boston, MA 02215, *philgona@math.bu.edu*

***Key Words:*** type I and II error, multiple biomarkers, marker-specific effects, Framingham Heart Study, Cox models, predicting chronic disease outcomes

Biomarkers are used to augment traditional risk factors for predicting chronic disease onset, progression or mortality. Different markers studied may be associated with different aspects of disease pathogenesis. When multiple markers are analyzed there is need to identify the more informative among several markers while limiting type I and II error. Type I error will be inflated if markers are analyzed one at a time whereas multi-marker models may lack the power to detect marker-specific effects. We conducted a simulation study of power and type I error of Cox models investigating 10 biomarkers for predicting CVD outcomes in Framingham Heart Study using multi-steps. Backward elimination was used to select most informative biomarkers. Multi-biomarker analysis may yield results that differ from those of single-marker analyses. Type I and II error should be considered to reconcile conflicts

### Robust Analysis of Biomarker Data with Informative Missingness Using a Two-Stage Hypothesis Test in an HIV Treatment Interruption Trial: AIEDRP AIN503 /ACTG A5217

❖ Karen Messer, University of California, San Diego, Moores UCSD Cancer Center, 3855 Health Sciences Dr 0901, La Jolla, CA 92093-0901, *kmesser@ucsd.edu*; Florin Vaida, University of California, San Diego; Christine Hogan, Columbia University

***Key Words:*** missing data, clinical trial, robust analysis

AIN503/A5217 studies whether a period of early treatment with antiretroviral therapy may lower the viral setpoint in subjects recently HIV-1 infected. However, even under the null hypothesis control arm subjects are more likely to be missing endpoint data because of disease progression, leading to bias. Also, two comparisons between arms are important: one at identical times post-randomization, and the other at the same amount of time off-therapy, as measured post-randomization. Both comparisons are important. The null hypothesis is exchangeability under a time-shift between arms, which captures clinically relevant information and allows us to formalize a two-stage hypothesis test using both comparisons. With this framework, we show that the two-stage test can be adjusted for the missing data using a simple worst-rank substitution.

### Evaluating Causal Effect Predictiveness of Candidate Surrogate Endpoints

❖ Peter Gilbert, Fred Hutchinson Cancer Research Center/University of Washington, 1100 Fairview Ave., N., PO Box 19024, Seattle, WA 98109, *pgilbert@scharp.org*; Michael G. Hudgens, The University of North Carolina at Chapel Hill

***Key Words:*** biomarker, causal inference, principal stratification, principal surrogate, randomized trial, selection bias

Most methods for evaluating biomarkers as potential surrogate endpoints measure validity in terms of net effects (i.e., treatment effects adjusted for the biomarker measured after randomization). Frangakis and Rubin (2002, Biometrics) criticized these approaches because net effects may reflect selection bias, and suggested an alternative definition of a surrogate endpoint (a principal surrogate) based on causal effects. This talk will introduce a “causal effect predictiveness surface” that can be used for evaluating and comparing the quality of biomarkers as surrogate endpoints. Assumptions and study designs that identify this surface, and parametric likelihood-based methods for estimating it, will be described. This work is motivated by the problem of assessing an immune response to an HIV vaccine as a surrogate endpoint for HIV infection, which will be used as an illustrative example.

### A Novel Approach for Proteomic Biomarker Identification Using Mass Spectrometry

❖ Chen Ji, Stony Brook University, 350 Circle Road, A311B Schomburg, Stony Brook, NY 11790, *ji_rock@yahoo.com*; Yeming Ma, Brookhaven National Laboratory; Xuena Wang, University of Hawaii; Esther Arkin, Stony Brook University; Joseph Mitchell, Stony Brook University; Wei Zhu, Stony Brook University

***Key Words:*** variance component analysis, biomarker detection, head and neck cancer

Biomarker detection is a critical step in disease diagnosis and knowledge discovery based on protein mass spectrometry (MS) data. We propose a novel variance component approach for biomarker detection, which is illustrated through a SELDI-TOF MS study on head and neck cancer. In comparison with our previous biomarker detection approaches such as stepwise discriminant analysis and the traditional peak detection strategy, we found that the new variance component approach can better distinguish cancer from non-cancer cases with a sensitivity of 86.36% and a specificity of 96.30%. Furthermore, the proposed approach can be generalized to handle the multiple-class (for example, disease stage 1, stage 2, stage 3 and normal controls) discrimination problem.

## 256 ASA Stat Bowl Session 1

The ASA, ENAR, IMS, SSC, WNAR

**Tuesday, August 8, 10:30 am–12:20 pm**

There are 16 players in the tournament. There will be four first round games, each game consisting of four players. The placement of the players in the games was random, with the restriction of no one being in the same game as a person from the same university.

### Game 1

❖ Upasana Santra, University of Florida, 103 Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611-8545, *usantra@stat.ufl.edu*; ❖ Samiran Ghosh, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269, *samiran@stat.uconn.edu*; ❖ Susanta Tewari, University of Georgia, 210 Rogers Road, Apt. Q214, Athens, GA 30605, *statsusant@yahoo.com*; ❖ Xinwei Deng, Georgia Institute of Technology, School of Industrial and Systems Engineering, 765 Ferst Drive, NW, Atlanta, GA 30332-0205, *xdeng@isye.gatech.edu; gtg721v@mail.gatech.edu*

### Game 2

❖ Vivekananda Roy, University of Florida, Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611-8545, *vroy@stat.ufl.edu*; ❖ Ying Hung, Georgia Institute of Technology, 350187 Georgia Tech Station, Atlanta, GA 30332, *yhung@isye.gatech.edu*; ❖ Dipankar Bandyopadhyay, University of Georgia, 204 Statistics Building, Dept. of Statistics, University of Georgia, Athens, GA 30602, *dban@stat.uga.edu*; ❖ Jeffrey Lidicker, Temple University, Center for Statistical and Information Science, 623 Jones Hall, 3307 North Broad Street, Philadelphia, PA 19140, *lidicker@temple.edu*

### Game 3

❖ Fang Yu, University of Connecticut, Department of Statistics, 215 Glenbrook Road Unit 4120, Storrs, CT 06269, *fangyu@stat.uconn.edu*; ❖ Alicia Graziosi, Temple University, 12 Yale Road, Marlton, NJ 08053, *alicia.graziosi@temple.edu*; ❖ Arunava Chakravartty, University of California, Riverside, 200 W. Big Springs Road, Apt 4, Riverside, CA 92507, *achak001@ucr.edu*; ❖ Rajarshi Dey, Kansas State University, 820 Osage Street, Manhattan, KS 66502, *rajarshi@ksu.edu*

### Game 4

❖ Ke Zhang, Kansas State University, 2212 Prairie Glen, Manhattan, KS 66502, *kezhang@ksu.edu*; ❖ Tanujit Dey, Case Western Reserve University, 323 Yost Hall Department of Statistics, 10900 Euclid Avenue, Cleveland, OH 44106-7054, *txd30@case.edu*; ❖ Christopher Rigdon, Southern Illinois University, 5 Stephens Court, Glen Carbon, IL 62034, *crigdon@siue.edu*; ❖ Satrajit Roychoudhury, New Jersey Institute of Technology, 20 Marshall Street, Apt 10 H, Irvington, NJ 07111, *sr62@njit.edu*

## 257 Hurricane Katrina and Economic Data ● ✪

Business and Economics Statistics Section

**Tuesday, August 8, 10:30 am–12:20 pm**

### Current Population Survey Response to the Hurricane Katrina Disaster

❖ Edwin L. Robison, Bureau of Labor Statistics, 1809 Belvedere Blvd., Silver Spring, MD 20902, *robison.ed@bls.gov*; Anne Polivka, Bureau of Labor Statistics; Diane Herz, Bureau of Labor Statistics; Lawrence Cahoon, U.S. Census Bureau; Richard Ning, U.S. Census Bureau; Maria Reed, U.S. Census Bureau; Greg Weyland, U.S. Census Bureau

***Key Words:*** sample survey, disaster, CPS

The Current Population Survey (CPS) publishes monthly labor force estimates for the national Civilian Noninstitutional Population and provides vital data series for state labor force models produced by the Local Area Unemployment Statistics (LAUS) program. Hurricanes Katrina and Rita caused widespread devastation and unprecedented population displacements. This disrupted CPS field operations, processing, and estimation for September 2005 and had a continuing effect. This paper covers the CPS program's response to the challenges presented by the Katrina disaster. Procedures were developed to track population shifts. Questions were added to the survey instrument to capture persons displaced by Hurricane Katrina and to the supplement on displaced workers. Some analysis of the effect of Hurricane Katrina on labor force estimates is included.

### Accounting for Katrina Effects in State Labor Force Estimates

❖ Richard Tiller, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 4985, Washington, DC 20212, *tiller_r@bls.gov*; Sharon Brown, Bureau of Labor Statistics

***Key Words:*** small area estimation, state-space models, outliers, seasonal adjustment

The Bureau of Labor Statistics uses state-space time series models to produce state labor force estimates from the monthly Current Population Survey (CPS). Covariates from state unemployment insurance data and payroll employment from the Current Employment Statistics program are used to help estimate trend movements in the CPS. The use of models produces estimators with much smaller variances than the survey estimates, but the models can be slow to adapt to external shocks to the economy. Hurricane Katrina presented a special challenge because it resulted in major under coverage of the population in affected states due to the displacement of persons in the CPS sample during the disaster. In short, Katrina was unobserved in the CPS sample. This paper describes how information from the covariates was used to estimate the effects of Katrina on state labor force values.

### Effects of Hurricane Katrina on the BLS Payroll Survey

❖ Patricia Getz, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 4840, Washington, DC 20212, *getz.pat@bls.gov*; Richard Rosen, Bureau of Labor Statistics; Larry Huff, Bureau of Labor Statistics; Angela Clinton, Bureau of Labor Statistics

***Key Words:*** payroll employment, non-response adjustment, Hurricane Katrina

The Current Employment Statistics (CES) survey produces nonfarm payroll employment estimates at national-, state-, and metropolitan-area levels each month on a timely basis. The CES estimates, based on a sample of nearly 400,000 worksites, are among the most closely watched principal economic indicators. In the aftermath of Hurricane Katrina, users were keenly interested in the employment effects from the storm; at the same time, the unprecedented devastation in the Gulf Coast region presented severe challenges to conducting the survey. This paper will detail efforts by BLS to maximize survey response, adjust estimation procedures for nonresponse, and adjust business birth/death models to produce viable estimates in the months immediately following the hurricane. The paper also describes efforts to analyze the effects of Katrina on overall employment trends.

### Analysis and Procedures in the Census Bureau's Trade Surveys after Hurricane Katrina

❖ Jock R. Black, U.S. Census Bureau, 4700 Silver Hill Road, Room 2754-3, Washington, DC 20233-0006, *Jock.Reynolds.Black@census.gov*; Ruth E. Detlefsen, U.S. Census Bureau

***Key Words:*** sales, inventories, Hurricane Katrina, validity

The U.S. Census Bureau surveys approximately 16,000 firms across the United States each month to produce national estimates of sales and inventories for retail and wholesale industries. After the widespread destruction caused by Hurricane Katrina, the bureau was justifiably concerned about the validity of its yet-to-be-published estimates. This paper will present a scenario for analyzing survey data and gauging the sample's ability to measure totals accurately. It also will discuss modifications to data collection and post--data collection processes, such as response analysis, imputation, estimation, and seasonal adjustment.

## 258 Estimation of Treatment Effect for Clinical Trials with Group Sequential Designs ● ✪

Biopharmaceutical Section, Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Estimation Following Group Sequential Trials: a Bayesian View

❖ Gary L. Rosner, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *glrosner@mdanderson.org*

Most randomized clinical trials undergo some form of interim monitoring. It is well-known that sequential stopping rules affect the frequentist characteristics of estimators. Bayesian inference obeys the likelihood principle, so inference does not change if the stopping rules do not alter the likelihood. Often, trial organizers are interested in the frequentist characteristics of Bayesian study designs. In this talk, we will discuss characteristics of Bayesian estimates of treatment effects and covariate effects in a clinical trial that includes interim monitoring.

### Conditional Estimation of Treatment Effects for Clinical Trials with Group Sequential Designs

❖ Michael W. Lee, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., Raritan, NJ 08869, *mlee60@prdus.jnj.com*; Hui Quan, sanofi-aventis

***Key Words:*** interim analysis, early stopping, multiple endpoints, estimate bias

While preserving type I error rate, a group sequential design allows a trial to be stopped earlier if the treatment effect can be demonstrated through interim analyses based on the prespecified stopping rule. Thus, group sequential designs have been used widely, especially for long-term trials. However, researchers have demonstrated the biasness of the unadjusted estimate of the treatment effect. They proposed adjusted estimate to address the issue. In this presentation, we will propose a conditional approach for obtaining the adjusted estimate of treatment effect when the stopping rule relies on the outcomes of multiple primary endpoints. The proposed approach provides estimate with reduced bias and confidence interval with approximately the nominal coverage rate across stopping time. An example will be used to illustrate the application of our approach.

## 259 Advances in Phylogenetic Inference ● ✪

IMS, WNAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Analysis of Comparative Data with Hierarchical Autocorrelation

❖ Cecile Ane, University of Wisconsin-Madison, Department of Statistics, MSC, 1300 University Ave., Madison, WI 53706, *ane@stat.wisc.edu*

***Key Words:*** comparative methods, linear models, phylogeny

In comparative biology, observational units are usually species and typically do not yield i.i.d. observations. Rather, closely related species tend to be more alike than distantly related species. I will discuss methods applicable to these data and show some asymptotic properties of generalized least squares estimates in this setting.

### Using Bootstrap Support for Splits To Construct Confidence Regions for Trees

❖ Edward Susko, Dahousie University, Chase Building, Department of Mathematics and Statistics, Halifax, NS B3H 3J5 Canada, *susko@mathstat.dal.ca*

***Key Words:*** phylogenetics, statistical inference, bootstrap

Many estimated topologies in phylogenetic studies are presented with the bootstrap support for each split in the topology indicated. If phylogenetic estimation is unbiased, high bootstrap support for a split suggests there is a good deal of certainty that the split actually is present in the tree and low bootstrap support suggests one or more of the taxa on one side of the estimated split might be located with taxa on the other side. In the latter case, the follow-up questions about how many and which taxa reasonably could be placed incorrectly and where they might alternatively be placed are not addressed through the presented bootstrap support. We will present an algorithm that finds the set of all trees with minimum bootstrap support for their splits greater than some given value. The output is a ranked list of trees.

### The Threshold Model: an Underutilized Resource in Phylogenetic Inference

❖ Joseph Felsenstein, University of Washington, Department of Genome Sciences, Box 357730, Seattle, WA 98195-7730, *joe@gs.washington.edu*

***Key Words:*** phylogeny, quantitative genetics, MCMC, evolution

Morphological evolutionists make phylogenies for characters that have discrete states, such as 0/1 characters. They usually do so using a non-statistical "parsimony method" that minimizes the number of changes of state of the character along the tree. There has been some attempt to make statistical methods that ask whether two of these characters change along a tree in a correlated way. Papers by Pagel and Lewis have used a simple two-state Markov process to ask about correlated change and to infer phylogenies. This talk will emphasize a more realistic alternative, the threshold model of quantitative genetics, due to Sewall Wright (1934). An underlying continuous character, the "liability," wanders on a continuous scale, a developmental threshold converts this to the observed 0/1 character. Covariances of the underlying characters can be inferred using a MCMC method given a tree.

### Algebraic Statistics and Phylogenetic Inference: Establishing Identifiability of Mutation Models

❖ Elizabeth S. Allman, University of Alaska, Fairbanks, Department of Mathematics and Statistics, PO Box 756660, Fairbanks, AK 99775, *e.allman@uaf.edu*; John A. Rhodes, University of Alaska, Fairbanks

***Key Words:*** identifiability, molecular evolution models, algebraic statistics

For a model of molecular evolution to be useful for phylogenetic inference, model parameters must be identifiable. Ideally, from a joint distribution the model predicts, it should be possible to recover the tree topology, the base frequency distributions at nodes of a tree, and the rates at which various point mutations occur. We establish tree identifiability for a number of phylogenetic models, including a covarion model and a variety of mixture models with a small number of classes. We then consider methods of estimating parameters from a joint distribution arising from certain mixture models. These results stem from an algebraic investigation of probabilistic models of evolution. This viewpoint enables analysis of more complex models of mutation than previously possible. We describe the framework of algebraic statistics for understanding these and other recent results.

# 260 Personal Networks: Applications Using Data on Social Relationships ●

Section on Statistical Computing, Section on Statisticians in Defense and National Security, Section on Statistical Graphics

**Tuesday, August 8, 10:30 am–12:20 pm**

## Design and Analysis of "How Many Xs Do You Know?" Surveys

❖ Tian Zheng, Columbia University, Room 1005, MC 4690, 1255 Amsterdam Avenue, New York, NY 10027, *tzheng@stat.columbia.edu*; Andrew Gelman, Columbia University

We consider issues in the design and analysis issues of "How many X's do you know?" surveys as a means of studying social networks. We first discuss the analysis of such data using multilevel regression so that the properties of a network can be studied in relation to the characteristics of the individuals in the network and subgroups of interest in the population. In designing new surveys of this type, we must consider network effects and response effects, including the depth of social connections, the effects of imperfect recall, the choice of groups of known size to use for normalization, and methods for learning about small and moderately-sized subpopulations.

## Data Mining in Networks

❖ David Jensen, University of Massachusetts, 140 Governors Drive, Amherst, MA 01003, *jensen@cs.umass.edu*

Data mining is the process of constructing predictive models from large and complex databases. Counterterrorism data present unique challenges for data mining algorithms, including the relational and heterogeneous structure of the data, the fragmentary nature of the data, and the presence of relational autocorrelation. New research is addressing challenges, but applications are still three to five years away. Many design options exist with widely differing uses and impacts. Current popular understanding of data mining for counterterrorism is suffering from a few widely disseminated myths, including the necessity of a single large database and a vast new program of data collection.

## The Dynamics of Viral Marketing

Jure Leskovec, Carnegie Mellon University; ❖ Lada A. Adamic, University of Michigan, 304 West Hall, 1085 South University Ave., Ann Arbor, MI 48109, *ladamic@umich.edu*; Bernardo Huberman, Hewlett Packard Labs

***Key Words:*** viral marketing, information diffusion, networks, information cascades, recommender systems

We present an analysis of a person-to-person recommendation network, consisting of 4 million people who made 16 million recommendations on half a million products. We observe the propagation of recommendations and the cascade sizes, which we explain by a simple stochastic model. We then establish how the recommendation network grows over time and how effective it is from the viewpoint of the sender and receiver of the recommendations. While on average recommendations are not very effective at inducing purchases and do not spread very far, we present a model that successfully identifies product and pricing categories for which viral marketing seems to be very effective.

## Modeling Massive Dynamic Graphs

❖ Chris Volinsky, AT&T Labs-Research, Shannon Laboratory, Room D235, 180 Park Ave, Florham Park, NJ 07932, *volinsky@research.att.com*

***Key Words:*** graphs, fraud, networks, data mining

When studying large transactional networks---such as telephone call detail data, credit card transactions, or web clickstream data---graphs are a convenient and informative way to represent data. When the graph edges have a time stamp, we have a "dynamic graph," where the edges are born and die through time. I will present a framework for representing and analyzing dynamic graphs with a focus on the massive graphs found in telecommunications and Internet data. I also will demonstrate the application of this model to a telecommunications fraud problem, where we are looking for patterns in the graph associated with fraud.

# 261 Modeling Behavioral Data from Clinical Research on Smoking ●

Section on Health Policy Statistics, Biometrics Section

**Tuesday, August 8, 10:30 am–12:20 pm**

## Modeling Heaping in Self-Reported Cigarette Counts

❖ Daniel F. Heitjan, University of Pennsylvania, 622 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, *dheitjan@cceb.upenn.edu*

***Key Words:*** Bayesian analysis, heaped data, rounding, smoking cessation, addiction, clinical trial

Self-reported daily cigarette count data typically show large "heaps" at multiples of 20 cigarettes, the number of cigarettes in a pack in the United States. Evidence suggests smokers actually are consuming a variable number of cigarettes, which some then report rounded off to the nearest number of packs. Moreover, the propensity to report data in this heaped fashion appears to increase with the number of cigarettes smoked. Such heaping in the data generation process can cause profound biases if the data are taken at face value. We describe models for estimating the underlying cigarette count distributions from heaped data, which we fit using Bayesian methods to data from a smoking-cessation drug trial.

## A Cure Mixture Model for Multivariate Time-to-Event Data

❖ E. Paul Wileyto, University of Pennsylvania, School of Medicine, Psychiatry, Tobacco Use Research Center, 3535 W. Market St., Suite 4100, Philadelphia, PA 19104-3309, *epw@mail.med.upenn.edu*

***Key Words:*** cure-mixture models, recurrent events, multivariate survival, behavioral medicine, smoking cessation

The attempt to quit is, for most smokers, a series of alternating states, terminated by transitions that are either a return to smoking or recovery of abstinence. Ordinary recurrent failure time modeling leads to poor fit and biased estimates of treatment effects because long-term survivors may arise at each small step. We introduce a cure-mixture regression model for multivariate or recurrent failure time data based on Farewell's 1982 model, which used logistic regression to predict probability of membership in cured and noncured classes and Weibull regression to predict time to event in the noncured class. Parameter estimates are obtained using maximum likelihood, with standard errors adjusted to account for repeated measures using the cluster-correlated robust variance estimate. The model provides an excellent fit to data from a smoking cessation clinical trial.

### Modeling Variation in the Effects of Smoking Using Intensive Longitudinal Data

❖ Donald Hedeker, University of Illinois at Chicago, School of Public Health (MC 923), 1603 W. Taylor St., room 955, Chicago, IL 60612-4336, *hedeker@uic.edu*

***Key Words:*** mixed models, heterogeneous variance

Smokers begin their 'careers' with sporadic or irregular smoking, which progresses to a regular habit after long exposure. Here, we apply intensive longitudinal methods to an adolescent study in which subjects carry a palm pilot for a week and get random and event-triggered (i.e., when they smoke a cigarette) prompts for data. Using this Ecological Momentary Assessment (EMA) data, we describe how mixed models---including heterogeneous between- and within-subjects variance---can be used to address key questions in smoking research. For example, this kind of intensive measurement allows us to model differences in both types of variances, which permits us to investigate the degree to which more-experienced smokers exhibit less variability on several outcomes relative to less-experienced smokers.

### Exploring the Relationship between Adverse Events, Compliance, Dropout, and Abstinence in Pharmacotherapy Smoking Cessation Trials

❖ Joel A. Dubin, University of Waterloo, Waterloo, ON Canada, *jdubin@uwaterloo.ca*

***Key Words:*** abstinence success, compliance, dose-ranging, graphical methods, safety data, survival models

In trials of pharmacological aids to smoking cessation, the participants who comply (both not dropping out and adhering to their prescribed medication regimen) tend to exhibit higher rates of abstinence. Here, we apply specialized survival graphical methods to reveal this trend and subsequently implement complex survival modeling to identify predictors of dropout and overall noncompliance, including consideration of medication dosage strength, the experiencing and severity of adverse events, and abstinence success. We apply the methods to a dose-ranging clinical trial of naltrexone, combined with nicotine patch, where the study population was comprised of 400 current smokers who desired abstinence.

## 262 Complex Data Structures ●✪

SSC

**Tuesday, August 8, 10:30 am–12:20 pm**

### Issues Relating to Methods for Analysis of Survey Data

❖ David Binder, Statistical Society of Canada, 49 Bertona Street, Nepean, ON K2G 4G7 Canada, *dbinder49@hotmail.com*; Georgia Roberts, Statistics Canada

***Key Words:*** complex survey data, software for analysis, survey weights, variance estimation

In recent years, an increasing number of researchers have been able to access survey microdata files. These researchers perform various analyses to summarize the data and to describe relationships in a target population. Many of these researchers use analytic software without having a good understanding of the statistical underpinnings behind the methods being applied. Some of the issues facing the researchers include understanding the survey weights, understanding informative sampling, handling nonresponse, using variances that are model-dependent, incorporating survey design information into the modeling process, understanding the impact of model misspecification, inferring causality, and pooling/combining of survey files. In this paper we review many of these issues, focusing on pooling/combining of survey files.

### Spatial-Temporal Modeling for Marine Ecological Systems

❖ Joanna M. Flemming, Dahousie University, Department of Mathematics and Statistics, Room 122, Chase Building, Halifax, NS B3H 3J5 Canada, *flemming@mathstat.dal.ca*

***Key Words:*** Argos satellite telemetry, state-space models, behavior

Remotely sensed tracking data collected on animal movement is vastly underutilized due to a lack of statistical tools for appropriate analysis. Features of such data that make analysis particularly challenging include the presence of estimation errors that are non-Gaussian and vary in time, observations that occur irregularly in time, and complexity in the underlying behavioral processes. We develop a state-space framework that simultaneously deals with these features and demonstrate our method by analyzing seal pathways. We show how known information regarding error distributions can be used to improve inference of the underlying process(es) and demonstrate that our framework provides a powerful and flexible method for fitting different behavioral models to tracking data.

### A Stochastic Model for Forest Fire Growth

❖ Willard J. Braun, University of Western Ontario, Department of Statistics and Actuarial Science, London, ON N6A 5B7 Canada, *braun@stats.uwo.ca*

***Key Words:*** wildfires, spread, stochastic model

Forest fires often spread in unpredictable ways. Deterministic models of fire spread have been developed in both Canada and the United States. These models capture 'expected' behavior of fire spread fairly well, but are unable to give information about variability. In particular, spot fires are difficult to model using these deterministic approaches. We propose an interacting particle system as a way of incorporating stochasticity into the modeling of this highly variable phenomena. The model lives on a regular two-dimensional lattice. Each lattice site may be occupied by unburnt fuel, burning fuel, or burnt fuel; alternatively, a site may contain water or some other nonflammable material. State transitions are accomplished according to a continuous-time Markov chain. This talk will give details of this model and demonstrate how spot fires can be incorporated.

## 263 Effective Leadership for Statistical Thinking and Impact ● ✪

Council of Chapters, Section on Statistical Education, Section on Statistical Consulting

**Tuesday, August 8, 10:30 am–12:20 pm**

### Building Statistical Thinking into the School Curriculum: Obstacles and Opportunities

❖ Richard L. Scheaffer, University of Florida, 907 NW 21 Terrace, Gainesville, FL 32603, *rls907@bellsouth.net*

***Key Words:*** school mathematics, undergraduate statistics, mathematical thinking, statistical thinking

Statistics (as data analysis) is a key component of modern school (K--12) mathematics and one of the most highly enrolled mathematical science courses in most colleges. Thus, statistics is taught to many, but statistical thinking is still not a trait associated with most high-school or college graduates, even for those with strong mathematical thinking skills. Among the reasons for this are that much statistics is taught by mathematical thinkers and most courses are predominantly about techniques, rather than concepts. Examples of how statistics tends to be taught and assessed in schools and colleges, compared with examples of how it could be taught to enhance statistical thinking, will help shed light on the serious educational issue at hand.

### Seven Habits of Highly Effective Statisticians in Pharmaceutical Industry

❖ Frank Shen, Bristol-Myers Squibb Company, 509 Jenny Drive, Yardley, PA 19067, *frank.shen@bms.com*

***Key Words:*** effective leadership, high impact, pharmaceutical industry, collaboration

One of the best-selling books this decade, authored by Stephen Covey, highlighted the seven habits of highly effective people. While these seven habits build effectiveness in working with people, playing an effective and influential leadership role in pharmaceutical industry needs more than that. This is an industry that lives by scientific data and multidisciplinary talents. No individual owns the solution, and the statistician is not the only scientist who knows how to analyze and turn data into knowledge. Collaboration is the key to creating impact. An effective leader who knows how to stimulate such collaboration must have E cubed: emotion, energy, and edge. Together with the seven habits, a visionary and passionate leader inspires others to find their voices and clearly see their worth and potential to achieve a collaboration with much higher energy and impact.

### Effective Statistical Leadership in Government and Industry

❖ Fritz J. Scheuren, National Opinion Research Center, 1402 Ruffner Road, Alexandria, VA 22302, *scheuren@aol.com*

***Key Words:*** traits of leadership, application settings in human rights, quality management

There are many forms of leadership, including the intellectual leadership of ideas and gentle nurturing and mentoring of new professionals. Both of these are leadership skills good teachers and managers possess. There are other leadership traits: the example of staying current in your profession, of getting out and trying new techniques, of being there for a friend. Among those and other traits, this paper will cover the kind of leadership we sometimes call heroic---the leadership of speaking out on issues when courage is required. The focus, naturally, will be on what it means to put the two adjectives "effective" and "statistical" in front of the noun "leadership." My remarks will come from my experience in government and industry.

## 264 Bombs to Bullets, People to Planets: Exploring Imaging Applications with a Statistician's Eye ●

Section on Physical and Engineering Sciences,
Section on Statisticians in Defense and National Security,
Section on Statistical Graphics
**Tuesday, August 8, 10:30 am–12:20 pm**

### Do Guns Transfer Their DNA to Bullets?

❖ William F. Eddy, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, *bill@cmu.edu*

***Key Words:*** imaging, ballistics

If guns transfer markings to bullets and cartridges, it might be possible to build a database for guns that is similar to the Integrated Automated Fingerprint Identification System maintained by the FBI. Assuming guns do transfer markings, are there enough gun-to-gun differences to make it possible to distinguish individual guns? On the other hand, is there sufficiently little shot-to-shot variation within a single gun to be certain of the identity of that gun from the markings? These are but a few of the questions raised in a recent project conducted by the National Academies, titled "Assessing the Feasibility, Accuracy, and Technical Capability of a National Ballistics Database." This talk will describe the problem in more detail, discuss current systems used for firearm identification, and review the conclusions of the study.

### Using Image Grand Tour To Explore Multiangle, Multispectral Satellite Imagery

❖ Amy Braverman, Jet Propulsion Laboratory, Mail Stop 126-347, 4800 Oak Grove Drive, Pasadena, CA 91106, *Amy.Braverman@jpl.nasa.gov*; Edward Wegman, George Mason University; Wendy Martinez, Office of Naval Research; Juergen Symanzik, Utah State University; Brad Wallet, Automated Decisions

***Key Words:*** remote sensing, visualization, exploratory data analysis, grand tour

Remote sensing data are spatial arrays of p-dimensional vectors where each component corresponds to one of p variables. Applying the same R^p to R^d projection to all pixels creates d new images, which may be easier to analyze than the original because d < p. Image grand tour (IGT) steps through the space of projections, and d=3 outputs a sequence of RGB images, one for each step. In this talk, we apply IGT to multiangle, multispectral data from NASA's MISR instrument. MISR views each pixel in four spectral bands at nine view angles. Multiple views detect photon scattering in different directions and are indicative of physical properties of the scene. IGT allows us to explore MISR's data structure while maintaining spatial context; a key requirement for physical interpretation. We report results highlighting the uniqueness of multiangle data and how IGT can exploit it.

### Tomographic Imaging Using Background Cosmic Radiation

❖ Nicolas Hengartner, Los Alamos National Laboratory, Statistical Sciences Group, Mail Stop F600, Los Alamos, NM 87545, *nickh@lanl.gov*

***Key Words:*** tomographic imaging, inverse problems, design of instrument

Existing radiographic methods are inefficient for detecting shielded nuclear materials and potentially present radiation hazards to inspectors and vehicle passengers. Recent advances at the Los Alamos National Laboratory suggest a promising alternative: Use the natural scattering of muons---produced by the decay of cosmic rays showering Earth---as a passive radiographic probe to image passively dense objects. Muons interact with matter primarily through the Coulomb force. We model the net effect of that interaction as a random change in the path of the muon, with the magnitude of that change depending on the atomic number of the material traversed. In this talk, I will discuss the tomographic reconstruction of the scattering density of an object from measured changes in the paths of individual muons traversing that object.

### Using Statistics To Search and Annotate Pictures

❖ Nuno Vasconcelos, University of California, San Diego, Department of Electrical and Computer Engineering, 9500 Gilman Drive, Mail code 5603, La Jolla, CA 92093-0407, *nuno@ece.ucsd.edu*; Pedro J. Moreno, Google, Inc.

***Key Words:*** image search, image annotation, Gaussian mixtures, image tags

How do search engines look for images? How do you search if your query is an image? How do you annotate your picture collection automatically? Given the large number of existing image databases available, the ability to search images with particular attributes---such as containing people, vacation spots, or special events---becomes extremely useful. Clearly, simple schemes, such as manually labeling the images with textual tags, are massive and not practical endeavors. On the other hand, there are many annotated databases freely available. Is there a way to leverage statistical techniques to take advantage of both the images and their corresponding annotations? In this talk, we will describe what methods are used commonly for these tasks, what the metrics used to measure performance are, and what the current state of the art is.

## 265 Bayesian Spatio-Temporal Models ● ✪

WNAR, Biometrics Section, Section on Bayesian Statistical Science, Section on Statistics and the Environment

**Tuesday, August 8, 10:30 am–12:20 pm**

### Spatial Dynamic Factor Models

❖ Hedibert F. Lopes, The University of Chicago, Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, *hlopes@gsb.uchicago.edu*; Esther S. Salazar, Instituto de Matem·tica-UFRJ; Dani Gamerman, Instituto de Matem·tica-UFRJ

***Key Words:*** Gaussian processes, hyperparameters, spatial interpolation, factor loadings, Bayesian, RJMCMC

We propose a new class of spatial factor analysis. A number of factors are defined to drive the variability of the observations, as in standard factor models. Their loadings, however, are restricted to satisfy stochastic constraints based on their spatial structure. A few spatial structures can be considered, but we concentrate on those based on Gaussian processes---commonly used in geostatistics. Spatial dependence appears in the columns of the factor-loading matrix. A novel, reversible jump, Markov chain Monte Carlo algorithm is proposed to select the number of common factors. The new class of models is tested against synthetic and real data examples.

### Characterizing Invasions with Hierarchical, Rule-Based Systems

❖ Mevin B. Hooten, University of Missouri-Columbia, 146 Middlebush Hall, Department of Statistics, Columbia, MO 65211, *hooten@stat.missouri.edu*; Christopher K. Wikle, University of Missouri-Columbia

***Key Words:*** invasive species, hierarchical Bayesian models, dynamical systems

It is often the case in ecological monitoring projects that only binary data collection is feasible on large spatial and temporal domains. In such settings, the propagation of underlying ecological phenomena may still behave in a well-defined theoretical fashion, though limited information prohibits the implementation of conventional spatio-temporal process models. When there is simply not enough information to inform such scientifically based parameterizations, we propose the use of a scientifically naive, rule-based dynamic process model within a hierarchical framework. Utilizing only simple model specifications can still accommodate multiple sources of uncertainty yet exhibit complicated, large-scale, dynamical behavior for characterizing the spread of invasive species with only minimal data.

### Spatially Varying AR Processes Based on Discrete Convolutions

❖ Bruno Sanso, University of California, Santa Cruz, Applied Mathematics and Statistics, 1156 High Street, Mail Stop: SOE2, Santa Cruz, CA 95064, *bruno@ams.ucsc.edu*; Alexandra Schmidt, Universidade Federal de Rio de Janeiro; Aline Nobre, Universidade Federal de Rio de Janeiro

***Key Words:*** Bayesian methods, spatio-temporal models, process convolutions

We consider a class of models for spatio-temporal processes based on convolving independent processes with a kernel represented by a lower triangular matrix. We first consider a family obtained by convolving spatial Gaussian processes with isotropic correlations where the kernel is obtained from the covariance of an AR(p). A second family is based on considering convolutions of AR(p) processes and using the kernel to provide spatial interactions. The resulting random field corresponds to an AR(p) with spatially varying coefficients. Suitable priors on the parameters of the AR(p) guarantee the parameters satisfy the conditions for stationarity at each location. The proposed modeling framework provides a rich variety of covariance structures. We consider applications to the problem of detecting trends in environmental variables.

### Using Computer Models To Inform about Space-Time Fields

❖ Dave Higdon, Los Alamos National Laboratory, MS F600, PO Box 1663, Los Alamos, NM 87545, *dhigdon@lanl.gov*

***Key Words:*** computer model, Gaussian process, calibration

Almost all observable space-time fields result from some physical process that evolves over time. Examples range from the formation of ozone to the formation of the universe. Many of these physical systems can be simulated using computer models. However, utilizing such models in statistical inference is often problematic---models are usually quite computationally intensive and typically require large numbers of input settings that are not known a priori. Despite these challenges, the allure of detailed physical simulation models persists. Incorporating such models can lead to greater prediction accuracy and gives a mechanism for extrapolating well beyond observed experience. This talk focuses on approaches for utilizing detailed physics simulation models within the framework of space-time modeling. Applications in cosmology and weather will be considered.

## 266 Challenges Facing the Next Generation of Applied Statisticians

● ✪

Social Statistics Section, Committee on Applied Statisticians, Committee on Women in Statistics, Committee on Law and Justice Statistics, Section on Statistical Education, Section on Statistical Consulting

**Tuesday, August 8, 10:30 am–12:20 pm**

### Motivations and Challenges Facing Women Statisticians

❖ Kelly H. Zou, Harvard Medical School, 180 Longwood Ave., HCP, Department of Health Care Policy, Boston, MA 02115, *zou@bwh.harvard.edu*

***Key Words:*** gender parity, literature review, grant funding, descriptive statistics, multivariate regression analysis

The current preeminence of U.S. science and technology will face new challenges, such as the globalization of the workplace. This is an appropriate time to look at the prospects of our discipline, encourage more of the brightest students to study quantitative methods, and provide guidance for individuals now preparing for careers in statistics and related fields. This means looking inward in order to motivate and inspire the next generation of statisticians. In this research, a summary of the demographics in the statistical profession is presented based on a critical review of the available literature on membership surveys. In addition, a comparison of gender parity in research funding within a number of academic institutions is performed. Descriptive statistics, pooled and stratified analyses, and multivariate regression methods are conducted.

### Statistical Tools Employed in Legal Settings

❖ Mary W. Gray, American University, Department of Mathematics and Statistics, 4400 Massachusetts Avenue NW, Washington, DC 20016-8050, *mgray@american.edu*

***Key Words:*** forensic, statistics

Statistical analysis has been used for a long time to advance litigants' concepts of equity and justice. Topics such as jury selection, college athletics, identification evidence, employment practices, and massive tort liability have been subject to scrutiny from opposing statistical experts. The level of sophistication of the analyses presented has risen as the complexity of issues to be resolved has increased. This paper examines some of the emerging methodology being employed and the problem of educating nonstatisticians to deal with statistical evidence.

### Professional Accreditation for Statisticians

❖ Mary Batcher, Ernst & Young LLP, 1225 Connecticut Ave., NW, Washington, DC 20036, *mary.batcher@ey.com*

***Key Words:*** accreditation, focus group

In 2005, Fritz Scheuren appointed a task force on "Recognition of the Professionalism of Those in Business Practice" to take another look at accreditation of statisticians, now that accreditation programs have been implemented successfully by three other statistical societies. The task force reviewed the programs offered by the other statistical societies. They conducted focus groups of ASA members in three U.S. cities to assess the level of interest and solicit input on what such a program should include. In addition, a question on accreditation was included in the series of surveys conducted by the task force on self-awareness. The results of the focus groups and survey will be presented, along with a discussion of the pros and cons of accreditation, particularly as it might affect the next generation of statisticians.

## 267 Practical Issues in Conducting Statistics Education Research ●

Section on Statistical Education

**Tuesday, August 8, 10:30 am–12:20 pm**

### Practical Issues in Conducting Statistics Education Research

❖ Robert DelMas, University of Minnesota, 354 Appleby Hall, 128 Pleasant Street SE, Minneapolis, MN 55455, *delma001@umn.edu*; ❖ Beth Chance, California Polytechnic State University, San Luis Obispo, CA , *bchance@calpoly.edu*; ❖ Sterling Hilton, Brigham Young University, UT , *hiltons@byu.edu*; ❖ Lawrence M. Lesser, The University of Texas at El Paso, Department of Mathematical Sciences, 500 W. University Avenue, El Paso, TX 79968-0514, *Lesser@utep.edu*; ❖ Andrew Zieffler, University of Minnesota, Department of Educational Psychology, 206 Burton Hall, 178 Pillsbury Drive SE, Minneapolis, MN 55454, *Zief0002@umn.edu*

***Key Words:*** statistics education, research

The focus of this session is on the logistics of conducting statistics education research at the college level. Topics include ethical and practical issues, when and how to get IRB approval, using student assessment data as the basis of a research project, and making the transition from other disciplines to statistics education research.

## 268 Nonparametric Statistical Process Control Methods ●

Section on Quality and Productivity, Section on Nonparametric Statistics

**Tuesday, August 8, 10:30 am–12:20 pm**

### Minimum Control Charts

❖ Willem Albers, University Twente, Applied Math Department, Faculty EEMCS PO Box 217, Enschede, 7500AE The Netherlands, *w.albers@utwente.nl*

***Key Words:*** SPC, phase II control limits, order statistics

Shewhart control charts are sensitive to estimation effects and deviations from normality. Improvements have been derived, which are often adequate, but not always. In this talk, the remaining complication is attacked: What to do if a nonparametric approach is indicated, but too few Phase I observations are available? It is shown that grouping the observations during Phase II works well. Surprisingly, instead of using the group averages, it is preferable to compare the minimum for each group to a suitably chosen upper limit (in the two-sided case, also the maximum to an analogous lower limit). This 'minimum control chart' is demonstrated to be attractive; it is easy to explain and implement. Moreover, while it is truly nonparametric, its power of detection is comparable to that of the customary, normality assuming charts based on averages.

### Distribution-Free Quality Control Charts Based on Signed-Rank-Like Statistics

❖ Saad Bakir, Alabama State University, 2405 Reston Place, Montgomery, AL 36117, *bakir00@yahoo.com*

***Key Words:*** average run length, control sequence, nonparametric, robustness, Shewhart

In this article, we propose three distribution-free (or nonparametric) statistical quality control charts for monitoring a process center when an in-control target center is not specified. These charts are of the Shewhart-type, the exponentially moving average-type, and the cumulative sum-type. The constructions of the proposed charts require the availability of an initial reference sample taken when the process was operating in-control to calculate an estimator for the unknown in-control target process center. This estimated center is then used in the calculation of signed-rank-like statistics based on grouped observations taken from the process output periodically. As long as the in-control process underlying distribution is continuous and symmetric, the proposed charts have a constant in-control average run length irrespective of the process underlying distribution.

### Dispersion Control Charts Based on Ranks

❖ Raid Amin, University of West Florida, 11000 University Parkway, Dept of Math  Stat, Pensacola, FL 32514, *ramin@uwf.edu*

***Key Words:*** nonparametric control charts, quality control, dispersion-free, average run length

Nonparametric control charts are presented for the problem of detecting changes in the process variability when samples are taken at regular time intervals. It is standard practice to design control charts on the assumption that the distribution of the sample mean is (approximately) normal. When the distribution of the observations is not normal, there are potential problems with control charts designed based on normal theory assumptions. Normal theory calculations of properties of the chart are no longer reliable. The control limits will be inappropriate and the false alarm rate for the chart will be incorrect (usually too high). This problem is especially severe in control charts for variability. There may not be enough data to obtain a good estimate of the variance when the process is starting up, and the standard control limits will be inappropriate.

### Distribution-Free Multivariate Process Control Based on Log-Linear Modeling

❖ Peihua Qiu, University of Minnesota, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455, *qiu@stat.umn.edu*

***Key Words:*** discrete measurements, log-linear modeling, multivariate distribution, non-Gaussian data, nonparametric procedures, transformations

This paper considers statistical process control (SPC) when the process measurement is multivariate. Most existing multivariate SPC procedures assume the in-control distribution of the process measurement is known and it is a Gaussian distribution, which may not hold in applications. We demonstrate that results from conventional multivariate SPC procedures often are unreliable when the data are non-Gaussian. We suggest a methodology for estimating the in-control measurement distribution when a set of in-control data is available, which is based on log-linear modeling and takes into account the association structure of the measurement components. Based on the estimated in-control distribution, a CUSUM procedure for Phase II SPC also is suggested. This procedure does not depend on the Gaussian distribution assumption and thus is appropriate for most multivariate SPC problems.

### A Bayesian Approach in Modeling Shifts of the Mean/Variance of Count Data

❖ Panagiotis Tsiamyrtzis, Athens University of Economics and Business, 76 Patission Street, Dept of Statistics, Athens, 10434 Greece, *pt@aueb.gr*; Douglas M. Hawkins, University of Minnesota

***Key Words:*** Bayesian SPC by attributes, change point, gamma mixture

We consider a process producing count data from a Poisson distribution. Our interest is in detecting online whether the Poisson parameter (mean and variance) shifts to either a higher value (causing worst process performance) or a smaller value (good scenario). The necessity for drawing inference sequentially as the observations become available leads us to adopt a Bayesian sequentially updated scheme of mixture of gamma distributions. Issues regarding inference and prediction will be covered. The developed methodology is appealing in cases of short runs and/or Phase I count data.

## 269 Bayesian Computational Methods for Biomedical Applications

● ✪

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Bayesian Hidden Markov Modeling of Array CGH Data

❖ Subharup Guha, Harvard School of Public Health, 199 Park Drive, Apt 125, Boston, MA 02215, *sguha@hsph.harvard.edu*

***Key Words:*** genomic alterations, cancer, DNA, MCMC

Genomic alterations have been linked to the development of cancer. The technique of Comparative Genomic Hybridization (CGH) yields data that provide information about changes in DNA copy number. As increasing amounts of array CGH data become available, there is a growing need for automated algorithms that detect copy number gains and losses based on statistical considerations. We take a Bayesian approach, relying on the hidden Markov model to account for the dependence in the data. Localized amplifications (associated with oncogenes) and deletions (associated with tumor suppressors) are identified using posterior probabilities based on an MCMC sample. Publicly available data on pancreatic adenocarcinoma are analyzed. Comparisons are made with widely used methods to illustrate the reliability and success of the technique.

### A Hierarchical Bayesian Analysis of Longitudinal Frequency Data Using Piecewise Linear Regression

❖ John Kern, Duquesne University, 600 Forbes Ave., 440 College Hall, Pittsburgh, PA 15282, *kern@mathcs.duq.edu*; Nicholas Bernini,

Duquesne University; Sujit Ghosh, North Carolina State University

***Key Words:*** Poisson regression, knot locations, MCMC

Piecewise linear Poisson regression is used to model, for a specific individual, the mean frequency of a particular event as a function of time. Parameters of this model include the knot locations of the piecewise linear function and are estimated using MCMC sampling techniques. The hierarchical component of the model allows for the estimation of treatment group profiles and hence for direct group comparison. We apply this model to frequency data collected from breast cancer survivors receiving menopausal symptom relief treatment.

### Bayesian Spatial Modeling of Data from Bird Surveys

❖ Raymond Webster, North Carolina State University, Department of Zoology, Campus Box 7617, Raleigh, NC 27695, *rawebste@ncsu.edu*; Kenneth Pollock, North Carolina State University; Theodore Simons, North Carolina State University

***Key Words:*** bird surveys, count data, CAR model, detection probabilities, capture-recapture

Past analyses of large bird survey data have ignored variation in bird detection probabilities across space and spatial dependence in bird density. We present a unified framework for modeling bird survey data in the form of repeated counts, removal counts, or capture history data that accounts for both spatial dependence in density and variation in detection probabilities between survey sites. The hierarchical structure of the models makes them suited to Bayesian analysis using Markov chain Monte Carlo algorithms. For computational efficiency, we use a form of conditional autoregressive model proposed by Hrafnkelsson and Cressie (2003) for modeling spatial dependence. We apply our models to data from a large survey in the Great Smoky Mountains National Park. Our analyses have implications for survey design and show the limitations of fitting such complex models to sparse data.

### Bayesian Inference for NLME Models Involving ODEs

❖ Sujit Ghosh, North Carolina State University, 2501 Founders Drive, Department of Statistics, Raleigh, NC 27695, *sghosh@stat.ncsu.edu*; Lovely Goyal, North Carolina State University

***Key Words:*** Bayesian, mixed effects, nonlinear, MCMC, PKPD, ODE

Nonlinear mixed effects (NLME) models, such as PK/PD and PBPK, have been used widely in various biomedical fields. Such models are desirable because they allow us to study the dynamics of a drug within an individual represented by a set of ordinary differential equations (ODEs). Likelihood-based inference becomes challenging and computationally intensive when the system of ODEs is large and no analytical solution is available. We propose a new approach based on Euler approximation that allows us to obtain a tractable likelihood that approximates the original likelihood as the grid size approaches zero at a certain rate. We also develop efficient MCMC methods within a Bayesian framework to obtain parameter estimates. To illustrate our method, we apply it to data on HIV patients and present simulation studies for model validation and comparison to other competing approaches.

### Bayesian Modeling of Longitudinal Data with Nonignorable Missing Data

❖ Liansheng Zhu, North Carolina State University, 2727B Conifer Drive, Raleigh, NC 27606, *lzhu@stat.ncsu.edu*; Sujit Ghosh, North Carolina State University; Subhashis Ghosal, North Carolina State University

***Key Words:*** longitudinal, non-ignorable, missing, joint-modeling, Bayesian, pattern-mixture

Missing data are encountered often in longitudinal studies despite every attempt to collect full data. When the missingness is informative, it is generally difficult to model and analyze incomplete data because the distributional assumptions about missing data are not easily verifiable. Pattern-mixture models are used commonly in practice, in which patterns are determined by time to occurrence of missing data. Marginalization over patterns can be tricky, as outcomes are discrete. We propose a joint-modeling approach within a Bayesian framework in which patterns as random effects are marginalized within a generalized linear mixed model framework. Advantages of this approach include its being capable for both continuous and discrete responses, it is avoidant of the problem that not all parameters are estimable, it is computationally efficient, and it provides marginal estimates.

## 270 Recent Advances in Analyzing Agreement Data ●

Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Bivariate Modeling of Interobserver Agreement Coefficients

❖ Mohamed Shoukri, King Faisal Specialist Hospital, MBC#3, PO Box 3354, Riyadh, 11211 Saudi Arabia, *shoukri@kfshrc.edu.sa*; Allan Donner, University of Western Ontario

***Key Words:*** kappa coefficient, correlated binary data, bivariate beta distribution, pseudo likelihood estimation

Shoukri and Donner (2001) developed a model to estimate inter-rater and intra-rater agreement when each of two observers has the opportunity to obtain a pair of replicate binary measurements on each subject. A practical limitation of this model is that it accommodates only a single binary trait. In many studies, however, more than one trait may be of interest. For example, Becker et al. (2002) reported on a study of depression and anxiety as two outcome measures obtained on 173 patients evaluated by two methods, the first method using a patient health questionnaire and the second employing a structured clinical interview. In this case, although inferences on each trait separately are usually of main interest, a question also arises as to whether the level of inter-rater agreement is stable across traits.

### A Weighted Kappa for Agreement between Two Ratings with Different Ordinal Scales

❖ Stuart Lipsitz, Brigham and Women's Hospital, Division of General Internal Medicine, 1 Brigham Circle, BC3 002D, Boston, MA 02120, *lipsitz@clinepi.bwh.harvard.edu*

***Key Words:*** r x c contingency table

Agreement studies, where several observers may be rating the same subject for some characteristic measured on an ordinal scale, provide important information. The weighted kappa coefficient is a popular measure of agreement for ordinal ratings. However, in some studies, the raters use scales with different numbers of categories. For example, a patient quality-of-life questionnaire may ask "How do you feel today?" with possible answers ranging from 1 (worst) to 7 (best). At the same visit, the doctor reports his view of the patient's health status as very poor, poor, fair, good, or very good. The weighted kappa coefficient is not directly applicable here as the two scales have a different number of categories. We will discuss weighted kappa coefficients to measure agreement between such ratings.

### Coefficient of Agreement for Binary Data with Replications

❖ Huiman Barnhart, Duke University, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715, *huiman.barnhart@duke.edu*; Michael Haber, Emory University; Andrzej Kosinski, Duke University

***Key Words:*** agreement, kappa, method comparison, reliability

In method comparison studies, we are often interested in whether two methods can be used interchangeably or a new method can replace an existing method. For binary data, the kappa index traditionally has been used for assessing agreement between two methods. We present a new approach to define coefficients of individual agreement (CIA) between two methods for binary data with replications. Unlike the kappa index, which is scaled relative to agreement under independence, our coefficient is scaled relative to the agreement between replications. This is based on the intuition that if the methods are used interchangeably or one method is replacing the other, the difference between measurements from the two methods should not be much larger than the difference between replicated measurements. We present the relationship between kappa and CIA and illustrate the method with several examples.

### A Permutation Test for Dependent Concordance Correlation Coefficients

❖ John M. Williamson, Centers for Disease Control and Prevention, 4770 Buford Hwy., MS F22, Atlanta, GA 30341, *jow5@cdc.gov*; Hung-Mo Lin, The Pennsylvania State University; Sara Crawford, The Rollins School of Public Health of Emory University

***Key Words:*** resampling, agreement, concordance correlation, permutation test

The concordance correlation coefficient (CCC) is a popular index for measuring the reproducibility of continuous variables. Inference on correlated CCCs is problematic due to the difficulty in estimating the appropriate standard errors. We propose a resampling approach (permutation test) for conducting hypothesis tests on dependent CCCs. To illustrate the approach, we analyze data from a carotid stenosis screening study.

## 271 Extremes: Methods for Environmental and Meteorological Studies ● ✪

WNAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### How Good Are the Air Quality Standards?

❖ Sofia Aberg, Lund University, Box 118, Lund, 221 00 Sweden, *s_aberg@maths.lth.se*; Peter Guttorp, University of Washington

***Key Words:*** air quality standards, distribution of the maximum, random fields, Rice's formula

Air quality standards are set to protect public health, and their values typically are based on health effect studies without any statistical considerations. In order to judge if a standard is met, measurements of ambient air quality are collected at monitoring stations and used to decide whether the standard has been violated. In this paper, we examine the statistical quality of air quality standards by computing the distribution of the maximum of the ambient field conditional on a measured monitoring value at the value prescribed by the standard. The distribution of the maximum is computed using Rice's method and relies on a generalization of upcrossings of a level in one dimension to two dimensions.

### A Statistical Approach to North American Rainfall

❖ Jie Zhou, The University of North Carolina at Chapel Hill, 210 Smith Building, CB 3260, Chapel Hill, NC 27599, *jzhou@email.unc.edu*

The North American Rainfall Data collected by NCDC consist of rainfall values observed from about 6,000 stations each day throughout North America for the last 100 years. Based on these data, we would like to generate seasonal extreme value distributions for each station and find trends in extremes between different stations and different seasons. Based on these results, our objective is to make predictions on rainfall extremes for each location of North America. Statistical tests also were developed to examine whether the statistically generated ones match the data from NCEP. The NCEP data are grid-cell averages, while the observational data from NCDC are point-based. These statistical analyses provide a statistical relationship between these sets of data under appropriate uncertainty measures.

### Bayesian Hierarchical Modeling for Extreme Values in Spatio-Temporal Data

❖ Huiyan Sang, Duke University, 214 Old Chemistry Building, Duke University, Durham, NC 27708, *hs37@duke.edu*; Alan E. Gelfand, Duke University

***Key Words:*** CAR model, coregionalization, Markov chain Monte Carlo, precipitation data

We propose a hierarchical Bayesian approach for modeling a time series of high-resolution spatial extreme values. We assume the observations follow a Generalized Extreme Value (GEV) distribution for which the spatial dependence and temporal correlation are captured through the specifications of location, scale, and shape parameters. A conditionally autoregressive (CAR) coregionalization model is employed to specify the covarying parameters over the same region. The model is implemented using a Markov chain Monte Carlo (MCMC) algorithm to enable inference for parameters and provide spatio-temporal predictions. We apply the model to a set of the gridded interpolated precipitation data collected over a 50-year period for the Cape Floristic region in South Africa.

### Predicting Extreme Hurricane Winds in the United States

❖ Thomas Jagger, Florida State University, 3122 S. Leyden Street, Denver, 80222, *tjagger@blarg.net*

***Key Words:*** EVT, GPD, Bayesian, Pot

We demonstrate the use of POT (Peaks over Threshold) models to evaluate distribution of extreme winds within near-coastal regions of the United States. Initially, we use maximum likelihood methods to estimate the return levels for various long-range return periods for each coastal region. Next, we show that the return levels depend on climate variables such as ENSO and NAO. Finally, we demonstrate the usefulness of the Bayesian approach to POT modeling using WinBUGS software to model the relationship between climate variables and hurricane intensity. Additionally, we show that the Bayesian approach is useful for extending the historical record, managing measurement error, imputing missing values, and simulating future hurricane intensities.

### Statistical Analysis of a POT Model with Time-Varying Extreme Parameters

❖ Marta Nogaj, Commission of Atomic Energy, LSCE, Orme des Merisiers Bat 701, Gif sur Yvette, 91191 France, *marta.nogaj@cea.fr*

***Key Words:*** non-stationary extremes, climate change, non-stationary return levels, Pot model

In a changing climate, the stationary hypothesis of the EVT needs to be reconsidered. We thus examine the properties of a nonstationary peak over threshold (POT) model consisting of two complementary submodels: a GPD with a time-varying scale parameter for the values above a threshold and a Poisson distribution with a varying parameter describing the frequency of occurrences of extreme events. We investigate the theoretical accuracy of our approach and revisit and generalize the notion of return levels. We use this methodology to assess the changes in severity and frequency of extreme high temperatures during the last 50 years (from the NCEP reanalysis data) over the North Atlantic. This approach brings insight on the connection between average and extreme changes.

## 272 Outcome-Dependent Sampling ● ✪

Section on Statistics in Epidemiology

**Tuesday, August 8, 10:30 am–12:20 pm**

### The Combination of Ecological and Individual-Level Data

❖ Jonathan Wakefield, University of Washington, Health Sciences Building, Department of Biostatistics, Seattle, WA 98195, *jonno@u.washington.edu*; Sebastien Haneuse, Group Health Cooperative

***Key Words:*** ecological data, outcome-dependent sampling, ecological bias, case-control studies

In an ecological study outcome and exposure/confounder, data are available on groups of individuals, rather than on the individuals themselves. Such studies are logistically appealing, as they may make use of routinely available data and offer increased power and exposure contrasts. However, they suffer from a number of problems. Ecological bias can appear in simple situations. For example, the Scottish lip cancer data has been analyzed by numerous authors who have used an inappropriate mean model when viewed from an individual-level perspective. In this talk, various study designs and estimation methods will be described and compared. In particular, the ecological embedded case-control study of Haneuse and Wakefield (2005) will be outlined; a likelihood will be developed that acknowledges that the case-control data are drawn from the finite sample ecological data.

### Two-Phase Methods for Ecological Studies

❖ Sebastien Haneuse, Group Health Cooperative, 1730 Minor Ave., Suite 1600, Seattle, WA 98101, *haneuse.s@ghc.org*; Jonathan Wakefield, University of Washington

***Key Words:*** two-phase study, ecological study, Bayesian methods, biased sampling

Ecological studies are designs where the sampling unit is a group of individuals. Their use is limited due to a variety of biases, which may lead to the ecological fallacy. To avoid these difficulties, it is necessary to collect individual-level data. We consider the two-phase design as a means to combining group- and individual-level information. Current inferential techniques rely on asymptotics, and their small-sample properties have not been extensively examined. We develop a Bayesian framework for inference. The latter will help with small-sample situations and acknowledge spatial structure inherent in geographic-based applications.

### The Optimal Ratio of Cases to Controls for Estimating the Classification Accuracy of a Biomarker

❖ Holly E. Janes, Johns Hopkins University, Bloomberg School of Public Health, 615 North Wolfe Street, Office E3536, Baltimore, MD 21205, *hjanes@jhsph.edu*; Margaret S. Pepe, Fred Hutchinson Cancer Research Center/University of Washington

***Key Words:*** sensitivity, specificity, case-control study, ROC curve, efficiency, sample size

The case-control design is used frequently to study the classification accuracy of a screening or diagnostic biomarker. The appropriate case-control sampling ratio has never been determined. It is common to sample equal numbers of cases and controls, a strategy that can be optimal for studies of association. However, considerations are different when the biomarker is to be used for classification. We provide an expression for the optimal case-control ratio when the accuracy of the biomarker is quantified by the ROC curve and integrate it with sample size calculations to yield an efficient study design with specified power and type-I error. We also derive the optimal case-control ratios for estimating the area under the ROC curve and the area under part of the ROC curve. Our methods are applied to a study of a new marker for adenocarcinoma in patients with Barrett's esophagus.

### An Efficient Outcome-Dependent Sampling Design for Longitudinal Binary Response Data

❖ Jon Schildcrout, Vanderbilt University Medical Center, Department of Biostatistics, S-2323 Medical Center North, Nashville, TN 37212, *jonathan.schildcrout@vanderbilt.edu*; Patrick Heagerty, University of Washington

***Key Words:*** study design, binary response, time-varying covariates, outcome dependent sampling, longitudinal data analysis, epidemiology

Regression analysis of longitudinal data often seeks to estimate the prospective relationship between an exposure and expected response over time. With administrative, panel study, or cohort study data, we may have complete response vectors on individuals, but the exposure of interest is expensive to ascertain. We will describe a study design for the scenario where binary response vectors are acquired easily but exposure measurement is costly and therefore limiting. In this design, individuals are selected based on their response profiles. We can acknowledge the selection mechanism properly using a conditional likelihood, which leads to consistent parameter estimation. Under realistic scenarios, we show that outcome dependent sampling designs and subsequent maximum conditional likelihood analysis yields near fully efficient estimates relative to collection of data on all subjects.

# 273 The Transition from the Long Form to the American Community Survey: Findings from The National Academies Panel ●

Section on Survey Research Methods
**Tuesday, August 8, 10:30 am–12:20 pm**

## Functionality and Usability of Information from the American Community Survey

❖ Connie Citro, Committee on National Statistics, Keck Center, Room 1139, 500 5th St NW, Washington, DC 20001, *ccitro@nas.edu*; Graham Kalton, Westat; Michael L. Cohen, Committee on National Statistics

***Key Words:*** survey design, data products, small-area data

The American Community Survey is about to issue its first round of data products for states and large counties and cities. These products represent a major innovation for data users accustomed to the long form. This talk will provide the general structure of the ACS sampling design and its strategy for data collection, discuss the use of weighting techniques to reduce sampling variance, address nonresponse and undercoverage, and touch on the data products and strategy for their release---especially the issuance of one-, three-, and five-year estimates for various jurisdictions. It also will set the stage for two remaining talks on the impact of the ACS on users and the methodological issues the ACS raises and present the major findings and recommendations of a panel of the Committee on National Statistics that reviewed the ACS from the user perspective.

## The ACS and the Data User Community

❖ Joseph Salvo, New York City Department of City Planning, 3307 Ampere Ave., Bronx, NY 10465, *jsalvo@planning.nyc.gov*; Nancy Dunton, The University of Kansas Medical Center; Charles Purvis, Oakland Metro Transportation Commission

***Key Words:*** American Community Survey, non-sampling error, multi-year estimates, sampling error, population controls, data collection

The recent introduction of the ACS represents a major innovation for data users who have been accustomed to using the decennial census long form. In many ways, the ACS presents us with a golden opportunity, given the more timely release of products, the higher quality of data, and the time series of estimates. However, higher sampling variability, the issuance of competing estimates for areas of different size, the change in residence rules, the disclosure and suppression rules, and the lack of a point-in-time representation of estimates will present challenges for users. Some specific issues addressed include the treatment of monetary amounts, the impact of changes in geographic boundaries, and questionnaire wording on the utility of multi-year estimates, as well as the impact of population controls on areas with strong seasonal population changes.

## Methodological Issues in the ACS

❖ Graham Kalton, Westat, 1650 Research Blvd., Rockville, MD 20850, *grahamkalton@westat.com*; Tim Holt, Retired; Sharon Lohr, Arizona State University; Hal Stern, University of California, Irvine

***Key Words:*** population estimates, weighting adjustments

Each year, a sample of housing units is selected for the ACS from the Master Address File and allocated into monthly samples. Data from the monthly samples are then accumulated to produce one-, three-, and five-year estimates. The weighting for the one-year estimates includes an adjustment to make ACS data conform to mid-year population estimates by major demographic group at the county level. The effect of error in the population estimates and the effect on areas with seasonal populations are discussed. The effects of changes in the size and characteristics of an area on the area's multi-year estimates are examined under the planned weighting scheme and alternative schemes. The effects of the cost-of-living adjustments made to dollar amounts collected in ACS are reviewed. Other methodological issues also are discussed.

# 274 Student Paper Award Winners

Section on Statistical Computing, Section on Statistical Graphics
**Tuesday, August 8, 10:30 am–12:20 pm**

## Efficient Computation and Variable Selection for the L1-Norm Quantile Regression

❖ Youjuan Li, University of Michigan, 1737 Cram Circle, Apt. 8, Ann Arbor, MI 48015, *youjuan@umich.edu*; Ji Zhu, University of Michigan

***Key Words:*** degrees of freedom, L1-norm, quantile regression, variable selection

Classical regression methods have focused on mainly estimating conditional mean functions. However, in recent years, quantile regression has emerged as a comprehensive approach to the statistical analysis of response models. We consider the L1-norm regularized quantile regression (L1-norm QR), which uses the sum of the absolute values of the coefficients as the penalty. The L1-norm penalty has the advantage of simultaneously controlling the variance of the fitted coefficients and performing automatic variable selection. We propose an efficient algorithm that computes the entire solution path of the L1-norm QR. We also derive an unbiased estimate for the degrees of freedom of the L1-norm QR model, which allows convenient selection of the regularization parameter.

## An Algorithm for Regression of Scalars on Images

❖ Philip Reiss, Columbia University, Department of Biostatistics, 722 W. 168th Street, 6th Floor, New York, NY 10032, *ptr2003@columbia.edu*; R. Todd Ogden, Columbia University

***Key Words:*** brain imaging, b-splines, dimension reduction, functional data analysis, image regression, partial least squares

When neuroimaging techniques are used to measure a quantity at each voxel (volume unit) of the brain for each of a set of subjects, it is common practice to relate this quantity to a scalar variable of interest through a "mass-univariate" approach: Regress the imaged quantity on the scalar variable separately at each voxel and identify the voxels at which a significant effect is observed. This paper adopts the opposite approach; we treat the scalar as the outcome and regress it on the imaged quantity at all voxels simultaneously to produce a coefficient image. As this single model typically has many more predictors than cases, one must reduce the dimension of the predictors drastically. We present an algorithm combining three dimension-reduction techniques: projection onto a spline basis, selection of the most relevant predictor components, and thresholding.

### Clustering with Confidence

❖ Rebecca Nugent, University of Washington, Box 354322, Department of Statistics, Seattle, WA 98195-4322, *rnugent@stat.washington.edu*; Werner Stuetzle, University of Washington

***Key Words:*** plug-in estimation, cluster tree, connected components, confidence sets

We present a plug-in method for estimating the cluster tree of a density. The method takes advantage of the ability to compute exactly the cluster tree of a piecewise constant density estimate. We then introduce "clustering with confidence," an automatic pruning procedure that assesses significance of splits in the cluster tree; the only user input is the desired confidence level.

### Kernel Regularization and Dimension Reduction

***Key Words:*** dissimilarity measures, kernel, regularization, convex cone programming

It is often possible to use expert knowledge or other sources of information to obtain dissimilarity measures for pairs of objects, which serve as pseudo-distances between the objects. When dissimilarity information is available as the data, there are two types of problems of interest. The first is to estimate full position configuration for all objects in a low-dimensional space while respecting the data. In the second type of problem, the high-dimensional data points are assumed to lie on a low-dimensional manifold, and the goal is to unfold the manifold to recover the underlying intrinsic low-dimensional structure. We provide a novel, unified framework called Kernel Regularization that can be used to solve either problem. Various favorable operating characteristics and properties of the method will be illustrated via simulated and real data.

## 275 Measuring and Improving Data Quality ●

Section on Government Statistics
**Tuesday, August 8, 10:30 am–12:20 pm**

### How Do We Know If We Aren't Looking? An Investigation of Data Quality in the SCF

❖ Arthur Kennickell, Federal Reserve Board, 20th C Street, NW, Mail Stop 153, Washington, DC 20551, *Arthur.Kennickell@frb.gov*

***Key Words:*** interviewers, data quality

"Data quality" is a term often used vaguely to indicate some cluster of desirable traits. This paper argues that the most useful notion of data quality is one that turns on the utility of data for the analytical purposes for which it was intended. Although the standard attributes, such as missing data rates, are important, more subtle matters can be critical. Unfortunately, many such factors are difficult---even impossible---in the absence of advanced AI technology, to identify mechanically. This paper focuses on an exercise undertaken with the 2004 Survey of Consumer Finances (SCF) involving review of the data by subject matter experts and the use of feedback to interviewers based on that review.

### An Economic Analysis of Survey Response Quality

❖ Julia Lane, National Opinion Research Center, 1350 Connecticut Ave., NW, Suite 500, Washington, DC 20036, *lane-julia@norc.org*; Leslie Athey, National Opinion Research Center; Suzanne Bard, National Opinion Research Center; Judith Petty, National Opinion Research Center

***Key Words:*** data quality, interviewer recruitment, interviewer training, interviewing, feedback

Policymakers and researchers rely on large, complex, federally funded, in-person surveys to provide accurate information about a range of issues affecting American life. The quality of information collected in individual data items is as important a measure of the success of these surveys as the more traditional metrics of sample size, response rate, and cost. Interviewers are central to obtaining clear and accurate answers to survey questions, and recruiting the appropriate individuals and training and motivating them to produce high-quality data is a critical element in minimizing survey error. This paper begins by providing an overview of the issues involved. It uses an economic framework to examine the relationships among interviewer characteristics, training, ongoing feedback on data quality, and measures of response quality using a large longitudinal field study database.

### Effects of Late-Stage Completions on Data Quality

❖ Karen Grigorian, National Opinion Research Center, 55 E. Monroe Street, Suite 1800, Chicago, IL 60603, *grigorian-karen@norc.uchicago.edu*; Thomas B. Hoffer, National Opinion Research Center; Ronald S. Fecso, National Science Foundation

***Key Words:*** survey, quality, response rate, data collection, schedule, cost

Survey contractors try to obtain OMB-suggested response rates of 80% or higher, but often find they spend a high and disproportionate amount of the survey's resources completing late-stage cases. This paper uses the 2003 Survey of Doctorate Recipients (SDR) to assess the impact on data quality if data collection were stopped sooner. The 2003 SDR achieved a 79% unweighted and 80% weighted response. We first compares the demographic characteristics used in sample stratification for cases completed at the 75% and 79% unweighted response rates. Second, we compare the two weighted datasets on key nonstratification variables to determine the extent to which basic survey results change if data collection is stopped sooner. Next, we evaluate data quality at the item-level for late-stage cases. Last, we assess schedule and cost implications of ending early and offer recommendations.

## 276 High Dimensional Modeling and Model Selection

Section on Bayesian Statistical Science
**Tuesday, August 8, 10:30 am–12:20 pm**

### Priors for High-Dimensional Covariance Models

❖ Charles Curry, University of California, Santa Cruz, 265 Union Ave., C1062, Santa Cruz, CA 95008, *ctc@ams.ucsc.edu*

***Key Words:*** covariance modeling, Bayesian statistics, MCMC, reference prior

We extend empirical orthogonal function analysis and optimal fingerprint detection with Bayesian probability models and Markov chain Monte Carlo computational techniques. Using our technique, we fully characterize the uncertainty in climate system properties, as modeled by the MIT 2DLO model. To improve performance of our methodology, we approximate the response of the climate model with a fast, statistically equivalent model and incorporate the additional error added

by this approximation into the analysis. Additionally, we include in our parameter set the signal and noise covariance structure of the diagnostics. This goes beyond the maximum likelihood and significance testing strategies employed by optimal fingerprint detection, as it includes the uncertainty in the noise components in the final parameter distributions.

### High-Dimensional Modeling and Model Selection

❖ Carlos Carvalho, Duke University, Durham, NC 27708-0251,
*carlos@isds.duke.edu*

***Key Words:*** sparse factor models, evolutionary search

We describe latent factor models for multivariate analysis in very high dimensions and classes of models that couple this framework with factor regressions for predictive modeling of multivariate response variables. We use sparse factor models---relationships between high-dimensional variables and underlying, lower-dimensional latent factors are sparse---created using sparsity-inducing priors. Model search and fitting are addressed through stochastic simulation (MCMC) and a novel evolutionary search. The latter computational approach explores, defines, and fits models for higher-dimensional problems through an evolutionary process that gradually expands the dimension of the sample space. Examples are drawn from studies in breast cancer genomics, where the sparse factor models represent observed relationships in gene expression of thousands of genes.

### Sparsity in Linear Regression for High-Dimensional Problems

❖ Christopher Hans, The Ohio State University, 1958 Neil Ave.,
Cockins Hall Room 404, Columbus, OH 43210-1247,
*hans@stat.ohio-state.edu*

***Key Words:*** marginal likelihood, model uncertainty, variable selection, Bayesian regression, prior distribution

Characterizing linear regression model uncertainty when there are many possible predictor variables often involves the use of prior distributions that encourage "sparse" models---those with relatively few predictor variables. Sparsity is induced explicitly through the prior distribution on the model space, where a large penalty can be imposed for the inclusion of a variable in a model. Sparsity also is induced implicitly through the marginal (integrated) likelihood, which is a function of the prior distribution on the parameter space (regression coefficients). We present results regarding marginal likelihoods for several common classes of parameter space priors that help us compare the amount of sparsity they induce. The results also are used to aid in Bayesian learning about the amount of sparsity supported by the data.

### A Bayesian Nonparametric Approach for Expression Proteomic Analyses

❖ Leanna House, Duke University, Box 90251, Duke University,
Durham, NC 27705, *house@stat.duke.edu*; Merlise Clyde, Duke
University; Robert Wolpert, Duke University

***Key Words:*** Bayesian, nonparametric, over-complete, proteomics, Levy process

Expression proteomic analyses that use profiles generated by Matrix Assisted Laser Desorption Ionization, Time-of-Flight mass spectrometers (MALDI-TOF MS) require three components: quality control, feature extraction, and inference. We simultaneously address the components by assuming an over-complete functional data analysis approach that capitalizes on an underlying levy process and uses information across multiple spectra. Specifically, we assert that one, m/z, and resolution-dependent stochastic gamma process influences every within-population, multi-modal spectrum; yet, random, biological, or measurement error force spectra to deviate from the process parameters. Under this assertion, a Bayesian hierarchical approach naturally models data quality control variables and peak parameters while leading to posterior predictions of experimental-group status.

### Making Sense of Estimating Integrated Likelihoods: Bridges, Paths, and Other Routes

❖ Russell Steele, McGill University, 805 Rue Sherbrooke O,
Montreal, H3A 2K6 Canada, *steele@math.mcgill.ca*

***Key Words:*** Bayesian model selection, Monte Carlo integration, importance sampling, Markov chain Monte Carlo

In this paper, we will examine the usage of advanced Monte Carlo integration techniques---such as annealed importance sampling, bridge sampling, and path sampling---in the context of common model selection problems. We will show how the strengths and weaknesses of each method for model selection for linear mixed-effects models, meta-analysis, and neural network models. The focus of the talk will be on diagnosing problems that occur in various situations and suggesting potential solutions.

## 277 Topics in Function Estimation Using Isotonization and Smoothing Techniques ✪

IMS, Section on Nonparametric Statistics
**Tuesday, August 8, 10:30 am–12:20 pm**

### Nonparametric Density Estimation from Covariate Information

❖ Ryan Elmore, Colorado State University, Department of Statistics,
Fort Collins, CO 80523, *elmore@stat.colostate.edu*; Peter G. Hall,
Australian National University; Vladimir Troynikov, Department of
Primary Industries

***Key Words:*** calibration, deconvolution, density estimation, functional data analysis, kernel methods, mixture models

An increasing number of statistical problems arise in connection with functional calibration. In each case, inexpensive, indirect data in a particular context are combined with direct, expensive-to-acquire data from different but related settings, so as to estimate quantities in the former case. Suppose we observe data that give us access to the distribution of U given V, and, from these and data on U, we wish to estimate the density of V. The motivating real datasets are of age and covariate information in fish populations. We suggest two methodologies, each based on transforming the problem to one that involves inversion of a symmetric, linear operator. Our techniques have connections to methods for functional data analysis and a variety of mixture and deconvolution problems, as well as to calibration techniques.

### Limiting Distribution of the NPMLE for Current Status Data with Competing Risks

❖ Marloes Maathuis, University of Washington, Department
of Statistics, UW Campus Box 354322, Seattle, WA 98195,
*marloes@stat.washington.edu*

***Key Words:*** current status data, competing risks, asymptotics, nonparametric estimation, maximum likelihood estimator, survival analysis

Current status data with competing risks naturally arise in cross-sectional survival studies when there are several failure causes. Given such data, we study the nonparametric maximum likelihood estimator (NPMLE) of the subdistribution functions. We focus on asymptotic properties of the NPMLE, which are nonstandard due to the current status censoring. In particular, we discuss global and local consistency, the global and local n^{1/3}-rate of convergence, and the local limiting distribution.

### Penalized Likelihood Ratio Method for the Spiking Problem in Nonincreasing Density Estimation

❖ Jayanta Pal, University of Michigan, 1929 Plymouth Road, Apt 3031, Ann Arbor, MI 48105, *jpal@umich.edu*

***Key Words:*** likelihood, monotone, density, penalization, Brownian

The maximum likelihood estimation of a decreasing density, f, on the positive real line creates inconsistent estimate of f(0+). This is known as the spiking problem. Woodroofe and Sun (1993) considered a penalized likelihood and achieved consistent estimates. In this article, we characterize the restricted MLE under the null hypothesis H_0 : f(0+) =c. The target is to find out the (asymptotic) distribution of the (penalized) likelihood ratio under the null distribution. It will be useful to construct asymptotic level-alpha confidence interval for the endpoint. The limit distribution of the likelihood ratio with a chosen penalization is shown to be universal, and the quantiles can be numerically tabulated. This problem has applications in renewal theory and astronomical examples.

### Fractile Graphical Analysis with Multiple Covariates

❖ Bodhisattva Sen, University of Michigan, 439 West Hall, 1085 South University, Ann Arbor, MI 48109, *bodhi@umich.edu*

***Key Words:*** fractile graphical analysis, fractile graphs, geometric quantile, Mahalanobis quantile, kernel regression

The concept of fractile graphical analysis (FGA) was introduced by Mahalanobis (1960). It is a method to compare the regression functions for two bivariate populations (X,Y). The method is particularly useful for comparing two regression functions where the covariates (X) for the two populations are not necessarily on comparable scales. Kernel-smoothed version of fractile regression functions (fractile graphs) are considered in Sen (Sankhya 2005). In this talk, we extend the notion of FGA to deal with the multiple covariate set up (when X is a vector). We develop smooth estimates of fractile graphs and study their statistical properties. In the process, we discuss suitable notions of multivariate quantiles (the geometric and Mahalanobis' quantile). The method has been applied to real data collected from the Reserve Bank of India to draw interesting and useful conclusions.

## 278 Sample Survey Quality III ●

Section on Survey Research Methods

**Tuesday, August 8, 10:30 am–12:20 pm**

### An Evaluation of Methods To Compensate for Noncoverage of Phoneless Households Using Information on Interruptions in Telephone Service and Presence of Wireless Phones

❖ Meena Khare, National Center for Health Statistics, CDC, 3311 Toledo Road, Room 3218, Hyattsville, MD 20782, *mkhare@cdc.gov*; Sadeq Chowdhury, National Opinion Research Center; Kirk Wolter, National Opinion Research Center; Karen Wooten, National Immunization Program; Stephen Blumberg, National Center for Health Statistics

***Key Words:*** National Health Interview Survey, propensity score adjustments, ratio adjustments, weighting methods, NHIS

For legal and operational reasons, wireless phone numbers are usually excluded from RDD surveys. Wireless-only households are rapidly growing and about 6-10% of the U.S. households are wireless-only. Estimates from RDD surveys are subject to potential bias due to noncoverage of households without landline telephones. We use combined data from the 2003-2004 NHIS to compare characteristics of adults living in phoneless households with those living in wireless-only households or households with interruptions in telephone service. We select a random sample of telephone households and evaluate ratio- and propensity-based weighting methods to compensate for noncoverage of phoneless and wireless-only households using information on interruptions in landline telephone service and presence of wireless telephones. To assess bias, resulting estimates are compared with the NHIS estimates.

### Using Record Linkage Techniques To Improve Online Genealogical Research with Census Records

❖ John Lawson, Brigham Young University, 223B TMCB BYU, Department of Statistics, Provo, UT 84602, *lawson@byu.edu*

***Key Words:*** census index records, online search, family history, probabilistic record linkage, multinomial distribution, EM algorithm

Since the first edition of Alex Haley's book, Roots, there has been a fascination with tracing family origins. A Google search for "genealogy" results in millions of hits. Some web sites offer users online searches of census index records, which are electronic databases containing summary information from census heads of households. When searching these records, results can be frustrating unless a user enters the exact name, spelling, and age recorded in the census index. This paper describes the implementation of probabilistic record linkage to improve the search. By sampling the 1910 and 1920 census index records, record linkage weights were developed that allow intelligent searches recognizing reasonable variations in search criteria. Using these weights, the chance of missing a record was reduced from more than 50% to less than 2%.

### Integrating Culture Industries into Canada's National Economic Surveys Program

❖ Mary March, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A 0T6 Canada, *marcmar@statcan.ca*; Paddison Wong, Statistics Canada

***Key Words:*** integration, backcasting, frame reconciliation, annual survey

The first wave of a project to integrate Statistics Canada's cultural surveys program with its economic data program is nearing completion. The project faced numerous challenges including reconciliation of two different target populations and frames, redesign of questionnaires and concepts, attempting to deal with user requirements for data not normally collected by an economic survey, introduction of statistical sampling, introduction of new quality measures and developing a backcasting methodology (using data collected by the previous program, administrative data and data collected by the newly integrated survey) to provide some longitudinal data corresponding to the new target population. The paper discusses analyses that were undertaken to better understand culture statistics program needs and will explain and evaluate the methodological solutions that were developed.

### The Internet Response Method: Impact on the Canadian Census of Population Data

❖ Laurent Roy, Statistics Canada, R.H. Coats Building, 15th Floor, Section O, Tunneys Pasture, Ottawa, ON K1A 0T6 Canada, *Laurent.Roy@statcan.ca*

***Key Words:*** internet, response method, mode effect, questionnaire, census

The option to complete the census questionnaire via the Internet is one of the major changes for the 2006 Canadian Census of Population. The 2004 Census Test (CT) tested most systems and operations in preparation of the 2006 Census, including the Internet collection method. After the CT, some analyses were carried out to measure the impact of the Internet collection method on the data as compared with the traditional paper response method. The impact was analyzed on the basis of item response rates and response content. The analysis focused on the comparison of data for persons who were reported on the Internet during the 2004 CT but were reported on paper during the 2001 Census. These comparisons helped us to measure the Internet mode effect on item response rates and content. The presentation will discuss some features of the Internet questionnaire and the analysis of the 2004 CT data.

### Good Item or Bad: Can Latent Class Analysis Tell? Examining the Effectiveness of the Latent Class Analysis Approach to Item Evaluation

❖ Frauke Kreuter, University of Maryland, 1218 Lefrak Hall, Joint Program in Survey Methodology, College Park, MD 20742, *fkreuter@survey.umd.edu*; Roger Tourangeau, University of Maryland; Ting Yan, Institute for Social Research

***Key Words:*** latent class analysis, measurement error, social desirability bias, mode effects, questionnaire pretesting

Latent class analysis (LCA) has been used to model measurement error, to identify flawed survey questions, and to compare mode effects in the survey context. Compared to other approaches, LCA has the advantages that it doesn't require true values ("a gold standard") or the stringent assumptions usually assumed in classical measurement error analysis. However, there has been no evaluation of this technique yet in terms of its accuracy and effectiveness for detecting bad questions. This paper seeks to determine whether LCA yields plausible results that agree with accepted procedures for assessing questionnaire items. Specifically, we will examine, in the presence of a gold standard, how well the results from LCA converge with those from a comparison of survey responses with a gold standard.

### Population Coverage in the National Survey on Drug Use and Health

Susan R. Hunter, RTI International; ❖ Katherine B. Morton, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *bowman@rti.org*; James Chromy, RTI International; Peilan C. Martin, RTI International

***Key Words:*** household surveys, response rates, coverage rates

The National Survey on Drug Use and Health (NSDUH) is an annual national survey of the civilian, non-institutionalized population aged 12 or older in the 50 United States and the District of Columbia. This paper examines the coverage and response rates for various populations in the NSDUH using data from the 1999-2001 National Household Survey on Drug Abuse (NHSDA; prior to 2002 the NSDUH was called the NHSDA) and the 2002-2004 NSDUHs. Among the findings is that coverage rates of the Hispanic and 'Other' race/ethnicity populations tend to be higher than the Non-Hispanic Black and Non-Hispanic White groups. However, some of the higher coverage rates among Hispanic and 'Other' race/ethnicity groups may be an artifact of low intercensal projections due to the growing Hispanic and multiracial populations rather than a true measure of frame coverage.

## 279 Survey-Based Estimation III

●

Section on Survey Research Methods
**Tuesday, August 8, 10:30 am–12:20 pm**

### New Cell Collapsing Strategies Based on Collapsing Adjustment Factor

❖ Jay J. Kim, National Center for Health Statistics, 3111 Toledo Road, ORM, Hyattsville, MD 20782, *pzk3@cdc.gov*; Richard Valliant, University of Michigan; Wenxing Zha, National Institutes of Health

***Key Words:*** cell collapsing criteria, coverage ratio, censoring, bias, mean square error

One of the cell collapsing criteria is a ratio criterion, where the ratio is the poststratification factor or inverse coverage ratio. Quite often, if the ratio for a cell is greater than 2 or less than Ω, then the cell is collapsed with another cell. However, this can introduce bias in a poorly covered group. Censoring (or truncation) approach in collapsing was proposed and implemented in a simulation study. The simulation study showed that the censoring approach is better than the approach mentioned above. In this paper, we propose four new collapsing strategies, two of which is based on the conditional local bias and the others on the conditional local mean square error.

### Stratification for Radio Listening Estimation

❖ Richard Griffiths, Arbitron Inc., 9705 Patuxent Woods Drive, Columbia, MD 21046, *richard.griffiths@arbitron.com*

***Key Words:*** stratification, raking, mean square error, stochastic population controls

To form weights for the sample respondents to its quarterly surveys, Arbitron uses a raking (or, sample balancing) procedure. For years, there have been three marginal variables in this raking procedure. Recently, a fourth variable has been added to the mix. This paper examines issues related to adding the new marginal variable. In particular, we examine different stratification configurations of the new variable and their effect on the statistical properties (e.g., mean square error) of radio listening estimators. Additionally, the population controls for this new variable are based on a survey with relatively small sample sizes. We examine the effect of these stochastic population controls on the precision of the estimators.

### Calculating Effect Sizes Using Survey Data

❖ Jun Liu, Research Triangle Institute, 3040 Cornwallis Road, RTP, NC 27709, *junliu@rti.org*

***Key Words:*** effect size, survey data analysis, clustering, unequal weighting, variance component

Effect Sizes are widely used in comparative studies to measure the impact of an intervention over the control. Although the definition of the Effect Size is well known and generally accepted, its estimation is less so. Furthermore, when data is collected from complex surveys, where clustering and stratification are common, it is not clear how the Effect Size should be calculated. We present a formula for calculating Effect Sizes that takes into consideration of the survey design features and is design consistent. We use simulations to examine its properties against other alternatives.

### Estimation for Two-Phase Panel Surveys

❖ Jason Legg, Iowa State University, Department of Statistics, 204 Snedecor, Ames, IA 50011, *jlegg@iastate.edu*; Wayne Fuller, Iowa State University; Sarah M. Nusser, Iowa State University

***Key Words:*** longitudinal, panel survey, two-phase sampling

Two-phase panel surveys are conducted to study trends over time. The outcome of these studies is often a dataset containing characteristic values and weights for a set of observations. Replication weights are often included in the dataset to allow variance estimation for nonlinear functions. We propose a general linear model as a basis for estimating means and totals at each survey time point. A consistent replication variance estimator is provided as well as a central limit theorem for use in constructing confidence intervals for functions of means or totals.

### Evaluation of Collapsing Criteria in Sample Weighting

❖ Linda Tompkins, National Center for Health Statistics, 3311 Toledo Road, Room 3115, Hyattsville, MD 20782, *lit3@cdc.gov*; Jay J. Kim, National Center for Health Statistics; Wenxing Zha, National Institutes of Health

***Key Words:*** poststratification, weighting, coverage rate, domain estimation

Most surveys combine Asian, Indian, Pacific Islander and multiple-races with White in performing poststratification weight adjustment. In general, small minority groups have lower coverage rates, but are usually pre-collapsed with White. As a result, their weights are redistributed to White, which leads to underestimation of these minority groups. Different collapsing approaches have been proposed to offset the shifting of weights from smaller poststratification cells to larger cells. This study compares estimates from the National Health Interview Survey using the current collapsing method with those from alternative collapsing approaches.

### A Pseudo-Empirical Likelihood Approach for Stratified Samples with Nonresponse

❖ Fang Fang, University of Wisconsin-Madison, Department of Statistics, 1300 University Ave, Madison, WI 53706, *ffang@stat.wisc.edu*; Quan Hong, Eli Lilly and Company; Jun Shao, University of Wisconsin-Madison

***Key Words:*** nonresponse, empirical likelihood, stratified samples

Nonresponse is common in surveys. When the response probability of a survey variable Y is conditionally independent of Y given an observed auxiliary categorical variable Z, a simple method often used in practice is to use Z categories as imputation cells and construct estimators by imputing nonrespondents or re-weighting respondents within each imputation cell. This simple method, however, is inefficient when some Z categories have small sizes. Assuming a parametric model on the conditional probability of Z given Y and a nonparametric model on the distribution of Y, we develop a pseudo empirical likelihood method to provide more efficient survey estimators. Asymptotic distributions for estimators of population means are derived. For variance estimation, we consider a bootstrap procedure and its consistency is established. Some simulation results are provided.

### Estimation Methods To Produce Preliminary Statistics in an Employment Survey

❖ Anders Holmberg, Statistics Sweden, Department of Research and Methodology, Orebro, SE-70189 Sweden, *anders.holmberg@scb.se*; Natalie Jansson, Statistics Sweden; Martin Odencrants, Statistics Sweden

***Key Words:*** survey estimation, improving timeliness, preliminary statistics, PRN

To publish preliminary estimates is one way of satisfying the demand of faster official statistics. The Swedish Short Term Employment Survey is a quarterly survey of businesses published 45 days after the reference quarter. Data is collected monthly and the samples are coordinated over time in a national system of permanent random numbers. In this paper we present estimation methods to produce preliminary statistics that are approximately 20 days faster than the ordinary publication. This increase in timeliness is gained by implicitly modeling the third month of a quarter. We evaluate the quality of the proposed methods and discuss point and variance estimation principles in repeated surveys involving permanent random numbers.

## 280 Hot Topics in Sports

Section on Statistics in Sports, Section on Statistical Education

**Tuesday, August 8, 10:30 am–12:20 pm**

### A Compound Logistic Regression Model for Third Down Efficiency in the NFL

❖ Christopher Rigdon, Southern Illinois University, 5 Stephens Court, Glen Carbon, IL 62034, *crigdon@siue.edu*; Steve Rigdon, Southern Illinois University

***Key Words:*** logistic regression, EM algorithm

A logistic regression model for third down efficiency in the NFL is proposed and studied. The probability of converting a third down is assumed to depend on the ability of the defensive team as well as the ability of the offensive team. An algorithm similar the EM algorithm is used to estimate the parameters in the model.

### Using Spatial Analysis as a Coaching Tool in Ice Hockey

❖ Robin Lock, St. Lawrence University, Department of Math CS Statistics, Canton, NY 13617, *rlock@stlawu.edu*; Travis Gingras, St. Lawrence University; Chris Wells, St. Lawrence University; Michael Schuckers, St. Lawrence University

***Key Words:*** hockey, GIS, spatial, sports

We describe the use of a Geographic Information System (ArcGIS) to enter data and analyze patterns in shot selection and results for collegiate men's ice hockey. Using videos of games during the 2005-6 season at St. Lawrence University, we create a spatially-referenced database of all shot attempts, including variables to describe the game situation, player, shot type, and outcome. Results are displayed over a "map" of the offensive zone of a hockey rink to allow coaches and players to investigate patterns and tendencies that might help improve scoring ability. We consider models for smoothing the data and including covariates for predicting shot success. We also look at a goalie-specific map that gives information on possible areas for improving the goalie's ability to stop opponents from scoring.

### An Interesting Classroom Exercise

❖ Laurence Robinson, Ohio Northern University, Department of Mathematics, 262 Meyer Hall, Ada, OH 45810, *L-Robinson.1@onu.edu*

Prior to the beginning of a first week NFL football game, a commentator made a statement to the effect that winning the first game of the season is critical to a team's chances of making the playoffs. In my introductory statistics courses this statement is examined and evaluated, and some interesting ideas and questions arise.

### Identifying and Evaluating Contrarian Strategies for NCAA Tournament Pools

❖ Jarad Niemi, Duke University, 114 Old Chemistry Building, Box 90251, Durham, NC 27708, *jarad@stat.duke.edu*; Bradley P. Carlin, University of Minnesota; Jonathan Alexander, Duke University

***Key Words:*** basketball, march madness, office pool, point spread, team ratings

The annual NCAA men's basketball tournament inspires many individuals to wager money in office pools that require entrants to predict the outcome of every game prior to the tournament's onset. Coupled with the haphazard team selection behavior of many casual players, office pools' complexity suggests the possible existence of well-informed strategies that are profitable in the long run. In this paper, we seek to identify strategies that are contrarian in the sense that they favor teams that have a high probability of winning, yet are likely to be underbet by our opponents relative to other teams in the pool. Using 2003-2005 data from a medium-sized ongoing Chicago-based office pool, we show that such strategies can outperform the maximum expected score strategy in terms of expected payoff. Additionally, we developed approaches to predict opponent betting behavior using downloadable data.

### Improved College Football Scheduling using a Pseudo-Swiss System

❖ David Annis, Naval Postgraduate School, Operations Research Department, 1411 Cunningham Road, Monterey, CA 93943, *annis@nps.edu*; Samuel Wu, University of Florida

***Key Words:*** football, tournament design, scheduling

Division I-A football remains the only intercollegiate sport which does not determine its champion using a postseason tournament. The large number of teams (119) combined with their short (11-12 games) and disparate schedules result in frequent controversy in choosing the participants in a championship game. We propose a variant of the current schedule which includes some aspects of a Swiss tournament yet preserves the spirit of conference play and traditional rivalries. While the intent of this scheduling revision is to mitigate potential controversies, it also provides additional transparency over the current system. Simulation studies are used to validate the design and to compare its actual (probability of correctly identifying the best team) and perceived (probability of controversy) performance relative to the current system. Preliminary results suggest an improvement.

### Regression Modeling Using Multiple Sources To Detect Group-Level Differences in Pitching Performance

❖ Liam O'Brien, Colby College, Department of Mathematics, 5838 Mayflower Hill, Waterville, ME 04901, *lobrien@colby.edu*; Matthew Aschaffenburg, Colby College

***Key Words:*** covariance modeling, pitching, multiple sources

There are many different statistics used as metrics of pitching performance - both at the individual level and at the team level. We use three metrics as benchmarks of pitching performance for starting pitchers over a five-year period. Of interest is whether there are group level (i.e., league or handedness) differences in how these outcomes vary through time. Difficulty arises in modeling the association amount the metrics within player. To circumvent this, we use a Kronecker product covariance structure to reduce the number of association parameters. This makes estimation of the effects of interest more efficient and detection of league differences through the years easier.

### A Cross-Country Comparison of Efficiency of Soccer-Betting Markets

❖ Ruud H. Koning, University of Groningen, Department of Econometrics, PO Box 800, Groningen, 9700AV The Netherlands, *r.h.koning@rug.nl*

***Key Words:*** rating models, betting odds, efficient market hypothesis

Sports and betting have grown together to become multibillion dollar industries. The advent of the internet has globalized the betting market even more, with a few clicks it is possible to put a wager on events far away. Betting odds are also routinely used as indicators of winning probabilities or indicators of relative strengths. This assumes that betting markets process information efficiently. In this paper, we test this hypothesis for soccer betting markets in eleven different countries, and in markets for different divisions within the same country. In our analysis, we focus in particular on variation of efficiency between different countries.

## 281 Pedigree/Population Analyses

● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Family-Based Liquid Association Analysis on Atherosclerosis in C3H/HeJ.ApoE-/- x C57BL/6J.ApoE-/- Cross

❖ Tun-Hsiang Yang, University of California, Los Angeles, 11140 Rose Ave., #104, Department of Statistics, Los Angeles, CA 90034, *yangth@ucla.edu*; Ker-Chau Li, University of California, Los Angeles; Shinsheng Yuan, University of California, Los Angeles; Aldons J. Lusis, University of California, Los Angeles; Wei Sun, University of California, Los Angeles; Susanna Wang, University of California, Los Angeles

***Key Words:*** liquid association, allele sharing, linkage analysis, ANOVA, atherosclerotic lesions, aortic lesions

The traditional Linkage Analysis methods have been only modestly successful in identifying genes involved in the disease. The use of the Liquid Association method in conjunction with the idea of the Allele Sharing may improve the ability to identify the candidate genes for the complex disease. Here we propose the Family-based Liquid Association method for the disease genes mapping by taking into account gene interaction effects. Our data came from an F2 intercross between C57BL/6J and C3H/HeJ mice on the ApoE-/- background and fed in a "western" diet for the study of advanced atherosclerotic lesions. Information from a high-density genome scan with 1353 SNPs markers on the quantified aortic lesions trait of all 334 F2 mice is employed. We provide an ANOVA point of view to explain why the genes that mediate most other genes are likely to be linked to the trait.

### Evaluating the Impact of Family Structure on Estimating Genetic Association Parameters in Family Studies

❖ Stefan Boehringer, National Cancer Institute, 6120 Executive Blvd., Room 8105 MSC 7244, Rockville, MD 20852, *jsm@s-boehringer.org*; Ruth Pfeiffer, National Cancer Institute

***Key Words:*** statistical genetics, genetic association, penetrance, likelihood, residual correlation, Alzheimer's

Most work on family based association studies between genetic markers and a disease focuses on the problem of testing the null hypothesis of no association. We present a likelihood based approach to estimating association parameters in family-based studies for a true, unobserved disease locus. We show that, assuming the baseline risk of the disease is known, one can estimate the penetrance function under the proper genetic model, the allele frequency of the (unobserved) true disease allele and its linkage disequilibrium with the observed markers. These parameters are not identifiable based on observations on unrelated cases and controls. Extensions can be made to account for additional unlinked disease loci. We evaluate the impact of family structure and ascertainment schemes on the parameter estimates in simulations and apply the method to data on Alzheimer's disease.

### Evaluation of Three Approaches To Correct for Ascertainment of Pedigrees for Random-Effects Cox Proportional Hazard Linkage Analysis

❖ Susan Slager, Mayo Clinic College of Medicine, Division of Biostatistics, 200 1st Street, SW, Rochester, MN 55905, *slager@mayo.edu*; Shannon K. McDonnell, Mayo Clinic College of Medicine; Vernon S. Pankratz, Mayo Clinic College of Medicine; Antje Hoering, Cancer Research and Biostatistics; Terry M. Therneau, Mayo Clinic College of Medicine; Mariza de Andrade, Mayo Clinic College of Medicine

***Key Words:*** genetics, age-at-onset, linkage analysis

Many diseases have variable age of onset. By accounting for this variability in a linkage analysis, one may be able to obtain more statistical power to identify susceptibility loci than had one analyzed the data as a binary trait. Several analytical age-at-onset linkage approaches exist, one of which is the random effects Cox proportional hazard model. This approach is an extension of the Cox proportional hazard regression that includes both fixed and random effects. Recently, Pankratz et al. have found parameter estimates obtained from this random-effects hazard model are biased. This bias may be due to several factors, including pedigree ascertainment. We hypothesize that correcting for ascertainment may improve the parameters estimates. We investigate three approaches to correct for pedigree ascertainment and compare them to the approach that makes no correction.

### A Probability Model for Recent Tb Transmissions in Genetic Clusters

❖ Nong Shang, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, *nms6@cdc.gov*

***Key Words:*** genetic cluster, recent transmission, dependent Bernoulli trial, family history model

DNA fingerprinting of Mycobacterium tuberculosis is used to link tuberculosis into genetic clusters. Researchers have used such clusters as surrogates of recent TB transmissions. Most previous researchers have focused on patient-level characteristics to predict recent transmission. However, it is also important to consider the predictive power of cluster-level characteristics. Most of practical questions at the cluster level can be answered through modeling the probability that a new case is associated with recent transmission. Family history models were applied. Each of the parameters in the family history model is further modeled by the features of the clusters, resulting in two simultaneously estimated regression equations with observations as the sums of dependent Bernoulli random variables. Numerical methods were developed to estimate the parameters and to find the "best" models.

### Human Population Structure Detection via Multilocus Genotype Clustering

❖ Xiaoyi Gao, North Carolina State University, Bioinformatics Research Center, Campus 7566 / 1500 Partners II Building, Raleigh, NC 27695-7566, *xgao4@ncsu.edu*; Bruce S. Weir, University of Washington

***Key Words:*** population structure, SNP, hierarchical clustering, hapmap, Perlegen

We describe a hierarchical clustering algorithm for using Single Nucleotide Polymorphism (SNP) genotype data to assign individuals to populations. Our method does not assume Hardy-Weinberg equilibrium and linkage equilibrium among loci in sample population individuals. We show that the algorithm can assign sample individuals highly accurately to their corresponding ethic groups: CEU, YRI, CHB+JPT in our tests using HapMap SNP data and it is also robust to admixed populations when tested on Perlegen SNP data. Moreover, it can detect fine-scale population structure as subtle as that between Chinese and Japanese by using genome-wide high diversity SNP loci. This is the first successful separation of Chinese and Japanese samples using high diversity SNP loci with strong statistical support.

### Testing for Familial Aggregation When the Population Size Is Known

❖ Yixin Fang, Columbia University, 415 W. 118 Street, Apt. 1, New York, NY 10027, *yf2113@columbia.edu*

***Key Words:*** familial aggregation, random effect, asymptotic efficiency

We treat the problem of testing for familial aggregation when sampling from a population of a known size. We consider the setting where all families in the population containing at least one affected member are obtained. Both the situation where a random sample of control families with no affected individual is obtained and the situation where no such family is obtained or examined. Methods are developed and the relative efficiencies of different designs are considered. The results are compared to the setting where the population size is unknown and a sample of families containing at least one affected member, with or without a sample of control families, is obtained. It is found that knowing the population size has a large impact on relative efficiency, but that obtaining control families has a smaller impact.

### A Study of an Allele-Sharing Statistic on Extended Pedigrees

❖ Saonli Basu, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware Street SE, Minneapolis, MN 55455, *saonli@umn.edu*

***Key Words:*** permutation test, linkage analysis, MCMC, allele-sharing methods

Linkage analysis using allele-sharing methods is a trait-model-free approach for linkage detection between a set of markers and a trait. In linkage studies, the dependence in the joint inheritance of the trait and marker alleles are studied in pedigrees. Unavailability of marker data on individuals makes the computation of possible inheritance vectors infeasible. Markov chain Monte Carlo (MCMC) techniques provide a solution to this problem as they produce in real time a good estimate of the probabilities of inheritance vectors consistent with the marker data. The performance of an allele-sharing statistic in detecting linkage is verified by applying MCMC technique on extended pedigrees. A permutation test within the MCMC framework is proposed to deal with the issue of conservativeness due to missing marker data in the standard normal approximation of the standardized test statistic.

## 282 Disease Modeling: Theory and Applications ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Model Choice in Time Series Studies of Air Pollution and Mortality

❖ Roger Peng, Johns Hopkins University, 615 N. Wolfe Street, E3527, Department of Biostatistics, Baltimore, MD 21205, *rpeng@jhsph.edu*; Francesca Dominici, Johns Hopkins University; Thomas A. Louis, Johns Hopkins University

***Key Words:*** NMMAPS, air pollution, time series, particulate matter

Multi-city time series studies of particulate matter (PM) and mortality and morbidity have provided evidence that daily variation in air pollution levels is associated with daily variation in mortality counts. These findings served as key epidemiological evidence for the recent review of the United States National Ambient Air Quality Standards (NAAQS) for PM. As a result, methodological issues concerning time series analysis of the relation between air pollution and health have attracted the attention of the scientific community and critics have raised concerns about the adequacy of current model formulations. We use a simulation study and analyses of a large multi-city database to characterize model uncertainty in adjusting for seasonal and long-term trends in time series studies of air pollution and mortality.

### A Generalized Threshold Mixed Model for Analyzing Non-Normal Nonlinear Time Series: Plague in Kazakhstan as an Illustration

❖ Noelle I. Samia, The University of Iowa, Department of Statistics and Actuarial Science, 241 Schaeffer Hall, Iowa City, IA 52242, *nsamia@stat.uiowa.edu*; Kung-Sik Chan, The University of Iowa; Nils C. Stenseth, The University of Oslo

***Key Words:*** binomial distribution, delay, epizootic events (plague outbreaks), exponential family, stochastic regression

We introduce the generalized threshold mixed model for piecewise linear stochastic regression with possibly non-normal time series data. We assume the response variables conditionally belong to an exponential family, whose canonical parameters are linked to a piecewise linear stochastic regression function. We study the specific case where the response variable equals zero in the lower regime. Some large-sample properties of a likelihood-based estimation scheme are derived. Our approach is motivated by the need for modeling nonlinearity in serially correlated epizootic events. Data coming from monitoring conducted in a natural plague focus in Kazakhstan are used to illustrate this model by obtaining biologically meaningful conclusions regarding the threshold relationship between prevalence of plague and covariates, including past abundance of great gerbils and climatic variables.

### Statistical Determination of the Length of Quarantine Periods in an Epidemic

❖ Xiaojun You, Johns Hopkins University, E3031 Biostatistics Department, 615 N Wolfe St, Baltimore, MD 21205, *xyou@jhsph.edu*

***Key Words:*** quarantine, SARS, incubation period, truncated density

Quarantine of contacts of known SARS patients proved effective at controlling further transmissions during the 2003 outbreak. However, little work has been done on determining the optimal quarantine length. Our initial work analyzed the impact of using the maximal reported incubation period from an outbreak to guide quarantine policy. To do this, we quantified the probability of an infectious case transmitting disease after such quarantine. Methods were developed using order statistics theory with validation via a simulation study. Later we derived methods to calculate the desired quarantine length based on different criteria using maximum likelihood estimation with a truncated density function. The effectiveness of these latter methods was evaluated using simulation studies. Current work involves application and evaluation of the method using data from an actual in-flight SARS outbreak.

### Bayesian Analysis of the 1918 Influenza Pandemic in Baltimore, MD

❖ Yue Yin, Johns Hopkins University, 615 N. Wolfe Street, Dept of Biostatistics, Baltimore, MD 21205, *yyin@jhsph.edu*

***Key Words:*** influenza pandemic, basic reproductive number, infectious disease modeling, disease transmissibility

Around 30 million people were killed in the influenza pandemic of 1918. Though this pandemic has been thoroughly studied, several features remain unexplained. The change of disease transmissibility is suspected to have caused the multiple temporal waves in incidence, which has important implications for pandemic preparedness. With an emphasis in modeling the disease transmissibility, we analyze the population level time series data of influenza in Baltimore using Bayesian method. We found a time-varying disease transmissibility which peaks in the middle of the epidemic.

### Changes in Infant Mortality by Socioeconomic Status: U.S. Residents, 1995--2000

❖ Jay H. Kim, Centers for Disease Control and Prevention, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, *jhk0@cdc.gov*; Joe Fred Gonzalez, Jr., National Center for Health Statistics; Paul D. Williams, National Center for Health Statistics

***Key Words:*** infant mortality rates, socioeconomic status, health service areas

Infant mortality rates (IMR) have been declining among the general population in the United States for decades. However, the relationship between death rates and socioeconomic status (SES) has been seldom examined, especially in the late 1990's. This paper will explore the relationship between infant mortality rates for sex-race groups by SES of US residents from 1995 to 2000. Health service areas (HSAs) were used as the geographical units of analysis. There are 805 HSAs in the U.S., and each consists of an average of four counties. Mortality data were obtained from NCHS. A SES (scale 1-5) was constructed by principal component analysis applied to 14 variables that were made available from the 2000 U.S. Decennial Census. Several statistical methods will be investigated to compare differences of IMR between sex-race groups and SES over time.

### BMI, Race, and Prostate Cancer

❖ Negasi Beyene, National Center for Health Statistics, P.O. Box 1696, CDC, Hyattsville, MD 20788, *neb9@cdc.gov*

***Key Words:*** cancer, prostate, BMI

Prostate cancer is a major public health problem, which over a lifetime will affect an estimated one in five American. The American Cancer Society estimated that there were about 232,090 new cases of prostate cancer in the United States in the year 2005. About 30,350 men will die of this disease. Prostate cancer is the second leading cause of cancer death in men, exceeded only by lung cancer. While one man in six will get prostate cancer during lifetime, only one man in 32 will die of this disease. The death rate for prostate cancer is going down. More people are diagnosed earlier with this disease than used to be. Different studies conclude differently if BMI and race have any relation to having prostate cancer. Using the National Health Interview Survey of 1997-2003, I am trying to find out if there is any relationship between prostate cancer, BMI, and race.

### Standardized Risk and Description of Results from Multivariable Modeling of a Binary Response

Bong-Rae Kim, University of Florida; ❖ Randy L. Carter, University at Buffalo, Department of Biostatistics, Farber Hall Rm 249, 3435 Main St., Buffalo, NY 14214-3000, *rcarter@buffalo.edu*; P. V. Rao, University of Florida

***Key Words:*** logistic regression, post hoc comparisons, adjusted odds ratios, standardized proportions, lead poisoning, academic problems

Descriptions of significant associations found from a logistic regression analysis typically are based on adjusted odds ratios. Unfortunately, odds ratios provide no information about the prevalence of response. In this paper, we justify and recommend using standardized risks, i.e., standardized probabilities, which do provide information about prevalence, in addition to adjusted odds ratios, for pairwise comparisons of the levels of a significant factor. We illustrate the advantages of generally reporting standardized risk estimates, in the context of assessing the effect of blood lead levels during the preschool years on occurrence of academic problems in kindergarten. Results are more meaningfully interpreted when accompanied by standardized risk estimates.

## 283 Portfolios and Derivations

Business and Economics Statistics Section

**Tuesday, August 8, 10:30 am–12:20 pm**

### Mimicking Portfolios with Conditioning Information

❖ Andrew Siegel, University of Washington, Business School, Box 353200, Seattle, WA 98195-3200, *asiegel@u.washington.edu*

***Key Words:*** finance, portfolio, correlation

Mimicking portfolios have long been useful in asset pricing research. In most empirical applications, the portfolio weights are assumed to be fixed over time, while in theory they may be functions of the economic state. This paper derives and characterizes mimicking portfolios in the presence of predetermined state variables, or conditioning information. The results generalize and integrate multifactor minimum variance efficiency (Fama, 1996) with conditional and unconditional mean variance efficiency (Hansen and Richard (1987), Ferson and Siegel, 2001).

### Population Segmentation for Portfolio Strategies

❖ Timothy H. Lee, 595 Croydon Lane, Alpharetta, GA 30022, *tlee080551@aol.com*; Otto Schwalb,

***Key Words:*** segmentation, homogeneity, separation pattern

A large portfolio is often partitioned into subgroups for assignment of strategies in such a way that homogeneity among individuals within a subgroup is enhanced. Two different cases of homogeneity can be considered in, i) profile characteristics for each individual, ii) functional relationship between profile characteristics and target behavior such as creditworthiness or likelihood of response to some solicitations. Partitioning criteria should be selected depending on the segmentation purpose. In the study, Recursive tree and regression analysis was applied using re-sampling technique. Also included is discussion about issues of score strategies when multiple scoring models are developed on a segmented portfolio and each of the models has different separation pattern from others.

### Synthesis-Efficient Portfolio of Investment Projects

❖ Alex Fedorov, Novosibirsk State Technical University, Lenin str. 35 - 45, Berdsk, 633010 Russia, *New_Okela@mail.ru*; Dmitri Shubin, Novosibirsk State Technical University; Anatoly Naumov, Novosibirsk State Technical University

***Key Words:*** investments, risk, decision theory, portfolio, Monte Carlo approach, Markowitz's bullet

It is considered the finite set of initial investment projects, which are submerged into stochastic environment. Basing on this set it is necessary to build the set of project-constructors and choose the best of them, which will be the efficient portfolio of projects. We solve this problem in the framework of decision theory under indefinite conditions. Our efforts are determined by conceptual scheme of selection the best project-constructor. We calculate activities, absolute and relative risks for each project-constructor. On the criteria plane (risk/profit) we draw the set of points, which characterize obtained project-constructors. We build Markowitz's bullet for project-constructors and determine that basic mechanisms of investing process are still rightful: as higher profit of project, then higher risk of this project; as longer time of life of project, then risk also is higher.

### Assessing the Effects of Variability in Interest Rate Derivative Pricing

❖ Michael Crotty, North Carolina State University, Department of Statistics, 2501 Founders Dr, Raleigh, NC 27695, *mtcrotty@stat.ncsu.edu*; Peter Bloomfield, North Carolina State University

*Key Words:* interest-rates, term structure, derivatives, spline, bootstrap

A commonly traded interest rate derivative is an interest rate cap, which pays an amount equal to a notional amount multiplied by the amount a specified interest rate exceeds the initially agreed upon strike rate at periodic intervals until maturity. This research examines pricing variability in the Hull-White model, which can be used to price interest rate derivatives. First, the variability in the term structure of the zero rate curve, which is modeled with splines, is examined. This spline approach is then bootstrapped to determine the variability of the spline at various maturities. Next, the variability of the zero rate curve is propagated into the pricing model, also with a bootstrap approach. This bootstrap includes the entire method of derivative pricing from the calculation of a zero curve using a set of bond prices to inputting that zero curve into the pricing model.

### Structural Models of Corporate Bond Yields by Nonconvex Penalized Least Squares

❖ Yuejiao Ma, The Pennsylvania State University, 326 Thomas Building, University Park, PA 16802, *yxm131@psu.edu*

*Key Words:* nonconvex penalized least squares, covariance estimation, LASSO, SCAD

This paper analyses multivariate high frequency financial data to exam the determinations of bond yields. Plenty of literatures have been work on this topic by using different economic models and linear regression. A new approach we adopt is to use the class of non-convex penalized least squares proposed by Fan and Li (2001) to estimate and select variables for bond yield model. On the other side, we also study the covariate matrices to gain information of bond yields to the certain bond in later period by using covariate patterns at earlier stage. Estimation of covariance matrices usually involves high dimension dataset and conflicts the positive-definiteness constrains. In this paper, we use a new approach to select covariance based on penalized likelihood for any multivariate distributions.

### Toward Identification of Shocks in State-Space Models: Application to Stochastic Volatility

❖ Nour Meddahi, Montréal University, Department of Economics, CP 6128 succ Centreville, Montréal, H3C 3J7 Canada, *nour.meddahi@umontreal.ca*; Stéphane Gregoir, CREST/INSEE

*Key Words:* state-space models, identification, moments, specification tests, characteristic function, stochastic volatility

State-space models are commonly used due to the development of simulation methods (MCMC, particle filters). Their efficient estimation as well as the computation of non-linear functions of the observables (such as Value-at-Risk in finance) depends crucially on the correct specification of the error terms. On the one hand, simple approaches test for a specific distribution of one error term under a maintained hypothesis for the second one. A misspecified maintained hypothesis may lead to wrong rejections of the null. On the other hand, non-constructive methods test jointly for the specification of the two shocks. In this paper, we use the characteristic function of the observed process to derive moment-based specification tests of the distribution of a shock without specifying the second one. A test for the presence of fat tails in stochastic volatility models illustrates our approach.

### Estimating the Commonality in Volatility-Volume Relationship

❖ Lei Zhang, Syracuse University, The Martin J. Whitman School of Management, 721 University Avenue, Syracuse, NY 13244, *lzhang06@syr.edu*; Raja Velu, Syracuse University; Tze Leung Lai, Stanford University; Haipeng Xing, Columbia University

*Key Words:* market microstructure, mixture distribution hypothesis, hidden Markov models

The stochastic behavior of price changes and trading volume of a stock has been studied using bivariate mixture models, with the unobservable price- relevant information serving as the mixture variable. The behavior of this mixing variable, which is shown to be autocorrelated determines the dynamic behavior of the volatility-volume relationship. While most of the existing studies deal with single stocks individually, we present a model that can account for cross-stock common factors. The models for individual stocks are estimated via Simulated Maximum Likelihood or via Markov Chain Monte Carlo techniques. For the multivariate model, we approach the estimation via Hidden Markov Models. The empirical results based on daily data for thirty stocks will be presented.

## 284 Preclinical Design and Analysis ●

Biopharmaceutical Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Some Results Concerning Multiplicities in Animal Carcinogenicity Studies

❖ Mohammad Rahman, U.S. Food and Drug Administration, 20335 Swallow Point Road, Montgomery Village, MD 20886, *rahmanat@cder.fda.gov*; Karl Lin, U.S. Food and Drug Administration

*Key Words:* carcinogenicity, Peto test, poly-k, multiple testing

The regular rodent carcinogenicity study has concerns regarding its length and multiplicity of tests. Adjustment for multiplicity is usually done using some standard or ad hoc procedure. One of such ad hoc methods is to use ?=0.005/ 0.025 for common/rare tumors (Lin and Rahman) for two year two species study using analysis suggested by Peto et al. The ICH guidance outlines approaches for conducting a short term along with a long-term study. New statistical methods also have been suggested for the analysis of data. It is not clear if the method suggested by Lin and Rahman is still applicable for the newly proposed design and analysis methods. We will compare the Poly-k method and method suggested by Peto et al. in analyzing data from two long term studies and a short term transgenic mouse along with a long term study. New procedures for the adjustment of multiple testing may be proposed.

### Predictive Fingerprint Modeling in Systems Biology

❖ Lei Zhu, GlaxoSmithKline, 5 Moore Drive, MAI T205, Research Triangle Park, NC 27709, *lei.a.zhu@gsk.com*; Daniel Parks, GlaxoSmithKline; Xiwu Lin, GlaxoSmithKline; Kwan Lee, GlaxoSmithKline; Amit Bhattacharyya, GlaxoSmithKIine; Edit Kurali, GlaxoSmithKline; Amber Anderson, GlaxoSmithKline; Dilip Rajagopalan, GlaxoSmithKline

*Key Words:* predictive modeling, validation, partial least squares discriminant analysis, transcriptomics, metabolomics, lipidomics

Recent technological development on omics platforms generates high dimensional data (e.g. blood chemistry, transcripts, lipids, metabolites, etc.). This enables the systems biology approach to understand biological systems' response to drug and disease at various molecular levels. A small set of predictive fingerprints are often of great interest for diagnostic or drug development purposes. We present an externally cross validated Partial Least Square Discriminant Analysis model. The algorithm searched for the smallest set of markers that explained the most variation in the full data; and built a prediction model on the reduced data with selected markers. Due to the prediction model was built in two steps, proper validation was necessary to avoid selection bias. An example will be demonstrated from a systems biology study in the drug discovery process.

### Evaluation of a Novel Metric for Quality Control in an RNA Interference High-Throughput Assay

❖ Xiaohua Zhang, Merck Research Laboratories, 770 Sumneytown Pike, West Point, PA 19486, *xiaohua_zhang@merck.com*; Namjin Chung, Merck Research Laboratories; Amy Espeseth, Merck Research Laboratories; Marc Ferrer, Merck Research Laboratories

***Key Words:*** strictly standardized mean difference, RNA interference, high throughput screening, quality control, coefficient of variability in difference

The application of genome scale RNA interference (RNAi) relies on the development of high quality RNAi high throughput screening (HTS) assays. An important quality control (QC) characteristic in an HTS assay is how well the positive controls, samples, and negative controls can be separated from each other in the assay. Signal-to-noise ratio (S/N) and Z-factor have been adopted to evaluate this QC characteristic in small molecule HTS assays. However, none of them have clear probability meanings and strong statistical bases. A pair of easily interpretable and theoretically based metrics, SSMD and CVD, were proposed. We investigated the use of SSMD to evaluate the quality of RNAi HTS assays and compared it with other measures. Applying this pair of metrics to an in-house RNAi HTS assay demonstrated that SSMD is superior to S/N and Z-factor for QC evaluation.

### On Methods To Utilize HIV-RNA Data Measured by Two PCR Assays

❖ Joshua Chen, Merck & Co., Inc., P.O. Box 4, BL 3 2, West Point, PA 19486, *joshua_chen@merck.com*; Chunpeng Fan, University of Wisconsin-Madison; Jing Zhao, Merck & Co., Inc.

***Key Words:*** HIV-RNA, limit of quantification, clinical trials

HIV-RNA level is measured by Polymerase Chain Reaction (PCR) assays which have limits of reliable quantification (LoQ). Commercially available Amplicor Standard assay has a reliable range of 400-750,000 copies/mL while Ultrasensitive assay has a range of 50-75,000 copies/mL. Values below the lower LoQ are reported as categorical variables such as "< 400 copies/mL" for the Standard assay and "< 50 copies/mL" for the Ultrasensitive assay. Standard assay is usually used as the first tool to measure the HIV-RNA levels; if a value of "< 400 copies/mL" is reported, the sample is re-tested using the Ultrasensitive assay. For the calculation of change in log10 HIV-RNA, the additional data measured by Ultrasensitive assay are usually ignored. In this talk, we introduce a simple imputation approach using the fact that the two assays are highly correlated and compare it to existing methods.

### Consideration of Parallel-Line and Logistic models for Potency estimation

❖ Bill Pikounis, Centocor R&D, Inc., 200 Great Valley Parkway, Maildrop C 4 1, Malvern, PA 19355, *billpikounis@earthlink.net*; Ken Goldberg, Centocor R&D, Inc.

***Key Words:*** bioassay, dose-response

Standards such as the European Pharmacopoeia outline statistical methodology for the study of potency in assays. The 'validity' of the parallel-lines model is key. Alternatives such as the four-parameter logistic model are considered when that validity is questionable. This presentation examines the consequences of fitting these two models for the objective of estimating potency.

### A Linear Model Approach To Detect Outliers in Quantitative PCR

❖ Michael Man, Pfizer Inc., 2800 Plymouth Road, Ann Arbor, MI 481055, *michael.man@pfizer.com*

***Key Words:*** outlier, linear model, RT-PCR, QPCR

Quantitative polymerase chain reaction (QPCR) or reverse transcriptase polymerase chain reaction (RT-PCR) is one of the most used technologies in biological research. It is well-suited for measuring gene expression as mRNA level from a tiny amount of sample. Coupled with other technologies, such as micro-dissection, it is an indispensable tool in biomarker research. In practice, duplicate or triplicate samples generally are used to control and minimize technology variation. How to detect outliers is an exciting challenge that needs be addressed. This paper will describe a linear model approach for detecting outliers and discuss the implication for planning a QPCR study.

### Quantification Limits in HIV Clinical Trials

❖ Michael Nessly, Merck Research Laboratories, BL 1 3, Blue Bell, PA 19422, *nesslym@merck.com*

***Key Words:*** clinical trials, assay limits, polymerase chain reaction, HIV, censored data, HIV RNA

HIV RNA via PCR assay is the primary outcome parameter of many HIV clinical trials. Increased replicate assay variability at the extremes of the dynamic range results in limits of reliable quantification. Within approved labeling, nominal results beyond these limits may not be reported. There is also a detection threshold where no signal greater than a blank control is detected. Qualitative results are available from these thresholds which represent ordering in magnitude of residual viral load. In this presentation, the criteria for the different thresholds are discussed. Post hoc subgroup data explorations to determine if the difference in viral load between the assay thresholds has clinical or prognostic utility are presented. How the additional information may be utilized in future statistical modeling is discussed.

## 285 Sample Size ●

Biopharmaceutical Section
**Tuesday, August 8, 10:30 am–12:20 pm**

### Sample Size Considerations in Noninferiority Trials

❖ Qi Jiang, Amgen Inc., One Amgen Center Drive, 24 2 C, Thousand Oaks, CA 91320, *qjiang@amgen.com*; Steven Snapinn, Amgen Inc.

***Key Words:*** noninferiority, synthesis approach, sample size

When the use of a placebo is unethical, the efficacy of an experimental therapy must be inferred from noninferiority trials. The usual approach involves defining a fixed noninferiority margin as a function of the lower confidence bound of the estimate of the effect of the active control relative to placebo based on a meta analysis of historical trials. An alternative approach, sometimes called a synthesis approach, involves combining the estimate of the effect of the treatment relative to the control based on the noninferiority trial with the estimate of the effect of the control relative to placebo. This presentation will consider sample size considerations for these two approaches, focusing on issues such as the relative sizes of the historical and noninferiority trials, preservation of the control's effect, and the assumed benefit of the treatment relative to the active control.

### Determining Sample Size for Proportions When the Information on a Prognostic Factor Is Unavailable

❖ Bong S. Kim, Neumann College, One Neumann Drive, Aston, PA 19014, *kimb@neumann.edu*

***Key Words:*** sample size, power, proportions, prognostic factor

Sample size and power for proportions are presented when an important prognostic factor is not observed. We first find the conditional variance of the number of patients with the prognostic factor in the control and experimental groups, and then we find the unconditional variance by integrating over the distribution of the prognostic factor. Thus the variance of the test statistic is larger than it would have been had the prognostic factor been observed, and a large sample size will be required. We show how the sample size and the power vary according to the distribution of the prognostic factor. We illustrate this methodology with a sample size calculation for a clinical trial in which the information on a prognostic factor is unavailable.

### Sample Size Re-estimation in a Dose-Response Study

❖ Kazuhiko Kuribayashi, Pfizer Inc., 3 22 7 Yoyogi Shibuyaku, Tokyo, 151-8589 Japan, *kazuhiko.kuribayashi@pfizer.com*; Mike D. Smith, Pfizer Inc.

***Key Words:*** adaptive design, treatment effect estimation, type I error

The sample size determination is an important element in the design of a clinical trial. Re-estimation of the sample size based on an observed treatment effect during a trial can ensure an adequate power and potentially save a large amount of time and resources. The sample size re-estimation, however, prones to inflate the type I error and to exaggerate the magnitude of the treatment effect. Whereas some methodologies to control the type I error at the nominal level have been developed, issues of treatment effect estimation have not been fully resolved at the moment. In a dose-response study, it is important to estimate the treatment effect accurately. We consider a weighting approach to reduce the bias of treatment effect estimation while controlling the type I error at the nominal level in a dose-response study.

### Sample-Size Estimation for Repeated Measures Analysis in Randomized Clinical Trials with Missing Data

Kaifeng Lu, Merck & Co., Inc.; Xiaohui Luo, Merck & Co., Inc.; ❖ Pei-Yun Chen, Merck & Co., Inc., 126 E. Lincoln Ave, Merck Research Laboratories, Rahway, NJ 07065, *pei-yun_chen@merck.com*

***Key Words:*** sample size, missing data, repeated measures

In designing longitudinal studies, researchers must determine the number of subjects to randomize based on the power to detect a clinically meaningful treatment difference and a proposed analysis plan. In this talk, we present formulas for sample size estimation and assessment of statistical power for a two-treatment repeated measures design allowing for subject attrition. These formulas can be used for comparing two treatment groups across time in terms of linear contrasts. Subjects are assumed to drop out of study at random so that the missing data do not alter the parameters of interest.

### Sample-Size Calculations in Logistic Regression: Commnets on Hsieh, Bloch, and Larsen's Method

❖ Mohammed K. Alam, University of Cincinnati, CTR for Genome Information, 3223 Eden Ave Ml 0056, Cincinnati, OH 45267, *mohammed.alam@uc.edu*; Marepalli Rao, University of Cincinnati; Rupa Mitra, Minnesota State University

Hsieh, Bloch, and Larsen (1998) have proposed a novel method of calculating sample size in the environment of Logistic regression. Suppose Y is a binary random variable with single covariate X, which has a standard normal distribution, with gamma1 the slope in the model. Let MU1 and MU2 be the expected values of X given Y=1 and Y=0 respectively. It can be shown that gamma1 is zero iff the conditional distributions are standard normal. Thus to test the null hypothesis that gamma1 is zero is equivalent to testing MU1 is equal to MU2. Hsieh, Bloch, and Larsen used two sample methodology to calculate sample size for given level and power for testing the null hypothesis that gamma1 is zero against the alternative that gamma1 is a specified number. In their calculations, they made a number of assumptions. In this presentation, the focus is on the assumptions they have made.

### Extracting Information from an Ongoing Blinded Trial

❖ Jitendra Ganju, BiostatWorks, 45 Cleary Court, No 7, San Francisco, CA 94109, *jganju@biostatworks.com*

***Key Words:*** missing data, difference in proportions, relative risk, variance re-estimation

Data analysis for a blinded trial usually begins once the database has been unblinded. However, the analysis can actually begin while the data are still blinded. To effectively extract information, the statistician needs to cultivate an "analysis mindset" for blinded trials. Methods will be described to analyze blinded data to check the assumptions on which the sample size calculations were based.

## 286 Clinical Trial and Microarray Design ●

Biometrics Section, ENAR

**Tuesday, August 8, 10:30 am–12:20 pm**

### Adaptive Design on Adjusting Sample Size without Inflating Type I Error

❖ Jialu Zhang, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Bldg 22 Room 4242, Silver Spring, MD 20993-0002, *Zhangjia@cder.fda.gov*; John Lawrence, U.S. Food and Drug Administration

***Key Words:*** sample size adjustment, type I error, conditional power

Chen, DeMets and Lan (2004) showed that in a sequential trial, with the conditional power greater than 50% under the current trend, increasing sample size at the interim analysis will not inflate type I error. The idea is appealing because increasing sample size under the condition specified in the paper, regular test statistics with no adjustment can still be applied and the type I error will not be inflated. We generalized this idea from testing null hypothesis ? = 0 to ? = ?0. Valid confidence interval and point estimate of ? will be constructed. Further, the method proposed by Cui, Wang and Hung will be compared with this approach.

## Designs for Phase I Clinical Trials with Multiple Courses of Subjects at Different Doses

❖ Shenghua Fan, California State University, East Bay, Department of Statistics, California State University East Bay, Hayward, CA 94542, *kelly.fan@csueastbay.edu*; You-Gan Wang, CSIRO Mathematical and Information Sciences

***Key Words:*** phase I trials, multiple courses, isotonic estimates

Existing phase I designs assign each subject to one dose level that depends on responses from previous subjects. Yet some of the subjects with neither toxicity nor efficacy responses might be treated at higher dose levels. In addition, for some trials, it might be possible to obtain multiple responses (repeated measures) from a subject at different dose levels. A non-parametric estimation method for such studies will be presented in this talk. We also explore how designs of multiple courses of subjects may be implemented to improve design efficiency. The gain of efficiency from ``single dose per subject” to ``multiple doses per subject” are evaluated for several scenarios and for different drop-out rates. Our numerical study show that using ``multiple doses per subject” and the new estimation method increases the efficiency substantially.

## Bayesian Decision-Theoretic Adaptive Designs for Clinical Trials

❖ Yi Cheng, Indiana University South Bend, 1700 Mishawaka Ave., South Bend, IN 46530, *ycheng@iusb.edu*

***Key Words:*** clinical trial, group sequential methods, adaptive design, Bayesian analysis, utility functions, conditional power

In this talk, I propose a Bayesian decision theoretical approach by assessing utilities on group sequentially collected data. The loss function consists of the penalties of making wrong decisions and the cost of enroll patients in the trial. We incorporate estimated conditional power or predicted power into the decision procedure to accommodate desire for good performance in terms of frequentist characteristics regarding the rates of type I and type II errors. The performance of the proposed design is evaluated by numerical studies and by comparison to existing group sequential methods.

## Use of Hadamard Matrices in Designs for Two-Color Factorial Microarray Experiments

❖ Yu Ding, Temple University, 2002B South John Russell Circle, Elkins Park, PA 19027, *yur@temple.edu*; Damaraju Raghavarao, Temple University

***Key Words:*** Hadamard matrix, experimental design in microarray, factorial experiments,

In this presentation, we will demonstrate a method of constructing a factorial experiments with m factors at two levels each, balancing the levels of the factors on conditions and getting a diagonal dispersion matrix for estimated effects. It is interesting to focus on the construction of orthogonal factorial experiments in general.

## Minimax Estimation of Means in Large-Scale Experiments

❖ Tiejun Tong, Yale University, 170 Park Street, Apt. 3, New Haven, CT 06511, *tiejun.tong@yale.edu*; Liang Chen, Yale University; Hongyu Zhao, Yale University

***Key Words:*** minimax estimator, multiple test, false discovery rate, shrinkage

The development of microarray technologies has revolutionized biomedical research, and microarrays have become a standard tool in biological studies. Due to the cost and/or experimental difficulties, it is common that thousands of genes are measured only with a small number of replicates. In particular, the gene-specific estimators of means and variances are not reliable and the corresponding tests have low power. Some approaches to improving variance estimation have emerged. However, the impact of the gene-specific means to tests has not caused much attention. In this paper we propose some shrinkage estimators for means. We prove that our proposed estimators are minimax under Stein loss function. We further investigate the performance of the proposed estimators on improving the tests. A real data study is also used for illustration.

## Tradeoffs between Power and ‘Confounding’ in a Small Microarray Experiment

❖ Mary Putt, University of Pennsylvania, 621 Blockley Hall, 423 Guardian drive, Wayne, PA 19104, *mputt@cceb.upenn.edu*; Thomas Cappola, University of Pennsylvania

***Key Words:*** microarrays, study design, gene expression, confounding, cardiology

In a heart rejection study, microarrays were prepared from 7 cases during rejection and post-rejection, and from 7 controls without rejection. Initial SAM analyses of within-subject data (during versus post-rejection) showed little differential expression while between-subject comparisons (during-rejection versus controls) pointed to a number of differentially expressed genes. Of interest; do genes identified from between-subject comparisons reflect ‘confounding’ or are they directly involved with rejection? We propose a permutation test that combines the within- and between- subject comparisons to better identify genes directly associated with rejection. Power analyses suggest the study design may be useful for detecting differentially expressed genes when within-subject differences are somewhat smaller than those between-subjects, and when the level of confounding is small or moderate.

## Likelihood Inference for Survival Analysis in Two-Stage Randomization Designs

❖ Abdus Wahed, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, 15217, *wahed@pitt.edu*

***Key Words:*** survival analysis, likelihood estimation, two-stage designs, clinical trials

Two-stage randomization designs employ randomization in two stages. Subjects in the first stage are randomized to one of several primary treatments and then depending on the intermediate outcome of the exposure to the primary treatment, are randomized to one of several maintenance treatments in the second stage. We consider such a two-stage randomization design where the goal is to compare the mean survival time under various treatment combinations. Previous work includes IPW estimation, semi-parametric efficient estimation and PH models. The estimators proposed in these methods, especially in the first two are complicated. Considering the popularity of likelihood methods, we propose maximum likelihood estimators and compare them with the available estimators through simulation. The proposed estimators are easy to apply and have efficiency comparable to the existing estimators.

# 287 Statistical Methods in Assays and Cell Line Experiments ●

Biometrics Section, ENAR
**Tuesday, August 8, 10:30 am–12:20 pm**

## Evaluation Parallelism Testing Methods in Immunoassay

❖ Lanju Zhang, MedImmune, Inc., One MedImmune Way, Gaitherburg, MD 20878, *zhangla@medimmune.com*; Harry Yang, MedImmune, Inc.; Iksung Cho, MedImmune, Inc.

***Key Words:*** bioassay, dilution, parallelism

The potency of a biological product is often measured through immunoassays such as ELISA (enzyme-linked immunosorbent assay). Due to large amount of variability inherent to bioassays, the potency is normally determined relative to a known standard. It measures how much the test preparation is "diluted" or "concentrated" relative to the standard. A fundamental assumption in this type of assays for determining relative potency is that the test preparation behaves as a dilution of the standard preparation. In clinical testing literature, this is referred to as parallelism. While several model-based parallelism testing methods were available, there is little work done in evaluating the appropriateness of using these methods. In this paper, both through theoretical derivation and simulation, we establish results that can be used to guide the selection of parallelism testing methods.

## Censored Poisson Regression with Normal Random Effects with an Application to a Dilution Assay

❖ Jorge Quiroz, Wyeth Research, 401 N. Middletown Road, Pearl River, NY 10965, *quirozj@wyeth.com*; Jeffrey R. Wilson, Arizona State University

***Key Words:*** BLUP, censored count data, REML estimates

It is customary to find censored data in the analysis pertaining to health-related data. Modeling these types of data while ignoring the censoring often results in biased parameter estimates. In addition to censoring, correlated observations also may be present. In this presentation, we propose a censored Poisson regression with normal random effects suitable to adjust for censoring and overdispersion in count data. Data from a dilution assay with count response were analyzed using a censored Poisson regression model with normal random effects. We also conducted a simulation study to compare the performance of a censored Poisson regression model with normal random effects to other approaches.

## LD50/IC50 Estimation for in vitro Cytotoxicity

❖ Wenquan Wang, The University of Alabama at Birmingham, 1802 6th Ave., S., NP 2540, Birmingham, AL 35294, *wwang@uab.edu*

***Key Words:*** LD50, IC50, in vitro cytotoxicity, Sigmoidal curve, NLIN

LD50 or IC50 is an important measurement of anti-tumor effect for anticancer agents. In vitro cytotoxicity is another way of measuring this besides animal study. To evaluate the effect of a certain anticancer agent on cancer cell viability, ATPLite luminescence-based assay (Packard Instruments, Meriden, CT) is used to measure cellular ATP levels in this study. Sigmoidal curves (cell viability vs. dose) are fitted in SAS using NLIN Procedure. Certain adjustments need to be made in this procedure when the estimated highest or lowest cell viability is not 100% or 0% respectively.

## Application of Pharmacokinetic Modeling To Guide Clinical Sample Testing

❖ Andrew Chen, MedImmune, Inc., One MedImmune Way, Gaithersburg, MD 20878, *chenxi@medimmune.com*; Harry Yang, MedImmune, Inc.; Lanju Zhang, MedImmune, Inc.; Iksung Cho, MedImmune, Inc.; Ed O'Connor, MedImmune, Inc.

***Key Words:*** pharmacokinetic modeling, dilution, clinical testing

In clinical testing, it is a common practice to perform an initial dilution of a test sample to increase the probability that the final test result will be in a quantifiable range of the assay. However, choosing an optimal starting dilution is not always straightforward. The absence of proper methods guiding the starting dilution often results in repeated testing of the same sample, and increases the overall cost of testing. In this paper, we introduce a multiple dosing pharmacokinetic model that can be used to optimize the choice of initial dilution. The model parameters were estimated based on a single-dose study. This model allows us to customize the initial dilution scheme for each individual sample. A simulation study showed the model effectively reduced the amount of repeated testing.

## Error Variance Estimation with Few Replicates in Cell Line Experiments

❖ David Jarjoura, The Ohio State University, 320 W. 10th Ave., Columbus, OH 43210, *jarjoura.2@osu.edu*; Soledad Fernandez, The Ohio State University; Keding Hua, The Ohio State University

***Key Words:*** smoothing error variance, resampling residuals

Precisely controlled cell line experiments usually use very few replicates. Coefficients of variation are typically under 30% and effects are large justifying the use of just 3 replicates. The small number of replicates leaves few options for the statistician for calculating p-values. In a project, several distinct experiments using the same cell lines are run independently. Although each experiment asks a different question, the measures will typically be the same. We propose taking advantage of the several experiments to "borrow" information from each other when determining error variances. The approaches we propose are similar to smoothing error variances in micro-array studies. Cell line experiments are distinct from micro-array studies, because of the smaller number of replicates, well-defined hypotheses, and fewer distinct experiments for smoothing variances.

## Theoretical Relationship between Direct and Indirect Potency Assays for Biological Product of Live Virus

❖ Harry Yang, MedImmune, Inc., One MedImmune Way, Gaithersburg, MD 20878, *yangh@medimmune.com*; Iksung Cho, MedImmune, Inc.

***Key Words:*** plaque assay, 50% tissue culture infectivity dose (TCID50), titer, serial dilution, Poisson modeling

The potency of a biological product of live virus can be determined using either a direct plaque assay or an indirect assay such as 50% tissue culture infectivity dose (TCID50). The former estimates number of infecting particles, while the latter measures the reciprocal of the dilution of virus resulting in infection of 50% of host cells. In this paper, the theoretical relationship between the plaque and TCID50 assays was investigated, through mathematical derivations. The method assumes that the number of infecting particles in a test sample follows a Poisson distribution, and that each unit of the plaque assay titer corresponds to one infecting particle. Under these assumptions, it is shown 1) for the same test sample, the two assays give rise to different titer values; 2) on log10 scale, the difference in titer varies between 0.23 and 0.27 with an average difference being approxima

## 288 Analysis of Imaging and Spatial Data

Biometrics Section, ENAR
**Tuesday, August 8, 10:30 am–12:20 pm**

### Applications of Copulas To Improve Covariance Estimation for PLS

❖ Gina D'Angelo, University of Pittsburgh, Department of Biostatistics, 130 DeSoto Street, 303 Parran Hall, Pittsburgh, PA 15216, *gmdst17@pitt.edu*; Lisa Weissfeld, University of Pittsburgh; Scott Ziolko, University of Pittsburgh; Chester Mathis, University of Pittsburgh; William Klunk, University of Pittsburgh; Steven DeKosky, University of Pittsburgh; Julie Price, University of Pittsburgh

***Key Words:*** covariance, PLS, brain imaging, PET imaging, Alzheimer's disease

Dimension reduction techniques such as partial least squares (PLS) are currently being applied to classification problems in the area of genetics. These methods have also been applied to PET imaging data with the goal of creating summary measures and examining relationships between voxel-level data and covariates of interest. Previously, we have examined the use of standard PLS techniques for the analysis of amyloid deposition in AD and control subjects using PIB PET imaging techniques (Ziolko et al., 2005). The present work extends PLS to accommodate the unique correlation structure of this data set, for which the distribution of PIB voxel intensity values is a mixture of normal distributions while that of FDG PET is a single normal distribution. This extension is implemented by using a copula to estimate the covariance structure and illustrated in the PIB/FDG data set.

### Establishing the Utility of MRI Parameters in Predicting Disease Activity in RRMS Patients Using GMDM

❖ Welfredo R. Patungan, University of the Philippines, Diliman, Quezon City, Metro Manila, 1101 Philippines, *welfredo.patungan@up.edu.ph*; Daniel Bonzo, Serono, Inc.

***Key Words:*** MRI, RRMS, hidden Markov chain, latent structural model, GMDM

Magnetic Resonance Imaging (MRI) parameters have long been considered as important secondary efficacy parameters in evaluating disease progression in subjects with relapsing remitting multiple sclerosis (RRMS). However, their predictive value on clinical endpoints such as qualifying relapse rate and time to first exacerbation has not been fully explained. Various models such as Hidden Markov Chains and Latent Structural Models have been proposed in the past to explain their predictive value. In this paper, we propose using a longitudinal generalized linear model based on the so-called general mixed-data model (GMDM) to explain the predictive link between MRI and clinical parameters. We will establish this using data from a RRMS patient database. The GMDM maximally exploits the existing data structure and hence is considered a better model in explaining complex interrelationships.

### Application of Randomized Singular Value Decomposition Techniques to the Analysis of Imaging Data Using Partial Least Squares

❖ Scott Ziolko, University of Pittsburgh, UPMC Presbyterian, PET Facility, 200 Lothrop Street, Department of Radiology, Pittsburgh, PA 15213, *ziolkosk@upmc.edu*; Lisa Weissfeld, University of Pittsburgh; Chester Mathis, University of Pittsburgh; William Klunk, University of Pittsburgh; Steven DeKosky, University of Pittsburgh; Julie Price, University of Pittsburgh

***Key Words:*** covariance, PLS, brain imaging, PET imaging, Alzheimer's disease

The application of PLS to PET imaging data requires the computation of the singular value decomposition (SVD) of the covariance matrix of the image data with the covariate data. This approach is computationally intensive in simpler settings where a single modality, such as PET, is used and no longer feasible when using an additional imaging modality in place of the covariate data. Building on previous work (Ziolko et al., 2005), we examine the application of randomized SVD techniques (Drinea, Drineas, Huggins, 2005) to PLS. We compare the use of weighted selection techniques, random sampling techniques, and deletion of voxels with small t-values to reduce the dimensionality of the problem. We focus on the results obtained from PLS using these different approaches.

### A Comparison of Different Methods for Identifying Outliers in MRS Data

❖ Sandra B. Hall, The University of Kansas Medical Center, 3901 Rainbow Blvd., MS 1008, Kansas City, KS 66160, *shall@kumc.edu*; Mihai Popescu, The University of Kansas Medical Center; Anda Popescu, The University of Kansas Medical Center; Niaman Nazir, The University of Kansas Medical Center; Thomas Malone, The University of Kansas Medical Center; Robin Aupperle, The University of Kansas Medical Center; Allan Schmitt, The University of Kansas Medical Center; JoAnn Lierman, The University of Kansas Medical Center; William M. Brooks, The University of Kansas Medical Center

***Key Words:*** brain imaging, MRS, outliers

Magnetic Resonance Spectroscopic Imaging provides a non-invasive quantification of metabolite concentrations throughout the brain. This multi-voxel sampling method, however, yields some voxels with non-representative outlying data that should be identified and excluded since they might bias the average concentrations. On the other hand, if we classify too many data points as outliers then we will under estimate the standard deviation for the concentration of each metabolite. Further, if we fail to accurately identify actual outliers this will result in our overestimating the standard deviation. Both of these problems will lead to an inaccurate representation of the concentrations. Using a dataset from two groups of individuals, one with traumatic brain injury and another without injury we will compare different methods that can be used in order to classify data as outliers.

### A Central Limit Theorem for High-Dimensional Spatially Correlated Processes

❖ Danielle Harvey, University of California, Davis, 1442 Gravink Court, Woodland, CA 95776, *djharvey@ucdavis.edu*; Qian Weng, University of California, Davis; Evan Fletcher, University of California, Davis; Charles DeCarli, University of California, Davis; Laurel Beckett, University of California, Davis

***Key Words:*** central limit theorem, spatial data, MRI, orthogonal decomposition

Alzheimer's disease and other forms of dementia affect the underlying structure of the brain. Magnetic Resonance Imaging (MRI) enables researchers to observe tissue atrophy, strokes, and white matter damage in the brain. These images may be broken up into tiny volumes, or voxels, of information, of which there are hundreds of thousands in a single image. Most analyses have reduced these data to a one-number summary, such as a volume. However, such a summary may be too limited to capture adequately, the ways in which images vary. Alternative summaries may include decompositions of the voxel data. We present a general central limit theorem for decompositions of correlated spatial data, followed by simulations of a spatial decomposition of the data. We then apply the decomposition to actual MRI data from two groups of subjects to illustrate differences in location and extent of disease.

## Image Analysis by Spatial Point Process Modeling in Irregular Area

❖ Weimin Zhang, Texas A&M University, 6602 Harbor Town Drive, Apt. 201, Houston, TX 77036, *zhangwm@stat.tamu.edu*; Suojin Wang, Texas A&M University

***Key Words:*** area-interaction process, maximum pseudolikelihood estimate, Markov chain Monte Carlo, Monte Carlo maximum likelihood estimation, simulation study

Spatial point pattern analysis in an irregular area with heavy edge effects is generally computationally expensive due to heavy dependence on Monte Carlo simulations. In this paper, an aggregated point pattern model, the area-interaction process model, was applied to a lattice image data analysis application in an irregular area. The spatial distributions are examined by fitting area-interaction processes to the data. The Monte Carlo maximum likelihood estimation is considered. The statistical properties of the estimation procedure are investigated by a group of simulation studies. Despite the complexity of the formulation, the simulation research shows that the estimator is consistent. We compared this with the maximum pseudo-likelihood estimation.

## Identifying Risk Factors for Encephalitis from West Nile Virus Infection Using Scan Statistics

❖ Sarah Baraniuk, The University of Texas School of Public Health, 1200 Herman Pressler, Suite E835, Houston, TX 77030, *Mary.Baraniuk@uth.tmc.edu*; Kristy Lillibridge, The University of Texas School of Public Health

***Key Words:*** scan statistics, logistic regression, West Nile virus

While logistic regression provides reasonable methods for identifying risk factors, scan statistics allow us to identify possible temporal or geographical risk factors. The motivation for this work stems from a retrospective medical chart review completed on 172 confirmed West Nile virus (WNV) cases that were hospitalized in Houston between 2002 and 2004. The risk factors for developing encephalitis from WNV have not been clearly identified (with the exception of advanced age). It was found that of the 172 cases 113 were found to have encephalitis, 47 had meningitis and 12 were uncomplicated fever cases. A multiple logistic regression model identified various risk factors and scan statistics were used to assess the statistical significance of the clusters of WNV cases recorded.

# 289 Robust Methods

Section on Statistical Computing

**Tuesday, August 8, 10:30 am–12:20 pm**

## On Weighted Least Squares for Missing Data

❖ Sergey Tarima, Medical College of Wisconsin, 817 Watertown Plank Road, Office H2460, Milwaukee, WI 53226, *starima@mcw.edu*

***Key Words:*** parameter estimation, weighted least squares, missing data, hierarchical structures

A robast form of weighted least squares for missing data is considered. This form uses a hierarchical structure for estimating a multivariate vector. At i^th step of this procedure estimates of parameters for non-missing components of the vector is based on combining information in the subset of observations with the non-missing components with updated estimates of the location parameters from all subsets with even more missing components in an iterative fashion. This approach was applied to several MCAR and MAR simulation examples. One of them is median difference estimation for a bivariate exponential case, which shows robustness of the approach to some deviations from normality in the presence of MAR missing data.

## Testing Equality of Covariance Matrices When Data Are Incomplete

❖ Mortaza Jamshidian, California State University, Fullerton, Department of Mathematics, 800 N State College Blvd, Fullerton, CA 92834, *mori@fullerton.edu*; James Schott, University of Central Florida

***Key Words:*** likelihood ratio test, missing completely at random, robust tests, test of homogeneity of covariance matrices, Wald test, missing data

In this talk we propose statistical tests for equality of covariance matrices when data are missing, a problem that seems not have received attention in a general setting. A Wald test and a likelihood ratio test (LRT), based on the assumption of normal populations are developed. As in the complete data case, these tests perform poorly for non-normal data. This has led us to construct a robust Wald test for our problem. A simulation study is carried out to assess the performance of the proposed tests in terms of their observed significance level and the power. It is found that one of our tests performs particularly well for both normal and non-normal data in both small and large samples. In addition to their usual applications, we have discussed the application of the proposed tests in testing whether a set of data are missing completely at random.

## Robust Diagnostics for Multivariate Mixed Continuous and Categorical Data

❖ Tsung-Chi Cheng, National Chengchi University, 64 Section 2 ZhiNan Road, Departemnt of Statistics, Taiepi, 11605 Taiwan, *chengt@nccu.edu.tw*; Atanu Biswas, Indian Statistical Institute

***Key Words:*** maximum trimmed likelihood estimator, multiple outliers, robust Mahalanobis distance, robust diagnostics

In this article, we apply the maximum trimmed likelihood (MTL) approach (Hadi and Luceno 1997) to obtain the robust estimators of multivariate location and shape, especially for data mixed with continuous and categorical variables. The forward search algorithm (Atkinson 1994) is adapted to compute the proposed MTL estimates. A simulation study shows the proposed estimator outperforms the classical maximum likelihood estimator when outliers exist in data. Real datasets also are used to illustrate the method and results of the detection of the outliers.

### A Markov Chain Monte Carlo Approach for Finding the Minimum Volume Ellipsoid

❖ J. Brian Gray, The University of Alabama, Department of ISM, PO Box 870226, Tuscaloosa, AL 35487-0226, *bgray@cba.ua.edu*; Eric B. Howington, Coastal Carolina University

***Key Words:*** data mining, outlier, minimum covariance determinant, robust diagnostics, robust covariance estimation

The minimum volume ellipsoid (MVE) is a useful construct for computing robust multivariate outlier diagnostics and robust covariance matrix estimates. Exact computation of the MVE is impractical for all but the smallest data sets. A variety of approaches have been developed for approximating the MVE. In this presentation, we introduce a new MVE approximation method based on a Markov Chain Monte Carlo approach. This method can also be used to find the minimum covariance determinant (MCD) estimator. The proposed method is compared to existing techniques using real and simulated data.

### Clustering of Outlier Structure Using Minimal Spanning Tree Rankings and Minimum Volume Ellipsoids

❖ Mark W. Lukens, George Mason University, 4061 Cardinal Crest Drive, Woodbridge, VA 22193, *mark.lukens@comcast.net*; James Gentle, George Mason University

***Key Words:*** clustering, outliers, minimal spanning tree, minimum volume ellipsoids, multivariate ranking

One method to explore high dimensional data structure is through the use of minimal spanning trees. Multivariate rankings might be used in conjunction with a minimal spanning tree as a way of ordering the data. The edge distance between rankings is one way to separate the tree into clusters. This technique, however, depends upon how the ranking is done. Use of rankings from minimal spanning trees might encounter difficulties when multiple groups appear in the data. A different approach is to use minimum volume ellipsoids and a minimal spanning tree to rank the multivariate data. The minimum volume ellipsoids define the main structure and then the minimal spanning tree is fitted to the remainder of the outlier data. This not only provides a boundary between the main structure and any outlier groupings but can also assist in the identification of clusters.

### Spanning Trees as Data Analysis Tools

❖ Adam Petrie, Rensselaer Polytechnic Institute, 149 Corinth Street, North Adams, MA 01247, *petria@rpi.edu*; Thomas R. Willemain, Rensselaer Polytechnic Institute

***Key Words:*** tests for normality and stability, resistant estimators of location, outliers, clustering, implicit dimension, spanning trees

Spanning trees are efficient but often overlooked alternative tools for the analysis of large, multivariate metric datasets. They are easy to compute and can reduce high-dimensional correlated, clustered, and arbitrarily shaped data clouds to easily visualizable distributions of arc lengths, node degrees, and topological distances. These distributions reflect the inherent structure and characteristics of the dataset. For example, these properties of the spanning tree can be used to deduce the implicit dimension of a dataset, to differentiate between data generating distributions, to find outlying observations, and to detect the presence of clustering or clumping. We report computational experiments in high dimensions on (1) tests for normality and stable laws (2) resistant estimates of location (3) detection of clustering (4) tests of implicit dimension.

### A Nonparametric Approach to Descriptive Measures of Multivariate Massive Data Based on Convex Hull Peeling Depth

❖ Hyunsook Lee, The Pennsylvania State University, 265 Blue Course Drive, Apt. 23B, State College, PA 16803, *hlee@stat.psu.edu*

***Key Words:*** convex hull peeling, descriptive statistics, nonparametric multivariate analysis, massive data

Visualizing multidimensional data is often biased due to projection on less dimensions. Quantizing shapes of multivariate data is generally limited to the normal distribution assumption. In this presentation, we introduce descriptive statistics to measure the skewness and kurtosis of multidimensional data based on convex hull peeling depth. These convex hull peeling algorithms do not require calculating moments but provide a mapping to one dimensional scale curve, characterizing shapes of multivariate distributions. As diagnostic tools, convex hull peeling algorithms are exemplified with Monte Carlo simulations and data sets from Sloan Digital Sky Survey.

## 290 Analysis of Factorial Experiments ●

Section on Physical and Engineering Sciences, Section on Quality and Productivity

**Tuesday, August 8, 10:30 am–12:20 pm**

### Iterative Stepdown Tests: Analysis of Orthogonal Saturated Factorial Designs

❖ Daniel Voss, Wright State University, Mathematics and Statistics, 3640 Colonel Glenn Hwy, Dayton, OH 45435-0001, *dan.voss@wright.edu*

***Key Words:*** iterative test, adaptive test, effect sparsity, familywise error rate, unreplicated factorial, multiple test

Strong control of the experimentwise error rate has been established for a broad class of iterative stepdown tests for analysis of orthogonal saturated factorial designs. We present this result, an illustration that nonadaptive methods applied iteratively may be implicitly adaptive, and the results of a power study comparing methods covered by the result and some not. Lenth's method and a variation are considered, for example.

### Error Sum of Squares Comparison for Model Search, Identification, and Discrimination

❖ Hongjie Deng, University of California, Riverside, 3434 Kentucky Street, Riverside, CA 92507, *hdeng001@student.ucr.edu*; Subir Ghosh, University of California, Riverside

***Key Words:*** class of models, error sum of squares, fractional factorial plans, interaction effects, linear model

We consider model search, identification, and discrimination between models within a class of models with common and uncommon parameters. We use the error sum of squares (SSEs) as our criterion function. We present general results for comparing SSEs of different models. We illustrate our method and results for comparing models in a factorial experiment involving m factors at two levels. The models considered have the common parameters as the general mean and the main effects and the uncommon parameters as two two-factor interactions, one in each model. Several fractional factorial plans are used.

### Power Comparisons for Model Search, Identification, and Discrimination

❖ Ying Luan, University of California, Riverside, 4997 Brookhill Place, Riverside, CA 92507, *aying78@hotmail.com*; Subir Ghosh, University of California, Riverside

***Key Words:*** equivalence, fractional factorial plans, interaction effects, linear model, permutation matrix, power

This presentation considers the problem of model search, identification and discrimination between models from a specified class of models. The models within this class have some common parameters and one distinct parameter in each model. The power for testing uncommon parameter is used as our criterion function. We present some characterizations for the equivalence of models based on the components of the non-centrality parameter for the distribution of the test statistic. The characterizations are given in terms of a vector h dependent only on the design matrix. As an illustrative example, four fractional factorial plans for a 2^5 factorial experiment are considered. The common parameters are the general mean & the main effects, and the uncommon parameters are two factor interactions one for each model.

### A Comparison of Three Approaches to Modeling a Multivariate Response in a Designed Experiment

❖ Steven LaLonde, Rochester Institute of Technology, 98 Lomb Memorial Drive, Center for Quality and Applied Statistic, Rochester, NY 14623-5604, *smleqa@rit.edu*; Peter Bajorski, Rochester Institute of Technology

***Key Words:*** principal components, multivariate regression, designed experiment, response curve modeling

This paper examines three approaches to modeling a response curve as a function of factors in a designed experiment. The response data is a 21 step density log exposure curve. The experiment is a 2 level full factorial with 5 variables. The first approach is multivariate regression of the original twenty-one density variables. The second approach is a three step procedure: 1)reduce the number of response variables using principal components, 2)model the principal components using regression, and 3)reconstruct original twenty-one density variables from predicted principal components. The third approach is similar to the second, except that the usual principal components method is replaced by a non-negative principal components method described by Bajorski and Hall (2006). The paper addresses such issues as noise reduction, interpretability of solutions, and variable selection.

### Bayesian Analysis of Split Plot Failure Amplification Experiments

❖ Oksoun Yee, Schering-Plough Corporation, 132 Spruce Mill Lane, Scotch Plains, NJ 07076, *oi@utk.edu*; Robert W. Mee, University of Tennessee

***Key Words:*** binary response, generalized linear model, mixed model, model selection, split-unit, optimization

Joseph and Wu (2004) proposed use of an “amplification factor” to increase the information from experiments with a binary response variable. In addition to the amplification factor having a known effect, Joseph and Wu recommended that, for convenience of experimentation, this factor be taken as an easy-to-change, split-unit factor. In such cases, the analysis should take into account the possibility of both whole-unit and split-unit error variation. We present such an analysis here, where the Bayesian approach not only permits proper accounting of the error structure, but also facilitates the subsequent optimization step.

### Deriving Optimal Conditions for Large-Scale Controlled Synthesis of Nanostructures Using Statistical Methods

❖ Tirthankar Dasgupta, Georgia Institute of Technology, 765 Ferst Drive, NW, Atlanta, GA 30332, *tdasgupt@isye.gatech.edu*; Christopher Ma, Georgia Institute of Technology; Roshan J. Venghazhiyil, Georgia Institute of Technology; Zhong L. Wang, Georgia Institute of Technology; C. F. Jeff Wu, Georgia Institute of Technology

***Key Words:*** nanotechnology, experimental design, generalized linear model, robust process

In this paper, an effort is made to systematically investigate the best process conditions that ensures large-scale synthesis of different types of one dimensional cadmium selenide nanostructures. Through a designed experiment and rigorous statistical analysis of experimental data, models linking the probabilities of obtaining specific morphologies to the process variables are developed. A new iterative algorithm for fitting a Multinomial GLM is proposed and used. The optimum process conditions that maximize the above probabilities and, at the same time, make the synthesis process less sensitive to variations of process variables around set values are derived from the fitted models using Monte-Carlo simulations.

### Application of Two-Level Unreplicated Factorial Designs in Agricultural Field Experiments

❖ Tessema Astatkie, Nova Scotia Agricultural College, 61 Coldstream Drive, Truro, NS B2N 5B2 Canada, *tastatkie@nsac.ca*

***Key Words:*** compost, seaweed, Lenth method, organic pasture, normal probability plot of effect estimates

Unreplicated 2^k factorial designs are common in industrial experiments, but not in agricultural field experiments. The main reasons are large experimental errors due to weather- and soil-related factors and the unwillingness of agronomists to consider unreplicated designs. In this study, an unreplicated 2^5 factorial design was used to determine the effect of nitrogen, phosphorus, potassium, compost, and seaweed extract on dry matter yield. The Normal Probability Plot of effect estimates and the Lenth method were used in the first phase of the analysis. In the second phase, ANOVA was completed by either projecting the design to a replicated factorial with fewer factors or reducing the model by moving insignificant high-order interaction effects to the error. The study revealed up to four-factor significant interaction effects, despite the unusually dry weather during the study years.

## 291 Spline-Based Methods ●

Section on Nonparametric Statistics, Biometrics Section

**Tuesday, August 8, 10:30 am–12:20 pm**

### Log-Density Functional ANOVA Model Estimation and Nonparametric Graphical Model Building

❖ Yongho Jeon, University of Wisconsin-Madison, 104 Eagle Heights, Apt. C, Madison, WI 53705, *yjeon@stat.wisc.edu*

***Key Words:*** density estimation, functional ANOVA model, graphical model

The (undirected) graphical model uses graphs to compactly display the conditional dependence among random variables and have become popular, but has mostly been studied in the parametric framework. To

enhance the scope of applicability of the graphical model, we consider the building of nonparametric graphical model through its connection with log-density functional ANOVA model. We propose a new method for fitting the log-density ANOVA model based on a penalized M-estimation formulation with a novel loss function. With the smoothing spline type penalty, our method achieves the optimal rate in nonparametric estimation. With a sparsity-inducing penalty, we obtain a sparse solution in terms of function components, which provides a practical way to construct and estimate the nonparametric graphical model.

## Model Selection in Accelerated Failure Time Models with Nonlinear Covariate Effects

❖ Chenlei Leng, National University of Singapore, 6 Science Drive 2, DSAP, Singapore, 117546 Singapore, *stalc@nus.edu.sg*; Shuangge Ma, University of Washington

***Key Words:*** accelerated failure time, COSSO, Stute's estimator

As an alternative to the Cox model, the accelerated failure time (AFT) model assumes that the event time of interest depends on the covariates through a regression function. We investigate the AFT model with nonparametric covariate effects, when model selection is desirable. Formulated in the framework of smoothing spline ANOVA, the COSSO method with the Stute estimate can achieve a sparse representation of functional decomposition, by utilizing a reproducing kernel Hilbert norm penalty. Computational algorithms and theoretical properties of the method are investigated. The usefulness of the methodology is demonstrated via simulation studies and a real clinical data set.

## Efficient Polynomial Spline Estimation of Partially Linear Models for Clustered Data

❖ Lan Xue, Oregon State University, Kidder 78, Corvallis, OR 97331, *xuel@stat.oregonstate.edu*; Li Wang, Oregon State University; Lijian Yang, Michigan State University

***Key Words:*** asymptotic distribution, knot, longitudinal data, milk protein data, root-n consistency, semiparametric model

We consider estimation of the partially linear models for clustered data using polynomial spline smoothing. The estimation procedure characterizes the infinitely dimensional nonparametric function by a slowly growing number of parameters. Thus the computation is comparable to parametric least squares. On the other hand, it incorporates the within-cluster correlation properly. The resulting estimators are a "polynomial spline version" of both the profile-kernel (PK) estimators (Lin and Carroll 2001) and backfitting estimators (Zeger and Diggle 1994), replacing kernel smoothing by polynomial spline smoothing. They have the same asymptotic property as the PK estimators. A simulated example demonstrates that the proposed estimators are computationally efficient and as accurate as the PK estimators. Application to milk protein content data is described.

## Shape-Restricted Spline Regression

❖ Xiao Wang, University of Maryland Baltimore County, Department of Math and Statistics, 1000 Hilltop Circle, Baltimore, MD 21250, *wangxiao@umbc.edu*

***Key Words:*** isotonic regression, second order cone programming, reversible jump MCMC

The isotonic smoothing spline regression problem is considered. Such estimator is characterized as a nondecreasing natural cubic spline. We give a necessary and sufficient condition for a cubic function nondecreasing over an interval. Estimation of the unknown parameters is formulated into a second-order cone programming problem. The resulting estimated regression function is preserved nondecreasing in the whole domain and also has enough smoothness. Simulation results suggests the method performs well and we illustrate the method by ASA car data. If time permits, we also propose a new Bayesian approach for monotone curve estimation. We treat the number and locations of knots as free parameters and use reversible jump Markov chain Monte Carlo to obtain the posterior samples of knot configurations.

## Smoothing Spline Estimation for Skew-Symmetric Density Function

❖ Sheng-Mao Chang, North Carolina State University, 2300 Avent Ferry Road, Apt. M4, Department of Statistics, Raleigh, NC 27606, *schang2@unity.ncsu.edu*; Hao Zhang, North Carolina State University

***Key Words:*** profile likelihood, emiparametric efficiency, skew-symmetric distribution, smoothing spline

For skew-symmetric distributions, estimating the skew function is as important as estimating location, scale and other parameters. We propose a nonparametric smoothing spline estimator for the skew function, and jointly estimate other parameters efficiently in the context of semiparametric skew-symmetric density functions. Asymptotic properties of our estimators are investigated. Performance of the proposed method is demonstrated via numerical examples.

## Multivariate Time-Dependent Spectral Analysis Using Cholesky

❖ Ming Dai, The University of North Carolina at Charlotte, 9201 University Blvd., Charlotte, NC 28213, *mdai@uncc.edu*; Wensheng Guo, University of Pennsylvania

***Key Words:*** bootstrap, Cholesky decomposition, locally stationary time series, smoothing spline, spectral analysis

In this paper, we propose a nonparametric method to analyze the spectrum of a multivariate locally stationary process. In order to ensure that the final spectral estimate is positive definite while allowing enough flexibility each of its elements, we propose to smooth the Cholesky decomposition of an initial spectral estimate and the final estimate is reconstructed from the smoothed Cholesky elements. The final estimate is smooth in time and frequency, has a global interpretation, and is consistent and positive definite. We show that the Cholesky decomposition of the spectrum can be used as a transfer function to generate a locally stationary time series with the designed spectrum. This not only provides us much flexibility in simulations, but also allows us to construct bootstrap confidence intervals. A numerical example and an application to an EEG data set are used as illusions.

## Bivariate Binomial Spatial Modeling of Loa loa Prevalence in Tropical Africa

❖ Ciprian M. Crainiceanu, Johns Hopkins University, 615 N. Wolfe Street, Department of Biostatistics, Baltimore, MD 21205, *ccrainic@jhsph.edu*; Peter Diggle, University of Lancaster; Barry Rowlingson, University of Lancaster

***Key Words:*** geostatistics, low-rank, thin-plate splines

We present a state-of-the-art application of smoothing for dependent bivariate binomial spatial data to Loa loa prevalence mapping in West Africa. This application is special because it starts with the non-spatial calibration of survey instruments, continues with the spatial model building and assessment and ends with robust, tested software that will be used by the field scientists of the World Health Organization for online prevalence map updating. From a statistical perspective several important methodological issues were addressed: (a) building spatial models that are complex enough to capture the complex structure of the data but remain computationally usable; (b) reducing the computational burden that could be excessive with the our very large covariate data sets; (c) devising methods for comparing spatial prediction methods for a given exceedance policy threshold.

# 292 The Interface of Bayesian and Frequentist Methods

Section on Bayesian Statistical Science

**Tuesday, August 8, 10:30 am–12:20 pm**

## Admissibility and Minimaxity of Generalized Bayes Estimators for Spherically Symmetric Family

❖ Yuzo Maruyama, University of Tokyo, 2875 Osoyoos Crescent, Vancouver, BC V6T 2G3 Canada, *maruyama@csis.u-tokyo.ac.jp*; Akimichi Takemura, University of Tokyo

***Key Words:*** minimaxity, admissibility, regularly varying function, decision theory, harmonic prior, shrinkage estimation

We give a sufficient condition for admissibility of generalized Bayes estimators of the location vector of spherically symmetric distribution under squared error loss. We construct a very useful sequence of smooth proper prior densities approaching the target improper density fast enough for establishing the admissibility based on the method of Blyth. Compared to the known results for the multivariate normal case, our sufficient condition is very tight and is close to being a necessary condition. In particular we establish the admissibility of generalized Bayes estimators with respect to the harmonic prior and priors with slightly heavier tail than the harmonic prior. We also discuss conditions of minimaxity of the generalized Bayes estimator with respect to the harmonic prior.

## Adjusting Frequentist Results for Previous Information by Combining Prior and Objective Posterior Distributions

❖ David Bickel, Pioneer Hi-Bred International, Inc., 7200 NW 62nd Ave., PO Box 184, Johnston, IA 50131-0184, *david.bickel@pioneer.com*

***Key Words:*** objective Bayes, multiple experts, calibration, matching prior distribution, combining information, confidence distribution

While the ability to incorporate prior opinion is an advantage of subjective Bayesian analysis, prior information can be used instead with a p-value or confidence interval by combining the confidence distribution from the observed data with one or more confidence distributions that encode the prior information. Such combination is based on the objective posterior interpretation of the confidence distribution (Schweder and Hjort 2002) and the methodology for combining confidence distributions from independent data (Singh, Xie, and Strawderman 2005). A prior distribution elicited from an expert qualifies as a confidence distribution if either the expert is well-calibrated in the sense of generating correct coverage probabilities or the expert is assumed to have been randomly drawn from a space of experts such that the resulting coverage probabilities have the correct frequencies.

## Relationships between Frequentist and Bayesian Prediction Limits of the Poisson Process: Noninformative Priors

❖ Valbona Bejleri, University of District of Columbia, 4200 Connecticut Ave., NW, Washington, DC 20008, *vbejleri@udc.edu*

***Key Words:*** prediction, Bayesian, Poisson, prior, non informative

Despite the different conceptual meaning of the frequentist and Bayesian approaches, relationships between the solutions these two approaches give to a specific prediction problem exist. Relationships between Bayesian prediction limits for Poisson distribution derived from noninformative priors and limits constructed in a frequentist way are presented. Steps on derivation of Bayesian limits are included for completeness. It is shown that the Bayesian lower limit derived from a modified Jeffreys' non informative prior is equal to the frequentist lower prediction limit and the frequency that the prediction will be wrong equals the Bayesian probability of wrong coverage. It is also shown that for any proper prior, there exist possible values of the observed outcome and the rate parameter such that the Bayesian lower limits do not coincide with the frequentist lower limits.

## Reconciling Bayesian and Frequentist Evidence in the One-Sided Scale Parameter Testing Problem

❖ Athanasios Micheas, University of Missouri-Columbia, Department of Statistics, 146 Middlebush Hall, Columbia, MO 65211, *amicheas@stat.missouri.edu*; Dipak Dey, University of Connecticut

***Key Words:*** Bayesian p-values, posterior probability, predictive distribution, prior distribution

An interesting but controversial problem arises when Bayesian as well as frequentist methodologies suggest very often similar solutions. For over two decades, there has been an effort by several authors to assess when Bayesian and frequentist methods provide exactly the same answers when employed. We encounter this situation in the problem of hypothesis testing, where Bayesian evidence, such as Bayes factors and prior or posterior predictive p-values are set against the classical p-value. In this paper, we develop prior predictive and posterior predictive p-values for one sided hypothesis testing scale parameter problems. We reconcile Bayesian and frequentist evidence by showing that for many classes of prior distributions, the infimum of the prior predictive and posterior predictive p-values are equal to the classical p-value. The results are illustrated through standard examples.

## Combining Bootstrap and Bayesian Inferences

❖ Yan Zhou, University of Michigan, 1417 McIntyre Drive, Dept. of Biostatistics, Ann Arbor, MI 48105, *yzhouz@umich.edu*

***Key Words:*** bootstrap confidence interval, Bayesian credibility interval, discrepancy

In the case of independent identically distributed samples, the naive bootstrap yields confidence limits that are asymptotically correct to the first order, but have less certain confidence coverage in small samples. Bayesian credibility intervals based on the posterior distribution of the model parameters tend to perform better for small samples, but are more dependent on modeling assumptions than the bootstrap. A discrepancy statistic based on the difference of model and bootstrap estimates of standard error is used as a basis for combining bootstrap and Bayesian inferences. The goal is to achieve a compromise that combines the advantages of those two methods, yielding intervals that combine robustness with good small-sample confidence coverage. We assess properties of our method by some simple simulation experiments.

### Higher Order Semiparametric Frequentist Inference Based on the Profile Sampler

❖ Guang Cheng, University of Wisconsin-Madison, Department of Statistics, Medical Science Center 1300 Univ Ave, Madison, WI 53706, *cheng@stat.wisc.edu*; Michael Kosorok, University of Wisconsin-Madison

***Key Words:*** higher order frequentist inference, Markov chain Monte Carlo, profile likelihood, the Cox proportional hazards model, the proportional odds model, case-studies with a missing covariate

We consider higher order frequentist inference for the parametric component of a semiparametric model based on sampling from the posterior profile distribution. The first order validity of this procedure established by Lee, Kosorok and Fine (2005) is extended to second order validity in the setting where the infinite dimensional nuisance parameter achieves the parametric rate. Specifically, we obtain higher order estimates of the maximum profile likelihood estimator and of the efficient Fisher information. Moreover, we prove that an exact frequentist confidence interval for the parametric component at level alpha can be estimated by the alpha level credible set from the profile sampler with error of order $O_P(n^{-1})$. As far as we are aware, these results are the first higher order frequentist results obtained for semiparametric estimation. The theory is verified for three examples.

## 293 Contributed Posters

General Methodology, WNAR, Biometrics Section, Biopharmaceutical Section, Section on Statistics and the Environment, Section on Survey Research Methods

**Tuesday, August 8, 10:30 am–12:20 pm**

### Sample Size and Power Calculations in Case-Control Designs for Detecting Gene-Environment Interaction in the Presence of Environmental Correlation

❖ Amy Murphy, Harvard School of Public Health, Department of Biostatistics, 655 Huntington Avenue, Boston, MA 02115, *amurphy@hsph.harvard.edu*; Peter Kraft, Harvard University

***Key Words:*** gene-environment interaction, sample size, power, correlation

In planning any type of study, a crucial phase is determining the required sample size to detect the association of interest. In the context of genetic association testing, research on sample size estimation is focused on detecting gene-environment interactions, which are more difficult to detect. An implicit assumption in sample size calculations for case-control association designs is that the case and control exposures are independent. If their environmental exposures are correlated, the distribution of the environment and gene-environment covariates will be affected, requiring an adjustment to the overall sample size to maintain power. We have developed methodology and software that allows for environmental correlation. Its relevance is shown in a genetic study of vitamin D and prostate cancer, where matching subjects by time of blood draw induces an correlation in vitamin D levels.

### Examining the Multiple Dye-Swap Design for Efficient and Effective Microarray Studies

❖ Theresa Kim, University of Washington, Department of Biostatistics, Box 357232, Seattle, WA 98195, *tykim@u.washington.edu*

***Key Words:*** experimental design, gene expression, microarray

For 2-group comparison microarray experiments with biological replicates, Kerr (Biometrics, 2003) showed that a multiple dye-swap strategy is a very efficient, practical design. For more than two groups, the design extends naturally, but there is more than one generalization when there are four or more groups. We explore different options for the extension of the multiple dye-swap design for 3, 4, 5, and more groups and evaluate these options in terms of their efficiency for making pairwise group comparisons.

### Using the Maximum Cross-Correlation Statistic To Find Significant Voxel-Wise Activations in fMRI Experiments

❖ Kinfemichael Gedif, Southern Methodist University, 12920 Audelia Road, 269, Dallas, TX 75243, *kgedif@smu.edu*; Richard F. Gunst, Southern Methodist University; Qihua Lin, Southern Methodist University; William R. Schucany, Southern Methodist University

***Key Words:*** maximum cross-correlation, fMRI, HRF

Most statistical analysis on fMRI experiments have a common objective of identifying activated voxels due to some stimulus. One way this is done is by performing voxel-wise test of the null hypothesis that the observed response is not significantly related to an assigned theoretical hemodynamic response function (hrf). Statistical analysis based on fitting hrf models strongly depends on the adequacy of the fitted model on each voxel's time series. We investigate a technique that does not require fitting an hrf to the voxel time courses. The maximum cross-correlation statistic between the observed response and the ideal stimulus sequence is used to construct a voxel activation map. Resampling can be used to test for significant cross-correlation. Such a method not only avoids fitting an hrf but also handles the fact that the hemodynamic response is temporally blurred.

### Predict Gene Expression Using Logistic Regression

❖ Lei Guo, Harvard University, Department of Statistics, 1 Oxford street, Cambridge, MA 02138, *lguo@fas.harvard.edu*; Yuan Yuan, Harvard University; Lei Shen, GlaxoSmithKline; Jun Liu, Harvard University

***Key Words:*** logistic regression, gene regulation, transcription factor binding sites, cross validation

Transcription factors (TFs) play crucial roles in gene regulation by interacting with genomic DNA. It has been shown that the Bayesian network can be applied to learn regulatory network structures and predict expression patterns of genes by their sequence information (Beer and Tavazoie, 2004). Although the prediction accuracy is 73%, it is not clear whether other simpler models can do better. We experimented with a simple logistic regression model together with variable selection, and achieved higher prediction accuracy (78%) under the setting identical to that of Beer and Tavazoie, even without motif site orientation and location information. Furthermore, we showed how their incorrect cross-validation inflated the estimated prediction accuracy.

### A Nonparametric Method of Background Correction for Microarray Data Analysis

❖ Zhongxue Chen, Southern Methodist University, 3225 Daniel Ave., P.O. Box 750332, Dallas, TX 75275-0332, *zhongxue@mail.smu.edu*; Monnie McGee, Southern Methodist University; Richard Scheuermann, The University of Texas Southwestern Medical Center at Dallas

***Key Words:*** microarray, background correction, oligonucleotide arrays, nonparametric, gene

Probe level data preprocessing is very important for microarray data analysis. Background correction is one of the three steps of preprocessing and it has a great influence on the next steps. We propose a new background correction method, which will use information from both of PM and MM. We use the lowest q2 percentile of MM that associated with the lowest q1 percentile of PM to estimate the background noise. This new method is compared with other methods by using the spike in dataset. The results show that our method has very good performance compared with the three most commonly used background correction methods: MAS5.0, RMA and dchip.

## A Case Study for Finding Condition-Specific cis-Regulatory Motifs and Modules in the Mouse Genome

❖ Dongseok Choi, Oregon Health & Science University, Div of Biostatistics, Dept. of Public Health & Preventative Medicine, 3181 SW Sam Jackson Park Road, CB-669, Portland, OR 97239, *choid@ohsu.edu*; Yuan Fang, Oregon Health & Science University; William Mathers, Oregon Health & Science University

***Key Words:*** co-regulation, cis-regulatory motifs, cis-regulatory modules, microarray, cluster, transcription factor

Deciphering genetic regulatory codes remains a challenge. By combining cluster analysis, known cis-regulatory motifs search, and motifs/modules finding program, we could identify in vivo condition-specific co-regulation in the mouse genome. A resampling-based algorithm was adopted to cluster our microarray data of a stress response, which generated 35 tight clusters with unique expression patterns containing 811 genes out of 5652 genes significantly altered. Database searches identified many known motifs within the 3 kilobase regulatory regions of 40 genes from 3 clusters, and modules with 6 to 9 motifs that were commonly shared by 60-100 percent of these genes. CisModule program predictions were comparable with the results from database searches and found 4 potentially novel motifs. These motifs and modules could be responsible for gene co-regulation of the stress response.

## Identifying Responder Cells in High-Content Imaging

❖ Shuguang Huang, Eli Lilly and Company, 7220 Madison Village Court, Indianapolis, IN 46227, *huang_shuguang@lilly.com*; Adeline Yeo, Eli Lilly and Company

***Key Words:*** high-content screening, imaging, cell, statistics, molecular

Over the past few years, there is a surge of applications of light microscopy in bio-imaging, which allows molecule-specific (DNA, RNA, proteins) measurements to be made with remarkable sensitivity. High-Content Screening (HCS) platform analyzes images and provides measurements of the intracellular fluorescence intensity for multiple parameters on a cell-by-cell basis. Of frequent research interest is finger-printing different physiological state of the cell, which is generally represented by the detection of certain protein. This research investigates the statistical methods for identifying the responder cells among a mixture cell population with cells in and not in the physiological state. In particular, a novel method, maximum selected test statistics, is proposed and shown to work well.

## Simulation Study Investigating Nonrandom Recombination Events on Single Point LOD Scores

❖ Derek Blankenship, The University of Oklahoma Health Sciences Center, 936 NE 4th Street, Moore, OK 73160, *derek-blankenship@ouhsc.edu*

***Key Words:*** linkage analysis, recombination fraction, simulation, genetics, LOD score

Statistical techniques make linking human traits to loci possible. Stat procedures benefit from recombination (RC) of chromosomes during meiosis, typically assumed to occur at random. However, there are known “hot/cold” spots where RC is not random. We implemented a simulation study to investigate effects of nonrandom recombination on LOD scores. We demonstrate a simulation of an autosomal dominant trait on a chromosome containing 10 loci, equidistant with a recombination fraction (RF) of 0.11 using SAS. There were 100,000 replicates of 6 families with 4 offspring and varying RF up to 100% and down to 99% for each location. As expected, the amount of change in LOD scores and RF is a function of location and magnitude of non-randomness of the “hot/cold” spot. The “hot/cold” spot had to be between the marker and trait locus and differences increased with magnitude of non-randomness.

## A Method for Computing the Overall Statistical Significance of a Treatment Effect among a Group of Genes

❖ Taewon Lee, National Center for Toxicological Research, 3900 NCTR Road HFT20, Jefferson, AR 72079, *taewon.lee@fda.hhs.gov*; Robert Delongchamp, National Center for Toxicological Research; Cruz Velasco, Louisiana State University Health Sciences Center

***Key Words:*** DNA array, gene ontology, correlations among genes, meta-analysis

In studies that use DNA arrays to assess changes in gene expression, our goal is to evaluate the statistical significance of treatments on expressions for predefined sets of genes, e.g., sets of genes grouped by gene ontology (GO) terms. Computer simulations demonstrated that ignoring the correlations among genes overstates the significance assigned to GO terms. We propose statistical tests which are based on meta-analysis methods for combining p-values. With the tests, reliable corrections for the effect of correlations among genes on the significance level of a GO term can be constructed for an alternative hypothesis where all transcripts in the GO term increase (decrease) in response to treatment. For general alternatives, which allow some transcripts to increase and others to decrease, the bias of naÔve significance calculations can be greatly decreased although not eliminated.

## An Alternative Approach To Estimate Averages in Affymetrix Chips

❖ William R. Schucany, Southern Methodist University, P.O. Box 750332, 3225 Daniels Ave, Dallas, TX 75275-0332, *schucany@mail.smu.edu*; Md. Jobayer Hossain, Southern Methodist University; Julia V. Kozlitina, Southern Methodist University; Kinfemichael Gedif, Southern Methodist University

***Key Words:*** oligonucleotide array, summarization, robust, gene expression

Abstract: It is essential in the analysis of oligonucleotide microarrays to preprocess the data. Background correction, normalization and summarization are the common steps for data preparation. Affymetrix uses a one-step Tukey biweight (TBW) method to estimate average levels in two stages in their algorithm known as MAS 5.0. This TBW is a robust procedure to estimate location unaffected by outliers. Similarly, the Hodges-Lehmann estimator is a classical robust location estimator with good properties. The TBW one step from the median uses symmetric weights even when the data are not symmetric. However, the Hodges-Lehmann statistic need not be symmetric when the data are skewed. We investigate the performance of this robust summary statistic in the preprocessing steps of gene expression data as an alternative to TBW.

## Near-Infrared (NIR) Spectroscopy Coupled with Molecular Marker Data as a Tool for Gene Discovery

❖ Mervyn Marasinghe, Iowa State University, Department of Statistics, 117 Snedecor Hall, Ames, IA 50011, *mervyn@iastate.edu*; Paul Scott, Iowa State University/USDA-ARS

***Key Words:*** quantitative trait locus, NIR spectroscopy, calibration models, genotyping

NIR spectroscopy is a non-destructive, high throughput method that yields information about the chemical composition of biological samples. Many modern model selection procedures have been used to develop calibration models for predicting the concentration of a specific analyte. Quantitative trait locus (QTL) mapping is widely used for identifying genomic regions associated with quantitative traits in plants and animals. Thus these two techniques may be combined to identify regions of an organism's genome that control chemical composition. In the first approach, spectroscopic data from organisms with known molecular markers could be used as learning sets to develop classifiers for genotyping new organisms. In the second approach, spectroscopic data could be used as quantitative trait values in genetic mapping experiments to identify genetic loci controlling spectroscopic features.

## A Matrix Approach for Comparing Estimates of a Population Total under a Many-to-Many Frame Structure

Martin Levy, University of Cincinnati; ❖ ZhiYuan Dong, University of Cincinnati, P.O. Box 210130, Cincinnati, 45221, *dongzn@email.uc.edu*

***Key Words:*** imperfect frames, correspondence errors, weighting, simple random sampling, quadratic forms

We compare population total estimator precision under simple random sampling when frame units and population elements have a many-to-many correspondence. Three methods that adjust for multiplicities are: perfect the frame and use standard estimation methods (PF); adjust for imperfections using either Arc-Weight (AW), or Horvitz-Thomson (HT) estimators. We represent the underlying structure as a bipartite graph and express the variances of PF, HT, and AW as quadratic forms depending on the graph's incidence matrix permitting data-free dominance studies. We show that AW is a close relative to HT when AW is viewed using a new tool, the First Order Inclusion-Weight. A comprehensive search algorithm is developed to enumerate all non-isomorphic bipartite graphs associated with any feasible valence settings to investigate all systems of "small" size and identify non-trivial dominance results.

## Comparison of Alternatives for Controlling Group Quarters Person Estimates in the American Community Survey

❖ Lynn Weidman, U.S. Census Bureau, Washington, DC 20233, *lynn.weidman@census.gov*; Michael Ikeda, U.S. Census Bureau; Julie Tsay, U.S. Census Bureau

***Key Words:*** American Community Survey, group quarters, estimation, controls

Estimation for persons residing in group quarters (GQ) has been carried out only once with American Community Survey data. In 1999 there were 36 counties in sample and GQ stratification, sampling, and weighting was done separately for each county. For the full GQ implementation that began in 2006, a new GQ sort by type within state and selection across the whole state was used, making it possible to also weight GQ persons by state. A simulation study using Census 2000 data evaluated five alternatives for controlling estimates of GQ persons in the weighting: (1)no control, controlling by demographics for (2) GQ persons by state and for GQ and HU persons combined by (3) estimation area (large county or group of smaller counties) and (4) state, and (5) controlling by major GQ types for state. The simulation is described and analyses of the results are presented.

## American Community Survey (ACS) Variance Reduction of Small Areas via Coverage Adjustment Using an Administrative Records Match

❖ Donald Malec, U.S. Census Bureau, Statistical Research Division, room 3132 bldg 4, Washington, DC 20233-9150, *donald.j.malec@census.gov*; Elizabeth Huang, U.S. Census Bureau; Jerry J. Maples, U.S. Census Bureau ; Lynn Weidman, U.S. Census Bureau

In order to reduce variance and correct for coverage of ACS estimates, there is a desire for population control at the tract level. Currently intercensal population controls are based on "usual residence" and are not available at the tract level while the ACS produces intercensal estimates at the tract level that are based on "current residence". This project proposes a way to use new controls, obtained by matching the ACS sample to an administrative records file, and then controlling sample estimates of administrative record tract counts to known tract totals. By matching administrative records addresses, this procedure achieves a consistent residence rule between sample and control. To evaluate the effects of coverage error in the administrative records and matching error with the sample, the procedure is applied to the 2000 Census long-form, where the correct controls are known.

## A Comparison of Strategies for Reducing Item Nonresponse in a Web Survey

❖ Jeffrey Kerwin, Westat, 1650 Research Blvd., Rockville, MD 20850, *jeffreykerwin@westat.com*; Kerry Levin, Westat;
Andrew Wang, National Institute of Standards and Technology;
Stephen Campbell, National Institute of Standards and Technology;
Stephanie Shipp, National Institute of Standards and Technology

***Key Words:*** item nonresponse, web survey, missing data, data quality

One key advantage of a web survey over a mail survey is the possibility for implementing strategies for reducing item nonresponse. We are currently conducting a web survey in which we compare two strategies for reducing item nonresponse. One strategy presents a prompt at the end of the survey which informs respondents of the number of items left unanswered in each of seven sections of the instrument. The page asks if they would like to review these items. Respondents can return to these sections, where they find the unanswered items highlighted. The second strategy prompts respondents immediately after each section that items have been left unanswered, giving them the option of returning to the section to find the highlighted unanswered items. The experiment includes a control group with no prompt for nonresponse. Outcomes of interest are degree and patterns of item nonresponse.

### Estimating Variance of Double Sampling Using Jackknife and Bootstrap Methods

❖ Jing Wang, Sam Houston State University, 1610 Ave. O, Apt 6, Huntsville, TX 77340, *jxw003@shsu.edu*; Ferry Butar Butar, Sam Houston State University

***Key Words:*** double sampling, jackknife, bootstrap

When the variable of interest is relatively expensive to measure and a correlated auxiliary variable can be measured easily then it is efficient to employ a double sampling. The primary purpose of this paper is to compare the efficiency between the Jackknife method and the Bootstrap method when applied to double sampling. The variance of the population total or the population mean will be sought. The estimated variance will be compared by using the Jackknife and the Bootstrap methods, respectively. Using simulation study, we evaluate the efficiency of double sampling compare to simple random sampling.

## 294 Business and Economics Statistics Section Speaker with Lunch (fee event)

Business and Economics Statistics Section

**Tuesday, August 8, 12:30 pm–1:50 pm**

### Housing Correction or Crash

❖ Mark McMullen, Moody's Economy.com, Senior Economist/Director of Government Consulting, 121 North Walnut Street, Ste. 500, West Chester, PA 19380, *mmcmullen@economy.com*

The economy's near-term outlook will be shaped largely by the performance of housing and mortgage markets. The most likely scenario holds that housing and mortgage market activity will weaken in an orderly way and the broader economy will gracefully adjust to a slower, more sustainable rate of growth. The downside risks to this sanguine scenario are significant, however. This discussion considers housing's role in the broader economy and its near-term prospects. Also, the outlook for specific regional housing markets is examined using results from a structural econometric model of housing supply and demand.

## 295 Section on Bayesian Statistical Science Roundtable with Lunch (fee event)

Section on Bayesian Statistical Science

**Tuesday, August 8, 12:30 pm–1:50 pm**

### Objective Bayes Model Selection

❖ M. J. Bayarri, University of Valencia, Department of Statistics, Av Dr Moliner 50, Burjassot, 46100 Spain, *susie.bayarri@uv.es*

***Key Words:*** objective Bayes factors, objective priors, BIC, consistency, multiple testing, Occam's razor

Bayesian model selection has a number of desirable properties, including consistency, generality of application, automatic Occam's razor effect, and appropriate handling of multiplicities. Unlike inference under a single model, in the presence of model uncertainty, the choice of priors has a profound effect. However, in these situations, subjective elicitation is usually unfeasible. In this roundtable, we can debate 'automatic' or objective choice of 'good' priors for model selection (or model averaging). Questions to be considered could include (but are not limited to) the following: Which 'default' or 'automatic' priors should not be used? Which properties should an objective model selection procedure have? How do we numerically compute Bayes factors? When is BIC a sensible approximation? Are there generalizations for increasing parameter dimensionality?

## 296 Biopharmaceutical Section Roundtables with Lunch (fee event)

Biopharmaceutical Section

**Tuesday, August 8, 12:30 pm–1:50 pm**

### Contemporary Issues in Data Monitoring Committees

❖ David Kerr, Axio Research, 2601 4th Ave., Suite 200, Seattle, WA 98121, *davidk@axioresearch.com*

***Key Words:*** data monitoring committee, data safety monitoring board, DMC, DMCS, DSMB, DSMBS

Data Monitoring Committees (DMCs) protect the safety of study participants and ensure the credibility of the study and the validity of study results by reviewing data throughout the course of the study. As DMCs have become more commonplace, consensus has formed in some areas, but there are still some controversial topics. These include degree of blinding of the DMC, release of efficacy results to the DMC, legal liability of DMC members, use of external information in DMC decisionmaking, and ethical considerations in DMC decisionmaking. The speaker has participated in nearly 100 DMC meetings in the role of Independent Statistical Center.

### Controlling Error Rate in Safety Assessments from a Regulatory Perspective

❖ Qian Li, U.S. Food and Drug Administration, 10802 Whiterim Drive, Potomac, MD 20854, *qian.li@fda.hhs.gov*

***Key Words:*** error rate, no worse than placebo, protective effect, safety analysis, safety signals

In safety assessment, we should control the error rate of wrongfully rejecting a true safety signal and be less stringent to that of wrongfully accepting a false safety signal. Current safety analyses have emphasized detection of significant safety signals. It is obvious that statistically nonsignificant results do not necessarily mean there is no safety concern. It is important to emphasize that unless we can show no harm with certain confidence, a drug cannot be declared safe---even if there is no statistically significant signal. For this purpose, we may consider establishing three types of safety testing hypotheses: showing protective effect, demonstrating no worse than placebo, and detecting risk signals. We would like to discuss if it is feasible to test the hypotheses, how to perform the tests, and what would be the impact on drug labeling and safety management.

### The Use of Targeted Designs in Clinical Trials

❖ Gerald Crans, Eli Lilly and Company, 389 Arbor Drive, Carmel, IN 46032, *crans889@msn.com*; Matthew Rotelli, Eli Lilly and Company

***Key Words:*** targeted design, sample size, efficiency

The cost of drug development has increased markedly over the last decade. Unfortunately, the rate of new drug approvals has not kept up with these costs. This has aroused the ire of pharmaceutical executives and health wary consumers, as the availability of novel agents was not as expected and profits have not materialized. A potential improvement to clinical research---which could reduce costs, increase the probability of technical success, and enhance feasibility---is possible by using targeted designs. These designs use variables collected during screening to enroll only subjects expected to respond to treatment. This could lower costs by reducing the number of subjects needed for enrollment, potentially making viable those clinical trials that were impossible before. The advantages and disadvantages of these designs will be discussed at the roundtable.

### Statistical Concerns and Methodology Regarding Cognitive and Psychomotor Endpoints in Early Development (Phase I Trials)

❖ Cynthia Gargano, Merck & Co., Inc., 309 Washington Street, Apt 4207, Conshohocken, PA 19428, *cynthia_gargano@merck.com*

***Key Words:*** cognitive, test battery, correlation, multiplicity, phase I

It is common in Phase I clinical trials to administer a large battery of cognitive and psychomotor tests to assess safety or proof of concept, depending on the indication sought by the sponsor. As a statistician, one is faced with several interesting problems. First, these tests often vary from site to site due to the way in which subjects' responses are recorded and differing equipment. Second, there are often versions of the same test. Third, the battery is given often in the same order (i.e., no randomization to time). Finally, there are statistical issues such as correlation between various endpoints of the test battery, multiplicity, and the assumption of normality. We will discuss approaches to the above issues.

### Preferred Methods of Dealing with Missing Data in Clinical Trials

❖ Rukmini Rajagopalan, Abbott Laboratories, Global Health Economics and Outcomes Research, Abbott Park, IL , *rr16448@gmail.com*

***Key Words:*** missing data, clinical trials, interpolation, last value carried forward

In the statistical analysis of data from clinical trials, it is essential that data collected are complete. The integrity of the analysis and results depend on the completeness of the data on each subject of the sample because the specified power of the analysis will be achieved only if all subjects of the sample contribute values for the stipulated observations. The problem of missing data is insurmountable. A host of literature exists that advises statisticians to deal with the issue. Several methods for this have been examined and perfected, including interpolation, extrapolation, value averaging, and process of last value carried forward when there are missing values in the efficacy variables. Of course, depending on the study and the efficacy endpoint, each of these methods has advantages and disadvantages.

### Decision Rules Based on Multiple Endpoints in Clinical Trials

❖ Alex Dmitrienko, Eli Lilly and Company, 10913 W. 144th Street, Overland Park, KS 66221, *dmitrienko_alex@lilly.com*

***Key Words:*** multiple endpoints, clinical trials, multiplicity

This luncheon will focus on issues arising in clinical trials with multiple endpoints. We will begin with a general discussion of the role of multiple endpoints in early (proof-of-concept) and registration trials, relevant regulatory considerations, and popular testing strategies---including the reverse multiplicity problem (all endpoints must be significant to achieve a positive outcome). The discussion also will touch upon statistical issues, such as the control of the false-positive rate, analysis of multiple endpoints (overall effect versus analysis of the individual variables), and interpretation of secondary findings and their use in product labels (gatekeeping strategies). Examples from a real trial will be used throughout the discussion to illustrate the approaches being discussed.

### Patient-Reported Outcomes: Issues Related to the Collection and Analysis, Including Measurement Reliability and Sensitivity

❖ Tammy J. Massie, U.S. Food and Drug Administration, 1401 Rockville Pike, WOC I, Room 276S, HFM-217, Rockville, MD 20850, *Tammy.Massie@fda.hhs.gov*

***Key Words:*** patient reported outcomes, clinical trial, measurement reliability, measurement sensitivity

Patient reported outcomes are integral to many clinical trials. Creating and implementing questionnaires that adequately and accurately assess a patient's perspective can be difficult. In this roundtable luncheon, the leader will facilitate discussion by providing a perspective of both a statistician in a regulatory agency and a clinical trial participant.

### Statistician's Role in PK/PD Modeling and Clinical Trial Simulation

❖ Haiyuan Zhu, Merck Research Laboratories, 126 E. Lincoln Ave., RY34-A304, Rahway, NJ 07065, *haiyuan_zhu@merck.com*

***Key Words:*** PK/PD modeling, clinical trial simulation

PK/PD modeling and clinical trial simulation now are recognized generally as important components of drug development. Modeling and simulations require that pharmacokineticists, pharmacologists, clinicians, and statisticians understand one another's science to a higher degree than ever before in order to meet the multidisciplinary demands of pharmacometric analysis---which often involves advanced, complex statistical techniques, such as nonlinear mixed effect models and Bayesian methods. However, the role of statisticians in these analyses in medical research is often not clear, or sometimes ignored. The intention of this roundtable is for people interested in this area to share their experiences and ideas about the statistician's roles in PK/PD modeling and clinical trial simulations and to discuss opportunities for statisticians to play larger roles in the future.

### Retrospective Testing for Baseline Variables Comparability and Subsequent Covariate Adjustments for Significant Imbalance: Good Clinical Practice or Bad Statistical Practice?

❖ Abdul Sankoh, sanofi-aventis, 200 Crossing Blvd., Bridgewater, NJ 08807, *abdul.j.sankoh@aventis.com*

***Key Words:*** baseline variables, covariate adjustment, significant imbalance

Baseline covariate adjustment analyses often are carried out in randomized clinical trials when there is evidence that the treatment difference depends on certain patient characteristics for two main reasons. The first of these is to adjust for imbalance in baseline variables related to the outcome; such adjustment helps correct for the groups' predisposition to behave differently from the outset. The second reason is to reduce the underlying variability in the data so more precise comparisons can be made. It is, however, well-known that post-randomization testing for baseline imbalance and subsequent covariate adjustments for such imbalance has no theoretical justification, notwithstanding its practical appeal.

### Statistical Issues Regarding Experimental Medicine

❖ James Bolognese, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065, *james_bolognese@merck.com*

***Key Words:*** experimental medicine, dose-adaptive study design, n-of-1 trials, proof-of-concept, early drug development, adaptive design

In 2005, I was assigned to support experimental medicine studies. I now report to the head of early development biostat in Merck Research Labs. Littman & Williams [Nature Reviews, 2005] define experimental medicine as “the use of innovative measurements, models, and designs in studying human subjects for establishing proof of mechanism and concept of new drugs, for exploring the potential for market differentiation for successful drug candidates, and for efficiently terminating the development of unsuccessful ones.” While this effort in Merck is still forming, much of my work thus far has involved examination of how dose-adaptive study designs and N-of-1 trials can be implemented for proof-of-concept (PoC) trials and application of classical methods to specific clinical trials of novel endpoints. This luncheon offers an opportunity to share ideas in this area.

### Bayesian Dose-Finding Strategies for Phase I Oncology Trials

❖ Glen Laird, Novartis Pharmaceuticals Corporation, 1 Health Plaza, East Hanover, NJ 07936, *glen.laird@novartis.com*

***Key Words:*** dose escalation, Bayesian, continual reassessment method, overdose control

In recent years, Bayesian estimation methods have come into use for drug dose determination as part of Phase I oncology studies. These methods promise to provide more accurate estimation than traditional escalation schemes, which typically use only dose-limiting toxicity (DLT) information from the most recent cohort of patients. Starting from the Continual Reassessment Method for a single-arm trial based on DLTs, modifications, extensions, and competing methods have been developed. A partial list includes methods for multiple arms, overdose estimation, and joint toxicity-efficacy modeling. Implementation issues also exist, such as the choice of cohort size and stopping rule. The advantages and disadvantages of these methods will be examined and their relative merit discussed against a context of traditional methods and examples from Novartis Oncology.

### On the Evaluation of Benefit and Risk: the Tools We Have and How They Help

❖ Yili Pritchett, Abbott Laboratories, 100 Abbott Park Road, R436, Bldg. AP9A-2, Abbott Park, IL 60064-6116, *yili.pritchett@abbott.com*

***Key Words:*** benefit-risk evaluation, clinical trials, new drug development

The need for quantitative tools to evaluate the benefit-risk profile of a treatment has been recognized and discussed in recent years. Some methodologies (e.g., NNT/NNH, Global Benefit-Risk) have been proposed and applied to real data from clinical trials. PhRMA and FDA have had serious discussions on this topic. What about the day-to-day practice of biostatisticians in industry? Do biostatisticians use certain models and conduct this type of analysis before or after the registration? Do those analyses provide new perspective to the understanding of a therapy? Do those approaches add value to the course of drug development?

### Use of Adaptive Clinical Trial Designs in Clinical Development for Product Registration

❖ Laura Meyerson, Biogen Idec, 14 Cambridge Center, Cambridge, MA 01746, *laura.meyerson@biogenidec.com*

***Key Words:*** adaptive trial design

There has been recent recognition that adaptive trial designs can be useful in shortening the path of clinical development, thereby hastening patient access to efficacious products. Adaptive trial designs range in intent from selecting the optimal dose for which to continue the same study to selecting the correct sample size. Many companies have used these approaches in clinical development leading to a successful registration earlier than the more traditional design; others may have tried and failed. Of particular interest are those studies that have tried this approach for one or more of their registration trials (for example, a phase II/III trial). This luncheon is intended to bring together industry statisticians who have utilized or tried to utilize more adaptive approaches and to discuss their experiences, both successes in registration and/or failures.

## 297 Section on Statistical Consulting Roundtable with Lunch (fee event)

Section on Statistical Consulting
**Tuesday, August 8, 12:30 pm–1:50 pm**

### Extreme Consulting: a Novel Method of Improving Efficiency of Biostatistical Analysis

❖ Daniel Byrne, Vanderbilt University, AA3228 MCN, 1161 21st Avenue South, Nashville, TN 37232-2195, *daniel.byrne@vanderbilt.edu*

***Key Words:*** consulting, analysis, efficiency

Often, the interaction between physician-scientists and biostatisticians regarding the analysis of a medical research project is less than optimally efficient. Extreme consulting is a method of analysis in which both the physician and biostatistician jointly analyze a dataset in real time to avoid delays and communication problems. Our experience with this method has been positive.

## 298 Section on Statisticians in Defense and National Security Roundtable with Lunch (fee event)

Section on Statisticians in Defense and National Security
**Tuesday, August 8, 12:30 pm–1:50 pm**

### Epidemiology and Education: Using Public Health and National Security in Statistics Education

❖ Henry Rolka, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, *hrr2@cdc.gov*; Donna Stroup, Centers for Disease Control and Prevention

*Key Words:* teaching, epidemiology, public health, national security, statistical literacy

Scientific advances bring clear benefits to society, especially improvements in health. At the same time, creation of opportunities that enable all students to reach their potential for education is a high priority. In addition, foreign advances in science now rival or exceed those of the United States. Epidemiology, the science of public health, provides a compelling and relevant context for teaching statistics. This roundtable luncheon will discuss the role public health and national security can play in statistics education in the United States. The Centers for Disease Control and Prevention has made its considerable educational resources available to promote undergraduate coursework in public health in the nation's colleges. We will discuss these efforts and an initial evaluation to help develop and recruit qualified statisticians in the public health workforce.

# 299 Section on Statistical Education Roundtables with Lunch (fee event)

Section on Statistical Education

**Tuesday, August 8, 12:30 pm–1:50 pm**

## Using Technology in the Classroom and for Distance Education

❖ Michael Speed, Texas A&M University, 3143 TAMU, Department of Statistics, College Station, TX 77843-3143, *mspeed@tamu.edu*

*Key Words:* technology, distance education

This roundtable will focus on recent advances in using technology for both on-campus and distance students. We will share successes and failures. "Build it and they will come" is true. Come share your experiences.

## Teaching Introductory Statistics with (All) Sports Examples

❖ Robin Lock, St. Lawrence University, Department of Math CS Statistics, Canton, NY 13617, *rlock@stlawu.edu*

*Key Words:* education, sports, teaching

Many students have a keen interest in sports, either as athletes, fans, or both. Can we tap into that interest to motivate the study of statistics? While not appropriate for all students, including some who actively rebel at sports examples, suppose we offer one section of a traditional introductory course specifically advertised to focus on sports examples. Will students be attracted to such a course? Can we still cover all the traditional statistical topics and ideas? Where are good sources for data and problems students find interesting? Can we still give them a sense of the applicability of statistics to lots of areas (e.g., business, psychology, biology) while staying within the realm of sports-motivated examples? These questions and our experience with offering such a course will serve as the theme for this roundtable discussion.

# 300 Section on Statistics in Epidemiology Roundtable with Lunch (fee event)

Section on Statistics in Epidemiology

**Tuesday, August 8, 12:30 pm–1:50 pm**

## Design and Analytic Challenges of Conducting Biomedical Research with Older Populations: Steps Toward a Gerontologic Biostatistics

❖ Peter H. Van Ness, Yale University, 1 Church Street, 7th Floor, School of Medicine, Program on Aging, New Haven, CT 06510, *peter.vanness@yale.edu*; Heather G. Allore, Yale University

*Key Words:* biostatistics, gerontology, multicomponent trials, missing data, clinical tradeoffs

Multifactorial etiologies of geriatric health syndromes and multiple morbidities of elderly patients present challenges for designing and analyzing clinical research studies in which they participate. The objective of this roundtable is to discuss a set of statistical issues that, while not unique to gerontologic research, constitute a distinctive ensemble worthy of deliberate and systematic attention (e.g., design of multicomponent intervention trials, regression modeling with missing data, evaluation of clinical tradeoffs, and integration of qualitative studies with quantitative analyses). The session will conclude with discussion of a proposed subdiscipline of gerontologic biostatistics, whose pedagogical implementation should prepare statisticians for collaborating in clinical research with geriatricians, gerontologists, and elderly populations.

# 301 Section on Government Statistics Roundtables with Lunch (fee event)

Section on Government Statistics, Section on Health Policy Statistics, Committee on Gay and Lesbian Concerns in Statistics

**Tuesday, August 8, 12:30 pm–1:50 pm**

## Including Sexual Orientation Questions in Surveys: Issues and Approaches

❖ Sarah Boslaugh, Washington University in St. Louis, School of Medicine, Department of Pediatrics, 660 South Euclid Campus Box 8208, St Louis, MO 63108, *boslaugh_s@kids.wustl.edu*

*Key Words:* measurement issues, questionnaire design, survey research, sexual orientation

There is increasing interest within the social science community in studying sexual orientation as a demographic characteristic analogous to gender or race. However, measuring sexual orientation on a questionnaire raises a number of measurement issues, the most important of which is that there is neither philosophical nor practical agreement about what it means to be "gay" or "straight." The way sexual orientation is defined and the way the questions are worded can make a huge difference in what categories people select to define themselves. This roundtable will review differing ways major surveys have measured sexual orientation and some of the ways researchers have been able to glean information about sexual orientation from surveys that did not include questions specifically about sexual orientation.

### ASA and Volunteerism: New Special-Interest Group

❖ Roberta Sangster, Bureau of Labor Statistics, 2755 Ordway Street, NW, 508, OSMR, Washington, DC 20008, *Sangster_R@bls.gov*

*Key Words:* volunteerism, special interest group

This is your opportunity to learn about an exciting new special-interest group (SIG) on volunteerism. You will hear about current projects and how to become involved with the SIG. Current projects include the Innocence Project, dealing with analysis of DNA evidence and the death penalty, and the design of a graduate curriculum for Vietnam. Another group is looking at cross-national comparisons in education.

## 302 Section on Statistical Graphics Roundtable with Lunch (fee event)

Section on Statistical Graphics

**Tuesday, August 8, 12:30 pm–1:50 pm**

### Are Graphics/Interactive Graphics Useful for Getting Your Work Done?

❖ Matthias Schonlau, RAND Corporation, 201 N. Craig Street, Pittsburgh, PA 15213, *matt@rand.org*; Martin Theus, University of Augsburg

*Key Words:* interactive graphics, data, visualization, applied statistician

This lunch is a discussion about our experiences with using graphics in our daily lives. We are interested in hearing from both statisticians who enjoy using graphs and from statisticians who do not. What type of graphs do you find most useful? Do you enjoy interactivity when using graphs? Do you use colored graphs? Does the capability of your software determine how you use graphs? Are there specific tasks for which you use graphs most often? Do you find graphs helpful for communicating with nonstatisticians? Can you think of a time where a graphic really made a difference? Did you ever find yourself in a situation in which a collaborator used a graph that in your opinion was not appropriate? Do graphics help you get your work done? The target audience for this lunch is the applied statistician who analyzes data regularly.

## 303 Section on Health Policy Statistics Roundtables with Lunch (fee event)

Section on Health Policy Statistics

**Tuesday, August 8, 12:30 pm–1:50 pm**

### The Role of Prescription Drugs in Health and Disability

❖ Michele Connolly, U.S. Social Security Administration,
3532 Annex, 6401 Security Blvd., Baltimore, MD 21235-6401,
*michele.connolly@ssa.gov*

*Key Words:* disability, functioning, health outcomes, prescription drugs

This roundtable will explore the role of prescription drugs in counteracting disability and improving health and functioning. Often, health and disability status are improved by the use of prescription drugs, yet researchers tend to not consider this. On the other hand, clinical trials for prescription drugs often focus on short-term effects, rather than the long-term implications for people with disabilities or those with poor health and functioning.

### Peer Grouping Methodology in Hospital Profiling

❖ Chintamani Rairikar, BlueCross BlueShield of Illinois, 1111 Amaranth Drive, Naperville, IL 60564, *rairikarc@bcbsil.com*

*Key Words:* peer grouping, hospital profiling, urban, rural

Peer grouping in hospital profiling aims to create arrays of hospitals for comparison purposes. Hospitals were classified as urban or rural based on the population density at the hospital's location. Specialty hospitals (e.g., pediatric, rehabilitation, behavioral health, nonacute, long-term care, and government) were excluded. Seven measures (i.e., case mix, number of hospital beds, resident-to-bed ratio, transplants, open-heart surgery, trauma centers, and neonatal ICUs) were used to generate an overall point score reflecting the complexity of services, size, and patient mix for each urban hospital. Two measures (i.e., critical access status and number of hospital beds) were used to define the peer group for each rural hospital. Peer grouping hospitals is a critical component of the hospital profiling process.

## 304 Section on Physical and Engineering Sciences Roundtable with Lunch (fee event)

Section on Physical and Engineering Sciences

**Tuesday, August 8, 12:30 pm–1:50 pm**

### What's the Difference between Collaborative Research and Consulting?

❖ Roger W. Hoerl, GE Global Research, 1 Research Circle, Schenectady, NY 12309, *hoerl@crd.ge.com*

*Key Words:* statistical consulting, collaboration, research

Many statisticians, especially those in the private sector outside of pharma, work in roles primarily consultative in nature. In these roles, we typically provide guidance to engineers, business people, or other professionals to solve statistical aspects of their problems. Due to many radical changes in our environment, such as the growth of the internet in global enterprises, the demand for high-cost statistical consulting services has flattened---if not decreased. Conversely, some statisticians in research environments have migrated to a more proactive role, which some call collaborative research. A major uniqueness of this role versus consulting is that the statistician collaborates on the problem as an equal researcher, sharing responsibility for problem resolution, as well as for publication and patenting. We plan to discuss these roles in more detail.

## 305 Section on Quality and Productivity Roundtable with Lunch (fee event)

Section on Quality and Productivity
**Tuesday, August 8, 12:30 pm–1:50 pm**

### Graphical and Numeric Approaches to Selecting Effects in Two-Level Factorial Models

❖ Wayne Adams, Stat-Ease, 2021 E. Hennepin, Minneapolis, MN 55413, *wayne@statease.com*

***Key Words:*** half-normal, selection, DOE, factorial, Pareto, designs

Gather to discuss the various graphical and numeric methods for selecting the "vital few" effects when analyzing two-level factorial designs. Topics will include the half-normal or Daniels Plot; Pareto charts; and forward, backward, and stepwise selection. The discussion will consider various objections to these selection methods.

## 306 Section on Survey Research Methods Roundtables with Lunch (fee event)

Section on Survey Research Methods
**Tuesday, August 8, 12:30 pm–1:50 pm**

### Sharing Frame Information To Improve Coverage and Reduce Cost and Duplication of Effort

❖ Paula Weir, Energy Information Administration, 1000 Independence Ave., SW, Washington, DC 20585, *Paula.Weir@eia.doe.gov*

***Key Words:*** frames, coverage, efficiency, sharing, CIPSEA

The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 encourages federal statistical agencies to seek to improve their efficiency continuously. However, only the U.S. Census Bureau, Bureau of Labor Statistics, and Bureau of Economic Analysis are authorized specifically to share data---and only business data---for exclusively statistical purposes. The cost of frames construction and maintenance and the increased difficulty of frames construction in a more complex world at a time of limited resources demands efforts on sharing frame information be examined in a broader context and new approaches sought. This roundtable will provide a forum for an informal discussion of this topic to spur new ideas and identify partners seeking solutions that both protect confidential information and foster efficiency for sharing information for frames.

### Bayesian Inference for Complex Sample Surveys

❖ Roderick J. Little, University of Michigan, Department of Biostatistics, 1420 Washington Heights, Ann Arbor, MI 48109, *rlittle@umich.edu*

***Key Words:*** Bayes, superpopulation models, predictive inference

Finite population sampling is perhaps the only area of statistics where the primary mode of analysis is based on the randomization distribution, rather than on statistical models for the measured variables. The Bayesian modeling approach to survey inference often is rejected as being too subjective, but the approach can be used successfully in the large survey setting by formulating weak models that take into account the survey design and by including relatively noninformative priors. This roundtable will promote a dialogue about the strengths and weaknesses of this approach relative to design-based inference.

## 307 Social Statistics Section Roundtable with Lunch (fee event)

Social Statistics Section
**Tuesday, August 8, 12:30 pm–1:50 pm**

### Telephone Surveys: Where Do We Go from Here?

❖ Clyde Tucker, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 1950, Washington, DC 20212, *tucker.clyde@bls.gov*

***Key Words:*** RDD surveys, nonresponse bias, mode effects, multimode surveys, cell phones, CATI

Recently, telephone surveys have come upon hard times. Criticism of their response rates and representativeness, especially with the growing number of cell-only households, has undermined the legitimacy of one of the mainstays of survey research. To address the problems facing the telephone survey industry, the 2nd International Conference on Telephone Survey Methodology was held in Miami in January 2006. This roundtable will discuss some of the major themes of the conference, including multi-mode data collection, adapting surveys to cell phones, response rates, and nonresponse bias.

## 308 ASA Stat Bowl Session 2

The ASA, ENAR, IMS, SSC, WNAR
**Tuesday, August 8, 2:00 pm–3:50 pm**

### Round 2

❖ Winners from Session 1, Six players will advance from Round 1 to Round 2, , *meetings@amstat.org*

Round 2 will consist of two games. Six players will advance from Round 1 to Round 2, three players going to each game. The six players will be determined by the winners of the four games, plus the two non-winning players that score the highest. The winners of the two Round 2 games will face off in the Finals. Each player will receive team points depending upon how they finish (4 points for champion, 3 points for runner up, 2 points to semifinalists not advancing, 1 point to participants not advancing to round 2). Team accumulating the most points will be declared team champion. Ties for team totals are broken by the highest average score for all members of the teams in question.

## 309 Modern Statistics at Land Grant Institutions: Memorial Session for Holly Fryer ●

Memorial, Section on Statistical Education
**Tuesday, August 8, 2:00 pm–3:50 pm**

### The Life and Times of Holly Fryer

❖ John E. Boyer, Kansas State University, Department of Statistics, Dickens Hall, Manhattan, KS 66506, *jboyer@stat.ksu.edu*

***Key Words:*** Holly Fryer, biography, administration

Holly Fryer was the first PhD in statistics from Iowa State University and the founding father of statistics at Kansas State University. This presentation tells Holly's life story and details some of his professional, administrative, and academic accomplishments, including his successful efforts to establish a Department of Statistics and Statistical Consulting Service at Kansas State.

### Recent Developments on Multiplicative Interaction Models

❖ Dallas E. Johnson, Kansas State University, 101 Dickens Hall, Kansas State U., Manhattan, KS 66503-0802, *dallas@stat.ksu.edu*

***Key Words:*** multiplicative interaction, AMMI models, SAS, Tukey's model, Mandel's model, two-way experiments

This paper considers the analysis of two- and three-way treatment structures using multiplicative interaction models such as Tukey's Model, Mandel's Model, and a general class of multiplicative interaction models sometimes known as AMMI models. SAS Macros that were developed recently to aid researchers in using these kinds of models will be illustrated. In addition to computing meaningful test statistics, the macros provide many useful graphical displays to help identify a pattern in interaction effects when such a pattern exists. The purpose of this paper is to illustrate the macros by analyzing several datasets.

### Estimating Genetic Relatedness

❖ Bruce S. Weir, University of Washington, Department of Biostatistics, 1705 NE Pacific Street, Box 357232, Seattle, WA 98195-7232, *bsweir@u.washington.edu*

***Key Words:*** genetic relatedness, SNP, population genetics, identity by descent, quantitative genetics, forensic genetics

Membership in the same population or family leads to genetic similarities among individuals, and the characterization of these similarities underlies much of population, quantitative, and forensic genetics. The framework of 15 patterns of identity by descent at each locus for two related individuals has been known for many years, but it is only the recent explosion in the quantity of genetic marker data that has made estimation of these components feasible. It is now possible to use several hundred markers, such as single nucleotide polymorphisms (SNPs), and simple likelihood methodology to place people correctly into one of several populations or to distinguish between relatives with only small relationship differences. The first activity can help identify risk factors for human diseases, and the second can aid in the identification of remains following mass disasters.

### Signals and Noises: Statistical Models Based on Fundamental Tenets of the Environmental and Ecological Sciences

❖ Mark S. Kaiser, Iowa State University, Department of Statistics, Ames, IA 50011-0001, *mskaiser@iastate.edu*

***Key Words:*** mixture models, dynamic models, law of limiting factors, maximum likelihood, Bayesian analysis

Environmental and ecological scientists often indicate that data they obtain in field studies is exceedingly "noisy." This suggests a conceptualization of the problems under investigation as "signal" plus "noise," a familiar structure for statistical models with additive errors. But most environmental and ecological process actually consist of a number of "signals" and a number of "noises." A profitable approach to the formulation of statistical models for any number of problems in these areas is to use the basic ecological theory involved to define the primary "signal" of concern. This can lead to interesting statistical structures, as is illustrated with several problems that have arisen through interactions with applied ecologists and environmental scientists.

## 310 Recent Advances in Resampling Methods for Complex Data Structures ✪

Section on Nonparametric Statistics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A Nonparametric Plug-in Rule for Smoothing Parameter Selection

❖ Soumendra N. Lahiri, Iowa State University, 315 E. Snedecor Hall, Department of Statistics, Ames, IA 50011, *snlahiri@iastate.edu*

***Key Words:*** block bootstrap, optimal block size, smoothing parameter, plug-in rule

In this talk, we describe a nonparametric plug-in principle for selecting smoothing parameters in nonparametric curve estimation problems using suitable resampling methods. The key idea is to combine the bootstrap and other resampling methods suitably so the various population parameters appearing in the theoretical optimal value of the smoothing parameter are estimated implicitly (i.e., without explicit analytical expressions). This proposed method possesses the computational simplicity of a plug-in approach but without the analytical work on the part of the user. Usefulness and properties of the proposed method are illustrated in problems involving optimal block length selection for block bootstrap and optimal bandwidth selection for nonparametric regression function/density estimation.

### A Blockwise Empirical Likelihood for Spatial Data

❖ Daniel J. Nordman, Iowa State University, Department of Statistics, Ames, IA 50010, *dnordman@iastate.edu*

***Key Words:*** data blocks, empirical likelihood, pseudo likelihood, random field

In this talk, we give a spatial blockwise empirical likelihood (EL) for spatial observations laying on a grid inside a spatial sampling region. For a large class of weakly dependent spatial processes, the spatial EL method results in likelihood ratios that can be used to build nonparametric confidence regions for spatial parameters, including means, variograms, and parameters identifying conditionally specified spatial models. An advantage of the EL approach over other spatial likelihood formulations is that no direct estimates of variance or skewness are required. Maximum empirical likelihood estimators also are available for parameter estimation and testing spatial moment conditions (e.g., spatial goodness-of-fit tests). This method extends the blockwise EL developed by Kitamura (1997) for short-range dependent time series.

### The Limit of Finite Sample Size and a Problem with Subsampling

❖ Donald W. K. Andrews, Yale University, P.O. Box 208281, New Haven, CT 06520, *donald.andrews@yale.edu*

*Key Words:* subsample, discontinuous distribution, subsample test, asymptotic size, subsample confidence interval

This paper considers tests and confidence intervals based on a test statistic with a limit distribution that is discontinuous in a nuisance parameter or the parameter of interest. The paper shows standard fixed critical value tests and subsample tests often have asymptotic size---defined as the limit of the finite sample size---greater than the nominal level of the test. We determine precisely the asymptotic size of such tests under a general set of high-level conditions that are relatively easy to verify. Often, the asymptotic size is determined by a sequence of parameter values that approach the point of discontinuity of the asymptotic distribution. The problem is not a small sample problem. For every sample size, there can be parameter values for which the test over-rejects the null hypothesis. Analogous results hold for confidence intervals.

### Unit Root Testing via the Tapered Block Bootstrap

❖ Cameron Parker, University of San Diego, La Jolla, CA 92093, *ccparker@sandiego.edu*; Dimitris Politis, University of California, San Diego

*Key Words:* subsampling, unit root testing

In this talk, a nonparametric, residual-based test for detecting the presence of a unit root using the tapered block bootstrap will be proposed. It has the advantage of being valid for a wide class of weakly dependent processes and not based on any parametric assumptions in the data-generating process. We will look at both the large-sample asymptotics and small-sample performance.

## 311 Computational Challenges of Massive Datasets and Sources ● ✪

Section on Statistical Computing, Section on Statisticians in Defense and National Security, Section on Statistical Graphics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Statistical and Computational Issues Associated with Massive Public Health Databases for Detecting Adverse Events

❖ Kathe E. Bjork, Colorado Department of Public Health and Environment, 4300 Cherry Creek Drive, S., Emergency Preparedness and Response Section, Denver, CO 80246, *kathe.bjork@state.co.us*

*Key Words:* surveillance, event detection, cusum, situational awareness

The bioterrorist events of late 2001 and the current threats of global pandemic/panzootic influenza have resulted in increased requirements and funding from the CDC and DHS to states for planning, surveillance, and response. The rapid increase in the magnitude and scope of surveillance data has generated a growing body of statistical and epidemiologic applications and methods, with some unproven in their applicability for improving public health practice. CDPHE-EPRS is responsible for developing data management systems for emergency response in Colorado, including compiling, assessing, merging, and analyzing data from numerous---sometimes disparate---sources for health and threat intelligence. This talk will describe management of data streams, analytical methods, and epidemiologic and statistical issues and challenges for using traditional and nontraditional health data sources.

### Dimension Reduction of Large Datasets in the Atmospheric Sciences

❖ Barbara A. Bailey, University of Colorado, Department of Mathematical Sciences, P.O. Box 173364, Denver, CO 80217-3364, *bbailey@math.cudenver.edu*

*Key Words:* nonlinear PCA, neural networks

Principal component analysis (PCA) is a multivariate statistical method widely used to reduce the dimensionality of large fields of atmospheric data. A nonlinear extension of the PCA, denoted nonlinear principal component analysis (NLPCA), can explain more of the variance and extract nonlinear features. A feed-forward neural network is used as a flexible nonlinear model to allow nonlinear mappings, whereas PCA only allows for linear mappings. Model selection and statistical properties of the neural network model parameters are investigated. The technique is applied to the noisy Lorenz system and atmospheric science data. Visualization of the results of the NLPCA is discussed.

### Automated Metadata

Edward Wegman, George Mason University; ❖ Faleh Alshameri, George Mason University, 4201 Massachusetts Ave, NW, Washington, DC 20016, *alshameri@gmail.com*

Homeland security implies searching massive databases for information involving possible terrorists and the threats they are likely to bring. Many of these databases include free-form text such as intercepted emails and transcripts of phone calls. The implication is that these massive databases are sufficiently large that they cannot be examined thoroughly by humans. Generally, metadata involve information about the format of the data, but not necessarily the actual content. The concept of automated metadata is to use data mining tools to extract features from the data and attach the features to the data as digital objects in the form of metadata. Thus, an investigator could search for specific datasets having some desired features. We have employed this notion with a dataset involving 16,000 articles gathered from CNN and Reuters. This work is joint with Faleh Alsham.

## 312 Statistical Models in Evolutionary Biology ● ✪

IMS, Biometrics Section, WNAR

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Introductory Discussion

❖ Christopher Genovese, Carnegie Mellon University, Department of Statistics, Baker Hall 132, Pittsburgh, PA 15213, *genovese@stat.cmu.edu*

Introduction to the session by Christopher Genovese

### Mathematical Models of Speciation

❖ Sergey Gavrilets, University of Tennessee, 1416 Circle Drive, Knoxville, TN 37996, *gavrila@tiem.utk.edu*

Understanding how new species emerge is an important challenge facing evolutionary biology. I will review what mathematical models tell us about the dynamics of speciation. I will concentrate mostly on simple analytical models allowing one to evaluate the probability, waiting time to, and duration of speciation.

### Fitness Value of Information

❖ Carl Bergstrom, University of Washington, Department of Biology, Box 351800, Seattle, WA 98195-1800, *cbergst@u.washington.edu*; Michael Lachmann, Max Planck Institute for Mathematics in the Sciences

Biologists measure information in different ways. Neurobiologists and researchers in bioinformatics often measure information using information-theoretic measures, such as Shannon's entropy or mutual information. Behavioral biologists and evolutionary ecologists more commonly use decision-theoretic measures, such as the value of information, which assess the worth of information to a decisionmaker. Here, we show that these two kinds of measures are related intimately in the context of biological evolution. We present a simple model of evolution in an uncertain environment and calculate the increase in Darwinian fitness made possible by information about the environmental state. This fitness increase is a composite of both Shannon's mutual information and the decision-theoretic value of information.

## 313 Biometrics Editor Invited Session

ENAR, WNAR

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Are Flexible Designs Sound?

❖ Carl-Fredrik Burman, AstraZeneca Pharmaceuticals, R&D, Mölndal, SE-431 83 Sweden, *Carl-Fredrik.Burman@astrazeneca.com*; ❖ Christian Sonesson, AstraZeneca Pharmaceuticals, AstraZeneca R&D, Mölndal, SE-431 83 Sweden, *Christian.Sonesson@astrazeneca.com*

***Key Words:*** adaptive design, sample size reestimation, sufficiency, inference principles, clinical trial

Some adaptive designs are controversial. Sample size re-estimation (SSRE) is one, in particular when based on interim efficacy estimates. The idea is attractive: why not modify the sample size in light of new information? With a weighted test, the type I error level can be protected. However, this test is not based on the sufficient statistic. The unblinded SSRE has been criticized as being less efficient than certain group-sequential designs. Burman & Sonesson argue that the violation of the sufficiency principle makes the inference invalid. They claim the weighted test may lead to paradoxical results: The null hypothesis of zero mean can be rejected, although the average response is zero. This session will contain a presentation and invited discussion of Burman & Sonesson's paper. The discussion will appear in Biometrics.

## 314 Challenges for Early Detection of Cancer Using Genomic or Proteomic Profiles ●

General Methodology, Biometrics Section, WNAR, ENAR

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Analytic Opportunities: Moving New Diagnostics to the Clinic

❖ Robert Lipshutz, Affymetrix, 3280 Central Expressway, Santa Clara, CA 95051, *Rob_Lipshutz@affymetrix.com*

***Key Words:*** microarray, algorithms

High-density microarrays are providing the ability to interrogate genetic information at unprecedented levels of resolution and scale. Types of information range from genes and RNA expression to polymorphisms and methylation. Access to such data in well-characterized cohorts provides the opportunity to develop molecular signatures correlated with or predictive of specific medical conditions. These represent new diagnostic paradigms based on the detection and interpretation of complex patterns of multiple parameters. The rigorous validation of these patterns with the objective of developing commercial tests for widespread clinical application presents opportunities for the development of standards for evaluating complex methods and algorithms. Issues will be illustrated with a brief summary of emerging microarray and disease-specific applications required for commercialization.

### Platforms for the Early Detection of Cancer

❖ Andrew Quong, Thomas Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107, *andrew.quong@jefferson.edu*

***Key Words:*** microarry, mRNA, markers, cantilever

With the advent of omic-technologies such as microarrays and LC-MS, we are beginning to see a shift from the paradigm of a single marker for disease to that of a bio-signature. At the same time, systems biology approach for understanding the many levels of biological information available is shedding new light on the molecular mechanisms of cancer. These studies will play an important role in elucidating relevant pathways from which several markers can be identified as a signature for early disease detection. Using these results in a clinical setting presents a whole set of challenges ranging from the development of statistical methods to producing highly-multiplexed assays for measuring gene and protein expression. I will discuss our work in developing new computational algorithms and a detection platform based on microcantilevers.

### Comparing Adaptive and Nonadaptive Algorithms for Cancer Early Detection with Novel Biomarker

❖ Martin W. McIntosh, Fred Hutchinson Cancer Research Center, M2-B230, P.O. Box 19024, Seattle, WA 98109-1024, *mmcintos@fhcrc.org*

Serum early detection biomarker discovery and validation commonly begin by comparing biosamples from cases and controls ascertained in cross-sectional studies. Translating those findings to the clinic consumes multiple steps of validation and, at each step, which decisions must be made regarding which candidate biomarkers should be selected for future development. One approach is to use sophisticated mathematic procedures to populate the panel with an optimal set of markers chosen based on how they behave together in these cross-sectional studies. We present an argument for foregoing this optimization early in the development and discovery process, as choices made early may have little relevance downstream and lead to sub-optimal performance.

## 315 Statistical Careers in Government Agencies ● ✪

Section on Government Statistics, Section on Statistical Education

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Growing Challenges Facing Federal Statisticians and Agencies

❖ William G. Barron, Princeton University, Princeton University Survey Center, 8110 Gatehouse Road, Falls Church, VA 22042, *William.Barron@ngc.com*

***Key Words:*** civic engagement, democracy, public interest

The 2006 JSM theme, Statisticians for an Uncertain World: Meeting Global Challenges, underscores the fact that statisticians can make critical contributions by applying the expertise and standards of their profession to issues that confront modern society. Employment in the federal government affords many such opportunities to work on problems of great importance. However, it appears that the work of federal statistical agencies is not widely understood, appreciated, or respected. Already facing severe resource constraints, recent developments portend further declines in public support, interest, and participation. Can statistical agencies offset growing lack of confidence in government, coupled with indifference, anger, and concerns over privacy? This paper cites recent examples of public disengagement and suggests further steps be taken to support statistical agencies and programs.

### Recruitment, Training, and Retention of Government Statisticians

❖ Clyde Tucker, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 1950, Washington, DC 20212, *tucker.clyde@bls.gov*

***Key Words:*** career paths, statistical science, statistical technology, succession planning, mentoring

This paper will explore several practical issues in the recruitment, training, and retention of government statisticians. These issues include bureaucratic barriers to recruitment, successful recruitment methods, salary and lifestyle considerations, establishment of career goals, mentoring, lifelong learning, methods to improve retention, and succession planning.

### Meeting the Challenges of Science at the Frontiers: Statistics in a Government Science Agency

❖ Nell Sedransk, National Institute of Statistical Sciences, 19 TW Alexander Drive, Chevy Chase, MD 27709-4006, *sedransk@niss.org*

***Key Words:*** interdisciplinary research, high-dimensional data, complex systems, applied statistics, statistical models, probability modeling

Several common beliefs about statistics in government diametrically oppose the reality for science agencies. Science at the frontiers---in a federal agency, research institute, academia, or industry---is high-tech, computationally intensive work by teams with deep scientific knowledge and equally deep reservoirs of intelligence. The statistical challenges now arise from complexity of the science, high-dimensionality of data, massiveness of databases, and intricacies of interrelationships among variables and observations. Successful science depends on integration of statistical thinking into the research; successful statistics is driven by the science to achieve accurate statistical formulation of concepts at the core of the research and apply analytic tools that penetrate the complexity of the data. Examples illustrate the statistician's role as an interdisciplinary research team member.

## 316 Multilevel Modeling of Complex Survey Data ●

Section on Survey Research Methods, Section on Health Policy Statistics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Multilevel Modeling with Multistage Survey Samples

❖ Tihomir Asparouhov, MuthÈn & MuthÈn, 3643 Stoner Ave., Los Angeles, CA 90066, *tihomir@statmodel.com*; Bengt Muthen, University of California, Los Angeles

***Key Words:*** multistage designs, multilevel models, multiple group models, Mplus, LRT adjustment

We describe a multivariate, multilevel, pseudo maximum--likelihood estimation method for multistage stratified cluster sampling designs, including finite population and unequal probability sampling. We describe sample design-based adjustment of the likelihood ratio test statistic that has approximate chi-square distribution. Simulation studies are conducted to evaluate the performance of the proposed estimator and test statistic. We illustrate this technique with a multilevel multiple group model and describe some of the complications that arise when the grouping variable is not nested within the sampling design.

### Adjusting for Unequal Selection Probability in Multilevel Models: a Comparison of Software Packages

❖ Kim Chantala, The University of North Carolina at Chapel Hill, Carolina Population Center, University Square, CB #8120, 123 Franklin Street, Chapel Hill, NC 27516-3997, *kim_chantala@unc.edu*; C. M. Suchindran, The University of North Carolina at Chapel Hill

***Key Words:*** multilevel models, sample weights, software

Most surveys collect data using complex sampling plans that involve selecting both clusters and individuals with unequal probability of selection. Research in using multilevel modeling techniques to analyze such data is relatively new. Often, sampling weights based on probabilities of selecting individuals are used to estimate population-based models. However, sampling weights used for estimating multilevel models (MLM) need to be constructed differently than weights used for population-average models. This paper compares the capabilities of MLwiN, Mplus, LISREL, PROC MIXED (SAS), and gllamm (Stata) for estimating MLM using data collected with a complex sampling plan. We illustrate how sampling weights need to be constructed for estimating MLM with these software packages. Finally, we contrast the results from these packages using data collected with a complex sampling plan.

### Bootstrapping for Variance Estimation in Multilevel Models Fitted to Survey Data

❖ Milorad Kovacevic, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A0T6 Canada, *kovamil@statcan.ca*; Rong Huang, Statistics Canada; Yong You, Statistics Canada

***Key Words:*** informative design, multi-stage design, survey weights, variance components

For estimation of parameters of a multilevel model fitted to hierarchical survey data, typically the survey weights and their bootstrap replicates are available only for the level-one units, thus making impossible implementation of the estimation methods that require weights at the higher levels. Some approximations are needed. In a simulation study designed according to the Canadian Workplace and Employee Survey, we study the impact of this limitation on the properties of mixed-model parameters estimates and their standard errors estimated by the survey bootstrap method. Several existing methods are compared and alternatives suggested.

## 317 The Use of Quality Control Charts in Biosurveillance and Prospective Public Health Surveillance ● ✪

Section on Statistics in Epidemiology, ENAR
**Tuesday, August 8, 2:00 pm–3:50 pm**

### The Use of Control Charts in Health Care Monitoring and Public Health Surveillance

❖ William H. Woodall, Virginia Polytechnic Institute and State University, Department of Statistics, Blacksburg, VA 24061-0439, *bwoodall@vt.edu*

***Key Words:*** statistical process control, cusum, health care quality, prospective disease cluster detection

There are many applications of control charts in health care monitoring and public health surveillance. We introduce these applications to industrial practitioners and discuss applicable ideas in industrial monitoring. The advantages and disadvantages of the charting methods proposed in the health care and public health areas are considered. Additional contributions in the industrial statistical process control literature relevant to this area are given. There are many application and research opportunities available in the use of control charts for health-related monitoring.

### Cumulative Sum Methods for Spatial Surveillance

❖ Peter A. Rogerson, University of Buffalo, Department of Geography, Wilkeson Hall, Buffalo, NY 14226, *rogerson@buffalo.edu*

***Key Words:*** cumulative sum, resels, Gaussian random fields, smoothing

Often, there is interest in monitoring health within a study region where data are available for a number of subregions. One way to carry out monitoring is to maintain separate cumulative sum charts over time for each region. A drawback of this approach is that it does not account for the possibility of clusters occurring on larger spatial scales. In this paper, I describe how monitoring may be carried out for neighborhoods constructed around each subregion. Separate charts may be kept for each subregion and its surrounding neighborhood. However, adjusting cusum thresholds for the number of subregions is conservative, as nearby subregions will have correlated charts. Here, these correlations are accounted for; an adjustment for the number of effectively independent charts is made using the theory of smoothed Gaussian random fields, and the approach is evaluated.

### Performance of Residual-Based Control Charts Using Generalized Exponential Smoothing of Syndromic Data for Routine Health Surveillance

❖ Howard S. Burkom, Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, *Howard.Burkom@jhuapl.edu*; Galit Shmueli, University of Maryland; Sean Murphy, Johns Hopkins University Applied Physics Laboratory

***Key Words:*** process control, syndromic surveillance, exponential smoothing, regression control charts

Syndromic surveillance depends on the monitoring of consumer data sources for early warning of unspecified disease outbreaks. These sources include clinical data, such as counts of emergency department visits, and nonclinical data, such as over-the-counter remedy sales. For sensitivity and timeliness at practical alert rates, developers have tried to adapt chart-based methods of statistical process control. Obstacles are evolving, often nonstationary input data streams, target signal uncertainty, and cyclic or seasonal background data behavior. Reliable performance often requires a combination of modeling and process control. This presentation applies generalized exponential smoothing to make forecasts and measures predictive accuracy and the performance of residual-based charts. This approach is compared to conventional methods for sensitivity and robustness using several data types.

### A Neural Network Approach to Control Charts with Applications to Health Surveillance

❖ Benjamin M. Adams, The University of Alabama, ISM Department, Box 870226, Tuscaloosa, AL 35487-0226, *badams@cba.ua.edu*; Kidakan Saithanu, The University of Alabama; J. Michael Hardin, The University of Alabama

***Key Words:*** neural networks, public health surveillance, quality control

An evaluation of multivariate monitoring scheme performances is provided. Monitoring schemes include traditional multivariate techniques, such as the MEWMA and T2 control charts, and a technique using neural networks. The neural network approach is shown to be a viable monitoring technique for complex multivariate problems, such as those found in health surveillance.

## 318 Technometrics Invited Session ●

Technometrics, Section on Physical and Engineering Sciences
**Tuesday, August 8, 2:00 pm–3:50 pm**

### Simultaneous Variable Selection

❖ Berwin A. Turlach, The University of Western Australia, School of Mathematics and Statistics (M019), 35 Stirling Highway, Crawley, WA, 6009 Australia, *berwin@maths.uwa.edu.au*; William N. Venables, CSIRO Mathematical and Information Sciences; Stephen J. Wright, University of Wisconsin-Madison

***Key Words:*** subset selection, multivariate regression, variable selection, interior point algorithm, solution path algorithm

We propose a new method for selecting a common subset of explanatory variables where the aim is to model several response variables. The idea is a natural extension of the LASSO technique and based on minimizing the (joint) residual sum of squares while constraining the parameter estimates to lie within a suitable polyhedral region. We briefly comment on other constraints that may be imposed to achieve simultaneous variable selection. The properties of the convex programming problem resulting from our approach are analyzed for the special case of an orthonormal design. For the general case, we develop an efficient interior point algorithm and also describe an algorithm that calculates the complete solution path. Time permitting, the method will be illustrated on various datasets.

### A New Strategy for Variable Selection

Xiaohui Luo, Merck & Co., Inc.; Leonard A. Stefanski, North Carolina State University; ❖ Dennis A. Boos, North Carolina State University, Department of Statistics, Campus Box 8203, Raleigh, NC 27695-8203, *boos@stat.ncsu.edu*

***Key Words:*** false selection rate, forward selection, mean squared error, noise addition, phony variables, SIMEX

We describe an approach to variable selection wherein we calibrate a tunable selection method to achieve desirable properties of models selected. We focus on forward selection where the tuning parameter is alpha-to-enter, aka SLENTRY to SAS users. Calibration is achieved by adding “noise” to the data and tracking its effect on the models selected. In one version of the strategy, parametric bootstrap-like data sets are generated by adding Gaussian noise to the response variable. Then SLENTRY is tuned by tracking the effect of the added noise on the MSEs of models selected for different SLENTRY values. An alternative means of “adding noise” to the data gives rise to a second method wherein random phony predictor variables are appended to the data, and SLENTRY is tuned by tracking the proportion of falsely-included phony variables in the models selected for different SLENTRY values.

## 319 Statistical Challenges in Natural Resource Management ● ✪

Section on Statistics and the Environment

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Bayesian Analysis of Animal Community Structure

❖ Jeffrey A. Royle, U.S. Geological Survey/Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708, *aroyle@usgs.gov*

***Key Words:*** species richness, avian surveys, biodiversity, capture-recapture, animal sampling, hierarchical models of occurrence and abundance

Models of animal community structure are based on surveys of species’ presence or absence on a sample of spatial units. One consideration in conducting inference about community structure is that species are detected imperfectly. This leads to fewer species observed in the sample than exist in the community. Classical methods for modeling community structure do not preserve species identity, and thus don’t allow the development of predictive models of community composition. We describe a strategy based on species-specific models of occurrence, from which estimates of important summaries of community structure are derived by aggregating indicators of occurrence for all species observed or estimated to be in the community. We use a data augmentation approach to develop an efficient Bayesian procedure for estimation and prediction under this model using MCMC.

### A Vision for Forest Inventory

❖ Ronald E. McRoberts, U.S. Department of Agriculture, North Central Research Station, 1992 Folwell Avenue, St. Paul, MN 55108, *rmcroberts@fs.fed.us*

***Key Words:*** forest inventory, maps

In response to the traditional user question, “How much?”, the Forest Inventory and Analysis (FIA) program of the USDA Forest Service annually reports design-based estimates of forest attributes for counties across the country. In response to the more recent “Where?” question, the FIA program makes sample data available via the internet, which is an adequate solution for large AOIs and applications that do not require exact plot locations. However, the vision for forest inventory has two components: unbiased, multivariate maps of sufficient spatial resolution for user applications and interactive access to the maps via a medium that does not try users’ patience excessively. The presentation focuses on current techniques for realizing the vision, including nearest neighbor techniques, internet access, and fast, approximating algorithms.

### Challenges in Modeling Associations between Streams and Their Watersheds

❖ John Van Sickle, U.S. Environmental Protection Agency, 200 SW 35th, Corvallis, OR 97333, *VanSickle.John@epa.gov*

***Key Words:*** watershed land use, area of influence, flow path distance

Regressions between streams and their watersheds’ land uses confront several challenges, including complex spatial correlation structures of stream networks, highly multivariate responses for stream biology, and uncertain regressors derived from digital maps. Another challenge is to identify the areas within watersheds whose land uses best predict stream responses. GIS tools can be used to carve out candidate areas of influence, and the best-predicting candidate might be revealed by variable selection methods. Alternatively, I suggest summing the influences of all cells in the watershed’s land use grid coverage for use as a regressor. Cell influence is modeled as a decaying function of cell distance upstream and upslope, and estimated influence decay rates describe the watershed areas of greatest influence. The model is illustrated for 52 small watersheds in Oregon’s Willamette Valley.

### What Is a Realistic Fisheries Stock Assessment Model?

❖ Terrance Quinn, University of Alaska, P.O. Box 210585, Fisheries, Auke Bay, AK 99801, *Terry.Quinn@uaf.edu*

***Key Words:*** fishery, model, mortality, abundance

A fisheries stock assessment model is a model of the population dynamics of a harvested fish species that integrates information from the fishery with scientifically collected biological and survey data. Model parameters include cohort abundance, natural and fishing mortality, and calibration coefficients. A realistic model contains those biological and human factors that play a major role in population dynamics. At a minimum, a historical dataset comprised of total removals, age or size information, and a relative index of population abundance is necessary. Current models provide for realistic changes in cohort abundance and fishing mortality over time. More realistic models are needed for temporal variations in natural mortality, either through covariates (e.g., disease and environment) or explicit incorporation of multispecies interactions (e.g., predator-prey).

# 320 Beyond Your Parents' Models: Latent Variables as You May Not Yet Have Thought of Them ●

Biometrics Section, ENAR
**Tuesday, August 8, 2:00 pm–3:50 pm**

## Latent Variable Mixture Modeling with Genetic Applications

❖ Bengt Muthen, University of California, Los Angeles, Box 951521, Los Angeles, CA 90095, *bmuthen@ucla.edu*

***Key Words:*** latent class hybrid analysis, QTL, IBD, twins, siblings

Modeling of data from twins, siblings, and other family members rely on a well-measured phenotype. In many applications, the phenotype is best characterized as a latent variable measured by a set of observed, fallible indicators. This presentation will propose a factor mixture (finite mixture) model with latent classes representing different response profiles corresponding to different subtypes of disorders and continuous variables representing severity variation within type. Genetic influence on both class membership and severity will be considered. Maximum-likelihood estimation with categorical outcomes using EM and numerical integration will be discussed in the Mplus framework. Applications include sibling data concerning ADHD problems. The sibling data analysis uses quantitative trait locus analysis with genetic correlations varying by IBD proportion.

## Latent Variable Modeling

❖ Katherine Masyn, University of California, Davis, Department of Human and Community Development, One Shields Ave, Davis, CA 95616, *kmasyn@ucdavis.edu*; Tihomir Asparouhov, MuthÈn & MuthÈn; Bengt Muthen, University of California, Los Angeles

***Key Words:*** Cox regression, latent variables, frailty models, multilevel models, survival time models, mixture models

We describe a general multivariate, multilevel framework for continuous time survival analysis that includes joint modeling of survival time variables and continuous and categorical observed and latent variables. The proposed framework is implemented in the Mplus software package. The survival time variables are modeled with nonparametric or parametric proportional hazard distributions and include right censoring. The proposed modeling framework is flexible and includes many advanced survival time models, such as finite mixtures of Cox regression models with and without class-specific baseline hazards, multilevel Cox regression models, and single and multilevel frailty models. We illustrate this framework by modeling school dropout data. Comparison is made with discrete time survival models.

## Semiparametric Bayesian Structural Equation Models

❖ David B. Dunson, National Institute of Environmental Health Sciences, NIEHS, Biostatistics Branch, MD A3-03, P.O. Box 12233, Research Triangle Park, NC 27709, *dunson@stat.duke.edu*

***Key Words:*** nonparametric, latent variable, structural equations model, random effects, Dirichlet process, Bayesian

Structural equation models (SEMs) with latent variables are used widely in the social sciences and increasingly in epidemiology. SEMs characterize relationships in multivariate data through a measurement model and a latent variable, or structural, model. Typically, Gaussian linear models are used for both components. Because latent variables are not observed directly, model assessment can be challenging in SEMs, and sensitivity to parametric assumptions is a concern. We propose a Bayesian semiparametric approach based on nonparametric priors for the unknown latent variable and measurement error distributions. The proposed framework allows the unknown distributions to change flexibly with predictors using a new class of dependent nonparametric priors based on kernel-weighted mixtures of Dirichlet processes.

## Penalized Latent Class Regression: Incorporating Scientific Knowledge into Measurement Models

❖ Jeannie-Marie Sheppard, Johns Hopkins Bloomberg School of Public Health, 550 N. Broadway, Suite 308, Baltimore, MD 21287, *i.dream.of.jeannie@gmail.com*; Karen Bandeen-Roche, Johns Hopkins Bloomberg School of Public Health; Peter Zandi, Johns Hopkins Bloomberg School of Public Health; William Eaton, Johns Hopkins Bloomberg School of Public Health

***Key Words:*** latent variable, measurement, validity, penalization, depression, nosology

A primary challenge in psychiatric epidemiology is nosology. To establish an association between any risk factor and a disorder, one must define that disorder first. Building on work by Houseman and colleagues, we will use a penalty in the fitting of latent class regression (LCR) models to produce depression subtypes based on polymorphisms of the serotonin transporter (5-HTT) gene and individuals' patterns of symptoms. While the relationship between this functional polymorphism and depression has not been elucidated, serotonin has been implicated in affective disorders, and we therefore hypothesize genotype may refine depression classification usefully. We will discuss methods to fit these models and report on their application to the subtyping of depression. Ultimately, we aim to augment existing tools to incorporate scientific knowledge into measurement modeling.

## Principal Stratification Designs To Estimate Outcomes Missing Due to Death

❖ Constantine Frangakis, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 20723, *cfrangak@jhsph.edu*; Donald Rubin, Harvard University; Ming-Wen An, Johns Hopkins University

***Key Words:*** causal inference, censoring by death, missing data, potential outcomes, principal stratification

We address studies of individuals after a critical event, such as injury, to measure outcomes about the period preceding the event. These studies aim to characterize the outcomes and predict the critical events based on these outcomes. The outcomes, however, are missing for the individuals who die after the critical event. Standard methods cannot estimate such missing outcomes accurately before death if these outcomes differ from observed strata. To address these aims, we propose a design using data on externally controlled factors that could prevent deaths. Based on this design, we develop a method that allows the missing outcomes to differ from any observed strata and still estimate the missing outcomes using the framework of principal stratification. Using injury data, we show that our design and method can uncover dramatically different results from those of standard methods.

## 321 Innovative Uses of the Survey of Income and Program Participation for Social Program Evaluation ● ✪

Social Statistics Section

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Assessing the Effect of Allocated Data on the Estimated Value of Total Household Income in the Survey of Income and Program Participation (SIPP)

❖ Patricia Fisher, U.S. Census Bureau, HHES Division, mail stop 8500, Washington, DC 20233, *patricia.j.fisher@census.gov*

***Key Words:*** imputation, data quality, SIPP

In the Survey of Income and Program Participation (SIPP), item nonresponses are allocated using a hot-deck procedure, or, in some instances, are allocated logically using available information. These allocated values then are used in creating aggregate amounts, such as the value of total household income. In the SIPP, total household income is the sum of all income amounts reported or allocated by household members and includes wages and salaries earned, dividends, rental income, social security payments, income received from TANF/AFDC, SSI, and other. Any or all of these income sources could have been allocated, and, to SIPP users, it is not well-understood how these allocated amounts affect the created value of total household income. Using the 2001 SIPP panel, this paper looks at the individual components of total household income and discusses the proportion allocated for each.

### Welfare Reform Revisited: Leavers and Cyclers---How Are They Doing under the New Restrictive Time Limits?

❖ Carole L. Popoff, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-8500, *carole.l.popoff@census.gov*; Qi Wang, U.S. Census Bureau

***Key Words:*** multi-level modeling, survival analysis, welfare reform, welfare leavers

Following 1996, welfare caseloads declined by 50--60%. Using the Survey of Income and Program Participation 2001 panel, we will investigate whether leaving and/or returning to TANF is associated with changes in demographic characteristics, family structure, employment opportunities, and educational attainment using survival analysis (Popoff and Wang 2005). Also, we will add variables that capture general economic conditions, state-specific participation rules, and other variables that may have influenced welfare leavers' actions by employing a generalized linear mixed model (GLMM)---a multi-level approach---to the original survival analysis. This will add a broader perspective from which to determine factors at both the individual or family level and the higher aggregate level where economic and other conditions may influence behavior.

### Assessing Estimates of Program Participation: an Analysis of Matched SIPP and Administrative Data

❖ Scott Cody, Mathematica Policy Research, Inc., Suite 550, 600 Maryland Ave SW, Washington, DC 20024, *scody@mathematica-mpr.com*; Julie Sykes, Mathematica Policy Research, Inc.

***Key Words:*** SIPP, administrative data, matched records, poverty, social security, SSI

This paper examines the reliability of SIPP-based estimates of participation in the Social Security and Supplemental Security Income (SSI) programs. Using data from the 1996 SIPP panel matched to administrative data from the Social Security Administration, we examine how individuals' program participation status and benefits amounts---as reported in the SIPP---differ from their administrative records. We find a high frequency of inconsistencies between SIPP and administrative data, but many are small in terms of dollar amount. Imputation in the SIPP explains many, but not all, of the inconsistencies. We examine how the frequency and type of inconsistencies vary by age and income. The findings have important implications for efforts to use SIPP to measure poverty and simulate eligibility for programs such as the Food Stamp Program.

### Accounting for SIPP Design Effects: Testing Four Approaches with the 2004 Panel

❖ Arthur F. Jones, U.S. Census Bureau, 3831 Eisenhower Ave., Alexandria, VA, VA 22304, *arthur.jones.jr@census.gov*; Robert Nielsen, U.S. Census Bureau; Michael Davern, University of Minnesota; John Boies, U.S. Census Bureau

***Key Words:*** SIPP, variance, survey methods, replicate weights

A review of SIPP-based research suggests many users do not account for the survey's complex sampling design when estimating the variances of point estimates. This practice leads to inappropriate conclusions about the precision of variance estimates. We estimate the variance of several demographic and economic characteristics (e.g., age, race, ethnicity, public and private health insurance rates, employment rates, program participation, and poverty rates). The variances are estimated using four unique estimators: simple random sample approach, generalized variance parameter approach, a Taylor Series estimator, and a replicate weight approach. The results show not accounting for sample design will lead to biased variance estimates. We conclude by offering specific advice to the SIPP data user community about the best way to calculate standard appropriately.

## 322 POC/Adaptive Design

Biopharmaceutical Section

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Impact of Pharmacometrics Reviews on Drug Approval and Labeling Decisions

❖ Joga Gobburu, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Pharmacometrics, OCPB, Bldg 21, Rm. 4524, Silver Spring, MD 20993-002, *jogarao.gobburu@fda.hhs.gov*

***Key Words:*** PKPD modeling, simulation, FDA, drug development

Modeling and simulation of data pertaining to pharmacokinetic, pharmacodynamic, and disease progression often is referred to as the pharmacometrics analyses. The objective of the current report is to assess the role of pharmacometrics, at FDA, in making drug approval and labeling decisions. The NDAs submitted between 2000--2004 to the cardio-renal, oncology, and neuropharmacology drug products divisions were surveyed. Of about 244 NDAs, 42 included a pharmacometrics component. Pharmacometric analyses were pivotal in regulatory decisionmaking in more than half of the 42 NDAs. Of the 14 reviews that were pivotal to approval-related decisions, five identified the need for further trials, while six reduced the burden of conducting additional trials. The survey and case studies emphasize the need for early interaction between the FDA and sponsors.

### Adaptive by Design in Exploratory Development

❖ Brenda Gaydos, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *blg@lilly.com*

***Key Words:*** adaptive designs, early phase development, proof of concept, maximum tolerated dose, dose response

Adaptive designs can facilitate earlier and better data-driven decisions, with increased efficiency, improving the probability of technical success for compounds moving forward. Traditional early-phase studies often are fairly flexible, but lack rigor. Details on design changes and data analysis may be determined after the study is ongoing. This approach can minimize the timeline for protocol development; however, changes on the fly complicate implementation and interpretation. Adapting by design (prespecifying the adaptive mechanism) still allows for some flexibility while providing an increase in information in a more optimally efficient way. This presentation will provide an overview of adaptive design approaches (advantages and disadvantages) in early exploratory development. Clinical trial examples will be presented with a discussion of practical barriers encountered.

### Comparison of Bayesian and Frequentist Adaptive Designs for Combination Proof-of-Concept/Dose Response Modeling Trials in Phase II

❖ Nitin Patel, Cytel Inc., 675 Massachusetts Ave., Cambridge, MA 02139, *nitin@cytel.com*; James Bolognese, Merck Research Laboratories; Jerald Schindler, Cytel Inc.; Scott Berry, Berry Consultants; Yannis Jemiai, Cytel Inc.; Vipul Suru, Cytel Inc.

Traditional designs for phase II proof-of-concept trials are inefficient in allocating doses to patients, as they often assign subjects to doses not of interest. Dose-adaptive designs use accumulating information on responses to determine sequentially which dose should be allocated next. Such adaptive designs efficiently provide information to assess efficacy and characterize dose response. Various frequentist and Bayesian adaptive designs have been proposed in the literature. Direct comparison between these methods is difficult, as no simple solutions to the power/sample size calculations are available. We will present results from a methodical comparison of the performance of these designs using computer simulation experiments. Our objective is to provide insights to guide the deployment of adaptive designs in practical settings.

### Bayesian Modeling of Safety Data Using Databases of Placebo Patients

❖ Jose Pinheiro, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, *jose.pinheiro@novartis.com*; Chyi-Hung Hsu, Novartis Pharmaceuticals Corporation

***Key Words:*** early development, proof-of-concept, mixed-effects models, CRM, adverse events

Evaluation of safety signals in early clinical development trials is often a difficult statistical problem because of the small number of patients involved and the short study duration typically observed. To make matters more challenging, safety variables often are measured in dichotomized form (e.g., occurrence of lab parameter elevations beyond a clinically relevant threshold). We will discuss and illustrate the use of Bayesian mixed-effects models to represent longitudinal continuous safety data and combine them with information from databases of placebo patients to implement a CRM-like algorithm to identify safety signals in early development data more efficiently. Simulation results, based on real clinical trial data, will be used to illustrate the proposed methods.

## 323 Quality Initiatives in Establishment Surveys ●

Section on Survey Research Methods

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A Methodology for Evaluating Sufficiency of Survey Frames

❖ Howard Bradsher-Fredrick, Energy Information Administration, 1000 Independence Ave., EI70, Washington, DC 20585, *hbradshe@eia.doe.gov*

***Key Words:*** survey frame, establishment survey, data quality, performance measures

The Energy Information Administration (EIA) has continued its strategic planning efforts to maintain and improve data quality and its movement toward employing a greater number of performance measurements. As part of that effort, EIA evaluated its own survey frames for "sufficiency." This effort involved the development of a set of evaluation criteria that called for collecting a large quantity of information and data for 34 EIA master frames. An inter-office team developed the evaluation criteria and was responsible for the data and information collection. They also applied the criteria to evaluate the 34 master frames. The team eventually decided all but four of the frames were sufficient; the insufficient frames were deemed "insufficient" for a variety of reasons.

### An Evaluation of the 2002 MECS Measure of Size

❖ William Gifford, Energy Information Administration, 1000 Independence Ave., SW, Office of Energy Markets and End Use, Washington, DC 20585, *william.gifford@eia.doe.gov*

***Key Words:*** energy, measure of size, establishment, survey, manufacturing, evaluation

The Manufacturing Energy Consumption Survey (MECS) is an establishment survey that collects data on energy consumption, expenditures, and other energy-related topics from the manufacturing sector. The MECS frame is based on the business register, which is managed by the U.S. Census Bureau. The MECS takes advantage of the establishment-level information that comes with sharing a frame with the manufacturing section of the Economic Census. While the MECS is fortunate to have a measure of size assigned to each element in its frame, based on data recorded from those actual elements, statistical analysis is needed to evaluate how the measure of size compared with the energy consumption measures collected in the 2002 MECS. In this paper, the MECS measure of size will be compared with consumption measures graphically and with statistical modeling techniques.

### Quality Issues in a Regulatory Data Collection System

❖ Alan K. Jeeves, Bureau of Transportation Statistics, 400 7th Street, SW, Room 3430, Washington, DC 20590, *alan.jeeves@dot.gov*

***Key Words:*** data quality, editing, imputation, regulatory data

The Federal Aviation Act requires all air carriers to report traffic and financial information to the DOT. Although these data collection systems originally supported airline regulation by the Civil Aeronautics Board (CAB), the government and private sector still use these data to monitor industry competition and financial condition. A new impetus for data quality occurred with the recent enactment of the Rural Service Improvement Act, which established BTS airline data as the basis for determining carrier eligibility and payment rates for intra-Alaskan mail transport. Our research showed that although the scope of the edit procedures could be broadened, the main problem was that the number of records flagged exceeded the available resources to fix them. Maximizing data quality requires recognition that since not all errors can be fixed, the focus must be on the most egregious.

# 324 Statistical Methods on Networks and Graphs ✪

Section on Statisticians in Defense and National Security, Section on Statistical Graphics

**Tuesday, August 8, 2:00 pm–3:50 pm**

## Graphs for Streaming Text

❖ Elizabeth Hohman, Naval Surface Warfare Center, Code B10, 17320 Dahlgren Road, Dahlgren, VA 22448, *elizabeth.hohman@navy.mil*

***Key Words:*** streaming text

There are several obstacles to analyzing streaming text data that do not arise when the corpus of documents is static. When employing a vector space model for text processing, word weights usually are used that depend on the frequency of the word in the document and the frequency of the word in the corpus. Such word weighting must be revised to process a streaming collection of documents. This presentation discusses methods for word weighting streaming text and methods for representing a changing corpus with a dynamic graph. An example corpus is used that contains daily news articles from five categories. Graphs are used to represent the streaming corpus. Statistics on the graphs are calculated in order to determine changes in the corpus.

## Graph Theoretic Methods for the Identification of Article Associations

❖ Jeffrey L. Solka, Naval Surface Warfare Center, Code B10, 17320 Dahlgren Road, Dahlgren, VA 22448, *jeffrey.solka@navy.mil*; Avory Bryant, Naval Surface Warfare Center; Nicholas Tucey, Naval Surface Warfare Center

***Key Words:*** document, clustering, graph, network, associations

This talk will discuss recent work in the use of graph theoretic measures for the identification of article associations between document collections. It will examine the use of minimal spanning trees and article clustering methodologies to identify interesting associations between articles. The approach will be demonstrated on small collections of various article sets.

## Anomaly Detection in Genetic Networks

❖ Christopher Overall, George Mason University, 5316 Satterfield Drive, Woodbridge, VA 22193, *coverall@gmu.edu*; Jeffrey L. Solka, Naval Surface Warfare Center; J. W. Weller, George Mason University; Carey Priebe, Johns Hopkins University

***Key Words:*** scan statistics, genetic networks, microarrays, anomaly detection, chatter

This talk will detail our recent work on the application of graph-based scan statistics to genetic networks in order to detect anomalies in gene activity. An anomaly may manifest as a single gene with excessive connections to other genes or it may manifest as "chatter," in which genes centered about a particular gene exhibit excessive connections with one another. This work has been adapted from previous efforts by Priebe et al. (2005) to detect anomalies within the social network developed from the Enron email dataset. Our methodology uses multivariate model-based clustering to create a time sequence of graphs from time-series gene expression datasets. It then applies the graph-based scan statistics methodology to detect anomalies within the graphs for each time point. The procedure will be illustrated using a well-known Drosophila gene expression dataset.

## Network Traffic Profiling

❖ Patricia H. Carter, Naval Surface Warfare Center, Code B21, Building 1470, Room 2311, 17320 Dahlgren Road, Dahlgren, VA 22448, *patricia.h.carter@navy.mil*

***Key Words:*** network traffic, document space, clustering, graph models, anomaly detection, classification

The goal of this work is to characterize normal versus non-normal computer network traffic state at different temporal and spatial granularities. A graph theoretic representation is used in which the nodes are an abstraction of all traffic over a unit of time; this abstraction is the frequencies of groups of network connections of various types. The weight of an edge between two nodes is a measure of similarity based on document similarity measures. We apply techniques from document clustering and classification and scan statistics in this setting.

# 325 Spatial and Spatio-Temporal Bayesian Inference ✪

Section on Bayesian Statistical Science

**Tuesday, August 8, 2:00 pm–3:50 pm**

## Multivariate Spatial Modeling in Bayesian Hierarchical Settings

❖ Sudipto Banerjee, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware Street S.E., Minneapolis, MN 55455, *sudiptob@biostat.umn.edu*

***Key Words:*** geostatistics, Bayesian modeling, multivariate spatial data, cross covariances

Recent advances in Geographical Information Systems (GIS) and the availability of easily accessible databases and software enable statisticians and data analysts to experiment with richer models that capture spatial association in geocoded data. Recently, there has been much interest in the analysis of spatial data concerning multiple variables that arise from different data sources. This talk will present a flexible class of models that arise as space-varying linear transformations of tractable spatial processes. These models yield valid probability models that assign to each variable its own spatial structure. We discuss these methods in the contexts of analyzing multiple cancers over counties in Minnesota with each cancer type having its own spatial structure and for agronomy experiments where spatial association is deemed present at different resolutions.

## Flexible Spatial Modeling for Multivariate Geological Data Using Convolved Covariance Functions

❖ Anandamayee Majumdar, Arizona State University, Department of Math and Statistics, ASU, TEMPE, AZ 85287-1804, *ANANDA@math.asu.edu*; Alan E. Gelfand, Duke University

***Key Words:*** convolution, coregionalization, Fourier transforms, Gaussian spatial process, hierarchical model, spectral density

For analyzing multivariate geostatistical spatial data, there arises the need for flexible explanatory stochastic models. Here, we propose a general constructive approach for building suitable models based on convolution of covariance functions. We begin with a general theorem that asserts under weak conditions, cross convolution of covariance functions provides a valid cross-covariance function. We also obtain a result on dependence induced by such convolution. Because, in general, convolution does not provide closed-form integration, we discuss efficient computation. We then suggest introducing such specification through a Gaussian process to model multivariate spatial random effects within a hierarchical model. We note that modeling spatial random effects in this way is parsimonious relative to, say, the linear model of coregionalization.

### Gaussian Process Models for a Sphere with Application to Faraday Rotation Measures

❖ Margaret Short, Los Alamos National Laboratory, 3262 Walnut Street, Apt. A, Los Alamos, NM 87544, *mbshort@lanl.gov*; Dave Higdon, Los Alamos National Laboratory; Philipp Kronberg, Los Alamos National Laboratory

***Key Words:*** Gaussian process, spatial process, Markov chain Monte Carlo, Faraday rotation measures, error mixture model

Our primary goal is to obtain a smoothed summary estimate of the magnetic field generated in and near to the Milky Way by using Faraday rotation measures (RMs). The ability to estimate the magnetic field generated locally by our galaxy and its environs will help astronomers distinguish local versus distant properties of the universe. Each RM in our dataset provides an integrated measure of the effect of the magnetic field along the entire line of sight to an extragalactic radio source. RMs can be considered prototypical of geostatistical data on a sphere. In order to model such data, we employ a Bayesian process convolution approach that uses Markov chain Monte Carlo for estimation and prediction. Complications arise due to contamination in the RM measurements, and we resolve these by means of a mixture prior on the errors. This is joint work with D. Higdon and P. Kronberg.

### A Comprehensive Spatial-Temporal Analysis of Breast Cancer: First Primary, Second Primary, and Breast Cancer Survival

❖ Song Zhang, M. D. Anderson Cancer Center, 8450 Cambridge Street, Apt. 1175, Houston, TX 77054, *yszhang@wotan.mdacc.tmc.edu*

***Key Words:*** breast cancer, spatial-temporal analysis, Bayesian p-splines, hazard, survival analysis

We propose a comprehensive spatial-temporal analysis of breast cancer, including first primary occurrence, second primary occurrence, and breast cancer survival. We assume a Poisson model for first primary incidence. We are interested in two competing risks: developing second primary occurrence and death from first primary. For each risk, we define a semiparametric age-hazard function that allows us to piece together information from subjects of different ages and construct a much longer hazard curve. The unknown baseline age-hazard curve is modeled by a Bayesian P-spline prior. Random spatial and temporal effects are assumed for the log-rates of the first primary occurrence and for the hazard functions. They are modeled by MCAR priors such that each component is allowed to have a different spatial-temporal pattern. A simulation study and real data analysis are presented.

### Bayesian Nonparametric Mixture Modeling for Spatial Processes

❖ Michele Guindani, M. D. Anderson Cancer Center, *michele@wotan.mdacc.tmc.edu*; Alan E. Gelfand, Duke University; Sonia Petrone, Universit‡ Commerciale Luigi Bocconi

***Key Words:*** Dirichlet process mixing, latent processes, nonGaussian, nonstationary, mixture models

Recent Bayesian modeling of univariate spatial data has considered mixed effect models, where a residual stationary Gaussian spatial effect is assumed. Arguably, we might prefer the flexibility of a nonstationary, non-Gaussian specification. In a nonparametric setting, this can be accommodated by a spatial Dirichlet process (Gelfand, Kottas, and MacEachern 2005). However, a limitation of this class of processes is that the latent factor driving surface selection is defined globally. In this work, we propose a class of random distributions, which enables us to model the effect of the latent factor locally and is an extension of the class of finite-dimensional Dirichlet priors (Iswharan and Zarepour 2000). In a spatial setting, it is characterized as a mixture of Gaussian random fields with spatially varying weights. We detail model fitting and inference and compare with the SDP.

## 326 Least Angle Regression

Section on Statistical Computing

**Tuesday, August 8, 2:00 pm–3:50 pm**

### New Methods and Software for Variable Selection in Regression

❖ Chris Fraley, Insightful Corporation, 1700 Westlake Ave., N., Suite 500, Seattle, WA 98109, *fraley@insightful.com*; Tim C. Hesterberg, Insightful Corporation

***Key Words:*** LAR, least angle regression, LASSO, variable selection, software, prediction

Classification and regression problems with large numbers of candidate predictor variables occur in a variety of scientific fields, increasingly so with improvements in data collection technologies. Goals in model selection include accurate predictions, interpretable models, stability, and avoiding bias. Methods related to least angle regression (LAR) are among the most promising approaches to these issues. We describe our efforts to develop open-source Splus/R software for LAR, lasso and associated methods, in which we aim to establish a collaborative framework for future development and enhancement.

### Penalized Cox Regression Analysis in the High-Dimensional and Low Sample Size Settings with Application to Microarray Gene Expression Data

❖ Jiang Gui, University of Pennsylvania, 635 Blockley Hall, 423 Guardian Dr, Philadelphia, PA 19104, *jgui@cceb.upenn.edu*; Hongzhe Li, University of Pennsylvania

***Key Words:*** LARS, L1 penalized, survival analysis, microarray, high-dimensional

New high-throughput technologies are generating many types of high-dimensional genomic and proteomic data. These data potentially can be used for predicting clinical outcomes and studying gene regulatory subnetworks and interindividual differences in responses to drugs. In practice, however, the number of independent samples is usually small compared to these high-dimensional genomic data. As a result, many standard statistical methods cannot be applied directly or perform poorly in such high-dimension and low sample size settings. In this talk, I will present L1 penalized methods for relating microarray gene expression data to censored survival outcomes. I will demonstrate and evaluate the proposed method using both simulations and applications to real datasets.

### Grouped and Hierarchical Model Selection through Composite Absolute Penalties (CAPs)

❖ Guilherme Rocha, University of California, Berkeley, Department of Statistics, 367 Evans Hall, Berkeley, CA 94720, *gvrocha@gmail.com*; Peng Zhao, University of California, Berkeley; Bin Yu, University of California, Berkeley

***Key Words:*** variable selection, LARS, penalized regression, penalized classification

Recently, much attention has been devoted to model selection in regression and classification by use of a penalty function (Tibshirani 1996). In some cases, one may want to incorporate further natural groupings or hierarchical structures present within the regressors into the selected model. Our goal is to obtain model estimates that approximate the true model while preserving such structures. Letting (Y_i,X_i) be a set of observations of a response Y_i and corresponding explanatory variables X_i, we obtain our model estimates by modeling EY=g(Xb) and jointly minimizing a convex loss function L(b, Y, X) and penalty (CAP) function formed according to the desired structure. The CAP penalty is formed by suitably defining groups G_i, collecting the L_{\gamma_{i}} norm of the coefficients b_{G_i} into a new vector, and computing the norm of this new vector.

### Determination of Regularization Parameter Using L-Curve by LARS-LASSO Algorithm

❖ Leming Qu, Boise State University, 1910 University Drive, Boise, ID 83725, *lqu@boisestate.edu*; Partha Routh, Boise State University

***Key Words:*** l-curve, regularization, ill-posed problems, LARS, LASSO

Regularization is a common technique to obtain reasonable solutions to ill-posed problems. In Tikhonov regularization, both the data-fitting and the penalty terms are in L2 norm. The L-curve is a plot of the size of the regularized solution versus the size of the corresponding residual for all valid regularization parameters. It is a useful tool for determining a suitable value of the regularization parameter. LASSO replaces the L2 norm by L1 norm for the penality term. The LARS algorithm computes the whole path of the LASSO with a computational complexity in the same magnitude as the ordinary least squares. Thus, the L-curve for LASSO can be obtained efficiently by the LARS-LASSO algorithm. The tuning point of the L-curve is chosen as the value of the regularization parameter. We compare L-curve method with existing methods, including GCV and C_p.

## 327 Proper Compensation for Statistical Consulting Services Provided in a University Setting

Section on Statistical Consulting, Section on Statistical Education

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Successes (and Challenges) in Funding Campus Consulting Facilities

❖ Murray Clayton, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, *clayton@stat.wisc.edu*

***Key Words:*** consulting, biostatistics, biometry

Currently, the University of Wisconsin-Madison has two major consulting groups associated with the Department of Statistics: one in biostatistics and one in the College of Agricultural and Life Sciences (CALS). The funding model for the CALS facility was based on having several jointly appointed faculty and funding “off the top” for graduate student and other expenses. This model has been in existence for more than 25 years, although recent events have necessitated changes. I will describe the historical model, discuss the changes we have implemented, and try to forecast what might happen in the future.

### An Overview of The Ohio State University Statistical Consulting Service

❖ Christopher Holloman, The Ohio State University, 1958 Neil Ave., Columbus, OH 43210, *holloman@stat.ohio-state.edu*

***Key Words:*** consulting, university, management, funding, training

Recently, The Ohio State University Statistical Consulting Service (SCS) was reorganized under the direction of Thomas Bishop and Christopher Holloman. The SCS organization follows management models found in corporate and private consulting firms with modifications to accommodate an academic climate. The primary focus of the SCS is to support OSU faculty and graduate students in their research efforts. In addition, the SCS has captured a significant number of external consulting projects. We will discuss the current organizational and management structures, graduate student training, and funding strategy implemented to support the SCS. Finally, we will discuss a recent consulting project that demonstrates the value the SCS has brought to the OSU research community.

### Multiple Solutions to Funding Statistical Consulting within a University

❖ Linda Young, University of Florida, Department of Statistics, PO Box 110339, Gainesville, FL 32611-0339, *LJYoung@ufl.edu*

***Key Words:*** consulting, funding

Four colleges at the University of Florida provide statistical consulting for faculty and students: College of Liberal Arts and Sciences (CLAS), Institute of Food of Agricultural Sciences (IFAS), Public Health and Health Professions (PHHP), and College of Medicine (COM). The approach to funding differs within each college, ranging from free to fully fee-based. Four colleges, four approaches to funding. First, a description of the funding mechanisms within each college will be described. Then, the strengths and weaknesses of each approach will be discussed.

### Compensation Is More Than Money: Life as an Internal Statistical Consultant in a Medical School

❖ Sarah Boslaugh, Washington University in St. Louis, School of Medicine, Department of Pediatrics, 660 South Euclid Campus Box 8208, St Louis, MO 63108, *boslaugh_s@kids.wustl.edu*

***Key Words:*** statistical consulting, biostatistics, career development, statistics in medicine

As the sole biostatistician employed by the Department of Pediatrics at Washington University Medical School, I function as a sort of “internal statistical consultant”: my services are available without charge to every member of the department plus residents and medical students, and my salary is paid by departmental funds. The primary benefit to the department of this arrangement is my total availability to it for any statistical question or service. The advantages to me include the security of a salaried staff position, freedom from the tenure review process, and the variety of work that crosses my desk. The disadvantages include the absence of an immediate peer group, the constant need to assert my status as a professional, and the need to advocate for nonsalary compensation, such as authorship credit, professional development funds, and appropriate office space.

## 328 Approaches for Handling Missing Data and Measurement Issues ● ✪

Section on Statistics in Epidemiology, Biometrics Section
**Tuesday, August 8, 2:00 pm–3:50 pm**

### Missingness Screens and Regression Modeling in Clinical Aging Research

❖ Peter H. Van Ness, Yale University, 1 Church Street, 7th Floor, School of Medicine, Program on Aging, New Haven, CT 06510, *peter.vanness@yale.edu*

***Key Words:*** missing data, regression modeling, screening tools, older populations

Properly handling missing data is a challenge for clinical researchers working with older populations having high levels of morbidity and mortality. Screening tools have been developed by biostatisticians that enable researchers to understand the character of missing values in a set of variables. One test assesses whether values are missing completely at random--in a way depending neither on observed nor unobserved values. Another set of indexes assesses whether the values are missing at random--in a way depending on observed values but not on unobserved values. Use of such screening tools introduces complications into variable selection. In this presentation, we describe a model fitting process that incorporates the use of missingness screens, controls for collinearity, and selects variables based on model fit. We illustrate the process in a study of ICU delirium in an older cohort.

### Deletion Approach To Handle Missing Data in Longitudinal Data

❖ Cuiling Wang, Albert Einstein College of Medicine, 1300 Morris Park Ave., Belfer Building Rm 1303, Bronx, NY 10461, *cuwang@aecom.yu.edu*; Myunghee C. Paik, Columbia University

***Key Words:*** generalized estimating equation (GEE), imputation, inverse probability weighting, longitudinal studies, missing at random (MAR)

We propose a deletion approach and a closely related weighting approach to handle missing outcomes in generalized estimating equation (GEE) analysis. In deletion approach, we generate artificial observation indicators independent of the outcome, given the covariates, and draw inferences conditioning on the artificial observation indicators. This method provides valid estimate under MAR. The proposed weighting method has more stable weights than those of the inverse probability weighting method of Robins et al. We also apply the results of Robins, Rotnitzky, and Zhao (1994) to improve the efficiency of the proposed weighting estimator. The proposed weighting and improved weighting estimator show an advantage in efficiency compared to the corresponding inverse probability weighting method when the proportions of missing data are high.

### Impact of Missing Data on Building Prognostic Models and Summarizing Models across Studies

❖ Mahtab Munshi, Takeda Global Research and Development Center, Takeda Global Research and Development, 475 Half Day Road, Lincolnshire, IL 60069, *mmunshi@tgrd.com*; Daniel McGee, Sr., Florida State University

***Key Words:*** missing data, summary coefficients, logistic model, maximum likelihood estimation, coronary heart disease

As we progress in medical research, new covariates become available for studying a given outcome. While we want to investigate the influence of new factors on the outcome, we also do not want to discard the historical datasets not having information about these new markers. We propose methods to obtain prognostic models using data from multiple studies when one of the covariates is not contained in all studies. Our method is based on obtaining summary coefficients from individual logistic regression models fitted within each study. We use the example of addition of high density lipoproteins to existing equations for predicting death due to coronary heart disease. We perform simulations to show that our proposed method gives improved estimates for the completely observed covariates while still giving comparable estimates for the covariate that is not contained in all studies.

### Using Multiple Imputation To Improve Race-Specific Disease Rate Reporting in a National Active Surveillance System

❖ Elizabeth R. Zell, Centers for Disease Control and Prevention, 1600 Clifton Road, MS-C09, Atlanta, GA 30333, *EZell@cdc.gov*

***Key Words:*** incidence rate, racial disparities, multiple imputation, disease surveillance

Reducing health disparities is a Healthy People 2010 objective. CDC's Active Bacterial Core surveillance (ABCs) monitors racial disparities among invasive meningococcal and pneumococcal infections in the U.S. Race data are missing in 15% of reported ABCs cases. We have explored multiple imputation models which showed improvement over simple methods to account for unknown race. ABCs is an evolving system that adds variables over time to improve disease monitoring and for which specific variables included in case reports often varies among participating sites. A more complex approach to assigning missing race values is needed. To improve usefulness for larger epidemiologic evaluations we explore several imputation models. We evaluate multiple imputation using the multivariate normal and sequential regression multivariate models. A comparison of incidence rates by race will be presented.

### Efficacy Studies of Malaria Treatments in Africa: Efficient Estimation with Missing Indicators of Failure

❖ Rhoderick Machekano, University of California, Berkeley, Division of Biostatistics, Havilland Hall, Berkeley, CA 94710, *rodmach@berkeley.edu*; Alan Hubbard, University of California, Berkeley

***Key Words:*** multiple imputation, missing at random, inverse probability of censoring weight, doubly robust estimate

Efficacy studies of malaria treatments can be plagued by indeterminate outcomes for some patients. The study motivating this paper defines the outcome of interest (treatment failure) as recrudescence and for some subjects, it is unclear whether a recurrence of malaria is due to that or new infection. This results in a specific kind of missing data. The effect of missing data in causal inference problems is widely recognized. Methods that adjust for possible bias from missing data include a variety of imputation procedures, inverse weighting methods, and likelihood based methods. We compare multiple imputation to two inverse weighting procedures (the inverse probability of censoring weighted (IPCW) and the doubly robust (DR) estimators), and a likelihood based methodology. The DR estimator is most robust.

### Application of Errors-in-Variables to Model Variation between Studies in Regression Equations for GFR

❖ Tom Greene, The Cleveland Clinic, Department of Quantitative Health Sciences, 9500 Euclid Ave Wb4, Cleveland, OH 44195, *greenet@ccf.org*; Liang Li, The Cleveland Clinic

***Key Words:*** errors-in-variables, attenuation, generalizability, kidney function

Glomerular filtration rate (GFR) is the optimum measure of kidney function, but hard to measure. Thus, GFR is estimated from serum creatinine (SCR) and demographic factors using one of several equations developed by ordinary least squares regression. Recent reports have questioned the generalizability of these equations. We use errors-in-variables regression with study interactions to jointly relate SCR to GFR and demographic factors in 10 studies. We show that the majority of previously reported heterogeneity is due to predictable differences between studies in attenuation of regression coefficients resulting from differential error in SCR. The study interactions indicate remaining study variation not explained by attenuation. Our approach is applicable to settings in which a regression on easily measured predictors is used to estimate a more rigorous measure across populations.

### Dose-Response Errors and Detection of Biological Thresholds

❖ Michael E. Ginevan, Exponent, Inc., 307 Hamilton Ave., Silver Spring, MD 20901, *michael@ginevan.com*; Deborah K. Watkins, Exponent, Inc.

***Key Words:*** dosimetry errors, logistic regression, thresholds, Monte Carlo

The idea that random errors in dose estimates result in deflation of the estimated dose-response coefficient is well known. However these random errors can also mask the presence of thresholds in dose response models. The work presented here describes a Monte Carlo simulation approach to evaluating the impact of random dose errors on the commonly used logistic dose-response model. We consider the cases of both absolute and relative dosimetric error and examine the issue of the magnitude of errors relative to the range of doses in the study population. In each of these cases we evaluate the magnitude of reduction on the dose-response estimate and also the extent to which biological thresholds might be masked by a given error pattern.

## 329 Mode Considerations ●

Section on Survey Research Methods
**Tuesday, August 8, 2:00 pm–3:50 pm**

### Color, Labels, and Interpretive Heuristics for Response Scales

❖ Roger Tourangeau, University of Maryland, 1218 LeFrak Hall, College Park, MD 20742, *rtourangeau@survey.umd.edu*; Mick Couper, University of Michigan; Frederick G. Conrad, Institute for Social Research

***Key Words:*** measurement error, attitude scales, color

We carried out two experiments to investigate how the shading of the options in a response scale affected the answers to the survey questions. The experiments were embedded in two web surveys, and they varied whether the two ends of the scale were represented by shades of the same or different hues. The experiments also varied the numerical labels for the scale points and examined responses to both unipolar scales (assessing frequency) and bipolar scales (assessing favorability). We predicted that the use of different hues would affect how respondents viewed the low end of the scale, making responses to that end seem more extreme than when the two ends were shades of the same hue. The results were generally consistent with this prediction. When the end points of the scale were shaded in different hues, the responses tended toward shift to the high end of the scale.

### Efficient Multimode Data Collection

❖ David Dolson, Statistics Canada, 15th Floor, R.H. Coats Building, Tunneys Pasture, Ottawa, ON K1A 0T6 Canada, *ddolson@statcan.ca*

***Key Words:*** census, internet survey, data processing

This paper explains the multimode collection process used in the 2006 Canadian Census. Explained are the ways in which a central master control system, using a dynamic frame of dwellings and questionnaire tracking, resolves duplicate responses and ensures efficient control and coordination of all operations, whether run locally or centrally. Initial evaluations will be given. Data collection via five response channels must be coordinated. Questionnaires were delivered by either dropoff---based on field listing of dwellings---or by mail using a register-based frame of addresses. Respondents had a choice to respond by either internet or mail. Some data were collected by personal or CATI interviews. Also, respondents to the long questionnaire could either reply to the income questions or give Statistics Canada permission to link to their tax records to obtain these data.

### Improved Evaluation of the Quantitative Survey Response Variance

❖ Patrick Flanagan, U.S. Census Bureau, 1314 Bristol Ridge Place, Crownsville, MD 21032, *patrick.e.flanagan@census.gov*

***Key Words:*** surveys, concordance correlation coefficient, index of inconsistency, response error, response variance, reinterview

When evaluating the survey response variance of quantitative survey data using a reinterview, the Index of Inconsistency has an advantage over the Pearson correlation coefficient and Lin's concordance correlation coefficient in that it estimates the more illustrative ratio of the simple response variance to the overall variance (though under restrictive conditions they all produce the same results). The typical formula for the Index of Inconsistency for quantitative data is biased if the equal means or equal variances assumptions are violated. I propose an improved Index of Inconsistency "testimator" to adjust the estimation for violations of those assumptions, particularly focusing on the equal variance assumption. The results of simulations illustrate the effect of violation of assumptions on four possible estimators, comparatively illustrating their properties.

### Comparing Performance and Self-Reported Data on Ease of Use of Surveys

❖ Christine Rho, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, PSB Room 1950, Washington, DC 20212, *rho.christine@bls.gov*

***Key Words:*** self-report, performance measure, usability, cognitive interview

During pre-testing of survey questions and websites, qualitative data are collected based on participants' think-alouds and comments, as well as satisfaction and ease of use ratings. During cognitive and usability tests of the Bureau of Labor Statistics surveys and websites, discrepancies were observed between participants' self-reports about the clarity or ease of use of survey questions and their actual performance on those questions. For example, despite encountering difficulties and committing critical errors, participants still reported that a web-based instrument was easy to use. The discrepancy between self-reported and performance data may be attributed to social-desirability and good-subject bias. The analysis summarizes results from several different studies and argues for greater reliance on performance data to evaluate ease of use of surveys.

### Estimation of Measurement Error and Identification of Causes in the Absence of Validation Data

❖ Andy Peytchev, University of Michigan, 426 Thompson Street, ISR Room 4062, Ann Arbor, MI 48106-1248, *andrey@isr.umich.edu*

***Key Words:*** measurement error, interviewer effects, two-stage model

In order to identify survey design factors that induce measurement error, validation data are needed. Gold standards are difficult to obtain and are of limited types. In the absence of truth, analyses commonly revert to differences in means, such as the Kish and Hansen, Hurwitz, and Bershad models for interviewer variance, and similar ANOVA models for comparisons of response scales. Proposed is a two-stage approach - the measurement error is first estimated through a predictive model and then it is modeled using hypothesized covariates. This two-stage model will be demonstrated on cumulated waves of the Survey of Consumer Attitudes to examine interviewer effects on measurement error. Effects of interviewer experience, call history, and current call characteristics on measurement error will also be estimated.

### An Alternative Estimator for Multi-Frame Sample Designs

❖ Charles D. Palit, University of Wisconsin, 933 University Bay Drive, Madison, WI 53705, *cdpalit@wisc.edu*

***Key Words:*** dual frame, multi-frame, estimation

Estimates from dual frame designs have generally followed the pattern laid out by Hartley in 1962. In this paper we present an alternative estimation procedure for multi-frame sample designs and discuss some applications

### Is There a Homogeneity in Value Orientations within European Societies?

❖ Peter P. Mohler, ZUMA, P.O. Box 122155, Mannheim, 68072 Germany, *director@zuma-mannheim.de*

***Key Words:*** latent class, homogeneity, within country, cross-cultural, surveys, values

The Schwartz' value items are among the most widely used today. Schwartz assumes universal patterns across cultures, languages and social systems. This assumption has been tested using a short (21 item) Schwartz scale and the data from two consecutive waves of the European Social Survey. The data do not support the general universal pattern assumption, indeed. This leads to the suspicion that there are third, intervening variables which distort the universal pattern. Recently developed techniques (Latent Class) will be used to identify latent classes which are non-comparable, singular and thus culture specific, while other classes follow the universal pattern. The paper will identify such culture specific classes and discuss the issue of within culture homogeneity as a problem for cross-cultural surveys.

## 330 Bayesian Methods in the Social Sciences ● ✪

Section on Bayesian Statistical Science

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Latent Space Models for Irish Voting Data

❖ Isobel Claire Gormley, Trinity College Dublin, Department of Statistics, Trinity College, Dublin, BT45 5LD Ireland, *gormleyi@tcd.ie*; Thomas B. Murphy, Trinity College Dublin

***Key Words:*** latent space models, Plackett-Luce model, Irish elections, ranking data

Proportional representation by means of a single transferable vote is the electoral system employed in Ireland. The system involves voters ranking candidates in order of preference and gives rise to a wealth of ranking data. A latent space model is proposed where both candidates and voters are simultaneously located in a latent space; the distance between members of the space quantitatively describes their relationship. The Plackett-Luce model for ranking data is employed to model the ranked nature of the votes cast. The location of each voter and each candidate is inferred from the voter preferences. The relative location of the candidates is of interest in political science. The model is fitted within the Bayesian paradigm. Spatial issues such as the dimensionality of the latent space and identifiability are addressed.

### A Bayesian Approach to Correct Sample Selection Bias in Hierarchical Linear Models

❖ Jun Lu, American University, Department of Mathematics and Statistics, Gray Hall 201, 4400 Massachusetts Avenue NW, Washington, DC 20016, *lu@american.edu*; Labeed Mokatrin, American University

***Key Words:*** Heckman's two-stage estimator, sample selection bias, Bayesian models

Sample selection bias occurs when the observed samples are selected non-randomly from the population. Gronau (1974) formulated a sample selection model with two correlated linear processes: (1) a selection process, where subsamples are picked from a set of random samples by the selection rule, and (2) an outcome process where only the subsamples are observed. Under this model, Heckman (1979) proposed the well-known two-stage estimator to correct the selection bias. Here, we provide a Bayesian approach as an alternative to the two-stage estimator. Through simulation study, we compare the performance of the Bayesian approach and the traditional two-stage estimators. We also generalize the model by considering hierarchical linear models in the outcome process. We then illustrate the Bayesian model using education data.

### Exploring Teacher by Student Interactions in Longitudinal Achievement Data

❖ J. R. Lockwood, RAND Corporation, 201 N. Craig Street, Suite 202, Pittsburgh, PA 15213, *lockwood@rand.org*; Daniel McCaffrey, RAND Corporation

***Key Words:*** teacher effects, value-added models, cross-classified models, latent growth models

With increased emphasis on accountability in public education, analyses of longitudinal student achievement data attempting to isolate teacher contributions to student learning have become more common. To date, these so-called "value added" models have focused on estimating the average effect of a teacher across all of his/her students. Understanding to what extent teacher effects might vary across students is both substantively interesting and may have implications for the validity of models that estimate only average effects. In this talk we present results on our explorations of the variation of teacher effects across students using several longitudinal student achievement data sources. Our tools for investigating the interactions include both non-parametric approaches, as well as Bayesian cross-classified models.

### An Item Response Model for Balancing Personalization and Privacy in Online Browsing

❖ Steven L. Scott, University of Southern California, Bridge Hall 401H, 3670 Trousedale Parkway, Los Angeles, CA 90089, *sls@usc.edu*; Ramnath Chellappa, Emory University

***Key Words:*** item response, e-commerce, personalization, hierarchical model

We propose an item response model for describing a consumer's preferences about various product categories. The model encodes an individual's preferences about $k$ product categories using a latent $k$ dimensional Gaussian vector. Individual preference vectors are viewed as draws from a common multivariate normal distribution interpretable as a model for market preferences. A trusted party maintains information about individual preference vectors and the prior market parameters. Each individual may label preferences for each product category as public or private. The trusted party releases a customer's public preferences to online merchants wishing to provide a personalized browsing experience. For public release, private preferences are drawn from their conditional (market) distribution given public preferences.

### Modeling Student Ability in Low-Stakes Settings

❖ Eric Loken, The Pennsylvania State University, S150 Henderson Building, University Park, PA 16802, *loken@psu.edu*

***Key Words:*** adaptive testing, Bayesian methods, latent variables

Computer adaptive methods are routinely employed for high-stakes testing applications such as admissions and licensing tests. With certain model alterations, computerized adaptive methods also can be used for low-stakes settings. A fourth parameter is introduced to the standard three-parameter logistic model to reflect the non-zero probability that some questions are answered incorrectly. Under low-stakes conditions, the joint estimation of item and person parameters may be more challenging because some students may answer only a few questions, and questions may be delivered over a narrow range of ability. A Bayesian approach using Markov chain Monte Carlo methods to simulate the posterior distribution for item parameters after an initial pilot phase accurately estimates the question parameters. The methods are applied to data gathered from an Internet web site.

### A Bayesian Approach to the Calibration of New Test Items

❖ Tiandong Li, Westat, 206 Cherrywood Terrace, Gaithersburg, MD 20878, *tiandong_li@yahoo.com*; Chinfang Wong, University of Maryland; Ru Lu, University of Maryland

***Key Words:*** Bayesian approach, calibration, latent class model

Within the framework of probability-based evidential reasoning, new test items can be calibrated into the same scale as the old test items. The focus of this study is on inference about new tasks in assessment programs using different 'domain-behavior' structures in a latent class model. We examine and compare the precision of estimation on item parameters. Monte Carlo Markov Chain (MCMC) is applied to estimate the posterior distribution in inferences. The study simulates 2000 replicates under a binary latent class model with 4 latent variables and 16 old test items. The results show that the number of old items is more important than the 'Domain-behavior' structure in calibrating new items. During calibration, the items with higher discrimination produce better results.

## 331 Distribution and Robustness Theory

IMS, Section on Nonparametric Statistics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A New Approach to Robust Parameter Estimation against Heavy Contamination

❖ Hironori Fujisawa, The Institute of Statistical Mathematics, Minami Azabu, Minato ku, Tokyo, 106-8569 Japan, *fujisawa@ism.ac.jp*

***Key Words:*** bias, cross entropy, divergence, heavy contamination, Pythagorian relation, robustness

This talk focuses on a certain cross entropy and divergence, which enable us to reasonably deal with the case of heavy contamination. We see that the bias caused by outliers can become sufficiently small even in the case of heavy contamination. The proposed method can be shown to be a kind of projection from the viewpoint of a Pythagorean relation, which is why it works well. In addition, it is proved that the method of parameter estimation with a sufficiently small bias even in the case of heavy contamination is essentially unique under some conditions.

### Minimax Asymptotic Mean-Squared-Error of L-estimators of Scale Parameter

❖ Daniela Szatmari-Voicu, The University of Texas at El Paso, 6022 Caprock Court, Apt. 102, EL PASO, TX 79912, *dsvoicu@utep.edu*

***Key Words:*** robust estimation, scale parameter, l-estimators, asymptotic mean-squared-error

Consider the scale estimation problem under the location-scale model Xπ,X≤,..,Xn~iid Fn((x-?)/s), where -8< ?< 8 and s>0 are unknown and Fn is an unknown member of the shrinking neighborhood Fen ={Fn❡Fn=(1-en)Fo+enG, where en=k/Vn, k=cst>0, Fo known error distribution and G unknown} and the class of L-estimators of scale that are location-scale equivariant and Fisher consistent at Fo. On this class of estimators we derive the AMSE (maximal asymptotic mean-squared-error) in terms of the mixing distribution on the quantiles which defines the L-estimator. For non-Normal Fo, we next find estimators which have minimum AMSE over the subclass of (a) a - interquantile ranges and (b) mixtures of at most two a - interquantile ranges. We next derive the AMSE of the L-estimator of scale symmetrized about the median. When Fo symmetric, the symmetrized and non-symmetrized L-estimators have equal AMSE.

### Order Statistics of Concomitants of Subsets of Order Statistics and Applications

❖ Ke Wang, The Ohio State University, Department of Statistics, 1958 Neil Ave., Cockins Hall Room 404, Columbus, OH 43210, *kewang@stat.ohio-state.edu*; H. N. Nagaraja, The Ohio State University

***Key Words:*** order statistics, concomitants, copula function, asymptotic theory, extremes

Let (X_i, Y_i), i=1,\ldots,n, be a random sample from a continuous bivariate distribution F(x,y) with positively associated components. Suppose the Y-concomitants of the top s order statistics of the X-sample values are observed. We are interested in the event that at least t of these concomitant values are among the top k values of the entire Y-sample. A closed form expression for the desired probability is derived using conditioning argument. It is shown that the desired probability depends on the population distribution, F(x,y), only through the associated copula function. Approximation of the desired probability for large samples is also explored. Finally application of these results to some real situations is discussed.

### Power Transformation Toward a Linear Regression Quantile

❖ Yunming Mu, Texas A&M University, 3143 TAMU, College Station, TX 77843, *ymu@stat.tamu.edu*

***Key Words:*** quantile regression, Box-Cox transformation, cusum process, empirical process, v statistic, lack-of-fit

We consider the linear quantile regression model with a power transformation on the dependent variable. Like the classical Box-Cox transformation approach, it extends the applicability of linear models without resorting to nonparametric smoothing, but transformations on the quantile models are more natural due to the equivariance property of the quantiles under monotone transformations. We propose an estimation procedure and establish its consistency and asymptotic normality under regularity conditions. The objective function employed in the estimation also can be used to check inadequacy of a power-transformed linear quantile regression model and to obtain inference on the transformation parameter. The proposed approach is shown to be valuable through illustrative examples.

### Some Moment Relationships for Skew-Symmetric Distributions

❖ Dale Umbach, Ball State University, Department of Mathematical Sciences, Ball State University, Muncie, IN 47306, *dumbach@bsu.edu*

***Key Words:*** skew-symmetric, skewness, moments

If f is a density function that is symmetric about 0 and G is the distribution function of a distribution that is symmetric about 0, then 2f(x)G(x) is a density function. Such distributions are called skew-symmetric distributions. Properties of moments of these distributions, especially skew, as they relate to f are discussed. In particular, even moments are independent of the choice of G, and the first moment always falls between 0 and the mean of folded distribution 2f(x) for x>0. Such results are extended to the parametric family, 2f(x)G(dx) for real d. It is shown, for example, that odd moments are increasing functions of d, but there is no general-order relations skew as defined by the central third moment.

### On the Decomposition of the Skew-Symmetric Family of Probability Distributions

❖ Jose A. Sanqui, Appalachian State University, 121 Bodenheimer Drive, ASU Box 32092, Boone, NC 28608, *sanquijat@appstate.edu*; Arjun K. Gupta, Bowling Green State University; Truc T. Nguyen, Bowling Green State University

***Key Words:*** skew-normal, skew-symmetric, characteristic function

In the paper of Gupta et al. (2004), a conjecture was given regarding the decomposition of the skew-symmetric family. The conjecture states that if a certain linear combination of two independent random variables belong to a member of the skew-symmetric class and one of the random variables is standard normally distributed, then the other variable must also belong to the same member of the skew-symmetric family. This conjecture is shown to hold for the skew-normal distribution. In this paper, we will present results for other members of the skew-symmetric family. Reference: Gupta, A.K., Nguyen, T.T and Sanqui, JAT (2004). Characterization of the skew-normal distribution, Annals of the Institute of Statistical Mathematics, 56, 351-360.

## 332 Estimation for Government Data Collection

Section on Government Statistics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### New Models for Estimating Health Insurance Coverage for Small Areas

❖ Donald M. Bauder, U.S. Census Bureau, 2301 E. Street, NW, Apt A714, Washington, DC 20037, *donald.m.bauder@census.gov*

***Key Words:*** small area estimation, hierarchical models, Bayesian models, health insurance

The Small Area Health Insurance Estimates (SAHIE) program at the Census Bureau provides model-based estimates of health insurance coverage. In 2005, the program developed experimental estimates of health insurance coverage at the county level, using fully Bayesian models. Here, we describe these models and results from the estimation. We also present more recent developments in the models.

### Small-Area Estimation of Health Insurance Coverage at the Sub-State Level: a Hierarchical Bayes Model

❖ Steven Riesz, U.S. Census Bureau, 4700 Silver Hill Road Stop 8700, Washington, DC 20233-8700, *steven.riesz@census.gov*; Robin Fisher, U.S. Census Bureau

***Key Words:*** small area estimation, hierarchical Bayes, administrative records

We develop a hierarchical Bayes model to estimate the number people without health insurance for each state, within categories defined by sex, age, race/ethnicity, and income. Our specific goal is to estimate, for each state, the number of women without health insurance, between the ages of 40 and 64, who earn less than 200% of the poverty line, within 3 race/ethnicity categories. Our model is noteworthy in two respects: (1) we model direct survey estimates of 2 different quantities, namely, the proportion of the population within certain income categories, and the proportion of the population with health insurance, and (2) we model auxiliary data from administrative records, instead of using them as predictors. The estimation is performed by MCMC simulation, using WinBUGS. We evaluate the model fit with posterior predictive p-values. We produce estimates and their standard errors.

## Estimation of Standardized State-Level Food Stamp Participation Rates

❖ Elizabeth Stuart, Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024, *estuart@mathematica-mpr.com*; Allen Schirm, Mathematica Policy Research, Inc.; Alan M. Zaslavsky, Harvard Medical School; Laura Castner, Mathematica Policy Research, Inc.; Scott Cody, Mathematica Policy Research, Inc.

***Key Words:*** hierarchical model, food stamp participation, logistic regression, multiple data sources

Food stamp participation rates vary widely across states. The policy implications of this variation depend on how much is due to different characteristics of the eligible households across states and how much is due to differences in state economic conditions and policies. We fit a two-stage model to investigate this variation. The first stage estimates state-level participation rates for a standard population similar to the national population of eligible households. The second stage describes the relationship between the standardized state-level participation rates and state policies and economic conditions. We provide an overview of this procedure, focusing attention on the standardization procedure. An unusual feature of our logistic regression model for household-level participation is that information on eligibility and on participation come from two different sources.

## Estimating School District Poverty with Free and Reduced Priced Lunch Data

❖ Craig Cruse, U.S. Census Bureau, 7112 Split Rail Lane, Laurel, MD 20707, *craig.s.cruse@census.gov*; David S. Powers, U.S. Census Bureau

***Key Words:*** small area estimates, poverty, school districts, free and reduced price lunch

The Small Area Income and Poverty Estimates program of the U.S. Census Bureau produces model-based estimates of the number of children ages 5-17 in poverty in school districts. The Department of Education, under the No Child Left Behind Act of 2001, uses these estimates in allocation formulas for Title I funding . The National Academy of Sciences (NAS) favorably evaluated the program's methodology, and recommended research using free and reduced price lunch (FRPL) eligibility counts in the poverty model. FRPL data are collected annually and are available from the National Center for Education Statistics. This research systematically addresses the NAS recommendation using regression-based techniques, and evaluates the influence of FRPL data as a predictor of the number of children in poverty in school districts.

## An Analysis of Gender Differences in Vehicles Miles Traveled Using Nonparametric Methods

❖ Brian Sloboda, Bureau of Transportation Statistics, 8750 Georgia Ave., 112B, Silver Spring, MD 20910, *brian.sloboda@dot.gov*; Wenxiong V. Yao, University of Arkansas at Little Rock

***Key Words:*** nonparametric methods, trip data, vehicle miles traveled, gender differentials

Traditionally, women make shorter work trips, greater use of public transit, more trips for the purpose of serving another person's travel needs, and drive far fewer miles per year than men. These differences in travel are delineated by the separate social responsibilities of men and women. However, in recent years, women have been participating more in the labor force, but still retain their family obligations as nurturers, shoppers, and homemakers. The National Household Travel Survey (NHTS) can be used to understand these current trends in their travel patterns, which often are attributed to changes in labor force participation, household structure, and attitudes. This paper will analyze the differences in the vehicle miles traveled between men and women using nonparametric methods from the NHTS data.

## Improved Preliminary Estimation of Total Employment Change for the U.S. Current Employment Statistics Survey

❖ Bogong Li, Bureau of Labor Statistics, Washington, DC , *Li_T@bls.gov*; Partha Lahiri, University of Maryland

***Key Words:*** CES, Bayesian hierarchical model, weighted link relative estimator, composite estimator

The U.S. Bureau of Labor Statistics (BLS) releases estimates of total non-farm business payroll employment changes every month. These estimates have significant impact on US economic policy and financial market decision makers. However these estimates are produced based only on "preliminary" Current Employment Statistics (CES) survey results, final estimates are release two months later. In this paper, we develop a statistical model based on historical CES data with the following goals in mind: (1) validate factors that affect monthly employment changes, (2) measure the magnitude of these effects on monthly employment movement, (3) prove in part some underlying economic factors affecting national and regional employment changes. We then use prediction results from the model to produce improved preliminary estimates that is more accurate than the current preliminary estimates.

## Analysis of Recall Effect on the Reporting of Expenditures for the Consumer Expenditure Interview Survey

❖ Barry Steinberg, Bureau of Labor Statistics, 2 Massachusetts Avenue Rm 3650, Washington, DC 20212 United States, *Steinberg.Barry@BLS.gov*; Boriana Chopova, Bureau of Labor Statistics; Lucilla Tan, Bureau of Labor Statistics; Jared Ogden, Bureau of Labor Statistics; Pierre Bahizi, Bureau of Labor Statistics; Geoffrey Paulin, Bureau of Labor Statistics

***Key Words:*** recall, means

The Consumer Expenditure Interview Survey has a three-month recall period. Over the years, researchers have noticed that the level of expenditures is highest for the most recent recall month, lower for the middle month, and lowest for the most distant month. This pattern is generally believed to be caused by respondents gradually forgetting about their expenditures over time, although other theories have been put forth as well. To improve the accuracy of the reported expenditures, research was done to identify the cause of this recall phenomenon, measure its magnitude, and possibly recommend steps to solving it. This paper presents a model for quantifying the magnitude of various sources of the observed recall effect, along with a statistical analysis of the Consumer Expenditure Survey data. Findings are presented that show the estimated forgotten amount and its magnitude relative to the model's predicted, "true" expenditure.

# 333 New Approaches to Modeling Costs and Other Health Outcomes ●

Section on Health Policy Statistics, Biometrics Section
**Tuesday, August 8, 2:00 pm–3:50 pm**

## Performance of Statistical Models To Predict Mental Health and Substance Abuse Cost

❖ Maria Montez, Boston University, 51 Pershing Ave., Biostatistics Department, Acushnet, MA 02743, *montez@bu.edu*; Cindy Christiansen, Boston University; Susan L. Ettner, University of California, Los Angeles; Susan Loveland, Boston University; Amy K. Rosen, Boston University

***Key Words:*** cost analysis, GLM, log-normal, retransformation, risk adjustment

Our sample consisted of 525,620 VA patients with mental health (MH) or substance abuse (SA) diagnoses who incurred costs during FY99. We tested a log-normal (LN) model and generalized linear models on untransformed cost, defined by distributional assumption and link function: Gaussian Identity (GI); Gamma Log (GL); and Gamma Square Root (GSR). Risk-adjusters included age, sex, and 12 MH/SA categories. To determine the best model, predictive ability was evaluated using root mean square error (RMSE), mean absolute prediction error (MAPE), and predictive ratios among deciles of predicted cost. To study the effect of analyzing a random sample of the population on model choice, we recomputed these statistics using small size random samples. The GRS and the GI models had similar RMSEs; the GSR and the LN models had similar MAPEs. The GSR had some convergence problems with small samples.

## Use of Risk-Adjustment Models To Predict Future High Medical Cost Cases: Is the Model Performance Sensitive to the Time Intervals in Claims Data?

❖ Ya-Chen Tina Shih, M. D. Anderson Cancer Center, HSR Department of Biostatistics and Applied Math, 1515 Holcombe Blvd BOX 196, Houston, TX 77030, *yashih@mdanderson.org*; Lirong Zhao, M. D. Anderson Cancer Center; Ying Xu, M. D. Anderson Cancer Center

***Key Words:*** risk-adjustment model, prediction, high cost cases, claims data

Risk-adjustment models are used to identify persons that are likely to incur high medical costs in the future. They typically included last year's annual medical expenditures (expen) as a covariate; thus, their applications require at least having one full year of claims data while decision-makers often need feedback in a more timely fashion. We explored whether the performance of these models is sensitive to the time intervals of expen in claims. Using a 1% random sample from the Taiwanese National Health Insurance claims, we defined high cost cases as having expen at the top 1% of the distri. and employed logistic regressions to predict these cases. We compared the area under the curve from the ROC curves for various RA modes using claims at 3 time intervals -- 3, 6, 12 months, and concluded that models with shorter time intervals performed equally well as those with longer intervals.

## Two-stage Samples and the Minimum Sum Method for Medicare Fraud Investigations

❖ Iliana Ignatova, University of South Carolina, 322 Catawba Circle, Columbia, SC 29201, *iliana_ignatova@yahoo.com*

Random sampling of paid Medicare claims has been legally acceptable for investigating suspicious billing practices by health care providers since 1986. A population of payments made to a given provider during a given time frame is isolated and a sample selected for investigation. Edwards et al. (2005, Health Services & Outcomes Research Methodology) showed that the widely used CLT methods can fail badly. They developed an alternative - the Minimum Sum (MS) method. In this paper the sampling is performed in two stages. In case of little abuse in the first stage the investigation is stopped otherwise a second sample is examined. Based on this strategy a lower confidence bound for total number of payments in error and the corresponding MS lower bound for the total overpayment amount are defined. Criteria for choosing the sampling parameters are considered. Relative efficiencies are studied.

## Experiences with a Virtual Regionalization Model for Cardiac Surgery

❖ Edward Wegman, George Mason University, Department of Statistics, MSN 4A7, 4400 University Drive, Fairfax, VA 22030-4422, *ewegman@gmu.edu*; Yasmin H. Said, Johns Hopkins University; Shabib A. Alhadheri, SUNY Upstate Medical University

***Key Words:*** hypoplastic left heart syndrome, risk factors, pediatric cardiology

Serious medical conditions are often redirected from regional medical facilities to more experienced and larger central facilities. However, additional risk is incurred by such transfers. Two university-based medium-sized referring pediatric cardiac surgical centers have implemented a virtual regionalization model in which one full-time surgeon operates at both institutions using a common management protocol. In this paper we analyze the risk factors for Hypoplastic Left Heart Syndrome patients using the Virtual Regionalization Model.

## The LASSO-Patternsearch Algorithm and Its Application to Data from the Beaver Dam Eye Study

❖ Weiliang Shi, University of Wisconsin-Madison, Department of Statistics, 1300 University Ave., Madison, WI 53706, *shiw@stat.wisc.edu*; Grace Wahba, University of Wisconsin-Madison; Kristine Lee, University of Wisconsin-Madison; Ronald Klein, University of Wisconsin-Madison; Barbara E.K. Klein, University of Wisconsin-Madison

***Key Words:*** LASSO, GLM, risk factor patterns, myopia

The LASSO-Patternsearch is proposed, to identify clusters of multiple risk factors for outcomes of interest in large demographic studies. Many diseases are suspected of having multiple interacting risk factors acting in concert, and it is of interest to uncover these interactions. The method is related to H. Zhang et al, JASA99(2004), except that some variable flexibility is sacrificed to allow entertaining models with low as well as high order interactions among multiple predictors. A LASSO is used to pick out important patterns, being applied conservatively to have a high rate of retention of true patterns, while allowing some noise. Then the patterns picked by the LASSO are tested in the framework of generalized linear models to reduce the noise. The method is applied to data from the Beaver Dam Eye Study and is shown to expose physiologically interesting interacting risk.

### On the Equivalence of Medical Cost Estimators with Censored Data

Heejung Bang, Cornell University; ❖ Hongwei Zhao, University of Rochester, Department of Biostatistics and Computational Biology, Rochester, NY 14642, *Hongwei_Zhao@urmc.rochester.edu*; Phillip E. Pfeifer, University of Virginia; Hongkun Wang, University of Virginia

***Key Words:*** censoring, cost analysis, inverse probability-weighting, survival analysis

In clinical trials comparing different treatments and in health economics and outcomes research, medical costs are frequently analyzed to evaluate the economical impacts of new treatment options and economic values of health care utilization. Since Lin et al. (1997)'s first finding in the problem of applying the survival analysis techniques to the cost data, many new methods have been proposed. In this report, we establish analytic relationships among several widely adopted medical cost estimators that are seemingly different. Specifically, we report the equivalence among various estimators that appeared in Lin et al. (1997), Bang & Tsiatis (2000) and Zhao & Tian (2001). We identify conditions under which these estimators become identical and, consequently, the biased estimators achieve consistency.

## 334 Degradation Models and Other Topics in Reliability ✪

Section on Physical and Engineering Sciences

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A Statistical Method for Crack Detection in Thermal Acoustics Nondestructive Evaluation Data

❖ Chunwang Gao, Iowa State University, 46 Schilletter Village, Apt. D, Ames, IA 50010-8746, *cgao@iastate.edu*; William Q. Meeker, Jr., Iowa State University

***Key Words:*** nondestructive evaluation, thermal acoustics, probability of detection, principal components, robust fitting, data clustering

A systematic statistical method based on principle components analysis is developed to automatically analyze a sequence of images (movie) generated in thermal acoustic inspections to detect component anomalies such as internal cracks. The method uses principle components regression for dimension reduction in the data processing. Robust fitting and clustering are applied on principle components regression results to reduce the effect of the noise structure in the movie data and to help in setting up the detection rule. The method gives results that are consistent with a human expert detection by visual inspection of the movie after some simple signal processing. The Probability of Detection ( POD) is calculated according to the detection rules.

### Estimation of Flaw Size Distribution Parameters under Correlated Random Censoring

❖ Peter Hovey, University of Dayton, Department of Mathematics, 300 College Park Dr, Dayton, OH 45469-2316, *Peter.Hovey@notes.udayton.edu*; Alan Berens, Retired

***Key Words:*** safety, fatigue crack, censoring, nondestructive inspection, maximum likelihood

The U.S. Air Force plans for maintenance and retirement of aircraft based in part on fatigue crack growth models. Periodic inspections are used to help assess airworthiness and plan for future inspections. Nondestructive inspections are not perfect so some cracks are missed and the likelihood that an individual crack is detected is a function of the size of the crack when inspected. Additionally, the crack size distribution is related to the number of flight hours the aircraft has experienced, so not all inspection results come from the same distribution. Several incremental models are compared that account for the correlated random censoring imposed by the inspection system and the variation between aircraft and times of inspections.

### Modeling Unit Degradation from Exceedance Time Data

❖ Hui Fan, Rensselaer Polytechnic Institute, CII 5015, DSES Department, RPI, 110 8th Street, troy, NY 12180, *fanh2@rpi.edu*; Brock Osborn, GE Global Research; Thomas R. Willemain, Rensselaer Polytechnic Institute; Pasquale Sullo, Rensselaer Polytechnic Institute

***Key Words:*** degradation, exceedance, jet engine, reliability

Monitoring and predicting equipment degradation is an important task within reliability analysis. We consider the problem of exploiting degradation data censored to reveal only the times at which a performance variable exceeds a threshold. Our motivating example is the study of exhaust gas temperatures in jet engines, where safety regulations define exceedance thresholds and require expensive off-wing maintenance after a specified number of exceedances. Using both artificial data from a random walk model and real data, we develop methods for using data on exceedance times to estimate unit-specific degradation parameters and develop predictive distributions for the time until a specified number of exceedances will occur. We also assess the use of nonparametric methods to test for the presence of performance degradation.

### A Discrete Degradation Model for Ultra-Thin Gate Oxide Data

❖ Shuen-Lin Jeng, Tunghai University, No 181 Sec 3 TaichungKan Road, Dept. of Statistics, Taichung, 407 Taiwan, *sljeng@thu.edu.tw*; Min-Hsiung Hsien, Tunghai University

***Key Words:*** discrete degradation, compound Poisson, breakdown, nano meter

Using degradation measurements is getting more important in reliability study because of few failures observed during short experiment time. Assessing the reliability by the degradation processes of gate oxides has been investigated for decades in physics and microelectronics. Recently, discrete degradations have been observed in the breakdown of ultra-thin oxides, which are only few nano meters in thickness. Based on the physical properties of the gate oxides, we develop a nonhomogenous compound Poisson model incorporating different stress levels. We provide the likelihood function for model parameters and derive the first passage time distributions for breakdown times. Asymptotic confidence interval for quantiles of breakdown times is developed. A simulation study is set up for this discrete degradation model to investigate the finite sample properties.

### Bayesian Degradation Modeling with Covariate-Dependent Box-Cox Transformation of the Response Variable

❖ Fridtjof Thomas, VTI, Tunagatan 13 D 4tr, Borlange, 78434 Sweden, *fridtjof.thomas@vti.se*; Arzu Onar, St. Jude Children's Research Hospital; Bouzid Choubane, Florida Department of Transportation; Tom Byron, Florida Department of Transportation

***Key Words:*** accelerated testing, mixed-effects models, MCMC, pavements, transportation engineering

We discuss Bayesian degradation models that were developed for flexible road pavements based on accelerated pavement testing with the Heavy Vehicle Simulator (HVS). The models are fitted to data from the Florida Department of Transportation where rutting performance of three pavements was tested under three temperature settings. The analysis utilizes linear mixed-effects models for longitudinal degradation data where the parameter estimates and their posterior marginal distributions are obtained via a Markov chain Monte Carlo (MCMC) technique. The linearity in this model is achieved by utilizing a covariate-dependent Box-Cox transformation of the response variable, where the transformation parameter is estimated as part of the modeling procedure.

### Recurrent Events: Modeling and Inference

❖ Edsel A. Pena, University of South Carolina, Department of Statistics, Columbia, SC 29208, *pena@stat.sc.edu*

***Key Words:*** dynamic models, frailty models, repair models, marginal models, renewal models, event time modelling

Recurrent events occur in a variety of scientific areas, so probabilistic models for their occurrence and inference methods for model parameters are important. For a unit denote by $\{N(s): s > 0\}$ the stochastic process counting number of event occurrences on or before time s. Given an unobserved frailty Z and the history F(s-) just before time s, we study the class of models specified by the intensity process $P\{dN(s)=1|Z,F(s-)\} = ZY(s)\exp\{Q(s,N(s-);a)+R(s;b)\}l(E(s))ds$, where Y(s) indicates that unit is still under study; Q and R are bounded predictable processes, E(s) is an observable predictable process with piecewise monotone increasing and differentiable paths; l(.) is a hazard rate function; and Z has some distribution H(.;e). Model parameters are l(.), a, b, and e. Recent inference methods for this class of models, which subsumes many existing models, will be presented.

## 335 Classification Methods and Functional Data Analysis ●

Section on Nonparametric Statistics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Scrambling Method for Cluster Analysis Using Supervised Learning

❖ Oksana Shcherbak, Union Bank of California, 2673 Matera Lane, San Diego, CA 92108, *oksana.shcherbak@uboc.com*

***Key Words:*** cluster analysis, CART decision trees, supervised learning, multivariate analysis, data segmentation, outlier analysis

A general problem of finding high-density regions in the data space, also known as data segmentation, and is an example of unsupervised learning often arises in the field of multivariate analysis. In this paper we present a scrambling technique that allows the use of supervised learning methods such as CART decision trees in unsupervised learning. We found the performance of several CART models for different datasets obtained using the scrambling method to be quite strong and stable for high-dimensional problems. Examining the final models we detected informative behavioral patterns in our population by producing mutually exclusive dependency rules and discovered interesting departures from those resulting to an effective analysis of outliers. We conclude that the technique of scrambling analysis hold promise as a tool enabling effective detection of clusters and outliers in the data.

### Finding an Approximate Solution Path of Support Vector Machines for Large Datasets

❖ Zhenhuan Cui, The Ohio State University, 1958 Neil Avenue, Department of Statistics, Columbus, 43210, *zhenhuan@stat.ohio-state.edu*; Yoonkyung Lee, The Ohio State University

***Key Words:*** solution path, support vector machine, algorithm, classification, regularization, large datasets

The solution path of support vector machines (SVM) contains the entire set of solutions at every value of the regularization parameter that controls the complexity of a fitted model. The algorithm for the solution path has recently been extended from binary cases to multi-category cases. This algorithm greatly facilitates the computation of SVM by sequentially constructing the whole spectrum of solutions. However, large datasets and the choice of a flexible kernel may pose a computational challenge to the sequentially updating algorithm. In this paper, we borrow the idea of basis thinning to alleviate the computational load for large datasets and propose a method for approximate solution paths. In addition, some related computational issues are discussed and the effectiveness of the algorithm is demonstrated for some benchmark datasets.

### Bandwidth Selection for RBF Kernel in Kernel-Based Classification

❖ Jeongyoun Ahn, The University of North Carolina at Chapel Hill, 254 Butler Court, Chapel Hill, NC 27514, *jyahn@email.unc.edu*

***Key Words:*** discrimination, kernel method, machine learning, feature space, cross validation

The Gaussian RBF kernel functions are widely used for many kernel methods, including classification. The bandwidth parameter is often selected by some data driven tuning methods. One of the most popular tuning methods is cross validation, which is known to be subjective to sampling variation and also computationally expensive. We propose a new method for bandwidth selection, based on the geometrical understanding of kernel based classification: a nonlinear classification that is actually a linear one in the embedded feature space. Thus we find the bandwidth that makes this linear classification task the "easiest". This method empirically turns out to be remarkably robust to sampling variation, for both SVM and the nearest centroid method. It also yields highly competitive misclassification rates, especially compared to GACV and Xi-alpha method.

### Multivariate Functional ANOVA for Kriging Model in Computer Experiments

❖ Zhe Zhang, The Pennsylvania State University, 325 Thomas Building, University Park, PA 16802, *zxz118@stat.psu.edu*

***Key Words:*** functional ANOVA, kriging model, computer experiment

When conducting computer experiments, engineers frequently encounter functional data with sparse sampling rate. The computer experiments requires that the model needs to interpolate the data hence the Kriging model is frequently used. In order to find out which predictor has more important impact on the response, functional ANOVA is performed to analyze the response using multivariate Kriging model.

### Bootstrap Investigation of the Median Curve of a Functional Dataset

❖ David B. Hitchcock, University of South Carolina, Department of Statistics, 216 LeConte College, Columbia, SC 29208, *hitchcock@stat.sc.edu*

***Key Words:*** functional data, bootstrap, data depth, median curve

A definition of the median of a set of functional data is given. The sample median curve (or "most representative" curve) is defined as the data curve with the minimum average pairwise L1 distance to all other curves, and is a better representative of a "typical" curve than the pointwise mean or pointwise median curve. The median curve has connections to the sample median of a set of univariate or multivariate observations. An analogous definition of a population median curve is presented, along with methods for approximate inference about the population median curve. These methods---in particular, approximate confidence bands for the median curve---are based on a bootstrap approximation of the sampling distribution of the sample median curve.

### Analysis of Panic-Relevant Experimental Tidal Volume Curves: Wavelet-Based Functional Hypothesis Testing

❖ Sang Han Lee, Texas A&M University, 3143 TAMU, College Station, TX 77843, *hanul31@stat.tamu.edu*; Marina Vannucci, Texas A&M University; Eva Petkova, Columbia University; Maurice Preter, Columbia University; Donald Klein, Columbia University

***Key Words:*** wavelets, functional ANOVA, one-degree of freedom test, permutation test

We present results from statistical analysis designed for high-frequency tidal volume traces measured during an induced panic model. The hypothesis of the study is that impairing normal subjects' endogenous opioidergic system by naloxone makes them vulnerable to the respiratory effect of subsequent lactate paralleling the known effect of lactate specific to patient with panic disorder. The goal is testing if an opioidergin dysfunction may be the pathophysiological mechanism underlying panic disorder. We propose a functional hypothesis testing procedure to determine the significance of two curves with order restriction. One-degree of freedom test statistics summarizes the pointwise difference over the whole time lag. We use permutation to estimate the distribution of the test statistics concluding that there is a lactate effect which manifests itself as a steady tidal volume increase.

## 336 Categorical Data and Experimental Design

Section on Statistical Computing, Section on Quality and Productivity

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A Simple Method for Generating Multivariate Categorical Variates

❖ Hyunjip Choi, Kyonggi University, Department of Applied Information Statistics, Yiui dong Yeongtong gu, Suwon, 443-760 South Korea, *hjchoi@kyonggi.ac.kr*

***Key Words:*** random number, Peason chi-squared statistics, multidimensional contingency tables, log-linear models

We suggest a method for generating multivariate categorical variates based on the log-linear models. A linear combination approach by Lee(1997) is applied to get the joint distribution with the well known Pearson chi-squared statistics. By using the method, we can generate completely associated joint distributions which have the fixed association among three variables. Therefore the method could be extended to more higher dimension than three dimension.

### Testing Multinomial Categories with Sequential Sampling: Is a Wheel of Fortune Fair?

❖ Hokwon Cho, University of Nevada, Las Vegas, 4505 Maryland Parkway, Box 454020, Las Vegas, NV 89154-4020, *cho@unlv.nevada.edu*; Hai Zhen, University of Nevada, Las Vegas

***Key Words:*** testing, multinomial categories, Dirichlet type-II integrals, multiple decision theory, probability of correct decision, wheel of fortune

We study an inverse-type sampling procedure for testing k (< oo) multinomial categories with non-zero probabilities. An efficient stopping rule is devised to satisfy a pre-determined level of confidence, P*-condition. Dirichlet Type-II integrals will be primarily used for developing the procedure based on the decision theoretic point of view. Optimal sample sizes will be found along with the desirable probability level P{CD}, the probability of correct decision. Various configurations on "wheel of fortune" will also be considered for an illustration. Finally, a comparison with the existing method and the developed procedure will be discussed via Monte Carlo experimentation.

### A Comparison of Approximations for Logistic Regression

❖ Charity J. Morgan, Harvard University, Statistics Department, One Oxford Street, Cambridge, MA 02138, *cmorgan@stat.harvard.edu*; Donald Rubin, Harvard University

***Key Words:*** logistic regression

Estimation of the parameters of a logistic regression typically requires an approximation. Due to its computational simplicity, the approximation that is generally used is the asymptotic normal approximation to the logistic likelihood. Here, we will explore two other approximations, probit and robit likelihood approximations to the logistic likelihood, and compare those two to the asymptotic normal approximation. We find that we can improve upon the asymptotic normal approximation with very little additional computational effort. Also compared are the improvements that arise from using the SIR algorithm as an adjunct to improve any such approximation.

### Construction Methods for Balanced Incomplete Block Designs

❖ Jeffrey Shaffer, Tulane University, 5855 Walnut Creek Road, b210, River Ridge, LA 70123, *jshaffer@tulane.edu*; Sudesh Srivastav, Tulane University

***Key Words:*** diallel cross design, mating design, balanced incomplete block design, construction methods, breeding experiments, plant animal breeding

Diallel cross experiments as mating designs are used to select the best inbred lines with respect to general combining ability - the overall average performance of individual inbred lines in crosses - for additional breeding. Several methods of constructing optimal balanced incomplete block diallel cross designs are given. The methodology discussed involves finite field theory and supplementary difference set techniques. Construction methods are derived by using resolvable balanced incomplete block designs to select crosses from balanced incomplete block designs with nested rows and columns. Infinite series of particular classes of diallel designs are obtained.

### Asymptotic Properties of Effective Experimenting Strategies

❖ Anatoly Naumov, Novosibirsk State Technical University, P.O. Box 158, Novosibirsk-91, 630091 Russia, *a_a_naumov@mail.ru*

***Key Words:*** design of experiments, asymptotical effective design, sequential design, software, simulation

The purpose of this research consists in showing, that in conditions of the limited volumes of the data it is necessary to use effective planning of experiments. In particular, the influence of such volume of the data limitation on a modification of performances of the sequential strategies of experimenting with magnification of a total number of experiments is investigated. It is shown, that a posteriori variance of estimations of response function and variance of estimations of a vector of model parameters and the values, appropriate to them, criteria of an optimality of the designs of experiments can strongly differ from effective valuations, appropriate to them asymptotically, and, in particular, for small size of measurements (for small volume of the sample data). Algorithms and programs of construction of effective strategies on PC for the limited volumes of the data are considered.

### Misspecification Tests for Binomial and Beta-Binomial Models

❖ Marinela Capanu, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street, 3rd Floor, New York, 10021, *capanum@mskcc.org*; Brett Presnell, University of Florida

***Key Words:*** goodness-of-fit, lack-of-fit, misspecification, IOS, binomial, beta-binomial

The IOS test of Presnell & Boos (2004) is a general purpose goodness-of-fit test based on a ratio of in-sample and out- of-sample likelihoods. For large samples, the IOS statistic can be viewed as a contrast between two estimates of the information matrix that are equal under correct model specification. Both the IOS test and its large sample approximation are simple to compute and broadly applicable for testing the adequacy of parametric models. We compare the performance of IOS with existing goodness-of-fit tests for binomial and beta-binomial models. Our findings suggest that IOS is strongly competitive, not only with other general purpose tests, but even with tests designed especially for a specific model.

## 337 Teaching Statistics to Specific Audiences ●

Section on Statistical Education, Section on Statistical Graphics

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Teaching Effective Graph and Table Construction Needs More Attention in Statistical Education

❖ Thomas E. Bradstreet, Merck Research Laboratories, BL 3 2, 10 Sentry Parkway, Blue Bell, PA 19422, *thomas_bradstreet@merck.com*; Michael Nessly, Merck Research Laboratories; Thomas H. Short, Indiana University of Pennsylvania

***Key Words:*** effective communication, visual display, graphs and tables, interactive, workshop, statistics education

Communicating effectively through graphs and tables is a requisite skill for both students and working professionals, whether they are statisticians or nonstatisticians. Therefore, an important component of any statistical education curriculum is the mastery of basic definitions, strategies, and perceptual considerations in the design and construction of effective graphs and tables. Students should be introduced to these ideas prior to their undergraduate studies; reinforcement and expansion of the ideas should continue through their academic preparations and into their professional careers. Classroom instruction should be interactive with practical applications. We describe a new workshop-driven short course we constructed for professionals working in the pharmaceutical industry. The course content and activities generalize to other industrial, academic, and government settings.

### What Do M&Ms, Dahlias, Soil Erosion, and Data Analysis across the Curriculum Have in Common?

❖ Jerry Moreno, John Carroll University, Department of Mathematics and Computer Science, 20700 North Park Blvd, University Hts, OH 44118, *moreno@jcu.edu*

***Key Words:*** NCTM standards, k-12 curriculum, classroom visitations

The primary purpose of this presentation is to encourage you to become involved in your school's statistics offerings. Trust me; they need your help. By now, most states and local school districts are committed to NCTM's recommendations and principles (1989, 2000) on incorporating statistics into their K-12 mathematics curriculum, yet the implementation of such remains painfully slow (apart from AP Statistics whose growth has been truly amazing and inspiring). I will present my involvement in three classrooms over the last year and one school project with the hopes of your saying "I can do that, even better, and I will!"

### Instructional Tools in Educational Measurement and Statistics (ITEMS) for School Personnel: Evaluation of Two Web-Based Training Modules

❖ Rebecca Zwick, University of California, Santa Barbara, Gevirtz Graduate School of Education, Santa Barbara, CA 93106-9490, *rzwick@education.ucsb.edu*; Jeffrey C. Sklar, California Polytechnic State University, San Luis Obispo; Graham Wakefield, University of California, Santa Barbara

***Key Words:*** standardized testing, statistics education, psychometrics

In the current No Child Left Behind era, K--12 teachers and principals are expected to have a sophisticated understanding of standardized test results, use them to improve instruction, and communicate them to others. The goal of this project, funded by the National Science Foundation, is to develop and evaluate three web-based instructional modules in educational measurement and statistics to help school personnel acquire the "assessment literacy" required for these roles. The first module, "What's the Score?," was administered to 113 educators who also completed an assessment literacy quiz. Viewing the module had a small, but statistically significant, positive effect on quiz scores. In 2006, we will collect data on the effectiveness of our second module, which explains the impact of measurement and sampling error on test results.

### Interactive Animation for Learning IRT and Misfit Identification in Item Response Theory

❖ Chong Ho Yu, Arizona State University, 3S89 Computing Commons, Tempe, AZ 85287-0101, *chonghoyu@gmail.com*; Roger Freeman, Paradise Valley Unified School District; Angel Jannasch-Pennell, Arizona State University; Samuel DiGangi, Arizona State University; Chang Kim, Arizona State University; Victoria Stay, Arizona State University; Wenjuo Lo, Arizona State University; Lori Long, Arizona State University

***Key Words:*** item response theory, misfit, interface, assessment, measurement, Rasch model

Since the introduction of the No Child Left Behind Act, assessment has become a pre-dominant theme in the US K-12 system. However, making assessment results understandable and usable for the K-12 teachers has been a challenge. The objective of this presentation is to illustrate a well-developed interactive tutorial for understanding the complex concepts of assessment based on an interface study. For instance, there will be illustration of misfit identification in Item response theory, such as infit mean squares, outfit mean squares, infit standardized residual, and outfit standardized residual. User feedback is collected from item authors, teachers, and education majors. Usage of the tutorial is tracked by interface study procedures. Obstacles to understanding are identified for refining the tutorial delivery

### An Education Model for Secondary Students

John Aleong, University of Vermont; ❖ Chandra Aleong, Delaware State University, 1200 N. DuPont Hwy., Dover, DE 19901, *aleong@desu.edu*

***Key Words:*** transportation, secondary school, evaluation, case studies

The National Summer Transportation Institute (NSTI), which is funded by the U. S. Department of Transportation Federal Highway Administration (FHWA) provides opportunities for secondary school students to pursue careers in the Transportation industry. Students spend four weeks at several universities and colleges across the country engaged in practical workshops on statistics, mathematics, engineering, science and communications. Laboratory and field trips are an integral part of the program. Case studies of the program breath, history, evaluation and future will be presented. There will be a discussion of the unique emphases adopted at two completely different institutions and an attempt to develop an evaluation scheme for programs in successive years.

### A Pilot Survey on Ways Practicing Engineers Learn Statistics

❖ Jorge Romeu, Syracuse University, P.O. Box 6134, Syracuse, NY 13217, *jlromeu@syr.edu*

***Key Words:*** statistics education, data analysis, surveys

The statistics education of engineers is important because statistical methods are intimately linked to sound engineering practices. However, statistics is currently insufficiently, inadequately, and too theoretically taught in engineering schools. Hence many engineers have to learn it later. Very little is known about after-college statistical education. In his ICOTS-7 paper on Teaching Engineering Statistics to Practicing Engineers, Romeu develops this topic and conducts an initial survey (http://web.syr.edu/~jlromeu/SurveyICOTS.html) to support his research. Here, extensive statistical analysis of these data, including concomitant variables, is undertaken. We assess the effects of age, years of experience, academic level, areas of application, stats courses taken in college, etc. on the specific ways practicing engineers select to learn statistics on their own, after leaving college.

### Presenting Statistics Online to Nontraditional Business Students

❖ William Livingston, Baker College Center for Graduate Studies, 1116 W. Bristol Road, Flint, MI 48507-5508, *bill.livingston@baker.edu*

***Key Words:*** statistics education, online, non-traditional business students

The advent of online education has brought about particular challenges for statistics educators. This is particularly true when the audience consists of non-traditional business students who may lack traditional statistics courses in their undergraduate education. This article relates strategies that have proven effective in research and statistics courses offered to non-traditional business students in an online MBA program.

## 338 Innovative Approaches to Introductory Statistics ●

Section on Statistical Education

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Statistics: Telling the Whole Story

❖ Nancy Pfenning, University of Pittsburgh, 2316 Shady Ave., Pittsburgh, PA 15217, *nancyp@pitt.edu*

***Key Words:*** teaching strategies, big picture, global summary

Introductory statistics courses are typically so focused on presentation of individual topics that students “can’t see the forest for the trees.” A network of teaching strategies has been designed to convey to students the whole story, not just individual topics. Besides presenting these techniques, we report results for a survey that asked students to tell what they had learned overall in their statistics course. Students taught with the usual textbooks tended to report having learned specific topics, such as how to calculate a mean or do a z test. Students taught with the big-picture approach (using the author’s new textbook) were much more likely to provide a global summary, such as how data should be produced and summarized and how-to make inference about a population based on a sample. We will discuss benefits of training students to maintain a global perspective.

### Teaching a Writing-Centered Course in Introductory Statistics

❖ Deborah Lurie, Saint Joseph’s University, 1 West Levering Mill Road, Bala Cynwyd, PA 19004, *lurie@sju.edu*

***Key Words:*** education, writing, introductory statistics

Writing-centered courses at Saint Joseph University are courses in which writing serves as the primary means of learning and of evaluation. Assignments may include formal papers, journal writing, graded and upgraded reading responses and reaction papers and peer-editing exercises. Students receive instruction on their writing assignments and on peer editing and they are given the opportunity to revise some of their work. All writing centered courses are capped at 20 students so that students can receive extensive feedback. The courses are discussion -oriented, and they are designed to foster a community of writers within the classroom and a deeper understanding and appreciation of course content. This talk will highlight the lessons learned from using this approach to teach introductory statistics.

### Assessing the Gains from Concept Mapping in Introductory Statistics

❖ David Doorn, University of Minnesota, Duluth, 412 Library Drive, 165 SBE, Duluth, MN 55812, *ddoorn@d.umn.edu*; Maureen O'Brien, University of Minnesota, Duluth

***Key Words:*** concept map, interactive learning, introductory statistics, assessment

In an effort to improve active learning in introductory statistics, we introduce the use of concept mapping techniques as an integral and on-going part of the course. While previous papers have touted the use of this and other interactive teaching methods in statistics education, we add to this literature by providing additional assessment of its efficacy. This comes through an experimental design that involves a single instructor teaching two sections of the same statistics course over the same semester. Both cover the same material in the same way with the exception that concept mapping is used as an additional learning tool in one section, but not the other. Assessment of learning outcomes is done through the use of pre-tests and post-tests of understanding of statistical concepts. We also track changes in student's study habits over the semester through additional surveys.

### Further Assessment of Materials for Engaging Students in Statistical Discovery

❖ W. Robert Stephenson, Iowa State University, 327 Snedecor Hall, Department of Statistics, Ames, IA 50011-1210, *wrstephe@iastate.edu*; Amy Froelich, Iowa State University; William M. Duckworth, Iowa State University

***Key Words:*** active learning, activities, introductory level, laboratory

As part of an NSF funded project we developed new materials for the introductory statistics course designed to engage students in discovery. The materials are designed to actively involve students in the design and implementation of data collection and the analysis and interpretation of the resulting data. Previously, performance of students who used the materials was compared to the performance of students who did not use the materials. One drawback of that comparison was that there were different instructors. In this paper we compare the performance of students in a small honors section to that of students in a second section of introductory statistics. The same instructor taught both sections. Preliminary results on students' performance on test items and projects will be presented. Additionally, attitudes towards the new materials and statistics in general will be presented.

### Sequencing of Topics in an Introductory Course: Does Order Make a Difference?

❖ Christopher Malone, Winona State University, Department of Mathematics and Statistics, PO Box 5838, Winona, MN 55987, *cmalone@winona.edu*; John Gabrosek, Grand Valley State University

***Key Words:*** sequencing, introductory, topics, order

The procedural steps involved in completing a statistical investigation are often discussed in an introductory statistics course. For example, students usually learn about study design and data collection, appropriate descriptive summaries and graphical displays, and completing necessary inferential procedures. The traditional sequencing of topics in an introductory course places statistical inference in the latter portion of the course. As a result, students have limited opportunities to perform a complete statistical investigation. We propose a new sequencing of topics that may enhance a student's ability to perform a statistical investigation from beginning to end. Assessment items were given to students taught under the traditional sequencing of topics and the new proposed sequencing of topics. The results from this assessment will be discussed in detail.

### A Geometrical Approach to Introductory Statistics

❖ Daniel Kaplan, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105, *kaplan@macalester.edu*

***Key Words:*** geometry, education, regression

At Macalester College we have developed a new approach to introductory statistics with an emphasis on understanding complicated relationships among multiple variables. The course is oriented around multiple regression and multi-way analysis of variance and makes extensive use of modern computation. In our course, students learn about least-squares regression in terms of simple projections in two and three dimensions; they calculate correlation coefficients using a protractor; they perform analysis of variance using a ruler. The geometrical approach, in addition to being accessible to just about every student, makes it possible for students to reason about the results of statistical modeling in sophisticated ways. For example, Simpson's paradox does not seem like a paradox at all when presented geometrically --- it's rather obvious.

### Toward Statistical Literacy: Statistics in Everyday Life

❖ Hoke Hill, Jr., Clemson University, Applied Economics and Statistics Department, F-148 Poole Ag Center, Clemson, SC 29634, *hhill@clemson.edu*; William Bridges, Jr., Clemson University; Rose Martinez-Dawson, Clemson University

***Key Words:*** statistical literacy, general education

Statistics in Everyday Life is a course developed to meet a new Science and Technology in Society general education requirement. The purpose of the course is to develop students' statistical literacy through critical interpretation of statistical and probability statements encountered in society. This team-taught course has been offered in Spring 2005, Fall 2005, and Spring 2006. The course is organized into a series of topics of interest to students (polls, meteorology, advertising, lottery, etc.). The authors will discuss how the basics of probability, descriptive statistics, and statistical inference can be incorporated into such a course. The evolution of topic selection and organization will be discussed. Examples of class materials, student assignments, classroom activities, and student group developed web pages will be shown. Student feedback of the course will be presented.

## 339 Regression and Time Series

Section on Statistical Computing

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Automatic Detection of Outliers Based on the Forward Search

Matilde Bini, University of Florence; Bruno Bertaccini, University of Florence; ❖ Franco Polverini, University of Florence, Viale Morgagni 59, Department of Statistics, Florence, 50134 Italy, *polverin@ds.unifi.it*

***Key Words:*** cut off point, forward search, robustness

The article was stimulated by the work of Atkinson and Riani on the estimation of regression models using a robust methodology called "forward search" that seems to work well in the estimation of a variety of models particularly when part of the data are generated by models different from the one we intend to estimate. The methodology detects the presumed outliers and estimates the models without them. The weak point of the procedure is, in our opinion, that the choice of the subset of the data to use for this robust estimation relies on the behavior of some statistics as one adds observations to an initial small set of data. The aim of the paper is to propose some alternative and automatic ways to make this choice that would allows the running of simulations to assess the properties of the estimators and the power of the various tests used in the estimation of the models involved.

## A Novel Statistical Approach to Identifying and Limiting the Effect of Influential Observations

❖ Tamekia Jones, The University of Alabama at Birmingham, 327 Ryals Public Health Building, 1665 University Blvd, Birmingham, AL 35294, *tljf81@uab.edu*; David Redden, The University of Alabama at Birmingham

***Key Words:*** outliers, leverage, influential observations, robust regression, MCD

Outliers are observations with extreme standardized deviations between the observed dependent variables and their predicted values. Outliers can be detected using studentized residuals. Leverage is a measure of the standardized deviation of an observation's independent variable from the mean of the independent variables and is assessed using the diagonal of the hat matrix. An influential observation (outlier and leverage point) affects estimation of regression parameters. Detection of these points is difficult due to the masking effect which occurs when influential points are hard to identify using regression diagnostics. We present a robust regression method that extends Rousseeuw's concept using the minimum covariance determinant. Results utilizing our proposed method illustrate that our approach overcomes the masking effect by properly identifying influential observations.

## Robust Winsorized Regression Using Bootstrap Approach

Deo Kumar Srivastava, St. Jude Children's Research Hospital; ❖ Jianmin Pan, St. Jude Children's Research Hospital, 332 N. Lauderdale Street, Department of Biostatistics, Memphis, TN 38105, *jianmin.pan@stjude.org*; Ila Sarkar, Louisiana Health Care Review, Inc.

***Key Words:*** linear regression, winsorization, robustness, bootstrap

In linear regression the explanatory variables are customarily assumed fixed and one models the relationship between the response and explanatory variables. However, in practice, one collects information on several factors and models one as a function of others. The least squares (LS) estimation forms the backbone of classical regression analysis. However, this approach is highly sensitive to "outliers" in both response and explanatory variables and several methods to safeguard against them have been proposed. In this paper we propose Winsorized regression using bootstrapping for estimating and testing the parameters when the data are from symmetric populations and evaluate their asymptotic properties. The simulation results indicate that the new method provides significant improvement over LSE if the data are from non-normal models with minimal loss in power if normality holds.

## Bent-Cable Regression with Autoregressive Noise

❖ Grace Chiu, University of Waterloo, Department of Statistics and Actuarial Science, Waterloo, ON N2L 3G1 Canada, *gchiu@uwaterloo.ca*; Richard Lockhart, Simon Fraser University

***Key Words:*** autoregressive, bent-cable regression, change-point, time series

Bent-cable regression extends the popular piecewise linear (broken-stick) model, allowing for a smooth change region of any non-negative width. Existing bent-cable methodology assumes independent and identically distributed errors. In this talk, we investigate data that exhibit a bent-cable mean structure with noise generated by a low order autoregressive model. We suggest a somewhat unusual approach for developing asymptotics of both regression and time series parameter estimators. Preliminary results based on simulations and physiological and atmospheric datasets are presented.

## Multiresolution Outlier Detection for Long-Range Dependent Time Series

❖ Lingsong Zhang, The University of North Carolina at Chapel Hill, Department of Statistics, CB 3260 UNC Chapel Hill, Chapel Hill, NC 27599-3260, *lszhang@email.unc.edu*; Zhengyuan Zhu, The University of North Carolina at Chapel Hill

***Key Words:*** outlier, long range dependence, time series, multiresolution analysis

In the context of Internet traffic abnormally detection, we find that some outliers in a time series are difficult to detect in a given scale while rather obvious under a different scale. We explore the outlier effect when the underlying time series have long range dependence, and conclude that testing outliers at multiple time scales helps to reveal outlier structure in the time series. We propose a new visualization tool, the outlier map, which can help identify outliers in long range dependent time series.

## Restricted Error Regression

❖ James Cochran, Louisiana Tech University, P.O. Box 10318, Ruston, LA 71272, *jcochran@cab.latech.edu*

***Key Words:*** regression, optimization

We consider the ramifications of restricting regression analyses so the error terms satisfy the conditions necessary for valid inference (homoskedasticity, normality, independence, and fit). Specifically, we assess the difficulty and time required to solve the resulting constrained nonlinear optimization problem. We also consider measures of divergence of the resulting regression from the unrestricted regression and how to use this divergence to measure how well the unrestricted regression satisfies these conditions, as well as the validity of standard summary statistics and hypothesis tests for the restricted error regression.

## Local Linear Estimation for Single-index Conditional Quantiles

❖ Zhou Wu, University of Cincinnati, 2921 Scioto Lane 1005, Cincinnati, OH 45219, *wuzu@email.uc.edu*; Yan Yu, University of Cincinnati; Keming Yu, Brunel University

***Key Words:*** bandwidth selection, consistency, pilot estimator, single-index models

Quantile regression has recently gained much attention for its ability to model conditional distribution of the response variable given the covariates. This paper concerns single-index models for conditional quantiles, which inherit advantages of such models in mean regression context: the unknown link function allowing flexible curvature and the linear index allowing interpretability of parameters. A local linear approach to estimating such models is proposed by minimizing a double weighted check function. With the pilot estimator assumption together with other mild conditions, we study the large sample properties of estimators for both the nonparametric link function and parametric index parameters. Both real data example and simulation results are provided to illustrate the new methodology.

# 340 Misclassification and Measurement Error: Problems and Solutions ●

Biometrics Section, ENAR

**Tuesday, August 8, 2:00 pm–3:50 pm**

## Measurement Error Correction Methods Based on Cumulative Average Intake for Nutritional Data

❖ Weiliang Qiu, Harvard Medical School, 181 Longwood Ave., Channing Laboratory, Boston, MA 02115, *weiliang.qiu@gmail.com*; Bernard Rosner, Harvard Medical School

***Key Words:*** measurement error, regression calibration, nutritional data

In nutritional studies, surrogate measures such as the food frequency questionnaire (FFQ) often are measured with systematic and random error. Rosner et al. (1989) derived interval estimates to correct for measurement error at one time point based on regression calibration. In this article, we extend this approach to the case where FFQs are available at several time points. We build a logistic regression model to relate disease to cumulative average intake based on six FFQs obtained between 1980--1998 and propose a method to correct the RRs and CIs for measurement error using multiple imputation to handle missing FFQs at particular time points. The approach is illustrated with data from the Nurses' Health Study, relating incident breast cancer between 1980--2000 to calorie-adjusted saturated fat intake.

## A Robust Likelihood-Based Approach to Nonlinear Measurement Error Models with Application to Radiation Dose Effects on Leukemia-Specific Hazard Rate among A-Bomb Survivors

❖ Antara Majumdar, University at Buffalo, Department of Biostatistics, 264A Farber Hall, 3435 Main St. Bldg. 26, Buffalo, NY 14214, *aroy3@buffalo.edu*; Randy L. Carter, University at Buffalo

***Key Words:*** measurement error model, instrumental variable, maximum likelihood, Monte Carlo EM algorithm, radiation effects, leukemia survival data

We propose a robust likelihood-based method to estimate the parameters of measurement error models. We combine three key points: First, the errors-in-variables problem can be viewed as a missing data problem. Second, the measurement error problem can be solved with availability of an instrumental variable. Third, the likelihood based on piece-wise exponential survival data is the same as the likelihood based on Poisson distribution. We construct a robust multivariate likelihood of the manifest variables utilizing all the above information and use the MCEM algorithm to maximize the observed data likelihood. Our method accommodates other covariates that are error-free, as well as facilitates testing of regression parameters. This work was motivated by the ongoing research at RERF. We illustrate our method by analyzing the incidence of deaths due to leukemia among A-bomb survivors.

## Varying Coefficient Model with Measurement Error

❖ Liang Li, The Cleveland Clinic, Department of Quantitative Health Sciences, 9500 Euclid Ave Wb4, Cleveland, OH 44195, *lil2@ccf.org*; Tom Greene, The Cleveland Clinic

***Key Words:*** measurement error, varying coefficient model, local polynomial, chronic renal disease, semi-parametric

We propose a semi-parametric partially varying coefficient model to study the creatinine clearance process in various populations. The coefficient of GFR is specified as a function of age to allow its effect to be age-dependent. The GFR measurements are assumed to be surrogates for the true GFR, their relationships modeled in an additive error model. We use local corrected score equations to estimate parameters and coefficient functions, and use an expected generalized cross-validation method to select the kernel bandwidth. The proposed method avoids making assumptions on the distributions of true GFR and residuals. An application of the proposed model reveals that there might be a consistent creatinine clearance process among healthy kidney donors and chronic renal disease patients.

## Conditional Score Methods for Regression Models with Poisson Surrogates

❖ Alvin Van Orden, North Carolina State University, 208 N. Dixon Ave., Cary, NC 27513, *acvanord@ncsu.edu*; Leonard A. Stefanski, North Carolina State University

***Key Words:*** measurement error, Poisson surrogate, conditional score

We study estimation in linear models in which one covariate is unobservable and an observable surrogate for the latent variable has a Poisson distribution. Our model is similar to that of Li, Palta, and Shao (2004) who used the method of corrected score functions to derive consistent estimators. Our model is amenable to the conditional score method of Stefanski and Carroll (1987). We develop the associated estimators and study their properties analytically and via simulation.

## Effects of Misclassification on Exposure-Disease Association with Odds Ratio

❖ Jun-mo Nam, National Cancer Institute, EPS, Room 8028, 6120 Executive Blvd, Rockville, MD 20892-7240, *namj@mail.nih.gov*

***Key Words:*** bias, efficiency, interval estimation, misclassification, odds ratio

Effects of the exposure misclassification in inference on an exposure-disease association with the odds ratio are a major concern in case-control studies. When the misclassification is independent of case or control (non-differential misclassification), we expect the estimated odds ratio is biased towards to the null value. However, the estimate for a single study may overestimate the true odds ratio by chance. We investigate statistical methods for identifying non-differential misclassification. Regardless of differential or non-differential misclassification, it causes bias and a loss of efficiency in interval estimation of the odds ratio. If misclassification is present, the actual coverage probability of a confidence interval for the odds ratio is smaller than a nominal one and the expected width of the confidence interval is broader.

### Adjusting for Misclassification in Binary Response Variables

❖ Jun Zhai, Duke University, 2117 CIEMAS DUMC, Box 3382, Duke University, Durham, NC 27708, *jun.zhai@duke.edu*; John Olson, Duke University; Mike West, Duke University

***Key Words:*** binary, misclassification, adjusting, logistic

Binary regression models, such as logistic models, are widely used in biological research, in areas such as prognostic modeling in clinical genomics. However, when there is a significant rate of misclassification of the reported status of binary response variables, such models can fail miserably. We address this in utilization of a logistic model adjusting for misclassification of response variables. The adjusted model can be fitted either using MCMC methods for posterior evaluation, or using Newton-Raphson algorithm to directly compute MAP/ML estimates. As an example, we have applied and evaluated the adjusted model in predicting lymph node (LN) status of breast cancer patients. There is a 10-20% false-positive misclassification that this approach addresses and adjusts for. We discuss some computational issues in evaluating the adjusted models as well as this application.

## 341 QT/QTc ●

Biopharmaceutical Section
**Tuesday, August 8, 2:00 pm–3:50 pm**

### Noninferiority Testing in Thorough QT/QTc Studies

❖ Balakrishna Hosmane, Northern Illinois University, Division of Statistics, DeKalb, IL 60115, *bala@math.niu.edu*; Charles Locke, Abbott Laboratories

***Key Words:*** parallel design, cross-over design, thorough QT/QTC study, mixed effect model

An assessment of non-inferiority of the higher dose to placebo is performed in thorough QT/QTc studies by the intersection-union test within the framework of a linear mixed effects analysis with both parallel and cross-over designs. For the purposes of planning such a study, the joint distribution of the estimate of the difference in means of high dose of investigational drug and the placebo was derived. The power of 'thorough' QT/QTc study evaluated using the joint distribution and the simulation study were quite close.

### Bias in Estimates of QTc Prolongation by Timepoint-Wise Treatment Comparison

❖ Yibin Wang, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, *yibin.wang@novartis.com*; Guohua Pan, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** QT/QTC interval, bias, ICH E14, QT prolongation, primary endpoint

Recent step 4 E14 guidance proposed the primary end point for QTc assessment as the change from baseline in QTc interval at timepoint corresponding to the maximum increase in subject-averaged QTc interval compared with placebo. The choice of the primary end point assumes that the time of maximum increase occurs at the same timepoint for all subjects, and coincides with the maximum effect of the treatment. Use of the above primary end point may result in underestimation of drug effect on QTc prolongation because of variation of the tmax. On the other hand, selection of the timepoint may lead to upwards bias in estimates of treatment effect by random noise that contributes high on-drug values or low placebo values. This study examines and quantifies the potential bias in estimates of QTc prolongation based on the primary end point under different exposure-QTc-response relationships.

### Tolerance Regions for Multivariate Random Effects Model with Application to Beat-to-Beat QT-TQ Interval Data

❖ Kimberly Crimin, Pfizer Inc., Eastern Point Road, MS82602313, Groton, CT 06340, *kimberly.crimin@pfizer.com*; Joseph McKean, Western Michigan University

***Key Words:*** QT prolongation, GEE, tolerance regions, random effects

QT prolongation has been responsible for the withdrawal and re-labeling of numerous compounds, making it important to understand how the ECG signal varies over time. The development of automatic ECG signal acquisition systems has made it possible to collect continuous ECG data. The duration of any cardiac cycle is largely dependent on the preceding diastolic interval, measured by the TQ interval. When the TQ interval is extinguished, the probability of an arrhythmia increases. Tolerance regions are used in clinical applications to define the 'normal' range of clinical or laboratory measurements. We present a method to calculate tolerance regions for bivariate random effects models and apply these methods to continuous QT-TQ interval data collected on 'healthy' individuals. The variance was estimated using GEE. We investigate methods for computing tolerance factors in a MC study.

### A New Method for Sample Size Calculation of the Thorough QT/QTc Study

❖ Shu Zhang, Sepracor, Inc., 84 Waterford Drive, Marlborough, MA 01752, *shu.zhang@sepracor.com*

***Key Words:*** thorough QT/QTC study, sample size calculation, criteria for negative TQS, maximum change

The ICH has recently issued a guideline (E14) to the pharmaceutical industry to address QT interval prolongation. The guideline proposed conducting a "Thorough QT/QTc Study (TQS)" and defined two criteria for a "negative TQS". Several methods have been proposed for the sample size calculation of the TQS. However, most methods only considered mean change instead of the maximum change as defined by the guideline, and the impact of the two criteria on the traditional framework of hypothesis testing was not examined. In this work, we examine the impact of these criteria on the hypothesis testing framework, derive the distribution of maximum change, propose a unified approach for the sample size calculation specifically designed to address the two criteria, calculate the sample size based on different assumptions, and compare our results with those using some other approaches.

### Bayesian Measurement Error Approach to QT Interval Correction/Prolongation

❖ Jie Chen, Merck Research Laboratories, Investigational Research, BL32, Blue Bell, PA 19422, *jie_chen@merck.com*; Xin Zhao, Merck & Co., Inc.

***Key Words:*** heart rate correction, QTC interval prolongation, measurement error, hierarchical model

Traditionally regression methods are applied to obtain the heart rate corrected QT intervals (QTc) which are then compared between baseline and treatment groups. These regression models rely on a fundamental assumption that the predictor variable (heart rate or RR interval) is measured without error. However, in reality both QT and RR intervals measured in ECG are subject to not only measurement error, but also fluctuations caused by numerous factors. In this paper we propose a Bayesian measurement error model that does not require assuming that the RR intervals are measured without error. The proposed procedure provides a unified way for within-treatment subject-specific QT interval corrections and for statistical inference on QTc prolongation that might be related to a drug. Finally the procedure is illustrated using a real example and a simulation study.

### Designing and Analyzing Covariate-Adjusted Response Adaptive Randomized Clinical Trials

❖ Ayanbola Ayanlowo, The University of Alabama at Birmingham, RPHB 327M, 1530 3rd Ave South, Birmingham, AL 35294-0022, *aayanlowo@mail.dopm.uab.edu*; David Redden, The University of Alabama at Birmingham

***Key Words:*** response adaptive, covariate-adjusted, covariate by treatment interaction, randomized clinical trials

Response adaptive randomized clinical trial designs adjust the probability of treatment assignment based upon accrued treatment effect information. Recently in literature, the concept of covariate-adjusted response adaptive randomized designs has been proposed. This method adjusts for both the subjects' response and covariate values to determine the treatment assignment probabilities. We examine the statistical properties of covariate-adjusted response adaptive randomized designs in the presence of covariate by treatment interaction. Specifically, we present methods for analyzing covariate-adjusted response adaptive randomized trials and examine the statistical power of these designs in the presence of covariate by treatment interaction. Simulations will be presented illustrating the properties of these methods under different types of covariate by treatment interactions.

## 342 Methods for Design and Analysis of Microarrays ●

Biometrics Section, ENAR

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Detecting Differential Expressions in GeneChip Microarray Studies: a Quantile Approach

❖ Huixia Wang, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 S Wright St, Champaign, IL 61820, *hwang22@uiuc.edu*

***Key Words:*** gene expression, quantile regression, microarray data, random effect, rank score test

We consider testing for differentially expressed genes in GeneChip studies by modeling the quantiles of gene expression through probe-level measurements. By developing a robust rank-score test for linear quantile models with a random effect, we propose a reliable test for detecting differences in certain quantiles of the intensity distributions. By using a genome-wide adjustment to the test statistic to account for within-array correlation, we demonstrate that the proposed rank-score test is highly effective, even when the number of arrays is small. Our empirical studies show that detecting differences in the quartiles for the probe-level data is a valuable complement to the usual mixed-model analysis based on Gaussian likelihood. The methodology development in this paper is a first attempt at developing inferential tools for quantile regression in mixed models.

### Gene Selection Using a Modified Mahalanobis Distance

❖ Xiwen Ma, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, *xiwenma@stat.wisc.edu*

***Key Words:*** Mahalanobis distance, eigenvalue, eigenvector, maximum likelihood estimate, non-parametric

A fundamentally important question in microarray experiments is to locate differentially expressed genes for disease diagnosis. We consider the experiment which compares two sources of RNA between diseased and healthy subjects. The key problem is to well define a criterion of discrimination and to detect an optimal subset of genes according to it. We propose a modified Mahalanobis distance which is independent of distribution assumptions and can be used in detecting an optimal subset of genes. We also provide a stepwise method to locate the optimum subset and apply to well known benchmark public datasets and compare with existing approaches.

### Analysis of Microarray Data: Two Examples

❖ Meng Du, University of Toronto, Statistics, 100 St Georger St, Toronto, ON M5S 3G3 Canada, *meng@utstat.utoronto.ca*; Muni Srivastava, University of Toronto

***Key Words:*** DNA microarrays, high dimensional data, global mean test, multiple testing

Two microarray datasets are analyzed. The correlation structures of the datasets are first examined using the diagonality and sphericity tests by Srivastava (2005, 2006). A global mean test by Srivastava and Du (2006) is applied to check whether the genes of the two groups in each dataset are differentially expressed. And several newly proposed multiple testing procedures are used to discover which genes may have caused the difference. All the methods applied are for high dimensional data, particularly with the sample size smaller than the dimension.

### Multidimensional Local False Discovery Rate for Microarray Studies

❖ Yudi Pawitan, Karolinska Institutet, Nobels vag 12A, Dept of Medical Epi and Biostat, Stockholm, 17177 Sweden, *yudi.pawitan@ki.se*; Alex Ploner, Karolinska Institutet

***Key Words:*** gene expression, mixture model, permutation test, smoothing

Background: The false discovery rate (fdr) is a key tool for statistical assessment of differential expression (DE) in microarray studies. It is, however, well known that overall control of the fdr alone is not sufficient to address the problem of genes with small variance, which generally suffer from a disproportional high rate of false positives. Graphical tools and modified test statistics have been proposed for dealing with this problem, but there is currently no procedure for controlling the fdr directly. Methods: We generalize the local fdr as a function of multiple statistics, combining a common test statistic for assessing differential expression with standard error information. Results: Experience with simulated and real microarray data shows that the multidimensional local fdr allows an objective and powerful assessment of DE as a function of gene variability.

### Optimal Cross-Validation for Estimating Small Sample Classification Error Rate

❖ Wenjiang Fu, Michigan State University, West Fee Hall, Suite B601, East Lansing, MI 48824, *fuw@msu.edu*; Suojin Wang, Texas A&M University

***Key Words:*** .632+ bootstrap, cross-validation, leave-one-out, microarray data analysis, misclassification

Estimation of misclassification rate is important in many studies, especially in the analysis of high-throughput microarray data, where technology cost usually limits study sample size from 20 to 100. Although methods of estimating misclassification rates have been studied, many of them fail to yield satisfactory results due to small sample size, such as cross-validation, leave-one-out bootstrap (LOOBT), .632 bootstrap (BT632), etc. Although the .632+ bootstrap method (Efron and Tibshirani 1997) improves the .632 estimator by further correcting its bias, it does not perform well in terms of the overall mean squared error (MSE). In this study, we propose an optimal cross-validation method that aims to minimize the overall MSE. Results from simulation and microarray data analysis demonstrate that our novel method performs the best among its major competitors, including LOOBT and BT632.

### Gene Expression (Microarray) Analysis by Neural Networks

❖ David Booth, Kent State University, Management and IS Department, College of Business Admin, Kent, OH 44242, *dbooth@bsa3.kent.edu*; David Zhu, Kent State University; Richard Geoke, Kent State University; David Baker, Kent State University; James Hamburg, Kent State University

***Key Words:*** microarray, gene expression, leukemia, neural networks, cross validation, cancer

Three sets of optimal leukemia class predictors (genes) were obtained by three methods from three authors from the same set of data. We tested these optimal sets using back propagation neural networks (which were not used by the original authors) with three-fold cross-validation and leave-one-out cross-validation. We found that the predictor sets performed poorer with the neural networks than with the original methods, though not in all cases. We discuss this result and suggest methods for possibly taking advantage of this finding.

## 343 Analysis of Correlated Data and Mixed Models ●

Biometrics Section

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Regression Analysis of Panel Count Data with Dependent Observation Times

❖ Xin He, University of Missouri-Columbia, 146 Middlebush Hall, University of Missouri - Columbia, Columbia, MO 65211, *xhw4c@mizzou.edu*; Jianguo Sun, University of Missouri-Columbia; Xingwei Tong, University of Missouri-Columbia

***Key Words:*** estimating equation, informative observation times, mean function model, panel count data, regression analysis

Panel count data often occur in long term studies that concern recurrence rate of a recurrent event. Methods have been proposed for regression analysis of panel count data, but most of the existing researches focus on situations where observation times are independent of longitudinal response variables and therefore rely on conditional inference procedures given the observation times. This paper considers a different situation where the independence assumption may not hold. That is, the observation times and the response variable may be correlated. For inference, estimating equation approaches are proposed for estimation of regression parameters and both large and finite sample properties of the proposed estimates are established. An illustrative example from a cancer study is provided.

### Modellng Variability in Longitudinal Data Using Random Changepoint Models

❖ Annica Dominicus, Karolinska Institutet, Box 281, Stockholm, 17177 Sweden, *annicad@math.su.se*; Samuli Ripatti, Karolinska Institutet; Juni Palmgren, Karolinska Institutet

***Key Words:*** change-point model, non-linear mixed model, first order linearization, Gibbs sampling

Some cognitive functions exhibit multiple phases in old age, which motivates the use of a change point model for the individual trajectory from repeated measures data. The change point varies between individuals and is treated as random. We contrast the random change point model with linear and quadratic random effects models, focusing primarily on trait variability over age groups. The methods are illustrated using Swedish data on cognitive function in old age and through simulations. We show that the models impose different restrictions on the trait variance over age groups, and we demonstrate that the random change point model has favorable properties. As estimation methods we discuss the performance of approximate maximum likelihood estimation based on first order linearization of the random change point model as well as a Bayesian Gibbs sampling procedure.

### Semiparametric Modeling with Correlated Data

❖ Chun Han, The University of Kansas, 405 Snow Hall, 1460 Jayhawk Blvd, Lawrence, KS 66045, *chan@math.ku.edu*

***Key Words:*** correlated data, cross-validation, longitudinal data, marginal models, partial linear models, penalized likelihood

In longitudinal/clustered data analysis, the popular marginal partial linear models are Y_{ij}=f(t_{ij})+ X_{ij} \beta + \epsilon_{ij}, where Y_{ij} is the j-th measurement taken on the i-th subject at time t_{ij}, f is an unknown smooth function, X_{ij}'s are covariate variables, and \epsilon_{ij}'s are random mean 0 errors with possible within-subject correlation. Assuming that the correlation structure of the random errors depends on a few parameters, we consider two estimation procedures using smoothing splines and propose a method to jointly select the smoothing parameter and the correlation parameters. The asymptotic optimality of the estimates was established together with the \sqrt{n} convergence rates of the correlation parameters and the \sqrt{n} normality of the \beta estimates. Application to the CD4 data is also presented.

### Marginal Mixture Analysis of Correlated Bounded-Response Data with an Application to Ultrasound Risk Assessment

❖ Yan Yang, University of Illinois at Urbana-Champaign, 101 Illini Hall, MC374, 725 South Wright St, Champaign, IL 61820, *yanyang@uiuc.edu*; Douglas Simpson, University of Illinois at Urbana-Champaign

***Key Words:*** correlated data, generalized estimating equation, latent model, Tobit model, two-part model, zero-inflated

Data with bounded responses are common in many areas of application. Often the data are bounded below by zero with a substantial portion of zeros, so ordinary generalized linear models fail. Three methods in the literature for modeling zero-inflated data are left-censored regression models, two-part models, and latent mixture models. We develop a general class of zero-inflated mixture (ZIM) models that unifies and generalizes these three classes of models. Novel diagnostics are proposed for assessing the adequacy of a ZIM model. We extend ZIM models to correlated data by adapting the theory of generalized estimating equations, and provide risk threshold estimates for bounded adverse outcomes. We illustrate the issues and methodology in the context of an ultrasound safety study of the occurrence and extent of lung hemorrhage in laboratory animals due to focused ultrasound exposure.

### On the Analysis of Mouse Preference Data

❖ Yang Yang, University of Western Ontario, Department of Statistics and Actuarial Sciences, London, ON N6A 5B7 Canada, *yyang42@uwo.ca*; Willard J. Braun, University of Western Ontario

***Key Words:*** generalized linear models, ordinal predictor, data analysis

Data have been collected from simple choice experiments on female mice. The mice are placed at one end of a Y-shaped maze and monitored to see which arm of the maze they are attracted to. The main interest is in detecting a female's attraction or repulsion from the odor of a parasite-infected male. In particular, if another female's odor is combined with the male odor, does this make the infected male seem "safer"? How familiar does the female have to be to counteract the natural repulsion? And are there longitudinal effects? These kinds of questions are addressed through the analysis of these data.

### Equality of REML and ANOVA Estimators of Variance Components in Unbalanced Models

❖ Shaun Wulff, University of Wyoming, Department of Statistics, PO Box 3332, Laramie, WY 82071-3332, *wulff@uwyo.edu*

***Key Words:*** unbiased estimator, MINQUE, UMVUE, classification models, random effects

For completely balanced normal classification models, it is well known that REML and ANOVA estimators of variance components are equal. This talk will demonstrate extensions of this result to classes of partially balanced designs.

### Measuring the Mean Squared Error of the EBLUP in Linear Mixed Models

❖ Jamie McClave Baldwin, Info Tech, Inc., 5700 SW 34th Street, Suite 1235, Gainesville, FL 32608, *jamie.baldwin@infotechfl.com*; Ramon Littell, University of Florida

***Key Words:*** linear mixed models, EBLUP, MSEP, Kenward-Roger adjustment, propagated variance

In 1984, Kackar and Harville (JASA, 79) introduced a new method to handle the propagation of variance stemming from unknown variance parameters in linear mixed models. In the 20+ years since this development there have been many adjustments, expansions and new techniques stemming from their idea. In this paper, we review the (frequentist) procedures now available for adjusting the Mean Squared Error of Prediction (MSEP) of the EBLUP in the setting of Linear Mixed Models to account for estimating unknown covariance parameters. While much has been done to improve the estimation of the MSEP, many analysts do not understand when or why it is important to make these corrections or even what the corrections are. This paper attempts to answer the questions: What is the precision of the EBLUP? How can we measure it? When and how should we adjust for the extra variation?

## 344 Quality Applications and Approaches ● ✪

Section on Quality and Productivity, Section on Physical and Engineering Sciences

**Tuesday, August 8, 2:00 pm–3:50 pm**

### Economic Decisionmaking Challenges in Quality Applications

❖ John Barrett, University of North Alabama, 314 Wright Drive, Florence, AL 35633, *jdbarrett@una.edu*

***Key Words:*** statistical process control, economic design, elasticity of demand, process capability

The area of quality improvement, specifically statistical process control (SPC), has been assessed with respect to many metrics. The focus of the SPC literature has been improvement of control charting techniques, with primacy given to statistical considerations. Economic concerns often have been considered as a secondary issue, primarily with regard to process costs and control chart design. With respect to the larger corpus, it often is assumed that any efforts to improve quality are to be undertaken. The purpose of this paper is to review the high points of past work and offer insights in the area of economics in quality and SPC.

### Methods of Data Quality

❖ Thomas Herzog, U.S. Department of Housing and Urban Development, 1630 Autumnwood Drive, Reston, VA 20194, *thomas_n._herzog@hud.gov*; William E. Winkler, U.S. Census Bureau; Fritz J. Scheuren, National Opinion Research Center

***Key Words:*** missing data, logical restraints on data, name and address matching

This paper provides an overview of methods of quality. In addition to listing a number of basic data quality tools, it describes the Fellegi-Holt model of statistical data editing (1976) and imputation extensions due to Little and Rubin (2002) and the Fellegi-Sunter model of record linkage (1969). The paper provides a number of applications and examples involving retail establishments, medical and biomedical applications, highway safety, list frames and national registers, social insurance, and financial record systems. A number of the examples and ideas are from a forthcoming textbook by the authors.

### Quality Techniques in the Mill Benefit Young Trees in the Field

❖ Bonnie P. Dumas, MeadWestvaco Corporation, Forestry Division, P.O. Box 118005, Charleston, SC 29483, *bonnie.dumas@meadwestvaco.com*

***Key Words:*** quality, forestry, statistical process control (SPC), Six Sigma

A case study examines whether quality management tools useful to meet carton specifications in a mill assembly line are successful in a forestry research initiative to identify best quality loblolly pine trees. A research tree improvement goal is to identify trees having the best growth, form and disease resistance traits. Toward this end alpha lattice experimental designs are used to control variation in testing over a thousand loblolly varieties, costing many thousands of dollars, at multiple locations. Six Sigma methodology used in our packaging mills was applied to this environment. The result was identification of a quality goal to reduce misplacement of young trees in field planting spots, a problem in implementing the complex designs. The outcome is an improved process of study installation that insures the correct tree placement that is imperative for high quality research data.

### Interlaboratory Study of DEA Proficiency Test Using RExcel

❖ Hung-kung Liu, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *hung-kung.liu@nist.gov*; Adriana Hornikova, National Institute of Standards and Technology

***Key Words:*** uncertainty, interlaboratory study, proficiency test, Rexcel

The Office of Forensic Sciences of DEA performs proficiency tests to monitor the measurement quality of their field laboratories and to obtain uncertainty estimators for their future measurements. Over the past five years 125 samples of different chemicals with various concentrations were measured by different methods from multiple chemical laboratories. We performed detailed statistical analyses on a selected sample of 25 different concentrations of the drug methamphetamine. For a total of 515 observations collected in a very un-balanced design, we performed exploratory data analysis, multiple factor analysis of variance and random effects model analysis using both graphical and quantitative approaches. The goals of this study are to better control the quality of illegal drug measurements and to provide customized self-evaluating RExcel workbook for DEA field laboratory.

### Statistical Data Processing of GC-MS Data

❖ Shiying Wu, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *swu@rti.org*; Jun Liu, Research Triangle Institute; Robert J. Morris, RTI International; James Raymer, RTI International; Ye Hu, RTI International; Larry Michael, RTI International

***Key Words:*** GC/MS, biomarker, chromatogram data analysis

Full-scan GC/MS , GC/GC, or GC/XX techniques are standard analytical chemical analysis tools used in a wide range of research fields. It can identify and quantify thousands of chemical compounds in one analysis. It can also be used for identifying biomarker for various biochemical processes, biochemical profiling, and identifying other chemical compounds of interest. However, such techniques are subject to the influence of many experimental factors and have low reproducibility. Currently, most GC/XX type outputs, in the form of chromatograms, are analyzed manually. Comparing multiple chromatograms from either repeated analyses or different samples poses many challenges, and it becomes almost impossible if the number of samples or the number of replicates is large. We have developed a complete data processing and data analysis process for comparing a large number of chromatogra

### Dynamic Calibration for Instruments

❖ Charles Hagwood, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *charles.hagwood@nist.gov*

***Key Words:*** calibration, dynamic linear model, pressure instruments

Often during their usable lifetime instruments are calibrated and recalibrated several times. The goal is to take advantage of these sequentially collected calibration data sets. Instruments drift between calibrations and adapting the dynamic linear model to this calibration problem allows one to incorporate instrument drift into the calibration equation. It is shown that dynamic calibrators perform better than static calibrators.

## 345 Contributed Posters

Section on Physical and Engineering Sciences, Section on Statistics in Epidemiology, Section on Statistics and the Environment, Biopharmaceutical Section, Business and Economics Statistics Section, Section on Statistical Computing, IMS, Section on Survey Research Methods

**Tuesday, August 8, 2:00 pm–3:50 pm**

### A Monte Carlo Method To Simulate the Stochastic Behavior of a Field of Forces

❖ Elizabeth Martinez Gomez, National Autonomous University of Mexico, Heriberto Frias 238-1, Col.Narvarte,C.P.03020, Del.Benito Juarez, Mexico D.F., Mexico, *affabecameister@gmail.com*; Hector A. Perez de Tejada Jaime, National Autonomous University of Mexico ; Hector J. Durand Manterola, National Autonomous University of Mexico

***Key Words:*** stochastic process, Monte Carlo method, acceleration, energization, field of forces

The acceleration of charged particles to suprathermal energies is of potential interest in astrophysics. The simplest mechanism is the acceleration in an electrostatic field parallel to the magnetic field. Consequently, one is led to the concept of stochastic acceleration in which the electric field is random in some sense and, although particles both gain and lose energy in elementary interactions, they gain energy on average. This random process (statistical treatment) usually is based on either a Fokker-Planck or a quasilinear approach. We propose a model in which the equation of motion includes, among others, the force due to a stochastic field simulated by a Monte Carlo method. We apply this model to Saturn's magnetosphere and find that charged particles injected with low energies can be accelerated strongly to reach much higher energies and transported outward.

### Examples of Estimating Confidence Intervals for the Arithmetic Means of Lognormally Distributed Data from Mixed or Random Effects Linear Models

❖ Stanley Shulman, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, MS R3, Cincinnati, OH 45226, *sas2@cdc.gov*

***Key Words:*** generalized confidence intervals, variance components

In occupational health studies, it frequently happens that the measurements are lognormally distributed , but because government standards are based on arithmetic means, appropriate means and confidence intervals must be estimated. Past work on the confidence interval problem has mostly been devoted to simple random samples from lognormal populations. The methods used include the bootstrap and generalized confidence intervals. When confidence intervals are required for the arithmetic mean of a factor in a mixed or random effects linear model, the methods from simple random sampling may be used by grouping the data to obtain random samples. However, possible complications arise when confidence intervals must be provided for averages of arithmetic means, for which results are given. This work compares results obtained by the above methods.

## Statistical Design of Computer Experiments for a 3D Chemical Microanalysis Imaging System

❖ Juan Soto, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *juan.soto@nist.gov*; James J. Filliben, National Institute of Standards and Technology; John H. Scott, National Institute of Standards and Technology

***Key Words:*** microscopy, tomography, experimental design, exploratory data analysis, sensitivity analysis, spectral analysis

An increasing number of nanoscale applications (e.g., biosystems and materials fabrication) depend heavily on computational models to generate insight into underlying mechanisms. NIST researchers are building a novel system for complete 3D chemical microanalysis of nanoscale samples. Understanding how elemental detection limits vary with analysis conditions is an important step in building a robust system. Before reliably producing 3D chemical images, it is necessary to accurately enumerate existing chemical elements above such limits, which in turn guides the selection of spectral windows and other details needed for 3D reconstruction algorithms. Using the NISTMonte simulator, we build synthetic 3D models to create synthetic spectral data. We carry out a series of sensitivity experiments (and analyses) to determine the dominant factors and optimal instrumental settings.

## A Study of the Type I and Type II Error Rates of Tests for Species Diversity Based on Shannon and Simpson's Indexes of Diversity

❖ Lewis VanBrackle, Kennesaw State University, 1980 Towne Manor Drive, Kennesaw, GA 30144, *lvanbrac@kennesaw.edu*; Kristina Corts, Kennesaw State University

***Key Words:*** hypothesis tests, type I error, power, species diversity

Ecologists routinely test for equality of species diversity in two communities using test statistics with approximate t and Normal distributions based on Shannon's and Simpson's diversity indices. However, the type I and type II error rates of these tests are not well known. We performed simulation studies of these tests under broken-stick, geometric and extreme population models to examine their achieved alpha levels and power. For populations with few species, we found the tests to be conservative and to have rather low power for detecting levels of differences in species diversity that may be of interest to ecologists.

## Projecting the Risk of Future Climate Shifts

❖ Luis Cid, Universidad de ConcepciÛn, Departamento de Estadistica, UNIVERSIDAD DE CONCEPCION, CONCEPCION, 01 Chile, *lucid@udec.cl*; David B. Enfield, National Oceanic & Atmospheric Administration

***Key Words:*** multidecadal climate variability, paleoclimate reconstruction, climate shift risk

Recent research has shown decadal-to-multidecadal climate variability associated with environmental changes have important consequences for human activities. How do we convert these relationships into decision support products useful to managers, interested in knowing when future climate regime shifts will likely occur? Numerical models are not able to make deterministic predictions for future climate shifts. Recent development of paleoclimate reconstructions based on tree ring proxy for the Atlantic multidecadal oscillation (AMO) allow us to estimate probability distribution functions from long climate index series and to calculate the probability of future regime shifts. We show how probabilistic decision support tools can be developed for a specific climate. The methods are robust and can, in principle, be applied to any climate mode for which a sufficiently long index series exists.

## Equivalence of Fractional Factorial Designs

❖ Tena Katsaounis, The Ohio State University, 1185 Yorkwood Road, Mansfield, OH 44907, *katsaounis.1@osu.edu*; Angela Dean, The Ohio State University

***Key Words:*** factorial designs, isomorphism, linear codes, non linear codes, orthogonal arrays, Hamming distance distribution

Many methods have been proposed in literature for classifying and ranking experimental designs. Complete classification and ranking is desirable, as equivalent designs share the same statistical properties, but nonequivalent designs might have different statistical properties under different models. We discuss equivalence of regular and nonregular fractional factorial symmetric designs. We examine combinatorial equivalence, which is appropriate for qualitative factors. Clark and Dean (2001) gave a necessary and sufficient condition for combinatorial equivalence. Since then, a number of authors have suggested alternative methods, which are supposedly faster than the Clark and Dean method, but, as we show, none yields a sufficient condition for equivalence. We examine the efficiency of candidate methods as fast screens for nonequivalent designs.

## The Oklahoma Oral Health Needs Assessment: Comparison of Sampling Methods Using Monte Carlo Simulations

❖ Emily Leary, The University of Oklahoma Health Sciences Center, P.O. Box 5293, Norman, 73070, *Emily-Leary@ouhsc.edu*

***Key Words:*** simulation, sample design, multistage sampling

The objective of this study is to compare sampling, weighting, and variance estimation methods for the Oklahoma Oral Health Needs Assessment. The proportion of third grade children who have received protective dental sealants on at least one permanent molar tooth is the binary variable used to perform the simulations. The sealant population was built to simulate the proportion of third graders with dental sealants for each simulated sampled school. Weighted and unweighted estimates from a regionally stratified sample, a probability proportional to size sample, and a two-phase stratified method using counties are reported. Appropriate weighting methods were used for each sampling design as well as two variance estimation techniques: linearization and jackknifing. The distributions of all weighted and unweighted estimates are compared to population values.

## Design of Experiments for Parameter Estimation in Compartmental Models

❖ Michael DeVasher, The University of Alabama, P.O. Box 864676, Tuscaloosa, AL 35486, *mdevasher@ctl.ua.edu*; J. Michael Hardin, The University of Alabama

***Key Words:*** compartmental models, nonlinear design, pharmacokinetics

The design of experiments to estimate the parameters of a nonlinear model requires prior knowledge of these same unknown parameters. As such, an optimum experimental design hinges on the ability of the researcher to incorporate previously gained expertise into the new research. Little work has been done in the design of experiments aimed at estimating the parameters of a compartmental model, especially incorporating prior information on the parameters. Through the use of smoothing splines and locally weighted regression, we construct designs using more informative prior distributions that provide for a decrease in the variability of the estimated parameters. An analysis of the improved efficiency of such designs is discussed. Also, the resistance of the proposed designs to various inaccuracies in the prior distributions is explored.

## Validation of the K-Statistic for Accelerometer Data Quality

❖ James Slaven, National Institute for Occupational Safety and Health, 1095 Willowdale Road, MS 4050, Morgantown, WV 26501, *cto8@cdc.gov*; Michael Andrew, National Institute for Occupational Safety and Health; John Violanti, SUNY University at Buffalo; Cecil Burchfiel, National Institute for Occupational Safety and Health; Bryan Vila, Washington State University

***Key Words:*** actigraphy, quality, sleep, statistics

Accelerometer data quality can be inadequate due to data corruption or noncompliance of study protocols. We discuss a statistical test, the K-statistic, to determine the quality of accelerometer data. This method was evaluated using 105 subjects who wore accelerometers to assess sleep quality. This method uses the overall average distance between consecutive time points and the overall average mean amplitude of consecutive time points. Using leave-one-out cross-validation and calibration-testing methods of discrimination statistics, the error rates for this method were 0.0333 and 0.0167, respectively. An additional group of 37 subjects, whose data were collected at a later time, were used to validate the method. Discrimination analysis was performed using the original dataset as a calibration set and the new dataset as the test set, giving an error rate of 0.0263.

## The Relationship between the T-Square Statistics of a Phase I and Phase II Operation

❖ John Young, McNeese State University, 1750 Bilbo Street, Lake Charles, LA 70601, *johncyoung@bellsouth.net*; Robert L. Mason, Southwest Research Institute; Youn-Min Chou, The University of Texas at San Antonio

***Key Words:*** multivariate statistical process control, generalized variance, scatter matrices

A T2 statistic serves two major functions when implementing a MVSPC. When used in the construction stage, or a Phase I operation, the statistic helps locate special events in a preliminary data set, such as potential outliers or atypical points. When used in the monitoring stage, or a Phase II operation, the statistic helps detect signals, which are observations that do not conform to the baseline conditions established in the Phase I operation. For the unknown parameter case, the T2 statistics used in Phase I and Phase II have different distributions. This difference has been attributed to the specialized functions performed by the separate T2 statistics. However, in this paper, we show mathematically that the T2 statistic used in Phase I is actually a function of the T2 statistic used in Phase II, and either statistic can be used as a substitute for the other.

## Number of Replications Required in Control Chart Monte Carlo Simulation Studies

❖ Jay Schaffer, University of Northern Colorado, 501 20th Street, Greeley, CO 80639, *jay.schaffer@unco.edu*; Myoung-Jin Kim, University of Northern Colorado

***Key Words:*** control charts, Monte Carlo simulations, minimum replications

Monte Carlo simulations have been used extensively in studying the performance of control charts. Researchers have used various numbers of replications in their studies, but almost none of them provided justifications for the number of replications used. Currently, there are no empirically based recommendations regarding the required number of replications to ensure accurate results. This research examined six recently published studies to develop recommendations for the minimum number of replications necessary to reproduce the reported results within a specified degree of accuracy. The results of this study indicated that using 10,000 replications was unnecessarily large and a smaller number of replications could be used to reproduce the target ARLs within the 2% error bands satisfying the modified Mundfrom's criteria. In many cases, only 5,000 replications or fewer were required.

## Is Average Run Length a Good Measure of Performance?

Jay Schaffer, University of Northern Colorado; ❖ Chad Eshelman, University of Northern Colorado, Applied Statistics and Research Methods, McKee 530, Greeley, CO 80639, *chad.eshelman@gmail.com*

***Key Words:*** average run length, median run length, control chart performance, X Bar, cusum, EWMA

Average run length (ARL) has long been the benchmark characteristic used to evaluate a control chart's performance. However, Monte Carlo simulations indicate a control chart's run length distribution is often highly positively skewed. Under these conditions, the ARL tends to be "distorted" and may not provide an accurate measure of what the typical run length might be. This research examines the run length distributions of the X-Bar, CUSUM, and EWMA control charts under various process conditions. Median run length and ARL are compared in detail. The results suggest median run length is a more accurate measure of run length performance.

## A MATLAB Software Implementation for Time-Series Analysis by State-Space Methods

❖ Jyh-Ying Peng, Academia Sinica, Institute of Statistical Science, 128 Academia Road Sec 2, Taipei, 115 Taiwan, *jypeng@stat.sinica.edu.tw*; John Aston, Academia Sinica

***Key Words:*** statistical software, time series, state space models

A MATLAB software toolbox for doing time series analysis by state space methods is presented. The software features interactive construction and combination of models and model components, supporting univariate and multivariate models, dynamic models, non-Gaussian models, and standard models such as ARIMA and structural time-series models. The software includes functions for Kalman filtering and smoothing, simulation smoothing, likelihood evaluation, parameter estimation, signal extraction, and forecasting, with exact initialization for filters and smoothers, and support for missing observations. The software also includes TRAMO model selection and Hillmer-Tiao decomposition. The software provides a general toolbox for doing time series analysis using the MATLAB platform, allowing users to take advantage of its readily available graph plotting and general matrix computation capabilities.

## An Interpolated Periodogram-Based Metric for Comparison of Time Series with Unequal Lengths

❖ Jorge Caiado, CEMAPRE/ISEG and IPS, Rua do Quelhas 6, Lisboa, 1200-781 Portugal, *jcaiado@esce.ips.pt*; Nuno Crato, CEMAPRE/Technical University of Lisbon; Daniel Peña, Universidad Carlos III de Madrid

***Key Words:*** classification, ARMA and ARFIMA models, time series, periodogram, cluster analysis

The comparison of time series has been studied in literature using both time and frequency domain methods. However, existing spectral methods for discrimination and clustering analysis of time series cannot be applied directly to series with unequal lengths. Some studies use time series of unequal length and had to truncate time series spectra to compare them. We then try and develop a method without this drawback. We propose a periodogram-based method for classifying times series with different lengths. For such cases, we know that the Euclidean distance between the periodogram ordinates cannot be used. One possible way to deal with this problem is to interpolate linearly one of the periodograms in order to estimate ordinates of the same frequencies. The interpolated periodogram-based is used to compare different stationary and near-stationary processes.

### Recent Developments in Seasonal Adjustment Software at the U.S. Census Bureau

❖ Brian C. Monsell, U.S. Census Bureau, 4700 Silver Hill Road, Mail Stop 9100, Washington, DC 20233-9100, *brian.c.monsell@census.gov*

***Key Words:*** model based signal extraction, seasonal adjustment diagnostics, accessibility, automatic model identification

Early in 2006, the United States Census Bureau released two seasonal adjustment programs: Version 0.3 of the X-12-ARIMA program, and a beta release of the X-13A-S program, a seasonal adjustment package that optionally produces model based seasonal adjustments from SEATS or X-11 seasonal adjustments. This program, developed in collaboration with the current developers of SEATS, allows users to generate X-11 and SEATS seasonal adjustments using the same interface, and compare these seasonal adjustments using a common set of diagnostics. This session will demonstrate new features of these programs (focusing on X-13A-S) for providing accessible output and also XML output. It will also give results of an evaluation of the revised model identification procedure and diagnostics for model-based seasonal adjustments. Further directions for this work will also be discussed.

### Elevators as an Excitation Source for Structural Health Monitoring in Buildings

❖ Jong-ho Baek, University of California, Los Angeles, 3165 S. Sepulveda Blvd., 307, Los Angeles, CA 90034, *andrew@stat.ucla.edu*

We have investigated the feasibility of using elevators as a statistically repeatable source for use in structural health monitoring of a building.

### Predicting Integrals of Transformed Gaussian Random Fields

❖ Zhengyuan Zhu, The University of North Carolina at Chapel Hill, Department of Statistics, CB3260, University of North Carolina, Chapel Hill, NC 27599-3260, *zhuz@email.unc.edu*; Alf Harbitz, Institute of Marine Research, Observation Methodology

***Key Words:*** kriging, conditional simulation, abundance estimation, geostatistics

In many applications such as marine resource survey and mining, one sample a finite number of sites to observe the variable of interest within an area, with the goal of predicting the integral of such variable over the region of interest, and estimating the precision of such predictor. We develop inference procedures for predicting such integrals when the variable of interest is correlated in space and can be modeled as a realization of a transformed Gaussian random field with unknown transformation. Our work is motivated by the problem of estimating the abundance of a marine species within a given area, which will be used as an illustrating example.

### Analysis of Mixed Frequency Data: a Bayesian Model Averaging Approach

❖ Gavino Puggioni, Duke University, 222 Old Chemistry Building, Box 90251, Durham, 27708-0251, *gavino@stat.duke.edu*; Abel Rodriguez, Duke University

***Key Words:*** Dow-Jones, time series, dynamic linear models, model selection, model averaging, stochastic volatility

We introduce methods to model economic and financial time series using regressors sampled at finer frequencies. The framework is the general class of dynamic linear models, which provides a great level of flexibility and facilitates direct interpretation of the results. The problem of collinearity of infra-period observations is addressed using Bayesian model selection and model averaging. We include applications to the growth rate of US GNP and to the Dow Jones Index returns.

### Data Augmentation within a Conditionally Specified Gaussian Spatial Model

❖ Brooke Fridley, Mayo Clinic College of Medicine, 200 First Street, SW, Division of Biostatistics, Rochester, MN 55905, *fridley.brooke@mayo.edu*; Philip Dixon, Iowa State University

***Key Words:*** censored, data augmentation, spatial, Bayesian

Censored data occurs in numerous areas of application. When one adds the complexity of spatial dependency between observations, methods for handling censored observations become numerically challenging. Often in environmental studies, censoring occurs when contamination values fall below a level of detection. A common practice for handling censored observations is to set the censored observations equal to a function of the level of detection. Instead of using this single imputation approach for censored observations, data augmentation for the censored observations can be implemented. The use of data augmentation within a Bayesian conditionally specified Gaussian spatial model will be illustrated. In doing so, results from an analysis of a dioxin contamination site and results from a simulate study will be presented.

### Spatio-Temporal Precipitation Modeling Based on a Hidden Random Field

❖ Oleg Makhnin, New Mexico Tech, 801 Leroy Place, Socorro, NM 87801, *olegm@nmt.edu*

***Key Words:*** precipitation estimation, spatial statistics, moisture flux direction, random field, MCMC

We present a precipitation modeling approach based on “moisture flux direction” (MFD) random field. The precipitation data are obtained from point (gauge) measurements in a mountainous terrain in Southwest US. We analyze the spatial distribution of moisture flux direction, as inferred from the data, and evolution of this and other model parameters (e.g. regression coefficients) in time. We also relate the inferred MFD to the existing weather indices (ENSO, PDO).

### Real-Time Multivariate Analysis Onboard a Wi-Fi--Controlled Vehicle

❖ Jason Minter, Sam Houston State University, 500 Lakeshore Drive, Coldspring, TX 77331, *jminter@shsu.edu*; Cecil Hallum, Sam Houston State University

***Key Words:*** multivariate, real-time analysis, remote sensing, Wi-Fi control

The process of clustering and discriminate analysis on digitized imagery can be enhanced via vector augmentation to increase accuracy of the pixel discrimination into distinct classes. Our paper discusses an improved method of vector augmentation that permits greater accuracy when one seeks to identify specialized entities, like rivers, roads, etc. To obtain the digital imagery, an inexpensive Wi-Fi controlled vehicle was mounted with a digital camera and web-camera. In place of a handheld remote controller, a payload comprised of a miniaturized computer (the OQO, a 14 ounce computer) was added and used to do real-time processing. Using remote control of the OQO from a laptop, via a Wi-Fi/Bluetooth connection, control of operations was accomplished aboard the vehicle, and real-time multivariate analyses were conducted on digital images acquired by a high-resolution camera.

## 346 Deming Lecture

Deming Lectureship Committee, The ASA, ENAR, WNAR, IMS, SSC, Section on Statistical Consulting
**Tuesday, August 8, 4:00 pm–5:50 pm**

### Making Another World: a Holistic Approach to Performance Improvement

❖ Ronald D. Snee, Tunnell Consulting, 10 Creek Crossing, Newark, DE 19711, *Snee@TunnellConsulting.Com*

***Key Words:*** W. Edwards Deming, Six Sigma, statistical thinking, holistic improvement, lean manufacturing

Business improvement methodology, in which Deming played a leadership role, has been evolving for 100 years. Today, there is a number of effective improvement methods available, including Six Sigma, Lean Manufacturing, Baldrige Assessments, and ISO 9000. These options often result in nonproductive discussion about which improvement approach to use. But the prudent person uses all approaches, selecting the right tools for the problem at hand. To implement such a philosophy, one needs a holistic approach that integrates the useful tools from all approaches into a common methodology. Setting today's needs in the context of Deming's work, this talk will present an overview of a holistic approach, discuss methodology and several examples, and address the critical role of the statistician in improvement initiatives.

## 347 ASA Presidential Address and Awards

The ASA
**Tuesday, August 8, 8:00 pm–9:30 pm**

### Presentation of Awards

❖ Fritz J. Scheuren, National Opinion Research Center, 1402 Ruffner Road, Alexandria, VA 22302, *scheuren@aol.com*

Certificates of Appreciation for Retiring Editors - Fritz J. Scheuren Samuel S. Wilks Memorial Award - Mark P. Becker, Chair Samuel S. Wilks Memorial Medal Committee Goffried E. Noether Awards - Regina Y. Liu, Chair Noether Awards Committee Statistics in Chemistry Award - Daniel C. Park, Chair SPES Committee on Chemometrics Award of Outstanding Statistical Application - Karen Kafadar, Chair Award of Outstanding Statistical Application Committee W. J. Youden Award in Interlaboratory Testing - Neil R. Ullman, Chair W. J. Youden Award in Interlaboratory Testing Committee Edward C. Bryant Scholarship Award - Elizabeth A. Stasny, Chair Edward C. Bryant Scholarship Award Committee Gertrude M. Cox Scholarship in Statistics Award - Holly B. Shulman, Chair Subcommittee to Committee on Women in Statistics and the Caucus for Women in Statistics SPAIG Award - George W. Williams, Chair Statistical Partnerships among Academe, Industry, and Government Award Committee

### From Data to Policy: Scientific Excellence Is Our Future

❖ Sallie Keller-McNulty, Rice University, Brown School of Engineering, Houston, TX 77005-1892, *sallie@rice.edu*

Science, engineering, technology, and people must come together to support the growing complexity of today's global challenges, ranging from international security to space exploration. As scientists and engineers, it is essential that we develop the means to put our work into a decision context for policymakers. Scientific and technical progress requires interdisciplinary teams, as it is impossible for an individual to have the knowledge to solve many of today's problems. By the nature of our training, statisticians are well-poised to assume the role of science and technology integrator. To be successful, this will place statisticians closer to policy pressures and politics. This talk will focus on the growing expectations facing statistical sciences and how we, as statisticians, must take responsibility for separating the scientific method from the politics of the scientific process.

### Presentation of Founders Awards and New ASA Fellows

❖ Fritz J. Scheuren, National Opinion Research Center, 1402 Ruffner Road, Alexandria, VA 22302, *scheuren@aol.com*

Founders Award, Fritz J. Scheuren, Chair Founders Award Committee New ASA Fellows, M. Elizabeth Halloran, Chair Committee on Fellows

## 348 Section on Bayesian Statistical Science Roundtable with Coffee (fee event)

Section on Bayesian Statistical Science
**Wednesday, August 9, 7:00 am–8:15 am**

### Model Selection in Hierarchical Models

❖ David B. Dunson, National Institute of Environmental Health Sciences, NIEHS, Biostatistics Branch, MD A3-03, P.O. Box 12233, Research Triangle Park, NC 27709, *dunson@stat.duke.edu*

***Key Words:*** model selection, random effects, variance components, homogeneity test, Bayes factor, BIC

Traditional methods of model selection, such as likelihood ratio tests and AIC/BIC selection criteria, may not be appropriate for comparing hierarchical models with different covariance structures or numbers of latent variables. Some of the difficulties arise due to constraints on the parameter space. We will discuss emerging methods for dealing with model uncertainty in hierarchical models, such as random effects models and structural equation models. Some possibilities include new approximations to the Bayes factor and stochastic search methods. Important issues include prior selection and methods for efficient computation.

# 349 Section on Statistical Education Roundtable with Coffee (fee event)

Section on Statistical Education
**Wednesday, August 9, 7:00 am–8:15 am**

### An Open Discussion about Quantitative and Qualititative Research in Statistics Education

❖ Jackie Miller, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43210, *miller.203@osu.edu*

***Key Words:*** statistics education, research, quantitative, qualitative

As statisticians, we know the importance of well-designed and well--carried out quantitative research. But there are other types of research that can be and are done in statistics education. Much of the research that has been published has involved small action research studies. Few large-scale quantitative studies have been published. This talk will be a forum for us to discuss questions, including (but not limited to): What is the place of quantitative, qualitative, and mixed methodology research in statistics education? What are the advantages/disadvantages of these types of research? What types of research questions can be answered with which type(s) of research? Can we say one type is better than another? The questions are limitless. This coffee table will be an exciting chance to talk about research in statistics education!

# 350 Section on Statistics and the Environment Roundtables with Coffee (fee event)

Section on Statistics and the Environment
**Wednesday, August 9, 7:00 am–8:15 am**

### Keeping Our Jobs: Relevance of Statistical Research in a Production Environment

❖ Gretchen Moisen, U.S. Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT 84401, *gmoisen@fs.fed.us*

There is no shortage of statistical challenges in ecological and environmental applications. However, budgets for natural resource work are in decline, and individual scientists---as well as entire research projects---are being cut if not deemed 'relevant' by their parent organizations. Statisticians and quantitative scientists are being forced to ask a number of questions. Are we solving the most critical problems? Are we effectively marketing our ideas? What are the most effective ways to communicate our results? Are our proposed solutions being transitioned into a production stream? How should we facilitate organizational change? What are our metrics for success? This roundtable will provide an opportunity to compare and discuss strategies for ensuring the relevance of statistical research during lean budget years.

### Current Issues in Space-Time Modeling of Environmental Data

❖ Montserrat Fuentes, North Carolina State University, 8203 Campus Box, Statistics Department, Raleigh, NC 27695, *fuentes@stat.ncsu.edu*

***Key Words:*** Bayesian spatial inference, spatial likelihood, environmental spatial analysis, software for spatial data, spatial temporal modeling, covariance

This roundtable focuses on the statistical challenges to model, predict, and estimate the complex spatial-temporal structure of environmental data. One of the main challenges is to calculate a Gaussian likelihood for large space-time problems, and modern approaches to this problem use spectral methods and algorithms to approximate the determinant and the inverse of a large matrix. However, the impact these approximations have on inference is not yet clear. Another issue is how to model a complex space-time dependency structure, and current work involves extensions of separable models and coregionalization models (in the multivariate case). These are popular models, but are they realistic? What is the best software to handle large spatial-temporal problems? What are the advantages/drawbacks of WinBugs, SAS, R, and all the R packages available to handle space-time data?

# 351 Section on Statistics in Epidemiology Roundtable with Coffee (fee event)

Section on Statistics in Epidemiology, Section on Statistical Consulting
**Wednesday, August 9, 7:00 am–8:15 am**

### Developing and Cultivating Successful Collaborations

❖ Robert F. Woolson, Medical University of South Carolina, 135 Cannon Street, PO Box 2580835, Charleston, SC 29425, *woolson@musc.edu*

***Key Words:*** collaboration, statistical consulting

Collaboration is, for many biostatisticians, the essence of their careers. Successful collaborations yield many professional and personal rewards, but the cultivation of a sound collaborative environment requires many ingredients. There are numerous factors to help develop a positive culture for research collaborations, including investigators' appreciation of the role of biostatistics in research, biostatistician's enthusiasm for collaboration, a reward system for collaboration, and the ability of both biostatistician and investigator to raise the level of interaction from simple consultation to a full exchange of ideas. We will solicit personal experiences from participants on the role of institutional culture, the role of mentorship within the biostatistical group---including ways to promote such mentorship, and the importance of maintaining both methodological and collaborative activities.

# 352 Section on Physical and Engineering Sciences Roundtable with Coffee (fee event)

Section on Physical and Engineering Sciences
**Wednesday, August 9, 7:00 am–8:15 am**

### Catching up on Wavelets: Recent Advances, Future Directions

❖ Donald B. Percival, University of Washington, Applied Physics Laboratory, Box 355640, Seattle, WA 98195-5640, *dbp@apl.washington.edu*

***Key Words:*** time series analysis, image analysis, spatial analysis, function estimation, signal estimation, wavelet transform

The pioneering work of Donoho, Johnstone, and coworkers in the 1990s on wavelet shrinkage for nonparametric function estimation demonstrated to the statistical community the usefulness of wavelets for tackling difficult statistical problems. Whereas wavelets are functions that have a burst of activity and then damp down to zero, the subject of wavelets has not followed this pattern, as evidenced by a seemingly increasing tidal wave of papers reporting interesting applications of wavelets and extensions of the basic theory. his coffee roundtable will provide a forum for reviewing recent work and what the future might hold in the ongoing marriage of statistics and wavelets.

## 353 Statistical Society of Canada Roundtable with Coffee (fee event)

SSC, Section on Statistical Consulting

**Wednesday, August 9, 7:00 am–8:15 am**

### Tips for Academic Medical Statisticians

❖ Karen Kopciuk, Alberta Cancer Board, Division of PHI, 1331 29 Street NW, Calgary, AB T2N 4N2 Canada, *karen.kopciuk@cancerboard.ab.ca*; Rhonda Rosychuk, University of Alberta

***Key Words:*** academic medical statisticians, career choice, practical tips, medical research, statistical consulting

The life of an academic medical statistician involves consulting and collaborating on medical research projects. In some settings, the statistician also may be expected to have an independent research program. With its nearly exclusive research focus, this career choice offers many unique opportunities and challenges not found in a traditional academic position. The moderators of this roundtable have a combined experience of 10+ years as academic medical statisticians, and both are recent recruits to this setting. They will discuss practical issues, such as creating and running statistical consulting units, making time for your own research, dealing with isolation, finding good collaborators, obtaining funding for research, and teaching statistics to medical researchers. The roundtable is targeted for statisticians early in their career, although people in any career stage are welcome.

## 354 Late-Breaking Session #2: What is the Role of Statistics in Public Policy Debates about Climate Change?

The ASA, ENAR, IMS, SSC, WNAR

**Wednesday, August 9, 8:30 am–10:20 am**

### The Kyoto Accord, the 2001 IPCC Third Assessment Report, and the Academic Papers Underpinning Them

❖ Edward Wegman, George Mason University, Department of Statistics, MSN 4A7, 4400 University Drive, Fairfax, VA 22030-4422, *ewegman@gmu.edu*

***Key Words:*** PCA, paleoclimate reconstruction, climate change, global warming

The Kyoto Accord focused on reducing greenhouse gasses and was supported by the report of the Intergovernmental Panel on Climate Change, "2001 Third Assessment Report." This report featured the "hockey stick" millennium temperature reconstruction based on a number of proxy variables. The academic papers that developed the reconstruction used a principal components analysis, which has been challenged by critics due to inappropriate use of PCA and a lack of independent verification of the findings. We discuss the methodology used, the use of potentially problematic data, and the social network of authors in temperature reconstruction. The implications of global warming are important for the financial and human dimensions, and public policy decisions must be made on a statistically sound, uncontroversial basis. D. Scott, D. Brillinger, Y. Said, J. Rigsby, D. Reeves, and A. Braverman contributed to this report.

### National Research Council Report on the 'Hockey Stick Controversy'

❖ J. Michael Wallace, University of Washington, Department of Atmospheric Sciences, 106 King Building, Box 354235, Seattle, WA 98195-4235, *wallace@atmos.washington.edu*

***Key Words:*** climate change, global warming, paleoclimate, paleoclimate reconstruction

A series of 1000 to 2000-year-long paleoclimate reconstructions published within the past decade suggest that global temperature has risen dramatically over the past century compared the variability earlier in the record. A graphic based on one of these reconstructions, popularly referred to as the "hockey stick curve", was featured in the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). The analysis on which this graphic is based has been criticized on statistical grounds. Findings and conclusions of a recent U.S. National Research Council report on surface temperature reconstructions that bear upon this controversy will be presented and discussed.

### The CCSP Report on Temperature Trends in the Lower Atmosphere

❖ Richard L. Smith, The University of North Carolina at Chapel Hill, Department of Statistics and Operations Research, 201 Smith Building, CB#3260, Chapel Hill, NC 27599-3260, *rls@email.unc.edu*

***Key Words:*** global warming, time series regression, uncertainty estimation

The Climate Change Science Program (CCSP) was set up by the Bush administration in 2002 with a charge of producing 21 reports on different aspects of climate change. The first report, "Temperature Trends in the Lower Atmosphere", claimed that the lower atmosphere was indeed growing warmer, in contrast to some earlier published studies, and that there was "clear evidence of human influences on the climate system." Despite this apparently clear-cut conclusion, the report raised many further issues. This talk will review the whole report with particular reference to statistics and the report's approach to assessing uncertainty.

# 355 Introductory Overview Lectures: Image Statistics and Bootstrap

The ASA, ENAR, IMS, SSC, WNAR
**Wednesday, August 9, 8:30 am–10:20 am**

## Introduction to Bootstrapping

❖ Tim C. Hesterberg, Insightful Corporation, 1700 Westlake Ave., Suite 500, Seattle, WA 98109-3044, *timh@insightful.com*

***Key Words:*** resampling, permutation test, teaching statistics, statistical education, inference

This talk is intended for people who have completed at least one course in statistics, from introductory statistics on up. I'll give an overview of the basic ideas and methods of bootstrapping and permutation tests, with an emphasis on how statisticians can use these methods for communicating with others---teachers to help students understand concepts such as sampling distributions, standard errors, confidence intervals, and P-values and consultants to help clients understand statistical results. I'll show how to use the bootstrap for statistics such as an x intercept in regression or a difference in trimmed means where classical methods are hard. And I'll show examples where the bootstrap is more accurate than classical methods because it doesn't require assuming normality or equal variances; in one example, classical t-tests give P-values that are off by a factor of four!

## Image Statistics

❖ Eugene Demidenko, Dartmouth Medical School, 7927 Rubin, DHMC, Lebanon, NH 03756, *eugened@dartmouth.edu*

***Key Words:*** image processing, t-test, Kolmogorov-Smirnov, cancer imaging, image registration, histogram equalization

It is well known that statistics is an interdisciplinary science. It would be an overstatement to say that statistics does not play a role in imaging. However, it would be correct to say that statistics is not presented at the level it deserves. Most authors of image science are from engineering and computer science departments and there are only a few occasions when images are treated from a statistical perspective. There exists a great demand in image analysis from a statistical point of view. Especially important is the task of statistical image comparison. I consider three example of statistical image analysis: statistical comparison of four groups of cancer treatment using histology imaging, reduction of a color image to a grey scale image, and hypothesis testing for post-reconstructed and post-registered medical images.

# 356 Estimation and Inference for Models with Many Parameters

Business and Economics Statistics Section
**Wednesday, August 9, 8:30 am–10:20 am**

## Objective Bayesian Analysis for Multivariate Dynamic Models

❖ Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65210, *sund@missouri.edu*; Shawn Ni, University of Missouri-Columbia

***Key Words:*** objective Bayesian, state-space models

It is well-known that dynamic or state-space models have been successful in practice. Bayesian multivariate dynamic models have received a lot of attention lately. Due to the large number of unknown parameters included in the models, the Bayesian estimates are sensitive to the choice of the priors. In this paper, we review the objective Bayesian analysis for a univariate dynamic model and explore the various choices of objective priors for a multivariate dynamic model.

## Bayesian Inference for Gaussian Copula Models

❖ Robert J. Kohn, University of New South Wales, UNSW, Sydney, 2052 Australia, *r.kohn@unsw.edu.au*

***Key Words:*** covariance selection, multivariate analysis, non-Gaussian data, Markov chain Monte Carlo, graphical models

A Gaussian copula regression model gives a tractable way of handling a multivariate regression when some of the marginal distributions are non-Gaussian. Our paper presents a general Bayesian approach for estimating a Gaussian copula model that can handle any combination of discrete and continuous marginals and generalizes Gaussian graphical models to the Gaussian copula framework. Posterior inference is carried out using a novel and efficient simulation method. The methods in the paper are applied to simulated and real data and demonstrated to work well.

## Spatial Bayesian Variable Selection with Application to Functional Magnetic Resonance Imaging

❖ Michael Smith, University of Sydney, Econometrics and Business Statistics, Merewether Building (H04), Sydney, NSW 2006 Australia, *michael.smith@econ.usyd.edu.au*; Daniel Smith, University of Sydney

***Key Words:*** binary MRF, model averaging, MCMC

In this talk a procedure to undertake Bayesian variable selection and model averaging for a series of regressions that are located on a lattice is proposed. For those regressors which are in common in the regressions, we consider using an Ising prior to smooth spatially the indicator variables representing whether or not the variable is zero or non-zero in each regression. This smoothes spatially the probabilities that each independent variable is non-zero in each regression, and indirectly smoothes spatially the regression coefficients. The approach is applied to the problem of functional magnetic resonance imaging in medical statistics, where massive datasets arise that need prompt processing. The relative strengths and weaknesses of using the Ising prior over alternative binary Markov random fields will also be discussed.

## Statistical Inference for Highly Parameterized Models for Discrete-Valued Data

❖ Sylvia Frühwirth-Schnatter, Johannes Kepler University, Altenbergerstrasse 69, Linz/Austria, A-4040 Austria, *sylvia.fruehwirth-schnatter@jku.at*

***Key Words:*** discrete data, mixture auxiliary sampling, data augmentation

In this talk, we will be concerned with statistical inference for discrete-valued data when modeling is based on complex generalized linear models, such as state-space models for count data or multinomial random-effect models. First, we will discuss MCMC estimation for these types of models, which is based on an approximate, but accurate, new mixture auxiliary sampler that introduces two sequences of artificial latent variables. This mixture auxiliary sampler leads to a conditionally linear Gaussian model. Next, we will show that auxiliary mixture sampling also is useful for model choice and variable selection.

## 357 Statistical Models in Computational Biology ● ✪

WNAR, Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

### Spatial Smoothing To Map HIV Recombination Hotspots: Associations with RNA Secondary Structure

❖ Marc A. Suchard, University of California, Los Angeles, David Geffen School of Medicine, 650 Charles E. Young Drive, Department of Biomathematics, Los Angeles, CA 90095-1766, *msuchard@ucla.edu*; Vladimir N. Minin, University of California, Los Angeles; Karin S. Dorman, Iowa State University

***Key Words:*** phylogenetics, spatial models, genomics

We introduce a Bayesian hierarchical framework for simultaneous analysis of multiple HIV recombinants and elucidation of breakpoint spatial preferences. A dual multiple changepoint model constructs the likelihood for individual sequence alignments. Recombination breakpoints are drawn from a common hyper-prior characterized by location-specific recombination log-odds. Understanding of HIV recombination mechanisms suggests these log-odds should not differ drastically at adjacent locations. We employ this information by recruiting an intrinsic Gaussian Markov random field prior. Application of our model to the HIV gag coding region provides strong evidence for a recombination hotspot on the boundary between matrix and capsid genes. RNA stem-loop elements, located in this region, support the hypothesized involvement of local secondary structure in promoting recombination.

### Computational and Statistical Algorithms for Parentage Inference with Single Nucleotide Polymorphisms

❖ Eric C. Anderson, Southwest Fisheries Science Center, 110 Shaffer Road, Santa Cruz, CA 95060, *eric.anderson@noaa.gov*

***Key Words:*** parentage, paternity, importance sampling, Monte Carlo, Baum algorithm, genetics

Single nucleotide polymorphisms (SNPs) are molecular markers abundant in the genomes of most organisms. Because SNPs can be genotyped in a high-throughput setting with minimal genotyping error, they may be quite useful for parentage inference in large populations of intensively monitored or managed species. I review recent developments in importance sampling methods, allowing the calculation of false positive rates in large-scale, likelihood-based, parentage inference. Then, I present algorithms that combine exclusion and likelihood-based methods for performing parentage inference in very large problems involving billions of potential parent pairs and many thousands of offspring.

### Detecting Positive Selection in Protein-Coding DNA Sequences in Absence of Substantial Phylogenetic Information

❖ Raquel Prado, University of California, Santa Cruz, Baskin School of Engineering, Applied Math & Statistics, 1156 High St MS: SOE2, Santa Cruz, CA 95064, *raquel@ams.ucsc.edu*; Daniel Merl, University of California, Santa Cruz

***Key Words:*** positive selection, GLMS, hierarchical models, structured priors, model comparison

Detecting site-specific rates of substitution is a challenging problem when dealing with polymorphic data. These data are characterized by low divergency, and, therefore, little phylogenetic signal is available. We use Bayesian-generalized linear models to describe the substitution patterns in polymorphic alignments. Our models are hierarchical and allow us to include biologically relevant prior information. Once the models are fitted to the data, it is possible to summarize the posterior distributions of key quantities that describe the patterns of evolution at the molecular level, such as the nonsynonymous to synonymous rates ratio for each amino acid site and the transition to transversion rates ratio. In addition, we present tools for comparing models that support neutral evolution in the data versus models that assume the data show evidence of positive and negative selection.

### Inferring Maximum-Likelihood Species Phylogenies under Coalescence

❖ Laura S. Kubatko, University of New Mexico, 1 University of New Mexico, MSC03 2150, Department of Mathematics and Statistics, Albuquerque, NM 87131-0001, *salter@stat.unm.edu*

***Key Words:*** phylogenetics, coalescence, gene tree, species tree

The coalescent process is known to generate discord in gene and species phylogenies. It can be shown that inference of the species phylogeny using traditional methods that do not model the coalescent process explicitly can be statistically inconsistent as the number of genes increases. In this talk, we describe the likelihood of the species phylogeny under coalescence and give the details of an algorithm for finding the maximum-likelihood species tree. The method is applied to real data for which modeling of the coalescent process is particularly important.

### A Bayesian Approach to Gene Tree Concordance

❖ Bret Larget, University of Wisconsin-Madison, Department of Statistics, Madison, WI 53706, *larget@stat.wisc.edu*

***Key Words:*** phylogeny, MCMC, gene tree, Bayesian statistics, evolution, phylogenetics

Bayesian phylogenetics involves the estimation of evolutionary relationships from genetic data. It is not unusual for different genes to support different evolutionary histories. The two most common strategies to deal with this are combining all of the data into a single analysis that assumes a single common tree or making separate independent estimates for each gene. We describe an approach for the estimation of several gene trees between these two extremes that accommodates the possibility of multiple different gene trees, but also incorporates information from all genes through a prior distribution on the set of gene trees for each individual gene tree estimate. We use a novel two-stage Markov chain Monte Carlo approach for calculations for this problem.

## 358 Complex Sampling Designs and Related Inference Issues in Epidemiologic Studies ●

Section on Statistics in Epidemiology

**Wednesday, August 9, 8:30 am–10:20 am**

## Complex Case-Control Sampling Methods: Solutions to Some Diverse Problems in Epidemiological Research

❖ Bryan Langholz, Keck School of Medicine of USC, Department of Preventive Medicine, 1540 Alcazar St, CHP-220, Los Angeles, CA 90089-9011, *langholz@usc.edu*

***Key Words:*** case-control studies, conditional logistic regression, risk estimation, Breslow estimator, study design, risk sets

A nested case-control data model provides a natural connection between failure time data and case-control data. The semiparametric analysis methods lead to 'standard' conditional logistic regression, but also apply to a large class of individually matched case-control sampling methods. However, this framework does not tell us what designs may be useful for the given study goals and data resources. In this talk, I will illustrate how the approach was used to solve three design and analysis problems: a two-stage design, absolute risk estimation when the number at risk is not known, and the appropriate analysis of studies with random digit dial controls. My focus will be on developing intuition about the estimators, general characteristics of successful designs, and how the methods easily extend to accommodate special situations.

## Confounding of Genetic Association Studies by Population Structure

❖ Alice S. Whittemore, Stanford University, Department of Health Research and Policy, HRP Redwood Building, 259 Campus Drive, Stanford, CA 94305-5405, *alicesw@stanford.edu*

***Key Words:*** case-control studies, confounding, genomic control, inbreeding, population isolates, population structure

Genetic association studies can be confounded by undetected population structure. Such confounding can lead to spurious associations and failure to detect causal associations. Genomic Control (GC) and family-based methods have been proposed to deal with population structure. They are based on two models for population structure: Isolates and Inbreeding. I will show that confounding by population structure can occur under the Isolates Model but not the Inbreeding Model. Thus, GC methods adjust for variance distortion, but not confounding. Conversely, family-based association studies control for confounding, but not variance distortion.

## Case-Control Studies of Gene-Environment Interaction: a Bayesian Approach

❖ Bhramar Mukherjee, University of Florida, P.O. BOX 118545, Department of Statistics, Gainesville, FL 32611-8545, *mukherjee@stat.ufl.edu*; Li Zhang, University of Florida; Malay Ghosh, University of Florida

***Key Words:*** ascertainment bias, conditional likelihood, Dirichlet process, meta-analysis

In case-control studies of gene-environment association with disease, when genetic and environmental exposures can be assumed to be independent in the underlying population, one may exploit the independence in order to derive more efficient estimation techniques than the traditional logistic regression analysis (Chatterjee and Carroll, Biometrika, 2005). However, covariates that stratify the population, such as age, ethnicity and alike, could potentially lead to sources of non-independence. We provide a novel semiparametric Bayesian approach to model stratification effects under the assumption of gene-environment independence in the control population. The results reflect that the semiparametric Bayesian model allows incorporation of key scientific evidence in the form of a prior and offers a flexible, robust alternative when standard parametric model assumptions do not hold.

# 359 Rare Trait Inference ●

Section on Survey Research Methods

**Wednesday, August 9, 8:30 am–10:20 am**

## Network Sampling: a Potential Tool for Survey Estimates about Rare Populations

❖ Iris M. Shimizu, National Center for Health Statistics, 3311 Toledo Road, Room 3123, Hyattsville, MD 20782, *ims1@cdc.gov*; Monroe G. Sirken, National Center for Health Statistics

***Key Words:*** sampling design, multiplicity estimators

Network sampling uses multiplicity rules for linking individual observation units to multiple selection units. The number of selection units linked to an observation unit is the multiplicity of that unit. In a traditional survey, each observation unit is countable at one, and only one, selection unit, to which it is linked by a unitary counting rule---a rule for which the 'multiplicity' is one. Because multiplicity rules spread individual observation units over more selection units than a unitary rule, the network sampling technique can be more efficient than a conventional sample of rare populations and/or elusive populations, which are difficult to survey at their usual residences. This paper discusses the network sampling methodology and some of its advantages and disadvantages. Selected comparisons between network and traditional sample surveys will be included.

## Design and Likelihood-Based Inference for Sample Surveys on Rare Traits

❖ Steve Thompson, Simon Fraser University, Department of Statistics and Actuarial Science, 8888 University Drive, Burnaby, BC V5A 1S6 Canada, *thompson@stat.sfu.ca*

***Key Words:*** link-tracing designs, adaptive sampling, Markov chains, design-based, model-based, sampling

Sampling populations with rare, hidden, or hard-to-reach traits often requires innovative methods such as link-tracing techniques, in which social links are followed to bring more members of the hidden population or more persons with the rare traits into the sample. Although such methods can be effective, they require inference methods that, in some respects, compensate for the design to give an unbiased or accurate picture of the population. Efficient design-based approaches involve Rao-Blackwell or Markov chain resampling methods. The most effective model-based approaches involve likelihood or Bayes methods and generally involve Markov chain Monte Carlo methods for feasible computation. A third approach is to use a targeted random walk design to obtain simple estimates that are, in a sense, robust against the lack of control of some aspects.

## Testing Logistic Regression Coefficients with Clustered Data and Few Positive Outcomes

❖ Sally Hunsberger, National Cancer Institute, Executive Blvd., Room 8120, Bethesda, MD 20892, *sallyh@ctep.nci.nih.gov*; Barry I. Graubard, National Cancer Institute; Edward Korn, National Cancer Institute

***Key Words:*** bootstrap, generalized Wald statistics, goodness-of-fit, clustering, generalized score statistics, survey methods

A simulation-based method is developed for testing logistic regression coefficients with clustered data when there are few positive outcomes. This testing methodology is shown to compare favorably to generalized Wald and score tests and a bootstrap hypothesis test in terms of maintaining nominal levels. An application is given that analyzes the association of asthma with various demographic variables and risk factors using data from the third National Health and Nutrition Examination Survey. The proposed methods also are useful when testing goodness-of-fit of logistic regression models using deciles-of-risk tables.

### Analytic Issues for Rare Events in the NHANES Survey

❖ Lester R. Curtin, Centers for Disease Control and Prevention,
3311 Toledo Road, Hyattsville, MD 20782, *lrc2@cdc.gov*

***Key Words:*** exact binomial tests, rare events, degrees of freedom, repeated cross-sectional surveys

While all complex surveys are designed to achieve a specified level of analytic precision, the NHANES survey has three important aspects that complicate analyses: the number of degrees of freedom is small, some outcome measures are based on subsamples of the original selection, and there are a number of important measured health characteristics that push the analytic needs for the survey beyond its original design specifications. While some analytic issues can be handled by combining multiple years of data, many of the statistical methods adopted for complex survey data are based on asymptotic results and are not adequate for rare outcomes. This presentation summarizes recent CDC research on exact confidence intervals for small proportions and percentiles. Also, test statistics---adjusted for small degrees of freedom---for comparing rare events are explored.

## 360 New Directions in Statistical Machine Learning ● ✪

Section on Statistical Computing, Section on Statisticians in Defense and National Security, Section on Nonparametric Statistics

**Wednesday, August 9, 8:30 am–10:20 am**

### Margin-Based, Semisupervised Learning

Junhui Wang, University of Minnesota; ❖ Xiaotong Shen,
University of Minnesota, School of Statistics, 224 Church Street S.E.,
Minneapolis, MN 55455, *xshen@stat.umn.edu*

***Key Words:*** generalization, grouping, missing labels

In classification, semisupervised learning occurs when a large amount of unlabeled data is available with only a small number of labeled data. In such a situation, the objective is to use unlabeled data to enhance predictability of classification. This talk presents a novel margin-based, semisupervised learning methodology, utilizing the grouping information from unlabeled data together with the concept of margins in a form of regularization controlling the interplay between labeled and unlabeled data. In addition, the generalization error is estimated using both labeled and unlabeled data for tuning. The methodology is implemented for support vector machines (SVM) as well as psi-learning through difference convex programming, which reduces to sequential quadratic programs. Our theoretical and numerical analyses suggest the proposed methodology achieved the desired objective.

### Classification and Feature Selection for High-Dimensional Data

❖ Hao Zhang, North Carolina State University, Campus Box 8203,
Department of Statistics, Raleigh, NC 27613, *hzhang2@stat.ncsu.edu*

***Key Words:*** high dimensional low sample size, support vector machines, nonconvex penalty, microarray gene expression

Classification for high-dimensional low sample size data becomes more pertinent in bioinformatics. One fundamental problem in cancer study is to identify which groups of genes are differentially expressed in normal and cancerous cells and which leads to a better understanding of genetic signatures in cancer and the improvement on cancer treatment strategies. The high-dimension low sample size structure of microarray data demands more flexible and powerful statistical tools for analysis. We will introduce a novel type of regularization in support vector machines (SVM) to identify important genes for cancer classification. A special nonconvex penalty is imposed on the hinge loss function in the SVM. By systematically thresholding small estimates to zeros, the new procedure eliminates redundant genes automatically and yields a compact and accurate classifier.

### Bias and Variance of Bagging Based on Subsampling with and without Replacement

❖ Andreas Buja, University of Pennsylvania, The Wharton
School, 471 Jon M Huntsman Hall, Philadelphia, PA 19104-6340,
*buja@wharton.upenn.edu*

Bagging is a device intended to reduce the prediction error of learning algorithms by reducing their variance. In its original form, bagging draws bootstrap samples from the training sample, applies the learning algorithm to each bootstrap sample, and averages the resulting prediction rules. Variants of bagging are obtained by letting the resample size, M, be different from the sample size, N, where both M<N and M>N are possible. Still, other variants are obtained when “resampling with” is replaced with “resampling without,” in which case one needs M<N. We investigate bagging in a simplified situation: the prediction rule produced by a learning algorithm is replaced by a simple, real-valued u-statistic of i.i.d. data. U-statistics of high order can describe complex dependencies, and yet they admit a rigorous asymptotic analysis.

## 361 Causal Inference and the Estimation of Neighborhood Health Effects ●

Section on Health Policy Statistics, Biometrics Section

**Wednesday, August 9, 8:30 am–10:20 am**

### The (Mis)estimation of Neighborhood Effects: Identification Problems and the Multilevel Model

❖ J. Michael Oakes, University of Minnesota, 1300 S. 2nd Street,
Suite 300, Minneapolis, MN 55454, *oakes@epi.umn.edu*

***Key Words:*** causation, identification, context, multilevel

Epidemiology has long had an interest in the effect of neighborhood environments on health outcomes. In recent years, a fervor has erupted as investigators have aimed to disentangle the effects of such environments from the background of residents. Hundreds of papers have been published, but little of this work has sustained methodologic examination. Drawing from a recent paper (Oakes. 2004. Soc Sci Med

58:1929-52), this talk discusses fundamental problems in the identification of meaningful neighborhood effect parameters. I show that except in special cases where sampling mechanisms yield exchangeable subjects, none of the published estimates merits attention. The fervor is a case of "statisticism," a term used to describe a ritualistic appeal to significance testing and sampling error when such issues are a relatively minor concern.

### Response to "The (Mis)estimation of Neighborhood Effects"

❖ Jay S. Kaufman, The University of North Carolina at Chapel Hill, Department of Epidemiology, 2104C McGavran-Greenberg Hall, Chapel Hill, NC 27599-7435, *Jay_Kaufman@unc.edu*

***Key Words:*** causation, identification, spatial, confounding, epidemiology, effect decomposition

Oakes raises many formidable issues about causal inference that remain challenging for public health researchers, although many are not unique to estimation of neighborhood effects. No overall solution to identification problems is realistic in this or any setting, but some incremental improvements can be made in the conduct and interpretation of observational multilevel research. Because a common focus is on the decomposition of effects into compositional and contextual components, clarification of this strategy is a worthwhile goal. Due to problems discussed by Oakes, such as omitted variables and emergent effects, it may be better to conceptualize the estimated contextual effect as a kind of residual effect, not the effects of measured predictors at the individual level.

### Forming Better Guesses about Neighborhood Effects on Health

❖ Brian Krauth, Simon Fraser University, 8888 University Drive, Economics Department, Burnaby, BC V5A 1S6 Canada, *bkrauth@sfu.ca*

***Key Words:*** neighborhood effects, causation, epidemiology

A recent stream of econometric research (Altonji, Elder, and Taber 2005, Krauth 2006) uses the statistical relationship among the observed explanatory variables in a regression to generate a plausible guess about the relationship between observed and unobserved variables. The usual assumption of orthogonality between observed and unobserved variables is a special case - i.e., a bad guess. This paper extends this approach to the standard linear omitted variables problem, and demonstrates that this approach may be particularly useful for the measurement of neighborhood effects on health.

### Causal Diagrams To Express Identification of Place Effects Using Multilevel Models

❖ M. Maria Glymour, Harvard School of Public Health, Landmark Center W., Room 403J, 401 Park Drive, Boston, MA 02215, *mglymour@hsph.harvard.edu*; S. V. Subramanian, Harvard School of Public Health

***Key Words:*** causal diagrams, multilevel models, neighborhood effects, causal inference, back-door criterion, ecologic exposures

We apply a causal inference framework based on directed acyclic graphs (DAGs) to identification of effects of exposures measured at an ecologic level. We give example DAGs representing various types of place effects and plausible back-door paths. We show that multilevel models are useful for causal inference because they can be used to block back-door paths mediated via individual and ecologic covariates. Indeed, multilevel models are even more pertinent for mechanistic studies where heterogeneous treatment effects are of as much interest as the population average treatment effect. However, their value is limited when important back-door paths remain unmeasured. Under such conditions, the front-door criteria or instrumental variable approaches (e.g., community interventions or natural experiments) are discussed as useful alternatives for identification-of-place effects.

## 362 Detecting Anomalies in Dynamic Multivariate Data ✪

Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences, Section on Statistical Graphics

**Wednesday, August 9, 8:30 am–10:20 am**

### Aggregation Queries at Streaming Speeds

❖ Divesh Srivastava, AT&T Labs-Research, NJ 07901, *divesh@research.att.com*

***Key Words:*** data streams, aggregation, quantiles, heavy hitters, IP network traffic

Measuring and monitoring complex dynamic phenomena, such as IP network traffic evolution, produces highly detailed and voluminous data streams. The monitoring applications that analyze these massive data streams require sophisticated, user-defined aggregation queries (which summarize the data distributions in these streams) to identify, for example, anomalous activity. Motivated by the nature of IP network traffic data, we present two families of aggregation queries: hierarchical heavy hitters (hierarchically organized large-valued regions) and biased quantiles (totally ordered skewed ranks). We present deterministic algorithms to approximate these aggregates in a single pass over the data stream. Our experimental results highlight the importance of fast, space-efficient, nonblocking implementations to make them useful in practice.

### Dynamic Thresholds: Monitoring Streams of Counts Online

❖ Diane Lambert, Google, Inc., 1440 Broadway, New York, NY 10018, *dlambert@google.com*; Chuanhai Liu, Purdue University

***Key Words:*** anomaly detection, networks, reliability, online testing

Streams of counts on traffic volume and processing errors are compared continually to thresholds to detect network degradation. Thresholds often are set by hand, which is tedious and error-prone. This talk will describe how data analysis led to a simple, statistically principled approach to automated thresholding that self-initializes, tracks long-term trends and cyclical timing patterns, and monitors its own performance, yet has just a few parameters that can be kept up-to-date as the data flow by.

### Monitoring Massive Streams Simultaneously: a Holistic Approach

❖ Deepak K. Agarwal, AT&T Labs-Research, 180 Park Ave., Florham Park, NJ 07932, *dagarwal@yahoo-inc.com*

***Key Words:*** Kalman filter, prospective, multiple testing

A holistic approach to prospective anomaly detection for massive number of streams is proposed. The method works by building a baseline model to capture normal behavior. Any baseline model that provides a p-value for the observed, relative to the predicted can be used. Anomalies are detected by tracking normal scores derived from p-values. A flexible and fast five-parameter Bayesian model adjusts for multiple testing at each time point. Methods to delete uninformative streams from the monitoring process are also discussed. The method is illustrated on a real application where our baseline model is built using a state space approach.

### Two-Dimensional Variable Window Scan Statistics

❖ Joseph Glaz, University of Connecticut, Department of Statistics, U-4120, 215 Glenbrook Rd., Storrs, CT 06269-4120, *joseph.glaz@uconn.edu*

***Key Words:*** Bayes factor, biosurveillance, minimum p-value statistic, repeated p-values, simultaneous testing, spatial statistics

In this talk, we will discuss recent developments for variable window scan statistics for data modeled by discrete and continuous distributions in a two-dimensional rectangular region. These scan statistics can be viewed as spatial statistics for detecting local clustering of events. We will discuss Bayesian testing procedures, minimum p-value statistics, repeated p-value statistics, and scan-score statistics. Numerical results will be presented to evaluate the performance of these spatial scan statistics. Two-dimensional variable window scan statistics have applications in several areas of science and technology, including biosurveillance, ecology, epidemiology, and reliability theory. Future directions for research and challenging open problems in this area will be discussed as well.

## 363 Design and Analysis of Experiments for Complex Computer Simulators ●

Section on Physical and Engineering Sciences

**Wednesday, August 9, 8:30 am–10:20 am**

### Designs for Integrated Computer and Physical Experiments

❖ C. Shane Reese, Brigham Young University, 230 TMCB, Department of Statistics, Provo, UT 84602, *reese@stat.byu.edu*; Derek Bingham, Simon Fraser University; Wilson Lu, Simon Fraser University

***Key Words:*** Gaussian processes, industrial statistics

Physical experimentation is an expensive endeavor. Computer experiments, which are often a less-expensive alternative, provide a reasonably accurate representation of physical experimental results. This work proposes an approach to designing experiments when the researcher has a limited amount of data collected on each of the physical and computer experiments and wishes to allocate future experiments to be tested at inputs for the physical experiments, the computer experiments, or experiments to be run at combinations of both physical and computer models. This hybrid experimental situation relies on Gaussian process formulations for the modeling of the experimental outputs as functions of the experimental inputs. We demonstrate the methodology with both simulated and actual experimental situations.

### Sequential Experiment Design for Contour Estimation from Complex Computer Codes

❖ Pritam Ranjan, Simon Fraser University, Department of Statistics and Actuarial Science, 8888 University Drive,, Burnaby, BC V5A1S6 Canada, *pritamr@cs.sfu.ca*

***Key Words:*** computer experiment, Gaussian process, inverse problem

In many engineering applications, one is interested in identifying the values of the inputs in computer experiments that lead to a response above a prespecified threshold, or in a region of interest. In this talk, we will introduce statistical methodology that identifies the desired contours (i.e., iso-surfaces) in the input space. First, a stochastic model is used to approximate the global response surface. Then, for expensive black box functions, we will propose a sequential approach for the selection of trials directed toward improvement of the estimation of the contours. Finally, extraction of the contours from the estimated surface will be outlined. Application also includes multiple contour estimation and highest posterior density region.

### Uncertainty Quantification for Combining Experimental Data and Computer Simulations from Multiple Data Sources

❖ Brian J. Williams, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545-0001, *brianw@lanl.gov*; Dave Higdon, Los Alamos National Laboratory; Jim Gattiker, Los Alamos National Laboratory

***Key Words:*** calibration, computer experiments, Gaussian process, functional data analysis, uncertainty quantification, predictive science

This work focuses on combining observations from field experiments with detailed computer simulations of a physical process to carry out inference. This typically involves calibration of parameters in the simulator and accounting for inadequate physics. We consider physical applications for which the field data and simulator output are multivariate. Multivariate data lead to computational challenges for implementing the framework. We consider adaptive basis methods to achieve significant dimension reduction. This methodology is extended to incorporate multiple sources of field data and simulator output into a joint calibration and prediction analysis. Different sources of data inform on specific calibration parameter subsets, which are not required to be disjoint. We illustrate the proposed methodology with experimental data and simulations that inform on gas equations of state.

## 364 Status of Disability Information in Surveys ●

Committee on Statistics and Disability

**Wednesday, August 9, 8:30 am–10:20 am**

### Expert Panel on Disability Survey Data

❖ Susan Schechter, Office of Management and Budget, NEOB, Room 10201, 725 17th Street, N.W., Washington, DC 20503, *Susan_Schechter@omb.eop.gov*; ❖ Jennifer Madans, National Center for Health Statistics, 3311 Toledo Road, Beltsville, MD 20782, *jennifer.madans@cdc.hhs.gov*; ❖ Martin Gould, National Council for Disability, 1331 F Street, NW, Suite 1050, Washington, DC 20004-1107, *mgould@ncd.gov*; ❖ Philip Rones, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212, *Rones_p@bls.gov*; ❖ Mary Grace Kovar, National Opinion Research Center, 1350 Connecticut Ave., NW, Suite 500, Washington, DC 20036, *Kovar-MG@norc.org*

***Key Words:*** disability, survey, health, federal programs

The Expert Panel on Disability Survey Data will focus on the need for more reliable and consistent disability measurement. Expenditures for the many federal disability programs comprise nearly 10% of federal outlays. Furthermore, disability rates are increasing as the population ages. Yet, prevalence estimates range from 5 million to 50 million, depending on the definition used. As a first step, ASA involvement will help federal policymakers and program administrators better address disability issues and improve the lives of people with disabilities. The panel consists of Philip Rones (Deputy Commissioner, Bureau of Labor Statistics), Mary Grace Kovar (National Opinion Research Center), Susan Schecter (Office of Management and Budget), Jennifer Madans (National Center for Health Statistics), and Martin Gould (Director of Research and Technology, National Council for Disability).

## 365 National Science Foundation Invited Session

National Science Foundation, Section on Statistical Education

**Wednesday, August 9, 8:30 am–10:20 am**

### Funding for Statistics and Probability Research at the National Science Foundation and the National Institutes of Health

❖ Cheryl Eavey, National Science Foundation, SBE/SES/NSF, 4201 Wilson Blvd, Arlington, VA 22230, *ceavey@nsf.gov*; ❖ Dean Evasius, National Science Foundation, Division of Mathematical Sciences, 4201 Wilson Blvd, Arlington, VA 22230, *devasius@nsf.gov*; ❖ Wen C. Masters, National Science Foundation, Division of Mathematical Sciences, 4201 Wilson Blvd, Arlington, VA , *wmasters@nsf.gov*; ❖ Ronald S. Fecso, National Science Foundation, 4201 Wilson Blvd, Suite 965, Arlington, VA 22230, *rfecso@nsf.gov*

***Key Words:*** funding opportunities, grant applications, proposal review, co-funding

A panel of NSF and NIH program directors will present a range of funding opportunities for researchers in statistics and probability. You will learn about NSF funding opportunities in Mathematical Science Priority Areas (updated), co-funding with programs in other areas, international collaboration, conferences, workshops, and special years. You also will learn about the NSF joint program with NIH/NIGMS and funding opportunities at NIH. Additionally, the panel will discuss review and funding procedures and point out common pitfalls in proposal writing.

## 366 Sparse Inference and Multiple Comparisons ✪

IMS

**Wednesday, August 9, 8:30 am–10:20 am**

### False Discovery Rates for Spatial Signals

❖ Ruth Heller, Tel-Aviv University, , *rheller@post.tau.ac.il*; Yoav Benjamini, Tel-Aviv University

***Key Words:*** signal detection, FDR, multiple testing, hierarchical testing, weighted testing procedures, functional MRI

We suggest a new approach to multiple testing for signal presence in spatial data that tests cluster units, rather than individual locations. This approach leads to increased signal-to-noise ratio within the unit tested and to a reduced number of hypotheses tests compared. We introduce a powerful adaptive procedure to control the size-weighted FDR on clusters (i.e., the size of erroneously rejected clusters out of the total size of clusters rejected). Once the cluster discoveries have been made, we suggest 'cleaning' locations in which the signal is absent by a hierarchical testing procedure that controls the expected proportion of locations in which false rejections occur without loosing control of the cluster error rate. We discuss an application to functional MRI that motivated this research and demonstrate the advantages of the proposed methodology on an example.

### Large Dimensional Covariance Matrix Estimation Using a Factor Model

❖ Jinchi Lv, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544, *jlv@princeton.edu*

***Key Words:*** factor model, diverging dimensionality, covariance matrix estimation, consistency, asymptotic normality, portfolio allocation and risk management

Large dimensionality is a common feature in modern portfolio management. We examine the covariance matrix estimation in the asymptotic framework that the dimensionality p_n diverges with the sample size n. Motivated by the CAPM, we propose to use a K-factor model to reduce the dimensionality and estimate the covariance matrix among those assets. Under basic assumptions, we establish the n^b-consistency with b=min(1/2,1-a/2) when p_n=O(n^a), 0< =a< 2, and asymptotic normality for the covariance matrix estimator. The performance is compared with the sample covariance matrix. We identify the cases when the factor approach can gain substantially the performance and when the gains are only marginal. Furthermore, the impacts of the covariance matrix estimation on portfolio allocation and risk management are studied, respectively. The results are supported by a thorough simulation study.

### A Bayesian Approach for Incomplete Paired Data

❖ Feng Liang, Duke University, Box 90251, Durham, NC 27708, *feng@stat.duke.edu*; Woncheol Jang, Duke University; Fei Liu, Duke University

***Key Words:*** Bayesian hypothesis testing, incomplete paired data

DNA microarray technology is applied often in a matched pair study. Failure of experiment causes the observations to be unmatched. We provide a Bayesian method to test the differences in paired gene expressions when there are unmatched data on both sides. Our approach naturally incorporates the information from the unmatched observations for a better estimation.

### Quantile Coupling for Median and Its Application to Nonparametric Robust Estimation

❖ Harrison Zhou, Yale University, 24 Hillhouse Ave., Department of Statistics, New Haven, CT 06511, *huibin.zhou@yale.edu*

***Key Words:*** quantile coupling, robust estimation, function estimation, asymptotic equivalence, large deviation, KMT/Hungarian construction

We give a sharp quantile coupling bound for a median statistic with a normal random variable, which improves the classical quantile coupling bounds with a rate under certain smoothness conditions for the distribution function. Its application to nonparametric robust wavelet estimation is discussed.

### Sparse Principal Component Analysis

❖ Hui Zou, University of Minnesota, 313 Ford Hall, 224 Church Street S.E., Minneapolis, MN 55455, *hzou@stat.stanford.edu*

***Key Words:*** PCA, SPCA, regularization, L_1 penalty, Procrustes rotation, LARS

Principal component analysis is used widely in data processing and dimensionality reduction. However, PCA suffers from each principal component being a linear combination of the original variables. Thus, it is often difficult to interpret the results. We introduce the SPCA criterion, which results in a principled method---called SPCA---for producing modified principal components with sparse loadings. We also propose an efficient algorithm based on reduced-rank Procrustes rotation and the LARS for computing the SPCA. The proposed methodology is applied to real and simulated data with encouraging results.

## 367 Adaptive Dose Response

Biopharmaceutical Section, Biometrics Section, ENAR
**Wednesday, August 9, 8:30 am–10:20 am**

### Implementing Bayesian Adaptive Dose-Response Finding Studies: a Clinical Perspective

❖ Michael Krams, 25 Dairy Hill Road, Madison, CT 06443, *michael_krams@comcast.net*

***Key Words:*** dose-response, Bayesian, adaptive

Some drug development programs fail due to insufficient attention to selecting dosages that optimize response. We discuss efficient approaches to designing clinical dose-response finding studies. A first step is to understand the disease by analyzing historical databases. We need to explore drug exposure-response relationships in a way that allows us to adjust our approach to dosing as trials progress, in light of initial results. We should continually reassess the choice of treatments we offer in trials so we optimize the acquisition of new information. Our trial designs have sequential termination rules that stop the trial as soon as a clear conclusion is reached (and not earlier). We may conclude that the treatment has too small an effect, or so large an effect that minimal dosage levels suffice---or that only a few specific dosage levels seem effective.

### Adaptive Dose-Response Phase II Trials for Clinical Development

❖ Qing Liu, Johnson & Johnson, 25 Wellington Drive, Long Valley, NJ 07853, *qliu2@prdus.jnj.com*

***Key Words:*** adaptive designs, dose-response trials, multiple-trend tests

We propose a two-stage adaptive dose-response design for phase II clinical trials. A primary objective is to demonstrate that, within the dose-range, there is a positive trend toward better efficacy for higher doses. A secondary objective is to select the lowest dose for phase III trials. To increase the efficiency and protect patient's safety, the design allows addition or deletion of doses in the second stage upon knowledge of the efficacy and safety dose-response relationships at the interim analysis. Once the doses are selected for the second stage, the sample size for the second stage also is calculated. The design is suitable for applications where the efficacy endpoint is not immediately assessable.

### Evaluating Rolling Dose Designs and Methods

❖ Amit Roy, Bristol-Myers Squibb Company, Route 206 and Provinceline Road, Princeton, NJ 08540, *amit.roy@bms.com*; Frank Shen, Bristol-Myers Squibb Company

***Key Words:*** adaptive dose finding, dose-response, phase II, target dose

The FDA has highlighted suboptimal dose selection as a key factor in the decline of successful NDA filings and adverse post-approval findings. An initiative sponsored by PhRMA to address this issue is the Rolling Dose Studies (RDS) Working Group, which investigates the efficiency of an innovative class of adaptive dose-finding designs. In these designs, the number of doses and the allocation of patients to doses are allowed to change during the study by incorporating information in the accruing efficacy and safety data. This presentation will describe results of a comprehensive comparison of a conventional dose-finding design and four alternative methods of designing and implementing RDS. The statistical operating characteristics of these methods were evaluated by simulation, under a variety of trial scenarios and dose-response functions.

## 368 Biomarker

Biopharmaceutical Section, Biometrics Section, ENAR
**Wednesday, August 9, 8:30 am–10:20 am**

### Biomarker Analysis of Medical Imaging and Radiotelemetry Signals

❖ Christopher Tong, Merck Research Laboratories, Biometrics Research RY33-300, P O Box 2000, Rahway, NJ 07065, *christopher_tong@merck.com*; Yevgen Tymofyeyev, Merck Research Laboratories; Karim Azer, Merck Research Laboratories; Philip E. Brandish, Merck Research Laboratories; Hongxing Chen, Merck Research Laboratories; James C. Hershey, Merck Research Laboratories; Matthew Walker, III, Merck Research Laboratories; Barry R. Campbell, Merck Research Laboratories; Kaijie Fang, Merck Research Laboratories; Donald S. Williams, Merck Research Laboratories; Alexandre Coimbra, Merck Research Laboratories

***Key Words:*** biomarkers, radiotelemetry, medical imaging

In this talk, we focus exclusively on quantitative biomarkers that are measured in animal models of diseases. Biomarkers from medical imaging modalities---such as ultrasound, magnetic resonance imaging (MRI), and functional MRI---and radiotelemetry signals are discussed in the context of cardiovascular, neurological, and bone diseases. These biomarkers are in vivo measurements that allow for longitudinal tracking of structural and/or functional changes with disease progression. In some situations, a biomarker must be selected from a set of plausible candidates; in others, our understanding of the physiological processes and the nature of the measurement technology allows us to formulate a potential biomarker we will evaluate. Statistical challenges are discussed, including hierarchical models and problems of large-scale time-series data.

### Statistical Considerations for Protein Biomarker Discovery from Human Plasma and Cerebrospinal Fluid

❖ Richard Higgs, Eli Lilly and Company, MS 1930, Lilly Corporate Center, Indianapolis, IN 46285, *higgs@lilly.com*

***Key Words:*** proteomics, biomarker, protein, variability, pharmaceutical

Pharmaceutical and biotechnology companies are pursuing biomarkers as a means to increase the productivity of drug development. Quantifying differential levels of proteins from complex biological samples, such as plasma or cerebrospinal fluid, is one approach being used to identify markers of drug action, efficacy, and toxicity. This presentation will enumerate and discuss the statistical issues relevant to the identification and quantification of peptides and proteins using the so-called bottom up LC-MS/MS experiment using proteolytic digests of complex biological samples. Reproducibility and variance components results of a proposed method will be presented along with an assessment of inter- and intra-subject variability of proteins in human plasma and cerebrospinal fluid from a prospectively designed proteomics clinical trial.

### Model Selection and Cross-Validation for Biomarker Discovery and Validation

❖ Annette Molinaro, Yale University School of Medicine, 60 College Street, Room 201, New Haven, CT 06520, *annette.molinaro@yale.edu*

***Key Words:*** biomarker discovery, genomic data, prediction, tissue microarrays, cross-validation, prediction error

Clinicians aim toward a more preventative model of attacking cancer by pinpointing and targeting specific early events in disease development. These early events can be measured as genomic, proteomic, epidemiologic, or clinical variables. The measurements are then used to predict clinical outcomes such as primary occurrence. Numerous methods are available to unearth biologically driven associations between variables and clinical outcomes. Given any of these methods, a model that elucidates relevant patterns of association is selected. The next challenge is to assess how well this model will predict outcomes in an independent validation sample. We will explore a general framework for comparing methods, selecting models, and assessing prediction error in the presence of censored outcomes. In addition to simulations, we will evaluate our approach in tissue microarray data on breast cancer.

### Application of RandomForest as a Variable Selection Tool on Biomarker Data

❖ Katja Remlinger, GlaxoSmithKline, 5412 Silver Moon Lane, Raleigh, NC 27606, *katja.s.remlinger@gsk.com*

***Key Words:*** cross validation, classification, selection bias, prediction model, supervised learning, high dimensional data

In supervised learning problems involving high-dimensional data, it is often desirable to reduce the number of variables. Biomarker data is usually high-dimensional and therefore a good candidate for variable selection approaches. Identifying a small number of important markers is not only essential to achieve high accuracy for the prediction model, but also necessary to allow an easy interpretation of the model from a biological point of view. Using biomarker data from an obesity study, we will illustrate the challenges involved in the model building and marker selection process. In particular, we will use a modified version of the RandomForest Wrapper Algorithm (Svetnik et al. 2004) to build models based on a small number of important markers and make comparisons to other approaches.

## 369 Ranked Set Sampling II ●

Section on Nonparametric Statistics

**Wednesday, August 9, 8:30 am–10:20 am**

### Confidence Intervals for Quantiles Based on Ranked Set Samples

❖ Tao Li, St. Francis Xavier University, P.O. Box 5000, Antigonish, NS B2G 1V5 Canada, *tli@stfx.ca*; Narayanaswamy Balakrishnan, McMaster University

***Key Words:*** ordered statistics, ranked set sample, confidence interval, expected width, percentage reduction, ordered ranked set sample

Confidence intervals for quantiles and tolerance intervals based on ordered ranked set samples (ORSS) will be discussed. We first derive the cdf of ORSS and the joint pdf of any two ORSS. In addition, we obtain the pdf and cdf of the difference of two ORSS. Then, confidence intervals for quantiles based on ORSS are derived and their properties discussed. We compare, with approximate confidence, intervals for quantiles and show that these approximate confidence intervals are not accurate. However, when the number of cycles in the RSS increases, these approximate confidence intervals become accurate, even for small sample sizes. We also compare with intervals based on usual order statistics and find the confidence interval based on ORSS becomes considerably narrower than the one based on usual order statistics when size n becomes large.

### Missing Data and Consequences in Ranked Set Sampling

❖ Jessica Kohlschmidt, The Ohio State University, 5980 OSweeney Lane, Dublin, OH 43016, *kohlschmidt.1@osu.edu*; Elizabeth Stasny, The Ohio State University; Douglas Wolfe, The Ohio State University

***Key Words:*** ranked set sampling, ranking, sampling, missing data, imputation, weighting

Missing data is a natural consequence of sampling. Some researchers ignore the missing data. If the data is missing completely at random, this procedure provides meaningful estimates. In many situations, the data missingness is due to a phenomenon occurring in the population. Ignoring the pattern of missingness can lead to inaccurate conclusions. We often have missing data on the variable of interest, but, in RSS, we have additional information available about each unit we wish to quantify. This extra information can be used to help determine estimation methods that may be superior to current methods. We investigate methods to deal with missing data that use the ranking information we have for each unit and the observed measurements of the variable of interest.

### Ranked Set Sampling for Ordered Categorical Variables

❖ Haiying Chen, Wake Forest University, Medical Center Blvd., Winston Salem, NC 27106, *hchen@wfubmc.edu*; Elizabeth Stasny, The Ohio State University; Douglas Wolfe, The Ohio State University

***Key Words:*** imperfect ranking, optimal allocation, ordinal variable, perfect ranking, relative precision

The application of ranked set sampling (RSS) to binary variables has been studied. The objective of this paper is to extend the application of RSS, both balanced and unbalanced, to ordered categorical variables with the goal of estimating the probabilities of all categories. We use ordinal logistic regression to aid in the ranking of the ordinal variable of interest. We also propose an optimal allocation scheme and methods for implementing it under either perfect or imperfect rankings. Results from a simulation study using a substantial dataset, the third National Health and Nutrition Examination Survey dataset, indicate the use of ordinal logistic regression in ranking leads to substantial gains in precision for estimation of population proportions.

### Order-Restricted, Randomized Designs for Linear Models Using L1 Norm

❖ Shannon Markiewicz, The Ohio State University, 955 Mount Pleasant Ave., Columbus, OH 43201, *markiewicz.8@osu.edu*; Omer Ozturk, The Ohio State University

***Key Words:*** ranked set sampling, order restricted randomized designs

Ranked set sampling is not well-suited for the design of experiments because it requires more experimental units and the role of randomization is not well-defined. To resolve these concerns, order-restricted, randomized (ORR) designs have been developed recently and their properties discussed in the literature. In this talk, we will develop nonparametric statistical inference for linear models in the context of order-restricted, randomized designs based on L-1 norm. We will derive the asymptotic distribution of the parameter estimates and develop a drop and Wald test for the contrast parameter. It will be shown that ORR design performs better than classical designs in linear models. A simulation study will show that test and estimators perform reasonably well, even for moderately large sample sizes.

### Two-Sample, Ranked-Sum Test for Order-Restricted Randomized Designs

❖ Yiping Sun, The Ohio State University, P.O. Box 3646, Columbus, OH 43210, *suny1963@yahoo.com*; Omer Ozturk, The Ohio State University

***Key Words:*** Mann-Whitney-Wilcoxon test, Pitman efficacy, clinical trials, judgment ranking, ranked-set sampling

It has been shown that the order-restricted randomized (ORR) design that relies on subjective judgment ranking to compare subsets of experimental units, improves statistical inference for the contrast between two levels of a treatment. This paper develops statistical inference (for Mann-Whitney-Wilcoxon statistics) for the location shift between control and treatment populations in an ORR design. It is shown that the asymptotic null distribution of the new test statistic is normal and it has desirable efficiency properties under high- and low-quality ranking. A simulation study indicates the proposed test performs well, even for small sample sizes. An example is provided to illustrate the use of the proposed test statistic.

## 370 From Policy to Application: a Health and Mortality Case Study

Section on Government Statistics
**Wednesday, August 9, 8:30 am–10:20 am**

### The National Longitudinal Mortality Study

❖ Norman Johnson, U.S. Census Bureau, DSMD, Room 3716 FB 3 MS 8700, Washington, DC 20233, *norman.j.johnson@census.gov*

***Key Words:*** longitudinal data, current population survey, 1980 census data, National Death Index, mortality differentials

The National Longitudinal Mortality Study (NLMS) is a sponsor-funded database developed by the U.S. Census Bureau to study the effects of demographic and socioeconomic characteristics on differentials in U.S. mortality rates. The NLMS consists of bureau data from Current Population Surveys, Annual Social and Economic Supplements, and the 1980 Census combined with death certification information from the National Death Index. The study currently consists of approximately 3.3 million records with more than 250,000 identified mortality cases. The socioeconomic variables offer researchers the potential to answer questions on mortality differentials for a variety of important socioeconomic and demographic subgroups not extensively covered in other databases. This paper will provide a general overview of the NLMS and discuss current activities to update longitudinal records.

### The National Death Index: an Overview

❖ Robert Bilgrad, National Center for Health Statistics, 3311 Toledo Road, Room 7318, Hyattsville, MD 20782, *rbilgrad@cdc.gov*

***Key Words:*** mortality ascertainment, death record linkage, deaths, mortality data

The National Death Index (NDI) is a file of identifying death record information for all U.S. deaths occurring since 1979. This computer-matching service assists health researchers in determining whether specific study subjects have died, and, if so, provides researchers with the states and dates of death and death certificate numbers. The NDI Plus service also provides the cause of death codes derived from the decedents' death certificates. Since 1982, the NDI has performed more than 3,800 searches involving more than 46 million records submitted by researchers involved in a variety of activities---including clinical trials, post-marketing drug surveillance, occupational health studies, cancer, and other disease registries---and longitudinal studies involving large population groups.

### The NLMS: Data Stewardship Policies at Work

❖ Wendy Alvey, U.S. Census Bureau, OAES Room 2430, FB 3, MS 3700, Washington, DC 20233, *wendy.l.alvey@census.gov*

***Key Words:*** data stewardship, privacy, confidentiality, title 13 data, National Longitudinal Mortality Study, policies

The U.S. Census Bureau's Data Stewardship Program provides the legal and ethical framework for ensuring confidentiality is maintained and respondent privacy is respected. The Data Stewardship Program is centered on four privacy principles. Data stewardship policies have been developed to support those principles and ensure current practice is compatible with them. This session will demonstrate how data stewardship is applied for a specific data production effort at the bureau---the National Longitudinal Mortality Study (NLMS). This paper focuses on two data stewardship policies that relate to the NLMS - Reimbursable Project Acceptance and Controlling Nonemployee Access to Title 13 Data. Both policies are critical components in the bureau's efforts to ensure respondent privacy and confidentiality are maintained.

### U.S. Census Bureau Administrative Record Data Stewardship Policies for Administrative Records Use

❖ Patricia Melvin, U.S. Census Bureau, 2403 Forest Edge Court, Unit J, Odenton, MD 21113, *patricia.l.melvin@census.gov*

***Key Words:*** administrative records, data stewardship, record linkage, National Longitudinal Mortality Study, policies, procedures

As part of its data stewardship program, the U.S. Census Bureau requires adherence to specific policies and procedures for the acquisition, use, and linkage of administrative records. Central to these policies and procedures is a formal project review and approval process that includes a specific policy for record linkage activities. The National Longitudinal Mortality Study utilizes administrative records, and is, therefore, subject to the requirements above. This paper provides an overview of the bureau's project review and approval process with a discussion of its Record Linkage Policy and the specific tools developed for ensuring compliance with the administrative records policies and procedures.

# 371 Visual Sampling Plan Software for Designing Environmental Sampling Plans for Chem/Bio/Rad and Munitions Contamination ●

Section on Statistics and the Environment

**Wednesday, August 9, 8:30 am–10:20 am**

## Visual Sample Plan (VSP) Software: What Is It, and How To Use It?

❖ John Wilson, Pacific Northwest National Laboratory, Grand Junction, 81505, *john.wilson@pnl.gov*; Lisa Nuffer, Pacific Northwest National Laboratory; Brent A. Pulsipher, Battelle-PNNL

***Key Words:*** VSP software, environmental sampling, data quality objectives (DQO) systematic planning

Visual Sample Plan (VSP) is a user-friendly, widely used graphic software tool for helping nonstatisticians determine statistically defensible number and location of environmental samples for various sampling objectives and designs for environmental contaminants. VSP is a data quality objectives--based systematic planning tool developed with funding from multiple government agencies since 1997. The Statistical Sciences group at the Pacific Northwest National Laboratory continues to expand the capability and functionality of VSP, which now has thousands of users worldwide. This paper provides an overview of VSP, including determination and placement of sampling locations on site maps for various sampling objectives and designs; data analysis capabilities; and the automatic generation of a report that documents the developed design, assumptions, and results.

## Sampling Designs for Surfaces within Buildings

❖ Brett D. Matzke, Battelle-PNNL, , *brett.matzke@pnl.gov*; Brent A. Pulsipher, Battelle-PNNL; John Wilson, Pacific Northwest National Laboratory; Richard Gilbert, Pacific Northwest National Laboratory; Lisa Nuffer, Pacific Northwest National Laboratory; Nancy Hassig, Pacific Northwest National Laboratory; John Hathaway, Battelle-PNNL

***Key Words:*** sample design, terrorist, software, visual sample plan, VSP, sampling design

The threat of a biological, chemical, or radiological terrorist attack is an increasing concern. Sampling will play a key part in estimating the magnitude and extent of contamination during different phases of the cleanup process should such an attack occur within a building. A visual sample plan (VSP) module was developed to provide sampling designs and statistical analysis for walls, floors, ceilings, and other surfaces within a building. This presentation demonstrates the options available for developing and visualizing sampling schemes for multiple surfaces within a room. Floor plans can be imported or drawn in detail. Analytical methods will be demonstrated for different design objectives, including the ability to compare measurements to a threshold for detecting hotspots and determining the extent of contamination when the data distribution is normal or unknown.

## Geophysical Survey Designs To Aid the Identification and Remediation of Ordnance-Related Contaminants

❖ John Hathaway, Battelle-PNNL, P.O. Box 999, MSIN K608, Richland, WA 99352, *john.hathaway@pnl.gov*; Brent A. Pulsipher, Battelle-PNNL; John Wilson, Pacific Northwest National Laboratory; Richard Gilbert, Pacific Northwest National Laboratory; Brett D. Matzke, Battelle-PNNL

***Key Words:*** sample design, UXO, compliance sampling, visual sample plan, software, transects

Currently, millions of acres of formerly used Department of Defense sites (FUDS) contain or may contain munitions or explosives of concern (MEC), such as unexploded ordnance (UXO). In an effort to reduce costs of finding UXO target areas and assessing the degree of confidence in UXO clean-up efforts, the Statistical Sciences group at the Pacific Northwest National Laboratory developed two UXO modules in the Visual Sample Plan (VSP) software (available free at http://dqo.pnl.gov/vsp). We explore how a Monte Carlo simulation gives the probability of detecting a target area as a function of transect spacing and the density of anomalous geophysical-detector readings in the target area. We also apply compliance sampling to MEC clean-up using a Bayesian and/or frequentist framework. A demonstration of how these modules are being used on FUDS lands will be given.

## A Practical Application of VSP to an Environmental Question: Abraham's Creek

❖ Kelly Black, Neptune & Company, 2031 Kerr Gulch Road, Evergreen, CO 80439, *kblack@neptuneinc.org*; Michele Wolf, Neptune & Company

***Key Words:*** VSP, software, environment, sample size, case study

Visual Sample Plan (VSP) is user-friendly free software developed cooperatively among several branches of the federal government used to provide statistical solutions for environmental sampling design problems. VSP helps the user select the number and location of samples to collect so a required confidence level for decisionmaking can be achieved. We applied VSP to Abraham's Creek, for which the Quantico Marine Corps Base is a potential contaminant source. Previous studies identified elevated levels of DDx and PCBs in sediment. However, substantial spatial data gaps led to a desire to collect additional data to better characterize the contamination. VSP was used to determine optimal sample size and location. This presentation is designed to illustrate both the ease with which VSP was applied to this project and the care that should be taken when applying this tool.

# 372 Statistics in the Aerospace Industry: Human Factor Studies ● ✪

Biometrics Section, Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences

**Wednesday, August 9, 8:30 am–10:20 am**

### Using Statistical Methods in the Design of the 787 Cabin Environment

❖ Martin Meckesheimer, The Boeing Company, 500 Wall Street, 1615, Seattle, WA 98121, *martin.meckesheimer@boeing.com*

***Key Words:*** 787, cabin interior, human subjects testing

Boeing has unveiled innovative architectural and technological features in the 787 Dreamliner that will completely redefine the passenger experience. These features include wider seats and aisles, larger windows, more spacious luggage bins, innovative lighting, improved ride quality, reduced cabin altitude, and improved cabin humidity. In this presentation, we will describe some of the studies conducted, showing how statistical methods have been used to guide decisions concerning the 787 cabin environment.

### Predicted Arterial Oxygenation at Commercial Aircraft Cabin Altitudes

❖ Mike Muhm, The Boeing Company, P.O. Box 3707, MC 7A XH, Seattle, WA 98072, *mike.muhm@boeing.com*

***Key Words:*** hypoxia, altitude, hypobaric hypoxia, arterial blood gas

The degree of hypoxia manifest by airline passengers during flight is not well-characterized. The relationship between age and PaO2 at sea level (PaO2 sl) and between PaO2 alt and PaO2 sl, PCO2 at sea level (PCO2 sl) and pulmonary health status were investigated using linear regression techniques to analyze previously published data. In persons with normal pulmonary health, the relationship between PaO2 sl (mm Hg) and age (years) was PaO2 sl = 105.9 - 0.44 * age (R2 = 0.582, MSE = 25.314); PCO2 sl (38.1 ± 2.8 mm Hg) was not related to age over the range 18--75 years. In persons with chronic obstructive lung disease (COPD), neither PaO2 sl (78.2 ± 11.3 mm Hg) nor PCO2 sl (40.5 ± 5.7 mm Hg) were related to age (77.0 ± 9.0 yrs). The relationship between PaO2 alt and PaO2 sl, PCO2 sl and altitude (feet) will be described.

### Protocol Development of the Cabin Altitude Study

❖ Dianne McMullin, The Boeing Company, P.O. Box 3707, MC 02 10, Seattle, WA 98124-2207, *Dianne.L.McMullin@boeing.com*; Mike Muhm, The Boeing Company; Stephen P. Jones, The Boeing Company; I-Li Lu, The Boeing Company; Paul Rock, Oklahoma State University

***Key Words:*** altitude, hypoxia, protocol

The effect of hypoxia on mountain climbers and aviators at altitudes above 10,000 feet has been well-studied. However, the effect of hypoxia on symptoms and well-being in individuals at altitudes encountered in pressurized aircraft cabins have not been studied for the age and sex distributions characteristic of passengers and cabin crew on commercial airlines. To determine the effects of hypoxia encountered during commercial flight, 502 healthy volunteers (21--78 years of age) were exposed to ground-level or one of four higher altitudes in a hypobaric chamber for 20 hours in groups of 12 during a single-blind, sequentially updated split-split plot factorial design. Self-reported symptoms were assessed using the Environmental Symptoms Questionnaire IV (ESQ IV). Performance measures including memory, vision, and hand-eye coordination tests.

### Statistical Methods in Cabin Altitude Study

❖ Stephen P. Jones, The Boeing Company, P.O. Box 3707, MC 7L-22, Seattle, WA 98124-2207, *stephen.p.jones@boeing.com*

***Key Words:*** experiment, altitude

A frequent, but often unrecognized, exposure to 'altitude' occurs during commercial airline flight. Cabins of commercial aircraft are pressurized so that, at cruise altitude, the cabin pressure is no lower than the pressure equivalent of 8,000 feet. Reduced partial pressure of oxygen at 8,000 feet induces a level of hypobaric hypoxia generally tolerated by healthy individuals but possibly stressful to those with compromised cardiopulmonary function. Boeing conducted an experiment to determine the effect of cabin pressures equivalent to altitudes up to 8,000 feet on oxygen saturation and discomfort in volunteers selected to represent commercial airline passengers during a simulated 20-hour flight in an altitude chamber. In this paper, we describe the development of statistical methods for the analysis of the data and present the findings of the experiment.

### Applications of Structural Equation Models: Case Studies in Biomedical and Aerospace Engineering Research

❖ I-Li Lu, The Boeing Company, P.O. Box 3707, MC 7L-22, Seattle, WA 98124-2207, *I-Li.Lu@boeing.com*

***Key Words:*** LISREL, time-series, cross-sectional, dynamic, path analysis, simultaneous equations

Data acquired in longitudinal form often exhibit strong association in response variables. This phenomenon sometimes may be explained by theoretical reasoning or can be attributed to the existence of structural relations that linked the variables. When the latter is alleged, one usually attempts to discover the underlying structure and hence provide an explanation for the relationship. It is natural to specify a model that will allow adequate description of interdependency of variables. In this talk, we will illustrate the model building and estimation process of structural equations in studies of three types of data. We will start with the usual experimental data collected from a human research study, then the retrospective observational data extracted from an air traffic management database, and finally the data obtained from an industrial human factor experiment.

## 373 Medical Expenditures: Data Collection, Estimation, and Evaluations ● ✪

Section on Survey Research Methods, Section on Health Policy Statistics, Biometrics Section, ENAR
**Wednesday, August 9, 8:30 am–10:20 am**

### Evaluation of the Completeness of Household Reports of Medical Expenditures for Visits to Physician Offices

Steven R. Machlin, Agency for Healthcare Research and Quality; ❖ Diana Wobus, Westat, 1650 Research Blvd., Rockville, MD 20850, *dianawobus@westat.com*; David Kashihara, Agency for Healthcare Research and Quality

***Key Words:*** medical expenses, response accuracy, survey quality

The Medical Expenditure Panel Survey (MEPS) collects data on health care utilization and expenditures in the U.S. civilian noninstitutionalized population. Unfortunately, it is difficult to obtain complete information about medical expenditures from household respondents. This methodological analysis uses 2003 MEPS data to examine and compare the extent to which household respondents provide complete data on out-of-pocket payments and, when applicable, on private insurance payments for visits to doctors' offices. Descriptive and multivariate statistical methods are used to analyze variation in completeness of household responses for these two payment sources, as well as for total expenses.

### Evaluation of the Accuracy of Household Reports of Medical Expenditures for Visits to Physician Offices

Steven R. Machlin, Agency for Healthcare Research and Quality;
❖ David Kashihara, Agency for Healthcare Research and Quality, CFACT DSRM, 540 Gaither Road, Rockville, MD 20850, *dkashiha@ahrq.gov*; Diana Wobus, Westat

***Key Words:*** response accuracy, medical reports, survey quality

A primary objective of the Medical Expenditure Panel Survey (MEPS) is to collect data on health care expenditures from households in the United States. Because it is difficult to obtain complete and accurate information from these respondents, MEPS also collects expenditure data in a Medical Provider Component (MPC) from a sample of providers who treated MEPS sample persons. These data generally are considered to be more accurate because they reflect records of payments received from patients, insurers, and other sources. The MPC data provide a unique opportunity to assess the extent to which household reported data correspond to provider reported data. Using 2003 MEPS data, multivariate methods are used to analyze variation in accuracy of household reports of payments made by selected sources for visits to doctors' offices.

### The Impact of Medical Expenditure Predictors in MEPS Nonresponse Adjustments

❖ Lap-Ming Wun, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, *lwun@ahrq.gov*; Trena Ezzati-Rice, Agency for Healthcare Research and Quality; Steven B. Cohen, Agency for Healthcare Research and Quality; William Yu, Agency for Healthcare Research and Quality

***Key Words:*** nonresponse, logistic regression, weight adjustment, response propensity

The MEPS is a large and complex sample survey sponsored by the Agency for Healthcare Research and Quality. The sample is drawn from respondents to the previous year's National Health Interview Survey. The MEPS, like most sample surveys, experiences unit nonresponse. Weighting adjustment is a general method to compensate for nonresponse. Currently, in the MEPS, a weighting class nonresponse adjustment based on relevant covariates modeled through the software CHAID is used. An alternative method of using response propensities calculated from logistic regression models has been investigated and reported previously. This paper reports on the investigation of the impact of including a high medical expenditure predictor as a covariate in the logistic model for nonresponse adjustment. The impact is evaluated by applying the adjusted weights to selected analytical variables.

### Evaluation of Expenditure Estimates When Including a High-Expenditure Predictor in Nonresponse Adjustments

Lap-Ming Wun, Agency for Healthcare Research and Quality;
Steven B. Cohen, Agency for Healthcare Research and Quality;
❖ Trena Ezzati-Rice, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, *tezzatir@ahrq.gov*; William Yu, Agency for Healthcare Research and Quality

***Key Words:*** nonresponse, logistic regression, weight adjustment

The MEPS is a large and complex sample survey sponsored by the Agency for Healthcare Research and Quality. The sample is drawn from respondents to the previous year's National Health Interview Survey. The MEPS, like most sample surveys, experiences unit nonresponse. Weighting adjustment is a general method to compensate for nonresponse. Nonresponse adjustments often are carried out based on relevant covariates. This paper reports on the investigation of the impact of including a high medical expenditure predictor as a covariate in the logistic model for nonresponse adjustment at the dwelling unit (DU) level. The DU nonresponse adjusted weight is carried through the post-stratification process to develop the final annual person-level weight. The impact of these modified person weights on expenditure estimates is evaluated and compared with current estimation procedures.

### Approximation of Skewed Health Care Expenditure Distribution Using a Mixture Model

❖ William Yu, Agency for Healthcare Research and Quality, 540 Gaither Road, Room 5054, Rockville, MD 20850-6649, *wyu@ahrq.gov*

***Key Words:*** healthcare expenditures, skewness, mixture model, MEPS

The expenditure data from the Medical Expenditure Panel Survey (MEPS) have been shown to exhibit a marked positive skewness, characterized with a few high-expenditure respondents and many zero-expenditure respondents. Any approximation of the distribution of MEPS expenditure will need to capture the bimodality feature of the MEPS expenditure data. A mixture model with a point mass at zero and the positive half of the real line is proposed. This mixture model captures the bimodality feature of the MEPS expenditure distribution and allows for evaluation of the probability that a MEPS respondent has zero expenditure varies significantly with a selected set of variables. The positive values of MEPS expenditure may be assumed to follow one of the Weibull, Gamma, or Log-normal distributions. The mixture model will be estimated and evaluated for goodness-of-fit with 2003 MEPS expenditure data.

## 374 Analyses of Studies Using Biomarkers ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

### Array Spatial Variability and Normalization Techniques for Microarray Gene Expression Signals

❖ Samir Lababidi, U.S. Food and Drug Administration, 1350 Piccard Drive, HFZ 550, Rockville, MD 20850, *Samir.Lababidi@fda.hhs.gov*; Daya Ranamukhaarachchi, U.S. Food and Drug Administration

***Key Words:*** microarrays, gene expression, spatial variability, normalization

We previously showed that microarray gene expression data can have substantial variability based on spatial arrangement of genes on the chip and that such data, without correction for array spatial variability, may not be reliable enough to be used in disease diagnosis and prognosis. In this talk, we will present a chip design to study array spatial variability for two-color arrays across the chip surface and different times of hybridization while accounting for possible dye effect. To explore these sources of variability, appropriate statistical models were developed and multiple testing procedures implemented to take into account the high-dimensionality of multiple genes. These models can be used to obtain a method for normalization of gene expression in which measurements for these genes would be more reliable.

### Statistical Issues in Incorporating and Testing Biomarkers in Clinical Trials

❖ Daniel Sargent, Mayo Clinic College of Medicine, Kahler 1A, 200 1st St SW, Rochester, MN 55905, *sargent.daniel@mayo.edu*; Sumithra Mandrekar, Mayo Clinic College of Medicine

***Key Words:*** biomarker, clinical trial, phase I, phase II, phase III

Novel therapies are challenging drug development. Many new therapies have mechanisms of action aimed at specific molecular targets whose efficacy is expected to be limited to patients with a certain biomarker, and/or have a purported biomarker to measure the agent's efficacy. This raises issues in phase I, II, and III clinical trial design. We will present novel designs to identify optimal doses of a single or two-drug combination in Phase I trials utilizing both toxicity and efficacy. Next, we will discuss single-arm or randomized Phase II trials where a biomarker dictates patient selection and/or treatment. Finally, we will discuss confirmatory biomarker-based Phase III trials where the drug's mechanism of action is known and a reliable assay is established. We believe rational use of biomarkers will facilitate drug development and ultimately help deliver individualized therapy.

### Statistical Design and Multiple Testing Analysis of Microarray

❖ Jane Chang, Bowling Green State University, Department of ASOR, Bowling Green, OH 43403, *changj@cba.bgsu.edu*; Jason Hsu, The Ohio State University

***Key Words:*** microarray, sensitivity, specificity, design, multiple testings

Microarray experiments are no longer for discovery only. There are microarray-based products marketed for cancer recurrence prognosis. Microarrays for clinical use should meet FDA joint CDRH/CDER statistical requirements. But, it has been observed that reported sensitivity and specificity of biomarker-based cancer prognostics from even major studies are not necessarily reproducible. We will first describe DESIGN issues with microarray experiments that likely have contributed to this apparent lack of reproducibility. Recommendations for statistically designing microarrays and sample hybridization toward reproducible results will be given. Multiple testings ANALYSIS issues of gene expression levels also will be discussed.

### Prediction Modeling Using Survival Data for Gene Expression Prognostic Test for Breast Cancer

❖ Kit Lau, Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, CA 94502, *kit.lau@celeradiagnostics.com*; Alice Wang, Celera Diagnostics; John Sninsky, Celera Diagnostics; Trevor Hastie, Stanford University

***Key Words:*** gene expression, prognostic test, prediction modeling, survival data, breast cancer

Gene expression profiles have been shown to be prognostic for cancer. A supervised principal components method (Bair 2004) using survival endpoint was used to build a breast cancer signature for distant metastasis. A metastasis score (MS) from gene expression was derived from the Cox model. Probabilities of distant metastasis at any time can be calculated from the MS. To avoid over-fitting, "pre-validation" technique (Tibshirani 2002) was used to estimate test performance. The MS was calculated with 10-fold cross-validation (MS(CV)). Time-dependent AUC (Heagerty 2000), hazard ratios, and survival rates of stratified groups from MS(CV) were evaluated. Univariate and multivariate Cox regression were performed for the MS(CV) and Adjuvant!, an online prognosticator. An integrated MS that combines Adjuvant! and gene expression data had an improved AUC over both.

## 375 What Is Feminist Statistics?

● ✪

Social Statistics Section, Caucus for Women in Statistics, Section on Government Statistics, Section on Statistical Education

**Wednesday, August 9, 8:30 am–10:20 am**

### What Is Feminist Statistics?

❖ Martha Aliaga, American Statistical Association, 732 North Washington Street, Alexandria, VA 22314, *martha@amstat.org*; ❖ Arlene Ash, Boston University, School of Medicine, 720 Harrison Ave., Suite 1108, Boston, MA 02118, *aash@bu.edu*; ❖ Eduardas Valaitis, American University, Department of Mathematics and Statistics, Washington, DC 20016, *valaitis@american.edu*

***Key Words:*** women, feminism, pedagogy, career, social perspectives, experience

In light of the recent controversy about women's innate ability (or lack thereof) to do science and mathematics, we ask whether, in fact, there is a connection between feminism and statistics. Feminist pedagogy and its broader benefits, the use of statistics to support or undermine feminist issues, career paths for feminists in statistics, and whether feminist statistics exists will be addressed. Speakers will present generational and societal perspectives and speak of their own research and experience.

## 376 Mentoring Statisticians ●

Section on Statistical Education, Committee on Career Development

**Wednesday, August 9, 8:30 am–10:20 am**

### Mentoring Statisticians

❖ Amy Froelich, Iowa State University, 324 Snedecor Hall, Ames, IA 50011-1210, *amyf@iastate.edu*; ❖ Sastry Pantula, North Carolina State University, 2501 Founders Drive, 201E Patterson Hall, Raleigh, NC 27695-8203, *pantula@stat.ncsu.edu*; ❖ Sally C. Morton, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *morton@rti.org*; ❖ Cynthia Clark, U.K. Office for National Statistics, Royal Belgrave House, Flat 32, Hugh Street, London, SW1V 1RR UK, *cynthia.clark@ons.gov.uk*; ❖ Ronald Menton, Wyeth Research, 641 Ridge Road, Chazy, NY 12921-2420, *mentonr@wyeth.com*

***Key Words:*** mentoring

Formal or informal mentoring has occurred for several years among members of the statistics profession. In 2004--2005, an ASA task force examined mentoring and nurturing statisticians. This session will present examples. Amy Froelich will discuss how graduate assistants develop their teaching skills through an apprenticeship-like process at Iowa State University. Sastry Pantula will discuss the NC State program for junior faculty and NSF VIGRE program for postdocs. Cynthia Clark will discuss programs she implemented at the U.S. Census Bureau and UK Office for National Statistics. Sally Morton will discuss the challenges of constructing and implementing mentoring programs at nonprofit research institutions, drawing on her experience at RTI International and RAND Corporation. Finally, Ron Menton will discuss examples of mentoring programs in the pharmaceutical industry.

## 377 Capture-Recapture and Other Problems in Environmental and Ecological Statistics ●

Biometrics Section, ENAR
**Wednesday, August 9, 8:30 am–10:20 am**

### Semiparametric Models for Capture-Recapture Experiments with Behavoral Response

❖ Wen-Han Hwang, Feng Chia University, Department of Statistics, Taichung, 40724 Taiwan, *whhwang@fcu.edu.tw*; Richard Huggins, University of Melbourne

***Key Words:*** capture-recapture, population size, semiparametric model

We introduce a semiparametric framework for capture-recapture experiments to extend the parametric capture-recapture models of Huggins (1989) to a broader range of applications. The resulting models for the capture probabilities include parametric components for some effects (e.g., time effect) and nonparametric components for both individual heterogeneity and behavioral response. In particular, the behavioral response is framed as a smooth function of an individual characteristic covariate, providing a new strategy for modeling. The method is evaluated in simulations and applied to several real datasets.

### Population Estimation for Noninvasive Trapping

❖ Emily Murphree, Miami University, Department of Mathematics and Statistics, Oxford, OH 45056, *murphres@muohio.edu*

***Key Words:*** DNA mark-recapture, grizzly, hair trap, multinomial, population size, capture probability

Biologists are increasingly using non-invasive trapping techniques in their attempts to gauge the size of an animal population. I will outline a technique for estimating population size which is tailored to these "hair-trap" sampling schemes. One important feature of these schemes is that the number of visits to the trap at each sampling stage is not known. Under reasonable assumptions, numbers of distinct new animals captured at stages 1, 2, ..., S will have a multinomial distribution depending on N, the population size. Also, the number of distinct "repeaters" and the number of new animals captured at each stage will share the same capture probability, which can vary from stage to stage. I will discuss a maximum likelihood estimator of N based upon these results.

### Assessing Similarity of Two Assemblages with Unseen Species in Samples

❖ Tsung-Jen Shen, National Chung Hsing University, 250 Kuo Kwang Road, Department of Applied Mathematics, TaiChung, 405 Taiwan, *tjshen@amath.nchu.edu.tw*; Anne Chao, National Tsing Hua University; Robin L. Chazdon, University of Connecticut; Robert L. Colwell, University of Connecticut

***Key Words:*** beta diversity, biodiversity, forest succession, species overlap

A variety of similarity indexes for comparing two assemblages based on species incidence data have been proposed in the literature. These indexes are generally in terms of two simple incidence counts: the number of species shared by two assemblages and the number of species unique to each. We provide a new, probabilistic derivation for any incidence-based index that is symmetric and homogeneous. The probabilistic approach is further extended to formulate abundance-based indexes. Thus, any symmetric and homogeneous incidence index can be modified easily to an abundance-type version. Applying the Laplace approximation formulas, the proposed estimators that adjust for the effect of unseen shared species on our abundance-based indexes will be presented. Data on successional vegetation in six tropical forests are used for illustration.

### Estimating the Species Richness by a Poisson-Compound Gamma Model

❖ Ji-Ping Wang, Northwestern University, 2006 Sheridan Road, Evanston, 60208, *jzwang@northwestern.edu*

***Key Words:*** species richness, Poisson mixture

Suppose \$D\$ distinct species are observed from a infinite population consisting of total \$N\$ (unknown) species. It has been well recognized that \$N\$-estimate is unbounded if infinitely rare species are possible. Here we consider to estimate the number of species whose abundance is above a threshold. We model the probability of observing \$X\$ individuals from each species by a Poisson-mixed Gamma model, i.e. \$f[X;Q(\lambda)]=\int\frac{e^{-\lambda}\lambda^x}{x!}dQ(\lambda)\$ where \$Q\$ itself is a mixed Gamma. The Gamma is estimated by nonparametric maximum likelihood method. Instead of estimating \$N\$, we estimate \$\sum I(\lambda_i>\lambda_0)\$ based on the posterior distribution of \$\lambda|X\$. Confidence interval is constructed based on a resampling procedure. This method is illustrated using simulated and real data.

### Mixtures of Exponential Distributions To Describe the Distribution of Poisson Means in Estimating the Number of Unobserved Classes

❖ Kathryn Barger, Cornell University, 301 Malott Hall, Ithaca, NY 14853, *kjb34@cornell.edu*

***Key Words:*** species richness, EM algorithm, finite mixture model, Aitken's acceleration, microorganisms

In many fields of study scientists are interested in estimating the number of unobserved classes. For instance, a biologist may want to find the number of rare species of an animal population in order to study biodiversity. Based on the traditional negative binomial model for this data (Fisher, Corbet, and Williams 1943), we propose a model assuming that the number of individuals from each class are independent Poisson samples, and that the means of these Poisson random variables follow a finite mixture of exponential distributions. We fit the proposed model using a nested EM algorithm and incorporate Aitken's acceleration in a mixture of three exponentials. Microbial samples from the coast of Massachusetts Bay near Nahant, Massachusetts are analyzed to demonstrate that the model produces reasonable estimates and fits the data satisfactorily.

### Principle Component Analysis as a Statistical Tool To Investigate the Role of Specific Habitat Variables in Lyme Disease Ecology

❖ Haiyan Chen, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, *hchen@jimmy.harvard.edu*

***Key Words:*** principal component analysis, b. burgdorferi prevalence, small mammal, habitat variable

To investigate the impact of habitat variables on B. burgdorferi prevalence within reservoir mammals, forest and mammal surveys were conducted at 12 sites in New York between 1998 and 2001. Spearman rank correlations between each of 100 habitat variables collected and average species-adjusted B. burgdorferi infection rates (SAR) of the mammals in a site were calculated, 13 of which had p-value < 0.10 and were selected for a principal component analysis (PCA). Shrub, fern, tree, and water are deterministic factors regarding habitat characteristics in a given area. Backward multiple regression analysis was conducted to further identify the best predictors of the SAR. This study found that PCA is an efficient approach to condense the information contained in a large number of original variables into a smaller set of new composite dimensions with a minimum loss of information.

## 378 Models for Multivariate (Longitudinal) Data ●

Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

### Alternative Structural Models for Analyzing Multivariate Longitudinal Data

❖ Feng Gao, Washington University School of Medicine, 660 S. Euclid Ave., Campus Box 8067, St Louis, 63021, *feng@wustl.edu*; Paul Thompson, Washington University School of Medicine; Chengjie Xiong, Washington University School of Medicine; J. Philip Miller, Washington University School of Medicine

***Key Words:*** multivariate longitudinal data, structural equation modeling, cross-lagged regression model, latent growth curve model

Multivariate longitudinal data provides a unique opportunity in studying the joint evolution of multiple response variables over time. However, the analysis of multivariate longitudinal data can be challenging because the errors are likely to be correlated for the same marker measured at different occasions and the errors are also likely to be correlated among markers measured at the same time. Structural equation modeling (SEM) is a comprehensive statistical approach to identify patterns of directional and non-directional relationships among a set of variables. In this talk, with application to a real-world study to evaluate the joint evolution of the biomarkers for renal structure and function, 3 alternative SEMs are presented and compared: a) a cross-lagged regression model, b) a latent growth curve model and c) a dynamic model based on latent difference scores.

### Unconstrained Models for the Covariance Structure of Multivariate Longitudinal Data

❖ Chulmin Kim, University of Minnesota, Morris, 704 Imperial Drive, #309, Morris, MN 56267, *kimc@umn.edu*; Dale Zimmerman, The University of Iowa

***Key Words:*** multivariate longitudinal, covariance matrix, unconstrained parameterization, Cholesky decomposition, antedependence, maximum likelihood estimators

The constraint that a covariance matrix must be positive definite presents difficulties for modeling its structure. In a series of papers published in 1999 and 2000, Mohsen Pourahmadi proposed a parameterization of the covariance matrix for univariate longitudinal data in which the parameters are unconstrained. This unconstrained parameterization is based on the modified Cholesky decomposition of the inverse of the covariance matrix. We extend this idea to multivariate longitudinal data. We develop a modified Cholesky block decomposition that provides an unconstrained parameterization for the covariance matrix. A Newton-Raphson algorithm is developed for obtaining maximum likelihood estimators of model parameters. The results along with penalized likelihood criteria such as BIC for model selection are illustrated using a real multivariate longitudinal dataset and a simulated data set.

### Multilevel Flexible Models for Mixed Longitudinal Data

❖ Nuoo-Ting Molitor, University of Southern California, , *nuootinl@usc.edu*; Kiros Berhane, University of Southern California

***Key Words:*** splines, mixed effects models, mixed outcomes, Gibbs sampling, air pollution

The proper understanding of the joint and inter-related long-term effects of air pollutants on pulmonary functions and respiratory symptoms (e.g. bronchitis) is of the interest in protecting children's respiratory health. To this end, we propose a flexible multi-level modeling technique that uses splines, whenever appropriate, to model bivariate mixed longitudinal outcomes. A latent variable approach is used to connect both outcomes within a subject. This technique allows for the examination of the inter-relationship between pulmonary functions and respiratory symptoms, after adjusting for subject-specific growth-curve parameters. A Gibbs sampling approach is implemented to estimate the posterior distribution of the parameters and latent variables. Ecologic inference is then conducted in a multi-level setting. We illustrate the technique via analysis of data from the Southern California.

### Conditional Estimation for Joint Models for a Primary Endpoint and Multivariate Longitudinal Data

❖ Erning Li, Texas A&M University, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, *eli@stat.tamu.edu*; Naisyin Wang, Texas A&M University; Nae-Yuh Wang, Johns Hopkins University School of Medicine

***Key Words:*** multivariate random effects model, generalized linear model, asymptotic bias, conditional score, variance components, measurement error

We study the association between a primary endpoint and features of multiple longitudinal processes using joint models. In the joint model, subject-specific random effects from a multivariate linear random effects model for the multiple longitudinal processes are used as predictors in a generalized linear model for the primary endpoint. An asymptotic bias analysis indicates that the estimators obtained by Li et al. (2004, Biometrics 60, 1-7), which make no distributional assumption on random effects but assume independent within-subject errors in the longitudinal covariate process, can yield biased inference when these within-subject errors are in fact correlated. To overcome this drawback, we generalize their results to joint models with more flexible multivariate longitudinal covariate processes and develop inferential methods for the regression parameters which are easy to implement.

### A Bayesian Approach to Modeling Associations between Pulsatile Hormones

❖ Nichole Carlson, Oregon Health & Science University, Division of Biostatistics, 3181 SW Sam Jackson Park Road, Mail Code CB669, Portland, OR 97239, *carlsoni@ohsu.edu*; Timothy D. Johnson, University of Michigan; Morton B. Brown, University of Michigan

***Key Words:*** hormone data, biological modeling, birth-death MCMC, pulse association

Many hormones are secreted in pulses. The pulsatile relationship between hormones regulates many biological processes. To understand endocrine system regulation, time series of hormone concentrations are collected. The goal is to characterize pulsatile patterns and associations between the hormones. When the signal-to-noise ratio is large, pulse detection and parameter estimation is difficult with existing approaches. We present a bivariate deconvolution model of pulsatile hormone data using a Bayesian approach, which addresses these issues. We describe a model for a one-to-one, driver-response case and show how birth-death MCMC can be used for estimation. We exhibit that using known pulsatile associations and bivariate fitting improves estimation of the pulse locations and the parameters for each hormone. An example is presented using luteinizing and follicle stimulating hormones.

## 379 Robust Solutions ●

Business and Economics Statistics Section

**Wednesday, August 9, 8:30 am–10:20 am**

### Competitiveness Analysis of the Italian Firms: Use of Robust Classification Methods

❖ Matilde Bini, University of Florence, Viale Morgagni 59, Department of Statistics, Florence, 50134 Italy, *bini@ds.unifi.it*; Luigi Biggeri, Italian National Statistical Institute

***Key Words:*** competitiveness, cluster analysis, forward search, multivariate transformation, robustness

In Italy, during the last ten years there as been a fall of the competitiveness in comparison with most important European and extra European countries. The main causes of this phenomenon are due to the poor development of the total productivity of the firms, related to the low investments in innovation of processes, products and organizational structure of the firms, and to the revaluation of the Euro in comparison with the dollar. The aim of this work is to carry out some analyses and produce results that could be used to implement intervention policies to improve competitiveness among firms. The hypothesis is that the Italian firms are very heterogeneous as competitiveness and that their results depend on their performances. Therefore our objective is also to try to detect groups of firms that reached different results in relations to their different characteristics.

### On Robust Forecasting in Dynamic Vector Time Series Models

❖ Pierre Duchesne, UniversitÈ de MontrÈal, DÈpartement de mathÈmatiques et statistique, CP 6128 Succursale Centre-Ville, MontrÈal, PQ H3C3J7 Canada, *duchesne@dms.umontreal.ca*; Christian GagnÈ, UniversitÈ de MontrÈal

***Key Words:*** multivariate time series, exogenous variables, prediction, robust estimators, additive outliers

Robust estimation/prediction in multivariate autoregressive models with exogenous variables (VARX) are considered. The conditional least squares estimators (CLS) are known to be non robust when outliers occur. To obtain robust estimators, the method introduced in Duchesne (2005) and Bou Hamad and Duchesne (2005) is generalized. The asymptotic distribution of the new estimators is studied and from this is obtained in particular the asymptotic covariance matrix of the robust estimators. The occurrence of outliers may invalidate the usual conditional prediction intervals. Consequently, the new robust methodology is used to develop robust conditional prediction intervals. In a simulation study, we investigate the finite sample properties of the robust prediction intervals under several scenarios for the occurrence of the outliers, and the new intervals are compared to non-robust intervals.

### A Test of the Modality of the Variance Function in Modulated Autoregressive Models

❖ Gabe Chandler, Connecticut College, Campus Box 5348, 270 Mohegan Ave, New London, CT 06320, *gjcha2@conncoll.edu*

***Key Words:*** modality, autoregression, heteroscedasticity, scans

A test of modality of the variance function for a modulated autoregressive model, i.e. autoregressive processes with errors whose variance depends on time, is presented. The test is based on the idea that, given two candidate modes, a function is bimodal if it is not monotonic between the two candidates. An estimate of a quantile q of the residuals is found, and the test statistic d(k,q) is the longest time span that contains at most k residuals which are larger than the estimate of q. Unimodality is rejected if d(k,q) is too large. The test is calibrated against the hypothesis that the function is flat on some interval between the two candidates. Under the null, should the errors and quantile be known, finding the p-value would amount to considering the distribution of the longest scan of type k.

### Multiple Imputation of Right-Censored Data: an Application to Wage Data and Understanding the Changing Wage Gap by Gender in Germany

❖ Hermann Gartner, Institute for Employment Research, Regensburger Strasse 104, Nurenberg, 90461 Germany, *hermann.gartner@iab.de*

***Key Words:*** multiple imputation, missing data, gender wage inequality, censored wage data

In many large data sets of economic interest, some variables, as wages, are top-coded or right-censored. For example the US Census's CPS, the Austrian social security records and the IAB-employment register (IABS). We treat this problem as one of missing data, where the missingness mechanism is not missing at random but rather missing by design (Little & Rubin,2002). Our approach is based on the multiple imputation (Rubin, 1978, 2004) of the right censored variables by drawing values from appropriate truncated distributions. This approach makes an important difference to realistic answers. We illustrate the approach with the German IABS, and analyze the gender gap by a Juhn-Murphy-Pierce decomposition. One important conclusion is that the main source for the narrowing gender wage gap between 1991 and 2001 in Germany is an improvement of women's position within the wage distribution.

### A New Approach to Univariate Unit Root Tests Robust to Structural Change

❖ Seong-Tae Kim, North Carolina State University, Department of Statistics, 2501 Founders Drive, 201 Patterson Hall, Raleigh, NC 27695-8203, *skim@stat.ncsu.edu*

***Key Words:*** unit root test, structural change, sequential limit

Using methodology in panel unit root tests we propose a new approach to univariate unit root tests. Our method leads to an asymptotically normal distribution of the OLS estimator and is robust to structural changes while the power of the test does not drastically worsen. The main idea is that under the assumption that the process has a unit root we transform a univariate time series process to a double-index process in such a way that the segments are independent, i.e., we divide the T observations into m sets of size n where m sets of data are independent of each other. For this transformed data, we apply the sequential limits to obtain the normal limiting distribution, which is the same as in Levin and Lin (1992). An advantage of this technique is that an undetected break has a relatively minor effect on the asymptotic results which, in fact, disappears as the segments increase.

### Testing for Threshold Moving Average with Conditional Heteroscedasticity

❖ Guodong Li, The University of Hong Kong, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, 00852 China, *ligd@hkusua.hku.hk*; Wai K. Li, The University of Hong Kong

***Key Words:*** GARCH errors, Gaussian process, likelihood ratio test, threshold MA model

The recent paper by Ling and Tong (2005) considered a quasi-likelihood ratio test for the threshold in moving average models with i.i.d. errors. This article generalizes their results to the case with GARCH errors and a new quasi-likelihood ratio test is derived. The generalization is not direct since the techniques developed for TMA models heavily depend on the property of p-dependence which is no longer satisfied by the time series models with conditional heteroscedasticity. The new test statistic in this article is shown to converge weakly to a functional of a centered Gaussian process under the null hypothesis of no threshold and it is also proved that the test has nontrivial asymptotic power under local alternatives. Monte Carlo experiments demonstrate the necessity of our test as a moving average time series has a time varying conditional variance.

### A New Approach for Calculating Sample Size To Detect Desired Difference between Treatment Groups with Intended Power

❖ Seemit Sheth, Capital One Financial Corporation, 140 East Shore Drive, Glen Allen, VA 23059, *seemit.sheth@capitalone.com*

***Key Words:*** test, sample size, significance level, power

Often, we wish to detect a certain minimum difference between two treatment groups. For example, a financial services firm may be interested in determining if there is at least a 10% increase in revenue with a new product over an existing product since a smaller increase may not be worth the cost of switching to a new product. The traditional approach is to use this minimum difference, power, significance level and standard deviations to arrive at the required sample size. First, I will discuss the limitations of the traditional approach and explain with examples why many tests fail to detect the desired difference even if it exists. Then I will propose an approach that utilizes our prior information on expected test performance to create tests that enable us to detect desired differences between treatment groups with intended power.

## 380 Flexible Methods for Longitudinal Data ●

ENAR, Biometrics Section

**Wednesday, August 9, 8:30 am–10:20 am**

### Efficient Estimation in Semiparametric Generalized Linear Model for Longitudinal Data

❖ Lu Wang, Harvard University, 655 Huntington Ave., Sph 2, 4th floor, Department of Biostatistics, Boston, MA 02115, *luwang@hsph.harvard.edu*; Xihong Lin, Harvard School of Public Health; Andrea Rotnitzky, Harvard University

***Key Words:*** semiparametric generalized linear model, longitudinal/clustered data, generalized estimating equation, iterative kernel GEE estimator, semiparametric efficient score, semiparametric information bound

We consider estimation in semiparametric generalized linear model for longitudinal or clustered data. Conventional profile-kernel method fails to yield a semiparametric efficient estimator for the coefficients of parametric covariates. Wang, Carroll and Lin proposed an iterative kernel generalized estimating equation (GEE) estimator, which accounted for within-cluster correlation and is more efficient. We derive the semiparametric efficient score function and the semiparametric information bound in general scenarios under the semiparametric conditional mean model. We discuss the asymptotic properties of the iterative profile-kernel GEE estimator and show that it is semiparametric efficient. Simulations will be performed to demonstrate our results. Further extensions to the missing data case will be considered.

### Varying-Coefficient Model with Unknown within-Subject Covariance for the Analysis of Tumor Growth Curves

❖ Robert Krafty, University of Pennsylvania, 501 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, *rkrafty@cceb.upenn.edu*; Wensheng Guo, University of Pennsylvania; Phyllis Gimotty, University of Pennsylvania; George Coukos, University of Pennsylvania

***Key Words:*** varying-coefficient model, smoothing spline, functional linear models, tumor growth curves

In this article, we develop a method for estimating the varying-coefficient model containing unknown within subject covariance. Extending the idea of iterative weighted least squares to the functional setting, we iterate between estimating the functional subject means conditional on the covariance and estimating the functional covariance conditional on the subject mean curves. Smoothing splines for correlated errors are used to estimate the functional parameters of the mean with smoothing parameters selected via the generalized maximum likelihood. The covariance is non-parametrically estimated using a penalized estimator with smoothing parameters chosen via a Kullback-Leibler criterion. The method is used in a study to analyze the effects of different chemotherapy treatments on the volumes of two classes of ovarian tumors in mice.

### Modeling Plasma HIV Viral Load by a Piecewise Polynomial Linear Mixed Model

❖ Hsiao-Chuan Tien, The University of North Carolina at Chapel Hill, CB 7420 Department of Biostatistics, UNC Chapel Hill, Chapel Hill, NC 27599 7420, *htien@bios.unc.edu*; Pai-Lien Chen, Family Health International

***Key Words:*** piecewise polynomial linear mixed model, Markov chain Monte Carlo

The copies of plasma HIV viral load is an important virologic marker for HIV infection. The level of viral load at the set point becomes a crucial index of predicting disease progression. In this study we propose a piecewise polynomial mixed effect model to estimate the level of viral load at set point and to model the plasma HIV viral load process. Our proposed approach is able to integrate the variability of the time of reaching the viral peak level and the viral setpoint among individuals and to incorporate other covariates into the model for evaluating their impact on the level of viral load. The Markov chain Monte Carlo (MCMC) algorithm is used to obtain the estimations of those parameters. Data from a HIV cohort study in Africa is used to illustrate the model.

### Nonparametric Inference in the Heteroscedastic Two-Way Random Effects Model Based on Ranks

❖ Zhe Shang, Wyeth Research, 500 Arcola Road, Collegeville, PA 19426, *shangz@wyeth.com*

***Key Words:*** nonparametric, R estimates, ranks, robust

Two rank-based estimators for analyzing heteroscedastic two-way mixed effects models, assuming only symmetric but otherwise arbitrary continuous distributions for the random effects and random errors, are proposed. One of the rank-based estimators is developed in a robust test for the random effect. Unlike the parametric analysis, heavy-tailed distributions without finite moments pose no problems in the rank approach. A real data set is examined using the proposed method and compared with the parametric approach.

### Marginal Regression Modeling under Irregular, Biased Sampling

❖ Petra Buzkova, The University of North Carolina at Chapel Hill, Department of Biostatistics and Lineberger Comp Cancer Center, University of North Carolina, CB#7295, Chapel Hill, NC 27599, *pbuzkova@bios.unc.edu*; Thomas Lumley, University of Washington

***Key Words:*** estimating equations, longitudinal data, biased sampling

In longitudinal studies, observations often are obtained at continuous subject-specific times. Frequently, the availability of outcome data may be related to the outcome measure or other covariates related to the outcome measure. Under such biased sampling designs, unadjusted regression analysis yields biased estimates. Building on the work of Lin and Ying (2001) that integrates counting processes techniques with longitudinal data settings, we propose a class of estimators that can handle biased sampling. We call those estimators "inverse-intensity-rate-ratio-weighted" (IIRR) estimators. The proposed class of closed-form estimators is root n consistent and asymptotically normal. It does not require estimating any infinite-dimensional parameters. The estimators and estimators of their variance are relatively simple and computationally feasible.

### Projected Multivariate Linear Mixed-Effects Models for Clustered Angular Data

❖ Daniel Hall, University of Georgia, Department of Statistics, Athens, GA 30602-1952, *dhall@stat.uga.edu*; Lewis Jordan, University of Georgia; Jinae Lee, University of Georgia

***Key Words:*** directional data, circular statistics, longitudinal data, repeated measures, EM algorithm, random effects

In this talk we extend the projected multivariate linear model of Presnell et al. (1998) to the clustered data case via the inclusion of cluster-specific random effects. For low dimensional random effects, we describe EM and Monte Carlo EM algorithms to accomplish maximum likelihood estimation in these models. For higher dimensional random effects, a penalized quasilikelihood type estimation approach is proposed. Finite sample properties of these estimation methods are investigated via simulation study. An example involving the microbial angle of loblolly pine, a property related to wood strength, is used to motivate and illustrate this class of models.

### Optimal Estimators from Generalized Estimating Equations (GEE) for Longitudinal Data

❖ Ioana Schiopu-Kratina, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A0T6 Canada, *ioana.schiopu-kratina@statcan.ca*; Raluca M. Balan, University of Ottawa

***Key Words:*** generalized estimating equations, longitudinal studies, asymptotic efficiency

Liang and Zeger (LZ) (1986) used marginal models with longitudinal data to implicitly express a response variable in terms of covariates and a main regression parameter. Estimators of this parameter obtained as solutions of GEEs were studied by LZ, J.N.K Rao(1998), Schiopu-Kratina (SK)( 2003), Xie and Yang (2003), Balan and Schiopu-Kratina (BSK) (2005). Asymptotically and in the context of Heyde 1997, a sequence of estimators corresponding to asymptotically efficient EEs are optimal in that the limiting confidence interval they define is minimal. This is a continuation of the work in BSK (The Annals of Statistics 2005, 33, 522-541). We show that the slightly modified estimators in SK and BSK are optimal and simplify the conditions in BSK.

## 381 Survival, Time to Event ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

### Predicting Time of Completion in Multiphase Survival Trials

❖ Dennis Sweitzer, AstraZeneca Pharmaceuticals, 438 Gum Tree Road, Coatesville, PA 19320, *dennis.sweitzer@astrazeneca.com*

***Key Words:*** clinical trial, survival, maintenance of effect, trial management, simulation

Studying maintenance of clinical effect typically requires clinical response for a minimum amount of time on treatment before randomization. If randomized, patients are followed until treatment failure or withdrawal, and the trial is halted after a prespecified number of events. For ethical and cost reasons, it is desirable to minimize the number of patients enrolled and randomized and to predict the time of the last event under multiple scenarios. We describe a data-driven stochastic simulation for two such trials in which each phase is modeled as a competing event process; distributions of event times are derived from Kaplan-Meier survival curves from available data; parameter uncertainty is modeled based on K-M survival estimates; withdrawals and events occur at similar overall rates, though at different times; and predictions are updated as information is accrued.

### Power Calculation for Log-Rank Test under a Nonproportional Hazards Model

❖ Daowen Zhang, sanofi-aventis, 906 Sunny Slope Road, Bridgewater, NJ 08807, *daowen.zhang@sanofi-aventis.com*; Hui Quan, sanofi-aventis

***Key Words:*** complete follow-up for ITT analysis, dropout and random censoring, piecewise exponential

The log-rank test is the most powerful nonparametric test for detecting a proportional hazards alternative; and thus, is the most commonly used procedure for analyzing time-to-event data in clinical trials. When the log-rank is used for data analysis, the power calculation should also be based on the log-rank test (Schoenfeld, 1983 Biometrics). In some clinical trials, treatment may not manifest its effect right after patients receive the treatment. Therefore, the proportional hazards assumption may not hold. We derive formulas for the asymptotic power calculation for the log-rank test under this non-proportional hazards alternative. Simulation studies indicate that the formulas provide reasonable sample sizes for a variety of trial settings. An example will be used to compare our methods with other methods that are suitable for a proportional hazards model or exponential distribution.

## Use of Life Tables To Extrapolate Survival from Clinical Trial Data

❖ W. J. Hall, University of Rochester Medical Center, Rochester, NY 14642-8630, *hall@bst.rochester.edu*; Hongyue Wang, University of Rochester Medical Center

***Key Words:*** device trial

Practical needs may require extrapolation of survival curves beyond the limited time span observed in a clinical trial - e.g., in economic forecasting of the potential impact of a new expensive device or therapy. The observation span may not be sufficient to permit modeling of the inevitable effects of aging on future survival. We use US life tables to develop a parametric model for survival in a subpopulation matched by age and sex to the clinical trial. Using the resulting cumulative hazard function as a time transformation, the subpopulation has exponential survival on an accelerated time scale. This facilitates modeling for the control arm of the trial by assuming proportional hazards to the US; modeling of treatment-arm survival can then be done to reflect assumptions about treatment efficacy on future survival. We illustrate with data from a recent device trial.

## Analyzing Change in Hazard for Time-to-Event Endpoints in Clinical Trials

❖ Rafia Bhore, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Room 6220, Silver Spring, MD 20993-0002, *Rafia.Bhore@fda.hhs.gov*; Sandra Gardner, Sunnybrook Health Sciences Centre

***Key Words:*** change point, hazard function, survival, time to event, exponential distribution, inference

Risk of a serious adverse event with long-term exposure to a drug may increase or decrease over time, or similarly, the efficacy of a drug may show changes in responses through time to onset of response, time to relapse of a disease, etc. Change-point methodology is relatively new in its application to clinical trials with censored survival data where the underlying hazard rate abruptly changes over time. This talk will review and compare likelihood-based methods in the literature for estimating change points in the hazard function. Through simulated data based on parametric survival models with change point, we show how to determine point estimates and interval estimates of the change point for different methods. Methods will be compared with respect to length and coverage of confidence intervals.

## Time to Rescue as a Surrogate Endpoint for Analgesic Efficacy in Acute Pain Studies

❖ Julia Wang, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., PO Box 300, Raritan, NJ 08869, *jwang@prdus.jnj.com*; Akiko Okamoto, Johnson & Johnson Pharmaceutical R&D; Surya Mohanty, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** surrogate endpoints, analgesic trials, Prentice's criteria, proportion of treatment effect explained

In controlled clinical trials involving acute pain, efficacy variables such as pain relief/intensity scores are collected longitudinally. The time course of these variables is vital for understanding analgesic effect of the test compound. However, subjects often require rescue with a supplemental analgesic medication, making it hard to characterize the analgesic activity of the new compound. We propose that the "time to rescue medication" be adopted as a primary endpoint in place of the conventional primary endpoints derived from the longitudinally collected pain relief/intensity scores by demonstrating the former to be a surrogate of the later. Validation of the proposed surrogacy will be demonstrated through various statistical methods, including graphical techniques and Prentice's criteria. Estimates of proportion of treatment effect explained by the surrogate will be provided.

## On Analysis of Time to Progression

❖ Zhiping Sun, Merck & Co., Inc., 608 Elmway Circle, Blue Bell, PA 19422, *linda_sun@merck.com*; Cong Chen, Merck & Co., Inc.

***Key Words:*** time to progression, survival, surrogate

Time to (disease) progression serves as a critical intermediate endpoint to survival. With unprecedented number of new oncology therapies being tested, the role of time to progression in oncology drug development is becoming increasingly important. In this presentation, statistical issues will be discussed when time to progression is considered as a clinical endpoint, a time-dependent covariate, or an intermediate endpoint to survival in clinical trials which have a primary objective of demonstrating superiority of a new therapy versus a comparator timely and reliably. Relevant clinical design issues will also be discussed.

## Testing for Change-Points in Waiting Time Distributions

❖ Thomas Hammerstrom, U.S. Food and Drug Administration, 12111 Willow Wood Drive, Silver Spring, MD 20904, *tom.hammerstrom@fda.hhs.gov*; Rafia Bhore, U.S. Food and Drug Administration

***Key Words:*** changepoint, piecewise exponential, bootstrap

This paper will present the results of some simulations on fitting and estimating parameters of piece-wise exponential (PWE) models for time to event data. It will focus on two particular problems: 1) the accuracy of the likelihood ratio in choosing the correct distribution for the times among exponential, PWE with one change-point, PWE with two change-points, and Weibull; 2) comparison of true and nominal coverage levels of confidence intervals for parameters computed by bootstrap and parametric methods. Results will include data with and without time varying covariates.

# 382 Clustering and Classification

●

Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

## Classification by Ensembles from Random Partitions of High-Dimensional Genomic Data

❖ Hojin Moon, U.S. Food and Drug Administration, 3900 NCTR Road, HFT20, Jefferson, AR 72079, *hojin.moon@fda.hhs.gov*; Hongshik Ahn, Stony Brook University; James J. Chen, U.S. Food and Drug Administration; Ralph L. Kodell, U.S. Food and Drug Administration

***Key Words:*** class prediction, classification tree, cross validation, majority voting, risk profiling

A robust classification algorithm is developed based on ensembles of classifiers, with each classifier constructed from a different set of predictors determined by a random partition of the entire set of predictors. The proposed method combines the results of multiple classifiers to achieve a substantially improved prediction compared to the optimal single classifier. By combining classifiers built from each subspace of the predictors, our algorithm achieves a huge computational advantage in tackling the growing problem of dimensionality. We investigate the performance of our method compared to widely used classification methods using real data. Our classification algorithm has many applications, including the early detection of disease in apparently disease-free individuals based on multiple biomarkers and the assignment of patients to drug therapies based on genomic and other profiles.

## Ensemble Methods for Classifying an Ordinal Response

❖ Kellie J. Archer, Virginia Commonwealth University, 1101 E. Marshall Street, Sanger Hall B1 069, Richmond, VA 23298-0032, *kjarcher@vcu.edu*

***Key Words:*** ensemble methods, machine learning, random forest, ordinal response, classification

Ensemble methods have been demonstrated to be competitive with other machine learning approaches for classification and described for nominal, continuous, and survival responses. In a large number of biomedical applications, the class to be predicted may be ordinal. Examples of ordinal outcomes include TNM stage (I, II, III, IV) and drug toxicity (none, mild, moderate, severe). While nominal methods may be applied to ordinal response data, in so doing some information is lost that may improve the predictive performance of the classifier. As ordinal classification methods have been neglected in the machine learning literature, extensions for predicting an ordinal response using a bootstrap aggregating framework are presented in this talk. These extensions will be evaluated against traditional nominal and continuous ensemble methods using simulated and benchmark datasets.

## Tree-Based Integration of One-versus-Some Classifiers for Multiclass Classification

❖ Yuejing Ding, Columbia University, Room 1005, MC 4690, 1255 Amsterdam Avenue, New York, NY 10027, *yd2122@columbia.edu*; Tian Zheng, Columbia University

***Key Words:*** tree-based classification, multi-class classification, gene expression analysis

In gene expression analysis, the classification of two classes (e.g., cancer versus normal) has achieved some success. When comes to multi-class classification, the extension is not straightforward. There has been two main directions: 1) hierarchical (or tree-based) classifiers and 2) integrated votes from one-versus-all (OVA) or pairwise classifiers. In this talk, we present a new classification method based on tree-integrated one-versus-some classifiers, in an attempt to combine advantages from both strategies. An application to a microarray data set will also be discussed.

## Clustering Genes in Genetical Genomics Experiments

❖ Joshua Sampson, University of Washington, Department of Biostatistics, CAMPUS MAIL BOX 357232, SEATTLE, WA 981957475, *sampsonj@u.washington.edu*; Steve Self, University of Washington

***Key Words:*** expression, QTL, clustering, pleiotropy

Traditionally, genes are clustered into groups that share a similar pattern of expression over time or under changing experimental conditions. Now, due to the emergence of genetical genomics, gene expression is also measured over varying genetic backgrounds. Standard genomics' clustering methods group genes based solely on expression profiles, ignoring linkage information. In contrast, other methods for genetical genomics treat expression measurements as quantitative traits in a linkage study. Therefore, we propose a new clustering method that groups genes together when they share a common set of QTLs.

## A Divisive Method via Multivariate Hypothesis Testing for Clustering Gene Expression Patterns

❖ Haiyan Wang, Kansas State University, 108 E. Dickens Hall, Dept of Statistics, Manhattan, KS 66506, *hwang@ksu.edu*

***Key Words:*** longitudinal data, high dimensional data, clustering, hypothesis testing, microarray data, time course data

Linear method to cluster genes based on expression patterns often need data preprocessing to achieve reasonable results though data preprocessing could actually corrupt the original data and introduce spurious temporal behavior (Taguchi & Oono 2005). Here we present a divisive method to cluster genes according to expression patterns. The measure of dissimilarity between groups is given by the p value from a nonparametric multivariate hypothesis testing for no group differences all at time points. The test statistic uses overall ranks of expressions so that the analysis is invariant to monotone transformations of data. Asymptotic distribution of the test statistic is obtained and used to obtain the p values. Simulation shows that the procedure can extract significant patterns with low error without any supervision or preprocessing even in case of small number of observations per cycle.

## On Comparing the Clustering of Regression Models Method with K-Means Clustering

❖ Li-Xuan Qin, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street, 3rd Floor, New York, NY 10021, *qinl@mskcc.org*; Steve Self, University of Washington

***Key Words:*** clustering, regression, microarray, gene clustering

Gene clustering is a common question addressed with microarray data. Previous methods, such as K-means clustering and the multivariate normal mixture model, base the clustering directly on the observed measurements. The clustering of regression models (CORM) method (Qin and Self 2006) bases the clustering of genes on their relationship to covariates and explicitly models different sources of variations. In this paper we discuss connections and differences between CORM and K-means clustering. We show that CORM tends to seek a partition of genes that has stable cluster centers across samples. Simulation results show that CORM outperforms K-means and an extended K-means when the assumed regression model is true and is robust to certain model misspecifications. We also use a microarray dataset to demonstrate a scenario where only CORM is appropriate.

# 383 Clustering

Section on Statistical Computing, Section on Nonparametric Statistics
**Wednesday, August 9, 8:30 am–10:20 am**

## Clustering by Intersection-Merging

❖ Qunhua Li, University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195-4322, *newstatistician@yahoo.com*; Marina Meila, University of Washington

***Key Words:*** model-based clustering, mixture model, expectation-maximization, hierarchical agglomerative clustering, simulation annealing

We propose Intersection-Merging (IM), a wrapper algorithm to improve the initial clusterings in model-based clustering. The algorithm takes a set of clusterings obtained e.g. by EM, breaks down the clusterings into subclusterings via an intersection step, and then agglomerates them via a merging step. We introduce two versions of merging: greedy (standard IM) and by simulated annealing (IMSA). We apply the algorithm to the Gaussian mixtures of several synthetic and real data sets. The results show that both IM and IMSA improve on the starting clusterings under a variety of criteria.

## Strategies for Scaling and Weighting Variables in Cluster Analysis

❖ Srinivas Maloor, Rutgers University, 23817 BPO WAY, Piscataway, NJ 08854, *srinigotmail@gmail.com*; Ramanathan Gnanadesikan, Rutgers University; Jon Kettenring, Drew University

***Key Words:*** data analysis, variable scaling, variable weighting, discriminant analysis

Scaling of multivariate data (to put all variables on an “equal footing”) prior to cluster analysis is important as a preprocessing step. Widely-used current methods for this are problematic.In this paper, we propose some intuitive alternatives, which we call “equalizers”, that are directed at avoiding a basic pitfall in the usual methods.In addition, we let the data suggest weights or “highlighters”, that emphasize those variables with most promise for revealing the latent cluster structure. The equalizing and highlighting schemes vary in degree of complexity from very simple weights based on order statistics to more complicated iterative ones. The methods are applied to simulated and real data sets, and their performances are compared to some currently used methods. The results indicate that, in many situations, the new methods are much better than the most popular method, autoscaling.

## Model-Based Projection Pursuit Clustering

❖ Jie Ding, GlaxoSmithKline, 709 Swedeland Road, UE0393, King of Prussia, PA 19406, *Jie.2.Ding@gsk.com*

***Key Words:*** model-based clustering, projection-pursuit, high-dimensional data

Clustering methods assign observations to clusters based on the measured characteristics of each observation. In high-dimensional space, the true structure of clustering sometimes is contained in a low-dimensional subspace, the remaining dimensions contain little or no information about the clusters. Clustering methods that use all variables are often confounded by statistical noise in the dimensions that are unrelated to the cluster model. It is shown that a dimension-reduction technique, known as projection-pursuit, can increase the performance of the model-based clustering methods. A model-based projection-pursuit clustering methodology is introduced, which combines model-based clustering with projection-pursuit. In addition, the model-based projection-pursuit clustering provides low-dimensional pictorial representations of the clustering contained in the high-dimensional data.

## Estimating the Number of Data Clusters via Agreement Measure--Based Statistics

❖ Heng Liu, University of Illinois at Urbana-Champaign, Illini Hall, 725 S Wright Street, Champagin, IL 61820, *hengliu@uiuc.edu*; Michelle Wang, University of Illinois at Urbana-Champaign; Douglas Simpson, University of Illinois at Urbana-Champaign

***Key Words:*** kappa statistics, fMRI time series, clustering, naive Bayes rule

Clustering and classification have been important tools to address a broad range of problems in fields such as image analysis, genomics, and other areas. Many different heuristic clustering criteria are available. Often the data partitioning suffers lack of consistency across different algorithms and criteria. In this paper we propose a Kappa-type of agreement measure to compare and combine results from different clustering methods, and to select the number of clusters. By maximizing the agreement on allocation of the observations between different clustering methods, the approach is intended to provide a robust consensus set of clusters. The favorable performance of the method is demonstrated in simulation studies and for real fMRI time series data. Finally the asymptotic properties of the Kappa statistics are discussed.

## Clustering of Longitudinal Data: a Functional Data Approach

❖ Jeng-Min Chiou, Academia Sinica, Institute of Statistical Science, 128 Academia Road Sec 2, Taipei, 11529 Taiwan, *jmchiou@stat.sinica.edu.tw*; Pai-Ling Li, National Tsing Hua University

***Key Words:*** clustering, functional data, functional principal component analysis, modes of variation, stochastic processes

A functional clustering method for longitudinal data is proposed. This approach accounts for both the means and modes of variation differentials among clusters. The cluster membership predictions are based on a nonparametric random effect model of the truncated Karhunen-Loeve expansion, coupled with a nonparametric iterative mean and covariance updating scheme. We show that under the derived identifiability conditions, the proposed clustering method can greatly improve cluster quality as compared to conventional clustering algorithms. Moreover, by exploring the mean and covariance functions of each cluster, the proposed method provides an additional insight into cluster structures, which facilitates functional cluster analysis. Practical performance of the proposed method is demonstrated through simulation studies and data applications.

## Initializing Optimization Partition Algorithms

❖ Ranjan Maitra, Iowa State University, Department of Statistics, Ames, IA 50011-1210, *maitra@iastate.edu*

***Key Words:*** clustering, k-means, E-M algorithm, singular value decomposition

Clustering datasets is a challenging problem in general, but needed in a wide array of applications. A large number of approaches exist, most of which can be broadly grouped into either the optimization partitioning or the hierarchical clustering class of algorithms. Common examples of the partitioning approaches are the iterative k-means and the expectation-maximization (EM) algorithms. Such algorithms are sub-optimal for multi-dimensional data and find local optima in the vicinity of their initialization. I propose a staged approach for finding starting values.

Results on test experiments indicate excellent performance. Applications to clustering mercury emissions data and bioinformatics are presented.

### A Scale-Independent Clustering Method with Automatic Variable Selection Based on Trees

❖ Samuel Buttrey, Naval Postgraduate School, Code OR Sb, Naval Postgraduate School, Monterey, CA 93950, *buttrey@nps.edu*

***Key Words:*** cluster quality, classification and regression trees, prediction strength

Clustering techniques usually rely on measurements of distances (or dissimilarities) among observations and clusters. These distances are often affected by variables' scaling or transformation, and do not provide for selection of "important" variables. We fit a set of regression or classification trees; each variable acts in turn as the response variable. Points are "close" to one another if they tend to appear in the same leaves of these trees. Trees with poor predictive power are discarded. "Noise" variables which appear in none of the trees have no effect on the clustering and can be ignored. The clustering is unaffected by linear transformations of the continuous variables and resistant to monotonic ones. Categorical variables are included automatically. We demonstrate the technique on well-known noisy data sets. This paper updates an idea proposed at JSM 2004.

## 384 Examples for the Statistics Classroom ●

Section on Statistical Education
**Wednesday, August 9, 8:30 am–10:20 am**

### Making Babies by the Flip of a Coin?

❖ Matthew Carlton, California Polytechnic State University, San Luis Obispo, Department of Statistics, San Luis Obispo, CA 93407-0405, *mcarlton@calpoly.edu*

***Key Words:*** probability, binomial model, chi-square test, genetics, gender

Probabilists (and some geneticists) treat gender of children like flips of a fair coin: two possible outcomes, each equally likely, with outcomes independent from trial to trial. But are these assumptions realistic? We will look at data from the National Center for Health Statistics to find out. This talk parallels our 2005 Teacher's Corner article on this topic, with an interesting follow up on the cited federal data.

### Illuminating the Confidence Interval Concept Activity

❖ Alicia Graziosi, Temple University, 12 Yale Road, Marlton, NJ 08053, *alicia.graziosi@temple.edu*; Jeffrey Lidicker, Temple University

***Key Words:*** confidence interval, sample mean, population mean, activity

One of the most difficult concepts to teach in an introductory statistics course is how to interpret a confidence interval, particularly the concept that a confidence interval may not include the true population mean as determined by the sample. To this end, we propose a fun classroom activity. This activity can be used when teaching Z or T distributions and proportional confidence intervals. At the end of this activity, an illustrative graph is created using confidence intervals calculated by each student. Discussion of this graph will demonstrate the dependency of a confidence interval on its sample; a confidence interval is the probability a given interval contains the mean (and not that the probability that the mean is contained in the interval), the intervals change, the mean does not, and correct calculations do not always produce accurate intervals.

### Learning Activities for Large Classes

❖ Patti Collings, Brigham Young University, 231 TMCB BYU, Provo, UT 84602, *collingp@byu.edu*

***Key Words:*** introductory statistics, activities, sampling distribution, statistical concepts

Ever wondered if you could do activities in classrooms of 200 students or more? This paper presents activities we have used in our introductory statistics course at Brigham Young University. These activities are used to help students learn difficult statistical concepts such as sampling distribution, level of confidence and statistical significance. They are specifically designed for large classes. We will share our experience of what works and what doesn't work. We will also give feedback from our students.

### How Low Can You Go?

❖ Paul Stephenson, Grand Valley State University, Department of Statistics, 1 Campus Drive, Allendale, MI 49401, *stephenp@gvsu.edu*; Mary Richardson, Grand Valley State University; John Gabrosek, Grand Valley State University

***Key Words:*** simulation, sampling distributions, order statistics, statistics education, sports, gaming theory

GOLO is a simple and addicting dice game that replicates golf using nine twelve-sided dice. The basic rules of GOLO are to roll all nine dice and remove your lowest score(s). You may remove as many dice as you wish, but you must remove at least one die each roll. Place the remaining dice in the cup and roll again. Continue to roll until all nine dice are removed and add up the score for the nine dice (which represents the golfer's nine hole score). In this talk the authors will briefly introduce the game of GOLO and demonstrate how simulation can be utilized to replicate the playing of the game. A variety of game strategies will be proposed and the authors will examine the performance of each game strategy by investigating the distribution of the resulting scores. The authors will then discuss how GOLO can be used in both introductory and advanced undergraduate statistics courses.

### Conditional Probability and "Who Wants To Be a Millionaire?"

❖ Diane Evans, Rose-Hulman Institute of Technology, 7248 Belfonte Lane, Terre Haute, IN 47802, *diane.evans@rose-hulman.edu*

***Key Words:*** conditional probability, teaching activity

In statistics classes, I never pass up the opportunity to share with students the Monty Hall problem when we are discussing probability. Besides the lively interaction of students proposing and challenging various arguments for (or against) the answer, some students begin to reconsider problems where the Monty Hall "dilemma" may have been hidden beneath their surfaces. Since "Who Wants to Be a Millionaire?" continues to be a popular game show, one question that has been raised in class is "Does the contestant's probability of choosing the correct answer change after the elimination of two incorrect answers by the 50:50 lifeline option?" This problem is of interest to students because it involves a game show where probabilistic choices can help determine a contestant's winnings. This talk will look at calculating these conditional probabilities as a classroom activity.

### Post-Hoc Analysis for a Class of Chi-Square Tests

❖ Edward Markowski, Old Dominion University, IT Decision Sciences Department, Norfolk, VA 23529, *emarkows@odu.edu*; Carol A. Markowski, Old Dominion University

***Key Words:*** introductory statistics, chi-square tests, post-hoc analysis

Chi-square tests such as the test of independence are standard methods in introductory statistics texts. When the null hypothesis is rejected, the conclusions of such tests are too generic to fully identify the sources of relationships identified. Authors of introductory texts address this issue in different ways and with varying sophistication. In this paper we point out several weaknesses of these common approaches and propose an alternative approach to further analysis of significant chi-square results. This analysis uses the components of the chi-square statistic in identifying critical cells for interpretation. In addition, it uses subsequent chi-square tests to explicitly determine the appropriate degree of investigation of cell comparisons.

### Keeping an Introductory Statistics Course Interesting: Use of Demonstrations, Examples, Rewards, and a Little Humor

❖ Harry Norton, Carolinas Medical Center, 5500 Lansing Drive, Charlotte, NC 28270, *jnorton@carolinas.org*

***Key Words:*** teaching, statistical education, demonstrations, examples

Keeping undergraduate biology students and medical residents interested in statistics, when the majority of is taking the class as a requirement, can be challenging. I have attempted to accomplish this goal by the use of demonstrations, interesting examples, rewards, and a few related jokes and cartoons. I will present examples of techniques I use in my course instruction.

## 385 Bayesian Design and High Dimensional Inference

Section on Bayesian Statistical Science

**Wednesday, August 9, 8:30 am–10:20 am**

### Automatic Estimation of Multivariate Spectra via Smoothing Splines

❖ Ori Rosen, The University of Texas at El Paso, Department of Mathematical Sciences, Bell Hall 221, El Paso, TX 79968, *ori@math.utep.edu*; David Stoffer, University of Pittsburgh

***Key Words:*** spectral matrix, Cholesky decomposition, MCMC

The classical method for estimating the spectral density of a multivariate time series is to first calculate the periodogram, then smooth it to obtain a consistent estimator. We suggest a Bayesian approach that uses Markov chain Monte Carlo techniques to fit smoothing splines to each component, real and imaginary, of the Cholesky decomposition of the periodogram matrix. The spectral estimate is then obtained by reconstructing the spectral estimator from the smoothed Cholesky decomposition components. Our technique allows for automatic smoothing of the components of the spectral density matrix. We illustrate our methodology with data on the Southern Oscillation Index, as well as with a DNA sequence.

### Bayesian Curve Estimation with Overcomplete Wavelet Dictionary

❖ Jen-hwa Chu, Duke University, Box 90251, Durham, NC 27708, *jenhwa@stat.duke.edu*; Merlise Clyde, Duke University; Feng Liang, Duke University

***Key Words:*** overcomplete representation, Bayesian inference, reversible jump Markov chain Monte Carlo, wavelet transform

We describe a Bayesian approach for curve estimation based on an overcomplete wavelet dictionary proposed by Abramovich, Sapatinas and Silverman (1999), where the function is modeled by a sum of the wavelet components at arbitrary locations and scales, on which prior distributions are imposed. We believe that by avoiding the dyadic constraints for orthonormal wavelet bases, the overcomplete wavelet dictionary will have greater flexibility to match the structure of the data, and give sparser representations. The main challenge here is to efficiently search over an infinite number of basis elements. We will discuss a reversible jump Markov Chain Monte Carlo algorithm for inference on the parameters, and strategies to give better convergence. We will also present simulation results using standard test functions, which show that the new method leads to greater sparsity and improved MSE.

### Bayesian LAGO for Statistical Detection Problems

❖ Wanhua Su, University of Waterloo, Apartment 708, 383 Albert Street, Waterloo, ON N2L 6E3 Canada, *wsu@uwaterloo.ca*; Mu Zhu, University of Waterloo; Hugh A. Chipman, Acadia University

***Key Words:*** average precision, k-nearest neighbor (KNN), support vector machine (SVM), locally adaptive go (LAGO) estimate, Markov chain Monte Carlo (MCMC), Laplace approximation

In a typical statistical detection problem, we have data {Yi,Xi}, where Xi is the predictor vector of the ith observation and Yi=0\1 is its class label. The objective of a statistical detection problem is to identify class-1 observations, which are extremely rare. Our previous work (LAGO) was demonstrated to be superior to methods such as KNN and SVM in a real drug discovery problem. One drawback of the existing LAGO is that it only provides a point estimate of a test point's possibility of being class 1, ignoring the uncertainty of the model. In this talk, we present a probabilistic framework for LAGO, referred as BLAGO. Uncertainty can be accommodated by assigning prior distributions to the parameters. MCMC is used to obtain the posterior distributions. BLAGO provides proper probabilistic predictions that have support on (0,1) and captures uncertainty of the predictions as well.

### Avoiding Bias from Feature Selection in Classification and Regression Models

❖ Longhai Li, University of Toronto, Department of Statistics, 100 St George Street - 6th Floor, Toronto, ON M5S3G3 Canada, *longhai@stat.toronto.edu*; Jianguo Zhang, University of Toronto; Radford Neal, University of Toronto

***Key Words:*** feature selection bias, optimistic, mixture models, factor analysis, gene expression, naive Bayes models

For many classification and regression problems, a large number of features are available for possible use. Often, for computational or other reasons, only a small subset of these features are selected for use in a model, based on some simple measure such as correlation with the response variable. This procedure may introduce an optimistic bias. We show how this bias can be avoided when using a Bayesian model for the joint distribution of features and response. The crucial insight is that we should retain the knowledge that the discarded features correlate with the response weakly. We describe implementations for naive Bayes models of real and binary data, and for factor analysis and two-component mixture models. Experiments with artificial data confirm that this method avoids bias due to the feature selection. We also apply these models to actual gene expression data.

### Dose-Schedule Finding in Phase I/II Clinical Trials Using Bayesian Isotonic Transformation

❖ Yisheng Li, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *ysli@mdanderson.org*; Nebiyou B. Bekele, M. D. Anderson Cancer Center; Yuan Ji, M. D. Anderson Cancer Center; John Cook, M. D. Anderson Cancer Center

***Key Words:*** dose-schedule finding, clinical trial, Bayesian design, phase I/II, isotonic transformation, partially-ordered constraints

Most phase I oncology trials aim to locate the MTD of an experimental treatment. One of the main considerations apart from determining the MTD is determining an appropriate schedule for administration of the treatment. Historically, schedules have been fixed prior to the start of dose finding. Recently, an increasing number of trials have been designed to determine a dose AND a schedule. We propose a Bayesian design for dose-schedule finding by jointly modeling the binary toxicity and efficacy outcomes. Assuming ordered toxicity probabilities between schedules, we apply an isotonic transformation approach to estimating the constrained parameters. We select dose-schedule combination based on the joint posterior distribution of efficacy and toxicity. Using a simulation study and a real cancer clinical trial, we examine the operating characteristics of the proposed design.

### Bayesian Two Stage Optimal Design for Generalized Linear Models

❖ Ying Zhang, Virginia Polytechnic Institute and State University, 1224 University City Blvd., Apt D47, Blacksburg, VA 24060, *yizhang9@vt.edu*

***Key Words:*** optimal design, Bayesian, two stage, generalized linear model

Bayesian optimal designs have received increasing attention in recent years. Chaloner and Larntz(1989) investigated the logistic model using Nelder-Mead algorithm. However, since Nelder-Mead algorithm is a local-optimization method, the choice of starting design points has great influence on the performance of the procedure in getting to the global optimum. We propose a hybrid computational method to find the Bayesian optimal design efficiently even for multi-variable generalized linear models. More specially, logistic and Poisson models are studied. Furthermore, the Bayesian two stage optimal designs are developed by implementing the hybrid computational method for generalized linear models, especially logistic and Poisson models. The Bayesian two stage optimal design approach is better in finding the optimal design points based on the data information observed in first stage responses.

## 386 Bayesian Applications to Genetics ● ✪

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Wednesday, August 9, 8:30 am–10:20 am**

### The Evolutionary Forest Algorithm

❖ Scotland Leman, Duke University, ISDS, 27707, *scotland@stat.duke.edu*

***Key Words:*** speciation, divergence, MCMC, Bayesian, data augmentation

In this application, we address the demographic history of closely related members of the Drosophila pseudoobscura group. We estimate the joint posterior distribution of the time since speciation, backward migration rates and effective population sizes of the extant and ancestral populations. We base estimation on the multi-locus regions of DPS2002 and Rhodopsin 1 simultaneously linked by a unique speciation event. Summary statistics, rather than entire nucleotide sequences, permits a compact description of the genealogy of the sample. Consideration of an augmented random variable, comprised of multiple genealogies or forest, dramatically increases sampling efficiency. This method of using a forest of genealogical histories converges marginally to the posterior distributions of interest and convergence is improved on compared to methods in which a single history is considered.

### Bayesian Inference for Estimating Migration Rate, Mutation Rate, and Population Size in Microsatellite Loci

❖ Seongho Song, University of Cincinnati, Department of Mathematical Sciences, 839 Old Chem, Cincinnati, OH 45221-0025, *seongho.song@uc.edu*; Dipak Dey, University of Connecticut; Kent E. Holsinger, University of Connecticut

***Key Words:*** multivariate logistic transformation, genetic diversity, microsatellite loci, migration rate, mutation rate, population size

We propose various plausible models for the allele frequency distribution having a specific mean and covariance structure studied in Song et al(2005). Several models are proposed to estimate evolutionary parameters including migration (m), mutation rate (mu) and population size(N). First, we suggest the method of moments approaches to match population covariances to sample covariance, as an ad hoc approach, and propose the statistical model using truncated multivariate Normal distribution and multivariate Logistic Normal distribution, respectively, with specific covariance structures within and among populations. We also illustrate the prior information about the evolutionary parameters which are migration (m), mutation rate (mu) and population size (N). As an application, microsatellite human data will be discussed.

### Bayesian Discovery of Regulatory Motifs using Reversible Jump Markov Chain Monte Carlo

❖ Min Li, University of Washington, Department of Biostatistics, University of Washington, Seattle, WA 98195, *sierrali@u.washington.edu*

***Key Words:*** gene regulation, motif discovery, Bayesian hierarchical model, dimensional change, reversible jump Markov chain Monte Carlo

De Novo identification of transcription factor binding sites (TFBSs) is one of the most important problems in computational biology, as it provides valuable information towards elucidating gene regulatory networks. We focus on the search for repetitive but not identical patterns, called motifs, overpresented in the sequence data. A statistical approach requires a model for motif compositions, sites locations, motif widths, the number of motifs and the number of motif sites. This article proposes a novel Bayesian hierarchical model to incorporate all unknown features in motif discovery; inference over all unknown factors is associated with dimensional changes of the parameter space. Reversible jump Markov chain Monte Carlo allows parameter spaces with different dimensions to be explored. We apply out model to a set of upstream sequences of 20 co-expressed Saccharomyces cerevisiae genes.

### Detection of Quantitative Trait Loci on Multiple Linkage Groups

❖ Patrick Gaffney, ImClone Systems, 22 ImClone Drive, Branchburg, NJ 08876, *PatrickGaffney@imclone.com*

***Key Words:*** quantitative trait loci, QTL, RJ-MCMC, reversible jump MCMC

This paper examines the problem of determining the number and putative location of quantitative trait loci (QTL) on multiple linkage groups. A method for model selection using the Bayes factor is outlined. This method is shown via simulation to have good frequentist properties across a large range of sample sizes and number of linkage groups. Reversible jump MCMC is used to estimate the Bayes factor. Convergence of the sampler is improved by permitting the QTL to move to another linkage group during the update step.

### Bayesian Clustering of SNP Genotypes

❖ Guohua Yan, The University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC V6T 1Z2 Canada, *gyan@stat.ubc.ca*; William J. Welch, The University of British Columbia; Ruben H. Zamar, The University of British Columbia

***Key Words:*** Bayesian, clustering, single nucleotide polymorphisms

The TaqMan assay is a popular high-throughput genotyping technology for SNPs (single nucleotide polymorphisms). Clusters of genotypes are often identified by algorithms applied to normalized 2-dimensional data. We are investigating working with un-normalized data and model-based approaches. Methods based on clusters around lines seem appropriate by looking at the data. We use a Bayesian approach to incorporate prior information about the lines locations. This work is joint with Professors William Welch and Ruben Zamar.

### Bayesian Clustering of Short Temporal Gene Expression Dynamics

❖ Ling Wang, Boston University, 401 Broadway, Apt 35, Cambridge, MA 02139, *wangling@bu.edu*; Paola Sebastiani, Boston University; Marco Ramoni, Harvard Medical School

***Key Words:*** Bayesian model, invariance, polynomial model, caged

We present an extension of CAGED (Clustering Analysis of Gene Expression Dynamics) to cluster gene expression profiles measured in short temporal/dose-response microarray experiments. In contrast to the initial version of CAGED, in which temporal expression profiles are modeled by autoregressive equations, our new algorithm describes the trend using polynomial models of time/dosage. Our Bayesian approach uses proper conjugate priors for the model parameters so that the algorithm is invariant to linear reparameterizations of time/dosage. We compare our approach with the recently introduced program STEM (Short Time-series Expression Miner) to show that our method can find the correct number of clusters and allocate gene expression profiles to the right clusters in simulation studies, and produce more biologically meaningful Gene Ontology enriched clusters in real dataset.

### A Bayesian Retrospective Classification Model

❖ Jingqin Luo, Duke University, ISDS Box 90251, Durham, NC 27708-0251, *rosy@stat.duke.edu*

***Key Words:*** Bayesian statistics, microarray data, high-dimensional analysis, retrospective model

High-dimensional Microarray data are more and more relied on in cancer study to classify subjects. While modeling feature space of high dimensionality has been a challenging issue, most statistical methods proceed statistical analysis under simplified assumptions. A practically successful method, naive Bayes classifier (NB) assumes that features are conditionally independent given class label. Besides accurate classification, exploration of feature dependency network and discovery of important features are also desirable research interests. We develop a Bayesian compositional retrospective model to achieve this goal, which slightly relaxes NB's simple assumption. Variable selection and model selection are integral components of our method. Simulation studies have been carried out to compare our method with several state-of-art methods. Application to real cancer dataset is illustrated.

## 387 Sample Survey Quality IV ●

Section on Survey Research Methods
**Wednesday, August 9, 8:30 am–10:20 am**

### Changing to Register-Based Statistics

❖ Anders Wallgren, Statistics Sweden, SE 701 89 Orebro, Orebro, SE-701 89 Sweden, *anders.wallgren@scb.se*; Britt Wallgren, Statistics Sweden

***Key Words:*** administrative data, system of statistical registers, register-based surveys

Many countries are investigating the possibility of using more administrative data for statistical purposes. It is necessary to reduce response burden and costs; increasing nonresponse in censuses and sample surveys also make this new strategy necessary. A new approach is necessary, and register-based surveys require that suitable statistical methods be developed. We have studied the requirements for register-based statistics through analysis of Statistics Sweden's system of statistical registers. Our general conclusion is that administrative data should not be used as they are. The register population should be created by using many sources to achieve good coverage. Derived variables of good measurement quality are created after editing where many sources are combined. Statistical registers also should be made integrable so data can be combined in new ways to satisfy new needs.

### Longitudinal Evaluation of Point and Variance Estimates in an Establishment Survey after Ratio Imputation

❖ Adriana Perez, The University of Texas Health Science Center at Houston, 80 Fort Brown SPH RAHC, Rm 200, Brownsville, TX 78520, *adriana.perez@uth.tmc.edu*

***Key Words:*** total estimate, variance estimation, establishment data, nonresponse, simulations

The US survey of Graduate students and postdoctorates in Science and Engineering is an annually cross-sectional establishment survey. Each year, a ratio imputation technique is used in combination with previous year's data to handle item-nonresponse. Reports are presented in a yearly longitudinal structure without accounting for its imputation uncertainty. A complete pseudo universe with data from 1998-2002 is used, including students enrolled as full-time, part-time, part-time women and part-time men. Bias and mean square errors from simulations were summarized to evaluate this imputation method. Sensitivity analyses describing departures from current observed percentages of missingness are presented. This imputation method provides bias estimates of the total and overestimation of its associated variance. The higher the percentage of missingness the larger the bias.

### Experimental Design for the 2006 American Community Survey Content Test

❖ Mark Asiala, U.S. Census Bureau, 4215 Sweetman Road, Marbury, MD 20658-2120, *mark.e.asiala@census.gov*; Alfredo Navarro, U.S. Census Bureau

***Key Words:*** content test, experimental design, split panel

The 2006 American Community Survey (ACS) Content Test is a large sample survey (over 60,000) designed to test proposed changes to the ACS questionnaire for 2008. The original base sample design is a stratified primary sample unit design that builds on the sample design of an earlier survey, the Census 2000 Supplementary Survey. In response to changing field requirements, several changes were implemented to address total workload size and other operational issues. Additional stratification and sampling for the personal follow-up phase of the survey was added late in the design phase to help reduce the total workload. A balanced split panel assignment was designed to help minimize the number of field workers necessary for the survey in the less densely populated areas avoiding duplication of geographic assignment areas for experimental and control panel FR's.

### Comparison of Key from Paper and Automated Data Capture in the American Community Survey

❖ Anthony Tersine, U.S. Census Bureau, 7779 Newington Woods Drive, Springfield, VA 22153, *anthony.g.tersine.jr@census.gov*; Donald Keathley, U.S. Census Bureau

***Key Words:*** data capture, American Community Survey

The American Community Survey (ACS) is an ongoing monthly survey of the population in the United States. Each month the ACS mails its questionnaires to a random sample of households; the survey uses telephone and personal visit contact attempts in subsequent months to try to obtain interviews from those whose questionnaire was not received in mail. All mail-returned questionnaires go through a data capture process, the ACS currently uses a key-from-paper (KFP) method for this operation. To make the survey operation more efficient we are studying an automated data capture system. This paper will compare the current KFP method of capturing data to an automated data capture system. In particular, this paper will attempt to answer the following question: can the automated data capture system maintain the same or improve on the level of data quality when compared to the KFP method?

### A Comparison of Two Ratio Edit Methods for the Annual Survey of Government Finances

❖ Elizabeth Cornett, U.S. Census Bureau, , *elizabeth.d.cornett@census.gov*; Joanna F. McLaughlin, U.S. Census Bureau; Carma R. Hogue, U.S. Census Bureau; Stephen D. Owens, U.S. Census Bureau

***Key Words:*** ratio edits, resistant fences, Hidiroglou-Berthelot edit

The Annual Survey of Government Finances (AFS) is conducted yearly to collect data on revenue, expenditures, debt, and assets of state and local governments. The questionnaire was redesigned for 2005. As a result, some data items were consolidated and others were split into multiple data items. These changes required a redesign of the ratio edits and the methods for establishing the ratio edit bounds. Historical edits were no longer possible. New current year edits had to be determined. We researched two methods for editing AFS data at the unit level: resistant fences and Hidiroglou-Berthelot (HB). The first is an editing technique that is designed to work well with different distributions. The second technique focuses on small changes in large units versus large changes in small units. In this paper, we compare these methods for the AFS.

### Evaluation of the Detectability and Inferential Impact of Nonresponse Bias in Establishment Surveys

❖ Randall Powers, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212, *powers_r@bls.gov*; John L. Eltinge, Bureau of Labor Statistics

***Key Words:*** attrition, incomplete data, item nonresponse, missing data, sensitivity analysis, simulation study

In construction of estimators from survey data, one often encounters important issues arising from nonresponse. For establishment surveys, methods to address these issues generally must account for important features of the sample design and weighting structure. For any given nonresponse adjustment procedure, an analyst makes implicit or explicit use of models for the nonresponse phenomenon and the outcome variables of primary interest. The performance of the adjustment procedure then depends on the extent to which the data deviate from the assumed models, the impact of these deviations on estimator bias, and the inferential power of diagnostics designed to detect these deviations. This paper presents a simulation study to evaluate trade-offs among the issues of model deviations, estimator performance and detectability for establishment surveys.

### A Coverage Profile of Area Frame Blocks on the United States Census Bureau's Master Address File

❖ Timothy Kennel, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-0001, *Timothy.L.Kennel@census.gov*

***Key Words:*** coverage, sample redesign, frame creation, master address file, current population survey, area frame listing

With the availability of national list frames, the future of area frame listing in the United States is uncertain. This paper uses a dependent area frame listing to explore characteristics of gross undercoverage and overcoverage on a list frame, the U.S. Census Bureau's Master Address File (MAF). The results presented show that undercoverage is not missing at random and can bias key statistics. Furthermore, areas where targeted listings can be used to improve coverage errors are presented.

## 388 Cluster Modeling and Cluster Detection ● ✪

Section on Statistics in Epidemiology

**Wednesday, August 9, 8:30 am–10:20 am**

### P-Values for the Besag-Newell Cluster Detection Test

❖ Ronald Gangnon, University of Wisconsin-Madison, 603 WARF, 610 N Walnut St, Madison, WI 53726, *ronald@biostat.wisc.edu*

***Key Words:*** cluster detection, likelihood principle, p-value adjustment

Besag and Newell (1991) proposed a focused cluster detection method based on the number of nearest neighbor geographic regions, l, that need to be aggregated to include a prespecified minimum number of cases, k. Besag and Newell calculated their p-value based solely on the number of regions included in the cluster (l), ignoring the actual number of cases (y) included in the identified cluster. We consider several methods for computing the p-value for the cluster identified by the

Besag-Newell procedure based on different orderings of the bivariate outcome space (y,l). We compare the methods with respect to the degree to which each captures the strength of evidence against the null hypothesis implied by the data. We conclude that the ordering of the sample space induced by the observed likelihood ratio statistic has the most desirable operating characteristics.

## On Detecting a Rate Increase Using a Bernoulli-Based Scan Statistic

❖ Michael Joner, Virginia Polytechnic Institute and State University, 406A Hutcheson Hall, Mail Code 0439, Blacksburg, VA 24061-0439, *mjoner@vt.edu*; William H. Woodall, Virginia Polytechnic Institute and State University; Marion Reynolds, Virginia Polytechnic Institute and State University

***Key Words:*** scan statistic, Bernoulli, cusum, surveillance, prospective monitoring, statistical quality control

Scan statistics are often used by public health specialists to detect disease clusters or increased incidence rates. Such events are signaled if there are an unusually large number of events close in time and/or space. Most of the work has been for the retrospective case, in which a single set of historical data are to be analyzed. In the temporal case it has been recommended that the scan statistic be used in a prospective manner to determine whether the underlying incidence rate has increased, preferably as soon as possible after such an increase. We evaluate the average number of times that the scan statistic must be calculated for the Bernoulli-process case in order to generate a signal that the incidence rate has increased. We show that the Bernoulli cumulative sum method can have much better performance than the scan-based methods.

## Spatial Survival Clusters of Patients Diagnosed with Lung Cancer and Late-Stage Colorectal Cancer in California

❖ Lan Huang, National Cancer Institute, 6116 Executive Blvd., Room 5043, Rockville, MD 20852, *huangla@mail.nih.gov*

***Key Words:*** spatial scan statistic, exponential model, survival cluster detection, socioeconomic status, health disparities, GIS

We develop the first detailed illustration of the use of a cluster detection method utilizing a spatial scan statistic based on an exponential survival model, applying it to study the spatial patterns of survival of cancer patients with stage III, and stage IV colorectal cancer and stage I/II, stage III, and stage IV lung cancer in the state of California and the county of Los Angeles diagnosed during 1988-2002. We compute nonparametric estimates of survival inside and outside of those detected clusters, which confirms the survival pattern detected by the spatial scan statistic in both areas. In LA County, we investigate the possible relationship between the detected clusters and demographic factors, such as race, gender and histology, and socioeconomic factors, such as education, employment, income and health insurance using nonparametric methods and logistic regression models.

## A Latent Model for Highly Skewed and Grouped Data

❖ Huichao Chen, Emory University, Department of Biostatistics, 1518 Clifton Road, Atlanta, GA 30322, *hchen4@sph.emory.edu*; Amita K. Manatunga, Emory University; Robert Lyles, Emory University; Michele Marcus, Emory University

***Key Words:*** grouped data, skewed, one-sided likelihood ratio test, Michigan female health study

Data arising from reproductive epidemiological studies often present analytic challenges. To model the non-smooth distribution of exposure consisting of heaps and coarse values, we propose a general latent model for highly skewed and grouped data. We assume that the observed exposure is determined by the value of an underlying unobservable continuous response that follows a Weibull distribution. To accommodate correlations in repeated true latent responses, we introduce general random effects from the power variance function (PVF) family of distributions (Hougaard, 2000). The resulting marginal likelihood has a closed form. The performance of the proposed model is supported by simulation studies and its application is illustrated using repeated polybrominated biphenyl (PBB) exposure data on participants of the Michigan Female Health Study.

## Local Likelihood Models for Disease Cluster Modeling: a Space-Time Extension

❖ Monir Hossain, University of South Carolina, 800 Sumter Street, Dept of Epidemiology and Biostatistics, Columbia, SC 29208, *hossain@gwm.sc.edu*; Andrew B. Lawson, University of South Carolina

***Key Words:*** local likelihood, space-time modeling, random effects, Bayesian hierarchical modeling

Data-dependent clustering (Lawson 2006 and Hossain and Lawson 2005) assumed that the relative risk of a disease is defined to be a function of the `local' concentration of cases. This is the idea of local likelihood and the model based on local likelihood has found a novel application to disease cluster modeling. This paper extends the spatial local likelihood model (Hossain and Lawson, 2005) to spatio-temporal settings. The approach is based on extending the circler window for the lasso parameter to a cylindrical window where the base representing space and height representing time. We employ a Bayesian hierarchical modeling approach with a joint implementation of Gibbs and Metropolis-Hasting MCMC computational methods to obtain posterior estimates of all model parameters. We use the lung cancer mortality in 88 counties of Ohio State for the year 1968-88 to illustrate the method.

## Approximating the Multiple-Width-Window Scan Statistic for Nonuniform Background

❖ Joseph Naus, Rutgers University, 8 Brian Road, East Brunswick, NJ 08816, *naus@stat.rutgers.edu*

***Key Words:*** scan, multiple window widths, clusters, non-uniform

Simple, accurate approximations exist for the scan statistic with a fixed-width window for a constant-rate null background. For a non-constant background, or for the case of simultaneously testing with multiple window sizes, approximations were lacking and simulations used. Recently, Naus and Wallenstein, in Methodology and Computing in Applied Probability 6 (2004), provide a simple formula for the scan test with multiple width windows for the constant background null case. Naus and Wallenstein, in Statistics in Medicine (2006), use a simple formula to determine critical cluster sizes for a fixed window scan, given non-constant-rate null background. The present paper derives approximations that can handle multiple window widths in the nonconstant null background temporal case simultaneously.

## Cluster Analysis Using Methods of Pairwise Weight on Mixed Type Attributes

❖ William Warde, Oklahoma State University, Statistics Department, 301 MSCS, Stillwater, OK 74078-1055, *billw@okstate.edu*

***Key Words:*** agglomerative clustering, Rand's C statistic, mixed-type objects, association coefficient

A new formulation of distance is proposed for clustering mixed numeric and multiple binary data. The level of within cluster resemblance as quantified by the similarity, that is pairwise weight between i-th and j-th objects, must be accounted for in cluster analysis. The performance of clustering algorithms with our proposed distance based on methods of pairwise weight on mixed numeric and multiple binary data gives competitive or superior recovery and agreement level to Gower's. The agreement is used to quantify consistency of resultant clusterings produced by cluster algorithms. It might be a natural basis for organizing objects measured on mixed-type attributes by examining resultant clusterings depending on the characteristics of the data. The result of clustering algorithms with proposed distance based on methods of pairwise weights is examined by using principal coordinates.

# 389 Methodology for Spatial Data

Section on Statistics and the Environment, WNAR

**Wednesday, August 9, 8:30 am–10:20 am**

## Spatial Multivariate EOFs: Discrete to Continuous Approximations

❖ Yonggang Yao, The Ohio State University, 1958 Neil Ave., Columbus, OH 43210-1247, *yao@stat.ohio-state.edu*; Noel Cressie, The Ohio State University

***Key Words:*** discrete-to-continuous approximation, empirical orthogonal function, multivariate random field, smoothing interpolation, smoothing spline

Multivariate empirical orthogonal function analysis of multivariate spatial data has been practiced in many disciplines. Due to having a finite number of observations, people often have to extend the mEOF to the whole multivariate random field by using some discrete-to-continuous approximation (DCA) algorithm (e.g., smoothing interpolation). However, the advantages and disadvantages of DCAs have not been explored adequately. This paper considers model-based criteria for choosing a DCA. As an example, the smoothing-spline DCA is considered and applied to Iowa (1970--1990) temperature data.

## Spatial Designs and Strength of Spatial Signal: Effects on Covariance Estimation

❖ Kathryn Irvine, Oregon State University, 1905 NW Dogwood Drive, Corvallis, OR 97330, *georgiti@science.oregonstate.edu*; Alix Gitelman, Oregon State University; Jennifer A. Hoeting, Colorado State University

***Key Words:*** exponential covariance, nugget-to-sill ratio, infill asymptotics

Interpretations of spatial covariance parameters in ecological settings can be used to describe the spatial heterogeneity in a landscape that cannot be explained by covariates. In this paper we investigate the influence of the strength of spatial signal on maximum likelihood (ML) and restricted maximum likelihood estimates (REML) of covariance parameters in an exponential with-nugget model, and we also examine these influences under different sampling designs---specifically, lattice designs and more realistic random and cluster designs---at differing intensities of sampling (n=144 and 361). We find that neither ML nor REML estimates perform well when the range is large, and that the best estimation of the covariance parameters comes under the random sampling design. ML underestimate the autocorrelation function and REML produces highly variable estimates of the autocorrelation function.

## Spatial Models, Spatial Integrals, and Green's Theorem

❖ Gary Simon, New York University, 8 Simpson Place, Stony Brook, NY 11790, *gsimon@stern.nyu.edu*

***Key Words:*** spatial statistics, spatial integrals, Green's theorem

Many types of data, including disease incidence, are reported as counts over m geographical regions. I propose modeling these counts as m independent Poisson random variables. A "force of disease" function with interpretable parameters will be suggested and the Poisson rate for any region will be computed as the average value of this function over the region. This requires, of course, spatial integration. The resulting technical complexities can be overcome with Green's Theorem and the judicious use of mapping software.

## A Variation on Spike-Time Distance Prototypes

❖ Katherine Tranbarger, Amherst College, Amherst College, Department of Math box 2239, Amherst, MA 01002, *ktranbarger@amherst.edu*

***Key Words:*** point process, prototype, distance metric

In the spike-time distance metric proposed by Victor and Pupura, the distance between two realized temporal point patterns is determined by the minimal cost sequence of operations necessary to transform one pattern into the other. Operations include inserting a point, removing a point, and moving a point through time. Using this metric, a prototype pattern can be determined such that the distance between the prototype and all point patterns in the data set is minimized. In previous applications the cost of removing a point is set equal to the cost of inserting a point. This work considers the possible advantages and implications associated with setting that requirement of the metric aside in prototype pattern applications.

## A Method for Multiscale Spatio-Temporal Analysis

❖ Mary Louie, National Center for Health Statistics, 3311 Toledo Road, Suite 3215, Hyattsville, MD 20782, *mlouie@cdc.gov*

***Key Words:*** multiscale, spatial, temporal, autocorrelation, Markov process, disease mapping

For the multiscale spatial process model that is based on a factorization of the Poisson likelihood, we outline a potential method for incorporating a temporal component. We specify a set of multivariate latent variables that are governed by a Markov process so as to incorporate autocorrelation in the prior structure on the multiscale parameters. We consider distributions for the latent variables that are conjugate to that of the prior structure, thereby facilitating posterior computation. Results are presented from a simulation study based on a disease mapping framework.

## A Weighting Class Adjustment Estimator in a Continuous Domain

❖ Breda Munoz, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194, *breda@rti.org*; Virginia M. Lesser, Oregon State University; Leigh Harrod, Oregon State University

***Key Words:*** environmental surveys, missing data, continuous domain, Horvitz-Thompson estimator, weighting class adjustment

Environmental phenomena are the result of random processes that evolve in space and/or time. Environmental surveys are not exempt of missing data issues. Therefore, analysis results when the missing data problem is ignored may be biased, depending on the missing data mechanism. The missing at random (MAR) mechanism assumes that given the observed data, the probability of missing data depends on covariates only. A weighting class adjustment is a common technique used by survey analysts for missing data when the missing mechanism is assumed MAR. In complex surveys, this technique has been used for finite population sampling data. However, the weighting class adjustment has not been used for continuous populations. We extend the concept of weighting class adjustment to the continuous domain and develop an unbiased weighting class adjustment estimator under a stratified sampling assumption.

## 390 Survey-Based Estimation IV ●

Section on Survey Research Methods
**Wednesday, August 9, 8:30 am–10:20 am**

### Effects of Uncontrolled Factors at the Collection Stage on the Canadian Nutrition Survey

❖ FranÁois Verret, Statistics Canada, R.H. Coats Building, 120 Parkdale Avenue, Ottawa, ON K1A 0T6 Canada, *francois.verret@statcan.ca*; Steven Thomas, Statistics Canada

***Key Words:*** Canadian nutrition survey, usual intake distribution, nutrient, uncontrolled collection factor

In 2004, Statistics Canada conducted the Canadian Nutrition Survey. The main goal of the survey was to estimate usual intake distributions of key nutrients for 15 age-sex groups at the provincial level. A sample of 35,000 respondents was selected to do a first dietary recall and 10,000 of those respondents were asked to do a second recall. It was decided to make a weighting adjustment to reduce the bias due to the day of the week the recall was given. However, this adjustment does not solve every problem related to the day of the week the recall was given. The distribution of respondents in terms of socio-demographic characteristics, from one day to another, was not controlled. This paper presents the results of an analysis measuring the effect of several factors on the estimation of usual intake distributions for several key nutrients.

### Regression Diagnostics for Survey Data

❖ Jianzhu Li, University of Maryland, JPSM, 1218 Lefrak Hall, College Park, MD 20742, *jli@survey.umd.edu*; Richard Valliant, University of Michigan

***Key Words:*** regression, diagnostics, complex sampling

Regression diagnostics is an important area in regression theory which intends to explore the characteristics of a fitted regression model for a given data set. Techniques for regression diagnostics have been developed theoretically and methodologically for conventional OLS linear regression models. However, in survey sampling, diagnostics have not been extensively studied. The application of conventional techniques to survey data becomes less straightforward because of features of complex sampling designs like stratification, clustering, and weights. The goal of this research is to adapt and extend some of the standard regression diagnostics to the survey setting.

### Design Effects in Randomized Experiments Based on Sample Surveys

❖ K. P. Srinath, Abt Associates Inc., 4550 Montgomery Ave., Suite 800 North, Bethesda, MD 20814, *kp_srinath@abtassoc.com*

***Key Words:*** complex sampling design, sampling weights, treatment and control groups

Evaluation of new social or educational programs is frequently done using randomized experiments based on samples from surveys that use complex designs. Such evaluations may involve specific geographic areas like state, Census regions or could be at the national level. Dividing a sample of last stage sampling units to treatment and control groups at random in a multi-stage design is common. A question of interest is what is an appropriate design in terms of the number of units at each stage, the overall probability of selection and the sampling weight that minimizes the design effect. The desire is to minimize the design effect as much as possible to maximize the power of statistical tests for testing differences. We examine alternative designs including the self weighting design both from an operational point of view and design effect.

### Using Census Data to Define Estimation Areas for the American Community Survey: a Case Study

❖ Joseph Powers, U.S. Census Bureau, P.O. BOX 1607, COLLEGE PARK, MD 20740, *joseph.c.powers@census.gov*; Alfredo Navarro, U.S. Census Bureau

***Key Words:*** reliability, naive clusters, statistical clustering, compactness, weighting, estimation areas

In January 2005, the American Community Survey (ACS) expanded to sample all 3,219 counties in the U.S. and Puerto Rico. The ACS weighting and estimation methodology requires estimation areas to meet a minimum population size so that the observed sample size is big enough to produce estimates with adequate reliability. Counties below the threshold size must be grouped or clustered prior to estimation. A simple method groups the counties based on adjacency and then assess all the clusters using a predefined criterion. A better, automated algorithm was also developed. The algorithm is an iterative method that uses a set of Census long form characteristics to define a similarity index based on the Euclidean distance metric. This paper describes the naÔve method, the algorithm, and a statistical assessment. The results of the two schemes are compared for Puerto Rico and Texas.

### Estimating Birth Counts for Small Geographical Domains Used for Control Totals in Raking Adjustment

❖ Amang Sukasih, Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024, *asukasih@mathematica-mpr.com*; Donsig Jang, Mathematica Policy Research, Inc.; Mary Edith Bozylinsky, Mathematica Policy Research, Inc.; Barbara L. Carlson, Mathematica Policy Research, Inc.

***Key Words:*** weighting, estimation, natality, city of birth, CDC, fragile families

Raking is often used in survey weighting to control the distribution of the sample so that it is consistent with some known population distributions or totals by key variables. Such control totals are usually obtained from the sample frame or external data; however, domains available are sometimes not exactly matched to domains of interest. CDC natality data is a good source for population birth totals. Data on place of birth (state and county) are available. However, the same variable for smaller geographical units like city is not available, yet often some policy research studies need such birth totals by city of birth. The Fragile Families Survey is one example where births were sampled within hospitals within selected cities, and city-level analysis was of primary interest. We will explain how weighting, raking, and creating synthetic estimates of total births by city were done.

### Iteration of Second-Stage and Composite Procedures in the Current Population Survey

❖ Samantha Cruz, Bureau of Labor Statistics, 300 M Street, SW, N413, Washington, DC 20024, *Cruz.Samantha@bls.gov*; Edwin L. Robison, Bureau of Labor Statistics; Tamara S. Zimmerman, Bureau of Labor Statistics

***Key Words:*** iterative proportional fitting, raking, weighting

Labor force estimates of the U.S. civilian noninstitutional population are derived through a number of weighting steps in the Current Population Survey (CPS) estimation procedure. Currently, the weight for each interviewed person includes a second-stage ratio adjustment. This second-stage estimation process uses Iterative Proportional Fitting (IPF), or raking, in order to simultaneously match three sets of independent population controls and to create second-stage weights. Upon completion of the second-stage process, the weights are subjected to a composite estimation process which also uses a three-way rake to match composite population controls; but the procedure does not reproduce all second-stage controls. This paper explores more complex raking methods to determine if a single set of weights can be produced that simultaneously preserves all second-stage and composite controls.

### Estimation and Reliability Issues of Health Estimates from the Behavioral Risk Factor Surveillance System for U.S. Counties Contiguous to the United States-Mexico Border

❖ Joe Fred Gonzalez, Jr., National Center for Health Statistics, 3311 Toledo Road, Room 3121, CDC, Hyattsville, MD 20782, *jgonzalez@cdc.gov*; Machell Town, National Center for Chronic Disease Prevention and Health Promotion; Jay J. Kim, National Center for Health Statistics; Sam Notzon, National Center for Health Statistics; Juan R. Albertorio, National Center for Health Statistics

***Key Words:*** unbiased estimation, relative standard error, relative root mean square error

The Behavioral Risk Factor Surveillance System (BRFSS) is a State telephone based survey of the adult civilian non-institutionalized population residing in the United States. Consequently, the BRFSS final weights that are currently available in the data files are designed to produce unbiased estimates of socio-demographic and health characteristics at the State level. In addition to State level BRFSS estimates, there is interest in the health status of adults residing in the 25 U.S. counties contiguous to the United States-Mexico Border Region. The purpose of this paper is to investigate alternative ways of producing border region weights, estimates of BRFSS socio-demographic and health characteristics and their reliability in terms of relative standard error and relative root mean square error.

## 391 Noether Award Invited Session ● ✪

Noether Award Committee, Section on Nonparametric Statistics
**Wednesday, August 9, 10:30 am–12:20 pm**

### Doing Thousands of Hypothesis Tests at the Same Time

❖ Bradley Efron, Stanford University, Department of Statistics, Stanford, CA 94305, *brad@stat.Stanford.edu*

***Key Words:*** empirical null, false discovery rate, empirical Bayes

Classic multiple comparisons theory, beginning with the Bonferroni bound, was aimed at preserving frequentist p-values when simultaneously dealing with several testing problems. “Several” meant two to perhaps a dozen. Modern scientific technology---microarrays being the prototype---routinely confront the statistician with thousands of simultaneous tests. Statistical inference can be qualitatively different in such situations, with Bayes or empirical Bayes ideas forcing themselves upon even dedicated frequentists. I will discuss examples and a bit of current theory.

### Bayesian Mixed Models for Functional Data

❖ Jeffrey S. Morris, M. D. Anderson Cancer Center, Department of Biostatistics and Applied Mathematics, 1515 Holcombe Blvd Box 447, Houston, TX 77030, *jeffmo@odin.mdacc.tmc.edu*

***Key Words:*** functional data analysis, wavelets, Bayesian methods, nonparametric regression, proteomics, mixed models

Many studies yield functional data, with the ideal units of observation curves and observed data sampled on a fine grid. These curves frequently have irregular features requiring spatially adaptive nonparametric representations. We discuss new methods for modeling these data using functional mixed models, which treat the curves as responses and relate them to covariates using nonparametric fixed and random effect functions. The Bayesian wavelet-based approach yields adaptively regularized posterior samples for all model parameters that can be used for any desired Bayesian estimation, inference, or prediction. We illustrate this method on four applications yielding spiky functional data and describe how it can be extended to deal with incomplete functional data for which some regions of some functions are missing and to model higher-dimensional functional data (e.g., images).

## 392 Government Statistical Agencies Are Now Offering Electronic Reporting to Their Respondents, but Is It Worth the Effort? ● ✪

Section on Government Statistics
**Wednesday, August 9, 10:30 am–12:20 pm**

### Qui Bono: Who Benefits from Web Data Collection?

❖ Michael D. Levi, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Suite 5110, Washington, DC 20212, *levi.michael@bls.gov*

***Key Words:*** electronic data collection, web data collection

This paper will reflect on the Bureau of Labor Statistics’ experience with web-based data collection. It will analyze the benefits and drawbacks of this collection mode from the perspectives of the respondent, the statistical agency, and the consumer of statistical output. In the course of discussion, the paper will address the tradeoff between security and respondent burden, usability issues that arise when converting a paper schedule to the web, solicitation strategies and pick-up rates, the curious case of email forms, and unanticipated resource demands.

### The Development of Electronic Data Collection Techniques

❖ Toni Räikkönen, Statistics Finland, *toni.raikkonen@stat.fi*

***Key Words:*** XML, EDR, data collection, survey

Most web surveys are quite user interactive. The respondent usually gets to see a web questionnaire that is fully customized and provides at least some validation routines. Implementing this kind of system is easy when using traditional web programming techniques. The problem is that it is quite difficult to create such a system so flexible that it can handle different types of data collections without the need for additional programming. We have developed in Statistics Finland a generic server application that can take care of most web surveys, typically without the need of any programming. This generic application is heavily XML-based and uses XML for the description language of the questionnaires and survey logic. The power of the approach is the way XML and the related techniques are used throughout the application.

### Statistics Canada's Electronic Data Reporting Experience

❖ Jocelyn Burgess, Statistics Canada, 120 Parkdale Ave., Ottawa, ON K1A 0T6 Canada, *jocelyn.burgess@statcan.ca*

***Key Words:*** electronic data reporting, collection methods, new strategies, challenges at Statistics Canada

Statistics Canada has continued to move toward collection methods that are more efficient, secure, and responsive to the needs and preferences of respondents. In support of these objectives, the agency aggressively embarked five years ago on the development of the internet Electronic Data Reporting (EDR) capacities for each of approximately 50 business and agriculture surveys. During the past year, however, as a result of low take-up rates and increasing technical difficulties, Statistics Canada has been re-evaluating its EDR work to date from a technical, security, and cost point of view. New strategies are now emerging that will position the bureau to offer a more streamlined, technically simple, and cost-effective electronic mode of collection to respondents, where judged to be appropriate. This paper will outline the challenges Statistics Canada has faced during the past four years and present new ideas for delivering EDR solutions to respondents.

### Improving the Provider Experience: the Vision for Multi-Modal Data Collection in Australia

❖ Sean Thompson, Australian Bureau of Statistics, 639 Wickham Street, Fortitude Valley , 4006, Australia, *sean.thompson@abs.gov.au*

In 2002, the ABS adopted a strategy of "Total Approach Management" in relation to its business data providers. At its simplest, the strategy provides a framework for viewing interactions with data provider holistically---across collections and work areas and over time---rather than as a discordant series of uncoordinated channels of communication. More recently, we have developed a strategy for Multi-Modal Data Collection (MMDC), which focuses on developing communications infrastructure, especially for electronic data reporting (EDR), that is consistent with the philosophy of Total Approach Management. It could be argued that the "hasten slowly" policy on EDR served the ABS reasonably well for some time, as many of our early investment opportunities in EDR look less than ideal in hindsight.

## 393 Bayesian Methods in Cancer Genomics ● ✪

WNAR, Biometrics Section, Section on Bayesian Statistical Science, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Variable Selection in Regression Mixture Modeling for the Discovery of Gene Regulatory Networks

❖ Joseph G. Ibrahim, The University of North Carolina at Chapel Hill, Department of Biostatistics, McGavran-Greenberg Hall, Chapel Hill, NC 27599, *ibrahim@bios.unc.edu*; Mayetri Gupta, The University of North Carolina at Chapel Hill

***Key Words:*** transcription regulation, motif discovery, hierarchical model, evolutionary Monte Carlo, importance sampling, Bayesian model selection

The availability of diverse types of genomic data---such as DNA sequence, gene expression microarray, and proteomic data---has led to a rapid growth of statistical research in the effort to decipher gene regulatory networks---the interactions between genes (or groups of genes) in regulating a biological process. A natural way to address these issues is to combine gene clustering and motif discovery in a mixture framework, with unknown components representing the latent gene clusters and genomic sequence features linked to the resultant gene expression through a multivariate hierarchical regression. We demonstrate a hierarchical regression mixture model for genomic sequence and expression data and propose a Monte Carlo method for simultaneous variable selection (motifs) and clustering (genes).

### Variable Selection in Clustering via Dirichlet Process Mixture Models

❖ Marina Vannucci, Texas A&M University, College Station, TX 77843-3143, *mvannucci@stat.tamu.edu*

***Key Words:*** clustering, Dirichlet process mixture, microarrays

The increased collection of high-dimensional data in various fields has raised a strong interest in clustering algorithms and variable selection procedures. In this talk, a model-based method that handles the two problems simultaneously will be proposed. Dirichlet process mixture models are used to define the cluster structure, and a latent binary vector is introduced to identify discriminating variables. Inference on the cluster structure is done via Gibbs sampling, and the variable selection index is updated using stochastic search techniques. Performance of the methodology is explored on simulated and DNA microarray data.

### Nonparametric Models for Proteomic Peak Identification, Quantification, and Classification

❖ Merlise Clyde, Duke University, ISDS Box 90251, 223E Old Chemistry Building, Durham, NC 27708-0251, *clyde@stat.duke.edu*; Leanna House, Duke University; Robert Wolpert, Duke University

***Key Words:*** Bayesian, Levy process, kernel regression

We present model-based inference for proteomic peak identification, quantification, and classification from mass spectroscopy, focusing on nonparametric Bayesian kernel models. Using experimental data generated from MALDI-TOF mass spectroscopy (Matrix Assisted Laser Desorption Ionization Time of Flight), we model observed intensities in the spectra with a nonparametric model for the expected intensity as a function of time-of-flight. In particular, we express the unknown function as a sum of kernel functions, which provide a natural choice of basis functions for modeling peaks. We give interpretations of model parameters in the context of the problem and how to place priors on the unknown function using Levy random field priors. Extensions of the model to multiple spectra will be presented.

# 394 Building Statistical Capacity in Developing Countries ✪

Committee on International Relations in Statistics, Section on Statistical Education

**Wednesday, August 9, 10:30 am–12:20 pm**

## Statistics in Argentina

❖ Diana Kelmansky, University of Buenos Aires, Acoyte 136, Piso 20, Dto "D", Capital Federal, 1405 Argentina, *dkelman@ic.fcen.uba.ar*

This presentation will provide a brief overview of the current state of statistics in Argentina. Data are provided on the numbers of academic statistics departments in the country and the typical number of graduates and how these patterns have changed over time. Breakdowns are provided regarding the number of statisticians working in industry, academia, and government. The speaker will describe her own experiences during this past year as an educational ambassador.

# 395 Latent Class Models for Disease Classification ●

ENAR, Biometrics Section, WNAR

**Wednesday, August 9, 10:30 am–12:20 pm**

## Diagnosing Sepsis in Patients with SIRS

❖ Klaus Larsen, University of Copenhagen, Hvidovre Hospital, Kettegaard AllÈ 30, DK - 2650 Hvidovre, Denmark, *klaus.larsen@hh.hosp.dk*

***Key Words:*** LCA, ROC curve, sepsis, SIRS

Biomarkers are used routinely for diagnosing sepsis among patients with systemic inflammatory response syndrome (SIRS). However, it is still unclear how useful they really are for diagnostic purposes, to some extent due to a lack of statistical methodology. While classical approaches consider only one parameter at a time, latent class regression models allow for the inclusion of several biomarkers as indicators of a latent class variable, which may be interpreted as the true disease classification. Other variables, such as body temperature and respiratory frequency, may be included as covariates predicting the latent class membership. The ability of the proposed model to classify the patients correctly is compared with existing methods, such as the ROC curve, on a dataset of SIRS patients collected at Hvidovre University Hospital in Copenhagen, Denmark.

## Penalized Latent Class Methods for Disease Classification

❖ Andres Houseman, Harvard School of Public Health, Department of Biostatistics, 655 Huntington Ave, Boston, MA 02115, *ahousema@hsph.harvard.edu*; Brent A. Coull, Harvard School of Public Health; Rebecca Betensky, Harvard School of Public Health

***Key Words:*** cancer, LASSO, loss of heterozygosity, mixture models, penalized likelihood, ridge regression

There is increasing interest in classifying disease on the basis of a moderate number of genetic factors. For binary outcomes, latent class analysis (LCA) is a natural approach. However, LCA is complicated when the number of subjects is of small to moderate size. We present a regularization approach, penalized LCA, that improves estimation. We consider an LCA model where, conditional on unobserved membership in one of k classes, item responses are independent. Using a family of penalties, including the ridge and lasso, we extend this model to address higher-dimensional problems. Penalization may be imposed across items when the responses have a natural ordering, or may be imposed across classes when it is of interest to identify items that distinguish class membership. We illustrate our methods on loss-of-heterozygosity data from brain tumors and hypermethylation data in bladder tumors.

## Latent Class Measurement of Frailty and Dysregulation in Older Adults

❖ Karen Bandeen-Roche, Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, JHBSPH, 615 N. Wolfe St., Baltimore, MD 21205-2179, *kbandeen@jhsph.edu*

***Key Words:*** construct, validity, conditional independence, aging, prediction, constraint

Frailty is a health state that geriatricians assert as recognizable, but has proven illusive to classify rigorously. It has no gold standard, thus classification must be driven by theory on the underpinnings and clinical presentation of the syndrome. This paper proposes methodology for the measurement of frailty and related states of dysregulation that integrates latent class modeling of the data with the biological theory, hence primarily targeting internal validity. Utility of latent class modeling is evaluated and compared side by side with methods that invoke fewer distributional assumptions and are focused primarily on predictive validity. Measurement strategies that counterbalance the two strategies are proposed, evaluated, and illustrated. The paper aims to improve methodology for the measurement of syndromes lacking a gold standard.

# 396 Semiparametric Inference in Practice

IMS, Section on Nonparametric Statistics

**Wednesday, August 9, 10:30 am–12:20 pm**

## Semiparametric Approaches To Model the Survival and Longitudinal Data Simultaneously

❖ Jane-Ling Wang, University of California, Davis, 1 Shields Ave., Department of Statistics, Davis, CA 95616, *wang@wald.ucdavis.edu*; Jimin Ding, University of California, Davis; Fushing Hsieh, University of California, Davis

***Key Words:*** EM-algorithm, Monte-Carlo integration methods, random effects, profile likelihood, nonparametric maximum likelihood estimates, joint modelling

In many longitudinal studies, the event-times of interest (termed failure times) are collected with baseline and longitudinal covariates. The relationship between a failure time process and some longitudinal covariates and the understanding of the patterns of the longitudinal covariate processes are of key interest. Several difficulties arise when covariates are measured intermittently, possibly with error, and no measurements are available after the event-time, which triggers nonignorable dropout. Semiparametric joint likelihood approaches have emerged as effective ways to model both processes. However, there are challenges both on the computational front and for statistical inference. In this talk, we address both issues, which are intertwined, and discussed several practical solutions.

### Semiparametric Models with Data Missing by Design and Inverse Probability Weighted Empirical Processes: Partial Results and Open Problems

❖ Jon A. Wellner, University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195-4322, *jaw@stat.washington.edu*

***Key Words:*** Horvitz-Thompson processes, exchangeable bootstrap, missing data, sampling without replacement

Horovitz-Thompson or inverse probability weighted (IPW) versions of likelihood equations provide one simple method of estimation for two-phase stratified sampling designs. In this talk, I will discuss results for the basic Horovitz-Thompson empirical processes, which allow study of the asymptotic behavior of estimators for semiparametric models with nuisance parameters---which can be estimated at rate square-root n. The results are based on theorems for exchangeably weighted bootstrap methods obtained in the early 1990s. This talk will complement a related talk by Norman Brewlow on different aspects of the same problem.

### Statistical Inference for Variable Importance

❖ Mark van der Laan, University of California, Berkeley, Division of Biostatistics, School of Public Health, 140 Warren Hall 7360, Berkeley, CA 94720, *laan@stat.berkeley.edu*

***Key Words:*** causal effects, efficient influence curve, estimating function, variable importance, adjusted variable importance

Many statistical problems involve learning importance/effect of a variable for predicting an outcome of interest based on observing a sample of n independent and identically distributed observations on a list of input variables and an outcome. For example, though prediction/machine learning is, in principle, concerned with learning the optimal unknown mapping from input variables to an outcome from the data, the typical reported output is a list of importance measures for each input variable. The typical approach in prediction has been to learn the unknown optimal predictor from the data and derive, for each of the input variables, the variable importance from the obtained fit. In this article, we propose a new approach that involves for each variable separately defining the wished variable importance as a parameter and deriving the efficient influence curve.

## 397 Balanced Sampling ●

SSC

**Wednesday, August 9, 10:30 am–12:20 pm**

### Balanced Sampling by Means of the Cube Method

❖ Yves Tillé, University of Neuch,tel, Espace de l'Europe 4, CP. 805, Neuch,tel, 2002 Switzerland, *yves.tille@unine.ch*

***Key Words:*** survey sampling, balanced sampling, unequal probability, stratification

This talk will describe the main principle of the cube method. A balanced sampling design has the important property that the Horvitz-Thompson estimators of the totals for a set of auxiliary variables are equal to the totals we want to estimate. Therefore, the variances of all variables of interest are reduced, depending on their correlations with the controlled variables. Recently, Deville and TillÈe developed the cube method, which allows selecting balanced samples with equal or unequal probabilities---the sample being balanced on a large set of auxiliary variables. The cube method is actually a family of algorithms that satisfies exactly a given vector of inclusion probabilities.

### Stochastic Imputation Using Balanced Sampling

❖ Jean-Claude Deville, CREST/ENSAI, Laboratoire de Statistique d'EnquÍte, Campus de Ker-Lann, BRUZ, 35370 France, *deville@ensai.fr*

***Key Words:*** stochastic imputation, balanced sampling, martingale, nonresponse, cube method

Dealing with nonresponse is a necessity in survey data analysis. Weighting is a nearly perfect technique because, in particular, it allows a consistent estimation of the distribution function. Imputation is practically mandatory in the case of partial nonresponse. Deterministic imputations are invalid for estimating nonlinear transformations of the variable (the distribution function). One often add a random term to the predictor with , which adds a parasite variance to all estimations. Using the estimating equations of the prediction model and some techniques of balanced sampling, this extra-variance can be considerably reduced. In the case of donor imputation, coherency between imputation and weighting can be achieved. For discrete variables, balance is obtained with the cube method. For numerical variables, new ideas based on elementary properties of martingales are necessary.

### Use of Balanced Sampling in the Framework of the Master Sample for French Household Surveys

❖ Marc Christine, Institut National de la Statistique et des Études Économiques, 18 Bd Adolphe Pinard, Paris, 75014 France, *marc.christine@insee.fr*

***Key Words:*** household surveys, balanced sampling, primary units, simulations, geographical spreading out

French national household surveys are based on a sample of dwellings referred to as "master sample." An initial selection stage consists of drawing primary units (PU). In order to reduce variance due to this first stage, balanced sampling has been implemented---the PU being selected with balancing conditions at a "super-regional" level, using age and income variables. The statistical properties of drawn samples have been studied---first with simulations, then by measuring the 'representativeness' of the sample regarding other variables of interest highly correlated with those collected in surveys. Extensions of the technique allow for drawing samples with a good geographical spreading out. Other applications will be implemented for future samples derived from the new French rotative census, especially for the Labour Force Survey.

### Sampling and Estimation Strategies for the Canadian Unincorporated Business Population

❖ Wisner Jocelyn, Statistics Canada, 120 Parkdale Ave., 11th Floor, R.H. Coats Building, Ottawa, ON K1A 0T6 Canada, *wisnjoc@statcan.ca*

***Key Words:*** balance sampling, calibration, estimation, sampling plan

Canadian unincorporated businesses complete their income tax returns either on paper or electronically (internet, diskette, etc.). More than half of the returns are submitted electronically; we now are exploring ways of using only the electronic portion of the universe to produce estimates covering the entire population. Statistical tests on some of the key variables have shown paper filers and electronic filers are significantly different. We can not therefore assume the electronic filer file is a random sample of the entire, unincorporated universe. This paper illustrates the use of balance sampling on the electronic filer

subpopulation for inference. That strategy is then compared to a traditional sampling approach.

# 398 Statistical Challenges in Analyzing Highly Stratified Data

Biometrics Section, WNAR

**Wednesday, August 9, 10:30 am–12:20 pm**

## Semiparametric Transformation Models with Random Effects for Highly Stratified Survival Data

❖ Danyu Lin, The University of North Carolina at Chapel Hill, 3101E McGavran-Greenberg Building, CB#7420, Department of Biostatistics, Chapel Hill, NC 27599-7420, *lin@bios.unc.edu*

***Key Words:*** correlated failure times, frailty model, maximum likelihood, proportional hazards, semiparametric efficiency, survival analysis

We propose a class of semiparametric transformation models with random effects to formulate the effects of possibly time-dependent covariates on highly stratified survival times. We show the nonparametric maximum likelihood estimators are consistent, asymptotically normal, and asymptotically efficient. We also develop the corresponding likelihood-based inference procedures. Our theory covers all commonly used transformation models, including proportional hazards and proportional odds models, and allows a variety of random-effects distributions such as Gaussian and log-gamma distributions. Simulation studies demonstrate the proposed methods perform well in practical situations. An illustration with a well-known diabetic retinopathy study will be provided.

## Connections between Bayesian and Conditional Inference in Matched Studies

❖ Kenneth Rice, University of Washington, F-600 Health Sciences Building, Department of Biostatistics, Campus Mail Stop 357232, Seattle, WA 98195-7232, *kenrice@u.washington.edu*

***Key Words:*** matched case-control, Bayesian, conditional likelihood, semi-parametric, non-parametric, highly stratified

For matched case-control studies, conditional logistic regression is the 'industry standard' method, producing well-behaved, attractively robust inference. Given the implicit notions of exchangeability in such highly stratified data, Bayesian approaches are also attractive and offer substantial flexibility for applications over the conditional framework. However, the Bayesian approach requires one to assume priors for the many nuisance parameters associated with stratification; in general, this choice actively influences the analysis and requires some care. We present a class of nuisance-parameter priors that, marginally, produce the conditional likelihood exactly. This allows advocates of the standard approach access to Bayesian flexibility, while Bayesians gain the benefits of conditioning. Further attractive but nonstandard features of the Bayesian model also are explored.

## Model-Based Profile Confidence Intervals for Stratified Contingency Tables

❖ Joseph B. Lang, The University of Iowa, Department of Statistics and Actuarial Science, Iowa City, IA 52242, *joseph-lang@uiowa.edu*

***Key Words:*** categorical data, multinomial-Poisson homogeneous models, profile likelihood confidence interval

Consider a parsimonious model that imposes several restrictions on contingency table probabilities. Two facts are agreed upon generally: (1) When the model holds, restricted-model estimation is more efficient than unrestricted-model estimation; and (2) profile confidence intervals often have better coverage properties than their Wald counterparts. Motivated by these facts, this talk will describe restricted-model profile likelihood confidence intervals for a broad class of contingency table parameters. In particular, a simple and broadly applicable computational algorithm will be described. As an illustrative example, counts resulting from a stratified random sample will be used to estimate a common gamma coefficient in K rxc contingency tables. The proposed method is based on the theory of multinomial-Poisson homogeneous models.

## Fixed-Effects Models for Longitudinal Binary Data with Drop-Outs Missing at Random

❖ Paul Rathouz, The University of Chicago, Department of Health Studies, 5841 S Maryland Ave., MC2007, Chicago, IL 60649, *prathouz@uchicago.edu*

***Key Words:*** subject-specific model, missing data, conditional logistic regression, double robustness, semi-parametric efficiency

We consider the problem of attrition under a logistic regression model for longitudinal binary data in which each subject has his own intercept parameter, eliminated via conditional logistic regression. This fixed-effects model exploits the longitudinal data by allowing subjects to act as their own controls. By conditioning on the drop-out process, we develop a valid but inefficient conditional likelihood using the complete-record data. Noting the drop-out process is ancillary in this model, we use a projection argument to develop a score with improved efficiency over the conditional likelihood score and embed both of these scores in a more general class of estimating functions. We propose a member of this class that approximates the projected score while being more computationally feasible. We present a small simulation and an example analysis from aging research.

# 399 Human Perception and Statistical Graphics ●

Section on Statistical Graphics, Section on Statisticians in Defense and National Security

**Wednesday, August 9, 10:30 am–12:20 pm**

## Attention, Consciousness, and Data Display

❖ Ronald A. Rensink, The University of British Columbia, Department of Psychology, 2136 West Mall, Vancouver, BC V6T 1Z4 Canada, *rensink@psych.ubc.ca*

***Key Words:*** human vision, change perception, change blindness, information visualization

Recent advances in our understanding of human vision have shown we can be blind to major events in our environment, unless we attend to them. Recent advances also show we have severe limits as to how much we can attend. This talk will review these findings and propose ways of dealing with these limits to achieve effective visual display of information.

### Perception of Scene Spatial Layout and Complex Visual Displays

❖ Aude Oliva, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, Building: NE20-463, Cambridge, MA 02139, *oliva@mit.edu*

***Key Words:*** image perception, scene perception

One of the remarkable aspects of image perception is that human observers are able to recognize the gist of a complex scene very quickly and independently of the complexity of the image. The gist of a scene refers to the visual information processed during just a glimpse of a new image. This talk will review findings in human perception and computational modeling that help us understand which mechanisms we use to achieve fast image perception and rapid perceptual grouping of visual information. It also will describe how to use our understanding of these mechanisms for the display of complex visual information.

### Toward Integrating Perception, Cognition, and Visual Statistical Analytics in Quantitative Visualization

❖ Daniel B. Carr, George Mason University, Fairfax, VA 22030, *dcarr@gmu.edu*

***Key Words:*** visual analytics, comparative analysis, homeland security, health, environment, education

This talk will address the development of visual analytic software that uses methods from statistics. Many of today's information visualization approaches to quantitative data focus on overviews and drilling down to a data subset or to individual items. This contrasts with the statistics paradigm that partitions data, compares distributions, and makes decisions. The Department of Homeland Security recently called attention to the pressing need for visual analytics. A comparative analysis statistical paradigm is applicable to selected homeland security visual analytic needs and to health, environment, and a host of other applications. This talk will show recent software that endeavors to integrate statistical methodology with guidance from cognitive science, usability, and visual design communities to reach communities with little statistical background.

## 400 Haplotype Analysis ● ✪

Section on Risk Analysis, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Haplotype Analysis in Related Individuals

❖ Hongyu Zhao, Yale University, New Haven, CT 06520, *hongyu.zhao@yale.edu*; Ning Sun, Yale University

***Key Words:*** haplotype, linkage, association, statistical genetics, complex traits

The completion of the first two phases of the HapMap project has led to rich information on haplotype structures in different genomic regions across diverse populations. Such information coupled with high-density genotyping platforms has made it feasible to identify regions associated with diseases through genome-wide, haplotype-based analysis. Although many methods have been developed in recent years for haplotype analysis in unrelated individuals, few are available for haplotype analysis for samples consisting of related people. In this presentation, we propose statistical methods that can utilize high-density marker data for haplotype analysis of such samples. The usefulness of these methods will be demonstrated through their applications to real data.

### A Comparison of Haplotype-Based and Tree-Based SNPs Imputation in Association Studies

❖ James Y. Dai, University of Washington, F-600 Health Sciences Building

1705 NE Pacific Str, Department of Biomathematics, Seattle, WA 98195, *yud@u.washington.edu*; Ingo Ruczinski, Johns Hopkins University; Michael LeBlanc, Fred Hutchinson Cancer Research Center; Charles Kooperberg, Fred Hutchinson Cancer Research Center

***Key Words:*** missing SNPS, association inference, haplotype, nonparametric tree regression

Missing single nucleotide polymorphisms (SNPs) data are quite common in genetic association studies. Often, subjects with missing SNPs are discarded in analyses. Here, we develop two haplotype-based imputation approaches and one regression tree-based imputation approach for association studies. The goal is to assess the imputation accuracy and evaluate the impact of imputation on parameter estimation. We apply the methods to simulated data and a case-control study on developmental dyslexia. Our results suggest imputation generally improves over the standard practice of ignoring missing data in terms of bias and efficiency. The haplotype-based approaches slightly outperform the tree-based approach, but the latter has a substantial computational advantage. Finally, we demonstrate that utilizing the disease status in imputation generally helps to reduce the bias.

### Haplotype and SNP Analyses in Genetic Epidemiology with Application to Longitudinal Data

❖ M. Daniele Fallin, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Room W6509, Baltimore, MD 21205, *dfallin@jhsph.edu*; Kelly S. Benke, Johns Hopkins Bloomberg School of Public Health

***Key Words:*** haplotype, SNP, longitudinal, risk, phase

Haplotype-based analyses have been shown to increase power for genetic association studies, but require knowledge of phase, which is not measured directly in most studies. Statistical methods for estimation of phase and haplotype risk in a GLM framework have been proposed, but not widely applied for longitudinal data. We evaluate the relative performance of different approaches for longitudinal data, including a two-stage approach where haplotype phase is estimated first and then incorporated into longitudinal risk modeling, a single-stage approach where haplotype phase and regression parameters are updated iteratively, and using single-SNP data without phase assignment. Methods are evaluated through simulation and applied to empirical data.

## 401 Bayesian Hierarchical Modeling of Exposure Pathways ●

Section on Statistics and the Environment, Section on Bayesian Statistical Science

**Wednesday, August 9, 10:30 am–12:20 pm**

### Regional Spatial Modeling of Toxic Metals in Various Environmental Media

❖ Catherine A. Calder, The Ohio State University, Department of Statistics, 1958 Neil Ave., Columbus, OH 43221, *calder@stat.ohio-state.edu*

***Key Words:*** environmental science, exposure, hierarchical Bayesian modeling, spatial statistics

Characterizing the variation in human exposure to toxic substances over large populations often requires an understanding of the geographic variation in environmental levels of toxicants. This knowledge is essential when the primary routes of exposure are through interaction with environmental media, as opposed to more individual-specific exposure routes (e.g., occupational exposure). In this study, we focus on modeling the spatial variation in the levels of heavy metals (arsenic, cadmium, chromium, and lead) in air, soil, and water. We draw on a variety of different types of relevant data, including emission inventories and ambient environmental measurements, and develop hierarchical Bayesian models that allow us to combine these different types of information in a coherent and scientifically appropriate manner.

### Characterizing Human Exposure to Toxic Metals Using a Bayesian Pathways Model

❖ Nancy J. McMillan, Battelle, 505 King Ave., Columbus, OH 43201, *mcmillann@battelle.org*

***Key Words:*** structural equation modeling, pathways and routes of exposure, environmental exposure, Bayesian pathways model

Human exposure to some toxic substances can be characterized by levels of that substance measured in personal biological samples (biomarkers). The complex mechanisms by which humans are exposed environmentally to these substances can be explored by structural equation modeling (SEM) of coincident environmental media and biomarker measurements. SEM explores associations between correlated environmental media measurements and a biomarker to assess a hypothesized set of pathways of exposure. A hierarchical Bayesian statistical model, akin to SEM, for assessing the pathways and routes by which people are exposed to toxic substances is proposed. Using heavy metals exposure data from the National Human Exposure Survey (NHEXAS), results of applying the Bayesian pathways model are described. Assessments are made of the relative importance of each pathway on resulting biomarker levels.

### Computational Strategies for Fitting and Learning from Complex Bayesian Hierarchical Models

❖ Peter F. Craigmile, The Ohio State University, 404 Cockins Hall, 1958 Neil Avenue, Columbus, OH 43214, *pfc@stat.ohio-state.edu*

***Key Words:*** Markov chain Monte Carlo (MCMC), model diagnostics, sensitivity analyses, spatial statistics

The exposure pathways models that relate biomarker concentrations of toxic metals to local and global environmental exposures, while hierarchical in nature, often involve thousands of parameters. Because closed form expressions for the joint posterior distributions are unavailable, we sample from these distributions using Markov chain Monte Carlo (MCMC). We will discuss the computational methods we use to fit complicated hierarchical Bayesian models using MCMC. A key ingredient in making this work is the Battelle MCMC library, which allows a statistician to focus on the modeling, rather than the intensive data management required in large problems. We demonstrate methods for diagnosing and assessing the fit of these complicated models and consider sensitivity analyses that describe the impact of different parts of the model on the posterior distributions of the parameters.

## 402 Statistical Consulting for Clinical Research ●

Section on Statistical Consulting, Section on Statistical Education

**Wednesday, August 9, 10:30 am–12:20 pm**

### Statistical Consulting for Clinical Research

❖ Jeff Sloan, Mayo Clinic College of Medicine, Kahler 1A, 200 First Street, SW, Rochester, MN 55905, *sloan@mayo.edu*; ❖ Todd Nick, Cincinnati Children's Hospital Medical Center, Center for Epidemiology and Biostatistics, 3333 Burnet Avenue, Cincinnatti, OH 45229, *Todd.Nick@cchmc.org*; ❖ Felicity B. Enders, Mayo Clinic College of Medicine, , *enders.felicity@mayo.edu*; ❖ Michael Griswold, Johns Hopkins Bloomberg School of Public Health, , *mgriswol@jhsph.edu*

***Key Words:*** consulting, career building, publishing, grantsmanship

The session covers three major topic areas: pragmatic aspects of consulting, working with nonstatisticians (especially coping with challenging collaborators), and shepherding the statistician's career. Topics to be covered include how to write grant/statistical sections, how to review an article or grant, statistical collaboration vs. fee-for-service, discussing authorship, and ensuring reproducible research. The primary target audience is consulting statisticians, especially those in applied environments and young investigators. We anticipate clinical researchers from fields such as epidemiology, psychology, economics, and psychiatry also may have interest.

## 403 Statistics for Weather Forecasting I: Challenges and Opportunities ● ✪

Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences, Section on Statistics and the Environment

**Wednesday, August 9, 10:30 am–12:20 pm**

### General Introduction to Weather Prediction

❖ Clifford Mass, University of Washington, Department of Atmospheric Sciences, Box 351640, Seattle, WA 98195, *cliff@atmos.washington.edu*

***Key Words:*** weather, prediction, forecasting

This talk will trace the development of the technology of weather prediction and review the current state of the science. Atmospheric sciences began as a probabilistic "art," but rapidly turned deterministic with the success of numerical weather prediction. Today, with the rapid develop of ensemble prediction techniques and improved post-processing, weather prediction is returning to its probabilistic roots.

## Local Bayesian Model Averaging for Calibrated Weather Forecast Probabilities

❖ Eric Grimit, University of Washington, Box 351640, Seattle, WA 98055, *epgrimit@atmos.washington.edu*

***Key Words:*** probabilistic weather forecast, mesoscale ensemble, constrained optimization, continuous ranked probability score

Bayesian model averaging (BMA) is a standard approach to statistical inference recently used to generate calibrated probability distributions for weather variables, given an ensemble of dynamical model forecasts. In previous work, a set of BMA parameters is estimated using available observation data within the entire model domain. This leads to overestimation of the forecast uncertainty in regions where the errors are consistently smaller than average and underestimation where the errors tend to be larger. This paper presents results for BMA applied with separate parameter sets for each model grid point. The local BMA parameters are estimated with training data from a neighborhood around each point. The neighborhoods are defined through optimization of the continuous ranked probability score with constraints based on both geophysical and model characteristics.

## Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts

❖ Veronica Berrocal, University of Washington, Box 354322, Department of Statistics, Seattle, WA 98195, *veronica@stat.washington.edu*

Probabilistic weather forecasts are obtained by running numerical weather prediction models with varying initial conditions and/or model parameters, resulting in ensembles of deterministic forecasts. However, forecast ensembles are often underdispersive and therefore uncalibrated. We introduce a statistical postprocessing technique, called Spatial Bayesian model averaging (Spatial BMA), to calibrate forecast ensembles of whole weather fields. Spatial BMA provides statistical ensembles of weather field forecasts that take the spatial structure of observed fields into account and honor the information contained in the original ensemble. The technique was applied to 48-h forecasts of surface temperature over the North American Pacific Northwest using the University of Washington mesoscale ensemble with good results.

## Use of Uncertainty Information in Deterministic Weather Forecasting Decisions

❖ Susan Joslyn, University of Washington, Department of Psychology, Box 351525, Seattle, WA 98195, *susanj@u.washington.edu*

***Key Words:*** psychology, decision-making, uncertainty information, weather-forecasting

The psychological research reported here investigates people's use and understanding of information about weather forecast uncertainty in deterministic forecasting decisions. A series of experiments demonstrates that such information is beneficial to threshold decisions (e.g.,. whether to post a high wind advisory), but that presentation format and wording are crucial to people's understanding.

# 404 Dose-Finding

Biopharmaceutical Section, Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

## Hypothesis Testing and Bayesian Estimation Applied to Sparse Dose Response Designs

❖ Neal Thomas, 61 Dream Lake Drive, Madison, CT 06443, *snthomas99@yahoo.com*

***Key Words:*** dose response, Sigmoid Emax model, phase II design

Application of a sigmoid Emax model is described for the assessment of dose response applied to designs containing a small number of doses (e.g., three to six). The model will be evaluated following a strategy proposed by Bretz, Pinheiro, and Branson (2005). The sigmoid Emax model is used to create several contrasts that have high power to detect trends. The model, combined with a prior distribution for its parameters, suggests methods for selecting the contrasts representing the likely dose response. Weak FWER control is achieved using multiple comparison methodology. Bayesian methods are used then to estimate the dose response curve from the sparse dosing design. The Bayesian estimation applied to the sigmoid model represents the uncertainty in model selection missed when a single, simpler model is selected from a collection of non-nested models.

## Phase I Studies of Chemotherapeutic Agents in Cancer Patients: a Review of the Designs

❖ Douglas Potter, University of Pittsburgh, Department of Biostatistics, Suite 325, Sterling Plaza, 201 North Craig Street, Pittsburgh, PA 15213, *potter@upci.pitt.edu*

***Key Words:*** phase I, continual reassessment method, dose-finding, clinical trial

The designs available for phase I dose-finding studies of chemotherapeutic agents in cancer patients will be reviewed. The designs, which have the objective of defining a "maximum tolerable dose," are based on the assumption that both efficacy and toxicity increase with dose, and thus attempt to minimize the number of patients treated at low dose and the chance patients will be treated at excessively toxic or lethal doses. The designs fall into two classes: rule-based and model-guided. Rule-based designs can always determine a reasonable maximum tolerable dose, but when model assumptions are not satisfied, many model-guided designs will not.

## Design and Analysis of Dose-Finding Studies Combining Multiple Comparisons and Modeling Procedures

❖ Frank Bretz, Novartis Pharma AG, Biostatistics, Basel, 4002 Switzerland, *frank.bretz@novartis.com*; Jose Pinheiro, Novartis Pharmaceuticals Corporation; Bjoern Bornkamp, University of Dortmund

***Key Words:*** contrast test, minimum effective dose, sensitivity analysis, proof-of-concept, dose response study, multiple testing

The search for an adequate dose involves a complex series of decisions in developing a clinically viable product. We consider a unified strategy for designing and analyzing such dose-finding studies, including the testing of proof-of-concept (PoC) and the selection of promising dose(s) for further development. The methodology combines the advantages of multiple comparison and modeling approaches, consisting of a multistage procedure. PoC is tested in the first stage using multiple comparison methods to identify a statistically significant dose-response signal. Once this is established, the best dose-response model is used for dose selection in subsequent stages. This talk will describe practical considerations related to the implementation of this methodology. We will discuss how to determine sample sizes and perform power calculations based on the PoC step.

### Optimal Dose Response Studies with Potentially Missing Observations

❖ Weng Kee Wong, University of California, Los Angeles,
Department of Biostatistics, 10833 Le Conte Ave, Los Angeles, CA
90095, *wkwong@ucla.edu*; InYoung Baek, Stony Brook University;
Wei Zhu, Stony Brook University

***Key Words:*** dose response study, binary response, l-optimality, uniform design, d-optimality, missing information

We propose optimal designs that incorporate information about potentially missing observations at the onset of the study. We assume parametric models with a binary or continuous outcome and provide single- or multiple-objective optimal designs for estimating model parameters or percentiles. The design strategy is general when it is reasonable to use a response probability function to model the probability of a missing observation across the design space. We consider Bayesian and non-Bayesian approaches and evaluate efficiencies of uniform designs under various scenarios when there are missing data and efficiencies of optimal designs when information about potentially missing observations is ignored.

### Experimental Design for Experiments with Toxicity and Efficacy Response Functions

❖ Nancy Flournoy, University of Missouri-Columbia, 146
Middlebush, Columbia, MO 65211, *flournoyn@missouri.edu*

***Key Words:*** phase II clinical trials, contingent response functions, continuation ratio model, optimal design, trinomial responses

We describe recent results on optimal designs for experiments with toxicity and efficacy response function with associated implementation and analysis issues. When toxicity causes subjects to go off-study, there are three outcomes: toxicity, efficacy with toxicity, and efficacy without toxicity. Assuming the toxicity rate and the rate of efficacy-given-no-toxicity increase with dose, we say data follows a contingent response model. We describe recent work in optimal experimental design, extending results in optimal designs for contingent response models by H. Rabie and N. Flournoy (2004). Furthermore, because these models are nonlinear, they cannot be implemented directly. Thus we also discuss sequential implementation strategies that approximate the optimal designs.

## 405 Modeling and Adjustment of Economic Time Series ●

Business and Economics Statistics Section
**Wednesday, August 9, 10:30 am–12:20 pm**

### Reference Week Adjustment of Labor Force Series with X-12-ARIMA

❖ Zhao-Guo Chen, Statistics Canada, 120 Parkdale Ave., Ottawa,
ON K1A 0T6 Canada, *chenzha@statcan.ca*; Thierno A. Balde,
Statistics Canada; Benoit Quenneville, Statistics Canada; Helen Fung,
Statistics Canada

***Key Words:*** labor force survey, seasonal adjustment, X-11-ARIMA, X-12-ARIMA, reference week, regression

The Labor Force Survey (LFS) collects information on employment, unemployment, and actual hours worked during a reference week of the months. The data are affected by events such as a sporadic holiday occurring in the reference week. Also, data are affected by the reference week being earlier or later in the month. Statistics Canada currently uses X-11-ARIMA for seasonal adjustment. These reference week effects are estimated with a regression model on the irregular component of the series obtained from a first run of X-11-ARIMA. The estimated factors are then used as permanent prior-adjustment factors in a second run of X-11-ARIMA. We are planning to switch to X-12-ARIMA for the seasonal adjustment of LFS series. We study how to set the regression model with X-12-ARIMA properly and which is convenient for future operation. We also compare the new method with the previous one.

### Issues in Identifying Easter Effects in Economic Time Series

❖ Kellie Wills, U.S. Census Bureau, 520 N. Street, SW, S317,
Washington, DC 20024, *kellie.c.wills@census.gov*

***Key Words:*** RegARIMA, holiday, Easter, X-11

To decide whether the regARIMA model for an economic time series should include an Easter effect, seasonal adjusters at the U.S. Census Bureau use the AICC test in X-12-ARIMA, as well as subject matter information. However, the AICC test may be sensitive to model issues, such as incorrect trading day decisions or outliers. Series affected by Easter should show patterns of X-11 extreme values or regARIMA outliers in March and April, particularly in years when Easter falls in March, but such patterns have not been investigated systematically. This paper will investigate how the X-12-ARIMA AICC test for Easter might be combined with analysis of X-11 extreme values and regARIMA outliers for more robust Easter effect decisions. These methods will be evaluated using simulated series with known Easter effects. The paper will then examine census retail sales and manufacturing shipments series.

### A New Time Series Model for Seasonally Adjusting Economic Data with Trend-Cycle Movement and Irregular, Sharply Pronounced Seasonal Fluctuations

❖ Stephanus Arz, Deutsche Bundesbank, Wilhelm Epstein Str 14,
Frankfurt Main, 60431 Germany, *Stephanus.Arz@bundesbank.de*

***Key Words:*** seasonal adjustment, calendar adjustment, over-adjustment, multiplicative-additive model, X-12-ARIMA

Usually, seasonal adjustment is based on time series models, which decompose an unadjusted series into the sum or the product of four unobservable components (trend-cycle, seasonal, working-day, and irregular). In the case of clearly weather-dependent output in the West German construction industry, traditional considerations lead to an additive model. However, this results in an over-adjustment of calendar effects. An alternative is a multiplicative-additive mixed model, the estimation of which is illustrated using X-12-ARIMA. Finally, the relevance of the new model is shown by analyzing selected time series for different countries.

### An ARIMA Model--Based Approach To Estimate Evolving Trading Day Effect

❖ Xichuan Zhang, Australian Bureau of Statistics, ABS House, 45
benjamin Way, Belconnen ACT, 2617 Australia, *mark.zhang@abs.gov.au*; Anna Poskitt, Australian Bureau of Statistics

***Key Words:*** trading day effect, seasonal adjustment, ARIMA, random coefficient

An important aspect of the seasonal adjustment of monthly economic flow time series is the estimation and correction for nonregular calendar variation, termed trading day effects. A static trading day effect often is assumed and estimated based on a regression-ARIMA framework in seasonal adjustment software such as X-12-ARIMA and TRAMO. However, this assumption is not always realistic. To improve seasonal adjustment quality within the regression-ARIMA framework, this paper presents a method utilizing a rolling window time-series span in conjunction with various smoothing techniques to estimate an evolving trading day effect. Quality assessment is made for the proposed method in comparison with other methods suggested in literature using simulated time series. The proposed method also is evaluated using Australian Bureau of Statistics monthly time series.

### Modeling CPS Labor Force Time Series in Selected Metropolitan Areas

❖ Jennifer Oh, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, NE, Washington, DC 20212-0001, *oh.jennifer@bls.gov*; Richard Tiller, Bureau of Labor Statistics

***Key Words:*** small area estimation, state-space models, seasonal adjustment

Time series modeling of survey estimates is an effective approach for producing small-area estimates when a long time series is available. Beginning in 2005, the Bureau of Labor Statistics began producing estimates for five metropolitan statistical areas (MSAs) using time series models. These models represent the idiosyncratic features of the CPS survey by explicitly modeling the survey error process with conventional modeling of the underlying true time series. An important benefit of this approach is that it produces seasonally adjusted series where the disturbing effects of survey errors have been suppressed. These MSA models are integrated into a nationwide system of state models. Prior to 2005, labor force estimates for five major MSAs were produced using a nonstatistical approach known as the "handbook method."

## 406 When Disaster Strikes: Responses from the Survey Community ● ✪

Section on Survey Research Methods, Section on Statisticians in Defense and National Security

**Wednesday, August 9, 10:30 am–12:20 pm**

### Conducting Surveys When Disasters Strike

❖ Rachel Harter, National Opinion Research Center, 55 E. Monroe Street, Suite 4800, Chicago, IL 60603, *Harter-Rachel@norc.org*; Judith Petty, National Opinion Research Center; Jenny Kelly, National Opinion Research Center

***Key Words:*** survey, disaster, nonresponse, redesign

Catastrophic events, whether natural disasters or human activities with tragic impact on societal affairs, disrupt the lives of survey participants and alter the plans of agencies conducting surveys. The effects on surveys can range from cancellations to redesigns to the initiation of new survey research. This paper provides an overview of the effects of disasters on surveys, illustrated with examples from NORC's experience.

### On the Use of Survey Methods in Assessing Large-Scale Human Rights Violations in Conflict Zones: Lessons-Learned from Timor-Leste and Sierra Leone

❖ Romesh Silva, Human Rights Data Analysis Group, 480 California Ave., Suite 201, Palo Alto, CA 94306, *Romesh.S@benetech.org*; Patrick Ball, Human Rights Data Analysis Group

***Key Words:*** survey methods, human rights, multiple systems estimation, conflict zones

Many methods have been used to assess the magnitude and patterns of human rights violations, most frequently via convenience samples of qualitative interviews. In recent work in Timor-Leste and Sierra Leone, parallel projects were conducted to measure past human rights abuses. Both projects used household probability survey sampling and estimation as well as multiple systems estimation techniques. This paper presents results from comparing the survey designs and estimation methods used in the projects. In particular, lessons learned are noted about how standard demographic techniques can be leveraged as a framework for estimating total conflict-related deaths.

### Impact of Gulf Hurricanes on the National Immunization Survey

❖ Kirk Wolter, National Opinion Research Center, 55 E. Monroe Street, Suite 4800, Chicago, IL 60603, *wolter-kirk@norc.org*; James Singleton, Centers for Disease Control and Prevention

***Key Words:*** sampling, survey estimation, immunization

The 2005 hurricane season produced four storms that impacted the operations of the National Immunization Survey (NIS). We will discuss the changes made in data collection operations to accommodate the storms and their aftermaths. Katrina devastated New Orleans---one of the areas targeted in the survey---and its surrounding areas and created lasting statistical issues. We will discuss the definition of the survey population, coverage of the population by the survey sampling frame, residency rules, feasibility of collecting the requisite data, missing data, and alternative estimation procedures to produce acceptable estimates of immunization coverage for New Orleans, Louisiana, and the nation. The NIS is sponsored by the CDC and designed to produce America's principle database of information about the immunization status of children ages 19 to 35 months.

### Local Data Coordination and Dissemination in Post-Katrina New Orleans

❖ Allison Plyer, Greater New Orleans Nonprofit Knowledge Works, 1600 Constance Street, New Orleans, LA 70130, *allisonp@gnonkw.org*

***Key Words:*** data dissemination, data coordination, data intermediary, Katrina, New Orleans

Before Katrina, the Greater New Orleans Community Data Center used local, state, and federal data to inform interactions between government, funders, researchers, nonprofits, and community-based organizations. Our theory was that if all stakeholders were using the same information, we could better tackle the city's many challenges. After Katrina, our audience expanded to include federal agencies and the media. We assisted their understanding of the data landscape in New Orleans and facilitated connections with local organizations. Now, population and impact estimates are being generated and updated by various researchers and government agencies. We actively scan the environment to find and assess all post-Katrina estimates and projections and publish them in a highly usable, web-based format. We'll share practical examples of data coordination and dissemination in post-Katrina New Orleans.

# 407 Statistical Issues in Diagnostic Devices Including ROC Methods ●

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

## ROC Graphs for Assessing the Ability of a Diagnostic Marker To Detect Three Disease Classes with an Umbrella Ordering

❖ Todd Alonzo, University of Southern California, 440 E. Huntington Drive, Suite 300, Arcadia, CA 91006, *talonzo@childrensoncologygroup.org*; Christos Nakas, University of the Aegean

***Key Words:*** diagnostic testing, ROC curve, non-parametric tests, ROC surface, gold standard, u-statistics

Receiver operating characteristic (ROC) curves and area under these curves are used to assess the ability of continuous diagnostic markers to classify subjects correctly as having disease. ROC surfaces and volume under the surfaces have been proposed for settings with more than two disease classes. These approaches, however, do not allow one to assess the ability of a marker to differentiate two disease classes from a third without requiring a monotone ordering of the classes. We propose the construction of an ROC graph that accommodates an umbrella ordering. It also is proposed that a summary measure for this graph can be used to summarize the classification accuracy and corresponding variance estimates can be obtained using U-statistics or bootstrap. Proposed methods are applied to a study assessing the ability of a methylation marker to classify lung specimens correctly.

## Random Effects Modeling Approaches for Estimating ROC Curves from Repeated Ordinal Tests without a Gold Standard

❖ Paul S. Albert, National Cancer Institute, 6130 Executive Blvd., Room 8136, Biometric Research Branch,, Bethesda, MD 20892, *albertp@mail.nih.gov*

***Key Words:*** diagnostic accuracy, latent class analysis, mixture models, ROC curves

Estimating diagnostic accuracy without a gold standard is an important problem in medical testing. Although there is much literature about this problem for the case of repeated binary tests, there is substantially less work for the case of ordinal tests. A noted exception is the work by Zhou et al. (2005), which proposed methodology for estimating ROC curves without a gold standard from multiple ordinal tests. An assumption in their work was that the test results are independent conditional on the true test result. We propose random effects modeling approaches that incorporate dependence between the ordinal tests. We show, through asymptotics and simulations, the importance of correctly accounting for the dependence between tests. We illustrate these modeling approaches by analyzing the uterine cancer pathology data analyzed in Zhou et al. (2005).

## Diagnostic Imaging Procedures: Defining and Analyzing Test Results To Account for Unknown Disease Loci

❖ Gene Pennello, U.S. Food and Drug Administration, 1350 Piccard Drive, HFZ-550, Division of Biostatistics, Rockville, MD 20850, *gene.pennello@fda.hhs.gov*; Brandon D. Gallas, U.S. Food and Drug Administration

***Key Words:*** partial gold standard, conditional independence, verification bias, Bayesian methods, free-response receiver operating characteristic curve, latent class

Diagnostic imaging is a vital tool in helping to identify and biopsy tissue suspicious of being diseased. Examples are use of imaging in colposcopy and mammography to search the cervix and breast for cancerous loci. At the subject level, common definitions for the diagnostic test result do not account properly for unknown disease loci. This causes logical problems, such as defining a diseased subject as a false positive because all of the biopsies are negative. The problems are apparent in studies with and without complete disease ascertainment (i.e., follow-up and nonfollow-up studies). We consider alternative definitions for subject-level diagnostic test results in an attempt to address unknown disease loci. Under these definitions, we explore statistical methods of analysis that can be used to estimate diagnostic performance (e.g., sensitivity, specificity, FROC).

## A Model-Free Approach to Combining Diagnostic Markers

❖ Ruth Pfeiffer, National Cancer Institute, 6120 Executive Blvd., EPS 8030, Rockville, MD 20852, *pfeiffer@mail.nih.gov*; Efstathia Bura, The George Washington University

***Key Words:*** dimension reduction, likelihood ratio, random matrix, save, singular value decomposition, NHANES III

A popular summary measure of the discriminatory ability of a continuous diagnostic marker for binary disease outcomes is the area under the ROC curve. We will present a model-free approach to combining several markers into a composite diagnostic test. Using sufficient dimension reduction techniques, we will replace the predictor vector with a lower-dimensional version obtained through linear transformations of markers that contain sufficient information for regression of the predictors on the outcome. We will combine the linear transformations using their asymptotic properties into a scalar diagnostic score via the likelihood ratio statistic and assess its performance by the ROC curve. We will derive the asymptotic distribution of the singular values of consistent estimates of a normally distributed random matrix.

## Statistical Issues in Diagnostic Devices Including ROC Methods

❖ R. Lakshmi Vishnuvajjala, U.S. Food and Drug Administration, 1350 Piccard Drive, Rockville, MD 20850, *rlv@cdrh.fda.gov*

***Key Words:*** diagnostic devices, PMA, gold standard, review checklist

The diagnostic devices branch in the Division of Biostatistics at CDRH developed a checklist of items to look for in diagnostic submission. We initially thought of it for pre-market applications, but found it to be useful for 510(k)s as well. Because the statistical methodology for diagnostic devices is not as developed as it is for therapeutic devices, we thought it useful for manufacturers also.

## 408 Innovative Uses of Longitudinal Panels, Information Documents, and Time-Series Analysis to Study the Impact of the U.S. Tax System ●

Section on Government Statistics

**Wednesday, August 9, 10:30 am–12:20 pm**

### Analysis of the Distributions of Income, Taxes, and Payroll Taxes via Cross-Section and Panel Data

❖ Thomas Petska, Internal Revenue Service, P.O. Box 2608, Washington, DC, 20013-2608, *thomas.b.petska@irs.gov*; Michael Strudler, Internal Revenue Service; Ryan Petska, Ernst & Young LLP

***Key Words:*** income distribution, tax burden, administrative records

Administrative records from tax returns are a reliable source of information for the distribution of personal income and tax burdens. This paper is the seventh in a series examining trends in the distribution of individual income and taxes based on a consistent and comprehensive measure of income derived from individual income tax returns. In this analysis, we examine changes in the income distribution and tax burdens between 1996--2003, including the effects of federal income and social security payroll taxes on the after-tax distribution of income. In addition, we estimate Lorenz curves and Gini coefficients to see how income inequalities have changed. For this study, we use a cross section of returns to study how incomes and taxes change for individuals. We also examine panel data using an imbedded subset of the cross-sectional returns for 1998--2003.

### Social Security Taxes, Social Security Benefits, and Social Security Benefits Taxation: 2002

❖ Peter J. Sailer, Internal Revenue Service, 500 N. Capitol Street, NW, 5th Floor, Washington, DC 20002, *peter.j.sailer@irs.gov*; Evgenia Lomize, Internal Revenue Service

***Key Words:*** social security, taxes, benefits

By combining individual tax returns (Form 1040) and information returns (e.g., Forms W-2 and 1099-SSA) in one database, the Statistics of Income (SOI) Division has made it possible to study trends in social security tax payments, social security benefits, and the taxation of those benefits. In this paper, we will show the impact of the social security system on individuals in various income classes and those in different age and gender groups. For tax return filers, benefits and taxes can be classified by the total adjusted gross income of the filing unit. For nonfilers, classification by total income reported on all types of information documents filed for the individual is possible.

### Longitudinal Analysis of the Earned Income Tax Credit

❖ Karen Masken, Internal Revenue Service, 1111 Constitution Ave. NW, RAS R CAM, Washington, DC 20024, *karen.c.masken@irs.gov*

While it is known that there is significant turnover in EITC claimants from one year to the next, the reasons for this are not understood. To better understand why taxpayers move in and out of the EITC population, the Office of Research is conducting a nine-year longitudinal study. In addition to tracking taxpayers who claimed EITC in at least one of the last nine years, the study will track the children claimed. This paper presents the analysis of taxpayer patterns in claiming EITC during the study period and the pattern of who is claiming the child for the same period. In addition to looking at the patterns, the study attempts to use administrative data to identify reasons why taxpayers move in and out of the EITC population and why children are claimed by different taxpayers from one year to the next.

### The 1999--2003 Individual Income Tax Return Panel: a First Look at the Data

❖ Michael Weber, Internal Revenue Service, P.O. Box 2608, Washington, DC 20013-2608, *michael.e.weber@irs.gov*

***Key Words:*** panel, longitudinal, income tax

Virtually all the data published by the Statistics of Income Division of the IRS are based on cross-sectional samples. Although SOI produced a limited number of longitudinal samples in the past---namely the 1979--2004 Continuous Work History Panel, the 1985--2001 Sales of Capital Assets Panel, and the 1987--1986 Family Panel---it has published only a few tabulations from only one of these studies. The latest SOI longitudinal study is the 1999--2003 Individual Income Tax Return Panel, the first attempt at developing a series of longitudinal tabulations based on individual income tax returns that can be published by SOI on a regular basis.

### Constructing a Panel of Income and Estate Tax Data for Wealthy Individuals: Creativity and Compromise

❖ Barry Johnson, Internal Revenue Service, P.O. Box 2608, Washington, DC 20013-2608, *Barry.W.Johnson@irs.gov*; Lisa Schreiber, Internal Revenue Service

***Key Words:*** panel, unit non-response, non-sampling error, sample weight adjustment

The Statistics of Income Division of the IRS created a longitudinal panel of income tax filers beginning in 1987, including returns for the primary filing unit and dependents. Beginning in 1995, SOI began collecting data from federal estate tax returns for individuals in the original 1987 income panel. This paper will document efforts to construct and use the income/estate tax panel. Specifically, it will examine the effects of missing returns (a form of unit nonresponse), data quality issues (nonsampling error), and changes in the filing tax unit---all of which pose complex sample weight issues. Preliminary estimates from this file will be used to illustrate the effects of decisions made in constructing the panel, with particular attention given to the trade-off between longitudinality and population coverage that is often a problem in panel datasets.

## 409 Statistical Phylogenetics

● ✪

IMS, Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### A Model of AFLP Evolution and Its Use in Bayesian Estimation of Phylogenies

❖ Ruiyan Luo, University of Wisconsin-Madison, Medical Science Center, 1300 University Ave., Madison, WI 53706, *rluo@stat.wisc.edu*; Bret Larget, University of Wisconsin-Madison

***Key Words:*** phylogeny, MCMC, Bayesian, Indel

Amplified fragment length polymorphism (AFLP) markers are a type of genetic marker that are useful to biologists. Some biological mechanisms, like substitution process and insertion-deletion process, affect the presence or absence of AFLP markers. Under such processes, the markers are not independent, as the loss of a marker is accompanied by the appearance of two new ones. In previous work, we studied AFLP data based on the substitution process only. Now, we consider both the substitution and indel process and build a model to characterize the gain and loss of AFLP markers. With this model, we construct a Bayesian procedure, then use MCMC method to infer the distribution of tree topologies. We compare the result to the phylogenies obtained by other methods, which assume the independence of markers. We also check the reasonability of the assumptions made in the model.

### A Random Duplication/Deletion Model in Genome Rearrangement

❖ Soowan Sohn, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, *ssohn@wisc.edu*; Bret Larget, University of Wisconsin-Madison

***Key Words:*** genome rearrangement, gene duplication/deletion, phylogeny, MCMC

Variation in the arrangements of genomes is an alternative source of information for estimating evolutionary relationships among species. Existing statistical approaches for the analysis of genome arrangement data are based on the mechanism of gene inversion to rearrange genomes. However, the molecular evolution literature suggests a process of gene duplication and subsequent deletion is an important mechanism for genome rearrangement. We develop a model for gene duplication/deletion in which a set of neighboring genes is duplicated before one of the two copies of each gene, selected at random, is deleted. This process is related to a card-shuffling model. We develop a Bayesian method to infer a phylogeny under a gene duplication/deletion model for genome rearrangement. We use the Markov chain Monte Carlo Method to compute posterior probabilities.

### Spatially Smoothed Change-Point Processes for Phylogenetic Mapping of Recombination Hot Spots

❖ Vladimir N. Minin, University of California, Los Angeles, David Geffen School of Medicine, Biomathematics, Los Angeles, CA 90095, *vminin@ucla.edu*; Marc A. Suchard, University of California, Los Angeles; Karin S. Dorman, Iowa State University; Fang Fang, Iowa State University

***Key Words:*** change-point model, MCMC, recombination, HIV, Gaussian Markov random fields, phylogenetics

We will present a Bayesian framework for inferring spatial preferences of recombination from multiple putative recombinant nucleotide sequences. The dual multiple changepoint model for phylogenetic recombination detection resides at the lowest level of our hierarchy under the umbrella of a common prior on breakpoint locations. The hierarchical prior allows for information about spatial preferences of recombination to be shared among individual datasets. To overcome the sparseness of breakpoint data, dictated by the modest number of available recombinant sequences, we a priori impose a biologically relevant correlation structure on recombination location log-odds via a Gaussian Markov random field hyper-prior. We analyze 42 HIV gag recombinants and identify a recombination hot spot in the Capsid gene.

### Reconstructing Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions

❖ Liang Liu, The Ohio State University, *liuliang@stat.ohio-state.edu*; Dennis K. Pearl, The Ohio State University

***Key Words:*** coalescent theory, species tree

In this paper, we propose a Bayesian hierarchical model to estimate the phylogeny of a group of species using multiple estimated gene tree distributions such as those that arise in a Bayesian analysis of DNA sequence data. The whole process can be represented as a two-step Markov chain, from species tree to gene tree and from gene tree to DNA sequences. Simulation study was performed to examine the behavior of posterior probability of the true species tree. The result indicated that the posterior probability of the true species tree had strong positive correlation with the number of genes and the proportion of gene trees matching the true species tree. Our model was applied to a real dataset of yeast. The data had 106 genes and eight taxa. The posterior distribution of species trees was found to be highly concentrated around a single, well-supported tree.

### Reconstructing Evolutionary Trees Using Amino Acid Substitution Models that Allow Rate Variation to Depend on Spatial Location

❖ Xueliang Pan, The Ohio State University, 3536 Chowning Court, Columbus, OH 43220, *xpan@stat.ohio-state.edu*; Dennis K. Pearl, The Ohio State University; Liang Liu, The Ohio State University; Dennis J. Pollack, The Ohio State University

***Key Words:*** phylogeny, Bayesian model, site variation, crystallographic structure

Bayesian inference of phylogeny using MCMC methods has been used increasingly over the last decade. As many researchers recognize that sites do not evolve independently at the same rate, more realistic models that allow site variation are now the norm. The most used is the discrete gamma model proposed by Yang. In this study, a model allowing site variation based on its spatial location is proposed. The amino acid sites were partitioned according to their spatial location, obtained from the crystallographic structure; the rate variation of each partition was estimated by Mrbayes. The convergence and quality of this model were evaluated and found to be superior to the gamma model in terms of computational time and likelihood. This study is based on a dataset regarding phosphoglycerate kinase (PGK), an enzyme important in fermentation that is found in species from all kingdoms.

## 410 Bayesian Spatial Models ● ✪

Section on Bayesian Statistical Science

**Wednesday, August 9, 10:30 am–12:20 pm**

### Hierarchical Bayes Estimation of Response Rates with Spatial Correlations

❖ Xiaoming Gao, Missouri Department of Conservation, 1110 S. College Ave., Columbia, MO 65203, *Sherry.Gao@mdc.mo.gov*; Chong He, University of Missouri-Columbia; Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia

***Key Words:*** attitude survey, response rate, Bayesian hierarchical model, spatial correlation, Gibbs sampling

A Bayesian hierarchical logistic linear mixed model is used to estimate response rates at the sub-area level for hunter attitude surveys. The model includes fixed mailing effects, hunter's characteristics, random geographic effects, and spatial correlations between neighboring sub-areas. The computation is done by Gibbs sampling and adaptive rejection sampling techniques. The method is illustrated using data from the 2001 Deer Hunter Attitude Survey in Missouri. The result shows that not only hunter's age and date of hunting permit purchase significantly influence response rates, but there are significant mailing differences and strong spatial correlations of response rates among counties.

### Reference Priors for Gaussian Processes with Spatial Correlation Structure

❖ Mi Hyun Lee, Virginia Polytechnic Institute and State University, 208 Turner Street, Apt. 220, Blacksburg, VA 24060, *mhot1977@yahoo.com*; Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia

***Key Words:*** Gaussian process, spatial correlation function, reference prior

In many of environmental applications, Gaussian random processes---which are determined by mean and covariance functions---are commonly used to model spatial data by specifying the spatial correlation structure of the covariance functions. In this paper, Bayesian analysis on the unknown mean and covariance parameters of the spatial Gaussian random fields are proposed. The reference priors for the unknown covariance parameters are developed when the mean function is assumed to be linear and the regression parameters in the mean function are considered to be nuisance parameters. Under the restricted situations, Berger, De Oliveira, and Sanso (2001) and Paulo (2005) derived the reference priors for the unknown covariance parameters of the spatial Gaussian random processes. The reference priors are studied under more general situation in this paper.

### Bayesian Spatial-Temporal Smoothing of Cancer Mortality Rates

❖ Gentry White, University of Missouri-Columbia, 146 Middlebush, Columbia, MO 65211, *gwhite@stat.missouri.edu*; Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia; Paul Speckman, University of Missouri-Columbia

***Key Words:*** Bayesian, thin-plate splines, spatio-temporal, cancer mortality rates

The high-quality data on cancer available covers a sufficient span of time for the consideration of temporal trends in incidence and mortality. Recent efforts toward more in-depth analysis of the data using spatial and temporal modeling include Clayton and Kaldor (1987)---who used a conditional auto-regressive (CAR) prior for the spatial effects---and van der Linde et al. (1995)---who used a thin-plate, spline-based prior for spatial effects. The model presented here implements a semiparametric spatio-temporal prior using a thin-plate spline prior for spatial effects and an intrinsic auto-regressive prior for temporal trends. A prior whose covariance matrix is defined as the Kronecker product of the spatial and temporal priors is introduced. Results are compared to a similar joint model, using a CAR prior for the spatial effects.

### Spatially Adaptive Bayesian Thin-Plate Splines

❖ Yu Yue, University of Missouri-Columbia, 2303 Whitegate Drive, APT 2B, Columbia, MO 65202, *yytc9@mizzou.edu*

***Key Words:*** function estimation, spatial adaptation, thin-plate splines, Gibbs sampling

There is a well-known Bayesian interpretation for function estimation by spline smoothing. In this paper, we derive a Bayesian hierarchical representation for a discretized thin-plate spline with a spatially adaptive smoothness penalty. We show how this allows the fitted model to adapt to various degrees of smoothness in the data. We propose a convenient form of the discretized prior to adaptively model the precision parameter and discuss efficient computational algorithms for implementing the Gibbs sampler.

### Bayesian Smoothing of Density Estimation via Hazard Rates

❖ Luyan Dai, University of Missouri-Columbia, 2405 Whitegate Drive, Apt 2F, Columbia, MO 65202, *ld9n9@mizzou.edu*

***Key Words:*** Bayesian smoothing, density estimation, IAR process priors, hazard rates, adaptive variance components

We consider estimating a bounded density by Bayesian smoothing techniques via hazards rates. The well-known intrinsic autoregressive process prior helps construct frequentists' smoothness by Bayesian methods. Instead of directly model on density, we investigate several transformations-based hazards for the purpose of estimation. Beyond this, we propose adaptive variance components model based on IAR priors to improve estimation. Under the proposed model, the Bayesian computation can be realized via MCMC and implemented with Gilk's adaptive sampler and ratio of uniform sampler. In addition, we present a discussion and comparison among priors, including the pareto prior, inverse-gamma prior, and adaptive variance component priors. We compare the density estimation under different models.

## 411 Multidimensional Scaling and Manifold Learning

Section on Statistical Computing, IMS, Section on Statisticians in Defense and National Security

**Wednesday, August 9, 10:30 am–12:20 pm**

### Parametric Mapping (PARAMAP): an Approach to Nonlinear Mapping

❖ Ulas Akkucuk, Bogazici University, IIBF Department of Management, Bebek, Istanbul, 34342 Turkey, *ulas.akkucuk@boun.edu.tr*

***Key Words:*** nonlinear mapping, parametric mapping, dimensionality reduction, measures of agreement, paramap

Dimensionality reduction techniques are used for representing higher-dimensional data by a more parsimonious and meaningful lower-dimensional structure. Such methods have potential application to visualizing and interpreting high-dimensional data. In this paper, we will study Carroll's Parametric Mapping (PARAMAP). The PARAMAP algorithm relies on iterative minimization of a cost function measuring "smoothness" of the mapping from the low- to the high-dimensional space. We will develop a measure of congruence based on preservation of local structure between the input data and the mapped low-dimensional embedding and demonstrate the application of PARAMAP to various sets of nonlinear manifolds, including points located on the surface of a sphere, "Swiss Roll Data," and truncated spheres.

### Metric MDS to Surfaces

❖ David Johannsen, Naval Surface Warfare Center, Code B10, 17320 Dahlgren Road, Dahlgren, VA 22448, *david.johannsen@navy.mil*; Jeffrey L. Solka, Naval Surface Warfare Center

We will present our work on metric MDS to (closed and orientable) surfaces equipped with a (uniquely determined) compatible constant curvature metric. We will describe our algorithm for performing the MDS via the minimization of a loss function. We also will describe briefly the obvious generalization of these results to higher dimensions (i.e., MDS to space forms). Finally, we will discuss the issues of visualization (in the case of surfaces) and describe how one performs the minimization (via gradient descent on a Riemannian manifold).

### Local Multidimensional Scaling: a Nonlinear Dimension Reduction Method for Data Visualization

❖ Lisha Chen, University of Pennsylvania, 3730 Walnut Street, 400 Huntsman Hall, Philadelphia, PA 19104, *lisha@wharton.upenn.edu*; Andreas Buja, University of Pennsylvania

***Key Words:*** nonlinear dimension reduction, MDS, graph layout

In recent years, there has been a marked resurgence of interest in nonlinear dimension reduction methods. Among new proposals are “Local Linear Embedding” and “Isomap.” Both use local neighborhood information to construct a global low-dimensional embedding of a hypothetical manifold near which the data fall. In this talk, we will introduce a new nonlinear dimension reduction method called Local Multidimensional Scaling, or LMDS. Like the other proposals, LMDS uses only local information from user-chosen neighborhoods, but its novel feature is that it uses ideas from the area of “graph layout.” Our methods generally do a superior job in flattening the famous Swiss roll and some of the other illustrative datasets used in the LLE and Isomap papers. A meta-criterion will be introduced to compare the configurations generated by different methods.

### Classical Multidimensional Scaling and Laplacian Eigenmaps

❖ Michael W. Trosset, The College of William & Mary, Department of Mathematics, P.O. Box 879, Williamsburg, VA 23187-8795, *trosset@math.wm.edu*

***Key Words:*** classical multidimensional scaling, Laplacian eigenmaps, isomap, embedding, nonlinear dimension reduction, manifold learning

An eigenmap is a data matrix constructed from the eigenvectors of a centered inner product matrix, as in classical multidimensional scaling (CMDS or principal coordinate analysis). Eigenmaps are popular in manifold learning (e.g., Isomap applies CMDS to the shortest path distances of a certain graph, while Laplacian eigenmaps are constructed from graphs via centered inner product matrices known as graph Laplacians). I will describe several relations between CMDS and Laplacian eigenmaps and explore the implications of these relations for their proper use in embedding and nonlinear dimension reduction.

### Manifold Learning and Dimensionality Reduction for Classification

❖ Alfred Hero, University of Michigan, Department of EECS, 1301 Beal St, Ann Arbor, 48109-2122, *hero@umich.edu*; Raviv Raich, University of Michigan; Jose Costa, California Institute of Technology

***Key Words:*** Laplacian eigenmap, non-parametric classification, nonlinear PCA

In this talk, we present a modified Laplacian eigenmap (LE) dimensionality reduction method that explicitly incorporates classification constraints. This method is called classification-constrained dimensionality reduction (CCDR) and is designed to produce a reduced-dimension feature space that optimally trades off classification error for least squares manifold fitting error. We apply CCDR to several datasets, including LANDSAT hyperspectral satellite imagery. We demonstrate that CCDR reduction of the feature space can produce significant improvements in misclassification error. For example, in the case of the LANDSAT data application of the k nearest neighbors (k-NN) classifier to the CCDR, reduced-dimension feature space gives 10% lower misclassification error than k-NN applied to the full-dimensional data.

## 412 Multivariate Control Charts and Other Related Topics ●

Section on Quality and Productivity, Section on Physical and Engineering Sciences

**Wednesday, August 9, 10:30 am–12:20 pm**

### Single Variables Control Chart: an Overview

❖ Smiley Cheng, University of Manitoba, Department of Statistics, Winnipeg, MB R3T2N2 Canada, *smiley_cheng@umanitoba.ca*; Keoagile Thaga, University of Botswana

***Key Words:*** single chart, statistical process control, variables data, autocorrelated processes

Control charts are used widely in industries to monitor a process for quality improvement. When dealing with variables data, we usually employ two control charts to monitor the process location and spread. We will give an overview of control charts proposed in the last decade or so in an effort to use only one chart to monitor both process location and spread simultaneously. Two approaches have been advocated for using one control chart for process monitoring. One approach plots two quality characteristics in the same chart, while the other uses one plotting variable to represent the process location and spread.

### Multivariate Process Control for Improving Detection of Out-of-Control Conditions

❖ Amit Mitra, Auburn University, College of Business, Suite 516, Auburn, AL 36849-5240, *mitra@business.auburn.edu*

***Key Words:*** process control, multivariate processes, control charts, simulation

Monitoring of processes that involve multiple variables through control charts traditionally have used a Mahalanobis distance measure in the form of a T-squared control chart. The variables used normally have been the process output variables. However, the process output variables could be impacted by two categories of variables: process regulatory or input variables and the uncontrollable noise factors. While the process input variables are usually assumed to be fixed, this is not the case in practice. Hence, we propose the inclusion of these process input variables in the analysis of multivariate process control. For improved detection of out-of-control conditions, it is of interest to determine which of these categories of variables are potential causes. Through a simulation study, certain performance measures are investigated.

### The Multivariate Exponentially Weighted Moving Average

❖ Steve Rigdon, Southern Illinois University, , *srigdon@siue.edu*; Nicole Munden, University of Missouri

***Key Words:*** multivariate control chart, EWMA, quality, change point

The multivariate exponentially weighted moving average (MEWMA) control chart is now a widely used technique for monitoring a production process with many variables. It was designed to signal when a process is no longer operating in control. While doing this effectively, the chart was not designed to tell the operator when the process changed. We discuss the use of the chart to determine when the process mean has changed. The proposed method is compared to other change-point estimators.

### Monitoring Multivariate Process Variability for Individual Observations

❖ Baiyau Yeh, Bowling Green State University, Department of Applied Statistics, and Operations Research, Bowling Green, OH 43403, *byeh@cba.bgsu.edu*

***Key Words:*** multivariate variability, individual observations, exponentially weighted, trace, mean squared deviation, moving variances

Most of the existing control charts for monitoring multivariate process variability are based on subgroup sizes greater than one. In many practical applications, however, only individual observations are available, and the usual control charts are not applicable in these cases. In this paper, two new control charts are proposed to monitor multivariate process variability for individual observations. The proposed control charts are constructed based on the traces of the estimated covariance matrices derived from the individual observations. When there is only one quality characteristic, these two charts respectively reduce to the exponentially weighted mean squared deviation and exponentially weighted moving variance charts. It is shown, based on the simulation studies, that the proposed charts are superior in detecting increases in variance and changes in correlation.

## 413 Assessing Student Retention of Essential Statistical Concepts, Issues, and Topics ●

Section on Statistical Education

**Wednesday, August 9, 10:30 am–12:20 pm**

### Assessing Student Retention of Essential Statistical Concepts: Issues and Topics

❖ Leonard Gaines, Empire State Development, 30 S. Pearl Street, Albany, NY 12245, *lgaines@empire.state.ny.us*; ❖ Albyn Jones, Reed College, 3203 SE Woodstock Blvd., Portland, OR 97202, *jones@reed.edu*; ❖ Deborah Rumsey, The Ohio State University, Department of Statistics, Cockins Hall 1958 Neil Avenue, Columbus, OH 43210, *rumsey@stat.ohio-state.edu*; ❖ Jessica Utts, University of California, Davis, Department of Statistics, One Shields Ave, Davis, CA 95616, *jmutts@ucdavis.edu*; ❖ Karen Kinard, Tallahassee Community College, 3440 Thresher Drive, Tallahassee, FL 32312, *kinardk@tcc.fl.edu*

***Key Words:*** assessing student retention,

Do former students still know what you wanted them to know about statistics? The ASA's Section on Statistical Education has made important contributions toward the improvement of pedagogy over the years. The GAISE Report discusses how students should learn and be taught. The ARTIST project assesses statistical literacy, reasoning, and thinking for introductory courses. What still needs to be addressed, however, is what students are actually retaining. What could they actually be conversant in, or, better yet, actually apply at work or in their daily lives? The panel will address this so a broad-based general perspective is developed along with considerations of particular discipline-based knowledge that must be learned and retained. Once this is established, discussion turns toward how to best assess such retained knowledge. Audience participation is encouraged.

## 414 Unit Nonresponse in Surveys III ●

Section on Survey Research Methods

**Wednesday, August 9, 10:30 am–12:20 pm**

### A Comparison of a Model-Assisted Estimator and a Model-Based Estimator under Ignorable and Nonignorable Nonresponse

❖ Jill A. Dever, University of Maryland, 1218 Lefrak Hall, Joint Program in Survey Methodology, College Park, MD 20742, *jdever@survey.umd.edu*; Richard Valliant, University of Michigan

***Key Words:*** modified generalized regression estimator, best linear unbiased predictor, non-ignorable nonresponse

Theory suggests that survey statisticians use model-assisted estimators with non-ignorable sampling designs (due to non-ignorable nonresponse) and model-based estimators otherwise. What is the recommended approach when the survey is some variant of these conditions? Through a simulation study, we compared the merits of a modified generalized regression estimator (mGREG) with a best linear unbiased predictor (BLUP) for a population total under various design conditions. We provide two recommendations. (1) For ignorable designs, a BLUP should be used for continuous variables with (approximately) symmetric marginal distributions and an mGREG used for highly skewed distributions. (2) An mGREG is preferred to a BLUP for non-ignorable designs with a sample balanced on the first and second population moments. The difference in the estimators is related to the overall nonresponse rate.

### Modeling Nonresponse Adjustment Factors

❖ Hee-Choon Shin, National Opinion Research Center, 55 E. Monroe Street, Chicago IL 60603, IL 60603, *shinh@uchicago.edu*

***Key Words:*** weighting, nonresponse adjustment, mixed modeling

The most common method of adjustment for unit nonresponse is weighting, where respondents and nonrespondents are classified into adjustment cells based on some covariates known for all units in the sample, and a nonresponse adjustment factor is computed for all cases in each cell proportional to the inverse of the response rate in the cell. The estimated response rates depend on the overall cell structure. We will model the response rate in each cell as a function of covariates and uncertainty due to the arbitrary nature of cell structure. Empirical results will be presented.

### Use of Propensity Scores To Estimate and Adjust Nonresponse Bias in Complex Surveys

❖ Leigh Harrod, Oregon State University, P.O. Box 1032, Corvallis, OR 97339, *harrod@science.oregonstate.edu*; Virginia M. Lesser, Oregon State University

***Key Words:*** propensity scores, nonresponse, missing data, nonignorable missingness, not missing-at-random (NMAR), subsampling

Propensity score methodology has been applied in the analysis of survey data to account for differences in covariate distributions between respondents and nonrespondents. When a subsample of nonrespondents is obtained, response propensity scores may be used to account for a nonignorable missing-data mechanism. For a complex survey of elk hunters and a binary outcome, we discuss testing the missingness mechanism, adjusting estimates for nonresponse bias, and model selection techniques.

### Are Refusal Conversions Different from Willing Respondents on Demographic, Cardiovascular, and Sensitve Items? National Health and Nutrition Examination Survey, 1999--2002

❖ Margaret Carroll, National Center for Health Statistics, 3311 Toledo Road, Room 4413, Hyattsville, MD 20782, *mdc3@cdc.gov*; Yinong Chong, National Center for Health Statistics

***Key Words:*** non-response, sample weight, refusals

The National Health and Nutrition Examination surveys (NHANES) constitute a series of highly stratified multistage probability samples representative of the US noninstitutionalized population. Through these surveys a wide battery of health related data are collected during a home interview and MEC exam. NHANES has maintained high response rates thus minimizing the likelihood of non-response bias. During NHANES 1999-2002 a substantial number of converted refusals have been identified for the first time in an NHANES survey. Focusing on adults 20 years and older participating in this survey, this paper will 1) compare converted refusals at the interview level(n=2367) and those at the exam level(n=913) to willing respondants(n=7011) with respect to selected demographic, cardiovascular risk and sensitive items and 2) assess the impact of converted refusals on survey estimates.

### Response Process Models for Unit Nonresponse Adjustment

❖ Courtney Kies-Bokenkroger, Iowa State University, 3915 Brickman Ave., Ames, IA 50010, *cdkies@iastate.edu*; Sarah M. Nusser, Iowa State University

***Key Words:*** unit nonresponse, response process, response propensity

It is desirable to reduce the potential for nonresponse bias by applying adjustments during the estimation process. One approach to adjusting for unit nonresponse is to include the inverse of the probability of obtaining a response in the analysis weight for each responding unit. We explore models for estimating the response probability for a sample unit based on response process outcomes. To estimate the response probability for an individual, we develop a model for the probability of proceeding through a sequence of contact attempts. We derive estimators for model parameters and explore the properties of population parameter estimators that depend on the inverse of the estimated response probability.

### A Nonresponse Bias Analysis To Inform the Use of Incentives in Multistage Telephone Surveys

❖ Benjamin Skalland, National Opinion Research Center, 55 E. Monroe Street, Ste 4800, Chicago, IL 60603, *skalland-benjamin@norc.org*; Kirk Wolter, National Opinion Research Center; Hee-Choon Shin, National Opinion Research Center; Stephen Blumberg, National Center for Health Statistics

***Key Words:*** survey non-response, bias, incentives, NS-CSHCN

Increasingly, incentives are used to raise survey response rates. In a multistage telephone survey, incentives may be offered at different stages of the interview - e.g. before contact, after contact but before screening, or after screening. In this paper, we discuss how a non-response bias analysis can be used to identify the points in the interview where incentives might have the greatest effect on bias reduction. We also discuss the assumptions necessary for this approach and their plausibility. An example is then given using data from the National Survey of Children with Special Health Care Needs (NS-CSHCN), a multistage survey conducted by NORC as part of the National Center for Health Statistics' State and Local Area Integrated Telephone Survey (SLAITS) mechanism.

## 415 Sample Survey Quality V ●

Section on Survey Research Methods

**Wednesday, August 9, 10:30 am–12:20 pm**

### Efficiently Limiting Census Errors When Quality Control Parameters Range Freely

❖ Glenn Wolfgang, U.S. Census Bureau, 20 Valley Place, Huntingtown, MD 20639, *glenn.s.wolfgang@census.gov*

***Key Words:*** quality management, acceptance sampling, inspection efficiency

With increased focus in the Bureau of the Census on applying project management and Capability Maturity Model Integrated techniques, quality management and measurement become ever more important. Production tasks such as data keying and response coding are suited to acceptance sampling quality control plans, where, given three parameters (lot size, sample size, and an acceptance number), one can project an average outgoing quality limit, interpreted as an approximate outgoing error limit. Typically, all lots are created with a set of parameters chosen to produce a specified error limit. This paper reports quality control operations designed for enhancement in two ways: to assure a particular outgoing error limit when lot size and sample size vary and to optimize efficiency, minimizing inspection workloads using estimates of incoming process error.

### Precision of Compositional Data in a Stratified Two-Stage Cluster Sample: Comparison of the Swiss Earnings Structure Survey 2002 and 2004

❖ Monique Graf, Swiss Federal Statistical Office, Espace de l Europe 10, Neuchatel, 2010 Switzerland, *monique.graf@bfs.admin.ch*

***Key Words:*** complex survey, compositional data, linearization, confidence domain, coefficient of variation

Precision of released figures is not only an important quality feature of official statistics, it is also essential for a good understanding of the data. In this paper we show a case study of how precision could be conveyed if the multivariate nature of data needs to be taken into account. In the official release of the Swiss Earnings Structure Survey, the total salary is broken down into several wage components. For surveys 2002 and 2004, we first investigate the incidence of components and then follow Aitchison's approach for the analysis of compositional data, which is based on logratios of components. Different multivariate analyses of the compositional data are performed and compared between the years, whereby the wage components are broken down by economic activity classes. Then we propose a number of ways to assess precision.

## Independent School Survey Coverage Study

❖ Dedrick Owens, U.S. Census Bureau, 8700 Silver Hill Road, Mail Stop 8700, Suitland, MD 20233, *dedrick.l.owens@census.gov*

***Key Words:*** independent schools, vendor lists, coverage, costs

The U.S. Census Bureau updates its list of schools for the independent school list frame every two years. Sources from which schools are obtained include Catholic Diocese, local education agencies, non-Roman Catholic religious institutions, and the telephone book Yellow Pages. The Yellow Pages provide schools often missed by other sources. This study investigates cost and quality considerations associated with using a commercially provided list (vs the Yellow Pages) to identify independent kindergarten and elementary/secondary schools. Anecdotal evidence suggests the Yellow Pages procedure is costly relative to use of a vendor list, and statistical testing found no significant difference in coverage. This paper summarizes observed evidence and study results to make the case for using vendor data for school list frame development.

## Cluster Analysis for Outlier Detection and Its Application in a Large-Scale Survey

❖ Jianqiang Wang, Iowa State University, Department of Statistics, 204 Snedecor, Ames, IA 50011, *jqwang@iastate.edu*; Jean D. Opsomer, Iowa State University

***Key Words:*** hierarchical agglomerative clustering, outlier detection, distance measures, survey data collection

Cluster analysis is a popular data mining tool which helps researchers explore the structure of multi-dimensional data, find special groups in populations and seek associations between individual units. It can also be applied to detect unusual points in data. The National Resources Inventory is a longitudinal survey of natural resources information on nonfederal land in the US. One possible problem encountered during NRI data collection and processing is the existence of unusual observations and outliers. These observations need to be identified and evaluated for correctness, in order to ensure the quality of the NRI data. An exploratory study is conducted to investigate the use of clustering approaches for outlier detection in NRI. The performance of different hierarchical clustering methods is compared regarding their ability to isolate artificially constructed outliers.

## Using Evaluations To Plan and Integrate Survey Programs

❖ Shawna Waugh, Energy Information Administration, 3713 Evans Trail Way, Beltsville, MD 20705, *shawna.waugh@eia.doe.gov*

***Key Words:*** survey quality, evaluation, documentation, establishment surveys, forms design

Planning and integrating of survey instruments is required for survey programs, including the Petroleum Marketing Program sponsored by the Energy Information Administration. Throughout the survey phases---planning, data collection, processing, and dissemination---it is critical to ensure consistency of data collected for 11 surveys in this program. Recommendations from an evaluation of the program were used to make revisions to the letters, forms, instructions, and technical notes for this survey program. The survey evaluation provided numerous recommendations, including suggestions improving consistency in the survey instruments and documentation. This paper will focus on "best practices" for planning and implementing a family of surveys.

## Modeling Nonsampling Errors in Agricultural Surveys

❖ James Gentle, George Mason University, University Drive, Fairfax, VA 22030-4444, *jgentle@gmu.edu*; Charles R. Perry, National Agricultural Statistics Service; William Wigton, National Agricultural Statistics Service

***Key Words:*** nonsampling errors

Nonsampling errors present major problems in sample surveys. The existence of nonsampling errors often becomes apparent only when the results of two surveys disagree in excess of what could be accounted for by sampling variability. The National Agricultural Statistical Service of USDA conducts a census of US agriculture every five years in December, and conducts a survey of agriculture in June of every year. There is generally good agreement for most farms between the census and the surveys. When the June surveys of the years 2002 through 2005 are compared with the census of 2002, however, for a small number of farms, the deviations exceed what would be expected due to ordinary changes in farm operations. We are able to identify the types of farms whose records are likely to contain nonsampling errors. We are applying and extending methods of supervised classification to develop models

## Interviewer Burden and Its Effects on Data Quality in the Swedish Part of the European Social Survey (ESS)

❖ Lilli Japec, Statistics Sweden, Box 24300, Stockholm, 115 81 Sweden, *lilli.japec@scb.se*

***Key Words:*** interviewer interest, interviewer satisficing, multilevel regression analysis, probing

I introduce the concepts of interviewer burden and interviewer satisfying. The effects of interviewer burden, behavior, attitude and attributes are studied using the data from the Swedish part of the European Social Survey (ESS). Multilevel regression analysis is used to study the effects on a number of data quality indicators such as response rate, length of interview, probing frequency and "don't know frequency." The results show that the interviewer burden affects all the data quality indicators analyzed in this study.

# 416 Nonparametric Bayesian Methods

Section on Bayesian Statistical Science, Section on Nonparametric Statistics

**Wednesday, August 9, 10:30 am–12:20 pm**

## Bayesian Analysis for Quantile Regression of Correlated Data

❖ Chin-Hua Wang, Family Health International, P.O. Box 13950, Research Triangle Park, NC 27709, *lwang@fhi.org*; Pai-Lien Chen, Family Health International

***Key Words:*** quantile regression, Bayesian analysis, analysis for correlated data

Analysis of heterogeneous data where the conditional distributions vary differently with the covariates for the tails and the central location are often analyzed by quantile regression methods. Traditional quantile regression, based on minimizing the sum of absolute residuals, has been developed for independent data and is widely used. When repeated observations are taken on each unit, however, the assumption of independence is no longer valid and alternative methods are required. We propose a parametric Bayesian approach to model the population medians using the Markov chain Monte Carlo algorithm. The approach will be compared to maximum likelihood estimation and inference, an approach based on a transformation to log-normality, and a marginal model technique akin to GEE. Data from recent randomized clinical trials will be used to illustrate the methodology.

## Sequentially Allocated Merge-Split Sampler for Conjugate and Nonconjugate Dirichlet Process Mixture Models

❖ David Dahl, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, *dahl@stat.tamu.edu*

***Key Words:*** Bayesian nonparametrics, sequential importance sampling, partial conditioning, Markov chain Monte Carlo, Metropolis-Hastings algorithm

This paper proposes a new efficient merge-split sampler for both conjugate and nonconjugate Dirichlet process mixture (DPM) models. The latest generation of samplers for these Bayesian nonparametric models effectively update the model parameters, but can have difficulty updating the clustering of the data. To overcome this, merge-split samplers have been developed, but until now these have been limited to conjugate or conditionally-conjugate DPM models. This paper proposes a new MCMC sampler, called the sequentially-allocated merge-split (SAMS) sampler. The sampler borrows ideas from sequential importance sampling. Splits are proposed by sequentially allocating observations to one of two split components using allocation probabilities that condition on previously allocated data. The proposed sampler is substantially more efficient than existing conjugate and nonconjugate samplers.

## Fast Nonparametric Bayes Testing of Distribution Changes in Large Datasets

❖ Michael Pennell, National Institute of Environmental Health Sciences, P.O. Box 12233, Mail Drop A303, 111 TW Alexander Drive, RTP, NC 27709, *pennell@niehs.nih.gov*; David B. Dunson, National Institute of Environmental Health Sciences

***Key Words:*** Dirichlet process, dose-response, large data sets, nonparametric Bayes, toxicology, variational methods

In some toxicology studies, one may expect changes in the shape and location of an outcome distribution with dose. Unfortunately when the data consist of observations on hundreds-thousands of cells or small organisms, it can be difficult to fit models which are sufficiently flexible. To address this problem, we propose a computationally fast method for testing for distribution changes across an ordinal predictor. Our method is based on a Variational Bayes (VB) approach to fitting a Dynamic Mixture of Dirichlet Processes. Our approach begins by approximating the posterior in the lowest predictor group and then, using EM substeps, we successively add atoms to the later groups until there is no benefit in terms of the approximated marginal likelihood. Since VB approaches converge in relatively few steps, our method should be more computationally feasible than MCMC for large data sets.

## On the Random Functional of the Ferguson-Dirichlet Process

❖ Thomas J. M. Jiang, National Chengchi University, Department of Math Sciences, 64 ChihNan Road, Sec 2, Taipei, 11605 Taiwan, *jiangt@math.nccu.edu.tw*; Kun-Lin Kuo, National Chengchi University

***Key Words:*** inverse formula, Ferguson-Dirichlet process, c-characteristic function, Bayesian nonparametric

The random functional of the Ferguson-Dirichlet process has been subjected to research for decades. Univariate c-characteristic functions, which were first given by Jiang (1988), have been shown to have many properties similar to those of the traditional univariate characteristics function. These alternative c-characteristic functions are very useful, especially for those cases that are difficult to manage using the traditional characteristic function. Examples can be seen in Jiang (1988, 1991). Jiang, Dickey, and Kuo (2004) further extended them to be the multivariate c-characteristic functions. In this paper, we first give inversion formulas for these c-characteristic functions. Using them, we give a general expression of the random functional of the Ferguson-Dirichlet process. The exact prob. density functions of the random moments of some Ferguson-Dirichlet process are also given.

## Bayesian Circular Regression

❖ Barbara Jane George, U.S. Environmental Protection Agency, 109 TW Alexander Drive, Mail Code E205 02, Research Triangle Park, NC 27709, *george.bj@epa.gov*; Kaushik Ghosh, New Jersey Institute of Technology

***Key Words:*** directional data, MCMC, predictive density, Dirichlet process

Circular data, where observations are represented by points on the unit circle, arise in the modeling of diverse phenomenon such as time of peak blood glucose level or direction of ocean waves. In this talk, we discuss a semiparametric Bayesian approach for regression involving circular variables. The proposed approach uses Markov chain Monte Carlo techniques to fit a flexible regression model. We will illustrate the proposed method with an example.

## Nonparametric Bayesian Bootstrap in ROC Curve Study

❖ Jiezhun Gu, North Carolina State University, 2501 Founders Drive, Campus Box 8203, Raleigh, NC 27695, *jgu@unity.ncsu.edu*; Subhashis Ghosal, North Carolina State University

***Key Words:*** Bayesian bootstrap, asymptotic properties, ROC curve

Receiver operating characteristic (ROC) curve is applied widely in measuring discriminatory ability of diagnostic or prognostic tests, leading to parametric and nonparametric estimation methods. In this paper, we present the Bayesian bootstrap method to estimate ROC curves. Integrated absolute error of ROC curves is introduced as a measure of accuracy especially useful in simulation studies to evaluate the performance of different methods of estimation of the ROC curves and the estimate of the area under the curve. Flexibility and computational simplicity are the two main advantages of the Bayesian bootstrap. We also study the asymptotic properties of the Bayesian bootstrap estimate, providing a justification of this method.

# 417 Combining Information, Missing Data, and Hierarchical Bayesian Methods for Health Outcomes Data ●

Section on Health Policy Statistics, Biometrics Section, Section on Bayesian Statistical Science, ENAR
**Wednesday, August 9, 10:30 am–12:20 pm**

## Imputaion and Forecasting for State Mortality

❖ Guanhua Lu, University of Maryland/National Center for Health Statistics, 8261 Canning Terrace, Greenbelt, MD 20770, *ghlu@math.umd.edu*; Benjamin Kedem, University of Maryland; Rong Wei, National Center for Health Statistics

***Key Words:*** state mortality, average age pattern, mortality level, local regression, time series, imputation

For state mortality data, a 0 death count occurs at some ages due to the small population. If we take the log death rate into account, there will be missing values where the death count is 0. This makes the way to handle the state mortality quite different from the nation's. To handle this situation, first, we use the average of log death rates across years to capture the age-specific mortality pattern, and apply a local regression method to smooth this pattern. Then a model involving the average age pattern and parameters describing death rate change across age and mortality level across years is constructed. Parameters are estimated via optimality and a time series of the mortality level is used to perform future forecasting. Missing values are imputed through the model with estimated parameters.

## The Impact of Using Different Imputation Methods for Missing Quality-of-Life Scores on the Estimation of the Cost-Effectiveness of Lung Volume Reduction Surgery

❖ David Blough, University of Washington, Department of Pharmacy, Box 357630, Seattle, WA 98195, *dkblough@u.washington.edu*; Sean Sullivan, University of Washington; Scott Ramsey, Fred Hutchinson Cancer Research Center; Roger Yusen, Washington University School of Medicine

***Key Words:*** cost effectiveness analysis, missing data, imputation

Imputation methods were compared using a prospective cost-effectiveness analysis conducted alongside a randomized controlled trial (National Emphysema Treatment Trial). The trial compared lung volume reduction surgery plus medical therapy to medical therapy alone in patients with severe chronic obstructive pulmonary disease. Cost per quality-adjusted life-year gained was obtained, using the Quality of Well Being questionnaire. Different methods of imputation for missing quality-of-life data resulted in substantial differences in cost effectiveness ratios and differences in estimates of the uncertainty in the point estimates as reflected in the acceptability curves. Paradoxically, the use of a conservative single imputation method resulted in relatively less uncertainty. We recommend using a minimum of two imputation methods that always include multiple imputation.

## Combining Information from Various Data Sources To Improve Analyses of Adjuvant Cancer Therapies

❖ Yulei He, Harvard Medical School, Department of Health Care Policy, 180 Longwood Ave, Boston, MA 02115, *he@hcp.med.harvard.edu*; Alan M. Zaslavsky, Harvard Medical School

***Key Words:*** administrative records, hierarchical Bayesian model, measurement error, missing data, multiple imputation

Cancer registry records and administrative systems contain valuable data on provision of adjuvant therapies for cancer patients. Previous studies, however, have shown that these therapies are underreported in those systems. The Cancer Care Outcomes Research and Surveillance Consortium, a multi-center study on treatment and outcomes of colorectal and lung cancer patients, collects treatment data from various sources, e.g. a patient survey and medical records. We propose statistical strategies to combine information from the registry data and patient survey (both subject to undercoverage) and medical records data. Our hierarchical Bayesian models jointly model provision of multiple cancer therapies (e.g. adjuvant chemotherapy and radiation therapy) and reporting in various data sources. Multiple imputations for the true therapy status are hence created to facilitate improved analyses.

## Do Teenagers Always Tell the Truth? Bayesian Methods To Estimate the Prevalence of Adolescent Risk Behaviors from Self-Report

❖ Janet Rosenbaum, Harvard University, 15 Leonard Ave., Cambridge, MA 02139, *jerosenb@fas.harvard.edu*

***Key Words:*** adolescent, self-report, MCMC, teen pregnancy, public health, health risk behavior

Surveys measure adolescent risk behavior (ARB) prevalence; some adolescents over-report their true risk behaviors, while others underreport. Over- and under-reports can be detected if adolescents report having engaged in a risk behavior and then recant the earlier report. Bayesian MCMC methods can estimate the prevalence of a disease without a gold standard test, and the specificity and sensitivity of existing tests. These methods are modified to include covariates to estimate the specificity and sensitivity of self-report for several ARBs and their prevalence. The implications for cross-sectional data are explored.

## Bayesian Simultaneous Intervals for Small Areas: an Application to Variation in Maps

❖ Erik B. Erhardt, University of New Mexico, Department of Math and Statistics, MSC03 2150, Albuquerque, NM 87131, *erike@stat.unm.edu*; Balgobin Nandram, Worcester Polytechnic Institute; Jai Choi, National Center for Health Statistics

***Key Words:*** hierarchical Bayesian model, highest posterior density, Metropolis-Hastings algorithm, Nelder-Mead algorithm, Rao-Blackwellized estimator, simultaneous inference

Simultaneous intervals for parameters for small area inference are needed because these parameters are correlated. Our method finds simultaneous credible intervals for a large number of parameters, each for a single area. Our method uses a hierarchical Bayesian model, starting with either the 100(1-alpha)% credible interval or highest posterior density (HPD) interval for each area. Our method solves two simultaneous equations, one accounting for the probability content, 100(1-alpha)% of all the intervals combined, and one containing an optimality condition like the "equal ordinates" condition in the HPD interval. We compare our method with one based on a nonparametric method, which as expected under a parametric model, does not perform as well as ours, but is a good competitor. We illustrate our method using an example on disease mapping using a standard Poisson regression model.

### Spatial Statistical Methods for Small-Area Health Data with Application to the Association of Breast Cancer Incidence and Local Power Plant Emissions

❖ Heather Watson, New York University, 324 E. 52nd Street, Apt 2C, New York, NY 10022, *hnw206@nyu.edu*; Judith D. Goldberg, New York University School of Medicine; Mengling Liu, New York University School of Medicine

***Key Words:*** spatial statistics, small area, aggregate, disease clustering

To analyze small area health data, several spatial statistical methods are compared. Data is aggregated to different administrative levels. Small area health data produces unstable rates and the scale of aggregation has an impact on the inferences. Smoothed disease maps using Bayesian methods and spatial clustering methods for localized disease clustering or focused disease clustering near an environmental hazard are compared. Applied to female breast cancer cases reported from 1994-2000 in Rockland County and parts of New York City at the census tract and zip code level, the methods assess breast cancer incidence rates in excess near local power plants. Although the techniques aggregate the local small area differences and the methods of analysis differ, the techniques provide complementary inferences about areas of excess disease in the geographic region.

## 418 Bioinformatics ● ✪

Section on Statistical Computing, Biometrics Section, ENAR
**Wednesday, August 9, 10:30 am–12:20 pm**

### Linker DNA Length Preference in Human Chromatin Revealed by a Two-State Duration Hidden Markov Model

❖ Guei-Feng Tsai, Northwestern University, 60208, *guei-feng-tsai@northwestern.edu*; Ji-Ping Wang, Northwestern University; Jonathan Widom, Northwestern University

***Key Words:*** dinucleosome, linker DNA, hidden Markov model, duration

The nucleosome is the repeating subunit of chromatin, formed by a short stretch of DNA, 147bp in length, wrapping around four pairs of histones. The linker DNA that connects two neighboring nucleosomes may favor particular values for a given cell type of a given organism to facilitate chromatin forming. The data we have contains 1002 experimentally obtained human dinucleosome DNA sequences. Each sequence roughly covers two nucleosomes in the two ends with a linker DNA in between, while their exact positions are undetermined. We develop a two state duration Hidden Markov model to determine the optimal path of nulceosome-linker-nucleosome structure under certain biochemical constraints. Our results show that the linker length distribution does favor a series of values roughly space by 5 or 10 bps in a range between 10 bp and 120 bps.

### A Systematic Benchmark of Dimension Reduction in Remote Homology Detection with Support Vector Machines

❖ Melissa M. Matzke, Pacific Northwest National Laboratory, MS K1-90, Richland, WA 99352, *melissa.matzke@pnl.gov*; Bobbie-Jo Webb-Robertson, Pacific Northwest National Laboratory; Christopher S. Oehmen, Pacific Northwest National Laboratory; Jorge F. Reyes Spindola, Pacific Northwest National Laboratory

***Key Words:*** bioinformatics, support vector machine (SVM), multivariate analysis, dimensionality reduction, homology

Biopolymer sequence comparison to identify evolutionarily related proteins is one of the most common and data intensive computing tasks in bioinformatics. One of the most accurate approaches implements support vector machines (SVMs) to classify proteins into families via vectorization of the protein by sequence similarity scores obtained from the Bayesian Algorithm for Local Sequence Alignment (BALSA). However, one primary computational issue with SVMs is the size of the variable set. In this study, the performance of the SVM built with the complete BALSA score set is assessed against a reduced dimensionality. Principal components analysis, sequential projection pursuit, independent component analysis and kernel principal components analysis are used for dimension reduction. The area under the ROC curve is used to compare model performance.

### Using Microarray Gene-Coexpression Networks To Increase Gene Screening Validation Success and To Build Accurate Classifiers

❖ Wei Zhao, University of California, Los Angeles, Department of Human Genetics, Los Angeles, CA 90095, *wzhao@mednet.ucla.edu*; Steve Horvath, University of California, Los Angeles; Paul Mischel, University of California, Los Angeles; Aldons J. Lusis, University of California, Los Angeles; Stanley Nelson, University of California, Los Angeles

***Key Words:*** cancer, prediction, microarray, co-expression network, screening, validation

Microarray gene expression profiles have begun to prove useful for classifying subsets of patients or tumors, and for predicting survival and response to therapy. However, genes identified as predictive in one microarray study have frequently failed to validate in other studies. Traditional gene screening methods often select genes by correlating the expression profiles with microarray sample trait information (e.g. patients survival or case-control status). Several groups demonstrated that selecting genes simply on the basis of a p-value or fold-change criterion may lead to a gene list with poor validation success Michiels et al. (2005). Further, standard classifiers (e.g. k-nearest neighbor) may have poor accuracy in independent data. We propose to use a gene network-based gene screening strategy for identifying gene candidates. We provide ample empirical evidence of the usefulness.

### ANOVA Model-Based Pattern Recognition Technique

❖ Yushu Liu, University of Kentucky, 200 Thomas Hunt Morgan Building, University of Kentycky, Lexington, KY 40506, *yushu@ms.uky.edu*; R. Lakshman Chelvarajan, University of Kentucky; Thomas Getchell, University of Kentucky; Subbarao Bondada, University of Kentucky; Arnold J. Stromberg, University of Kentucky

***Key Words:*** pattern recognition, two-way ANOVA, PCA (principal component analysis)

In microarray data analysis, clustering methods such as hierarchical, k-means are commonly used to identify similarities in gene expression profiles that may suggest commonalities in underlying functions. These methods are appropriate when there are no pre-conceived notions about the commonalities. In contrast, pair-wise comparisons are used when commonalities are intuited by investigator. Motivated by this incongruity, we developed a pattern recognition metric based on two-way

ANOVA and investigator's experimental insight. This method is more stable and more capable of grouping functionally related genes than is k- means clustering. Code is available in Splus. We also used PCA to resolve differences in the expression profiles for multiple probesets of the same gene. Our methods were applied successfully to identify unique expression patterns for genes expressed by splenic macrophages.

### Canonical Parallel View and Adjustment for the Difference between Paired High-Dimensional Datasets

❖ Xuxin Liu, The University of North Carolina at Chapel Hill, Department of Statistics and Operations Research, Smith Building, CB3260, Chapel Hill, NC 27599, *liux@email.unc.edu*

***Key Words:*** microarray, batch adjustment, high dimensional low sample size, parallel direction

High Dimensional, Low Sample Size data are emerging in a number of areas of science. A lot of them deal with the comparison between paired data sets. In this paper, we propose two directions in high dimensional space, the canonical parallel direction and the canonical orthogonal direction, which will show the difference between paired data sets in a noval and useful way. This parallel direction can also be used for linear adjustment of this difference. The view and adjustment for the NCI60 data set illustrate the good performance of this method.

### ProMAT: Protein Microarray Analysis Tool

❖ Amanda M. White, Pacific Northwest National Laboratory, P.O. Box 999, MS K6 08, Richland, WA 99354, *amanda.white@pnl.gov*; Don S. Daly, Pacific Northwest National Laboratory; Susan S. Varnum, Pacific Northwest National Laboratory; Kevin K. Anderson, Pacific Northwest National Laboratory; Nikki Bollinger, Pacific Northwest National Laboratory; Rachel M. Gonzalez, Pacific Northwest National Laboratory; Richard C. Zangar, Pacific Northwest National Laboratory

***Key Words:*** ELISA microarray, curve estimation, prediction

Enzyme-linked immunosorbent assays (ELISA) are used to estimate the concentration of a protein in a sample. Although traditionally used in titer plates, implementing ELISAs in a microarray format allows for the simultaneous estimation of multiple protein concentrations for small samples. Typically researchers use tools such as spreadsheets to estimate standard curves and protein concentrations, which is time-consuming since these tools are not suited for this. The increase in data generated by using microarrays requires more efficient methods of generating standard curves and estimating protein concentrations, thus we have created ProMAT, an open-source ELISA microarray analysis tool based in R and Java. ProMAT also estimates prediction errors on the concentration predictions. Methods used to fit standard curves and generate concentration prediction errors will be discussed.

### Uncertainty in Clustering Posterior Distributions of Gene Expression Levels Using MCMC Samples

❖ Tanzy Love, Carnegie Mellon University, Department of Statistics, Pittsburgh, PA 15213, *tanzy@andrew.cmu.edu*

***Key Words:*** clustering, posterior distributions, gene expression

In time series or multiple treatment microarray experiments, we are interested in locating groups of genes that react together. Subject matter theory designates these groups as coregulated by the same biologic pathways. The statistical problem is clustering genes based on their expression values over multiple treatments. We don't have values for gene expression, rather replicated measurements with error. To incorporate this uncertainty, we have modeled expression estimates using hierarchical models. This provides posterior probability distributions for quantities such as expression value and expression ratio for two treatments. We also can construct the joint posterior probability distribution of these quantities. We use multiple sampling from the posterior distributions of gene expression vectors to cluster genes and estimate the uncertainty in this clustering, an example with maize

## 419 Nonparametric Statistics with Censored Data ●

Section on Nonparametric Statistics, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Nonparametric Siginficance Tests for Sums of Censored Random Variables

❖ Golde Holtzman, Virginia Polytechnic Institute and State University, 406A Hutcheson Hall, Mail Code 0439, Blacksburg, VA 24061-0439, *holtzman@vt.edu*; Carl E. Zipper, Virginia Polytechnic Institute and State University

***Key Words:*** censored, seasonal Kendall analysis, nitrogen, nonparametric, hypothesis test, sums of censored variables

Nonparametric significance tests of variables derived as sums (or other functions) of censored random variables lose power if observations in which one or more of the variables is censored are discarded. The device to be presented makes it possible to perform common rank- and sign-based tests on such data without discarding partially censored observations. We include an example applying the Seasonal Kendall Test for trend in time-series of derived water-quality variables. Specifically, we consider total nitrogen, which is the sum of nitrite, nitrate, and total kjeldahl nitrogen.

### Nonparametric Tests for Covariate Effects with Multistate Survival Data

❖ Limin Peng, Emory University, Department of Biostatistics, 1518 Clifton Road, NE Rm 324, Atlanta, GA 30322, *lpeng@sph.emory.edu*; Jason P. Fine, University of Wisconsin-Madison

***Key Words:*** censoring, dichotomization, empirical process theory, integral tests, multiple event types, uniform representation

In clinical trials and observational studies, it is often of scientific interest to evaluate the effects of covariates on complex multistate event probabilities. With continuous covariates, a common approach is to arbitrarily discretize continuous covariates, which may lead to substantial information loss, or to formulate the covariate effects in a regression model, which may have low power or be biased under misspecification. We propose nonparametric tests not requiring arbitrary discretization. General asymptotic results are derived under the null and the alternative hypotheses. The tests are consistent under stochastic ordering, which arises naturally with multistate data. Simulation studies demonstrate the gains of the new testing procedure over those based either on categorization or on regression models. The method's practical utility is illustrated with three real examples.

### Estimation for Two-Sample, Location-Scale Models under Type I Censorship

❖ Xuewen Lu, University of Calgary, Department of Mathematics and Statistics, Calgary, AB T2N 1N4 Canada, *lux@math.ucalgary.ca*

***Key Words:*** asymptotic normality, location-scale model, strong consistency, type I censorship

The author considers a method-of-moments estimation approach to the two-sample, location-scale problem under type I censorship. The estimators for both the location shift and scale change parameters are given. It is shown that the two estimators are strongly consistent, asymptotically normal, and easy to compute. Simulation studies indicate that the proposed estimation procedure performs well in finite samples. A failure-time dataset is used as illustration.

### Nonparametric Maximum Likelihood Estimation of Hazard Function under Shape Restrictions

❖ Desale Habtzghi, University of Georgia, Department of Statistics, Athens, GA 30602, *dhabtzgh@stat.uga.edu*; Mary Meyer, University of Georgia; Somnath Datta, University of Louisville

***Key Words:*** shape constraints, hazard function, estimation, nonparametric, right censoring, simulation

The problem of estimation of hazard function has received considerable attention in the statistical literature. In particular, assumptions of increasing, decreasing, bathtub-shaped and convex hazard function are common in literature, but practical solutions are not well developed. In this talk we introduce a new nonparametric method for estimation of hazard functions under shape restrictions to handle the above problem. We adopt a nonparametric approach in assuming that the density and hazard rate have no specific parametric form with the assumption that the shape of the underlying hazard rate is known (either decreasing, increasing, concave, convex or bathtub-shaped). We also show how the estimation procedures can be used when dealing with right censored data. We evaluate the performance of the estimator via simulation studies and illustrate it on some real data set.

### Inference on the Quantile Function under Left Truncation and Right Censoring

❖ Sana Buhamra, Kuwait University, P.O. Box 5969, Safat, Kuwait, 13060 Kuwait, *buhamra@kuc01.kuniv.edu.kw*; Noriah Al-Kandari, Kuwait University

***Key Words:*** quantile function, shrinkage, kernel estimation, truncated and censored data, Monte Carlo

We consider estimating the quantile function when data are left-truncated and right-censored (LTRC). Asymptotic statistical testing and estimation procedures are established for the quantile function from one sample and multiple-sample LTRC data. The proposed estimators are based on shrinkage estimation techniques assuming uncertain prior nonsample information on the value of the quantile. The asymptotic bias and risk of the proposed estimators are derived and compared with the benchmark estimator analytically and computationally. The proposed estimation strategy, which combines the sample and parameter information, performs better than a strategy based on sample information only. An application of the proposed methodology to the well-known Channing house data illustrates the theory.

### Empirical Likelihood Method for Heteroscedastic Linear Model

❖ Hua Zhu, University of Kentucky, 4070 Victoria Way, Apt. 84, Lexington, KY 40515, *huazhu@ms.uky.edu*; Mi-Ok Kim, University of Kentucky; Mai Zhou, University of Kentucky

***Key Words:*** Wilks theorem, empirical likelihood, variance modeling, right censored data

For a heteroscedastic regression model, variance modeling has been considered to improve the efficiency of mean parameter estimates or to estimate variance if it is interesting in its own right. Among others, Owen (2001) proposed the empirical likelihood (EL) method. We propose a new estimating equation and consider the EL method for the conditional means. We show that properly defined EL ratio statistic has an asymptotic chi-square distribution under null hypothesis. We also examine the possibility to extend the EL method to variance modeling in censored case with the proposed estimating equation.

### Empirical Likelihood and Marginal Confidence Interval

❖ Mi-Ok Kim, University of Kentucky, 817 Patterson Office Tower, Lexington, KY 40506-0027, *miokkim@ms.uky.edu*

***Key Words:*** empirical likelihood, censored regression

Empirical likelihood (EL) is a nonparametric inference method with results that are in general similar to those about likelihood ratio tests and Wilk's theorem in the parametric model. In regression the EL method faces a challenge of maximizing the likelihood in presence of nuisance parameters when a single component of the regression coefficients is of interest. We propose a simple solution to the problem in the least squares regression and investigate whether the proposed method can be generalized to censored case.

## 420 Time Series and Temporal Correlation with Regression Applications

Section on Physical and Engineering Sciences, Section on Statistics and the Environment

**Wednesday, August 9, 10:30 am–12:20 pm**

### Spectral Analysis of Nonstationary Time Series with Piece-Wise Monotonic Time-Varying Frequencies

❖ Md. Jobayer Hossain, Southern Methodist University, P.O. Box 750332, 3225 Daniels Ave, Dallas, TX 75275-0332, *mhossain@smu.edu*; Wayne A. Woodward, Southern Methodist University; Henry L. Gray, Southern Methodist University

***Key Words:*** LM-stationary, RM-stationary, general instantaneous period, general instantaneous frequency

One type of nonstationarity of a time series occurs when the frequency behavior of the process changes over time. Window based methods are commonly used for analyzing this type of nonstationary time series. Gray, Woodward et al have recently introduced techniques based on time deformation for analyzing nonstationary processes with monotonic time varying frequencies. It has been seen that their methods outperform window based methods in this case. We address the problem of using time deformation techniques in the case in which frequency change is not monotonic. We introduce piecewise M-stationary process which is capable of analyzing data with linear periodic change that is piece-wise monotonic. The instantaneous spectrum obtained using this model is able to capture the change of frequency behavior more clearly than existing methods.

### Nonparametric Kernel Estimates of Autocorrelation Structure from Single-Molecule Experimental Data

❖ Tingting Zhang, Harvard University, 1 Oxford Street, 608, Cambridge, MA 02138, *tzhang@fas.harvard.edu*; Samuel Kou, Harvard University

***Key Words:*** Cox process, autocorrelation function, bandwidth selection

The recently developed photon-by-photon approach (Yang and Xie 2002) for single molecule fluorescence experiments allows measurements of conformational fluctuation with high time resolution on a broad range of time scales. In this method, each photon arrival time is a data point. Of particular interest is the estimation of the autocorrelations of photon arrival rates, which provides crucial insights into the molecular structure and reaction dynamics, and in turn can be used to test models. Complementary to the parametric inference (Kou, Xie and Liu, 2005), in this paper we will introduce a nonparametric kernel based method to estimate the autocorrelation function. In addition to discuss the theoretical properties of our method, we will illustrate it through simulation examples, and apply it to analyze experimental data on protein conformation fluctuations.

### Exploring Statistical Correlations among Nonlinear Time Series/Signals

❖ Carolyn Morgan, Hampton University, Queen and Tyler Streets, Hampton, VA 23668, *carolyn.morgan@hamptonu.edu*; Morris H. Morgan, Hampton University

***Key Words:*** time series, correlation, nonlinear, stochastic

Methods for distinguishing chaotic from stochastic signals have been investigated. Interest has focused on detecting the change-over point between such signals. That work has lead to the development of several statistical metrics for assessing nonlinear time series features and a general interest in the problem of time series complexity measurements. This approach is based on using a metric that is a weighted linear combination of an order pattern statistic and a traditional distance based norm. Similar metrics have been used in the medical profession to assess brain and heart data patterns and to distinguish healthy subjects from sick ones. Examples of such metrics include but are not limited to Lyapunov exponents, fractal dimensions and permutation entropies. Our work focuses on a more robust measurement approach.

### Wavelet Variance Analysis for Random Fields

❖ Debashis Mondal, University of Washington, Department of Statistics, Box354322, Seattle, WA 98195, *debashis@stat.washington.edu*; Donald B. Percival, University of Washington

***Key Words:*** asymptotic normality, Daubechies wavelet filter, intrinsic random fields, local stationarity, maximal overlap discrete wavelet transform, cloud data

Wavelet-based statistical methods have been used for signal and image processing, edge detection, nonparametric regression and inverse problems. Wavelets also give rise to the concept of the wavelet variance, which decomposes the sample variance of a time series and provides a time- and scale-based analysis of variance. The wavelet variance has been applied to a variety of time series and is particularly useful for studying long memory processes, detecting inhomogeneity and estimating spectral densities. Here we extend the notion of the wavelet variance to random fields and discuss large sample inference for its estimation. To illustrate our theory, we explore some specific Gaussian random fields with long range dependencies and study a localized version of the wavelet variance to test for inhomogeneity and directionality. We illustrate our methodology using images of clouds.

### Wavelet-Based Estimation of Linear Regression Models with Two Errors: a Long Memory and a White Noise

❖ Kyungduk Ko, Boise State University, 1910 University Drive, Department of Mathematics, Boise, ID 83725-1555, *ko@math.boisestate.edu*

***Key Words:*** long memory, discrete wavelet transform, linear regression, EM algorithm

Linear regression models with long memory error have been useful in many areas, such as signal and image processing, climatology, and economics. Here we analyze a linear regression model with two errors, a long memory and a white noise. Discrete wavelet transforms are applied to the explanatory and response variables in order to simplify the dense variance-covariance matrix of the additive error structure of a long memory and a white noise. An EM algorithm is then adopted for the estimation of the model parameters. Performances are evaluated on simulated and real data.

### Calibrating OLS Estimators in Linear Regression with Long Memory Error

❖ Jaechoul Lee, Boise State University, 1910 University Drive, Department of Mathematics, Boise, ID 83725-1555, *jaechlee@math.boisestate.edu.*; Kyungduk Ko, Boise State University

***Key Words:*** degree of freedom, long memory error, ordinary least squares, regression

Linear regressions with long memory error have received considerable attention in many fields, such as engineering, medical imaging and economics, because of their wide range of possible applications. Use of maximum likelihood or generalized least square estimates for inferential purpose is practically limited due to the complexity and computational demands in evaluating the estimates and their variances by the nature of long range dependency. We propose an inferential tool to improve accuracy of ordinary least square estimate under the assumption that long memory parameter and innovation variance are both unknown a priori. The method uses the exact variance of OLS estimate and calibrates degrees of freedom in Student t critical value in order to achieve the pre-specified type I error. Simulation results are reported together with an application to the northern hemisphere series.

### On Improved Estimation in Linear Regression with Long Memory Errors

❖ Mohamedou Ould Haye, Carleton University, Hp Building, 1125 Colnel By Dr, Ottawa, ON K1S 5B6 Canada, *ouldhaye@math.carleton.ca*; A. K. Saleh, Carleton University

***Key Words:*** long memory processes, linear regression, preliminary test estimation, central limit theorem, time series, risk analysis

This paper investigates the asymptotic properties of various estimators of the slope parameter in linear regression with long memory errors, characterized by the slow decay of the covariance function, when vague nonsampling information on some components of the slope is available. In particular, we discuss the usual LSE (unrestricted estimator), restricted LSE (based on vague information), preliminary test estimator, shrinkage estimator, and the positive rule estimator of the slope. It is shown that near the prior information, the restricted, preliminary test, and shrinkage estimators perform better than the unrestricted estimator, while their superiority changes as the true hypothesis diverts from the prior information. The analysis is based on the asymptotic properties of the estimators under contiguous alternatives. Convergence rates will be much slower than in iid case.

# 421 QTL Analysis and Mapping ●

Biometrics Section
**Wednesday, August 9, 10:30 am–12:20 pm**

## Strategies for Fine Mapping of QTL in Complex Pedigrees Using Combined Linkage and Linkage Disequilibrium Method

❖ Natascha Vukasinovic, Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167, *natascha.vukasinovic@monsanto.com*; Fengxing Du, Monsanto Company

***Key Words:*** QTL, genetic marker, haplotype, linkage analysis, linkage disequilibrium

Advances in genome sequencing and genotyping have stimulated use of single nucleotide polymorphism (SNP) markers in fine mapping of quantitative trait loci (QTL). A novel LA/LD method that combines linkage analysis and linkage disequilibrium information is used for fine mapping of QTL in complex pedigrees. Polymorphic microsatellite and biallelic SNP markers are used in LA/LD analysis. However, when a large number of markers is used to construct haplotypes, inclusion of polymorphic along with biallelic markers results in overestimated QTL effects at positions near more informative markers. When only SNP markers are used, mapping precision is low, especially when a QTL is small. In this work, we explore strategies for efficient haplotype construction in complex pedigrees, regarding use of markers with various information content, marker spacing, and optimal number of loci in a haplotype.

## A New Statistical Model for Dissecting the Genetic Basis of Heterosis through Multiple Interval Mapping (MIM) in Design III Populations

❖ Antonio A. F. Garcia, Escola Superior de Agricultura "Luis de Queiroz", Avenida P·dua Dias, 11, CP 83, Piracicaba, SP, 13400-970 Brazil, *aafgarci@esalq.usp.br*; Zhao-Bang Zeng, Bioinformatics Research Center

***Key Words:*** hybrid vigor, QTL, plant breeding, marker assisted selection

Heterosis is a phenomenon in which hybrids have superior performance over their parents. Its causes are still not solved. Populations obtained through Design III with molecular markers have been used to study heterosis, but only in the framework of single marker analysis. To study the genetic basis of heterosis, we developed a Multiple Interval Mapping model for Design III, based on the genetic expectations of contrasts between marker genotypes. Our model allows estimation of genomic position, effects and interaction of QTL, and has more statistical power than the current ones. We applied our model to a maize data set and found 11 QTL and some digenic epistatic interactions, explaining about 80% of the phenotypic variation. The results suggested that dominance is the major cause of heterosis in maize. Additive by additive epistasis is responsible for a small fraction of this phenomenon.

## A Semiparametric Approach for Functional Genetic Mapping of Long-Term HIV Dynamics

❖ Song Wu, University of Florida, Department of Statistics, 103 GriffinFloyd Hall PO Box 118545, Gainesville, FL 32611-8545, *songwu@ufl.edu*; Jie Yang, University of Florida; Rongling Wu, University of Florida

***Key Words:*** AIDS, functional mapping, HIV dynamics, quantitative trait loci, semi-parametric

A series of statistical models in functional mapping framework have been proposed to characterize quantitative trait loci (QTL) on the host genome that affect short-term HIV dynamics with a parametric function characterizing earlier stage of viral load trajectories. However, no particular mathematical functions have been available to describe long-term HIV dynamic changes. Here we present a novel statistical model for functional mapping that analyzes the whole HIV load trajectories semi-parametrically. The new model allows for the test of differences in the genetic control of short- and long-term HIV dynamics, and detects the effect of the viral-host genome interaction. We perform extensive simulation studies to test the statistical behavior of the model. The new model will provide an important tool for genetic and genomic studies of human complex diseases and their pathological progre

## A Semiparametric Approach to K Mixtures of Two Components with Application to the Mapping of Quantitative Trait Loci

❖ Shiju Zhang, The University of Toledo, 1660 Vally Ave., Apt F, Homewood, AL 35209, *shiju.zhang@gmail.com*; Biao Zhang, The University of Toledo; Grier P. Page, The University of Alabama at Birmingham

***Key Words:*** density ratio, mixtures, quantitative trait loci

The problem of estimating K mixtures of two components is considered. Under the assumption that the ratio of the two component densities takes a parametric form, we obtain the maximum semiparametric likelihood estimators of the parameters and establish the large sample results of those estimators. We then develop an empirical likelihood ratio-based test for the mixing proportion. Simulation studies are conducted. Finally, we apply the approach to some real data and to the mapping of quantitative trait loci linked to genetic markers.

## Functional Mapping of Time-Warped Developmental Trajectories Based on B-Splines

❖ Xueli Liu, University of Florida, Department of Statistics, 102 Griffin Floyd Hall, Gainesville, FL 32611, *xueliliu@gmail.com*; Rongling Wu, University of Florida

***Key Words:*** alignment, b-splines, functional mapping, nonparametric function estimation, quantitative trait loci (QTL), time-warping

Dynamic quantitative traits whose phenotypic values change with time are of economic, biological and clinical importance. When the dynamics of development over time are at issue, subjects may experience events at a different temporal pace, which may be under genetic control. Previous genetic mapping of quantitative trait loci underlying dynamic traits has typically not been based on a model in which trajectories may be individually time-transformed. We propose an innovative model framework where we consider subjecting each trajectory to a time transformation in an attempt to reverse the warping of the time scale. Our model assumes a monotone time transformation for each trajectory. Subjects are aligned and the alignment helps to estimate the positions and effects of multiple QTL. The method is demonstrated and validated by simulated studies.

### Combining QTL Analysis and Bayesian Network Discovery Methods To Determine Genetic Relationships in a Micorarray/Marker Dataset

❖ Christine W. Duarte, North Carolina State University, Bioinformatics Research Center, Campus Box 7566, Raleigh, NC 27695-7566, *ckwoods@ncsu.edu*; Zhao-Bang Zeng, Bioinformatics Research Center

***Key Words:*** QTL, microarray, Bayesian networks

The knowledge of genetic regulatory mechanisms has important applications in medicine and biotechnology, but experiments for determining such mechanisms are labor-intensive and costly. The analysis of large-scale genomic datasets may represent an alternative path to learning genetic relationships, but first bioinformatic tools to mine this data must be developed. In this work, a method for the discovery of genetic networks has been developed which combines QTL analysis methodology and Bayesian Network learning algorithms. This method involves analyzing microarray and marker data from the progeny of a backcross of two genetically distinct lines. This method has been applied to analyze the genetic mechanism underlying the co-regulation of genes involved in the lignin biosynthesis pathway in Eucalptus trees. Preliminary results in the analysis of yeast data will be presented as well.

### A Statistical Approach for Genome-Wide Scan and Testing Imprinted Quantitative Trait Loci

❖ Yuehua Cui, Michigan State University, A411 Wells Hall, East Lansing, MI 48824, *cui@stt.msu.edu*

***Key Words:*** EM algorithm, genomic imprinting, maximum likelihood, imprinted quantitative trait loci, reciprocal backcross

Nonequivalent expression of alleles at a locus results in genomic imprinting. In this article, a statistical approach for genome-wide scan and testing imprinted quantitative trait loci (iQTL) underlying complex traits based on experimental crosses of inbred line species is developed. The genetic design is based on two reciprocal backcrosses. The proposed approach models genomic imprinting by considering imprinting as a probability measure. Simulation results shows that our model is very robust for identifying iQTL with various degree of imprinting ranging from no imprinting, partial imprinting to complete imprinting. Under various simulation scenarios, the model shows consistent parameter estimation with reasonable precision and high power in testing iQTL. The developed model provides a testable framework for testing and estimating iQTL involved in the genetic control of complex traits.

## 422 The Cox Model and Methods for Recurrent Events ●

Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Methods To Distinguish between the Cox's and Aalen's Model for Right-Censored Data

❖ Yinghua Zhang, Medical College of Wisconsin, Division of Biostatistics, 8701 Watertown Plank Road, Wauwatosa, WI 53226, *yhzhang@mcw.edu*; John Klein, Medical College of Wisconsin

***Key Words:*** Cox's proportional hazards model, Aalen's additive hazards model, censored survival data

In this paper, we compare three methods to distinguish between Cox's proportional hazards model and Aalen's additive hazards model in censored survival data. The statistics are based on distances between estimated survival curve and Kaplan-Meier survival curve, chi-squared goodness-of-fit test, and R2-type measures of explained residual variation, under each model. A Monte Carlo study was carried out to compare the performance of these models and the behavior of Cox's model and Aalen's model in the presence of model misspecification and an independent censoring mechanism.

### Asymptotic Theory for the Proportional Hazards Model with Random Effects

Anthony C. Gamst, University of California, San Diego; ❖ Michael Donohue, University of California, San Diego, 3855 Health Sciences Drive 0901, La Jolla, CA 92093-0901, *mdonohue@ucsd.edu*; Ronghui Xu, University of California, San Diego

***Key Words:*** correlated failure time data, proportional hazards, survival data

We will study the proportional hazards mixed effects model (PHMM) of Vaida and Xu (2000), which is a natural extension of the conventional proportional hazards model to handle clustered event time data. Maximum likelihood estimates under PHMM have been widely utilized without theoretical justification since being proposed. Under regularity and identifiability assumptions, we show consistency, asymptotic normality, and asymptotic efficiency of the maximum likelihood estimates of the model. The proof uses methods that Murphy (1994, 1995) applied to the frailty model, and Zeng, Lin, and Yin (2005) recently applied to the proportional odds mixed effects model.

### Comparing Two Crossing Hazard Rates by Cox Proportional Hazards Modeling

❖ Kejian Liu, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, *kejian.liu@novartis.com*; Peihua Qiu, University of Minnesota; Jun Sheng, University of Minnesota

***Key Words:*** Box-Cox transformation, proportional hazards regression, resampling techniques, survival analysis

Motivated by a clinical trial of Zinc nasal spray for the treatment of the common cold, we consider the problem of comparing two crossing hazard rates. A comprehensive review of the existing methods for dealing with the crossing hazard rates problem is provided. A new method, based on modeling the crossing hazard rates, is proposed and implemented under the Cox proportional hazards framework. The main advantage of the proposed method is the utilization of the Box-Cox transformation which covers a wide range of hazard crossing patterns. Simulation studies are conducted for comparing the performance of the existing methods and the proposed one, which show that the proposed method outperforms some of its peers in certain cases. Applications to a kidney dialysis patients data and the Zinc nasal spray clinical trial data are discussed.

### Robust Method for Analyzing Recurrent Events Data in the Presence of Terminal Events

❖ Rajeshwari Sundaram, The University of North Carolina at Charlotte, Department of Math and Statistics, 9435 Bonita Lane, Apt 1516, Charlotte, NC 28262, *rsundara@uncc.edu*

***Key Words:*** recurrent events, dependent censoring, survival, robust method, semiparametric inference

Recurrent event data are frequently encountered in biomedical studies: infections in AIDS patients and seizures in epilepsy patients. In addition to loss to follow up, such data are further complicated by the presence of terminal events like death, which obviously precludes subsequent recurrences. Here, we consider a family of semiparametric transformation models for the cumulative mean function of such recurrent events process over time in presence of terminal events. In the proposed method, we first model the survival times of the individuals and using them appropriately, define a class of robust estimators for the regression parameters. The asymptotic properties like consistency and asymptotic normality of the proposed estimators are established. Finite sample properties are examined through extensive simulations. We conclude with a real data example.

### Estimation of Gap-Time Distribution with Recurrent Event Data under an Informative Monitoring Period

❖ Akim Adekpedjou, University of South Carolina, Department of Statistics, Columbia, SC 29208, *adek@stat.sc.edu*; Edsel A. Pena, University of South Carolina

***Key Words:*** informative censoring, recurrent event, square integrable martingale, maximum likelihood, counting process, asymptotic efficiency

We consider a biomedical study which monitors the occurrences of a recurrent event for n subjects. We assume that the ith unit is monitored over a period [0,ti] and denote by Ti1,Ti2,..the successive event interoccurrence times for this unit, with ti independent of the Tij's. A total of Ki=max{k:Ti1+Ti2+.+Tik = ti} will be observed. In addition, we assume that the Tij's are IID from an unknown distribution F(t)=F(t;?) where ? is a p-dimensional parameter, and the ti's are IID from a distribution G. We address the problem of estimating ? and consequently the gap-time dist F(t), under the assumption that 1-G=(1-F)^fl for some unknown fl > 0,the so-called Koziol-Green (KG) model. We present finite and asymptotic properties of the estimators of ?,fl and of F(t), and compare the estimator of F(t) with the nonparametric estimator in Pena et al.(01,JASA)to ascertain efficiency gain achieved.

## 423 Binary Data ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### A Multiplicative Heteroscedasticity Model for Clustered Binary Data

❖ Mitchell Rosen, Omnicare Clinical Research, 630 Allendale Road, King of Prussia, PA 19406, *mitch.rosen@omnicarecr.com*

***Key Words:*** multiplicative heteroscedasticity, clustered binary data, maximum likelihood, generalized linear model, correlated binary data, ratio

A multiplicative heteroscedasticity (MH) framework for modeling clustered binary data is proposed. Parameter estimates and hypothesis tests are obtained using generalized linear models. Hypothesis tests for the differences between proportions are straightforward. The MH model is based on asymptotic normal theory.

### Design and Analysis of Active Control Noninferiority Trials with Binary Data

❖ Yu-Yun Ho, Johnson & Johnson Pharmaceutical R&D, 63 Hartley Lane, Basking Ridge, NJ 07920, *yho5@prdus.jnj.com*; Sudhakar Rao, Johnson & Johnson Pharmaceutical R&D; George Chi, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** non-inferiority, retention rate, rate difference, odds ratio, log of odds ratio, fixed margin

Determination of noninferiority margin in an active control noninferiority trial has been a much-debated problem. The margin must not be larger than the assured effect size of the control to guard against a false positive claim. On the other hand, it cannot be too small, leading to impractical sample sizes. In view of the controversies in margin determination, we propose a unified approach for testing retention of a prespecified fraction of the control effect (d0, 0< d0< 1) using rate difference, odds ratio, and log of odds ratio. It turns out that for each retention rate procedure, there is a fixed margin counterpart. We derive a formula to express noninferiority margin as a function of d0. The sample size formula is provided and verified with simulation. We also provide a general guideline for selecting a proper d0. The methodology is illustrated with an approved drug.

### A Note on Sample Size Re-estimation with Interim Binary Data for Double-blind Clinical Trials

❖ Xiaohui Luo, Merck & Co., Inc., Clinical Biostatistics, 181 Buttonwood Drive, Piscataway, NJ 08854, *edmund_luo@merck.com*; Peng-Liang Zhao, Kyowa Pharmaceutical, Inc.

***Key Words:*** type I error, power, interim analysis, dummy stratification

Shih and Zhao (1997) proposed a design with a simple stratification strategy for clinical trials with binary outcomes to re-estimate the required sample size during the trial without unblinding the interim data. A small increase in the effective type I error rate (increase about 0.01 at the significance level of 0.05) and the maintenance of an adequate effective power were shown for this design in the simulation study. Theoretical results were later derived by Govindarajulu (2005). In this talk we discuss how to choose the stratification factor in the design to control the effective type I error rate at the targeted significance level. The impact of the selected stratification factor on the effective power will also be presented.

### More Powerful Analyses of Stratified NonInferiority Trials with Binary Endpoints

❖ Devan V. Mehrotra, Merck Research Laboratories, UN-A102, 785 Jolly Road, Bldg C, Blue Bell, PA 19422, *devan_mehrotra@merck.com*; William W. B. Wang, Merck Research Laboratories

***Key Words:*** binary data, Cochran-Mantel-Haenszel weights, minimum risk weights, null variance, score test, stratified trial

Consider a clinical trial designed to establish the non-inferiority of two treatments based on a binary endpoint. If randomization is pre-stratified by a prognostic factor (e.g., gender), then the Z-statistic used to test for non-inferiority is a function of (i) the non-inferiority bound, (ii) the estimated overall treatment effect, d, obtained as a weighted average of the stratum-specific treatment effects, and (iii) the estimated variance of d, V_d. The power of the test for non-inferiority is influenced by the choice of weights, and by whether the null or observed stratum-specific variances are used to determine V_d. We use simulations to show the power gains achieved by using the minimum risk weights of Mehrotra and Railkar (2000) in tandem with observed variances, relative to common approaches such as the Cochran-Mantel-Haenszel weights in combination with null variances.

### Estimation of Multiple Response Rates in Clinical Trials with Missing Observations

❖ Myron Chang, University of Florida, 1705 NW 22nd Street, Gainesville, FL 32605, *mchang@cog.ufl.edu*

***Key Words:*** self-consistent principle, likelihood function, asymptotic variance, non-informative missing, relative efficiency

Clinical trials sometimes contain two or more binary endpoints, and a simultaneous estimation of multiple response rates is required. In practice, some objectives are inevaluable to one or more responses due to uncontrollable factors, such as insufficient cell quantity in samples for certain laboratory test. To estimate a response rate, the standard method is to compute the proportion of responders among all subjects who are evaluable for the corresponding response. Under the assumption of non-informative missing, we develop an estimation method based on the self-consistent principle. Our simulation studies show that our method is more efficient than the standard method when the binary endpoints are correlated and the number of missing observation is substantial. To compare our method to the standard method in large sample cases, the relative asymptotic efficiency is developed.

### Choice of Working Correlation Structure for a GEE-Based Analysis of Incomplete Longitudinal Binary Data

❖ Priya Kulkarni, Merck Research Laboratories, UN-A102, 785 Jolly Road, Blue Bell, PA 19422, *preeya@gmail.com*; Devan V. Mehrotra, Merck Research Laboratories; Xiaoming Li, Novartis Pharmaceuticals Corporation

***Key Words:*** longitudinal binary data, GEE, missing data, optimal correlation structure, compound symmetry, dropouts

Consider a randomized clinical trial in which subjects receive treatment A or B. For each subject, a binary response of interest (responder/non-responder) is measured at fixed time points. The objectives are to estimate the true proportions of responders at the scheduled end of the trial, p_A and p_B, and to test the hypothesis: p_A = p_B. The generalized estimating equations (GEE) approach can be used to address these objectives. The working correlation structure needed for the GEE-based analysis is either specified in advance (CS, AR(1), UN, etc.), or is chosen after looking at the data based on a pre-specified criterion. We will use simulations to compare the Type I error rate and power properties of these two competing strategies in the setting of incomplete longitudinal binary data, for both missing completely at random (MCAR) and missing at random (MAR) missing data mechanisms.

### Constructing Better Binomial Confidence Intervals by Remembering Three Lessons from Normal Data

❖ Craig Borkowf, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Mail Stop E48, Atlanta, GA 30333, *uzz3@cdc.gov*

***Key Words:*** Agresti-Coull method, binomial proportion, Clopper-Pearson method, confidence interval, SAIFS-T method, teaching statistics

Consider the established method for constructing confidence intervals (CIs) for the mean of normal data with unknown variance. First, we use the unbiased variance estimator with (n - 1) in the denominator, instead of the biased maximum likelihood (ML) estimator with n in the denominator (n = sample size). Second, we use the t-distribution instead of the standard normal distribution. Third, we implicitly benefit from the continuity of normal data. By contrast, we forget these three key lessons when we construct binomial CIs. First, we use the biased ML variance estimator, p(1 - p)/n, instead of the unbiased estimator, p(1 - p)/(n - 1), (p = sample proportion). Second, we do not use the t-distribution, despite the fact that the true variance is unknown. Third, we seldom adjust for discreteness. By applying these three lessons, however, we can obtain binomial CIs with near nominal coverage.

## 424 Missing Data ●

Biopharmaceutical Section, Biometrics Section

**Wednesday, August 9, 10:30 am–12:20 pm**

### Statistical Methods To Analyze Incomplete Clinical Trial Data

❖ Ohidul Siddiqui, U.S. Food and Drug Administration, 17905 Gainford Place, Olney, MD 20832, *Siddiquio@cder.fda.gov*

***Key Words:*** randomized trials, MMRM, LOCF, missing data

In analyzing longitudinal clinical trial datasets, a challenge for statisticians is to deal with missing data due to drop out from the trials before the study endpoints. Statisticians often impute missing values using last observation carried forward and use ANOVA models---or assume a particular missing data mechanism, either missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)---and likelihood-based approaches. Mixed effect model repeated measure analysis and random effect model analysis provide valid inferences in presence of MCAR or MAR, and pattern-mixture model analysis provides valid inferences in presence of MNAR. In this research, attempts are made to test the robustness of the stated approaches in analyzing simulated incomplete data and real data from NDAs of neurological and psychiatric drug products.

### Mixed Model: an Alternative to LOCF as Primary Analysis

❖ Cunshan Wang, Pfizer Inc., 50 Pequot Ave., MailStop 6025 B3212, New London, CT 06320, *cunshan.wang@pfizer.com*; Naitee Ting, Pfizer Inc.; Greg C. G. Wei, Pfizer Inc.

***Key Words:*** mixed model, likelihood, primary analysis, Macugen, LOCF

Last observation carried forward (LOCF) has been a common statistical tool for handling missing data in clinical trials. Numerous drugs have been approved based on LOCF analyses. Recently, newer statistical methods are available to better deal with missing data. In this article, we recommend a likelihood-based linear mixed effects model as the primary analysis for continuous longitudinal clinical data. This model uses all available data. Treatment and time are considered discrete to allow for different means over time without parametric modeling. Covariances among repeated measures are modeled through using subject specific random effects that can be expressed as functions of time. We propose to use it as an alternative to LOCF as primary analysis in protocols. Efficacy data from the Phase 3 Macugen studies are used as an example to illustrate its usage.

### Estimating Treatment Effect in Clinical Trials with Disease-Dependant Noncompliance

❖ Kuenhi Tsai, Vertex Pharmaceuticals Incorporated, 130 Waverly Street, Cambridge, MA 02139, *kuenhi_tsai@vrtx.com*; Peter Westfall, Texas Tech University; Stephan Ogenstad, Vertex Pharmaceuticals Incorporated; Miles Dunn, Vertex Pharmaceuticals Incorporated

***Key Words:*** noncompliance, clinical trials, simulations, covariate adjustment

Estimating treatment effect in clinical trials is often complicated by noncompliance with the nominal prescribed dosage. Incorporated noncompliance as a covariate becomes a trend in estimating treatment effect. However, the impact of noncompliance on treatment effect is dependent upon the therapeutic disease. We illustrate its impact in 3 therapeutic areas of HIV, rheumatoid arthritis, and Alzheimer. Due to the difference, the covariate adjustment of noncompliance in estimating the treatment effect has to be appropriately selected according to the natural history of the disease. Based on published PD models and available references, we simulate the data under various schemes of noncompliance on these 3 therapies. The simulated data are analyzed with and without noncompliance as a covariate in estimating treatment effect. Our results show the significant impact of noncompliance.

### Variability in Visit Times in Clinical Trials and the Implications for a Common Mixed Model for Repeated Measures

❖ Tristan Massie, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Bldg 22 Room 4249, Silver Spring, MD 20993-0002, *MassieT@cder.fda.gov*

***Key Words:*** repeated measures, visit windows

A popular mixed model for repeated measures in a clinical trial fits time as a class variable. When there is variability around planned visit times this assumes the overall response over time is a step function, constant within each visit window. This may be suboptimal for detecting treatment differences if the response is changing moderately fast around a particular visit and there is variability in the actual time of visit. In such a case one can also fit a local slope for each visit window at the price of only k (or a multiple of k) more parameters, where k is the number of post-baseline visits. Like the step function approach this obviates the need of specifying the functional form of the response over time and when the slopes are of moderate size it can also have better power than the step function approach for testing for a treatment difference at the planned visit time.

### Using Stochastic Differential Equations for Imputation of Missing Values in Longitudinal Clinical Data

❖ Naum Khutoryansky, Novo Nordisk, 100 College Road W., Princeton, NJ 08540, *nakh@novonordisk.com*

***Key Words:*** missing data, imputation, stochastic differential equations

Missing data are common in clinical trials. In longitudinal studies missing data are mostly related to dropouts. Some dropouts appear completely at random. The source for other dropouts is withdrawal from trials due to lack of efficacy. For the latter case standard analyses of the actual observed data can produce bias. This paper considers application of stochastic differential equations (SDE) for imputation of missing values for primary and secondary endpoints. Coefficients of the SDE (their expected values and variances) are estimated using available data at discrete time points. The next step is imputation of missing values (estimation of their expected values and variances) for each subject based on the SDE and available data for this subject. It is shown that there is a relationship between the SDE approach and the incremental mean imputation method introduced previously.

### What Is a Suitable Definition of Study Information in Longitudinal Clinical Trials?

❖ Guoguang Ma, Merck & Co., Inc., 10 Sentry Parkway, BL1 3, Blue Bell, PA 19422, *guoguang_ma@merck.com*; Michael Nessly, Merck Research Laboratories

***Key Words:*** study information, longitudinal trial

The total study information can be expressed either as the total number of patients or the total number of events in one sample problem. The repeated measurement setting, however, does not fit directly into this framework. The inverse of the variance of the estimated difference obtained at the planned end of the study is a logical choice. But it is dependent on the values of some unknown parameters. Wu and Lan (1992) introduced a surrogate information expression regarding the total number of measurements to be made as the total information. More recently, Kittelson et al. (2005) calculated the study information as a function of the covariance, the number of subjects, and the amount of follow up on each subject. In this work, we compare these approaches and discuss what could be a suitable definition of study information through simulation studies and a proof-of-concept study.

### A Local Influence Sensitivity Analysis for Incomplete Longitudinal Depression Data

❖ Shuyi Shen, Eli Lilly and Company, Indianapolis, IN, *Shen_shuyi@lilly.com*; Caroline Beucnkens, Limburgs Universitair Centrum; Craig Mallinckrodt, Eli Lilly and Company; Geert Molenberghs, Limburgs Universitair Centrum

***Key Words:*** sensitivity, local influence, missingness, depression

To analyze incomplete longitudinal clinical trial data, there has been a shift from simple ad hoc methods to more principled ignorable analyses, which are valid under the assumption of random missingness. While the occurrence of nonrandom dropout cannot be ruled out, models for data under nonrandom missingness have gained some popularity. Meanwhile, cautionary remarks have been issued regarding their sensitivity to often unverifiable modeling assumptions. Consequently, there is evidence for a shift towards using ignorable methodology, supplemented with sensitivity analyses to explore the impact of potential deviations of this assumption in the direction of missingness at random. One such tool is local influence, which tends to pick up a lot of different anomalies, not just deviations in the MNAR mechanism. This sensitivity analysis tool is applied to the depression data.

## 425 Time Series Outliers and Filters ●

Business and Economics Statistics Section

**Wednesday, August 9, 10:30 am–12:20 pm**

### Forecasting and Dynamic Updating of Time Series of Curves

❖ Haipeng Shen, The University of North Carolina at Chapel Hill, 304 Smith Building, Chapel Hill, NC 27599, *haipeng@email.unc.edu*; Jianhua Z. Huang, Texas A&M University

***Key Words:*** functional data analysis, principal component analysis, regularization, shrinkage, call center

We extend functional data analysis ideas to the case of a time series of curves, and develop time series models of functional data and new methods for forecasting and dynamic updating of curves. Our approach starts with dimension reduction through functional Principal Component Analysis, which is achieved via a regularized low rank approximation technique. Curve forecasting is then obtained using the principal components and time series forecasts of their coefficient series. For dynamic updating within a curve, a shrinkage approach is proposed to combine information from the previous curves and the early part of the current curve. A data-driven mechanism for selecting the shrinkage parameter is also developed, and appears to work well empirically. The methods are illustrated via a call center application, where both inter-day forecasting and intra-day updating of call volumes are needed.

### Tests for Changing Mean with Monotonic Power

❖ Ted Juhl, The University of Kansas, 213 Summerfield Hall,
Department of Economics, Lawrence, KS 66049, *juhl@ku.edu*

***Key Words:*** time series, structural change, nonparametric

Several widely used tests for a changing mean exhibit nonmonotonic power due to incorrect estimation of nuisance parameters under the alternative. In this note, we propose modified tests that have monotonic power against a wide range of alternatives of structural change. Instead of estimating the nuisance parameters based on ordinary demeaned data, the proposed tests use modified estimators based on nonparametric regression residuals. Asymptotic analysis for the proposed procedure is provided. It is shown that the estimator is consistent in the presence of structural breaks, as well as the case of constant mean. Tests for structural breaks based on such an estimator are able to avoid nonmonotonic power while still retaining the same asymptotic distribution under the null hypothesis of constant parameters.

### Some Ruin Problems with the Mixture Distribution

❖ Min Deng, Maryville University, 13550 Conway Road, St Louis,
MO 63141, *mdeng@maryville.edu*

***Key Words:*** aggregate claims, mixture distribution, ruin, surplus

This paper focus on one important aspects of Risk Theorem - the probability functions of ruin. In this paper we are going to discuss and exam some ruin problems with the mixture distribution. We also give the numerical examples to illustrate the results.

### An Asymmetric Information Modeling Framework for Ultra-High-Frequency Transaction Data: a Nonlinear Filtering Approach

❖ Yoonjung Lee, Harvard University, One Oxford Street, Cambridge,
MA 02138, *ylee@stat.harvard.edu*

***Key Words:*** non-linear filtering, high-frequency transaction data, asymmetric information modeling, market microstructure, maximum likelihood estimation, continuous-time finance

The paper proposes a new asymmetric information modeling framework that provides a theoretical explanation for some of the observed interactions among the key quantities in financial markets: the price impact of a trade, the duration between trades, and the degree of information asymmetry. In the model, a private signal is partially revealed through trades, while new public information arrives continuously at the market. The market maker utilizes a non-linear filtering technique to set a competitive price. The pricing rule depends on the actual sequence of order arrivals. The price impact of a trade tends to decrease when the duration between trades gets longer. I provide a procedure to estimate the parameters and discuss the sampling distribution of the parameter estimates.

### Variance Change in Time Series ARIMA Models

❖ Dongping Fang, SPSS Inc., 2166 Hidden Valley Drive, Naperville,
IL 60565, *dfang@spss.com*

***Key Words:*** ARIMA models, variance changes

For time series data, the variance may change with time. We consider situations where variance changes with time in a step function fashion. The goal is to detect if there are variance changes and where the change points are. This kind of variance change is studied by numerous authors (Wichern, Miller, and Hsu 1976, Hsu 1977, Tsay 1988, Inclan & Tiao 1994). Among existing detection methods, we did not find a satisfactory method. Some methods do very well if there is only one variance change. But when there are multiple variance changes, they do not perform very well any more due to masking effects. We propose a method that uses a moving window search criterion in an iterative procedure. We hope that this method can greatly reduce the masking effect because it is less likely to have more than one variance change in the same window. Simulation studies are done to compare several methods.

### Outlier Detection in Multiple Time Series by Projection Pursuit

Galeano Pedro, Universidad Santiago de Compostela ❖ Daniel Peña,
Universidad Carlos III de Madrid, Getafe, 28906 Spain,
*daniel.pena@uc3m.es*; Ruey S. Tsay, The University of Chicago

***Key Words:*** kurtosis coefficient, level shift, structural breaks, masking, projections, VARMA models

This article uses Projection Pursuit methods to develop a procedure for detecting outliers in a multivariate time series. We show that testing for outliers in some projection directions could be more powerful than testing the multivariate series directly. The optimal directions for detecting outliers are found by numerical optimization of the kurtosis coefficient of the projected series. We propose an iterative procedure to detect and handle multiple outliers based on univariate search in these optimal directions. In contrast with the existing methods, the proposed procedure can identify outliers without pre-specifying a vector ARMA model for the data. The good performance of the proposed method is illustrated in a Monte Carlo study and in a real data analysis.

### Longitudinal Microdata Outlier Detection Techniques

❖ Eric Simants, Bureau of Labor Statistics, 2 Massachusettes Ave.,
NE, Room 4840, Washington, DC 20212, *simants.eric@bls.gov*

***Key Words:*** micro-data, longitudinal-data, outliers

With more than 8.5 million new records processed each quarter the Bureau of Labor Statistics' Longitudinal Database (LDB) is one of the most comprehensive business registry lists in existence. The LDB contains business establishment data since 1990 and contains over 400 million records. Since the LDB serves as the sample frame for the Bureau's establishment based surveys, the publication of Business Employment Dynamics data, and as a research database for economists, the relevancy and usefulness of the data rely primarily on the timeliness and accuracy with which they are collected, cleaned, stored, and reported. The data are collected by each state from their Unemployment Insurance system. These data are reviewed through a rigorous process. One technique used is screening each record through a series of conditional edits based on deviations from prior values of the variable of interest.

## 426 Modeling of Genetic Data ●

Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Likelihood of a Particular Order of Genetic Markers and the Construction of Genetic Maps

❖ Susanta Tewari, University of Georgia, 210 Rogers Road, Apt. Q214, Athens, GA 30605, *statsusant@yahoo.com*

***Key Words:*** recombination, EM algorithm, simulated annealing, MLE

We model the recombination process of fungal systems for both random spore and tetrad data. Assuming no interference, a probability model framework was developed for two genes. MLE estimators for both random and tetrad data were studied. It was shown that the MLE estimator of recombination for tetrad data is uniformly more efficient than that from random spore data. The probability framework is generalized for an arbitrary number of genes, and the MLE was computed using the EM algorithm. Asymptotics for a goodness-of-fit measure, suitable for a sparse data situation, was derived. We implemented our model with data on linkage group-I of Neurospora crassa. Simulated annealing was used to search for the best order of genetic markers, and the goodness-of-fit value is consistent with the model assumptions.

## Estimating population structure

❖ Suvajit Samanta, North Carolina State University, 3223 Octavia Street, Raleigh, NC 27606, *ssamant@unity.ncsu.edu*; Bruce S. Weir, University of Washington

***Key Words:*** population structure, f-statistics, moment estimator, sampling error

The probabilities that two, three, four or two-pairs of alleles taken randomly from a population are identical by descent are parameters of great importance for population genetic theory. Only for the two-allele parameter is there a well-established estimation procedure. Moment estimation has been used when the distribution of allele frequencies over populations is not known, and maximum likelihood estimates have been given under the assumptions of normal or Dirichlet distributions. We now propose a moment estimator for the three-allele probability and we show that the large variances can be reduced by combining information over many loci. By comparing estimates of two- and three-allele probabilities we can address the relevance of the normal or Dirichlet distributions since the relationship between the two parameters is known for those distributions.

## Inference of Dynamic Activity of Transcriptional Modules

❖ Ron Yu, University of California, San Diego, 9500 Gilman Drive, MC 0359, La Jolla, CA 92093, *ronyu@modem.ucsd.edu*; Jie Liu, University of California, San Diego; Wei Wang, University of California, San Diego

***Key Words:*** transcription factors, regulatory network, gene expression regulation, motif binding, Saccharomyces cerevisiae, system biology

One of the major challenges in the post-genome era is to identify all the functional modules and decipher the regulatory network. Given a set of microarray experiments, we have developed a probabilistic model to integrate sequence motif, gene expression and ChIP-chip data to infer the activity and direct targets of transcription factors (TFs) simultaneously. The probabilistic model provides a flexible framework for data integration from different sources and removes arbitrariness commonly seen with target gene selections. As a validation, our model was applied to several yeast dataset and the results were promising. Special attention was paid to the robustness of the algorithm. The model is being applied to study other higher organisms including fly and human.

## Performance of Phylo-HMM for Evolutionary Conserved Element Detection in Promoter Region

❖ Xiaodan Fan, Harvard University, 1 Oxford Street, Statistics Department, Cambridge, MA 02138, *xfan@fas.harvard.edu*; Jun Liu, Harvard University

***Key Words:*** phylo-HMM, performance evaluation, evolutionary conserved element, promoter analysis, phylogenetic tree, sequence evolution model

An essential problem in comparative genomics is to identify functional elements through conservation analysis. Phylo-HMM was recently introduced to detect conserved elements based on multiple genome alignments, but usually it was used without strict statistical performance evaluation. We investigated the performance of phylo-HMM by simulation. In order to mimic the true promoter alignment, the parameters estimated from the true promoter alignment were used to simulate the data. It showed the performance of this approach depends on many factors beside the correctness of the model per se, such as the number of genomes used, the phylogenetic tree, signal abundance, expected signal length and the sequence evolution model. The results gave hints about where and how to use this method, and what is the performance we can expect.

## A Mixture Model Approach in Analyzing Genotype-Phenotype Association

❖ Jason Robarge, Indiana University School of Medicine, 1050 Wishard Blvd., RG4101, Division of Biostatistics, Indianapolis, IN 46202, *jrobarge@iupui.edu*; Lang Li, Indiana University; David Flockhart, Indiana University School of Medicine

***Key Words:*** mixture model, pharmacogenetics, genetic association

The motivation for pharmacogenetic studies is the potential to guide individual-specific therapy based on a clinical phenotype predicted from genetic variation. To detect true genotype-phenotype associations, analysis methods must control inflation of family wise type-I error in multiple comparisons and effectively reduce data dimensionality, while maintaining statistical power. A family of mixture models has been developed to analyze genotype-phenotype associations in genetic association studies. This method can classify a continuous or binary phenotype based on genotype and reduce the multiple-comparison burden when many genotypes are compared. The mixture model is an innovative way to formulate a multiple-comparison problem into a cluster analysis, which possesses real advantages over current methods.

## On the Detection of DNA Copy Number Changes

❖ Jie Chen, University of Missouri-Kansas City, Department of Mathematics and Statistics, 5100 Rockhill Road, Kansas City, MO 64110-2481, *chenj@umkc.edu*; Yu-Ping Wang, University of Missouri-Kansas City

***Key Words:*** ACGH, change point, DNA copy number

Cancer development is usually associated with DNA copy number changes in the genome. DNA copy number changes correspond to chromosomal aberrations and signify abnormality of a cell. Therefore, identifying statistically significant DNA copy number changes is evidently crucial in cancer research, clinical diagnostic applications, and other related genomic research. We propose to use the mean and variance change point model to study the DNA copy number changes from the array comparative genomic hybridization (aCGH) profile. The approximate p-value of identifying a change point is derived. The procedure to detect all changes in the data is carried out using binary segmentation procedure (BSP). The proposed method has been validated by Monte-Carlo simulation and applications to aCGH profiles from several cell lines such as the fibroblast cancer cell line and breast tumor cell line.

### Including Current Ages in Aggregation and Linkage Analysis of Longevity

❖ Jeanine Houwing-Duistermaat, Leiden University Medical Center, P.O. Box 9600, Leiden, 2300 RC The Netherlands, *j.j.houwing@lumc.nl*; Andrea Callegaro, Leiden University Medical Center; Marian Beekman, Leiden University Medical Center; Rudi Westendorp, Leiden University Medical Center; Eline Slagboom, Leiden University Medical Center; Hans van Houwelingen, Leiden University Medical Center

***Key Words:*** longevity, excess survival, linkage analysis

Dutch long-lived sibling pairs both older than 90 years from 100 families have been collected and genotyped for 6 microsatellite markers in a candidate region. Current age or age at death are available for the parents and siblings of the nonagenarain siblings. The standard mortality ratio using survival data for the Dutch population can be used to measure excess survival. Standard affected sibling pair linkage analysis can be used to analyze the genetic markers. We propose to extend these methods and include also information on the current ages of the siblings. To assess if the excess survival depends on the current ages of the long lived siblings, we used a Poisson model with as covariate a score based on the ages of the long-lived sib-pairs, namely the martingale residuals. For linkage analysis we propose a robust Haseman-Elston method also based on the martingale residuals.

## 427 Statistical Methods in Genetics ●

Biometrics Section, ENAR

**Wednesday, August 9, 10:30 am–12:20 pm**

### Regional Admixture Mapping and Structured Association Testing: Conceptual Unification Using a General Linear Model

❖ David Redden, The University of Alabama at Birmingham, RPHB 309 D, 1530 3rd Ave South, Birmingham, AL 35294-0022, *dredden@uab.edu*; Jasmin Divers, The University of Alabama at Birmingham; Kelly Vaughan, The University of Alabama at Birmingham; Hemant Tiwari, The University of Alabama at Birmingham; Mark Beasley, The University of Alabama at Birmingham; Jose R. Fernandez, The University of Alabama at Birmingham; Robert Kimberly, The University of Alabama at Birmingham; Rui Feng, The University of Alabama at Birmingham; Miguel Padilla, The University of Alabama at Birmingham; Nianjun Liu, The University of Alabama at Birmingham; Michael Miller, University of Minnesota; David B. Allison, The University of Alabama at Birmingham

***Key Words:*** errors in variables, linear model, admixture, association studies

Regional admixture mapping (RAM) uses individual genetic admixture estimates in order to identify genomic regions that may contain loci that influence phenotypes. Estimates of individual ancestry can be used in structured association tests (SAT) to reduce confounding due to population substructure. We provide a conceptual framework in which both RAM and SAT are special cases of a general linear model which allows for greater modeling flexibility, adaptation to multiple designs, inclusion of covariates and interaction terms, and multi-locus models. We clarify which variables are sufficient to condition upon in order to prevent spurious associations. This approach allows a far wider use of RAM and SAT models, using standard software, to address admixture as either a confounder of association studies or a tool for finding loci influencing complex phenotypes in diverse species.

### Data Normalization of Stable-Isotope Labeled Peptides in Mass Spectrometry

❖ Douglas Mahoney, Mayo Clinic College of Medicine, 201 1st Street, SW, Rochester, MN 55902, *mahoney.douglas@mayo.edu*; Ann L. Oberg, Mayo Clinic College of Medicine; Jeanette E. Eckel-Passow, Mayo Clinic College of Medicine; Terry M. Therneau, Mayo Clinic College of Medicine; Suresh T. Chari, Mayo Clinic College of Medicine; Unnikrishnan Gopinathan, Mayo Clinic College of Medicine; Lawrence E. Ward, Mayo Clinic College of Medicine; Xuan-Mai T. Persson, Mayo Clinic College of Medicine; Sreekumar Raghavakaimal, Mayo Clinic College of Medicine

***Key Words:*** mass spectrometry, normalization, stable-isotope labeling, proteomics, iTRAQ, O18

Mass spectrometry (MS) along with stable-isotope labeling (SIL) affords relative quantification of the global proteome in two or more biospecimens simultaneously. Unlike gene expression arrays which assay a predetermined set of genes with fixed probe sequences, MS requires algorithms that locate and quantify features that denote potential peptides per spectra. The nature of this global proteome scan results in a data matrix containing many censored values due to detection thresholds or peptides not being present in some biospecimens. This complicates normalization procedures which allow results to be compared across experimental runs. MVA plots reveal non-linear trends in bias between experimental runs similar to those seen in two color microarrays. Statistical issues encountered in extending normalization algorithms developed on microarray data to SIL MS experiments will be discussed.

### A Genome-Wide Study on Transcriptional Regulation of Protein Complex in Saccharomyces Cerevisiae

❖ Ching-Ti Liu, University of California, Los Angeles, 3767 Mentone Ave., Apt. 101, Los Angeles, CA 90034, *ctliu@stat.ucla.edu*; Shinsheng Yuan, University of California, Los Angeles; Ker-Chau Li, University of California, Los Angeles

***Key Words:*** protein complex, protein-protein interaction, microarray expression data

Characterization of protein-protein interaction has become a key step in understanding the cellular mechanism because most biological processes cannot be carried out without protein-protein interaction. This presentation will focus on a set of well-validated protein complex data from MIPS. It is commonly believed that functionally related genes tend to have similar patterns of variation in expression. However, we found notable contradictory evidence. Indeed, except for a few large protein complexes such as cytoplasmic or mitochondrial ribosomal proteins, the co-expression level of genes from the same complex is on the average not much higher than that from a pair of randomly selected genes. We find this result to be true for four well-known gene expression datasets. We also study ways of incorporating the ChIP-on-Chip data for protein complexes analysis.

## Comparative Validity and Power of Methods for Association-Testing with Related Individuals

❖ Hemant Tiwari, The University of Alabama at Birmingham, 1665 University Blvd., 327C RPHB, Birmingham, AL 35294, *htiwari@uab.edu*; Amit Patki, The University of Alabama at Birmingham; Mark Beasley, The University of Alabama at Birmingham; David B. Allison, The University of Alabama at Birmingham

***Key Words:*** association, within-cluster resampling, power, type 1 error, assoc, FBAT

Hoffman et el. (2001) proposed an elegant resampling method for analyzing clustered binary data. The focus of their paper was to perform association tests on clustered binary data using within-cluster-resampling (WCR) method. WCR can be easily extended to continuous data and is a computationally intensive but simple and highly flexible method. Considering family as a cluster, one can apply WCR to familial data. In this presentation, we compare WCR with other existing methods of association for familial data such as the maximum-likelihood-based (George & Elston, 1987) and the family-based association test (FBAT) (Laired et al, 2001). We evaluated these methods' performance based on type 1 error rates and power of the study for given genetic parameters using simulation. WCR outperformed all other methods with respect to both type 1 error rates & power in all scenarios considered.

## Measuring Deviations from Hardy Weinberg Equilibrium in SNP Data

❖ Vernon S. Pankratz, Mayo Clinic College of Medicine, 200 1st Street, Division of Biostatistics, Rochester, MN 55905, *pankratz.vernon@mayo.edu*; Robert Vierkant, Mayo Clinic College of Medicine

***Key Words:*** genetics, genotyping quality assessments, approximate inference

Hardy Weinberg Equilibrium (HWE) is rapidly achieved in a randomly mating population. Testing for deviation from HWE expectations can provide an assessment of the quality of genotyping results, and it can also provide information about other population genetic parameters. This check is often limited to a chi-square test comparing observed to expected genotype counts, although other possibilities exist. We present a measure of departure from HWE, and an approximation to its standard error. We also present results from simulations that assess the performance of the proposed measure in comparison to existing tests and measures. Partial results suggest that the approximations are close to empirical estimates. However, for small sample sizes the estimates can significantly underestimate the empirically derived values. We also assess small sample corrections to the proposed estimates.

## An algorithm for finite Markov Chain Imbedding Approach

❖ Lung-An Li, Academia Sinica, Institute of Statistical Science, Taipei, 11529 Taiwan, *lali@stat.sinica.edu.tw*

***Key Words:*** transition matrix, offspring

The finite Markov chain imbedding technique is a powerful tool to study the exact distribution of patterns in sequences. There are many practical applications by utilizing this technique. However, the key step of this method is to construct a transition matrix for the imbedded Markov chain, which is nontrivial. It heavily depends on the problem itself, especially case by case. This paper studies the feasibility to build an algorithm for such a construction shared commonly for most cases.

# 428 Contributed Posters

General Methodology, Section on Bayesian Statistical Science, Section on Statistics in Epidemiology, Biometrics Section, Section on Statistical Computing, Section on Health Policy Statistics, Section on Physical and Engineering Sciences, IMS

**Wednesday, August 9, 10:30 am–12:20 pm**

## The Relative Contribution Measures in Multilevel Modeling

❖ Liyi Cen, University of Pennsylvania, 3504 Shetland Way, Westville, NJ 08093, *lc8132@albany.edu*; Zhen Chen, University of Pennsylvania; Daniel E. Polsky, University of Pennsylvania; Kevin G. Volpp, University of Pennsylvania

***Key Words:*** variation, fixed effect modeling, multi-level modeling

The omega statistic was first introduced to compare the relative importance of two sets of explanatory variables in explaining the variation of the outcome. While there are many applications of omega statistic in fixed effect models, few exist in the literature in the context of multi-level modeling. When the data are nested, will the omega statistics be different between fixed effects models and multilevel models? How to calculate the omega statistics if the two sets of predictors are incomparable and correlated. In this study, we will address the above two questions by comparing the difference between standard logistic regression and multi-level modeling in predicting the variation in the response by patient level variables and hospital characteristics using simulations and Medicare file (MedPAR) in the context of examining mortality following hospital admission for a medical condition.

## Uncertainty Computation for the Virtual Cement and Concrete Testing Laboratory Measurements

❖ Blaza Toman, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *toman@nist.gov*; Charles Hagwood, National Institute of Standards and Technology; Adriana Hornikova, National Institute of Standards and Technology; Hung-kung Liu, National Institute of Standards and Technology; Nien Fan Zhang, National Institute of Standards and Technology

***Key Words:*** virtual measurements, uncertainty, Bayesian hierarchical models, mixture design

The Virtual Cement and Concrete Testing Laboratory (VCCTL) is a software system, being developed at the National Institute of Standards and Technology, that is to be used for evaluation and testing of new cement mixtures. This presentation will describe Bayesian statistical modeling of the virtual measurements which uses information from related physical measurements and models. The use of these statistical models, for validation of the software, and for computation of uncertainty estimates for the virtual measurements, will be described.

## Alternative Methods for Variable Selection in Generalized Linear Models with Binary Outcomes for Incomplete Data

❖ Gang Liu, University of California, Los Angeles, 3191 S. Sepulveda Blvd., APT 102, Los Angeles, CA 90034, *lgwj@ucla.edu*

***Key Words:*** SIAS & ITS, MCMC, model selection, missing data, GLM, multiple imputation

A crucial problem in building a multiple regression model is the selection of predictors to include. Here, we develop two strategies for variable selection in logistic and probit regression models with missing values on explanatory variables. One approach, which we call "impute, then select" (ITS), involves initially performing multiple imputation and then applying Bayesian variable selection to the multiply imputed data sets. The second strategy, which we call "simultaneously impute and select" (SIAS), is to conduct Bayesian variable selection and missing data imputation simultaneously within one Markov Chain Monte Carlo (MCMC) process. For illustration purposes, ITS and SIAS are applied to a data set from Los Angeles Foster Care Study where some explanatory variables have a substantial amount of missing data.

## Analysis in Opinions about the Death Penalty in U.S. States

❖ Shouhao Zhou, Columbia University, 1255 Amsterdam Ave., Room 1017, Department of Statistics Columbia Univ, New York, NY 10027, *zhou@stat.columbia.edu*; Andrew Gelman, Columbia University

***Key Words:*** time series, hierarchical modeling, Gibbs sampler, AR(p) error

We use hierarchical logistic regression model to analyze opinions about the capital punishment from two datasets by different polls in U.S. states from 1953. Some sampling methodology for posterior conditional distribution of time correlated variables is introduced, as well as model evaluation by counting the degree of freedom for different factors. Interesting estimates are reported and discussed.

## Join-Point Analysis of Survival Data

❖ Sandra Hurtado Rua, Northern Illinois University, Mathematical Sciences, DeKalb, IL 60115-2888, *hurtado@math.niu.edu*; Sanjib Basu, Northern Illinois University

***Key Words:*** Bayesian analysis, joinpoint regression, survival analysis, hazard function, cancer data

The identification of change of tendency over time in cancer data is important for determining agents that affect cancer. Join-point models help to describe continuous changes in trends of data and when such changes occur. Join-point models play an important role since they help to measure the impact of health policies and treatments, and in general, they help to locate and explain changes in the survival curve. We model join-point hazard and survival functions in cancer data that account for censored and uncensored observations. Given k join-points, the posterior distributions of the parameters of their models are estimated using Bayesian analysis with Markov chain Monte Carlo sampling. We further propose model selection approaches to determine the number of join-points in the model. We illustrate the method using monthly data from the SEER cancer database.

## Design-Based Calibration Estimators for Measurement Error in the Cox Model

❖ Thomas Lumley, University of Washington, Seattle, WA 98195, *tlumley@u.washington.edu*; Pam Shaw, University of Washington

***Key Words:*** raking, measurement error

Many techniques exist for using a small validation subsample to correct for measurement error in predictors when error-free data are available in the subsample. We present a technique that is valid when the validation subsample has classical measurement error and no assumptions are made about error model in the full cohort. The estimation is based on survey calibration for the cohort total of the estimating function, using biased estimating function in the cohort as calibration variables.

## ROC Curve Analysis in Osteoporosis Screening

❖ James Powers, The University of North Carolina at Chapel Hill, Dept of Biostatistics, UNC-CH, Chapel Hill, NC 27599, *powers.james@gmail.com*; Margaret Gourlay, The University of North Carolina at Chapel Hill; Kristine Ensrud, VA Medical Center

***Key Words:*** ROC curve, generalized linear model, osteoporosis, longitudinal study

Regression models for receiver operating characteristic (ROC) curve analysis, based on the theory of generalized linear models, have been proposed (Pepe 2000, Alonzo and Pepe 2002, Pepe and Cai 2004). These parametric distribution-free methods offer the ability to compare distributions of test results while accounting for important covariates. Correlation structures for clustered data also may be handled with this method. This poster applies ROC regression methods to bone density data from a large longitudinal observational study and discusses potential relevance to osteoporosis screening. Proposals for future statistical research are offered to supplement the data analysis.

## A Soft Endpoint for HIV-1Ba-l Growth in Cervical, Rectal, and Tonsular In Vitro Assays

❖ Sandra Senneke, BioStat Solutions Inc., 5 Ridgeside Court, Suite 202, Mount Airy, MD 21771, *ssenneke@biostatsolutions.com*; Nicola Richardson-Harman, BioStat Solutions Inc.; James Cummins, Southern Research Institute; Carol Lackman-Smith, Southern Research Institute; Christina Bromley, BioStat Solutions Inc.; Patricia Reichelderfer, National Institute of Child Health & Human Development

***Key Words:*** in-vitro, end point, time series, HIV, assay, multi-center

Measurement of virus growth over time is currently achieved by visual inspection of a growth curve. We introduce a soft endpoint calculation for in vitro HIV-1Ba-l growth in a time series. Virus replication was assessed using the HIV gag p24 protein assay. The objective of the soft endpoint is to determine the start of the phase of virus growth where maximum, and in some cases exponential growth, has been achieved. This was calculated as the last day when the change in p24 between two consecutive days was greater than the p24 SD for the entire assay. Calculations were based on data from a multi-center study, where virus growth was measured at approximately three day intervals for up to 15 days. The soft endpoint was then used in a binary growth categorization for each assay and candidate assay factors were tested using logistic regression analyses.

## Feasibility of Genome-Wide Haplotype Association Studies with Small Sample Size and Sparsely Spaced SNPs

❖ Shaokun Chuai, University of Pennsylvania, BAC 529C Blockley Hall, CCEB, 423 Guardian Drive, Philadelphia, PA 19104, *schuai@cceb.upenn.edu*; Nandita Mitra, University of Pennsylvania; Nathan Ellis, The University of Chicago

***Key Words:*** genome-wide scan, haplotype analysis, genetics epidemiology, SNP

We investigated the feasibility of localizing disease genes using sparse genome-wide panels of SNPs in the Ashkenazi Jewish population taking advantage of large linkage disequilibrium (LD) blocks. We performed two trials: (a) autosomal recessive Bloom syndrome targeting the BLM gene (3258 SNPs; 10 cases; 31 controls) and (b) autosomal dominant hereditary nonpolyposis colorectal cancer targeting the MSH2 gene (8549 SNPS; 13 cases; 63 controls). We performed sliding-window haplotype analyses using a generalized linear models framework. In both trials, significant associations were detected for several multi-locus haplotypes which are within a few million base pair regions that contained the targeted disease genes. This work demonstrates the power of the LD mapping approach in an isolated population and its general applicability to the identification of novel cancer-causing genes.

## Bayesian State-Space Models for Predicting Temporal Gene Expression Profiles

❖ Yulan Liang, University at Buffalo, 249 Farber hall, 3435 Main Street, Buffalo, NY 14214, *yliang@buffalo.edu*; Arpad Kelemen, Niagara University

***Key Words:*** Bayesian approaches, state space models, prediction, temporal gene expression, MCMC algorithms, deviance information criteria

Predictions the genome dynamic behavior is a challenging problem in genomic research and estimations the temporal correlation is one key. However, the unevenly short time courses with sudden change make the predictions difficult. In this paper, we develop two types of Bayesian State Space (BSS) models: multiple univariate time varying BSS and multivariate BSS models for predicting the gene expression profiles associated with diseases.In the univariate time-varying BSS model, we treat both stochastic transition and observation equations time-variant. In the multivariate model, we include temporal correlation structures with various prior settings. The unseen time points are treated as hidden state variables and are estimated by MCMC algorithms. We compared two type models using Deviance Information Criteria and applied our models to multiple tissues datasets.

## Handling Missing Data for Smoking Cessation with Bootstrap, Trees, and Multiple Imputation

❖ Jeff Thostenson, University of Arkansas for Medical Sciences, 4301 W. Markham Street, # 781, Little Rock, AR 77205-7199, *jdthostenson@uams.edu*; Lowell C. Dale, Mayo Clinic College of Medicine; Darrell Schroeder, Mayo Clinic College of Medicine; Heike Hofmann, Iowa State University

***Key Words:*** missing data, multiple imputation, bootstrap, classification trees

Due to incomplete follow-up data, smoking intervention studies can suffer missing outcomes. The current accepted approach for missing outcomes assumes treatment failure (i.e. smoking) in place of missing outcomes, with secondary analyses excluding incomplete cases. Analysis of a pilot study to assess the efficacy of a tobacco quitline to increase abstinence from smoking following an initial consultation at the Mayo Clinic Nicotine Dependence Center suffered missing data problems. Due to differing rates of missing data in the study arms, the two analyses produced conflicting results. Multiple Imputation (MI) analyses gave unreasonable predictions. A method of bootstrapping classification trees gave more believable predictions. Thus that prediction method was used as the imputation step of a new MI analysis, giving more reasonable predictions and using Rubin's Rules to assess variability.

## Analysis of Longitudinal Case-Control Zero-Heavy Data: Vaginal Shedding of HIV

❖ Leann Myers, Tulane University, Department of Biostatistics, Tulane SPHTM, 1440 Canal Suite 2001, New Orleans, LA 70112, *myersl@tulane.edu*; Hao He, Tulane University; Patricia Kissinger, Tulane University

***Key Words:*** longitudinal data, case-control, HIV, regression

A recent study of vaginal HIV shedding followed HIV-positive patients with trichomonas matched to HIV-positive trichomonas-free patients for three visits. Vaginal shedding of the HIV virus was measured at baseline, at one month (post-treatment for the trichomonas patients), and at three months. The response distribution at each time point was a mixture of a proportion of zeros and a highly skewed distribution for those with detectable shedding. The analytical challenge was how to analyze these data to address the question of whether trichomonas increases vaginal HIV shedding, given that the analysis must incorporate the matching scheme, the longitudinal aspects of the data, and the mixture distribution of responses. Two-part models, logistic regression approaches, and other strategies were compared in terms of point estimators, variability, assumptions, and hypothesis testing.

## SAS Estimation of Standard Errors for Partial Least Squares Regression

❖ April Grant, University of Arkansas for Medical Sciences, 4301 W. Markham Street, Slot 512 29, Little Rock, AR 72205, *grantaprils@uams.edu*; David K. Williams, University of Arkansas for Medical Sciences; Zoran Bursac, University of Arkansas for Medical Sciences; Geoffrey M. Curran, University of Arkansas for Medical Sciences

***Key Words:*** partial least squares regression (PLS), standard error calculation

Partial least squares (PLS) regression is a useful predictive modeling tool when complex correlation structures exist in data. PLS has well documented applications in the fields of econometrics and chemometrics. While the SAS PROC PLS is able to perform partial least squares regression, estimates of the standard error (SE) are currently unavailable through the procedure syntax or the output. Empirical (bootstrap and jackknife) and closed form solutions are among the possible methods to calculate SE. This research will present a SAS Macro used to estimate SE and show the comparability of the SE with some standard analytical methods through simulation and application examples. By making this macro available to practitioners, PLS can further be applied to other fields and prove a valuable tool to researchers.

## Assessment of Capture-Recapture Methods in Estimating Populations of Oklahomans with Birth Defects

❖ Robert C. Wild, The University of Oklahoma Health Sciences Center, Department of Biostatistics and Epidemiology, 801 NE 13th St CHB 309, Oklahoma City, OK 73109, *robert-c-wild@ouhsc.edu*; Barbara Neas, The University of Oklahoma Health Sciences Center

***Key Words:*** capture-recapture, Indian Health Service, multiple-records systems, log-linear models, bootstrap, population size estimation

Originally developed by ecologists, capture-recapture methods are used in epidemiology to obtain accurate estimates of true population sizes from independent but overlapping samples from the same population. The Oklahoma Birth Defects Registry (ODBR) is a population-based, active surveillance registration system that includes data from various standard sources and other specialty sources such as the Indian Health Service. Recently, capture-recapture methods have been developed that estimate populations sizes from samples that have heterogeneous inclusion probabilities. These methods can be applied to Oklahomans with birth defects. This paper addresses the key statistical techniques in capture-recapture methods that are potentially applicable to the ODBR. Log-linear model selection and variance estimation

methods that include MLE and empirical bootstrap techniques are also presented.

## Comparing Methods of Examining Trend Data for a Large Population

❖ Kristen Eberly, The University of Oklahoma, P.O. Box 26901, CHB309, Oklahoma City, OK 73190, *Kristen-Eberly@ouhsc.edu*; Barbara Neas, The University of Oklahoma Health Sciences Center; David M. Thompson, The University of Oklahoma

***Key Words:*** trend, graphical comparison, large sample size, confidence interval, moving average, test of trend

The objective of this study is to compare methods of evaluating time trend data to determine a meaningful trend. Deaths from lung cancer and related diseases in OK from 1980 to 2004 are shown for the entire population as well as stratified by sex and age. Evaluation of trends in large datasets can often produce statistically significant results due to the sample size. With the use of traditional alpha, non-meaningful trends are almost always statistically significant. Common methods such as confidence intervals, moving averages, and trend tests have been shown to be statistically significant. When looking at short time intervals compared to long intervals, statistical significance is still achieved when the actual trend may not be considered practically significant. In this study, graphical comparisons demonstrate the most meaningful interpretations.

## Repeated Measurement of Gastric Myoelectric Activity in Obese Participants

❖ Jiangyue Wang, The Pennsylvania State University, Drexelbrook Drive, Apt. 68, Room 6, Drexel hill, PA 19026, *juw120@psu.edu*

***Key Words:*** repeated measures analysis, mixed model, gastric myoelectrical activity

This study observed the reactions of the stomach to food viewing and food eating in obese people. 20 obese and 20 normal-weighted subjects viewed a video clip of food for 6 minutes; consumed food till full; and viewed the same video clip of food 30 minutes after the meal. Cutaneous electrogastrograms (EGGs) were recorded during the entire procedure. Repeated measures analysis showed that during the first food image viewing, obese participants had less of an increase in 3 cpm activity than normal-weighted participants. Profile analysis suggested that in the obese group, the percentage of 3 cpm activity decreased in the third and forth 6-min periods and increased to the highest level in the last 6-min period. These data suggested that obese people reacted to food images and to food eating differently from normal-weighted people.

## Salamander Cannibalism

❖ Jean Ellis, California State University, East Bay, 1334 E. 25th Street, Oakland, CA 94606-3242, *jellis@horizon.csueastbay.edu*; Erica Wildy, California State University, East Bay

***Key Words:*** controlled environment, density, food supply, cannibalism, salamanders

Dr. Erica Wildy, a professor at California State University East Bay (CSUEB), and a graduate Statistics student at (CSUEB), Jean C. Ellis, collaborated, at the recommendation of the Statistics Department Chair, Dr. Julia Norton, in the analysis of this data. Methodology They created the poster to investigate the behavior of salamanders in various environments. Dr. Wildy observed few salamanders thrived in more dense water than in larger bodies of water. There were no salamander body parts found in the more dense water area. This led to closer examination of salamanders and their surroundings to see if environmental conditions aid in the production of cannibalistic salamanders. Two factors are used to conduct this investigation, food supply and water density. High and low food supply and high and low density formed the four controlled environments used.

## Determining an Optimal Exposure Metric from a Bivariate Distribution of Asbestos Exposures (Length, Diameter) in a Cohort of South Carolina Textile Workers

❖ Stephen Gilbert, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Room 216 Mailstop C15, Cincinnati, OH 45226, *sgilbert@cdc.gov*; Leslie T. Stayner, University of Illinois at Chicago; Eileen D. Kuempel, National Institute for Occupational Safety and Health; John D. Dement, Duke University Medical Center

***Key Words:*** simulation, epidemiology, public health, risk assessment, modeling

What combination of asbestos fiber length/diameter cells best predicts human lung cancer? Original samples from a SC textile plant were reanalyzed using TEM to obtain more detailed bivariate (length, diameter) asbestos fiber size distributions by department and job. Separate Cox models were fit to cumulative exposure for all 36 length/diameter cells and other combinations of those 36 cells. While nearly all exposure cells and chosen combinations showed significant dose-response relationships, the fits to metrics with fibers >40μm long provide the best fits. Since exposure occurs in nearly all 36 cells we wanted to find the optimal combination of those 36 cells. So, single exposure metrics were computed as weighted sums of the 36 cells using random weights and simulated annealing. The simulation results support the notion that metrics that include long fibers provide the best fits.

## Analysis of Respiratory Viral Exposure and Timing of Exposure During Infancy with Development of Childhood Asthma

❖ Pingsheng Wu, Vanderbilt University School of Medicine, 1161 21st Ave., S., T 1218 MCN, Nashville, TN 37232, *pingsheng.wu@vanderbilt.edu*; Tebeb Gebretsadik, Vanderbilt University School of Medicine ; William Dupont, Vanderbilt University School of Medicine; Marie Griffin, Vanderbilt University School of Medicine; Rachel Enriquez, Vanderbilt University School of Medicine; Kecia Carroll, Vanderbilt University School of Medicine; Tina Hartert, Vanderbilt University School of Medicine

***Key Words:*** childhood asthma, respiratory viral exposure, timing of exposure

Asthma is one of the most common chronic conditions of childhood, and severe viral infection is one of significant risk factors for developing asthma. Our study objective is to identify the role of seasonal respiratory viral exposure and the timing of exposure during infancy on the risk of childhood asthma. We analyzed over 85,000 infants in the Tennessee Medicaid program from 1995 - 2004 and followed them to age 5 years. We applied the logistic regression model controlling for confounders. Restrictive cubic spline of infant age at the peak of viral circulation was used to describe the relationship of infant age and childhood asthma. Weekly bronchiolitis hospitalization counts were used as a surrogate marker of viral intensity. The model was further examined in a separate test dataset. Our results may lead to new strategies for both the prevention and treatment of childhood asthma.

## Evaluating the Predictiveness of a Continuous Marker in Case-Control Design

❖ Ying Huang, University of Washington, 13614 NE 11th Street, T12, Biostatistics, Bellevue, WA 98005, *ying@u.washington.edu*; Margaret S. Pepe, Fred Hutchinson Cancer Research Center/ University of Washington

***Key Words:*** predictiveness curve, ROC, case-control, monotone likelihood ratio

To describe the predictive capacity of a continuous biomarker, Pepe and Huang (2006) proposed "predictiveness curve", defined as a plot of risk versus the distribution function of the marker, and proposed estimating a monotone increasing predictiveness curve using a flexible parametric model for cohort data. However, case-control studies are most often performed in the early phases of biomarker development. In this report, we show the relationship between predictiveness curve and ROC curve given disease prevalence. Two strategies for estimating the monotone increasing predictiveness curve are proposed. The first one is to model the monotone likelihood ratio of biomarker. And the second one is to estimate a generalized predictiveness curve, defined as a plot of risk versus its distribution function. Semi-parametric likelihood approach is used for estimating the distribution of risk.

## Variable Selection in the Multivariate Adaptive Regression Splines (MARS)-Logit Models To Detect Gene-Gene Interactions

❖ Huiyi Lin, Louisiana State University Health Sciences Center, 221 W. Gatehouse Drive, Apt. A, Metairiie, LA 70001, *yi0407@yahoo.com*; Cruz Velasco, Louisiana State University Health Sciences Center; Jennifer J. Hu, Louisiana State University Health Sciences Center

***Key Words:*** variable selection, gene-gene interaction, MARS-logit, multivariate adaptive regression splines, simulation

Gene-gene interactions play critical roles in complex diseases. The MARS-logit models, which sequentially combine MARS and logistic models, can be applied for examining high-order interactions. A simulation study was carried out to assess the performance of the MARS-logit model using a number of variable selection approaches. Data on 20 single nucleotide polymorphisms (SNPs) were generated for 3000 subjects. Ten replicates were used to test 2- or 3-way interactions. Empirical guidelines for variable selection are provided. However, a software that combines both methods is needed for further evaluating the MARS-logit models. The MARS-logit models selected the correct SNPs, mode of inheritance and interaction type more often than MARS alone. Therefore, the MARS-logit hybrid model may be a better tool for exploring gene-gene interactions.

## A Comparison of Three Categorical Data Analysis Methods Applied to Survey Data

❖ Barbara Neas, The University of Oklahoma Health Sciences Center, 801 NE 13th Street, CHB 309, Oklahoma City, OK 73104, *barbara-neas@ouhsc.edu*; Hani Dimassi, The University of Oklahoma Health Sciences Center; David M. Thompson, The University of Oklahoma; Betty J. Pfefferbaum, The University of Oklahoma Health Sciences Center

***Key Words:*** categorical survey data, GEE, log-linear analysis, latent class analysis

This work was motivated from a secondary analysis of Oklahoma City bombing survey data of middle school students. We applied generalized estimating equations (GEE), log-linear analysis (LLA) and latent class analysis (LCA) to a subset of data and compared the results. These results provide the basis for a series of simulated data sets with different patterns of associations. The GEE analysis included school as a cluster variable and resulted in separate models for males and females with significance between the risk factor and outcome for males only. LLA showed the dependencies among the variables. LCA produced underlying constructs for the outcome. The construct representing presence of the outcome included the risk factor and a pseudo-age variable, while the construct representing absence of the outcome included the risk factor and gender. Other examples are discussed.

## Discrete Time Analysis of Mortality among Persons with Diabetes Using the Logit Model

❖ Edward Tierney, Centers for Disease Control and Prevention, 4770 Buford Hwy., NE, Mailstop K-10, Atlanta, GA 30341, *ext5@cdc.gov*

***Key Words:*** Bayesian logistic regression, mortality, discrete time analysis, BMI, diabetes

To assess the role of body mass index (BMI) in mortality, we modeled mortality by age, race, sex, and BMI for a sample of persons with diabetes from 1986 to 1994 with follow up through December 31, 1997. Data are from the National Health Interview Survey linked to the National Death Index and include 1,032 deaths among 28,783 person years of follow-up. We discretized the data by one year intervals from 19 to 85 years of age and used Bayesian logistic regression to model hazard rates. We used a cubic spline with five knots for age (as the time variable) and included terms for race, sex, and BMI. Odds ratios and 95% posterior intervals, compared to the normal BMI groups, are 2.3 (1.4, 3.3) for underweight, 0.89 (0.76, 1.0) for overweight, 0.85 (0.69, 1.0) for obese, and 0.95 (0.75, 1.2) for severely obese. Only the underweight group has significantly higher death rates than the normal BMI group.

## Permutation Test and Its Application to the Youth Partners in Care (YPIC) Study

❖ Lingqi Tang, University of California, Los Angeles, 10920 Wilshire Blvd., Suite 300, Health Services Research Center, Los Angeles, CA 90024-6505, *lqtang@ucla.edu*; Naihua Duan, University of California, Los Angeles; Joan Asarnow, University of California, Los Angeles

***Key Words:*** permutation, ANCOVA, nonresponse weights

Due to recent advances in computing power the permutation test has become a popular tool in the analysis of randomized controlled trials because it requires fewer assumptions than a parametric analysis. Although the permutation test is straightforward, there is no general agreement concerning an appropriate way of implementing the permutation test with analysis of covariance models (ANCOVA), especially when missing data complicate the problem. In this presentation, we demonstrate an application of permutation test to the Youth Partners in Care (YPIC) study. We conducted sensitivity analyses for intervention effects using a design-based permutation test to ascertain whether our findings from ANCOVA model are sensitive to model assumptions. Similar results from the permutation test and parametric model confirm that our findings are robust to model assumptions.

## Comparison of Resampling Methods to Least Squares and Maximum Likelihood Methods for Estimating Fracture Strength Parameters of Glass Test Specimens

❖ Gerald Shaughnessy, University of Dayton, 5834 Daffodil Circle, Dayton, OH 45449, *shaughne@notes.udayton.edu*; Peter Hovey, University of Dayton

*Key Words:* Weibull, resampling methods

The use of the Weibull model to characterize the failure of brittle materials has enjoyed widespread acceptance. The Weibull probabilistic model is founded on a "weakest link" mode of failure which allows the probability of failure to be related to specimen size. This paper will utilize resampling methods to estimate the strength characteristics of glass specimens. The results of the resampling estimation will then be compared to the traditional estimation procedures using least squares and maximum likelihood.

### Stepwise Permutation Tests Using Medians

❖ Scott Richter, The University of North Carolina at Greensboro, 1000 Spring Garden Street, MATHEMATICAL SCIENCES, GREENSBORO, NC 27402, *scott_richter@uncg.edu*; Melinda McCann, Oklahoma State University

*Key Words:* simultaneous inference, pairwise comparisons, median difference, permutation test

Richter & McCann (2005) presented a single step median-based multiple comparison procedure for assessing evidence of group location differences. The sampling distribution was based on the permutation distribution of the maximum median difference among all pairs, and provides strong control of the FWE. We extend this idea to develop a stepwise procedure for comparing group location. We show that the stepwise procedure provides greater power for detecting group location differences while still maintaining strong FWE control.

### Resampling Methods in Asymptotic Statistical Inferences

❖ Fassil Nebebe, Concordia University, 1455 de Maisonneuve Blvd., Montreal, PQ H3G1M8 Canada, *fnebebe@vax2.concordia.ca*; Tak K. Mak, Concordia University

*Key Words:* resampling designs, intraclass correlation, heteroscedasticity

The bootstrap is known to be useful in producing asymptotic variances of estimators which are otherwise difficult to obtain analytically. This application can be made more efficient by special resampling designs based on linear regression models. The present paper studies the use of these designs to derive efficiently the sampling distribution of any statistic for any sample size. The theory developed will be applied to some well known statistical inference problems, such as intraclass correlation estimation and the test for.

### Surface Shape Analysis with an Application to Brain Cortical Surface Analysis in Schizophrenia

❖ Christopher Brignell, University of Nottingham, School of Mathematical Sciences, University Park, Nottingham, NG72RD UK, *pmxcjb@nottingham.ac.uk*; Ian Dryden, University of Nottingham; William Browne, University of Nottingham

*Key Words:* cortical surface, high-dimensional, likelihood, neuroscience, shape, symmetry

In many application areas it is of interest to compare the shapes and sizes of high-dimensional surfaces, and to investigate symmetry. We focus on a particular application in neuroscience, investigating large scale cortical shape differences between control and schizophrenia patients. We introduce an automatic maximum likelihood method for brain registration, identifying the inter-hemispherical join, the anterior commissure, and the posterior commissure. Likelihood based inference is considered, and significant differences between the two groups are observed. The model is extended to account for curvature in the inter-hemispherical join. General practical issues in high-dimensional data analysis will be discussed.

## 429 Health Policy Statistics Speaker with Lunch (fee event)

Section on Health Policy Statistics

**Wednesday, August 9, 12:30 pm–1:50 pm**

### Mapping and Applications in Spatio-Temporal Public Health Data

❖ Bradley P. Carlin, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware Street S.E., Minneapolis, MN 55455, *brad@biostat.umn.edu*

*Key Words:* GIS, hierarchical models, mapping, MCMC, public health, spatio-temporal data

In the past decade, hierarchical Bayesian modeling of geographically referenced data has been abetted greatly by two significant developments in computing: geographic information systems (GISs) for the simultaneous graphical display and summary of the data and Markov chain Monte Carlo (MCMC) methods for the estimation of relevant posterior quantities. This talk surveys the author's work in this area, illustrated with examples drawn from public health practice. For instance, we will describe one study attempting to relate radon exposure and lung cancer near a former nuclear weapons facility in southwestern Ohio and another modeling pediatric emergency room visit counts for asthma as a function of ozone exposure and a range of sociodemographic variables among children in the Atlanta, Georgia, area.

## 430 Biopharmaceutical Section Roundtables with Lunch (fee event)

Biopharmaceutical Section, Section on Statistical Consulting, Section on Statistical Graphics

**Wednesday, August 9, 12:30 pm–1:50 pm**

### Procedures for Controlling the False Discovery Rate in the Multiple Comparison Problems Involving Multiple Endpoint Tests in Clinical Trials

❖ Arunava Chakravartty, University of California, Riverside, 200 W. Big Springs Road, Apt 4, Riverside, CA 92507, *achak001@ucr.edu*

*Key Words:* false discovery rate, multiple comparisons, family wise error rate, fallback tests, gatekeeper tests

Multiple comparison procedures are particularly important when performing simultaneous hypothesis testing on data with multiple endpoints rising out of the clinical trials. In testing multiple hypotheses in clinical trials, control of family-wise error rate must be considered. These multiple endpoints usually are classified as primary, secondary, and tertiary points, based on a certain criterion. Several methods are considered, many based on Bonferroni's method, including procedures proposed by Holm, Hochberg, Simes, and Hommel. There also exist certain fallback procedures proposed by Weins, where the Type I error rate is partitioned among various hypotheses of interest. Parallel gatekeeper strategies can be used to control the Type I error rate for multiple primary endpoints, while allowing consideration for the secondary points.

## Bayesian Approaches to Clinical Trials

❖ Dinesh Kumar, Eisai Medical Research, 38 Grayson Place, Teaneck, NJ 07666, *dinesh_kumar@eisai.com*

*Key Words:* Bayesian, trials, drug, development, rationale, applicability

Have you ever conducted or have thought about a trial that employed Bayesian methods. Bayesian methods offer a valuable chance to advance the development of trials in all phases as they get more complicated and a large amount of relevant data is accumulated. Bayesian analysis naturally facilitates use of sequential experiments potentially allowing for smaller, more informative trials. They are ideally suited for dose-finding studies, pediatric extension trials, modifying interim trial designs, trials requiring historical control, selection of biomarkers from pilot studies, and testing the sensitivity of results in later trial phases. They are being used in the medical device industry for efficacy trials and to address safety concerns in drug industry. Let us discuss the rationale underlying Bayesian trials, their pros and cons, and their applicability to the process of drug development.

## How Can Statisticians Better Contribute to Clinical Trials and Intervention Studies in Developing Countries?

❖ Craig Borkowf, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Mail Stop E48, Atlanta, GA 30333, *uzz3@cdc.gov*

*Key Words:* collaboration, consulting, developing country, HIV, infectious disease, Kisumu breastfeeding study

Statisticians must collaborate actively with other medical researchers in combating new and established infectious diseases in developing countries. Challenges include designing studies that reflect local realities, developing rigorous but practical statistical analyses, and dealing with cultural differences and long-distance communications. Rewards include travel opportunities to provide statistical support and training for local researchers and the chance to contribute to global public health. The organizer will share his experiences working on the Kisumu Breastfeeding Study (KiBS), a CDC/KEMRI collaboration. In this study, HIV-positive pregnant women living in western Kenya were given highly active antiretroviral therapy to suppress their viral loads and thereby reduce the risk of mother-to-child transmission. Participants are encouraged to share experiences from their own projects.

## Use of Propensity Scoring in Clinical Trials

❖ Daniel P. Reyner, Zimmer, Inc., Clinical Affairs, PO Box 708, Warsaw, IN 46581, *daniel.reyner@zimmer.com*

*Key Words:* propensity scoring, clinical trials, bias

Prior to analysis of clinical endpoints, the trial statistician may encounter significant and clinically meaningful between-group differences in distributions of potentially influential baseline and demographic characteristics. Such differences elevate concern of biased estimates of treatment effects. Examples may include studies with significant losses over follow-up, interim analyses, and post-approval studies. Discussion will focus on the use of propensity scoring in such situations, as well as various types of propensity scoring methods and related features.

## Increasing the Usage of Data from Cross-Over Studies

❖ Yonghua Wang, Bristol-Myers Squibb Company, Route 206 and Provinceline Road, Mail Stop E1211, Princeton, NJ 08543, *yonghua.wang@bms.com*

*Key Words:* cross-over, unbalanced, incomplete, analysis

It is still common to discard data if a subject did not finish a sequence in clinical pharmacology studies. The analysis of data from an unbalanced, incomplete cross-over design in SAS procedures may not include all available data to make maximum use of the data. These kind of studies are relatively small. The more data are used in the analysis, the better. If the FDA's guidance for industry, Statistical Approaches to Establishing Bioequivalence, is followed closely, subjects who did not complete the study sequences are automatically excluded from the analysis. For this reason, the subjects also are excluded from the summary statistics in many study reports. With the advance of general linear model research in the past decade, is this still a good practice?

## Statistical Methods for Optimization of Development of Combination Products

❖ Timothy Montague, GlaxoSmithKline, 1250 S. Collegeville Road, MS UP4315, Collegeville, PA 19426-2990, *timothy.h.montague@gsk.com*

*Key Words:* combination drug products

Polyfactorial approach to disease therapy is becoming more common within the practice of medicine. One example is in the treatment of HIV, where combination therapy is the standard of care. As such, there are significant opportunities in the pharmaceutical industry to develop combination therapies as fixed dose or variable dose combinations. Traditionally, development of combination therapies usually occurred following registration of the monotherapy. However, many new chemical entities (NCEs) are being targeted as combination therapies earlier in the development cycle. As such, a new paradigm for development of combination therapies is needed. The focus of this roundtable discussion will be the statistical methods that can help optimize the clinical development of combination products, focusing on clinical study design.

## How To Explain Complex Statistical Methods to Life Scientists

❖ Vadim Kutsyy, Cytokinetics, Inc., 280 E. Grand Ave., South San Francisco, CA 94022, *vadim@kutsyy.com*

*Key Words:* life science, biostatistics, model explanation

Being a statistician in a life science company requires continuing collaboration with life scientists. Life scientists have underlying hypothesis and use data to test these hypotheses. As a statistician, I am trying to analyze data in such way that a nonstatistician would understand ideas of the data analysis. Often, more complex statistical methods are needed to build high-quality models. How to explain these complex models to life scientists so they have enough understanding to trust the analysis?

## R-Based Data Analysis and Graphics Applications for Scientists: Moving beyond Excel

❖ Bert Gunter, Genentech, Inc., MS 333B, One DNA Way, South San Francisco, CA 94080-4990, *bgunter@gene.com*

*Key Words:* R, statistical software, data visualization, integrated applications, R GUIs, excel

Despite major advances in statistics and statistical graphics, most practicing scientists and engineers continue to use primitive and often unreliable software for data analysis and visualization. MS ExcelÆ is probably the most widely known and used example. We believe R can change this situation. It is free, trivial to install, and state-of-the-art. Unfortunately, it is also too difficult for most practitioners to learn. However, in research and industry, there is a great need for repetitive analyses performed as a routine part of normal work processes. For such circumscribed tasks, it is relatively easy to build GUIs in R or other software (including Excel) that access R's superior capabilities and incorporate them into structured analyses. We shall discuss our strategies and experience and those of others, focusing primarily on WindowsÆ OSs.

### Translational Medicine: from R to D and from D to R

❖ Chi-Hse Teng, Pfizer Inc., Science Center Drive, San Diego, CA 92121, *chi-hse.teng@pfizer.com*; Patricia English, Pfizer Inc.

***Key Words:*** translational medicine, translational research, genomics, biomarkers

The pharmaceutical industry is facing enormous pressure to increase productivity. One approach taken commonly by pharmaceutical companies is to implement translational medicine to bridge the gap between discovery and clinical development. Translational medicine has been characterized as a "two-way road: bench to bedside and bedside to bench," (Marincola 2003) "from accurate diagnosis to appropriate treatment" (Webb 2004). This luncheon is intended to provide a place for statisticians interested and/or working in this area to discuss opportunities, share their experiences, and map out necessary steps for statisticians to work in this area.

### Proof of Concept Studies

❖ Robert Smith, Bristol-Myers Squibb Company, 11 Bayberry Road, Princeton, NJ 08540, *Robert.A.Smith@bms.com*

***Key Words:*** proof of concept, pharmaceutical development

Many drug development programs currently include a "Proof of Concept Study" (i.e., a Phase IIa study on which pharmaceutical management bases a decision whether to continue clinical development). Following up on last November's FDA/PhRMA workshop on "good proof of concept studies," participants in this luncheon roundtable will discuss the exploratory versus confirmatory nature of PoC studies, hypothesis testing versus estimation or hypothesis generation, research hypotheses as Go/No Go criteria for continuing clinical development, sizing (powering) PoC studies, use of Bayesian methods, and use of "enriched" populations.

### Sensitivity Analysis for Missing Data in Clinical Trials

❖ Jie Zhang, Novartis Pharmaceuticals Corporation, One Health Plaza, Building 405, Room 2117, East Hanover, NJ 07936, *jie.zhang@novartis.com*

***Key Words:*** non-ignorable missingness, confirmatory studies, bias

When undertaking confirmatory, possibly model-based analyses of data from clinical trials, statisticians are confronted frequently with having to assess and address potential biases introduced by missing data. New developments for handling missing data have proliferated in the literature. Sensitivity analysis, which allows the assessment of the impact of a range of nonignorable missing patterns on the robustness of the statistical results, provides a unique and powerful alternative for analyzing missing data. An example utilizing sensitivity analysis in a large fracture trial will be described briefly. Practical issues associated with the application of sensitivity analysis will be discussed thereafter.

### Analysis of Cell-Based, High-Content Imaging Data

❖ Shuguang Huang, Eli Lilly and Company, 7220 Madison Village Court, Indianapolis, IN 46227, *huang_shuguang@lilly.com*

***Key Words:*** imaging, high-content screening, cellomics, statistics, cell cycle, molecular

Fluorescence-based microscopy technology has made it possible for cell-based, high-content imaging (HCI), which allows measurements of the intracellular fluorescence intensity of multiple targets on a cell-by-cell basis. In the context of drug discovery, HCI technologies are useful in profiling drug action in cellular milieu. Researches often are interested in whether a drug exerts a statistically significant effect, which is reflected by the perturbation of the distribution of certain phenotype. With the daunting amount of data, quantifying the treatment effect is a nontrivial challenge. This discussion will focus on background introduction of cell-based, high-content imaging and data analysis issues and statistical methodologies.

### Interval Censored Time-to-Event Data: Examples, Analyses, and Assumptions

❖ Daohai Yu, Duke University, 406 Somersview Drive, Chapel Hill, NC 27514, *dhyu8@yahoo.com*

***Key Words:*** interval censoring, conditional model, dependent censoring, discretionary visits, serial screening, survival analysis

The random censorship assumption is violated often in interval censored (IC) time-to-event data, such as when the data arise from a serial screening and the timing of screening potentially could depend on the patient's health status. Some examples of dependent IC event data will be discussed, along with data analysis methods currently available for IC time-to-event data. Underlying assumptions for each method will be examined and compared. In particular, a new likelihood-based approach coupled with a class of innovative conditional models for dependent, interval-censored, time-to-event data with a marker for discretionary visits will be discussed and compared with the current analysis methods assuming independent censoring.

## 431 Business and Economics Statistics Section Roundtable with Lunch (fee event)

Business and Economics Statistics Section

**Wednesday, August 9, 12:30 pm–1:50 pm**

### Economic Aspects of Alternative Energy

❖ Margaret F. Land, TeXas Environmental Studies and Analysis, LLC, 1713 Santa Monica Blvd., Kingsville, TX 78363, *margaret.land@tx-esa.com*

***Key Words:*** alternative, energy, economic

Discussion will center on economic and other comparisons of alternative, or "green," energy with traditional sources of energy.

# 432 Section on Statistical Computing Roundtables with Lunch (fee event)

Section on Statistical Computing
**Wednesday, August 9, 12:30 pm–1:50 pm**

### S-PLUS and R: Working Together?

❖ Tim C. Hesterberg, Insightful Corporation, 1700 Westlake Ave., Suite 500, Seattle, WA 98109-3044, *timh@insightful.com*

S-PLUS and R are two dialects of the S language. This roundtable is for those interested in discussing how the two languages and their corresponding communities can work together.

# 433 Section on Statistical Consulting Roundtable with Lunch (fee event)

Section on Statistical Consulting
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Statistical Consulting in (and around) the Bayesian Paradigm

❖ Alix Gitelman, Oregon State University, Statistics Department, 44 Kidder Hall, Corvallis, OR 97331, *gitelman@science.oregonstate.edu*

***Key Words:*** Bayesian statistics, statistical consulting

As a Bayesian statistician providing statistical consulting at a Land Grant University, I often avoid giving advice that is Bayesian in nature, not so much by choice as by necessity. Many consultations are one-time meetings in which people seek help with a sampling design, simple analysis, or response to a referee report. Unless consultants are willing to take on the burden of explaining the different paradigm and (often) the data analysis projects, it seems relatively difficult to convince people that the Bayesian approach is appealing. By contrast, some consulting projects turn into longer-term collaborative relationships, and these can provide excellent opportunities to introduce the Bayesian paradigm. This roundtable will provide an opportunity for Bayesian statisticians to compare notes on bringing Bayesian methods into the mainstream of statistical consulting.

# 434 Section on Statisticians in Defense and National Security Roundtable with Lunch (fee event)

Section on Statisticians in Defense and National Security, Section on Statistical Consulting
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Defense-Related Consulting Projects

❖ John Crown, RAND Corporation, 201 N. Craig Street, Suite 202, Pittsburgh, PA 15213-1516, *john_crown@rand.org*

***Key Words:*** defense, consulting, personnel policy, logistics, acquisition

This roundtable discussion will focus on topics for defense consulting projects, including personnel management policy, weapon system acquisition, logistics of the global war on terror, and the new frontiers of space. The defense industry remains a large consumer of research, and there are many successful contractors conducting research in widely varying fields supporting defense. Our goal will be to discuss the many analytical topics, what agencies have supported consultants, how to approach these agencies, and how to gain support for research efforts.

# 435 Section on Statistical Education Roundtables with Lunch (fee event)

Section on Statistical Education
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Analogies and Other Strategies To Help Students Use Intuition To Understand Concepts

❖ Lawrence M. Lesser, The University of Texas at El Paso, Department of Mathematical Sciences, 500 W. University Avenue, El Paso, TX 79968-0514, *Lesser@utep.edu*

***Key Words:*** analogy, representation, intuition, counterintuitive, teaching, concepts

Many strategies can be useful in helping students develop intuition for basic statistical concepts. Such strategies could include visual representations, analogies, or metaphors. These strategies can even help demystify important scenarios (e.g., Simpson's Paradox), perceived as counterintuitive initially. Teachers and texts vary greatly in when, how, and how much they choose and use these strategies. Statisticians, teachers, and education researchers are invited to join the dialogue about implementation and assessment issues and to identify examples and resources in this area. (To whet your appetite, you might enjoy "Engaging the Intuition in Statistics to Motivate" on the College Board AP Statistics Course homepage.)

### How I Use the ARTIST Web Site in My Teaching

❖ John Holcomb, Jr., Cleveland State University, 2121 Euclid Ave., Rhodes Tower 1515, Cleveland, OH 44115-2214, *j.p.holcomb@csuohio.edu*

***Key Words:*** assessment, evaluation, statistical literacy, statistical projects

During this roundtable, we will discuss implementation issues of the resources available at the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) web site. Discussion topics include the Assessment Builder, alternate assessment ideas, the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS), and other ideas involving assessment and the first statistics course.

## 436 Section on Statistics in Epidemiology Roundtable with Lunch (fee event)

Section on Statistics in Epidemiology
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Relative Risk Regression as an Alternative to Logistic Regression in Cross-Sectional and Longitudinal Studies

❖ Richard Kronmal, University of Washington, Building 29, Suite 310, 6200 NE 74th Street, Seattle, WA 98115, *kronmal@u.washington.edu*

***Key Words:*** relative risk regression, logistic regression, risk ratio, prevalence ratio

The most commonly used method for regression modeling of a binary variable in cross-sectional and longitudinal studies is logistic regression. Yet, it is the relative risk (prevalence ratio for cross-sectional and risk ratio for longitudinal studies) that is the quantity of interest in such analyses, not the odds ratio. We will discuss the methodological issues involved in relative risk regression.

## 437 Section on Government Statistics Roundtable with Lunch (fee event)

Section on Government Statistics
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Journal of Empirical Research on Human Research Ethics

❖ Joan E. Sieber, California State University, East Bay, 2060 Quail Canyon Court, Hayward, CA 94542, *joan.sieber@sbcglobal.net*

***Key Words:*** ethics, privacy, confidentiality, data sharing, publication, IRBS

Given issues of privacy, confidentiality, data sharing, and methodology that accompany human research, there are many ethics-related problems for statisticians to solve. Solutions need to be shared among the stakeholders concerned with human subjects' protection. A new peer-reviewed, international, nonprofit journal---the Journal of Empirical Research on Human Research Ethics (JERHRE)---began publication in March 2006. Without evidence-based problem solving, many conflicts are settled unsatisfactorily by applying a one-size-fits-all interpretation of principles or regulations. JERHRE seeks to create collaboration among stakeholders, stimulate research, and disseminate knowledge to foster intelligent application of ethical principles in research contexts worldwide. Come to this luncheon to learn about the journal and discuss these issues.

## 438 Section on Statistical Graphics Roundtable with Lunch (fee event)

Section on Statistical Graphics
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Biostatistical Graphics: Large, Weak Datasets

❖ Thomas Lumley, University of Washington, Seattle, WA 98195, *tlumley@u.washington.edu*

Classical statistical graphics dealt with strong relationships in small datasets. More recent techniques allow these relationships to be extended to datasets with many observations or variables. Seeing weak relationships in large datasets is still a challenging problem, whether the data are 100k SNPs on a few hundred people or a few dozen pulse waveform summaries on several thousand people. Is graphical exploration still useful, and how can we filter the signal from the noise? Can we get useful methods adopted in a field that still uses red-green contrasts to display basic data?

## 439 Section on Quality and Productivity Roundtable with Lunch (fee event)

Section on Quality and Productivity
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Communicating Statistics to Nonstatisticians in Industry

❖ Philip Scinto, The Lubrizol Corporation, 29400 Lakeland Blvd., Drop 152A, Wickliffe, OH 44092-2298, *prs@lubrizol.com*

***Key Words:*** communication, industry, customer, client, collaborator, colleague

As statistical consultants in industry, most of us find our work in solving problems and making decisions based on statistical analysis, logic, and thinking to be both challenging and rewarding. However, we must not overlook the importance of the equally challenging and rewarding necessity of communicating our analyses, logic, and decisions to our chemist, physicist, and engineer customers, collaborators, and colleagues. Even a high-level understanding of simple statistical concepts---such as estimation, bias, variability, power, and probability---by our customers would increase productivity by increasing the chances that analyses and decisions are applied correctly with or without the presence of a statistician. In this roundtable discussion, we will discuss and share our experiences and techniques for communicating statistics to nonstatisticians.

## 440 Section on Risk Analysis Roundtable with Lunch (fee event)

Section on Risk Analysis
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Results of a Post-Katrina Survey of New Orleans Residents

❖ David Banks, Duke University, ISDS Box 90251, Duke University, Durham, NC 27708, *banks@stat.duke.edu*

***Key Words:*** Katrina, survey, network model

Shortly after Katrina struck and public criticism of emergency management response began, the NSF funded a number of projects to study what happened in New Orleans. This roundtable will describe results from a household survey of New Orleans residents. Key questions include the determination of factors that affected their decision of whether and when to evacuate, what kind of post-Katrina experience they had, which agencies or groups provided useful help at different stages, and how those agencies and groups interacted.

## 441 Section on Survey Research Methods Roundtables with Lunch (fee event)

Section on Survey Research Methods
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Survey Research and Its Role in Improving the Health and Lives of the World's Poor

❖ David J. Fitch, Universidad del Valle de Guatemala, Apartado Postal 82, 01901, Guatemala, 01901 Guatemala, *dfitch@uvg.edu.gt*

***Key Words:*** USAID, WHO, World Bank, hand held computers, village sampling

There is too little statistical competence in USAID, WHO, and the World Bank. There are few members of the ASA within these organization---possibly none. We should inform ourselves and work for change so statistics might better contribute to improving the health and lives of the world's poor. U.S. legislation gives the chief statistician, EOP, responsibilities for the quality of survey work carried out with U.S. funds, but this does not include surveys funded by USAID. The poor, it seems, are not worthy of good statistical work. We have seen problems with USAID's DHS work, and they---not knowing evaluation---support inconclusive and illogical evaluation. Hand-held computers could increase efficiency greatly in village sampling, but they can't see it, so it hasn't been done. I'm hoping we can have time together to discuss developing country work.

### Cross-Cultural Issues in Survey Research

❖ Peter P. Mohler, ZUMA, P.O. Box 122155, Mannheim, 68072 Germany, *director@zuma-mannheim.de*

***Key Words:*** comparative, cross-cultural, multi-lingual, process, quality, surveys

Recognition of cultural diversity, acceptance of multilinguality within nation states, and ever-raising interest in comparative research led to a dramatic increase in cross-cultural survey research in the last decade. Also, one can observe a move from two-culture comparisons to simultaneous multiple culture/nation surveys. In turn, these surveys ask for new methods, processes, and analytic tools to handle the complex interplay of a multitude of factors, such as comparable sampling designs, equivalence of measurement, and local vs. global factors.

## 442 Social Statistics Section Roundtable with Lunch (fee event)

Social Statistics Section
**Wednesday, August 9, 12:30 pm–1:50 pm**

### The Next Survey Imperative: Being Proactive on Privacy

❖ Gerald Gates, U.S. Census Bureau, 4700 Silver Hill Road, Suitland, MD 20233, *gerald.w.gates@census.gov*

***Key Words:*** privacy, data stewardship, survey response

This discussion will highlight the range of approaches used by survey organizations to address growing public concern for privacy while facing increased pressure to halt declining survey response rates. The moderator is the chief privacy officer at the U.S. Census Bureau. He will share his experience in developing a privacy program at the bureau that integrates with the bureau's Data Stewardship Program. He will discuss privacy strategies that include benchmarking to other organizations, gaining senior staff commitment, creating a set of privacy principles and associated policies, conducting privacy impact assessments, training staff on privacy, communicating externally, and researching the public's attitude about privacy and how that attitude may influence response. Participants will share their views about how privacy impacts their own research.

## 443 Section on Teaching Statistics in the Health Sciences Roundtable with Lunch (fee event)

Section on Teaching Statistics in the Health Sciences
**Wednesday, August 9, 12:30 pm–1:50 pm**

### Distance Teaching and Learning in the Health Sciences

❖ T. Robert Harris, The University of Texas at Dallas, School of Public Health, Dallas Regional Campus, 6011 Harry Hines Blvd., v8.112, Dallas, TX 75390-9128, *TRobert.Harris@UTSouthwestern.edu*

***Key Words:*** distance learning, regression, interactive television, internet course, public health, graduate

This roundtable will offer a chance to share experiences, compare strategies and tactics, and talk about limitations of distance learning formats and possible opportunities. The convener has become a student and practitioner of distance learning using a hybrid of interactive television and internet media to teach multiple linear and logistic regression to graduate students in public health.

## 444 JASA Applications and Case Studies Invited Session

JASA, Applications and Case Studies, Section on Nonparametric Statistics
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Model-Assisted Estimation of Forest Resources with Generalized Additive Models

❖ Jean D. Opsomer, Iowa State University, Snedecor Hall, Ames, IA 50011, *jopsomer@iastate.edu*; F. Jay Breidt, Colorado State University; Gretchen Moisen, U.S. Forest Service; Goeran Kauermann, Universitaet Bielefeld

***Key Words:*** multi-phase survey estimation, nonparametric regression, calibration, systematic sampling, variance estimation

Multi-phase surveys often are conducted in forest inventories with the goal of estimating forest characteristics over large regions. We describe how design-based estimation of such quantities, based on information gathered during ground visits, can be made more precise by incorporating auxiliary information from remote sensing. The relationship between ground measurements and the remote sensing variables is modeled using generalized additive models. Model-assisted estimators utilizing these nonparametric fits are proposed and applied to forestry survey data from northern Utah. The design context of this study is two-phase systematic sampling from a spatial continuum, and we describe difficulties with the standard variance estimation approach. An alternative assessment of estimator performance based on a synthetic population is discussed.

## 445 New Methods for Modeling Choice in Marketing

Section on Statistics and Marketing

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Structural Estimation of Retail Demand and Inventory Decisions

❖ Andres Musalem, The Wharton School of the University of Pennsylvania, 3730 Walnut Street, 700 JMHH, Philadelphia, PA 19104, *amusalem@wharton.upenn.edu*

***Key Words:*** structural, demand, inventory, missing data, MCMC, data augmentation

In marketing, much recent research has been concerned with the resultant biased estimates from demand models when ancillary information is ignored (e.g., endogenously chosen prices on the part of the retailer). In this research, we focus on a problem that, when ignored, also may lead to wrong inferences: the occurrence of unobserved out of stocks. In fact, every time there is a stockout, the set of available alternatives to a consumer changes. Consequently, the assumption that each consumer may consider all the alternatives in her decision process is no longer valid. The estimation approach is based on simulating the unobserved evolution of the inventory on a retailer's shelf via MCMC and data augmentation methods. The proposed methodology is tested using simulated data and will be illustrated using a dataset from a retailer in Spain.

### Estimating Willingness To Pay with Random Coefficient Models

❖ Garrett Sonnier, University of California, Los Angeles, 110 Westwood Plaza, Los Angeles, CA 90095, *garrett.sonnier@anderson.ucla.edu*

***Key Words:*** choice, logit, heterogeneity

It is well-known that the choice probabilities derived from random utility models suffer from scale invariance. Only ratios of parameters can be identified, and the usual approach to this identification problem is to place some type of restriction on the parameter space. Generally, the specific restriction employed has been regarded as an arbitrary choice. In models with homogeneous coefficients or latent-class approaches to heterogeneity, the model fits are unaffected and point estimates of parameters obtained under one type of restriction can be recovered easily by transforming those from another restriction. This typically leads researchers to choose the parameterization that most easily facilitates estimation. We will demonstrate that the choice of parameterization in a random coefficient setting incorporating parametric distributions of heterogeneity matters.

## 446 Nonparametric Inference

IMS, Section on Nonparametric Statistics

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Statistical Inference for Evolving Periodic Functions

❖ Peter G. Hall, Australian National University, Center for Mathematics and its Applications, Canberra, 0200 Australia, *peter.hall@anu.edu.au*

***Key Words:*** astronomy, nonparametric regression

In the study of variable stars, where the light reaching an observer fluctuates over time, it can be the nature of the variation, unless it follows a regular pattern. In this respect, so-called periodic variable stars are particularly amenable to analysis. There, radiation varies in a perfectly periodic fashion, and period length is a major focus of interest. In this talk, we consider methodology for conducting inference about features that might account for departures from strict periodicity. In particular, we discuss the potential variation, over time, of the period or amplitude of radiation. We suggest techniques for estimating the parameters of this evolution and for testing the hypothesis that the evolution is present. This problem has a number of unusual features, including subtle issues of identifiability and strikingly fast convergence rates. The latter reflects the considerable impact that even small departures from exact periodicity can have on the visual appearance of the data.

### On Nonparametric Confidence Sets

❖ Tony Cai, University of Pennsylvania, 3730 Walnut Street, Department of Statistics/Wharton School, Philadelphia, PA 19104, *tcai@wharton.upenn.edu*; Mark Low, University of Pennsylvania

***Key Words:*** adaptive estimation, block thresholding, nonparametric confidence sets, nonparametric testing, quadratic functional, wavelets

In this talk, I will present some recent results on minimax and adaptive confidence balls and confidence bands in nonparametric regression and white noise model. Connections to nonparametric estimation of a quadratic functional and nonparametric hypothesis testing will also be discussed.

### Nonparametric Estimation of Eigenvectors

❖ Iain Johnstone, Stanford University, Department of Statistics, Sequoia Hall, Stanford, CA 94305, *imj@stanford.edu*

***Key Words:*** sparse representation, covariance matrix, nonparametric estimation, principal components

The talk will review some recent results on nonparametric estimation of the principal eigenvector of a large covariance matrix when it is assumed that the population covariance matrix differs from i.i.d. by a

fixed number of perturbations of sparse representation. Some connections with sparse estimation of mean vectors may be mentioned if time permits. This is joint work with Debashis Paul.

### Adaptive Generalized Likelihood Inferences for Additive Models

❖ Jianqing Fan, Princeton University, Department of ORFE, Princeton University, Princeton, NJ 08540, *jqfan@princeton.edu*; Jiancheng Jiang, Princeton University

***Key Words:*** Geleranized likelihood ratio, additive model, Wilks phenomenon, adaptive power property, bias correction, null distribution

The maximum likelihood test is fundamental and indispensable to classical parametric inferences. Despite its success in parametric inferences, the maximum likelihood ratio statistics might not exist in nonparametric function estimation. The generalized likelihood statistics were introduced in Fan, Zhang, and Zhang (2001) to overcome these drawbacks. In this talk, we further develop the technique for statistical inferences in additive models, a widely used model in learning theory. We will show the generalized likelihood ratio statistics follow asymptotically rescaled chi-square distributions and are distribution-free. We further prove the generalized likelihood ratio tests have an important adaptive power property and are asymptotically optimal in terms of rates of convergence for nonparametric hypothesis testing. Some bias correction properties also will be introduced.

## 447 Split-Plot Designs and Response Surface Analysis: the Interface ●

Section on Quality and Productivity, Section on Physical and Engineering Sciences
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Standard Error Calculations for Estimators of Regression Coefficients in Split Plot Designs

Shaun Wulff, University of Wyoming; ❖ Timothy Robinson, University of Wyoming, WY 82071, *tjrobin@uwyo.edu*; Christine M. Anderson-Cook, Los Alamos National Laboratory

***Key Words:*** split-plot, response surface, restricted randomization

Split plot designs have become recognized as a useful alternative to completely randomized designs in industrial experiments when there are practical difficulties in changing the levels of one or more factors due to cost or effort. These restrictions on the randomization lead to potential correlation structure that can affect inference procedures. In this talk, we examine estimation of the variance-covariance matrix, or standard errors, of the regression estimators. In many cases, the usual estimator is an underestimate of the true variance-covariance matrix of the regression parameter estimator since it does not account for the additional uncertainty attributed to the estimation of the correlation structure. We will discuss such cases as well as present commonly recommended adjustments to deal with this problem.

### A Bayesian Approach to the Analysis of Split-Plot Experiments

❖ Peter Goos, Universiteit Antwerpen, Faculty of Applied Economics, Prinsstraat 13, Antwerpen, 2000 Belgium, *peter.goos@ua.ac.be*; Steven G. Gilmour, Queen Mary, University of London

***Key Words:*** generalized least squares, variance component estimation, degrees of freedom

Industrial experiments often involve hard-to-change variables not reset for every run of the experiment. The resulting experimental designs are of the split-plot type. A proper classical statistical analysis requires the use of generalized least squares procedures and, hence, the estimation of the variance components in the statistical model. In most split-plot designs, the hard-to-change variables are reset a small number of times, such that estimation of the variance components corresponding to the whole plots stratum of the experiment is either impossible or inefficient. As a result, these variance components often are estimated to be zero and the generalized least squares inferences collapse to ordinary least squares ones. This problem can be avoided by incorporating prior beliefs regarding the magnitudes of the variance components in a Bayesian data analysis.

### Randomization-Based Analysis of Multistratum Response Surface Designs

❖ Steven G. Gilmour, Queen Mary, University of London, School of Mathematical Sciences, Mile End Road, London, E11 1SS UK, *s.g.gilmour@qmul.ac.uk*

***Key Words:*** split-plot design, response surface methodology, REML, randomization

Split plot designs and, more generally, multistratum designs arise due to restrictions in the randomization, so that different factors are randomized to different-sized experimental units. Thus, it is natural to explore how far randomization-based analysis, based on minimal model assumptions, can lead. For orthogonal multistratum designs, randomization-based methods give an essentially complete analysis, such as the usual split plot analysis. In response surface experiments, however, the designs are usually nonorthogonal. In this paper, it will be shown that randomization leads naturally to separate analyses in each stratum, but that there is no obvious way to combine them. Combination of information usually is done using REML and GLS. A modified method, which uses the estimates of pure error variance in each stratum, and GLS will be recommended instead as being more robust.

### Testing for Lack-of-Fit in Split Plot Response Surface Models

❖ Scott Kowalski, Minitab Inc., 1611 Silk Tree Circle, Sanford, FL 32773, *skowalski@minitab.com*; Geoff Vining, Virginia Polytechnic Institute and State University

***Key Words:*** split-plot, response surface, DOE

Split plot experiments are becoming increasingly more common in industry. Current research has focused on response surface methodology (RSM) for split plot scenarios. The application of RSM involves a sequential learning strategy. The traditional steps are screening experiments, followed by steepest ascent. After moving through the experimental region, one conducts follow-up experiments. It is important to check that a linear model still holds in these new regions. A common approach is to test for lack-of-fit. This talk will focus on constructing a test for lack-of-fit within a split plot framework.

## 448 Adaptive Designs in Clinical Trials ●

Biopharmaceutical Section, ENAR, WNAR
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Weighted and Unweighted Z-tests in Sample Size Re-estimation

❖ Kuang-Kuo G. Lan, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., Raritan, NJ 08869, *glan@prdus.jnj.com*

***Key Words:*** alpha level, sample size re-estimation

The weighted Z-test is a powerful tool in a two-stage design. The weights for information collected in the two stages of the studies are prespecified, and the alpha level can be well-controlled even after sample size re-estimation. Because the weighted Z-test may not assign weights according to the information fractions at each stage, some researchers have expressed concern about the use of this test in clinical trial data analysis. The unweighted Z-test, on the other hand, may inflate the alpha level after sample size re-estimation. This problem can be solved by incorporating futility stopping during interim analysis, which results in alpha reduction. We will provide examples to illustrate the pros and cons of both approaches.

### Considerations in Use of Adaptive Designs in Drug Development

❖ Hsien-Ming Hung, U.S. Food and Drug Administration, 10993 New Hampshire Ave., Bldg. 22, Room 4238, Mail stop 4105, HFD-710, Silver Spring, MD 20993-0002, *hung@cder.fda.gov*

***Key Words:*** adaptive design, flexible design, estimation, practicality, logistics, statistical testing

In the last decade, a great deal of research has been devoted to adaptive or flexible designs, particularly multistage designs, in the statistical literature. The concept behind multistage designs seems natural in dealing with the complex problems in drug development. Statistical methodology is well-developed for testing the effect of a test drug. In this talk, we will give an overview on the essence of statistical testing under such designs. Then, we will share our insights on the estimation problem and point out other problems, such as logistics and practicality, when such designs are used.

### Adaptive Designs in Clinical Trials

❖ Yu Shen, M. D. Anderson Cancer Center, 1515 Holcombe Bvld., Unit 447, Houston, TX 77030, *yshen@mdanderson.org*

Clinical investigations best serve the interests of public health through the use of efficient trial designs comprised of interim monitoring to preserve ethical, scientific, and economic concerns. In this talk, we will describe recent statistical methodology in clinical trial designs that are adaptive, allowing flexibility for sample size determination while maintaining scientific validity. The methods include the design and inference of continuous outcomes, censored survival data with staggered entry, and correlated outcomes with covariates. Important theoretical properties are investigated. Moreover, adaptive designs within the framework of a Bayesian decision theoretical approach will be discussed. The proposed designs and corresponding estimation methods are applicable to various areas of health research, especially cancer clinical trials.

## 449 A Tribute to Yehuda Vardi ● ✪

Memorial, IMS, Section on Physical and Engineering Sciences

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Statistical Inverse Problems in Active Network Tomography

❖ Vijay Nair, University of Michigan, Department of Statistics, 1085 South University 439 West Hall, Ann Arborm, MI 48109-1107, *vnn@umich.edu*

Network tomography, first introduced in Vardi (1996), characterizes two classes of large-scale inverse problems that arise in the modeling and analysis of computer and communications networks. This talk will deal with active network tomography where the goal is to recover link-level quality of service parameters, such as packet loss rates and delay distributions, from end-to-end, path-level measurements. Internet service providers use this to characterize network performance and monitor service quality. We will provide a review of recent developments, including the design of probing experiments, inference for loss rates and delay distributions, and applications to network monitoring. This is joint work with George Michailidis, Earl Lawrence, Bowei Xi, and Xiaodong Yang.

### Fast Functional MRI

❖ Cun-Hui Zhang, Rutgers University, Department of Statistics, Piscataway, NJ 08855, *czhang@stat.rutgers.edu*

We develop fast functional MRI methods to improve the time-resolution of the current functional MRI technology by sampling a small fraction of the Fourier transform of the spin density and using a prolate wave filter to approximately obtain not the usual susceptibility map, but the integral of this quantity over regions of interest in the brain at successive time-points. The aim of this space/time trade-off is to obtain at high time-resolution the total activity in these regions, which processes the specific stimulus/task and, more important in studying higher cognition, the sequence of occurrences of these processes. An fMRI experiment will be reviewed and discussed. This is joint work with Gary Glover, Martin Lindquist, and Larry Shepp.

### Duration Data: Poisson Process and Bias Correction

❖ Zhiliang Ying, Columbia University, Department of Statistics, 1255 Amsterdam Avenue, 10th Floor, New York, NY 10027-5927, *zying@stat.columbia.edu*

Yehuda Vardi made fundamental contributions to statistical analysis of duration data with selection bias. He also pioneered Poisson models for emission tomography and network tomography. In this talk, I will present Poisson-type models for duration data in the context of medical applications. I will show how useful estimators can be constructed that correct for selection bias.

## 450 New Statistical Methodology for Genomic Applications with Focus on Array CGH and Gene Networks ●

ENAR, Biometrics Section, WNAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### A Bayesian Hierarchical Model for Integrating Biological Data

❖ Shane Jensen, The Wharton School of the University of

Pennsylvania, Department of Statistics, 463 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, *stjensen@wharton.upenn.edu*

***Key Words:*** hierarchical models, Bayesian framework, gene expression, chip binding, clustering, promotor elements

A substantial focus of research in molecular biology is the network of factors which control the involvement of different genes in living cells. Previous statistical approaches for identifying gene regulatory networks have used gene expression data, ChIP binding data or promoter sequence data, but each of these resources provides only partial information. We present a Bayesian hierarchical model that integrates all three data types in a principled fashion. The gene expression data is modeled as a function of the unknown gene regulatory network which has an informed prior distribution based upon both ChIP binding and promoter sequence data. In this context, we discuss procedures for balancing multiple sources of prior information. Applications to both yeast and human data are presented and validated using several external sources of information.

### A Statistics Method for Array CGH Analysis

❖ Pei Wang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M2-B500, Seattle, WA 98109, *pwang@fhcrc.org*

***Key Words:*** array CGH, FDR

Array CGH is a powerful technique for genomic studies of cancer. It enables one to carry out genome-wide screening for regions of genetic alterations, such as chromosome gains and losses, or localized amplifications and deletions. In this paper, we propose a new algorithm---cluster along chromosomes (CLAC)---for the analysis of array CGH data. CLAC builds hierarchical clustering-style trees along each chromosome arm (or chromosome), then selects the 'interesting' clusters by controlling the False Discovery Rate (FDR) at a certain level. In addition, it provides a consensus summary across a set of arrays and an estimate of the corresponding FDR. We illustrate the method using an application of CLAC on a lung cancer microarray CGH dataset and a BAC array CGH dataset of aneuploid cell strains.

### On Detecting Chromosomal Aberrations Using Copy Number Data

Xuesong Yu, University of Washington; Tim Randolph, University of Washington; Hua Tang, Fred Hutchinson Cancer Research Center; ❖ Li Hsu, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M2-B500, Seattle, WA 98109, *lih@fhcrc.org*

***Key Words:*** copy number data, wavelet, statistical inference

Segmenting chromosomes into gains, losses, and no change based on copy number data from microarrays is of interest to biologist for identifying chromosomal aberrations. Due to technical and sample heterogeneity, the variance of an array varies greatly from one to another. It is therefore important to be able to provide biologists a measure of confidence level for the segmentation. We propose an estimator by using wavelets to quantify the scale of a clone being a break point for copy number gain/loss. Based on this estimator, we design a test statistic to assess the statistical significance of these breaks points. Simulation and a real data example will be used to illustrate the method.

## 451 The Role of Administrative Records in 21st-Century Surveys and Censuses ● ✪

Social Statistics Section, Section on Health Policy Statistics
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Methods and File Acquisitions Supporting the Expanded Use of Administrative Records

❖ Dean Resnick, U.S. Census Bureau, 4700 Silver Hill Road, Mailstop 9200, Washington, DC 20233-9200, *dean.michael.resnick@census.gov*

***Key Words:*** administrative records, file acquisition, record linkage, models

Over the past couple of years, the U.S. Census Bureau has been increasing administrative records use. This expansion is due largely to recent file acquisitions and several important methodological breakthroughs. This paper will discuss both the third-party files that substantially improved administrative records content and changes in approaches to processing and modeling that have likewise facilitated expansion. Important acquisitions include the Social Security Administration's Supplemental Security Income File and the United States Postal Service's National Change of Address File. Improvements in processing and modeling methods include reducing dependency on IRS information, converting a C-based record linkage software application to SAS, improving geographic capability, and developing models for use in federal poverty program analysis.

### Administrative Records and Survey Data Reuse: a Muse on Their Future

❖ Ronald Prevost, U.S. Census Bureau, 4700 Silver Hill Road, Suitland, MD 20746, *ronald.c.prevost@census.gov*

***Key Words:*** administrative records, survey data, questionnaire design, address frames, modeling

Administrative records and survey data present multiple, sometimes contrary, representations of society. The comparison of multiple data sources often can improve our understanding and measurement of social characteristics. The integration and reuse of data provides us the capacity to produce new information in response to needs while minimizing respondent burden and cost to taxpayers. The integration of administrative and survey data has been integral to important Census Bureau programs for many years. This paper presents a theoretical perspective on this integration and reuse of data and includes a discussion of the benefits of these applications, such as enhancements to current survey questionnaire design and wording, address and survey frames, cost modeling and statistical toolkits for field operations, survey controls, and the rapid delivery of statistical information.

### Why Are Survey Counts of Medicaid Enrollees Lower Than Administrative Enrollment Counts?

❖ Michael Davern, University of Minnesota, 2221 University Ave., SE, Suite 345, State Health Access Data Assistance Center, Minneapolis, MN 55414, *daver004@umn.edu*; David Baugh, Centers for Medicare and Medicaid Services; Christine Cox, National Center for Health Statistics; Kim Lochner, National Center for Health Statistics

***Key Words:*** Medicaid, administrative data, current population survey, National Health Interview Survey, 2001 supplemental survey

Survey counts of those enrolled in Medicaid are significantly lower than administrative data counts. Because surveys are the only source of information on key policy populations, such as the uninsured and those eligible for Medicaid but not enrolled, this disparity can have significant implications, depending on why the two counts are systematically different. We are examining matched survey data and administrative data from the Current Population Survey, the National Health Interview Survey, and the Medicaid Statistical Information System to better understand how people who are enrolled in Medicaid answer the two surveys. This will allow us to understand the roll that survey measurement error plays in explaining the difference. We also examine the impact of survey item non-response (imputation), survey editing, unit non-response and sample coverage have on explaining the difference.

### Combining Social Program Administrative Data with Census Bureau Survey Data

❖ Robert Goerge, The University of Chicago, Chapin Hall Center for Children, 1313 E. 60th St, Chicago, IL 60637, *bobg@uchicago.edu*

***Key Words:*** administrative data, income assistance, TANF, childcare subsidy program, eligibility, employment outcomes

Administrative and survey data provide two sets of facts. A number of research efforts have begun to address this problem by linking survey and administrative datasets. This paper reports on a joint effort by the U.S. Census Bureau, university research centers, and state government to combine census survey data (2001 Supplemental Survey) with wage report administrative data, income assistance program data (TANF), and child care subsidy program data at the household and individual person levels. The primary goal of the research is to understand how participation in the child care subsidy program, when eligible, affects employment outcomes of parents. Another goal is to develop a model by which eligibility can be estimated. The paper will address legal, political, and technical challenges associated with such an effort and present the outcome.

## 452 Using Empirical Likelihood Methods in Survey Sampling

Section on Survey Research Methods, Section on Nonparametric Statistics

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Empirical Likelihood Inference from Sample Survey Data

❖ Jon N. K. Rao, Carleton University, School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6 Canada, *jrao@math.carleton.ca*; Changbao Wu, University of Waterloo

***Key Words:*** empirical likelihood, complex surveys, auxiliary information, missing data, combining surveys

Hartley and Rao (1968) proposed a nonparametric likelihood approach to inference from sample survey data called the “scale-load” approach. This method was discovered 20 years later (Owen, 1988) in mainstream statistics under the name “empirical likelihood” (EL), and its theory is discussed in detail in Owen’s 2001 book. The EL approach is useful particularly for confidence interval estimation. Only recently, the EL has been revived in sample survey literature, and methods that take account of survey design features have been proposed in the context of using known population information and constructing confidence intervals on finite population parameters. This talk will present an overview of new developments in EL inference from sample survey data, focusing on known population information, imputation for missing data, integration of surveys, and confidence intervals.

### Variance Estimation for Empirical Likelihood Calibration Estimators in Unequal Probability Sampling

❖ Jae-kwang Kim, Yonsei University, Department of Applied Statistics, Sinchon-dong, Seoul, 120-749 Korea, *kimj@yonsei.ac.kr*

***Key Words:*** survey sampling, weighting, benchmarking, nonparametric maximum likelihood estimator

Empirical likelihood calibration estimation uses an empirical likelihood function to adjust design weights to fit calibration constraints. Under unequal probability sampling, two types of the empirical likelihood function have been considered. Chen and Sitter (1999) proposed using the pseudo-empirical likelihood function for calibration. Recently, Kim (2005) proposed using another type of empirical likelihood function for calibration estimation. In this paper, we compare the two empirical likelihood calibration estimators in detail and discuss the situations where one method is preferable to the other. Variance estimator using a modification of Binder (1983) linearization method is proposed. The proposed method is shown to be robust against the failure of the homogeneous design weight assumptions. Results from a limited simulation study also are presented.

### Empirical Likelihood Methods for Raking in Complex Surveys

❖ Randy R. Sitter, Simon Fraser University, Department of Statistics and Actuarial Science, Burnaby, BC V5A 1S6 Canada, *sitter@stat.sfu.ca*; Changbao Wu, University of Waterloo

***Key Words:*** raking ratio, generalized raking, marginal counts, Newton-Raphson

The classical raking ratio method of Deming and Stephan (1940) and the generalized raking procedures of Deville et al. (1993) are used for estimation in surveys with auxiliary information in the form of known marginal counts in a frequency table in two or more dimensions. We show the recently developed pseudo-empirical likelihood (PEL) methods can be used effectively for raking in complex surveys. Major advantages of the PEL approach include the availability of efficient and reliable computational algorithms with guaranteed convergence, the existence of simple linearization variance estimators, and the method of constructing confidence intervals through the PEL ratio statistics. We illustrate the proposed method using data from the California Health Interview Survey.

## 453 Advanced Statistical Methods in Psychological Research

● ✪

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Combining Speed and Accuracy To Assess Error-Free Cognitive Processes

❖ Mark Glickman, Boston University, Edith Nourse Rogers

Memorial Hospital, 200 Springs Rd., Bedford, MA 01730-1114, *mg@bu.edu*

***Key Words:*** competing risks, reaction times, speed/accuracy tradeoff, time-to-event model, working memory tasks

Both the speed and accuracy of responding are important measures of performance. A well-known interpretive difficulty is that participants may differ in their strategy, trading speed for accuracy, with no change in underlying competence. Another difficulty arises when participants respond slowly and inaccurately (rather than quickly but inaccurately) due to a lapse of attention. We introduce an approach that combines response time and accuracy information and addresses both situations. The modeling framework assumes two latent competing processes. The first, the error-free process, always produces correct responses. The second, the guessing process, results in all observed errors and some of the correct responses. The approach is applied to a dataset on response times in a working memory test.

## Long-Range Trends and Short-Range Dependencies in Response Time Data

❖ Mario Peruggia, The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210-1247, *peruggia@stat.ohio-state.edu*; Peter F. Craigmile, The Ohio State University; Trisha Van Zandt, The Ohio State University

***Key Words:*** ARMA models, Bayesian models, ex-Gaussian distribution, long range dependence, mixture models, wavelets

Human response time data are widely used in cognitive psychology to evaluate theories of mental processing. Typically, psychologists record the times taken by a subject to react to a succession of stimuli under varying experimental conditions. A careful analysis must distinguish and account for both local dependencies and overall trends. A traditional modeling approach that detrends the data first and then models the local dependencies can be unsatisfactory, especially when data from several subjects that appear to require varying degrees of detrending are considered at once. We build a hierarchical Bayes model that handles the various modeling choices concurrently and incorporates, in a unified fashion, the effects of experimental covariates and outlying observations. We use data from several experiments to illustrate and assess the performance of our modeling strategy.

## Hierarchical Bayesian Methods for Models of Memory Processes

❖ Paul Speckman, University of Missouri-Columbia, 134 Middle Bush Hall, Department of Statistics, Columbia, MO 65211, *speckmanp@missouri.edu*; Jeff Rouder, University of Missouri-Columbia; Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia; Jun Lu, American University

***Key Words:*** signal detection, nonlinear categorical models

Modern analysis of memory phenomenon often is based on small nonlinear categorical models (e.g., the theory of signal detection (Green & Swets, 1966) and Jacoby's Process Dissociation model (Jacoby, 1991)). While these models are sophisticated, conventional analysis in the psychological literature has ignored important sources of variability. It has been shown that ignoring these random effects can lead to biased estimates, which, in principle, can lead to erroneous theoretical conclusions. In our work, we introduce hierarchical models that account for important random effects---subject and item effects, for example. These models also may incorporate priors with correlated random effects in situations where these correlations have theoretical interest. We illustrate our work with an application to the Process Dissociation model.

# 454 A Statistician's Summer Reading List: Modern Topics To Check Out ●

Section on Statistical Education, Section on Teaching Statistics in the Health Sciences

**Wednesday, August 9, 2:00 pm–3:50 pm**

## Analyzing DNA Microarrays with Undergraduate Statisticians

❖ Johanna Hardin, Pomona College, Department of Mathematics, 610 N. College Ave., Claremont, CA 91711, *jo.hardin@pomona.edu*; Laura Hoopes, Pomona College; Ryan Murphy, Pomona College

***Key Words:*** microarray analysis, biostatistics, undergraduate research projects

With advances in technology, biologists have been saddled with high-dimensional data that need modern statistical methodology for analysis. DNA microarrays are able to measure thousands of genes (and their activity) simultaneously in a single sample. Biologists use microarrays to trace connections between pathways or to identify all genes that respond to a signal. The statistical tools we usually teach our undergrads are inadequate for analyzing thousands of measurements on tens of samples. The project materials include readings on microarrays and computer lab activities. The topics covered include image analysis, filtering and normalization techniques, and statistical methods. The course materials are designed for someone with little or no statistical background, but due to the novel concepts covered, they could be adjusted to accommodate practically any background.

## An Introduction to Incomplete Data Regression Methods Used in Practice

❖ Nicholas J. Horton, Smith College, Clark Science Center, Northampton, MA 01063-0001, *nhorton@email.smith.edu*; Ken P. Kleinman, Harvard Medical School

***Key Words:*** multiple imputation, incomplete data, missing data, missing at random, nonignorable nonresponse, maximum likelihood

Missing data arise in almost all real-world studies, and an extensive statistical literature has developed to address these complications. Software to account for missing values is now widely available and increasingly used in published studies. In this talk, we will review these methods, describe software implementations, and summarize the advantages and limitations of their use in practice.

## A Data Mining Reading List

❖ Richard De Veaux, Williams College, MA 01267, *deveaux@williams.edu*

What are the hot topics in data mining? We'll explore what's new in data mining and what statisticians need to know.

## Introducing Bayes in a First Statistics Course

❖ James Albert, Bowling Green State University, Department of Math and Statistics, BGSU, Bowling Green, OH 43403, *albert@bgnet.bgsu.edu*

***Key Words:*** hierarchical modeling, testing, prior information

Although Bayesian modeling is currently popular at the research level, relatively few universities offer classes on Bayesian inference. In this paper, we present several self-contained illustrations of Bayesian thinking suitable for a first-year mathematical statistics class. These examples focus on the advantages of Bayesian thinking, including the incorporation of prior beliefs, the combination of data from related sources, and the production of alternative measures of evidence.

## 455 Recent Advances in Brain Imaging ●

Biometrics Section, Section on Nonparametric Statistics, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Modeling State-Related fMRI Activity Using Change Point Theory

❖ Martin A. Lindquist, Columbia University, Department of Statistics, 1255 Amsterdam Ave MC 4690, New York, NY 10027, *martin@stat.columbia.edu*; Tor D. Wager, Columbia University

***Key Words:*** fMRI, brain imaging, change point, statistical process control

We apply ideas from statistical process control and change point theory to model slowly varying processes with uncertain onset times and durations. The analysis uses activity collected during a baseline period to estimate noise characteristics in the fMRI signal response and make inferences on whether, when, and for how long subsequent activity deviates from the baseline level. We tailor existing exponentially weighted moving average (EWMA) models to individual fMRI time courses, then develop a group analysis using a hierarchical model, which we term Hierarchical EWMA. Once a systematic deviation from baseline has been detected in the group, change point methods are used to estimate the time of change and recovery time. Variations across the brain in the onset and number of change points and in the duration of a shift away from baseline activity are used to classify brain regions.

### Wild Bootstrap for Functional Magnetic Resonance Imaging Data

❖ Hongtu Zhu, Columbia University and New York State Psychiatric Institute, Department of Child Psychiatry, NYSPI unit 74, New york, NY 10032, *hz2114@columbia.edu*; Bradley S. Peterson, Columbia University and New York State Psychiatric Institute

***Key Words:*** hypothesis test, functional MRI, multiple testing, statistical parametric mapping

Multiple testing problems arise commonly in the analysis of functional magnetic resonance imaging (fMRI) data. Much effort has been devoted to developing multiple testing procedures based on the familywise error rate to control overall type I errors. For instance, statistical methods---including random field theory and permutation method---have been used widely to calculate corrected p-values, accounting for tests. In this paper, we discuss wild bootstrap method and consider its application in fMRI data. The finite sample performance of wild bootstrap method is investigated with Monte Carlo experiments. The simulation results suggest the wild bootstrap is a good approach under certain conditions compared to other approaches (e.g., permutation). The results also are illustrated by applications to two real databases.

### Predicting Post-Treatment Brain Activity Using a Bayesian Hierarchical Model

❖ F. DuBois Bowman, Emory University, Department of Biostatistics, 1518 Clifton Road NE, Atlanta, GA 30322, *dbowma3@sph.emory.edu*; Ying Guo, Emory University

***Key Words:*** neuroimaging, fMRI, PET, hierarchical model, prediction

There is growing interest in the use of functional neuroimaging data to help inform medical decisionmaking. For example, knowing the impact of treatment on distributed patterns of brain activity, measured using fMRI or PET, may shed light on whether a treatment is appropriate for a particular patient. The complication is that post-treatment scans are not at a physician's disposal when a treatment decision is made. We develop a Bayesian hierarchical model that enables the use of pretreatment brain scans and subject-specific health characteristics to predict post-treatment brain function. The first level of the hierarchy models within-subject activation effects, and the second level models subject-specific effects in terms of population parameters. Estimation is performed using the EM algorithm. We evaluate the accuracy of our proposed prediction method using K-fold cross-validation.

### Power Calculations for Group fMRI Studies Accounting for Arbitrary Design and Temporal Autocorrelation

❖ Jeanette Mumford, University of Michigan, School of Public Health, 1420 Washington Heights, Ann Arbor, MI 48109-2029, *jmumford@umich.edu*; Thomas Nichols, University of Michigan

***Key Words:*** power analysis, group fMRI, two-level model

Statistical power calculations for group functional magnetic resonance imaging studies are an important study planning tool. Power estimation requires prior knowledge of the expected size of the group effect and its variability. Many factors contribute to these values, including study design, form and magnitude of the temporal autocorrelation, length of the first-level time series, length of the study, and between-subject variability. Desmond and Glover (2002) proposed a method using a paired t-test to analyze percent signal change across subjects, accounting for both between- and within-subject variability components. Although this approach is useful for planning experiments using paired t-test, it cannot complete power calculations for complicated noise and signal models. We introduce a more flexible group-model power calculation method that admits a wider variety of study designs.

### New Kernel Method on Unit Sphere and Its Application to Brain Imaging

❖ Moo Chung, University of Wisconsin-Madison, 1300 University Ave., Madison, WI 53706, *mchung@stat.wisc.edu*

***Key Words:*** spherical harmonics, Green's function, unit sphere, Gauss-Weistrass, kernel smoothing, spline

Spherical harmonics (SPHARM) have been used as building blocks for spline-smoothing on a unit sphere. We present a new general framework for kernel smoothing on the unit sphere using SPAHRM. The kernel is constructed as the Green's function of a partial differential equation (PDE) that has smooth solution. The data residing in the unit sphere serve as the initial condition of the PDE. Then, we approximate the solution of the PDE via the least squares method in the finite subspace of twice integrable function space. The subspace is constructed as a linear combination of the spherical harmonics. As the dimension of the subspace increases, the approximation converges to the solution of the PDE. As a particular example of this powerful technique, we show

the relationship between the Gauss-Weistrass kernel and an isotropic heat flow. This technique has been applied in brain shape analysis.

## 456 Measuring, Monitoring, and Evaluating Internal Revenue Service Data

Section on Government Statistics
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Monitoring SOI Samples

❖ Joseph Koshansky, Internal Revenue Service, P.O. Box 2608, Statistics of Income Division, Washington, DC 20013-2608, *joseph.koshansky@irs.gov*

***Key Words:*** case study, monitoring sampling, statistics of income

Statistics of Income (SOI) Division produces annual income, financial, and tax data based on samples of federal tax and information returns. Because the IRS does not store electronically all the information reported on returns, SOI selects documents the IRS processes at its centers across the country. This paper is a case study of the infrastructure SOI developed to monitor its samples of returns. Interviews, observations, documents, and physical information show how SOI procedures and information databases monitor and verify the control and timely processing of specific returns. Reducing the complexity associated with handling physical documents increased the likelihood SOI met program deadlines and transmitted more complete and accurate data to its customers. Major redesign of the SOI infrastructure is not necessary as electronic filings increase.

### Measuring Nonsampling Error in Exempt Organization Business Income Tax Data

❖ Tamara Rib, Internal Revenue Service, P.O. Box 2608, Statistics of Income Division, Washington, DC 20013, *Tamara.L.Rib@irs.gov*

***Key Words:*** data quality, net difference rate, index of inconsistency

Nonsampling errors are created when transcribing, processing, and editing data from IRS tax forms. This paper will examine one type of nonsampling error associated with Form 990-T, Exempt Organization Business Income Tax Return. A measurement error model will be used to estimate editor error from data collected by the Statistics of Income Division using a quality review sample. Statistics such as the net difference rate, index of inconsistency, and editor design effect will be presented and discussed.

### Customer Satisfaction Initiatives at IRS's Statistics of Income: Using Surveys To Improve Customer Service

❖ Ruth Schwartz, Internal Revenue Service, P.O. Box 2608, Washington, DC 20013, *ruth.a.schwartz@irs.gov*; Beth Kilss, Internal Revenue Service

***Key Words:*** customer satisfaction, survey results

IRS's Statistics of Income (SOI) Division conducts statistical studies on the operations of tax laws and publishes annual reports, including the quarterly "SOI Bulletin," which includes statistics produced from tax and information returns. As part of its commitment to deliver quality customer service, SOI's Statistical Information Services (SIS) office responds to thousands of data and information requests annually by providing SOI data with technical assistance. To ensure customer needs are being met through the SIS office and its flagship publication, SOI has been measuring customer satisfaction for both via customer satisfaction surveys. This paper focuses on three aspects of these surveys: the process by which customers were surveyed, the findings from the surveys, and the steps being taken to use the results to further improve products and services.

### Performance Measures within the Statistics of Income Division

❖ Kevin Cecco, Internal Revenue Service, P.O. Box 2608, Statistics of Income Division, Washington, DC 20013, *Kevin.Cecco@irs.gov*

***Key Words:*** performance, measurement

Developing performance measures continues to play an important role for many of the federal statistical agencies. The Statistics of Income (SOI) Division, within the Internal Revenue Service, relies on 12 performance measures for evaluating how well data and services are provided to customers. The performance measures were developed based on commitments to customer satisfaction, employee satisfaction, and business results. As a critical source of valuable information, these performance measures allow SOI to tailor its services and data dissemination efforts more effectively. This paper provides a perspective of the critical need of performance measures in federal statistical agencies, offers basic guidelines for constructing performance measures, and highlights results over the past three years associated with SOI's current performance measures.

## 457 Visualization of Large Datasets

Section on Statistical Graphics, Section on Statisticians in Defense and National Security
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Tours of Large Multivariate Data

❖ Dianne Cook, Iowa State University, Department of Statistics, 325 Snedecor Hall, Ames, IA 50011-1210, *dicook@iastate.edu*

***Key Words:*** multivariate data, data mining, dynamic graphics, visualization

This talk will describe binning of multivariate data for display using tours. We will examine the effect of binning in the multivariate space and binning the data projections shown by the tour. The types of patterns we are interested in exploring are multivariate density, anomalies---such as outliers---and clusters.

### Visualization of Features in Curve Estimates and Application to Genetic Loci Mapping

❖ Myung Hee Lee, The University of North Carolina at Chapel Hill, 210 Smith Building, CB 3260, UNC Chapel Hill, Chapel Hill, NC 27599, *mhlee@email.unc.edu*;
Ivan Rusyn, The University of North Carolina at Chapel Hill; David Threadgill, The University of North Carolina at Chapel Hill; J. Stephen Marron, The University of North Carolina at Chapel Hill

***Key Words:*** visualization, nonparametric smoothing, scale space, genetic loci mapping

Statistical smoothing methods are useful tools for exploratory data analysis. SiZer (based on studying statistical SIgnificance of ZERo crossings of smoothed estimates) is a visualization tool that provides insight as to whether the observed features (e.g., peaks and valleys) in a curve estimate are statistically significant. In this work, to study the genetic association with quantitative traits, we utilize a dense genotyping data obtained across a large panel of inbred mouse strains. As an exploratory analysis tool, SiZer enables us to provide insight into the density of the genetic markers on the chromosomes map genetic loci associated with a specific continuous phenotype measured in inbred mice.

### Upscaling Statistical Graphics

❖ Martin Theus, University of Augsburg, Universit‰tsstr. 14, Augsburg, 86135 Germany, *martin.theus@math.uni-augsburg.de*

***Key Words:*** large data, statistical graphics, interaction, alpha blending, data mining, visualization

This talk investigates what influence the size of a dataset has on standard statistical graphics. Area plots---like bar charts, histograms, and mosaic plots---are relatively invariant against the increasing size of a dataset. Point plots, all plots that plot a glyph for each observation, have to cope with the problem of overplotting and an increasing number of extreme outliers. The definitions of other plots (e.g., boxplots) are examined to see if they still seem to be appropriate when datasets get large. Problems, solutions, and modifications are presented, which are based on either more interactivity or on advanced plotting techniques.

### Visualization of Statistical Models on a Billion Cases

❖ Graham Wills, SPSS Inc., 233 S. Wacker Drive, Chicago, IL 60606-6307, *gwills@spss.com*

***Key Words:*** modelling, large data, visualization, distributed computation

Models for large datasets are complex in qualitatively different ways than models for normal-sized datasets. Distributed computation, the use of model updating rather than recalculation, and general issues of model management become critical to success. In SPSS, we are researching techniques to manage such models. In this talk, we will outline the issues and demonstrate our approach to solving the associated problems. The focus will be on visualizing the model artifacts---the models, their accuracy, their behaviors over time, the scored results, and the model-building process.

## 458 Topics in Univariate and Multivariate Time-to-Events Analysis ●

Section on Bayesian Statistical Science, Biometrics Section, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Bayesian Modeling of Multiple Episode Occurrence and Severity with a Terminating Event

❖ Amy Herring, The University of North Carolina at Chapel Hill, 407 N. Elliott Road, Chapel Hill, NC 27514, *amy_herring@unc.edu*

***Key Words:*** Bayesian, joint model, pregnancy, latent variable

In certain epidemiologic studies, one's health condition can affect the frequency and intensity of episodes that occur repeatedly and may be related to an event time. For example, vaginal bleeding during pregnancy may indicate problems predictive of preterm delivery. We propose a joint model for a multiple episode process and an event time. The frequency of the episode process is characterized by a latent variable model, which allows an individual's episode intensity to change dynamically over time. This latent episode intensity is then incorporated as a predictor in a discrete time model for the terminating event.

### Bayesian Semiparametric Inference for the Accelerated Failure Time (AFT) Model Using Hierarchical Mixture Modeling with N-IG Priors

❖ Alessandra Guglielmi, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133 Italy, *alessandra.guglielmi@polimi.it*; Raffaele Argiento, Universit‡ Commerciale Luigi Bocconi/CNR-IMATI; Antonio Pievatolo, CNR-IMATI; Fabrizio Ruggeri, CNR-IMATI

***Key Words:*** AFT regression models, Bayesian semiparametrics, mixture models, MCMC algorithms

We will pursue a Bayesian semiparametric approach for an AFT regression model, usually considered in survival analysis, when the baseline survival distribution is a mixture of parametric densities on the positive reals with a nonparametric mixing measure. A popular choice for the mixing measure is a Dirichlet process, yielding an MDP model for the error. Here, as an alternative to the Dirichlet process, the mixing measure is equal to an N-IG prior, built from normalized inverse-Gaussian finite dimensional distributions, as recently proposed in the literature. A comparison of the models will be carried out. MCMC techniques will be used to estimate the predictive distribution of the survival time, along with the posterior distribution of the regression parameters. The efficiency of computational methods also will be compared, using both real and simulated data.

### A Bayesian Dynamic Frailty Model for Recurrent Events

❖ Changhong Song, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269, *changhon@stat.uconn.edu*; Lynn Kuo, University of Connecticut

In many medical studies, times to occurrence of a specific event---such as recurring hospitalizations or tumors---have been collected. To analyze this kind of data, within subject association must be modeled to ensure correct inferences for the treatment effects. Some dynamic frailty models have been proposed to model this association and the evolution of individual effect over time. In this study, we propose a new family of dynamic frailty model to better describe each individual's risk that changes with age during the trial. In the new model, the individual frailty effect is modeled as a time-varying effect with unknown change points. Both the unobserved change points and the intensities of the frailty function are modeled as latent variables in the model. The implementation of Bayesian inference and model selection using a Markov chain Monte Carlo (MCMC) algorithm is developed.

### Multivariate Times-to-Events Analysis for Marketing Data Using Frailty Models

❖ Nalini Ravishanker, University of Connecticut, Department of Statistics, U-4120, 215 Glenbrook Road, Storrs, CT 06269-4120, *nalini.ravishanker@uconn.edu*; V. Kumar, University of Connecticut; Rajkumar Venkatesan, University of Connecticut

***Key Words:*** channel adoption, positive stable distribution, cross-ratio

For frailty models to analyze multivariate times to events, we assume that, given the unobserved frailty variable, the hazard for each time to event follows a modified proportional hazards model with the frailty, covariate effect, and baseline hazard acting multiplicatively. Inference under the Bayesian framework is described in the context of families of gamma and positive stable frailty distributions and useful dependence measures are computed. An illustration from marketing involves a framework for predicting the duration for multi-channel customers to adopt another transaction channel and to identify various drivers of channel adoption broadly classified as customer purchase behavior, supplier-specific factors, and customer heterogeneity. This is joint work with V. Kumar and Raj Venkatesan.

### Flexible Models for Quantile Regression

❖ Milovan Krnjajic, Lawrence Livermore National Laboratory, 7000 East Ave., L227, Livermore, CA 94550, *milovan@soe.ucsc.edu*; Athanasios Kottas, University of California, Santa Cruz

***Key Words:*** quantile regression, Bayesian nonparametric modeling, Dirichlet process mixtures, censored data, Markov chain Monte Carlo

We propose Bayesian nonparametric methodology to model error distribution in an additive quantile regression setting. Bayesian modeling enables full inference not only for the model parameters, but for any functional of interest of the response distribution. Moreover, nonparametric prior probability models allow the shape of error density to adapt to data, so provide more reliable predictive inference than models based on parametric error distributions. We develop three models based on Dirichlet process mixtures of uniform densities, which can capture the shape of any unimodal error density. The models are further extended to handle censored observations. We also develop quantile regression models, which allow the error distribution to change nonparametrically with the covariates. We use Markov chain Monte Carlo techniques for posterior simulation.

## 459 Statistical Approaches to Handling Data Quality: Issues and Evaluating Intervention Effectiveness in HIV/AIDS Research ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Assessing Mediation in HIV Intervention Studies

❖ Felicia Hardnett, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, MSE48, Atlanta, GA 30333, *fhardnett@cdc.gov*; Craig Borkowf, Centers for Disease Control and Prevention; Sherri Pals, Centers for Disease Control and Prevention; Ann O'Leary, Centers for Disease Control and Prevention; Jeffrey Parsons, City University of New York-Hunter College; Cynthia Gomez, University of California, San Francisco

***Key Words:*** mediation, asymmetric confidence limits, HIV intervention studies, intervention trials, intervening variable effect, indirect effect

Recent literature suggests the widely used approach for assessing mediation, the causal st eps method, can be severely underpowered in many research settings. In the present study, we compare results obtained by 95% asymmetric confidence limit (ACL) construction with results obtained by the causal steps method. ACL construction uncovered significant mediating factors not identified using the causal steps method. We also identified marginally significant suppressors. Constructing ACLs is preferred for this type of analysis because of its statistical power and because it does not require the intervention to have a significant effect on the outcome. This method is also preferred because the distribution of the mediation effect is asymmetric. Also, ACLs present the size of the mediating effect, rather than just a binary decision of significance.

### Are Higher Levels of Multilevel (Hierarchical) Models Necessary? Application to High-Risk Sexual Behavior Data

❖ DeMarc Hickson, Emory University/Centers for Disease Control and Prevention, 1600 Clifton Road, MS E48, Atlanta, GA 30329, *ahe5@cdc.gov*; Lance Waller, Emory University; Lillian Lin, Centers for Disease Control and Prevention

***Key Words:*** multilevel models, hierarchical data, deviance information criterion

Community Intervention Trial for Youth (CITY) project personnel conducted venue-based interviews of young men who have sex with men (MSM) for four consecutive summers in 12 communities throughout the United States as part of the evaluation of a community-level intervention field trial. This venue-based sampling introduced four levels to the data structure: individual, venue, community, and time. Multilevel (hierarchical) models represent datasets with two or more levels of clustering and allow investigators to examine cross-level effects and include level- or cluster-specific information (e.g., individual and/or macro group). The purpose of this research is to determine if higher levels of (hierarchical) data structures are necessary. We perform simulation studies to the degree to which the deviance information criterion (DIC) will assess the need for higher-data clusters.

### How Good Is Good Enough? An Investigation of the Effect of Uncertainty in Survey Parameters on Estimates of HIV Prevalence, Guyana 2004

❖ Maxine Denniston, Centers for Disease Control and Prevention, 4757 Banner Elk Drive, Stone Mountain, GA 30083, *mmd1@cdc.gov*; Nicole Seguy, Centers for Disease Control and Prevention; Wolfgang Hladik, Centers for Disease Control and Prevention

***Key Words:*** multi-stage sampling, estimation, HIV prevalence

Implementing a complex sampling design in a limited resource country presents many challenges. In response to a report of high levels of HIV infection, a survey of miners in Guyana was fielded in 2004 using two-stage cluster sampling to estimate HIV prevalence and risk behaviors. Major challenges included constructing first- and second-stage sampling frames, physical inaccessibility of sampled mines, and determining the number of miners present on the day a survey was conducted. This paper investigates the effect on the HIV prevalence estimate of uncertainties in survey parameters resulting from these challenges, such as errors in estimating the number of miners present at specific mines and differing rates of HIV infection in miners who refused testing or those working at inaccessible mines. Results should be useful in planning and interpreting results of future surveys.

### A Multiple-Recapture Approach When a Unique Identifier Is Not Available

❖ Ruiguang Song, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, MS E48, Atlanta, GA 30333, *RSong@cdc.gov*; H. Irene Hall, Centers for Disease Control and Prevention; John Gerstle, Centers for Disease Control and Prevention; Lisa Lee, Centers for Disease Control and Prevention

**Key Words:** capture-recapture, completeness, duplicate analysis, loglinear model, unique identifier

A condition for using capture-recapture methods to estimate the size of a population is that individuals captured by different sources can be identified. However, due to confidentiality concerns, data sources may not release information on variables that can identify a person uniquely. In this situation, duplicate analysis can be used to estimate the numbers of individuals captured by different sources. Two source methods are considered, compared, and applied to multiple-source models. Unlike the assumption for the two-source model, capture probabilities between multiple sources are not necessarily independent. A simulation study is conducted to evaluate the performance of the multiple recapture method when a unique identifier is not available. We apply this method to estimate the completeness of HIV/AIDS reporting.

## 460 Statistical Applications in Water Quality and Monitoring ●

Section on Statistics and the Environment

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Clustering Using Stressor-Response Relationships

❖ Samantha C. Prins, Virginia Polytechnic Institute and State University, Hutcheson Hall 0439, Department of Statistics, Blacksburg, VA 24061, *sbates@vt.edu*; Eric P. Smith, Virginia Polytechnic Institute and State University

**Key Words:** stressor-response, classification, model comparison, partition modeling, Voronoi diagram, model selection

Environmental data collected over large spatial regions often are analyzed using a single model though multiple models may be more appropriate. To find important models over smaller spatial regions, we will present a method for clustering sites to maximize the stressor-response relationships within clusters. The method uses Voronoi diagrams to assign sites randomly to one of K clusters based on values of nonstressor variables. Within each cluster, the method selects the optimal stressor-response relationship and fits the chosen model to all sites in that cluster. This form of model fitted is flexible and may be different across clusters. We will discuss methods to optimize the Voronoi assignment of sites to clusters, stressor-response model fitted within each cluster, and the number of clusters. We will present an application of the method to data from the mid-Atlantic Highlands.

### Composite Sampling for Environmental Variables

❖ Sylvia Esterby, The University of British Columbia, Mathematics Statistics and Physics, UBCO 3333 University Way, Kelowna, BC V1V 1V7 Canada, *sylvia.esterby@ubc.ca*; Abdel H. El-Shaarawi, National Water Research Institute

**Key Words:** composite sampling, environmental data, water quality

Composite sampling can be an efficient sampling method for detecting rare conditions and estimating characteristics of quantitative variables. In the collection of environmental data, one of the advantages of the method is the ability to increase spatial and temporal coverage without increasing the number of samples. It has been shown that composite sampling results in a gain in the efficiency when estimating the mean. This will not always be the case when estimating the variance, where kurtosis plays an important role. Efficient estimation is evaluated for several designs, including composite sampling, and for distributions relevant to water quality.

### Modified Power Priors with Multiple Historical Datasets in Water Quality Evaluation

❖ Yuyan Duan, Bristol-Myers Squibb Company, 22 Brandywine Court, Washington Crossing, PA 18977, *yyduan@yahoo.com*; Keying Ye, The University of Texas at San Antonio; Eric P. Smith, Virginia Polytechnic Institute and State University

**Key Words:** historical data, power prior, water quality standards

The modified power prior has been proposed to utilize additional information to improve the water quality evaluation. In this presentation, we investigate three power prior methods used in incorporating multiple historical datasets. One of the methods is chosen among the three by comparing the mean squared errors (MSE) in estimating the parameter of interest. The Bayesian statistical model using the selected power prior method is compared with the random effects model for the accommodation of potential heterogeneity between current and historical samples. Furthermore, we apply the modified power prior with a random effects model. This approach is shown to recognize the differences among groups the most, compared to using only power priors and only random effects models. Several sets of water quality data are studied to illustrate the implementation of the modified power prior method.

### Model-Based Clustering in a Brook Trout Classification Study within the Eastern United States

❖ Huizi Zhang, Virginia Polytechnic Institute and State University, Department of Statistics, 1216 University City Blvd., Apt C 34, Blacksburg, VA 24060, *huzhang@vt.edu*; Samantha C. Prins, Virginia Polytechnic Institute and State University; Eric P. Smith, Virginia Polytechnic Institute and State University

**Key Words:** model-based clustering, Voronoi tessellations, stressor-response, classification, brook trout

Cluster analysis is a commonly used technique on multivariate data that aims to group objects into clusters such that objects are similar within each cluster and dissimilar between different clusters. We developed a model-based clustering method for analysis of ecological data that groups objects by their empirical stressor-response relationship, rather than their attribute values, using Voronoi tessellations to find the optimal grouping solution. In particular, we extended the Voronoi approach to deal with categorical response data. This required the development of appropriate optimality criteria. This method is applied to a carefully compiled dataset of Brook trout presence/absence within the eastern United States. Results indicate fairly strong relationships that vary spatially.

### Nonparametric Harmonic Regression for Estuarine Water Quality Data

❖ Melanie Autin, University of South Carolina, 219D LeConte College, Department of Statistics, Columbia, SC 29208, *autinm@stat.sc.edu*; Don Edwards, University of South Carolina

**Key Words:** periodicity, generalized additive models, harmonic regression

Periodicity is omnipresent in environmental time series data. For modeling estuarine water quality variables, harmonic regression analysis has long been the standard for dealing with periodicity. Generalized additive models (GAMs) allow more flexibility in the response function, permitting parametric, semiparametric, and nonparametric regression functions of the predictor variables. We compare harmonic regression, GAMs with cubic regression splines, and GAMs with cyclic regression splines in simulations and using water quality data collected from the National Estuarine Research Reserve System (NERRS). The generalized additive models are more adaptive and require less user intervention.

## 461 Statistics for Weather Forecasting II: Challenges and Opportunities ● ✪

Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences, Section on Statistics and the Environment
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Probabilistic Forecasting in Meteorology

❖ Barbara Brown, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, *bgb@ucar.edu*

***Key Words:*** probability, weather, forecast, uncertainty

Expressing weather forecasts as probabilities has been part of operational meteorological forecasting in the United States since at least 1965, when the Weather Bureau produced its first probability of precipitation forecasts. However, the concept that weather forecasts are uncertain has been understood since the early days of weather forecasting (e.g., in the late 1800s, Cleveland Abbe---the “father” of weather forecasting in the United States---called his forecasts “probabilities”). Uncertainty in weather forecasts is estimated in a number of ways, including subjectively and statistically. More recently, probabilistic weather forecast information has been derived from multiple (ensemble) runs of dynamical weather prediction models. The development of probabilistic forecasting in meteorology and current approaches for estimating and expressing forecast uncertainty will be explored.

### Probabilistic Forecasts, Calibration, and Sharpness

❖ Fadoua Balabdaoui, Institut f¸r Mathematische Stochastik, Weender Strasse 80, Goettingen, 37073 Germany, *fadoua@math.uni-goettingen.de*; Tilmann Gneiting, University of Washington; Adrian E. Raftery, University of Washington

***Key Words:*** density forecast, ensemble prediction system, forecast verification, predictive distribution, PIT, scoring rule

We propose a diagnostic approach to the evaluation of probabilistic forecasts of continuous weather variables, such as temperature or wind speed, that take the form of predictive PDFs or CDFs. This approach is based on the paradigm of maximizing the sharpness of the predictive distributions subject to calibration. Calibration refers to the statistical consistency between the distributional forecasts and the observations, and is a joint property of the predictions and the events that materialize. Sharpness refers to the concentration of the predictive distributions, and is a property of the forecasts only. A simple theoretical framework suggests various modes of calibration. We propose and study tools for checking calibration and sharpness, among them the probability integral transform (PIT) histogram, marginal calibration plots, the sharpness diagram, and proper scoring rules.

### Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: the Regime-Switching Space-Time (RST) Method

❖ Tilmann Gneiting, University of Washington, Box 354322, Seattle, WA 98195, *tilmann@stat.washington.edu*; Kristin Larson, 3 Tier Environmental Forecast Group, Inc.; Kenneth Westrick, 3 Tier Environmental Forecast Group, Inc.; Marc G. Genton, Texas A&M University; Eric Aldrich, Duke University

***Key Words:*** weather prediction, continuous ranked probability score, spatio-temporal, predictive distribution, time series, calibration

With the global proliferation of wind power, accurate short-term forecasts of wind resources at wind energy sites are becoming paramount. Regime-switching space-time (RST) models merge meteorological and statistical expertise to obtain accurate, calibrated, fully probabilistic forecasts of wind speed and power. The model formulation is parsimonious, yet takes account of all the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal nonstationarity, conditional heteroscedasticity, and non-Gaussianity. The RST method uses geographically dispersed meteorological observations in the vicinity of the wind farm as offsite predictors, identifies forecast regimes, and fits conditional predictive models with good results for two-hour forecasts of hourly average wind speed at the Stateline wind energy center.

### Detection and Modeling of Long Memory in Biases of Daily Forecasts of Surface Air Pressure and Temperature

❖ Yulia Gel, University of Waterloo, 200 University Ave., W., Waterloo, ON N2L3G1 Canada, *ygl@math.uwaterloo.ca*; Bovas Abraham, University of Waterloo

***Key Words:*** weather forecasting, spatio-temporal modeling, time series analysis, long memory dependence

We consider a “deterministic” error, or bias, (i.e., forecast minus observed) of daily surface pressure and temperature forecasts with the lead time from one to 16 days ahead for different locations worldwide. The goal is to predict a future bias to decrease root mean square error (RMSE) of weather forecasts. We compare modeling of biases using long AR models and long-range dependence ARFIMA models for various lead times and locations.

## 462 Reducing the Risk of Data Disclosure through Swapping and Other Masking Procedures ● ✪

Section on Survey Research Methods, Section on Statisticians in Defense and National Security
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Reducing the Risk of Data Disclosure through Area Masking: Limiting Biases in Variance Estimation

❖ Inho Park, Westat, 1650 Research Blvd., Rockville, MD 20850, *InhoPark@westat.com*; Sylvia Dohrmann, Westat; Jill Montaquila, Westat; Leyla Mohadjer, Westat; Lester R. Curtin, Centers for Disease Control and Prevention

***Key Words:*** disclosure control, segment swapping, stratified multistage sampling, weighted mean, design effect, national health and nutrition examination survey

When preparing public files, variables used to compute variances are sometimes masked to limit the risk of data disclosure. For example,

in area surveys with a limited number of primary stage units (PSUs), the original PSUs are split and recombined to construct pseudo-PSUs with swapped second-stage units (segments). Masking PSU identifiers is an effective way to reduce data disclosure risks. However, such masking results in biased variance estimates, as observed in previous work (Dohrmann et al. 2002, 2004). This paper considers alternative masking strategies, with a focus on reducing the effects of segment swapping on the variance estimates, and the resulting bias. Data from the National Health and Nutrition Examination Survey is used to illustrate the impact of the new masking strategies on variance estimates computed using the masked PSUs.

### Disclosure Limitation in NAEP: a Probabilistic Approach to Swapping Sample Selection

❖ David Freund, Educational Testing Service, Rosedale Road, Princeton, NJ 0851, *dfreund@ets.org*; Andreas Oranje, Educational Testing Service

***Key Words:*** NAEP, data perturbation, data swapping, disclosure risk

NAEP data can be analyzed in detail using a publicly available online tool. With such access, disclosure risk increases sharply and government standards require preventive action, preferably by data swapping. Two characteristics complicate the implementation of a naÔve data swap for NAEP. First, random swaps are not only required to have no impact on several levels of margins of survey indicator distributions, but also of proficiency estimates. Second, the probabilistic nature of the sample puts certain subgroups at greater risk, even at the highest level of aggregation. The solution chosen in NAEP is a procedure that aims to protect two types of margins and incorporates probabilistic swap sample selection. Results from 2005 NAEP data show that relatively small swapping rates can provide reasonable disclosure risk prevention without severe reductions in data utility.

### Tactics for Reducing the Risk of Disclosure Using the NCES DataSwap Software

❖ Thomas Krenzke, Westat, 1650 Research Blvd., Rockville, MD 20850, *tomkrenzke@westat.com*; Stephen E. Roey, Westat; Sylvia Dohrmann, Westat; Leyla Mohadjer, Westat; Wen-Chau Haung, Westat; Steve Kaufman, Retired; Marilyn Seastrom, National Center for Education Statistics

***Key Words:*** confidentiality, statistical disclosure control

The National Center for Education Statistics (NCES) developed statistical guidelines for reducing the risk of disclosure prior to release of microdata. Among the standards is the requirement for data swapping, which is used to perturb (or mask) the data to reduce the risk of disclosing the identity of individuals or entities without impacting aggregate data. Working with NCES, Westat addressed several issues relating to complex surveys while enhancing the NCES standardized swapping software, DataSwap. This paper discusses data-swapping tactics for reducing the risk of disclosure (such as in high-risk domains, high-risk variables, hierarchical data structures), maintaining data consistency (e.g., handling skip patterns and recodes), and maintaining data quality (addressing missing data in swapping variables, controlled swapping approaches).

### Combinations of SDC Methods for Numerical Microdata

❖ Anna Oganian, National Institute of Statistical Sciences, 19 TW Alexander Drive, RTP, NC 27709-4006, *aoganyan@niss.org*; Alan Karr, National Institute of Statistical Sciences

***Key Words:*** statistical disclosure control (SDC), microdata, disclosure control methods, data utility, disclosure risk, combinations of SDC methods

Publication of microdata raises the problem of preventing disclosure of confidential information about particular respondents without significantly damaging the utility of the data. A number of methods have been proposed for microdata protection. These methods may be implemented with different degrees of intensity through the appropriate parameter setting. However, even parameter variation may not be a refined enough tool to control the damage to the data caused by the method. In this paper, we propose a new approach in the protection of numerical microdata: applying several stages of masking to the data in such a way that the released data becomes closer to the original on every new stage of masking. At the same time, this approach makes it harder for the intruder to get precise estimates of original values based on the released data.

## 463 Making Statistical History: Collecting, Preserving, and Providing Access to Records of the American Statistical Community

Committee on Archives and Historic Materials, General Methodology, Section on Statistical Education

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Making Statistical History: Collecting, Preserving, and Providing Access to Records of the American Statistical Community

❖ John Paul Deley, Energy Information Administration, Forrestal Building EI30, 1000 Independence Ave SW, Washington, DC 20585, *johnpaul.deley@eia.doe.gov*; ❖ Pat McClellan, American Statistical Association, 1429 Duke Street, Alexandria, VA 22314-3469, *pat@amstat.org*; ❖ Rich Allen, ASA Committee on Archives and History, 8006 Springfield Village Drive, Springfield, VA 22152-3435, *reisepaar2@earthlink.net*

***Key Words:*** history, biography, archives, research, bibliography, documentation

The preservation of ASA history constitutes an important investment in the Association's future. The Committee on ASA Archives and Historical Materials assists in the preservation, storage, and cataloging of the holdings of noncurrent institutional records of the ASA that are maintained at numerous repositories because of their historical importance and research value. This session will provide an overview of the diverse collection and the committee's recent work. It will focus on methodologies for placing the documentary heritage of the group into a national and international context by facilitating a dialog on increased technical, academic, and biographical use of these vital information assets.

## 464 Unit Nonresponse in Surveys IV ●

Section on Survey Research Methods

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Evaluation of Using a Model-Assisted Sampling Paradigm versus a Traditional Sampling Paradigm in a Nationally Representative Establishment Survey

❖ Marcus Berzofsky, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, *berzofsky@rti.org*; Brandon Welch, RTI International; Rick L. Williams, RTI International; Paul Biemer, RTI International

***Key Words:*** establishment survey, model-assisted sampling

National inference from a survey is traditionally based upon probability sample selection and survey weighting. For studies requiring estimates for a variety of rare subgroups with variable eligibility and response rates, a model-assisted approach might be considered to better control the subgroup sample sizes. Representative inference under the model-assisted sampling paradigm can be achieved using quotas combined with model-based weighting that does not depended on probability weighting. Using data from a national survey of establishments, we simulated a model-assisted paradigm and evaluated if the estimates were consistent with those under the traditional paradigm. Our findings suggest that the model-assisted approach offers advantages over the traditional sampling approach. For our purposes, a hybrid approach that captures the major advantages of both paradigms proved to be optimal.

### Using Telephone-Exchange Data To Adjust for Nonresponse: Application in an Establishment Survey

❖ Stephen R. Williams, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08540, *swilliams@mathematica-mpr.com*; Ronghua Lu, Mathematica Policy Research, Inc.

***Key Words:*** nonresponse, analysis weights, propensity modeling, establishment telephone surveys

We used logistic regression models to predict probabilities that a unit will be located and, if located, will respond. The inverse of the response propensity can be used to adjust the basic sampling weights. We have applied this method for computing weights in a range of settings, some using telephone-exchange data for household surveys, but know of no application to establishment surveys. This paper describes the application in a recent national survey of educational institutions. The sampling frame for this telephone survey was constructed from several list sources. As expected from such a frame, the survey staff was unable to contact and establish eligibility for a relatively large portion of the sample. With little auxiliary data about both the screened and unscreened sample units this survey offered an important application of propensity modeling using telephone-exchange level data.

### Nonresponse Adjustment Using Logistic Regression: To Weight or Not To Weight?

❖ Eric A. Grau, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08543, *egrau@mathematica-mpr.com*; Frank Potter, Mathematica Policy Research, Inc.; Stephen R. Williams, Mathematica Policy Research, Inc.; Nuria Diaz-Tena, Mathematica Policy Research, Inc.

***Key Words:*** unit nonresponse, weighting, propensity modeling, nonresponse adjustment, survey bias and precision, community tracking study

Unit nonresponse in sample surveys is accommodated by reallocating the weights of unit nonrespondents to respondents. One way of doing this is to develop logistic regression models to predict the probability of response. The inverses of the predicted probabilities from these models are then used to adjust the sampling weights. In rounds two and three of the Community Tracking Study (CTS) Household and Physician Surveys, nonresponse adjustments to the weights were carried out using weighted logistic regression models. In the fourth round of the survey, unweighted logistic regression models were used to adjust for nonresponse, with design variables, basic sampling weights, and higher order interactions included in the models, following a methodology introduced in papers by Vartivarian and Little (2003). In this paper, we compare nonresponse adjustments using the two methods.

### Response Rates and Response Patterns among New Enterprises: Results from the Kauffman Firm Survey

❖ Frank Potter, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08543, *FPotter@mathematica-mpr.com*; Yuhong Zheng, Mathematica Policy Research, Inc.; David DesRoches, Mathematica Policy Research, Inc.; Janice Ballou, Mathematica Policy Research, Inc.; Zhanyun Zhao, Mathematica Policy Research, Inc.

***Key Words:*** nonresponse, establishment surveys, respondent characteristics, web-based, CATI

Economic growth is related to continued formation of new enterprises. The Ewing Marion Kauffman Foundation sponsors research into entrepreneurship and new business formation. The Kauffman Firm Survey (KFS) is a national-level longitudinal survey of new businesses conducted for the Foundation by Mathematica Policy Research (MPR). The KFS consists of an initial survey of 5,000 enterprises formed in 2004 with annual follow-up surveys to collect data on the new business characteristics and financing needed to create and sustain them. Survey data will be available to researchers to improve the understanding of the dynamics of new businesses. MPR selected the sample using D&B data files and is using web-based and CATI data collection procedures to capture the information. The purpose of this paper is to describe the characteristics and determinants of response among this elusive population.

### Estimation of Attrition Biases in SIPP

❖ Eric Slud, U.S. Census Bureau, Statistical Research Division, Washington, DC 20233-9100, *Eric.V.Slud@census.gov*; Leroy Bailey, U.S. Census Bureau

***Key Words:*** attrition adjustment, nonresponse, longitudinal survey, variance estimation

This paper develops an estimator for the bias due to attrition nonresponse weighting within a longitudinal survey, with associated standard errors. The estimator was proposed by Bailey (2004) in connection with the Census Bureau's Survey of Income and Program Participation (SIPP), for cross-sectional survey items. It is generalized here to longitudinal survey items, and design-based variance formulas and estimators based on VPLX and the Fay (1989) method plus some alternative approximations are given. The methods are illustrated and results presented using data from the SIPP 1996 panel. Additionally, the paper explores the use of estimated attrition-weighting for some items in a longitudinal survey like SIPP in a regression setting to provide attrition-weighting for other items.

### Subsampling Nonrespondents: Issues of Calculating Response Rates

❖ Sonya Vartivarian, Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024, *SVartivarian@Mathematica-mpr.com*; Sameena Salvucci, Mathematica Policy Research, Inc.; Donsig Jang, Mathematica Policy Research, Inc.; Daniel Kasprzyk, Mathematica Policy Research, Inc.

***Key Words:*** sampling weights, nonresponse bias, subsampling, response rate, data quality

Resource considerations often contribute to the decision to subsample nonrespondents for follow-up. For example, the resources saved by limiting the number of nonrespondents to follow may allow for a more effective, and often more expensive, mode of data collection for those subsampled. We consider weighted response rates under subsampling schemes as a measure of data quality. When subsampling is used to follow-up nonresponse, AAPOR guidelines suggest a weighted response rate that sets the nonsampled unit weights to zero and weights the subsampled unit weights by the inverse of the subsampling fraction. We also consider variations of the calculation of weighted rates when subsampling occurs and when the nonsampled units are still followed, but through a different mode. Finally, we discuss the relationship between a weighted response rate and its sampling variability.

### Physician Survey Response Methods Research

❖ Catharine Burt, National Center for Health Statistics, 3311 Toledo Road, Room 3409, CDC, Hyattsville, MD 20782, *cburt@cdc.gov*; David Woodwell, National Center for Health Statistics

***Key Words:*** response rate, methodological studies, NAMCS

Health surveys targeting physicians historically have had difficulties in obtaining high response rates. Response rates to the National Ambulatory Medical Care Survey (NAMCS) have been declining over the last 10 years. In the NAMCS, a sample of office-based physicians are asked to provide a brief interview about the office practice and to complete a medical record abstract form on a sample of 30 office visits during their randomly-assigned week. NCHS conducted several studies between 2000-2002 to gauge the impact of different methods on physician cooperation. Samples of physicians were randomly assigned to treatment and control groups in 3 studies to test response effects: inclusion of a motivational insert with the introductory letter, offering monetary or token gift, and form length. Only form length had an effect on response rate. Nonresponse reduction & bias strategies are discussed.

## 465 Survey-Based Variance Estimation II ●

Section on Survey Research Methods

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Study of Volatility and Smoothing of Estimated Variances in the Employment Cost Index Program

❖ Meghan S. O'Malley, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212, *O'Malley.Meghan@bls.gov*; Daniell Toth, Bureau of Labor Statistics; Chester Ponikowski, Bureau of Labor Statistics

***Key Words:*** variance smoothing, exponential smoother, nonstationary time series, seasonality, Laspeyres index

The Employment Cost Index (ECI), one of the key products produced by the National Compensation Survey (NCS) and conducted by the Bureau of Labor Statistics (BLS), is a Laspeyres index that provides quarterly and annual measures of the rate of change in compensation per hour worked including wages and salaries, and employer cost of employee benefits. Point estimates for ECI variances are produced using the balanced repeated replication (BRR) procedure and provided to data users as smoothed estimates in the form of five-year moving averages of these BRR variance estimates. This paper presents an analysis of the volatility of the estimated variances as a time series, discussion of the appropriateness of smoothing, development of exponential smoothing procedures taking into account specific properties of the ECI variance estimates, and a comparison of the results.

### Estimation of Generalized Variance Functions for the 2003 Survey of Doctorate Recipients

❖ Michael Yang, National Opinion Research Center, 1350 Connecticut Ave., NW, Washington, DC 20036, *yang-michael@norc.org*; Yongyi Wang, National Opinion Research Center

***Key Words:*** variance estimation, GVF, SDR

A generalized variance function (GVF) is a mathematical model describing the relationship between the variance or relative variance of a survey estimator and its expectation. The SDR has been publishing GVF parameters for major analysis domains of interest since the early 1990s. This paper compares the 2003 SDR variance estimates derived from several GVF models. The purpose is to search for a potentially superior model than the simple linear model that has been used. The predicted variance from each model is compared with the directly estimated variance and a model is considered superior if the predicted variance is closer to the direct variance estimate. We further compare the estimated GVF parameters over time to explore the possibility of using prior estimates for current surveys.

### Stability of Jackknife Variance Estimates for Prescription Count Estimates over Time Intervals

❖ Christina Gaughan, IMS Health, 660 W. Germantown Pike, Statistics, Plymouth Meeting, PA 19462, *cgaughan@us.imshealth.com*; Heather Zuleba, IMS Health; Chris Boardman, IMS Health; Kennon Copeland, IMS Health

***Key Words:*** jackknife, survey methodology, variance estimation

In order to obtain variance estimates for product specific point estimates of retail filled prescription (Rx) counts at the national, territory, and precriber level, a jackknife methodology is utilized. The sample data used for jackknife variance estimates are for 400-some products obtained from roughly 70 data suppliers representing approximately 35,000 retail pharmacies. Because variance estimates are required for multiple time intervals (weekly, monthly, and quarterly), 14 weeks of data will be used. These 14 weeks allow for the unique opportunity to test stability of the variance estimates among product specific Rx counts over the time period. The Rx counts for the products will have varying trends over the time period. The relationship of the size and trends of the Rx counts to the stability of variance estimates will be explored.

### To Replicate (a Weight Adjustment Procedure) or Not To Replicate? An Analysis of the Variance Estimation Effects of a Shortcut Procedure Using the Stratified Jackknife

❖ Katherine Thompson, U.S. Census Bureau, 2700 Karen Drive, Chesapeake Beach, MD 20732, *katherine.j.thompson@census.gov*; Wesley Yung, Statistics Canada

*Key Words:* unit non-response, Bernoulli sample, linearization estimator

Many surveys employ weight adjustment procedures to compensate for unit non-response. These procedures can be viewed as bias corrections or as second-phase sampling adjustments (with the second phase as a Bernouilli sample of respondents from sampled units). The latter view couches the issue of unit non-response as a variance estimation problem. In this framework, we compare the effect of replicating unit non-response weight adjustment procedure on stratified jackknife variance estimates to those obtained using the corresponding "shortcut procedure" on the same data [Note: shortcut procedure variance estimates use replicate weights constructed from the full-sample unit non-response adjusted weights], both theoretically (via linearized jackknife variance estimators) and empirically (using data from the U.S. Census Bureau's Annual Capital Expenditures Survey).

### On Generalized Variance Functions

❖ Donsig Jang, Mathematica Policy Research, Inc., 600 Maryland Ave., SW, Suite 550, Washington, DC 20024, *Djang@Mathematica-Mpr.com*; Amang Sukasih, Mathematica Policy Research, Inc.; Xiaojing Lin, Mathematica Policy Research, Inc.

*Key Words:* design effect, variance approximation, complex survey, SESTAT

Generalized variance function (GVF) techniques would provide a few parameter estimates for a certain domain so that analysts can approximate variance estimates for estimates of their own interest. In this paper, we review GVF formulation and the underlying assumptions for GVF models. A typical GVF model assumes: (1) design effects are homogeneous for each GVF domain; and (2) proportions of people having attributes considered for GVF fitting are unrelated to the total number of people in the domain. GVF parameters are often estimated via least squares fitting methods. As an alternative, design effects, sample size, total domain size, and the variance of the total domain size estimate can be used to calculate GVF parameter estimates directly. Using the 2003 SESTAT data, we will compare the regression based GVFs with the design effect based GVFs.

### Generalized Variance Functions To Create Stable and Timely Variance Estimates for Prescription Count Estimates

❖ Kennon Copeland, IMS Health, 7216 Ni River Landing, Fredericksburg, VA 22407, *kcopeland@us.imshealth.com*; Christina Gaughan, IMS Health; Chris Boardman, IMS Health

*Key Words:* generalized variance function, jackknife variance estimates, prescription estimates, diagnostics

Variance estimates using jackknife methodology are created for product specific retail point estimates of filled prescription (Rx) counts at the national, territory, and precriber level. The jackknife variance estimates are created using ~70 data suppliers as the sampling strata. Due to computation time constraints and to provide stability over time periods generalized variance functions (GVF) are utilized to obtain variance estimates for the point estimates. The GVF developed for prescription estimates uses the jackknife variance estimates for ~400 products as the dependent variable and total Rx count and other product specific attributes (e.g. brand/generic) as the independent variables. Various GVFs from Wolter (1985) are considered. The diagnostic regression statistics with graphical representations for these models will be presented, as well as potential bias due to the use of GVF.

### Standard Error Estimation for County-Level Radio Listening

❖ Mandy Webb, Arbitron Inc., 9705 Patuxent Woods Drive, Columbia, MD 21046, *mandy.webb@arbitron.com*; Richard Griffiths, Arbitron Inc.

*Key Words:* standard error, generalized variance function

The County Coverage Service provided by Arbitron produces radio listening estimates at the county level. To retain MRC accreditation for this service, Arbitron seeks to find a robust methodology for estimating the precision of these radio listening estimates. Various methods for estimating standard errors are considered and compared. Arbitron started with the standard variance formula for a proportion and then included design effect adjustments that take into account the complex design and weighing procedures. Since the resulting standard error estimates will be provided to Arbitron's clients either in the County Coverage books themselves or on the Arbitron website, generalized variance functions (using the weighting, county and sample information) are also considered. Arbitron conducted an empirical study to compare the effectiveness of each of the various methods.

## 466 Studies in Air Quality and Pollution ●

Section on Statistics and the Environment, WNAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Comparing CMAQ to Observations

❖ Li Chen, The University of Chicago, 5734 S. Ellis Ave., Room 459, CISES, Chicago, IL 60637, *lichen@uchicago.edu*; Michael L. Stein, The University of Chicago

*Key Words:* aggregation, fractional bias, root normalized mean square error, variation decomposition

This talk compares observed hourly ozone concentrations to the CMAQ modeled ozone concentrations at three different spatial resolutions, 36km, 12km and 4km. Performance measures, e.g., fractional bias and root normalized mean square error, are calculated. The results show that higher resolution CMAQ model output does not necessarily provide smaller FB and RNMSE than the lower ones. Aggregation is performed to obtain new versions of lower resolution model output based on higher resolution model output. The aggregated lower resolution model output predicts better than either the unaggregated high resolution run or the low resolution run in terms of RNMSE. Variation decomposition is used to understand the statistical behavior of CMAQ model output at different resolutions. The temporal variation is well captured by CMAQ output, but spatial variation and space-time interaction are not.

### A Case Study in Estimating Percentage Detection Biases along a Recorded Ozone Profile

❖ Wendy Meiring, University of California, Santa Barbara, Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, *meiring@pstat.ucsb.edu*

*Key Words:* environmental, ozone, functional, instrument bias

We consider biases in recorded ozone partial pressures measured as functions of altitude by balloon-based instruments. Biases exist due to percentage detection errors that vary with altitude, due in part to changing instrument pump-efficiencies. Our goal is to estimate the altitude-dependent percentage detection biases. We study a sequence of ozone profile measurements over one geographical location, with each measurement profile being the entire recorded altitude-dependent ozone profile at a particular time. By incorporating other sources of ozone measurements on an integrated spatial scale, we frame estimation of the percentage detection profile as estimation in a functional linear model with a scalar response and functional explanatory variable. Simulations demonstrate improved estimation stability by incorporating physically motivated constraints on the percentage detection function.

## Statistical Conditional Simulation of a Multiresolution Numerical Air Quality Model

❖ Xiaofeng Shao, The University of Chicago, 5734 S. University Ave., Department of Statistics, Chicago, IL 60637, *shao@galton.uchicago.edu*; Michael L. Stein, The University of Chicago

***Key Words:*** conditional simulation, air quality, block bootstrap, multiscale

This paper addresses sub-grid variability, an issue that naturally arises in multiresolution numerical air quality models. A novel conditional simulation approach is proposed to produce an ensemble of high resolution runs based on the runs we have, and various criteria are used to assess whether our simulated high resolution runs capture the overall space-time variability of the original high resolution runs. The main idea of our algorithm is to apply a nonlinear filter to the high resolution runs based on the low resolution runs, then perform a time domain block bootstrap for the residuals simultaneously over space. Our algorithm works fairly well for the in-sample prediction at high resolution. A bias occurs for the out-of-sample prediction and an approach is suggested to reduce the bias.

## New Classes of Asymmetric Spatial-Temporal Covariance Models

❖ Man Sik Park, Colorado State University, 1877 Campus Delivery, Fort Collins, CO 80523-1877, *mspark@lamar.colostate.edu*

***Key Words:*** air pollution, asymmetry, nonseparability, spatial-temporal process, mat\'{e}rn covariance, spectral density function

Environmental spatial data often show complex spatial-temporal dependency structures that are difficult to model and estimate due to the lack of symmetry and other standard assumptions of the covariance function. In this study, we introduce certain types of symmetry in spatial-temporal processes: axial symmetry in time, axial symmetry in space, and diagonal symmetry in space, and propose new classes of asymmetric spatial-temporal covariance models by using spectral representations. also explain the relationship between symmetry and separability and introduce nonseparable covariance models. Finally, we apply our new classes of covariance models to estimate the spatial-temporal structure of fine Particulate Matter (PM2.5) over the northeastern region of U.S.

## Multivariate Spatio-Temporal Model for Speciated Fine Particle Matter

❖ Jungsoon Choi, North Carolina State University, 3820 Jackson Street, P219 ES King Village, Raleigh, NC 27607, *jchoi@ncsu.edu*; Montserrat Fuentes, North Carolina State University; Brian Reich, North Carolina State University

***Key Words:*** multivariate spatiotemporal processes, Bayesian inference, linear coregionalization model, air pollution, environmental statistics

Fine particle matter (PM2.5) is an atmospheric pollutant linked to serious health problems, including mortality. PM2.5 is a mixture of pollutants with five main components: sulfate, nitrate, total carbonaceous mass, ammonium, and crustal materials. These components have complex spatial-temporal dependency and cross-dependency structures. It is important to gain insight and understanding about the spatial distribution of each component of total PM2.5, and to estimate how the composition of PM2.5 might change with location and season. Here, we introduce a multivariate spatiotemporal model for speciated PM2.5 and propose a Bayesian hierarchical framework with spatiotemporally varying coefficients. A linear coregionalization model is also developed to account for dependency structures. We apply our model to speciated PM2.5 monitoring data in the U.S. for the year 2004 from EPA.

## Fast and Flexible Statistical Techniques for the Analysis of Space-Time Data with Complex Structures

Dana Draghicescu, City University of New York-Hunter College; ❖ Michael Porter, City University of New York, Graduate Center, 365 Fifth Ave., Department of Earth and Environmental Sciences, New York, NY , *mporter@gc.cuny.edu*

***Key Words:*** space-time dependence, environmental processes, model checking, visualization

This study is motivated by the growing demand for applied, interdisciplinary research. Due to the increased availability of high speed, inexpensive computing capabilities, researchers in many fields are employing complex statistical analyses. Within these fields there has been a recent recognition of the ways complex space-time dependencies in data can influence findings. Over the past decade statisticians have addressed these issues by developing flexible models for space-time stochastic processes. In this paper we discuss techniques that are statistically accurate, computationally efficient, and suitable for the exploration, visualization and model checking for data with complex structures. We consider a broad range of distributional characteristics, examine how they can be employed in the work of scientists and policy makers, and illustrate these ideas on air pollution data.

## Statistical Challenges in Comparisons of Measured Indoor and Outdoor Exposures in an Urban Setting

❖ Sorina Eftim, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Dept of Biostatistics E3039, Baltimore, MD 21205, *seftim@jhsph.edu*; Alison Geyh, Johns Hopkins Bloomberg School of Public Health; Patrick Breysse, Johns Hopkins Bloomberg School of Public Health

***Key Words:*** air pollution, environmental exposure, inhalation exposure, vehicle emissions, linear models

Black carbon, PAHs and Particle size/counts and NOx's are of public health importance as indicators of community exposures to particulate matter. Relying on time resolved indoor and outdoor measurements, the Baltimore Traffic Study was designed to evaluate the impact of urban traffic on community mobile source pollutants. Communities are potentially exposed to a complex array of traffic-related pollutants emitted from a variety of vehicle types with differing tailpipe emissions and road dust re-suspension characteristics. Environmental concentrations are likely to be highly variable in space and time due to the underlying variability in their determinants, including meteorology, traffic (volume, class, and operating condition) and source proximity. The ensuing exposure and health threat is manifested at the intersection of these environmental concentration human time activity patterns.

# 467 Quality Measures for Human Populations

Social Statistics Section, Section on Health Policy Statistics
**Wednesday, August 9, 2:00 pm–3:50 pm**

## Individuals with Disabilities: How They Impact Research

❖ Larry Featherston, University of Arkansas, 5738 W. Chattel Street, Fayetteville, AR 72704, *lfeathe@uark.edu*

***Key Words:*** disability, validity, reliability, methodology, sampling, errors

This presentation takes a critical look at research reliability and validity when studying or including individuals with disabilities in social and/or behavioral science research. With more than 54 million Americans suffering from one or more disabilities, individuals with disabilities create a unique challenge to all researchers, impacting sampling and sample representation, data collection, research conclusions, and research reliability and validity. The author will articulate ways to control for internal, external, and statistical threats to validity when studying individuals with disabilities. The author will discuss the findings of a 2005 study by Roessler, Williams, Featherston, and Featherston on the Social Security Administration's Ticket-to-Work program and the findings implications for other researchers in the behavioral sciences.

## Statistical Methodology for Longitudinal Social Network Data

❖ Anton Westveld, University of Washington, Box 354322, Seattle, WA 98102, *westveld@stat.washington.edu*; Peter Hoff, University of Washington

***Key Words:*** social networks, latent variables, structured covariances, longitudinal data, block Toeplitz

Social network data consist of measured relations occurring from interactions within a set of actors. This type of data allows for the empirical investigation of the interconnectivity of the actors, which is a cornerstone of social science theory. This paper will focus on data generated from the repeated interaction of pairs of actors (temporal dyadic data) resulting in an outcome for each actor at each time point. The network and temporal dependencies are modeled through a random effects approach resulting in a stochastic process defined by two weakly stationary covariance matrices. We will present several parameterizations of these matrices from a simple Markov structure to a very general block Toeplitz structure. A new latent variable approach to the estimation of general structured covariance matrices will also be discussed with the block Toeplitz structure presented as an example.

## Testing for Differential Responses in a Multiple Category Scale: a Case Study on Self-Rated Health among Foreign- and Native-Born Asian Americans

❖ Elena Erosheva, University of Washington, Box 354322, Seattle, WA 98195-4322, *elena@stat.washington.edu*; Emily C. Walton, University of Washington; David T. Takeuchi, University of Washington

***Key Words:*** propensity score matching, symmetry models, categorical data, self-rated health

Using data on Asian Americans from the National Latino and Asian American Study, we test whether immigrants are less likely to report the extreme ends of the 5-category self-rated health scale than their native-born counterparts. We use propensity score matching to derive groups who share similar demographic and health characteristics. Each native-born person is matched to a foreign-born of the same ethnicity by nearest available Mahalanobis metric within a caliper defined by the propensity score. Propensity score framework allows us to make descriptive comparisons of self-rated health responses by nativity status, controlling for background characteristics. We find that nativity is not associated with higher likelihood of reporting the extreme ends of the health status scale. In addition, we find no evidence of imbalances in endorsement of any particular category between the two groups.

## Quality Management at the National Center for Health Statistics (NCHS)

❖ Kenneth Harris, National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782, *kwh1@cdc.gov*

***Key Words:*** reliability, measurement, verification

NCHS is the federal agency responsible for the collection and dissemination of the nation's vital and health statistics. To carry out its mission, NCHS conducts a wide range of annual, periodic, and longitudinal sample surveys and administers the national vital statistics registration systems. Quality management at NCHS involves a variety of quality control procedures and quality enhancing activities. Over the years, NCHS has kept pace with changes and improvements in quality management. Two of the biggest processes in today's quality environment are Six Sigma and ISO 9000. This paper will describe steps being taken at NCHS to incorporate these processes into the NCHS Quality Management program.

## Achieving Clinical Satisfaction with the Desirability Function

❖ Terrence Murphy, Yale University, Program on Aging, 6 Hunting Ridge, Hamden, CT 06518, *terrence.murphy@yale.edu*

***Key Words:*** desirability, function, multiple, responses, medical, decision-making

As aging Americans increasingly take multiple medications, clinicians are faced with situations where medications that alleviate one condition may raise the risk of adverse effects or decrease quality of life. Patients also want their preferences regarding these decisions to be included in the clinical decision-making framework. We propose the use of a utility function called the desirability function to quantify these individualized medical decisions. Used in engineering and pharmacology, this function incorporates several characteristics germane to clinical decision making. These include targets for each outcome, upper and lower limits on acceptable patient response and an integration of patient preferences. We discuss the mathematical characteristics of the desirability function and demonstrate its potential utility with a clinical example.

## Temporary Help, Leased, and Contract Workers: Designing and Testing a Supplement to the Current Employment Statistic Survey

❖ Polly Phipps, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, rm 1950, Washington, DC 20212, *phipps_p@bls.gov*; Kathy Downey, Bureau of Labor Statistics; Christopher Manning, Bureau of Labor Statistics; Kirk Mueller, Bureau of Labor Statistics

***Key Words:*** establishment survey, survey design, cognitive interviews, alternative employment arrangements, flexible staffing

The CES is a monthly establishment survey producing data on employment, hours, and earnings. There are occasions when supplemental information from respondents would allow us to gain a deeper understanding of economic trends. The first undertaking in this effort is a survey of temporary help, leased, and contract workers (TLC). Use of TLC workers is a leading indicator of labor market trends, and a better understanding of industry hiring would help quantify gains and losses in payroll employment. The survey has gone through several development phases: consultation with industry experts on concepts; exploratory interviews with CES respondents to assess data availability/quality; and cognitive interviews testing the questionnaire, including respondents' understanding of survey terms. We present results from each phase, and outline quality measures to be used in evaluating survey results.

### Multilevel Structural Equation Model for Ordinal Responses

❖ Sophia Rabe-Hesketh, University of California, Berkeley,
3659 Tolman Hall, University of California, Berkeley, CA 94720,
*sophiarh@berkeley.edu*; Xiaohui Zheng, University of California,
Berkeley

***Key Words:*** multilevel model, mixed model, item response model, structural equation model, latent variable, GLLAMM

In a standard multilevel model for subjects nested in clusters, a single response variable is regressed on subject-level and cluster-level covariates and random effects represent any between-cluster heterogeneity not accommodated by the covariates. Here we extend these models to allow the response and/or covariates to be latent variables that are measured indirectly by several ordinal indicators. Item-response models for the relationship between indicators and latent variables are combined with structural models relating the latent response to the latent and observed covariates. The models are applied to the U.S. sample of the Program for International Student Assessment 2000 to investigate the relationship between the student-level latent variable 'reading ability' and the school-level latent variable 'teacher excellence', controlling for a number of observed student-level covariates.

## 468 Confidence Intervals and Hypothesis Testing

Business and Economics Statistics Section
**Wednesday, August 9, 2:00 pm–3:50 pm**

### New Tests for Joint Hypothesis of a Unit Root When There Is a Break in the Innovation Variance

❖ Amit Sen, Xavier University, 3800 Victory Parkway, ML 3212,
Cincinnati, OH 45207, *sen@xavier.edu*

***Key Words:*** unit root, innovation variance, break, Dickey Fuller f-tests

In a recent paper, Kim, Leybourne, and Newbold (2002) demonstrated the Dickey-Fuller t-statistic for the unit root null hypothesis is oversized when there is a fall in the innovation variance relatively early in the sample. Therefore, they develop t-statistics for the unit root null hypothesis using a modified GLS strategy based on the QMLE break-date estimator and the implied pre-break and post-break variance estimators. We develop F-statistics for the joint null hypothesis of a unit root when the data-generating process has a break in the innovation variance using the modified GLS strategy of Kim, Leybourne, and Newbold. We derive the asymptotic null distribution of the new statistics, tabulate their critical values, and present finite sample simulation evidence regarding their size and power.

### Easily Implemented Confidence Intervals and Hypothesis Tests for Sharpe Ratios under General Conditions

❖ J. D. Opdyke, DataMineIt, 40 Tioga Way, Suite 240, Marblehead,
MA 01945, *JDOpdyke@DataMineIt.com*

***Key Words:*** Sharpe Ratio, market performance, finance, hypothesis test, confidence interval, portfolio

For a quarter century (since Jobson and Korkie 1981), derivations of the asymptotic distribution of the Sharpe ratio that are practically useable for generating confidence intervals or for conducting one- and two-sample hypothesis tests have relied on the restrictive---and now widely refuted---assumption of normally distributed returns. This paper presents an asymptotic distribution that is valid under general conditions (stationary and ergodic returns) and consistent with that of Christie (2005), but it is far simpler and more intuitive and mathematically tractable (it can be implemented easily in a spreadsheet). Also generalized beyond the normality assumption is the small sample bias adjustment presented in Christie (2005). A thorough simulation study examines the small sample behavior of the derived one- and two-sample estimators under leptokurtosis and asymmetry.

### Parameters Estimation and Bias Corrections for Diffusion Processes

❖ Chengyong Tang, Iowa State University, 204 Snedecor Hall, Ames,
IA 50011, *yongtang@iastate.edu*; Song X. Chen, Iowa State University

***Key Words:*** diffusion, Vasicek processes, CIR processes, bias correction, jackknife, bootstrap

Various diffusion processes are popular models for different interest rates in financial applications. The existing methods of estimating the parameters of diffusion processes are subject to serious bias problem when the mean reverting of the processes is weak, which is often the case for the commonly used processes in interest rates modeling. The serious bias is potentially harmful in applications, e.g. in the pricing problems. This paper examines the performance of commonly used estimators for diffusion processes. Explicit expansion for mean and variance are given. Methods for bias correction are discussed.

### New Tests for Endogeneity in a Simultaneous Equation System with Discrete Endogenous Variable

❖ Xu Cao, University of Missouri-Rolla, 1301 High Street, A, Rolla,
MO 65401, *cao@umr.edu*; V. A. R. Samaranayake, University of
Missouri-Rolla

***Key Words:*** endogeneity, simultaneous EQN system, likelihood ratio test, Hausman test, maximum likelihood estimator

A class of simultaneous eqn models with discrete endogenous variable are considered in this paper. Likelihood ratio test for testing endogeneity has been constructed and simulated. Furthermore, two new tests based on maximum likelihood estimators are suggested ,and numerical experiment shows that these two tests have larger power than likelihood ratio test and one of them approaches UMP.In addition, we evaluate the performance of Hausman test for this class of models, and it turns out that Hausman test has less power than likelihood ratio test.

### Causality Tests in Cointegrated Systems and Temporal Aggregation of Multivariate Autoregressive Moving Average Processes

❖ Ceylan Yozgatligil, Temple University, 307 W. Side Drive, Apt. 301, Gaithersburg, MD 20878, *ceylany@temple.edu*; William W. S. Wei, Temple University

***Key Words:*** causality test, cointegrated systems, temporal aggregation, error correction representation, vector autoregressive moving average processes

Time-series data are usually sums or averages over time of data generated more frequently than the reporting interval and used routinely to test causality between variables. There are many studies on the distortion effect of temporal aggregation on causality. We develop a testing procedure based on the maximum likelihood estimation of the error correction representation to test causality in cointegrated system under temporal aggregation.

### LAD Estimation of ARFIMA-GARCH Models

❖ Wai K. Li, The University of Hong Kong, Department of Statistics and Actuarial Science, Pokfulam Road, Hong Kong, 00852 China, *hrntlwk@hku.hk*; Guodong Li, The University of Hong Kong

***Key Words:*** ARFIMA-GARCH models, diagnostic checking, LAD estimation, long memory

A least absolute deviation approach is considered to estimate fractionally autoregressive integrated moving average models with conditional heteroscedasticity. The time series generated by this model is short memory or long memory, stationary or nonstationary, depending on whether the fractional differencing parameter d ? (-1/2, 0) or (0,8),(-1/2, 1/2) or (1/2,8) respectively. Using a unified approach, the asymptotic properties of the least absolute deviation estimation are established. This article also derives the large sample distribution of residual autocorrelations and absolute residual autocorrelations and these results lead to two useful diagnostic tools for checking the adequacy of the fitted models. Some Monte Carlo experiments were reported. A real example is also included.

### A Note on the Inequality Constraints for the GARCH Models

❖ Henghsiu Tsai, Academia Sinica, 128 Academia Road, Sec. 2, Taipei, 115 Taiwan, *htsai@stat.sinica.edu.tw*; Kung-Sik Chan, The University of Iowa

***Key Words:*** absolutely monotone function, generating function, volatility

For a generalized auto-regressive conditional heteroscedastic (GARCH) model, we derive a necessary and sufficient condition for its conditional variance process to be non-negative. This condition is in terms of the generating function of the kernel defining the ARCH(8)representation of the GARCH model, which has a simple form. We discuss useful implications of this result and delineate the parametric region of stationarity and non-negative kernel for lower-order GARCH models.

## 469 Equivalence, Superiority, and Noninferiority ●

Biopharmaceutical Section

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Simultaneous Test for Superiority and Noninferiority Hypotheses in Active Controled Clinical Trials

❖ Joanne Zhang, Center for Drug Evaluation and Research, Room 5231, Building 22, 10903 New Hampshire Ave., Office of Biostatistics, Silver Spring, 20993, *zhangju@cder.fda*; Yi Tsong, U.S. Food and Drug Administration

***Key Words:*** cross-trial comparison, generalized historical control, simultaneous test, non-inferiority, superiority, active controlled clinical trials

Two-stage switching between testing for superiority (SUP) and non-inferiority (NI) has been an important statistical issue in the design and analysis of the active controlled clinical trials. Tsong and Zhang (2005) have shown that the type I error rates do not change when switching between SUP and NI with the traditional generalized historical control (GHC) approach; however, they may change when switching with the cross-trial comparison (X-trial) approach. They further proposed a simultaneous test for both hypotheses to avoid the problem. Tsong and Zhang (2005) proved the existence of the confidence interval for the simultaneous test. Through a simulation study, it is shown that the type I error rate and power between the simultaneous test and the conventional superiority test are smaller than the difference between the results of t-test and the normal asymptotic test.

### Testing Equality of Medians in Two Independent Lognormal Distributions

❖ Hongwei Wang, Merck & Co., Inc., Merck Research Lab, RY34-B370, P.O. Box 2000, Rahway, NJ 07065, *hongwei_wang@merck.com*; Arvind K. Shah, Merck & Co., Inc.

***Key Words:*** lognormal distribution, median, confidence interval, coverage, testing equality

The distribution of some biomarkers follows lognormal distribution, where median is more appropriate measure for making inference. We are testing the equality of medians through the confidence intervals (CIs) on the difference in the medians of two independent lognormally distributed variables. The performance of six CIs is evaluated using simulated as well as actual data. The first is based on log-transformation and Delta method, the next two are based on back-transforming the limits of a CI on the difference of means on a log scale, the fourth is based on the asymptotic property of the median itself, and the fifth and sixth are based on the inter-quartile range and Hodges-Lehman estimator respectively. The average width and coverage of these confidence intervals are reported for varying mean, variance, and sample size. Some recommendations based on the results are also given.

### Likelihood Ratio Tests for Equivalence Hypotheses

❖ Shun-Yi Chen, Tamkang University, Tamkang University, Department of Mathematics, Tamsui, 251 Taiwan, *sychen@mail.tku.edu.tw*; Ching-Feng Hsu, Tamkang University

***Key Words:*** clinical equivalence, average bioavailability, interval hypotheses, coverage probability, power

Likelihood ratio (LR) tests for the equivalence hypothesis and the bioequivalence hypothesis are derived. For both hypotheses testing, it is demonstrated that the LR test statistics are the same and can easily be constructed for interval hypothesis testing problems. The test is an exact procedure and always produces the nominal level for all possible mean values under the null hypothesis. As a method more powerful than the studentized range test, the test is simple and requires no more than current existing statistical tables; extra statistical tables or complicated computations are not needed. The LR test can also be applied to k>2 treatments.

### Simultaneous Testing of Noninferiority and Superiority Increases the False Discovery Rate

❖ Tie-Hua Ng, U.S. Food and Drug Administration, 1401 Rockville Pike, #200S, HFM-219, Rockville, MD 20852, *ng@cber.fda.gov*

***Key Words:*** active control, noninferiority, false discovery rate, simultaneous testing, multiplicity adjustment

It is well recognized that multiplicity adjustment is not necessary in simultaneous testing for noninferiority and superiority. However, Ng (2003; Journal of Biopharmaceutical Statistics, 13, 629-662) argued that there will be more experimental treatments that are expected to have the same effect as the active control tested for superiority in simultaneous testing than would occur if only one null hypothesis is tested, thereby increasing erroneous claims of superiority. The proportion of erroneous claims of superiority is analogous to false discovery rate. Therefore, simultaneous testing of noninferiority and superiority increases the false discovery rate of claiming superiority.

### New Tests for Null Hypotheses of Nonunity Relative Risk

❖ Kallappa Koti, U.S. Food and Drug Administration, 10903 New Hampshire Ave., Silver Springs, MD 20993-0002, *koti@cber.fda.gov*

***Key Words:*** clinical trials, active control, binary data, vaccine efficacy, screening tests

Testing for noninferiority and equivalence between an experimental and standard therapy in terms of the ratio of binomial proportions is considered. New tests based on the Fieller-Hinkley distribution of the ratio of random variables are proposed. Restricted maximum likelihood estimates of the null variances are used to derive the tests. Sample size determination is discussed. The proposed test procedures are extended to multiple tables. The tests are applied to numerical examples.

### Assessing the Superiority of a Combination Drug

Jianjun Li, Merck Research Laboratories; Steven Snapinn, Amgen Inc.; ❖ Guoyong Jiang, Cephalon, Inc., 41 Moores Road, Frazer, PA 19355, *jjiang@cephalon.com*

***Key Words:*** composite hypothesis, min test, nonmonotone test

We propose new tests for evaluating the superiority of a combination drug by modifying the critical value of the conservative min test. The critical values of these new tests are determined as continuous functions of the observed difference between the component effects. This approach offsets the conservativeness of the min test, resulting in a sizable improvement in power while maintaining the type-I error rate. Two real data examples are presented for illustration.

### To Permute or Not Permute

❖ Haiyan Xu, Johnson & Johnson Pharmaceutical R&D, 2121 Town Court, N., Lawrenceville, 08648, *hxu22@prdus.jnj.com*; Jason Hsu, The Ohio State University; Yifan Huang, H. Lee Moffitt Cancer Center & Research Institute; Violeta Calian, University of Iceland

***Key Words:*** permutation test, type I error rate, multiple testing

Permutation test is a popular technique for testing a hypothesis of no effect, when the underlying distributions are unknown. To test for difference between two populations, a permutation test might be based on the difference of the sample means in the univariate case, and the maximum of such test statistics in the multivariate case. The null distribution is then estimated by permuting the observations between the two samples. We show that if the purpose is to test for equality of means, then a permutation test can have inflated Type I error rate unless the distributions are identical. We also show that if the purpose is to test for identical marginal distributions, then a permutation test can have inflated Type I error rate unless the joint distributions are identical. Implications of these results in multiple testing of microarray data will be discussed.

## 470 Semiparametric Methods ●

Biometrics Section, Section on Nonparametric Statistics, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### The Efficiency of Multivariate Pseudo-Likelihood Estimation

❖ Park Bum Hee, Hankuk University of Foreign Studies, Department of Statistics, Yongin, 449-791 South Korea, *asckii@msn.com*; Park Heungsun, Hankuk University of Foreign Studies

***Key Words:*** GEE, multivariate pseudo-likelihood estimation, power-of-the-mean variance

Power-of-the-mean (POM) function is used widely in variance modeling procedures. Park and Cho (2005) suggested quasi-likelihood and variance function (QVF) for estimating POM in independent pest-count data. However, their approach cannot be used in correlated count data. We suggest multivariate pseudo-likelihood (MPL) estimator in the correlated POM variance structure. Simulation study illustrates the efficiency of MPL over QVF for various situations.

### Confidence Intervals Based on Non-Smooth Estimating Equations for Longitudinal Data Using Markov Chain Marginal Bootstrap

❖ Di Li, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 S. Wright St., Champaign, IL 61820, *dili@uiuc.edu*

***Key Words:*** confidence interval, estimating equations, Markov chain marginal bootstrap, longitudinal data, standard error

Statistical inference of longitudinal data based on non-smooth estimating equations can be accomplished by constructing large sample confidence intervals based on the asymptotic covariance matrix. However, this matrix often involves an unknown density function, which can make direct estimation difficult and unreliable. We introduce Markov chain marginal bootstrap (MCMB) (He and Hu 2002) as an alternative to construct confidence intervals. When there are p parameters, instead of solving a system of p-dimensional nonlinear equations at each step for classical bootstrap method, MCMB sequentially solves p one-dimensional equations, thus greatly reducing the computational difficulty, especially when p is large. Our simulation study shows favorable performance of the MCMB method in accuracy and reliability compared to the direct method of estimating the asymptotic covariance matrix.

### Hierarchical Quasi-Likelihood Approach to Bioavailability and Bioequivalence Analysis

❖ Changchun Xie, McMaster University, 65 Charlton Ave. W., Apt 202, Hamilton, ON L8P 2C2 Canada, *xie@ccc.mcmaster.ca*

***Key Words:*** bioavailability and bioequivalence, crossover data, hierarchical likelihood, hierarchical quasi-likelihood

A data set arising from a 2X2 crossover trial is analyzed for bioavailability and bioequivalence. The classical method is the mixed model. This is equivalent to the marginal likelihood approach in a normal-normal model. There are some limitations for this approach. It is not appropriate if the random effect is not normally distributed. On the other hand, there is no way to check whether the random effect is normal or not. In this paper, we introduce a hierarchical quasi-likelihood approach. Instead of assuming the random effect is normal, we make assumptions only about the mean and the variance function of the random effect. Our method is flexible to model the random effect. Since we can estimate the random effect for each individual, we can check the adequacy of the distribution assumption about the random effect. This method can also be extended to high dimensional crossover data.

### Analysis of Linear Transformation Models with Covariate Transformations

❖ Chunpeng Fan, University of Wisconsin-Madison, Department of Statistics and Department of Biostatistics and Medical Informatics, 1300 University Ave, Madison, WI 53706, *cfan@stat.wisc.edu*; Jason P. Fine, University of Wisconsin-Madison

***Key Words:*** linear transformation model, covariate transformation, nuisance parameter under alternative, semiparametric model

The linear transformation model is a natural extension of the linear regression model to permit nonparametric response transformation. There has been much work on such models without covariate transformations. However, in many applications, both response and covariate transformation may be required. We propose inferences of the regression and transformation parameters for the linear transformation model with unknown parametric covariate transformations. Since transformation parameters vanish under the null of no covariate effect, such tests are nonstandard. When covariate effects are non-zero, the estimated regression and transformation parameters are consistent and asymptotically normal. Simulation studies show that the tests and estimators perform well with realistic sample size. An application to the well known GAGurine data shows improved goodness of fit compared to earlier analyses.

### Smoothing Spline Anova Model for Bivariate Bernoulli Outcome

❖ Hyonho Chun, University of Wisconsin-Madison, 5639 Longford Terrace 104, Fitchburg, WI 53711, *chun@stat.wisc.edu*

***Key Words:*** smoothing spline ANOVA, log odds ratio, logit

The parameters of a bivariate Bernoulli density with canonical parameterization have forms which have nice interpretations - one for logit function and the other for log odds ratio. Usually, log odds ratio is estimated as a constant, but here special attention is on estimating the log odds ratio as a smooth function of covariates. Not only logit function but also log odds ratio function will be estimated in a very flexible way by using smoothing spline ANOVA model. This will reveal interesting behavior of log odds ratio function. The large sample behavior of the estimator will be investigated. This formulation will be utilized for assessing familial aggregation of a disease.

### Statistical Inference for Multivariate Outcome-Dependent Sampling Design

❖ Tsui-Shan Lu, The University of North Carolina at Chapel Hill, Department of Biostatistics, CB #7420, School of Public Health, 3105 McGavran-Greenberg Hall, Chapel Hill, NC 27599, *tlu@bios.unc.edu*; Haibo Zhou, The University of North Carolina at Chapel Hill

***Key Words:*** outcome-dependent sampling, empirical likelihood, semiparametric

An outcome-dependent sampling (ODS) (Zhou et al. 2000) scheme is a retrospective sampling scheme like the case-control study where one observes the exposure with a probability, maybe unknown, that depends on the outcome variable. Allowing the selection probability of each individual in the ODS sample to depend on the outcome can be a cost effective way to enhance study efficiency. We consider a design of ODS, where the sampling of a family depends on the aggregate of the outcomes with the family. We propose a semiparametric empirical likelihood method for such family-based ODS design. The proposed methods are semiparametric in the sense that the marginal distribution of covariates is treated as a nuisance parameter and is left unspecified. Our simulation results show that the proposed estimator provides a more efficient parameter estimate than it obtained using a simple random sample.

## 471 Inference and Models for Censored Data ●

Biometrics Section, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Statistical Analysis of Survival Data under Informative Truncation

❖ Shu-Hui Chang, National Taiwan University, Room 539, No. 17 Hsu Chow Road, Taipei, 100 Taiwan, *shuhui@ntu.edu.tw*

***Key Words:*** dependence, stationarity, truncated data

Truncated survival data arise when individuals are recruited according to a certain sampling criterion. In the random truncation model, existing statistical inferences about the survival function and regression analysis for the survival time require the key assumption of independence between the survival time and truncation time. In many situations, this key assumption of independence between the survival time and truncation time may not be satisfied. A log-rank-type method for testing the association between survival time and covariates is considered when survival time and truncation time are dependent. Examples are presented to illustrate the performance of the proposed method.

### Medical Cost Estimation under Dependent Censoring

❖ Wenqin Pan, Duke University, Outcomes/DCRI, Location: 6010, 2400 Pratt Street, Room 0311 Terrace Level, Durham, NC 27705, *wendy.pan@duke.edu*; Donglin Zeng, The University of North Carolina at Chapel Hill

***Key Words:*** cost analysis, double robust, semiparametric estimation, auxiliary covariates, efficiency, missing at random

We study the estimation of total medical cost when both dependent censoring and a large amount of auxiliary information are present. Under the missing at random assumption, we propose two working models, which can be semiparametric, to obtain condensed propensity information. The estimator of the total cost can be derived nonparametrically using this condensed information. We show that when either working model is correct, the estimator is consistent and asymptotically normal. The asymptotic variance can be consistently estimated. The small-sample performance of the proposed estimator is evaluated via a number of simulation studies and a real data application.

### A General Semiparametric Transformation Model for Survival Data

❖ Hao Liu, University of California, Davis, Division of Biostatistics, MS1C, School of Medicine, Davis, CA 95616, *ucdliu@ucdavis.edu*; Alexander Tsodikov, University of California, Davis

***Key Words:*** semiparametric model, transformation model, proportional hazard model, proportional odds model, cure model, prostate cancer

We study a general class of nonlinear transformation models for semiparametric regression for the survival data. The proposed model for the survival function separates analytically the baseline cumulative hazard function and the regression component. It includes proportional hazard model, proportional odds model, transformation model and cure model as the submodels. We develop the nonparametric maximum likelihood estimation (NPMLE) and prove the strong consistency and asymptotic normality of NPMLE estimators by the theory of empirical processes. The numerical calculation is dealt by the general framework of quasi-EM (QEM) algorithm. We illustrated the methodology by a simulation study and the analysis of a survival data for prostate cancer patients.

### Estimation of a Survival Curve with Unlinked Entry and Failure Times

❖ Yujun Wu, University of Medicine & Dentistry of New Jersey, 312 Hana Road, Edison, NJ 08817, *wuy5@umdnj.edu*; Weichung J. Shih, University of Medicine & Dentistry of New Jersey; Dirk Moore, University of Medicine & Dentistry of New Jersey

***Key Words:*** broken survival data, EM algorithm, self-consistency algorithm, study monitoring table, survival analysis

In monitoring a clinical trial or other observational study with survival endpoint, sometimes the numbers of patients entering and dying at each time point might be presented, but the connections between them were kept confidential. Hence the exact time to failure or censoring for each individual was missing. We refer to such a "study monitoring table" with missing pairing information between the entry and death times as "broken" survival data. In this paper we study the problem of estimating the survival distribution from a broken survival data set. We have developed two methods, likelihood-based estimation and self-consistency estimation, to estimate the survival curve parametrically and empirically, respectively. We use simulations to study the properties of these methods, and illustrate them with data from the STELLAR-3 trial.

### Comorbidity through the Life Span

❖ John Dixon, Florida State University, 630 W Virginia St. Apt. 312, Tallahassee, FL 32304, *dixon@stat.fsu.edu*; Eric Chicken, Florida State University; Myles Hollander, Florida State University; Dan McGee, Florida State University

***Key Words:*** survival, Framingham Heart Study

Which events occurring in early or middle ages of life are significant in terms of influencing (for example, increasing the probability of) possible consequences such as coronary heart disease in later decades? Using the Framingham Heart Study, we investigate this and related questions using covariate-based survival methods. We compare our results with those of Peeters, Mamum, Willekens and Bonneuz (2002).

### Goodness-of-Fit Tests for Left-Truncated and Right-Censored Data

❖ Yi-Ting Hwang, National Taipei University, 67 Sec 3 Ming Sheng E Road, Taipei, 103 Taiwan, *hwangyt@mail.ntpu.edu.tw*

***Key Words:*** censoring, chi-square test, random truncation

Survival data in many follow-up studies often are collected using cross-sectional sampling designs. Data of this type often are subject to left-truncation and right-censoring. The product limit estimator is the most commonly used nonparametric estimator for the variable of interest. However, under certain assumptions, it is known to be less efficient than the parametric or semiparametric estimator. A chi-square test is proposed to test the hypothesis that the truncation distribution follows a parametric family.

### A Goodness-of-Fit Test for Copula Models

❖ Antai Wang, Georgetown University, 6036 Richmond Hwy., Unit 602, Alexandria, VA 22303, *aw94@georgetown.edu*

***Key Words:*** goodness-of-fit, Archimedean copula models, survival data

I propose a goodness of fit test for Archimedean copula models for bivariate data. Based on an interesting result for the conditional distribution of $V=S(T_{1},T_{2})$ given the censoring patterns, I extend the proposed goodness-of-fit test procedure from the uncensored data to censored data. Simulation studies have shown that the test works quite well. Some illustrative examples will be presented.

## 472 Osteoporosis, Contraceptive, and Vaccine Trials ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Quantitative Risk-Benefit Assessment in the Multiple Outcomes of Raloxifene Evaluation (MORE) Trial: an Application of the Global Benefit-Risk Assessment

❖ Messan G. Amewou-Atisso, Eli Lilly and Company, Lilly Corporate Center, Drop Code 6134, Indianapolis, IN 46285, *amewou-atissomg@lilly.com*; Yili Pritchett, Abbott Laboratories

***Key Words:*** osteoporosis, selective estrogen receptor modulator (SERM), global benefit-risk assessment

The risk-benefit profile of 4-year treatment with raloxifene, a SERM indicated for the prevention and treatment of osteoporosis was assessed in the Multiple Outcomes of Raloxifene Evaluation (MORE) trial using the global benefit-risk (GBR) approach. The Women's Health Initiative (WHI) global index was applied to the MORE data that demonstrated a favorable benefit-risk profile for raloxifene compared to placebo. In the GBR approach, "benefit" was defined as the absence of events that raloxifene might reduce the incidence of, based on the mechanism of action and the results of raloxifene registration trials; "risk" was defined as occurrence of any of other serious adverse events. Statistically significant superiority of raloxifene over placebo was observed for different definitions of benefit-risk categories, and the results are consistent with those by the WHI global index.

## Quantifying the Effect of the Surrogate Marker by Information Gain

❖ Yongming Qu, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *qu_yongming@lilly.com*; Michael Case, Eli Lilly and Company

***Key Words:*** surrogate marker, Kullback-Leibler information, information gain, likelihood reduction factor

Statistical validation of a surrogate marker has been studied for more than a decade. Recently, Alonso et al. (Biometrics, 60:724-728) proposed a quantity called likelihood reduction factor (LRF) to evaluate the validity of a surrogate marker. However, as point out in the present paper, the LRF may not correctly validate a surrogate marker. Therefore, a new quantity, the proportion of information gain (PIG), based on the Kullback-Leibler information is proposed. Simulations show that under the given model assumptions, the PIG precisely reflects the role of a surrogate marker. The new approach is applied to a set of data from osteoporosis.

## Assessing Learning Effect and Nonrandom Dropout in a Contraceptive Device Trial

❖ Pai-Lien Chen, Family Health International, P.O. Box 13950, RTP, NC 27709, *pchen@fhi.org*

***Key Words:*** learning effect, nonrandom dropout, transition model

Participants in the new device group of a randomized control device trial might need time to become skilled at using the device. In addition, some participants who are not used to the new device might consequently drop out from the study. As a result, the learning effect and the nonrandom dropout can substantially influence the evaluation of treatment effects. Without considering those factors, the conclusion of the trail is vulnerable to potential bias. In this study we propose a transition model to assess the effect of learning and of nonrandom dropout from a randomized control device trial. Data from a randomized condom contraceptive trail is used to illustrate the proposed method. The impact of learning effect and nonrandom dropout of this condom trail on its surrogate endpoints and on the study conclusion regarding the primary endpoint are discussed.

## An Improved Exact Method for the Estimation and Testing of a Rate Ratio

❖ William W. B. Wang, Merck Research Laboratories, UN-A102, 785 Jolly Road, Bldg C, Blue Bell, PA 19422, *william_wang@merck.com*; Ivan Chan, Merck & Co., Inc.

***Key Words:*** exact method, Poisson distribution, rate ratio

In controlled clinical trials evaluating treatment or prevention of diseases with low incidence, drug or vaccine effects are often characterized by the ratio of two independent Poisson rates. Many methods have been proposed for the estimation and the testing of this ratio. These include the asymptotic methods by Miettinen and Nurminen (Stat. Med. 1985) and the conditional exact methods by Guess and Thomas (Epidemiol 1990) and by Martin and Austin (Epidemiol 1996). When the study is designed as case driven, the exact methods are widely used. In this presentation, we propose an improved exact method based on a more efficient statistic. Simulation studies are performed to compare the proposed exact method with other asymptotic and exact methods in terms of size, power and coverage probability. We will illustrate these methods using a real dataset from a large-scale vaccine efficacy trial.

## Utilizing Statistical Models To Predict the Duration of Protection of Vaccines

❖ Liwen Xi, Merck & Co., Inc., BL 3 2, P.O. Box 4, West Point, PA 19486, *liwen_xi@merck.com*

***Key Words:*** immunogenicity, titer, plateau, decay, bootstrap

For new vaccines to gain approval and enter the market, immunogenicity and efficacy need to be established through short-term clinical trials. The duration of the protection of a vaccine is expected to be longer than the duration of the clinical trials. Thus, long-term data that would indicate the duration of the protection are not available upon the approval of vaccines. In many instances, the biological mechanisms that lead to the decay of the titers are unknown, making it a challenge to model the long-term duration based on the short-term data. In this paper, nonlinear mixed models are utilized to predict the duration of the protection of vaccines. Three classes of functions are discussed: linear and nonlinear decay without plateau and nonlinear decay with plateau. The bootstrap method is utilized to construct confidence intervals for the mean duration of protection.

## A Statistical Framework for Quantile Equivalence Clinical Trials with Application to Pharmacokinetic Studies That Bridge from HIV-Infected Adults to Children

❖ Lixia Pei, Harvard University, Department of Biostatistics, HSPH, 655 Huntington Avenue SPH2 4th Floor, Boston, MA 02120, *lpei@hsph.harvard.edu*; Michael Hughes, Harvard University

***Key Words:*** bridging clinical trials, children, equivalence, HIV, pharmacokinetic, quantile

Many bridging trials are designed to evaluate whether a proposed dose for one population gives similar pharmacokinetic (PK) levels as an established dose in another population. For HIV bridging trials, because of the risk of viral resistance, the goal is often to determine whether the doses used in two populations result in similar percent of patients with low PK levels. For example, it may be desired to confirm that a proposed pediatric dose gives approximately 10% of children with PK levels below the 10th percentile of PK levels for the established adult dose. However, the 10th percentile for adults is often imprecisely estimated in small studies. Little attention has been given to such bridging studies. In this paper, a formal framework for the design and analysis of quantile-based bridging studies is proposed, and is used to illustrate some limitations of current trial designs.

## The Use of an Internal Unblinded Statistician with a Data-Monitoring Committee

❖ David Radley, Merck & Co., Inc., P.O. Box 4, West Point, PA 19486, *david_radley@merck.com*; Gregory Golm, Merck & Co., Inc.

***Key Words:*** data monitoring committee, unblinded statistician, internal, external

Phase III clinical trial programs often employ a Data Monitoring Committee (DMC) comprised of experts external to the trial sponsor, to ensure patient safety, oversee trial conduct and perform interim analyses. These committees are served by an unblinded statistician who generates the data summaries for review by the DMC. Many sponsors use external contract personnel to perform this role, and FDA guidance states this arrangement is best for the integrity of the trial. However, there are advantages to using a statistician internal to the sponsor, while the supposed advantages of an external statistician may be less than anticipated. The experiences of internal unblinded statisticians for two major vaccine programs will be described, along with organizational measures taken to ensure independence. We conclude that such a role is viable and is beneficial for the conduct of the trial.

## 473 Oncology Trials ●

Biopharmaceutical Section, Biometrics Section, ENAR
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Evaluating the Quality Reporting of Clinical Trials in Primary Treatment of Brain Tumors

❖ Lehana Thabane, McMaster University, Centre for Evaluation of Medicines, 105 main Street east Level P1, Hamilton, ON L8N 1G6 Canada, *thabanL@mcmaster.ca*; Rose Lai, Columbia University; Rong R. Rachel, The University of British Columbia; Michael Fraumeni, Juravinski Cancer Centre

***Key Words:*** systematic review, consort, quality of reporting, RCTS, primary treatment, brain tumors

The objectives of this study were to assess the reporting quality of randomized controlled clinical trials (RCTs) in primary treatment of brain tumors and determine predictors of the quality score. We performed a systematic review of the literature in primary treatment of brain tumors. We evaluated the quality of overall reporting based on the revised CONSORT statement. We used generalized estimating equations (GEE) to analyze the data. We retrieved 74 articles that satisfied our inclusion criteria. About 15 CONSORT methodologic items were reported in less than 50% of the trials. GEE results showed that impact factor, year of publication and sample size were associated with quality score. Conclusions: Despite improvements in general reporting quality, key methodologic items that safeguard against bias may still be poorly handled in reporting of RCTs in primary treatment of brain tumors

### Predicting Malignant Renal Lesions by Using Preoperative Color Doppler Ultrasonography: Building a Nomogram

❖ Alexia Iasonos, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd street, 3rd Floor, New York, NY 10021, *iasonosa@mskcc.org*; Ganesh V. Raj, Memorial Sloan-Kettering Cancer Center; Paul Russo, Memorial Sloan-Kettering Cancer Center

***Key Words:*** multivariate logistic regression, nomogram, ROC, cross-validation

It is important to accurately identify patients with benign renal masses who may not need surgery for their lesion. Current methods like CT and ultrasonography (US) cannot adequately differentiate between malignant and benign lesions. We recently found an association between vascular flow on CDUS and clear cell carcinomas. However, the lack of specificity limits the ability of CDUS alone to reliably predict histology. In conjunction with patient characteristics and radiographic parameters, we developed a nomogram to predict histology, using a retrospective cohort of 299 lesions. Regression techniques for multivariate models, including statistically versus clinically significant variables, will be presented. The predictive accuracy of the model was assessed by the area under the Recite model was validated internally (10-fold cross-validation), and externally for overfit adjustment.

### Statistical Properties of a Modified Accelerated Design for Phase I Cancer Clinical Trials

❖ Weili He, Merck & Co., Inc., 1129 Blvd., Westfield, NJ 07090, *weili_he@merck.com*; Jun Liu, Rutgers University; Hui Quan, sanofi-aventis

***Key Words:*** standard algorithm-based design, DLT, MTD, a+b design

In recent years, the accelerated design has been proposed as an improvement to the standard algorithm-based design (Storer 1989 and Simon et al. 1997). There may be situations under the proposed accelerated designs by Storer and Simon et al. in which the MTD may be declared with undesirable features. We propose a modified accelerated design. Lin and Shih (2001) studied several key statistical properties for the standard algorithm-based A+B design based on the exact mathematical derivation. In this paper, we derive similar key statistical properties for the accelerated design. Under certain design settings, the modified accelerated design requires fewer expected number of patients, hence less time---sometimes significantly less---to complete the study; however, it may incur a higher overall expected toxicity rate.

### Using Marginal Structural Model To Adjust for Post-Discontinuation Chemotherapy in Cancer Clinical Trials

❖ Yanping Wang, Eli Lilly and Company, Lilly Corporate Center, DC 6831, Indianapolis, IN 46285, *wang_yanping@lilly.com*; Jim Symanowski, Eli Lilly and Company

***Key Words:*** causal inference, survival analysis, cancer clinical trials

Comparison of survival between treatment arms in randomized cancer clinical trials is often complicated by post-discontinuation chemotherapy (PDC). We need to adjust for the effect of PDC so the estimated treatment arm effects have the desired causal interpretation. Methods such as the time-dependent proportional hazards model may not properly adjust for post-randomization therapy and may produce biased results. Alternatively, the marginal structural model (MSM) proposed by Robins has the potential to reduce or remove the bias. Because the results of the MSM are appropriately adjusted for potential effects of PDC, the isolated effect of experimental treatment on overall survival can be estimated with less or no bias. We applied this approach to Phase III cancer clinical trials and compared the results with those from the time-dependent proportional hazards model.

### On Dose Escalation Rules in Phase I Cancer Clinical Trials

❖ Susan Li, Centocor R&D, Inc., 965 Chesterbrook Blvd., Mail stop C42, Wayne, PA 19087, *sli19@cntus.jnj.com*

***Key Words:*** phase I cancer clinical trial, dose escalation, MTD

The primary goal of phase I cancer clinical trials is to select the appropriate dose and schedule for Phase II studies. Dose escalation rules are used in the trial to determine the maximum tolerated dose (MTD). In this presentation, a few commonly used dose escalation rules are reviewed and discussed. Rules for clinical trials of biologic treatments are recommended, they are compared with the traditional escalation rules, challenges and benefits are discussed.

### Identifying Patients with Newly Diagnosed; Histologically Proven; Untreated; Symptomatic Stage I, II, or III Myeloma Who May Benefit from Dexamethasone

❖ Keyue Ding, Queen's University, 10 Stuart Street, Kingston, ON K7L 3N6 Canada, *kding@ctg.queensu.ca*

***Key Words:*** bump-hunting method, clinical trial

Oncology Clinical Trials usually evaluate the efficacy of a new treatment by comparing the overall survival between all patients who are randomly assigned to the new treatment and those assigned to a standard/placebo treatment. But patients may react to the new treatment differently, some patients get significant benefit, some get less, while some may even be harmed by it. NCIC CTG trial of MY7, which investigate whether the combination of Melphalan and Dexamethasone (MD) is superior to the standard treatment of Melphalan and Prednisone (MP), showed that MD does not differ from MP in overall survival for study population. We undertake an exploratory analysis, using newly developed bump-hunting method, and find that male patients with ECOG performance of 0 or age of less than 60 were significantly benefit from MD, while female patients may be harmed by it.

### A Statistical Method To Integrate Independent Review and Investigator Review in Clinical Cancer Trial

❖ Xiaolong Luo, Johnson & Johnson Pharmaceutical R&D, 920 Route 202, S., PO Box 300, Raritan, NJ 08869, *xluo@prdus.jnj.com*

***Key Words:*** clinical trial, cancer, survival analysis, informative censoring, independent review, tumor assessment

In open label cancer trials, investigators are aware of treatment group assignment. This knowledge could consciously or unconsciously bias treating physician's assessment of the patient's treatment outcome. In order to avoid this bias, many phase III oncology trials are implemented with an independent review for tumor assessment. At the same time, a patient's treatment plan is driven by the protocol, leading to potential informative censoring. Furthermore, the data with informative censoring cannot be correctly analyzed and interpreted with procedures such as Kaplan Meier method and log rank test. The dilemma would be either taking investigator assessment with bias from knowing the treatment assignment or taking independent review with bias from the analysis procedures. We will illustrate this situation and solve the problem with actual clinical trials data and simulation analyses.

## 474 Network Analysis and Spatial Applications ✪

Section on Physical and Engineering Sciences
**Wednesday, August 9, 2:00 pm–3:50 pm**

### Dynamic Origin-Destination Matrix Estimation from Partially Observed Trajectories in a Large Network

❖ Jaimyoung Kwon, California State University, East Bay, Department of Statistics, Cal State East Bay, Hayward, CA 94542, *jaimyoung.kwon@csueastbay.edu*; Pravin Varaiya, University of California, Berkeley

***Key Words:*** origin-destination matrix, network, graph theory, partially observed data, space-time data

We recently developed Method of Moment and bootstrap methods for utilizing a novel data type, partially observed packet or vehicle trajectories, for making inferences about time-varying origin-destination (OD) matrices in computer or traffic networks. There are various challenges when the algorithm is applied to a large network with complicated topology. Put in graph theoretic terms, it stops working if the graph representing the network has cycles or is not uniquely traversable. In the current paper, we address those issues and present a systematic method for resolving the problem by incorporating trip pattern information in the model. Improving the estimator by smoothing over time and space will also be discussed. The new approach is applied to partially observed vehicle trajectories data from FasTrak infrastructure in a very large freeway network in San Francisco Bay Area.

### Network Tomography Problems

❖ Jiangang Fang, Rutgers University, 55 Prosper Street, New Brunswick, NJ 08901, *alexfang@stat.rutgers.edu*; Cun-Hui Zhang, Rutgers University

***Key Words:*** network tomography, traffic matrix estimation, delay distribution, deconvolution

We discuss two estimation problems based on network data. In the first problem (joint work with Yehuda Vardi and Cun-Hui Zhang), we propose ITGA, an iterative tomogravity algorithm for the estimation of network flow based on link data. We present certain experimental results to demonstrate the potential of ITGA. In the second problem (joint work with Cun-Hui Zhang), we consider nonparametric estimation of delay distribution in general multicast experiments. We study a simple estimator based on observations with small delay in certain paths, and propose an improvement of it with faster convergence rates. The problem is related to deconvolution.

### A Bayes/Empirical Bayes Approach for Service Level Network Reliability/Survivability Measure

❖ Cheng Chen, Texas A&M University, Math Department, 7MSC172, 700 University Blvd, Kingsville, TX 78363, *kfccc01@tamuk.edu*; Margaret F. Land, TeXas Environmental Studies and Analysis, LLC; Rajat Sethi, Texas A&M University-Kingsville

***Key Words:*** empirical Bayes, internet traffic, reliability

Due to the explosive IP traffic growth over the past five years, carrying real-time voice and video traffic over Internet (VOIP) over the non-real-time IPv4 network becomes the most pressing needs for telecommunication industry. This poses a great challenge for the telecom network carrier, which is further complicated by the stringent reliability/availability requirements (.99999) of traditional voice networks. Key to resolve the challenge is to accurately estimate traffic flow in the network so that a network with optimized capacity can be designed. An Empirical Bayes estimation procedure to estimate live traffic was developed. The "asymptotic" behavior (property) of the proposed estimator is also obtained. The proposed method can also be applied in diseased states where mortality/survival analysis is an important aspect of research and data analysis.

### Calibration and Prediction for Computer Experiment Output Having Qualitative and Quantitative Input Variables

❖ Gang Han, The Ohio State University, 577 Harley Drive, Apt. 7, Columbus, OH 43202, *han.191@osu.edu*; Thomas Santner, The Ohio State University; William Notz, The Ohio State University

***Key Words:*** calibration, Gaussian stochastic processes, power exponential correlation, kriging

We propose statistical models for prediction and calibration that allow both qualitative and quantitative input variables. The model allow prediction of a computer code at an untested set of qualitative and quantitative inputs as well as quantifying the uncertainty in the prediction. In the case of calibration, both the physical experiment and computer code are allowed to depend on both types of variables. A Bayesian Qualitative and Quantitative Variable (QQV) model is constructed and implemented by Markov Chain Monte Carlo methodology. This model is compared with a frequentist approach and a Bayesian independence model in several examples.

### Validity of Likelihood and Bayesian Inference for Gaussian Process Regression

❖ Bela Nagy, The University of British Columbia, Department of Statistics, 6356 Agricultural Road, Vancouver, BC V6T 1Z2 Canada, *nagy@stat.ubc.ca*; Jason Loeppky, The University of British Columbia; William J. Welch, The University of British Columbia

***Key Words:*** Gaussian process, kriging, simulation, profile likelihood, Bayesian inference, computer experiments

The traditional prediction variance formula for Gaussian processes underestimates the true variance because it doesn't incorporate the variability due to estimating the covariance parameters. Our results show that the difference can be quite substantial both with and without random error in the model, especially for small sample sizes. Corrections have been proposed, but they tend to be numerically cumbersome. Alternatives are explored and evaluated by simulations on performance and computational feasibility.

### Exploiting Spatial Information in Multivariate Calibration

❖ Brian Marx, Louisiana State University, Department of Experimental Statistics, 161 Ag Administration Bldg, Baton Rouge, LA 70803, *bmarx@lsu.edu*; Paul H. C. Eilers, Leiden University Medical Center

***Key Words:*** p-splines, tensor product, penalty, spectra, inverse regression, digit recognition

Many modern instruments deliver multi-dimensional signals, like spectra, images or frequency distributions. It often is of interest to use such data as regressors for classification or prediction, after "training" on complete observations. The problem is known as multivariate calibration (MVC), a very popular subject in the chemometric community. The result of MVC is a surface of regression coefficients. Classical methods, like principal component regression or partial least squares do not exploit the spatial characteristics of the data. We propose penalties to force neighboring regression coefficients to have similar values, using cross-validation is used to optimize the weight of the penalty. To make this scheme feasible in two or more dimensions, a tensor product basis of B-splines is introduced. Our penalized (generalized linear) regression framework allows non-normal responses.

### Representations of Spatial Surface Models

❖ James Yen, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *james.yen@nist.gov*

***Key Words:*** shape and surface estimation, uncertainty estimation, hierarchical model, mapping

The laser detection and ranging (LADAR) device, atomic force microscope (AFM), interferometer, coordinate measuring machine (CMM), and geometry measuring machine (GEMM) are all modern instruments that can measure the form of surfaces and objects. While these instruments can produce measurements with precisions ranging from nanometers to centimeters, they all face questions associated with quantifying the uncertainty of those measurements. In particular, topics may include registration and stitching of measurements into a coherent whole, modeling for uncertainty components, and visual representation of the measurements and models.

## 475 Dimension Reduction and Image Analysis

Section on Statistical Computing, Section on Statistical Graphics

**Wednesday, August 9, 2:00 pm–3:50 pm**

### Assessment of Influential Observations Using Alpha Factor Analysis

❖ Zenaida F. Mateo, University of Manitoba, Department of Statistics, 338 Machray Hall, Winnipeg, MB R3T 2N2 Canada, *zeny_mateo@umanitoba.ca*; Yutaka Tanaka, Nanzan University

***Key Words:*** alpha factor analysis, influence functions, Euclidean norms, common factors, unique variance, common factor decomposition

Alpha Factor Analysis (AFA) is one of the popular methods of factor analysis which was proposed based on the psychometric concept of generalizability. It's basic idea is to determine the common factor (fj) in such a way that they have maximum correlation with the corresponding universe common factors. The main objective in this study is to evaluate the influence of a small change of data on the values of the unique variance matrix and the common variance matrix. This was analyzed by deriving some theoretical influence functions of the common variance matrix and the unique variance matrix respectively in the common factor decomposition using perturbation theory. To assess the influential observation, some influential measures like the Euclidean norms were utilized using the proposed procedure. Some examples and comparisons will be presented to illustrate the present procedure.

### Factor Analysis for Multiattribute Ranked Data

❖ Philip L. H. Yu, The University of Hong Kong, Pokfulam Road, Dept of Statistics and Act Science, Hong Kong, 00852 China, *plhyu@hku.hk*; Wai Ming Wan, The University of Hong Kong

***Key Words:*** ranked data, factor analysis, MCEM

This paper develops factor models for ranked data in which items are ranked based on several attributes or criteria. In modeling multi-attribute ranking data, two sources of item-response dependencies have to be distinguished. Within-attribute dependence arises when items are evaluated on the same attribute and between-attribute dependence emerges when items are compared with respect to different attributes. We extend the factor model proposed by Yu, Lam and Lo (2005) for single-attribute ranked data so that it takes into account the between-attribute dependence as well. The Monte Carlo Expectation Maximization (MCEM) algorithm is used for parameter estimation. A bootstrap method is proposed for assessing the fitness of a model. Simulation studies are carried out to demonstrate the proposed estimation and goodness-of-fit methods.

### Sufficient Dimension Reduction, Regardless of (n, p) Relation

❖ Lexin Li, North Carolina State University, Department of Statistics, 2501 Founders Drive 220G, RAleigh, NC 27695-8203, *li@stat.ncsu.edu*

There has recently been a surge of interest in the analyses of large-scale, high-dimensional data problems. Examples include data mining application where the sample size n is often huge relative to the number of predictors p, and DNA microarray data where n is often much smaller than p. In this talk, we will propose a dimension reduction method that can handle both large-n-small-p and small-n-large-p problems. We demonstrate that the proposed method can effectively identify the active individual predictors, and extract the linear combinations of predictors that preserve all regression information. The method is based on an inverse regression formulation, and has been developed within the framework of sufficient dimension reduction.

### Linear Dimension Reduction in Image Analysis Using Geometrical Tools

❖ Evgenia Rubinshtein, Florida State University, 167 Crenshaw Drive, Apt 15, Tallahassee, FL 32310, *rubin@stat.fsu.edu*; Anuj Srivastava, Florida State University

***Key Words:*** image analysis, dimension reduction, optimization, linear dimension reduction

Image analysis requires dimension reduction before statistical analysis, in order to apply sophisticated procedures. A variety of criteria have been proposed: reconstruction error using PCA, class separation using FDA, non-Gaussianity using kurtosis, sparseness, mutual information, recognition of objects, and their combinations. We present geometric tools for finding linear projections that optimize a given criterion for a given data set. The main idea is to formulate a problem of optimization on Grassmann or Stiefel manifolds, and to use differential geometry of the underlying space to construct optimization algorithms. Purely deterministic updates lead to local solutions, and addition of random components allows for stochastic gradient searches that eventually lead to global solutions. We demonstrate these results using several image datasets, including natural images and facial images.

### Inferring Galaxy Morphology through Texture Analysis

❖ Kinman Au, Carnegie Mellon University, 132 Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213, *kinman@stat.cmu.edu*; Christopher Genovese, Carnegie Mellon University; Andrew Connolley, University of Pittsburgh

***Key Words:*** galaxy morphology, image analysis, texture analysis, astronomy

We give an approach to estimate galaxy morphology from digital images. In particular, our algorithm extracts textural information from galaxy images at difference scales. The multiscale information is then merged into an unified representation. By fitting a morphological model based on the textural information, we derive an quantitative and physically meaningful description of galaxy morphology. Such description will help astronomers to study galaxy evolution.

### Image Analysis Using the EM Algorithm with Stochastic Variation

❖ Xiaoxi Zhang, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, *xiaoxi@umich.edu*; Roderick J. Little, University of Michigan

***Key Words:*** expectation-maximization algorithm, stochastic approximation, Gaussian hidden Markov random fields

In this application, we use a Gaussian Hidden Markov Random Field to model the changes in brain/tumor vascular permeability induced by radiation as a noisy observation of an underlying Random Field. In the missing data scenario, the change in permeability is the observed outcome, and the hidden state label is the missing observation. The conditional expectation in the E-step requires summation over all possible realizations of the Random Field, the size of which increases exponentially. We thereby propose a stochastic approximation of the expectation. We use Louis's method to obtain the standard errors of parameter estimates for future comparison with their Bayesian posterior counterparts. Furthermore, the number of states in the Random Field can be selected using information criteria, AIC, BIC, etc. With satisfactory simulation results, we are now working on the real dataset.

### Validity Diagnostics for DTI Heterogeneity Models

❖ Meagan E. Clement, Rho, Inc., 1010 Goldmist Lane, Durham, NC 27713, *clement@email.unc.edu*; Keith E. Muller, The University of North Carolina at Chapel Hill; Guido Gerig, The University of North Carolina at Chapel Hill; Matthew Gribbin, The University of North Carolina at Chapel Hill; Joseph Piven, The University of North Carolina at Chapel Hill

***Key Words:*** diffusion tensor imaging, heterogeneity, bimodal mixture distribution

Widely used summary measures from diffusion tensor imaging (DTI) can be interpreted as statistical estimators of population properties for Gaussian stochastic processes. Despite concerns about an underlying assumption of homogeneity, one-to-one transformations of the measures have observed distributions accurately represented in terms of only two estimated parameters each. Furthermore, standard statistical methods provide diagnostic tools for checking the homogeneity assumption, as should be done in every analysis. DTI data from small regions of the brain illustrate the process. Inadvertently including both white and grey matter in the brain due to region definition generates bimodal mixture distributions. Kernel density estimation and related histogram tools allow detecting the problem.

## 476 The Practice of Statistical Consulting: Study Design and Sample Size

Section on Statistical Consulting

**Wednesday, August 9, 2:00 pm–3:50 pm**

### The Joys (and Perils) of Professional Statistical Consulting

❖ Nestor Rohowsky, IDCS, Inc., 5718 Laurel Ave., La Grange, IL 60525, *nrohowsky@idcsinc.com*

***Key Words:*** professional, consulting, communications, understanding, experiences, collaboration

Every practicing statistcian consults in one form or another, whether as an adjunct to his or her main occupation (e.g., professor, corporate staff member, or as someone whose career is to provide statistical expertise to clients in a variety of professional settings). The latter is referred to by the author as a professional statistical consultant. A professional statistical consultant is a special breed. He or she is asked to provide statistical expertise to satisfy a variety of client needs. Successful conclusion of such a project is professionally and personally satisfying. But along the way, there can be many overt and hidden perils the professional statistical consultant must learn to identify and overcome. This paper delves into the joys and perils of professional statistical consulting using the author's experiences as a guide.

## Impact of Effect Size, Sample Size, and Crossover Percent on Intention-to-Treat (ITT) Analysis: Do Subjects Need To Stay in the Group They Were Assigned?

❖ Thomas Wasser, Lehigh Valley Hospital, 5754 Loyola Street, Macungie, PA 18062, *Thomas.Wasser@lvh.com*; Christopher S. Hollenbeak, The Pennsylvania State University; Stephen Matchett, Lehigh Valley Hospital

***Key Words:*** ITT models, crossover, effect size, simulation study

Introduction: ITT models have been shown to control for subject crossover in randomized trials. But ITT methods prevent the true treatment effect from being known. Methods: Using R, a simulation study was conducted to determine the impact of Type I error given six crossover percentages (1 to 11%), four effect sizes (ES, 0.2 to 0.8 SD), and four sample sizes (50 to 300). Simulations included 1000 replications per combination. Type I error rates were computed. Results: When ES are small (< 5%) Type I rates were below 1%. When ES were larger, and crossover increased Type I rates increased linearly to a max of 4.6%. Large samples with high crossover and large ES had the highest Type I rates. Conclusions: When crossover rates are low and ES are small researchers can abandon ITT methods with little risk of additional Type I error. The benefit is that the true treatment effect can be determined.

## Comparison of Effect Size, Power, and Type I Error Rate in Simulated Efficacy and Effectiveness Trials

❖ Mary Z. Mays, Arizona State University, 4747 E. Olney Ave., Higley, AZ 85236, *mary.mays@asu.edu*; Jan Jirsak, University of Arizona

***Key Words:*** efficacy, effectiveness, effect size, power

Evidence-based practice in healthcare has increased interest in the types of evidence produced by efficacy and effectiveness trials. Although there is consensus about the importance of both types of trials, guidelines for reporting trials favor efficacy analysis. A simulated experiment was created by randomly sampling 45 subjects from a population of 10,000. Data from each of 20 replications were analyzed with both efficacy and effectiveness analyses. Effect sizes ranged from 0 to 0.19 in the efficacy analyses and from 0.23 to 0.64 in the effectiveness analyses. Power ranged from 0.05 to 0.87 in the efficacy analyses and from 0.82 to 1.00 in the effectiveness analyses. P values were less than 0.05 in five of 20 efficacy analyses and in all 20 effectiveness analyses. These results suggest that statistical analyses in small-scale efficacy trials should include mediators.

## Reporting Significant Results for a Large Sample Study

❖ Gloria Caldito, LSU Health Sciences Center, 1501 Kings Highway, PO Box 33932, Shreveport, LA 71130, *gcaldi@lsuhsc.edu*

***Key Words:*** effect size, statistical significance, substantive importance, confidence interval, confounding factors, p-value

When a large sample is used to compare two groups, small group effects can be detected as statistically significant. An index for the group effect referred to as the effect size is selected to reflect its substantive importance. Confidence intervals are the appropriate values to report. Whether the small but statistically significant effect sizes also have practical significance depends on their interpretability in the context of the study. While they cannot be attributed to chance, the potential of their being due to confounding factors exists; thus the need to adjust the effect sizes for such factors. Reporting both the unadjusted and adjusted effect sizes will give the reader an idea of how much confounding there is in the data. A case-control study of half a million US veterans to determine the protective effect of statins against different cancers is given as an example.

## Uncontrolled Variation in Multistage Experiments

❖ T. B. Bailey, Iowa State University, Department of Statistics, 121 Snedecor Hall, Ames, IA 50011-1210, *tbbailey@iastate.edu*

Experiments requiring several stages can be expected to have experimental errors associated with the stages. Examples of multi-stage experiments are given to illustrate the nature of uncontrolled variation. Control of this variation requires an understanding of the nature of the experimental errors in such experiments. The important roles of randomization and method of analysis in multi-stage experiments are discussed.

## Dose Ranging Studies in Acupuncture, Manipulative Therapy, and Mind Body Research

❖ Laura L. Johnson, National Center for Complementary and Alternative Medicine, 6707 Democracy Blvd., Suite 401, Bethesda, MD 20892-5475, *johnslau@mail.nih.gov*; Catherine Stoney, National Center for Complementary and Alternative Medicine; Partap Khalsa, National Center for Complementary and Alternative Medicine

***Key Words:*** phase I, phase II, dose ranging, clinical trials, complementary and alternative medicine, integrative medicine, integrative medicine

Dose-ranging studies are needed prior to Phase III trials of botanical and other biologically based complementary and alternative medicine (CAM) therapies. However, dose-ranging is equally important for other CAM therapies such as mind-body interventions, chiropractic treatments, massage, and acupuncture. Investigators need to identify the optimum number, frequency, length of sessions, and similar dose-related concerns to achieve efficacy and sustainability while minimizing negative consequences such as poor recruitment or patient non-adherence. Some therapies begin dose ranging studies with the commonly recommended dose and escalate. Choice of outcome, patient population, control group, and blinding are important study design decisions. Use of a suboptimal dose that is safe but ineffective does not serve the larger goals of CAM or integrative medicine community.

## A General Serial Gatekeeping Procedure To Control Studywise Error Rate

❖ Fang Xie, Cephalon, Inc., 41 Moores Road, PO Box 4011, Fazer, PA 19355, *fxie@cephalon.com*; Chung-Kuei Chang, Cephalon, Inc.; Guoyong Jiang, Cephalon, Inc.

***Key Words:*** parallel gatekeeping procedure, serial gatekeeping procedure, studywise error rate, strong control of familywise error rate

Parallel gatekeeping strategies have been proposed to control the studywise error rate for multiple primary and secondary endpoints without rejecting all primary endpoints. However, the studywise error rate has to be shared among the primary and secondary endpoints. Sample size calculation may become an issue due to inadequate information of the secondary endpoints. In addition, allocation of the type I error to each subset hypothesis could be computationally cumbersome. We show that if the overall type I error for testing multiple primary and secondary endpoints is respectively controlled at a significance level by a testing procedure, a serial gatekeeping procedure controls the studywise error rate at the same significance level. We also propose a 3-tier hierarchical testing procedure that can simplify the design of a study and is appealing for regulatory submission.

# 477 Dimension Reduction Methods

Section on Nonparametric Statistics

**Wednesday, August 9, 2:00 pm–3:50 pm**

## Sliced Inverse Moment Regression Using Weighted Chi-Squared Tests for Dimension Reduction

❖ Jie Yang, The University of Chicago, 5608 S. Harper Ave., Chicago, IL 60637, *jyang@galton.uchicago.edu*; Zhishen Ye, Eli Lilly and Company

***Key Words:*** dimension reduction in regression, weighted chi-squared test, SIR, PHD, save, SIMR

We propose a new class of dimension reduction methods using the first two inverse moments, called Sliced Inverse Moment Regression (SIMR). We develop corresponding weighted chi-squared tests for the dimension of the regression. Basically, SIMR are linear combinations of Sliced Inverse Regression (SIR) and a new method using candidate matrix $M_{zz'|y}$, which is designed to recover the entire inverse second moment subspace. Theoretically, SIMR, as well as Sliced Average Variance Estimate (SAVE), are more capable of recovering the complete central dimension reduction subspace than SIR and Principle Hessian Directions (pHd). Therefore it can substitute for SIR, pHd, SAVE or any linear combination of them at a theoretical level. Simulation study shows that SIMR using the weighted chi-squared test may have consistently greater power than SIR, pHd, and SAVE.

## Selecting Tuning Parameters in Dimension Reduction Methods in Regression

❖ Peng Zeng, Auburn University, 1449 Richland Road, Apt. 2Q, Auburn, AL 36832, *zengpen@auburn.edu*

***Key Words:*** dimension reduction, central subspace, nonparametric regression

Sufficient dimension reduction in regression is intended to identify a subspace of predictors such that all information of predictors about response is preserved in the subspace. The minimal sufficient dimension reduction subspace is referred to as the central subspace. Many methods have been proposed to estimate the central subspace. Among them the Fourier method proposed by Zhu and Zeng (2006) can guarantee that the central subspace can be exhaustively estimated. In this talk, we discuss how to optimally select the tuning parameters that are involved in the estimation procedure of the Fourier method. Similar idea can also been applied to select tuning parameters in other dimension reduction methods. Simulation examples and real datasets are used to demonstrate the performance.

## Projection-Directed Nonparametric Omnibus Test for the Multivariate Multisample Problem

❖ Xiaobin Yuan, St. Jude Children's Research Hospital, Department of Biostatistics, 332 North Lauderdale Street MS723, Memphis, TN 38105, *yxb88@yahoo.com*; Cheng Cheng, St. Jude Children's Research Hospital

***Key Words:*** projection, nonparametric omnibus test, principal component directions

A class of simple tests combining projection and univariate nonparametric tests is proposed for multisample multivariate problem. The direction for projection is based on principle components of the pooled sample. For the projected data at each direction, we can choose proper univariate tests for different alternatives. For the general multivariate problem, we choose the contingency table method (Massey's test), which has good rejection power for different alternatives and is easy to implement. Our proposed tests are translation and rotation invariant, have little asymptotic efficiency loss for testing multivariate normal means compared to the optimal T-square test, and possess competitive power in a range of alternatives. Comparisons of empirical powers between the projection method and other proposed methods are given.

## Aggregation of Nonparametric Estimators for Volatility Matrix

❖ Yingying Fan, Princeton University, Department of ORFE, Princeton, NJ 08544, *yingying@princeton.edu*

***Key Words:*** aggregation, nonparametric function estimation, volatility matrix of diffusion, factor, local time, affine model

An aggregated method of nonparametric estimators based on time and state domain estimators is proposed. To attenuate the curse of dimensionality, we propose a factor modeling strategy. We first investigate the asymptotic behaviors of the volatility matrix estimators in the time and state domains. The asymptotic normality is separately established for them. They are asymptotically independent. Hence, they can be combined to improve the efficiency of the estimated volatility matrix. The optimal dynamic weights are derived and it is shown that the aggregated estimator uniformly dominates the estimators using time or state domain smoothing alone. A simulation study, based on an essentially affine model, is conducted and it demonstrates convincingly that the new procedure outperforms both time and state domain estimators. Empirical studies endorse further the advantages of our method.

## Dimensionality Reduction of High-Dimensional Tables

❖ Siamak Noorbaloochi, VAMC, University of Minnesota, One Veterans Drive, CCDOR 152 2E, Minneapolis, MN 55417, *siamak@stat.umn.edu*; David Nelson, VAMC, University of Minnesota; Joe Grill, VAMC, Minneapolis

***Key Words:*** collapsibility, dimensionality reduction

Collapsibility is defined as equivalence of inferences based on a marginal table and its corresponding joint model. The marginalization is defined with respect to a set of variables. We present a method to extend the notion of marginalization with respect to partitioning the cells of the joint model such that collapsibility remains valid. Application of the method in high-dimensional tables with multinomial counts is investigated and a method for estimating the cell probabilities for sparse tables is introduced.

## Statistical Inference of Distributions on Manifold

❖ Wanli Min, IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, *wanlimin@us.ibm.com*

***Key Words:*** non-parametric, kernel, dimension reduction

It becomes increasingly common that large dataset of seemingly high dimension can be analyzed through reduced representation on an underlying manifold. This paper considers two related problems, namely, estimation of distribution density from which the random sample are drawn, and construction of the lower dimensional space representation upon which further statistical analysis can be performed . We propose an solution and establish asymptotic properties. An application to digital image analysis will be included.

# 478 Applied Bayesian Modeling

●

Section on Bayesian Statistical Science
**Wednesday, August 9, 2:00 pm–3:50 pm**

## Tree-Based and Bayesian Modeling of Food Web Collapse in the Permian Mass Extinction

❖ Steve C. Wang, Swarthmore College, Department of Math and Statistics, 500 College Ave, Swarthmore, PA 19081, *scwang@swarthmore.edu*; Peter D. Roopnarine, California Academy of Sciences; Kenneth D. Angielczyk, University of Bristol

***Key Words:*** paleontology, CART, MCMC, simulation-based inference, forward model, calibration

The Permian mass extinction was the most severe extinction event in the history of life, with up to 90% of species killed. Many causes have been proposed (asteroid impact, volcanism, etc.), but none is supported by incontrovertible evidence. To infer the true cause, it is important to determine the mode of extinction. E.g., did an initial extinction at the lowest level of the food web (plants) lead to the extinction of herbivores and then carnivores - a bottom-up mode of extinction? Or did animals at higher levels of the food web die first, causing extinctions at lower levels - a top-down mode? Or were all levels affected equally - a uniform mode? Here we use a stochastic forward model to simulate the effect of these modes of extinctions. Using simulated data from this model, we then attempt to infer the mode of extinction using (1) classification trees and (2) a Bayesian MCMC model.

## Bayesian Calibration Models for Obsidian Hydration Dating

❖ Andrew Schaffner, California Polytechnic State University, San Luis Obispo, Statistics Department, San Luis Obispo, CA 93407-0405, *aschaffn@calpoly.edu*

***Key Words:*** Bayesian methods, calibration, archaeology, dating

Archeological artifacts made of obsidian (e.g., projectile points) are difficult to date as their chemical composition prohibits traditional radiocarbon methods. Current hydration dating methods estimate artifact age using inverse regression techniques and a physical diffusion model. These models often violate common regression requirements and include a great deal of measurement error biasing the regression estimates. Calibration methods also have been used, but the estimation of predictive errors is non-trivial and often ignored. Instead, we demonstrate how hierarchical Bayesian models may be used to produce superior estimates of the age of obsidian artifacts including estimates of uncertainty. In addition, our models account for many known sources of error in the data used to fit the model as well as data distributions commonly associated with practical archeological data.

## Bayesian Semiparametric Analysis for a Single-Item Maintenance Optimization

❖ Elmira Popova, The University of Texas at Austin, 1 University Station, C2200, Austin, TX 78703, *elmira@mail.utexas.edu*; Paul Damien, The University of Texas at Austin; Timothy Hanson, University of Minnesota

***Key Words:*** Bayesian, nonparametric, optimization, maintenance

We address the problem of a finite horizon single item maintenance optimization structured as a combination of preventive and corrective maintenance in a nuclear power plant environment. We present Bayesian semiparametric models to estimate the failure time distribution and costs involved. The objective function for the optimization is the expected total cost of maintenance over the pre-defined finite time horizon. We show the effectiveness of our approach using real data from the South Texas Project Nuclear Operating Company (STPNOC) located in Bay City, Texas.

## Bayesian Modeling of the Effect of Four-to-Three-Lane Conversion on the Number of Crashes and Crash Rates for Iowa Roads

❖ Wen Li, Iowa State University, 112 Snedecor Hall, Ames, IA 50010, *shirley@iastate.edu*; Alicia Carriquiry, Iowa State University

***Key Words:*** ADT, crash frequency, crash rate, deviance

The goal of this study is to model crash data collected by the Iowa Department of Transportation. Six hierarchical Poisson regression models are used to fit the data, and deviance of each model is calculated to assess which is best. The best-fitting model allows each site to have its own set of parameters to account for conversion and time in the log link mean function. Moreover, using the best model, we estimate that the average reduction in crash rate at treatment sites after conversion is about 33.1% (compared to about 7.8% at control sites) for observed traffic volumes. When expressing rate as crashes per 1 million vehicles, the estimated reduction is 45.1% at treatment sites compared to 24.5% at control sites. As we expect, conversion is significantly associated to a reduction in crash rate.

## Bayesian Procrustes Analysis

Athanasios Micheas, University of Missouri-Columbia; ❖ Yuqiang Peng, University of Missouri-Columbia, Department of Statistics, 146 Middlebush Hall, Columbia, MO 65211, *peng_yq@yahoo.com*

***Key Words:*** Bayesian computation, complex normal distribution, Procrustes analysis, similarity transformations

We introduce a Bayesian framework upon which Procrustes analysis maybe conducted. We model shapes of objects in two dimensions and develop Bayesian methods for estimation of the parameters in the similarity transformations and compare with the classic Procrustes fit. We also discuss the Bayesian methodology for a full Procrustes analysis, where a population of shapes is considered and we desire estimation of mean shape. The Bayesian approach allows us to compute point estimates and credible sets for the Procrustes fit and the Full Procrustes superimposition. The method is illustrated through an example from hydrology, where shapes of storm systems are created from radar images. We then employ precipitation forecasts and the actual truth at specific times, in order to perform verification of the forecasting method using Bayesian Procrustes methods.

## Statistical Analysis of Single-Unit Firing-Rate

❖ Sam Behseta, California State University, 9001 Stockdale Highway, Department of Mathematics, Bakersfield, CA 93311, *sbehseta@csub.edu*; Robert E. Kass, Carnegie Mellon University

***Key Words:*** Bayesian functional data analysis, bars, multiple curve fitting, bootstrap, analysis of neuronal data, Gaussian filter

In this work, we consider the problem of comparing trial-averaged firing-rate functions across multiple experimental conditions. We are interested in comparisons within neurons and among populations of individually recorded neurons. This is a natural extension to our previous work (Behseta and Kass 2005), in which we developed methodologies for comparing two functions. We propose a series of likelihood ratio--based tests that may be used to perform such comparisons either pointwise or globally over the entire experimental time. An extended simulation study of power demonstrates the strength of these tests, even for moderate sample sizes. Finally, we implement these tests on a group of neurons recorded from the visual cortex of a monkey's brain.

# 479 Nonresponse Bias and Other Estimation Challenges

Section on Government Statistics

**Wednesday, August 9, 2:00 pm–3:50 pm**

## Nonresponse Bias in the Omnibus Household Survey

❖ Promod Chandhok, Bureau of Transportation Statistics, 8700 Ross Street, RITA, Bowie, MD 20720, *promod.chandhok@dot.gov*

***Key Words:*** travel surveys, bias, response error

Nonresponse bias of key variables in the Omnibus survey is estimated using data from other sources. Response bias in the survey is also discussed.

## Nonresponse Bias of Time-Use Measures' Interrelationships

❖ John Dixon, Bureau of Labor Statistics, 10308 Meredith Ave., Kensington, MD 20895, *dixon_j@bls.gov*

***Key Words:*** survey, nonresponse, time use

The American Time Use Survey measures the amount of time spent on many different types of activities. Other studies have examined the impact of nonresponse on the estimates of time spent in those activities. This study examines the potential bias in the estimates of the inter-relationships between categories. Do those who don't respond differ in the trade-offs they make in their time? The effect on the trade-offs due to hours worked and hours spent in child care are of particular interest.

## Using Survival Analysis To Predict Sample Retention Rates

❖ Andy Sadler, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3655, Washington, DC 20212, *sadler.andy@bls.gov*

***Key Words:*** price index, retention rates, quote allocation, life-table estimate, survivor function, censored observations

The International Price Program (IPP) publishes monthly price indexes which measure the average change over time in prices of internationally traded products and services. IPP's sampling methodology supports its published merchandise price indexes by attempting to maintain a minimum number of items, within each published stratum, over the stratum's two year life cycle. Stratum retention rates -- formerly calculated by using an averaging process -- are used to set minimum item allocations. This paper uses the Life-Table estimate of the survivor function applied to censored data to model retention rates. Our conclusions are: (1) The model produces a markedly superior fit to that produced from using simple averages; (2) The model is un-affected when samples are rotated; (3) The model allows us to use the Lower Confidence Limit which gives conservative estimates of the survival rates.

## Comparison of the Income Items from the CPS and Census 2000

❖ Bruce H. Webster, Jr., U.S. Census Bureau, 9348 Cherry Hill Road, Apartment 104, College Park, MD 20740, *bruce.h.webster@census.gov*

***Key Words:*** income, response errors, Census 2000, CPS ASEC

The Census Bureau conducts several surveys that collect income data. Each of these surveys gives slightly different results. This paper compares income responses given by the same respondent for the same timeframe to two separate surveys, the Census 2000 long form and the 2000 Current Population Survey Annual Social and Economic supplement (ASEC), focusing on those who gave different results. Both Census 2000 and the 2000 ASEC collected 1999 calendar year data. To evaluate the sources of these differences, the Census Bureau created a dataset that contains respondents that were in both surveys. By comparing fully reported responses in each survey, this paper tries to isolate the reported differences and provide possible reasons why those differences occurred.

## Multiple-Record Applicants in the Analysis of Hiring Disparity

❖ Charles McGhee, U.S. Department of Labor, 7600 Oak Knoll Drive, North Richland Hills, TX 76180, *mcghee.charles@dol.gov*; Marika Litras, Office of Federal Contract Compliance Programs; Michael Sinclair, Office of Federal Contract Compliance Programs

***Key Words:*** hiring discrimination, multiple records, regression assumptions

The Office of Federal Contract Compliance Programs is responsible for enforcing provisions that prohibit employment discrimination by federal contractors and subcontractors. A significant challenge in analyzing hiring disparity is developing statistical models from data that contain multiple-record applicants (MRAs). This issue is central to hiring evaluations where jobs are seasonal or have high turnover. How multiple-record applicants are analyzed, moreover, has important implications for defining applicants and meeting independence assumptions of logistic regression. This research outlines various settings in which MRAs are common, proposes opportunity- and person-based models for dealing with MRAs, and covers the strengths and weaknesses of these approaches. Case studies in low and high MRA settings are presented.

## The Use of the Peters-Belson Method in Hiring Discrimination Assessments

Michael Sinclair, Office of Federal Contract Compliance Programs; ❖ Shirong Leu, U.S. Department of Labor, Washington, DC , *shiro38@msn.com*; Arline Easley, U.S. Department of Labor

***Key Words:*** hiring discrimination, Peters-Belson, jackknife variance est.

The Peters-Belson method was developed to examine wage discrimination using linear regression analyses. In application, one conducts a regression analysis on the favored class and applies it to the nonfavored class to identify a disparity between the actual and predicted values. Recently, the method was extended to examine health care disparities and other forms of discrimination for binary outcomes via logistic regression. In this paper, we will examine the general properties of hiring discrimination evaluations as compared to a standard regression analysis as related to the applicant pool sizes and differences in the traits for the favored and nonfavored class members. We also will discuss philosophical issues surrounding the use of this approach relative to a standard regression analysis and the methodology for applying a jackknife variance estimator.

### Estimating Missing Prices in Producer Price Index

❖ Onimissi Sheidu, Bureau of Labor Statistics, 2 Massachusetts Ave NE, Room 3650, Washington, DC 20212 United States, *sheidu_o@bls.gov*

***Key Words:*** producer price index, weighted relatives, aggregate cell relatives, missing prices, absolute difference, simulation

The Producer Price Index (PPI) is a monthly estimate of average changes in prices received by domestic producers of goods and services in all stages of processing. Each month the PPI requests data for more than 100,000 price quotes. For data not received, the PPI must estimate a value. In this paper, we investigate whether there is an advantage in terms of improving the accuracy of the estimates by using only the weighted relatives of items in a cell with similar products to estimate an item's missing price, or to use higher aggregate cell relatives comprised of different product cells. We discovered the proposed method of estimating missing prices using detailed product cell relatives is superior only in cases where there are enough items, usually greater than 10, in a cell.

## 480 Stochastic Process and Mixture Models

IMS

**Wednesday, August 9, 2:00 pm–3:50 pm**

### A Class of Probability Measures on the Simplex

❖ Zach Dietz, Tulane University, 5420 Garfield Street, New Orleans, LA 70115, *zdietz@math.tulane.edu*

***Key Words:*** Dirichlet, reinforcement

The Dirichlet distribution plays a central role in controlling the behavior of certain Generalized Polya Urn schemes, and RAM (residual allocation model) phenomenon. We will demonstrate how the Dirichlet distribution may be elicited from a Markovian reinforcement scheme, and then generalize the procedure to introduce a class of distributions on the simplex.

### The Bahadur representation for sample quantiles under weak dependence

❖ Shuxia Sun, Wright State University, Department of Mathematics and Statistics, 3640 Colonel Glenn Hwy, Dayton, OH 45435, *shuxia.sun@wright.edu*

***Key Words:*** Bahadur representation, sample quantile, strongly mixing

In this paper, we give a Bahadur representation of sample quantiles based on strongly mixing random variables under polynomially decaying rate. Our results extend Theorems 1 and 2 of Yoshihara [1995. The Bahadur representation of sample quantile for sequences of strongly mixing random variables. Statist. Probab. Lett. 24, 299-304].

### Long Time Asymptotics for Constrained Diffusions in Polyhedral Domains

❖ Chihoon Lee, The University of North Carolina at Chapel Hill, 210 Smith Building, CB 3260, UNC Chapel Hill, Chapel Hill, NC 27599, *chlee@email.unc.edu*; Amarjit Budhiraja, The University of North Carolina at Chapel Hill

***Key Words:*** semi-Martingale reflecting Brownian motion, constrained diffusions, geometric ergodicity

We study long time asymptotic properties of constrained diffusions that arise in the heavy traffic analysis of multiclass queueing networks. We first consider the classical diffusion model with constant coefficients namely a semimartingale reflecting Brownian motion (SRBM), in a d-dimensional positive orthant. Under a natural condition on stability of a related deterministic dynamical system Dupuis and Williams (1994) showed that an SRBM is ergodic. We strengthen this result by establishing geometric ergodicity for the process. As consequences of geometric ergodicity we obtain finiteness of the moment generating function of the invariant measure in a neighborhood of zero, uniform time estimates on moments of the process, and functional central limit results. Similar long time properties are obtained for a broad family of constrained diffusion models with state dependent coefficients.

### The Mixture Labeling Problem: a Frequentist View

❖ Daeyoung Kim, The Pennsylvania State University, 326 Thomas Building, Department of Statistics, University Park, PA 16802, *dzk123@psu.edu*; Bruce G. Lindsay, The Pennsylvania State University

***Key Words:*** finite mixture model, labelling problem, asymptotic identifiability

This talk describes the labeling issue in the parametric finite mixture model in terms of a frequentist view and examines several relabeling methods which could be used in a simulation study. Although the parameter labels are not identifiable in the strict sense, there is a form of asymptotic identifiability which can provide reasonable answers when components densities are well separated, relative to the sample size. We compare several techniques for hard labelling (assigning specific labels) and soft labeling (probabilistic assignment) to see their effect on simulation studies of the parameter estimators' sampling distributions.

### Stability and Tail Properties of Nonlinear Stochastic Recursions with Application to Nonlinear AR-GARCH Models

❖ Daren B. H. Cline, Texas A&M University, Department of Statistics, College Station, TX 77843-3143, *dcline@stat.tamu.edu*

***Key Words:*** GARCH, ergodicity, nonlinear stochastic recursion, regular variation

We characterize the Lyapounov exponent and ergodicity of nonlinear stochastic recursion models, including nonlinear AR-GARCH models, in terms of an easily defined, uniformly ergodic process known as the collapsed process. Properties of this latter process also can determine the tail properties, including regular variation, and existence of moments for the stochastic recursion when it is stationary.

### Noncommutative Stochastic Convergence of the Bounded Besicovitch Sequence

❖ Larisa Shwartz, IBM, 19 Skyline Drive, Hawthorne, NY 10532, *lshwart@us.ibm.com*; Genady Grabarnik, IBM T. J. Watson Research Center

***Key Words:*** limit theorems, non-commutative probability, operator algebras

Non commutative probability becomes one of the most developing areas of probability and statistics. It investigates how probabilistic notions

behave in the new environment, environment that allows describing quantum mechanical effects. Different limit theorem where established in this settings. The settings also allow building meaningful stochastic calculus, etc. The paper is concerned with stochastic convergence results, in particularly with stochastic convergence of subsequence averages. We utilize stochastic Banach principle to get stochastic convergence of Chesaro averages over uniform subsequences and averages of bounded Besicovitch sequences.

# 481 Contributed Posters

General Methodology, Social Statistics Section, Section on Survey Research Methods, Biopharmaceutical Section, Section on Physical and Engineering Sciences, Section on Statistics and the Environment, Section on Statistics and Marketing, Section on Statistical Consulting, Biometrics Section, Section on Statistical Computing, Section on Statistics in Sports, Section on Statistics in Epidemiology

**Wednesday, August 9, 2:00 pm–3:50 pm**

## A Multivariate Statistical Analysis of Female Empowerment

❖ Janelle Jones, SUMSRI, 4208 Woodstock Drive, Lorain, OH 44053, *janellecj@hotmail.com*; Adrianne Demski, SUMSRI

*Key Words:* empowerment, factor analysis, discriminant analysis

As women of the world struggle for equality, there is a need for ways of measuring progress. We explore the empowerment of women using multivariate statistical techniques, such as factor analysis and discriminant analysis. We hope to classify countries into two populations, one where women are empowered and the other where women are not. We simplify this process by reducing the dimensionality of the data from 13 variables to a smaller collection of underlying factors. "We must remember that unless and until women are given their rightful place, no society or country can progress." --IK Gujral, Prime Minister of India

## Examination of Structure Coefficient Interpretation in Descriptive Discriminant Analysis: the Three-Group Case

❖ Mercedes Schneider, Ball State University, Department of Educational Psychology, TC 524, Muncie, IN 47306, *mschneider@bsu.edu*

*Key Words:* discriminant analysis, structure coefficient interpretation, Monte Carlo simulation, MANOVA post hoc tests

This simulation examined the usefulness of determining a structure coefficient (SC) value >= .3 as identifying a meaningful continuous variable in descriptive discriminant analysis (DDA) when DDA is used as a post hoc for a significant MANOVA. Unequal group sizes and inequality of variance were considered for the three-group case. Specifically, sample sizes included n = 10, 25, 50, 75, and 100, and unequal variance involved at least one half of the simulations violating a test for homogeneity of variance. General results indicated support for the rule that an SC value >= .3 identifies a continuous variable useful for group separation.

## Computer Literacy of Adolescents in Grades 9 to 12: an Exploratory Study

❖ Matthew Sink, Shoreline Christian High School, 16521 22nd Ave., NE, Shoreline, WA 98155, *csink@spu.edu*; Christopher A. Sink, Seattle Pacific University

*Key Words:* computer, literacy, survey research, high school, students

We report on a study examining whether there were gender differences in computer literacy among high school (HS) students. The US Dept. of Education studied the rates of computer use in children through the 12th grade reporting that 91% of each gender used the computer on a regular basis. It did not look at computer literacy. Our aim was to determine if there was a difference in computer literacy between HS girls and boys. We created a computer literacy survey measuring 6 categories of skills (i.e., word processing skills, Internet skills, instant messaging, downloading music, gaming, and computer programming). Eighty four HS students (43 F, 41 M) participated. Gender usage frequency did not differ. Boys had on average higher scores across each of the 6 literacy categories with the largest gap in programming. Findings suggest that girls do not possess the same literacy skills as boys.

## Partial Least Squares Regression and Its Application in Drug Discovery

❖ Jingjing Chen, Merck & Co., Inc., 531 Williamsburg Way, King of Prussia, 19406, *jingjing_chen@merck.com*

*Key Words:* partial least squares, PLSR, Selwood, drug discovery

Partial least squares is a variant of a family of least squares models of correlation matrices that has been a popular modeling, regression, discrimination, and classification technique in industrial applications. Partial least squares regression (PLSR) reduces the dimension of the regressors by extracting scores and regressing the scores on the response variables. The purpose of this paper is to illustrate using partial least squares in the context of a regression problem in the drug discovery field. This paper reviews the key concepts and major algorithms of PLSR and discusses and compares different approaches dealing with high-dimensional regressors, such as principal component regression (PCR), reduced rank regression (RRR), canonical correlation, and multiple factor analysis. The Selwood Data was used as an example to demonstrate the application of PLSR in the drug discovery field.

## Quasi-Probability Distributions Based on the Lagrange Expansions

❖ Shubiao Li, Central Michigan University, Department of Mathematics, Mount Pleasant, MI 48859, *li2s@cmich.edu*; Carl Lee, Central Michigan University; Felix Famoye, Central Michigan University

*Key Words:* Lagrange expansion, quasi probability distributions, mixture distribution

In this paper, we derive some new quasi probability distributions by relaxing some of the conditions of the Lagrange expansions for generating discrete probability distributions. With the relaxed conditions, some new properties of the Lagrangian probability distributions are obtained. We show that a mixture of Lagrangian distribution of the first kind and the second kind gives a Lagrangian distribution of the second kind. This result is used to generate some existing and new quasi probability distributions. Some properties of the new quasi probability distributions are studied.

## Data Analysis of Virtual Cement Measurements

❖ Adriana Hornikova, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *adriana.hornikova@nist.gov*; Charles Hagwood, National Institute of Standards and Technology; Hung-kung Liu, National Institute of Standards and Technology; Blaza Toman, National Institute of Standards and Technology; Nien Fan Zhang, National Institute of Standards and Technology; Edward J. Garboczi, National Institute of Standards and Technology; Jeffrey W. Bullard, National Institute of Standards and Technology

***Key Words:*** data analysis, virtual measurements, simulations, sensitivity analysis, verification and validation

Virtual measurements are the outputs of well-defined mathematical models based on theoretical principles and simulation algorithms. The VCCTL (Virtual Cement and Concrete Testing laboratory) is a software system built by the Material and Construction Research Division of NIST to perform these computations. Its intended use is as a research/exploratory tool to test new cement compounds for their properties. This presentation describes our efforts at verification and validation of the VCCTL software. We have performed sensitivity analysis of the VCCTL measurements of heat of hydration with respect to changes in various input variables. We have also examined the effects of choice of random seeds in the simulations. For a particular subset of Portland cements, it is possible to compare physical measurements obtained by large inter laboratory experiments to the VCCTL virtual measurements.

## Analysis of Number of Components in Mixture Model

❖ Yan Wang, Southern Methodist University, 6041 Village Bend Drive, Apt. 1404, Dallas, TX 75206, *yanw@smu.edu*; S. Lynne Stokes, Southern Methodist University

***Key Words:*** component, mixture model

Latent class model is widely used in social science to classify the similar objects into groups. But the number of classes is key in the latent class model. The arising question is how many classes exist given the data. Determination of number of classes is important but difficult. Several measures of goodness-of-fit are proposed to determine the number of classes. We compare the performance of these measures under latent class models.

## The Impact of Erroneous Inclusion and Exclusion of Variables in Multivariate Inference

❖ Youfeng Nie, Sam Houston State University, 1701 Bobby K Marks Drive, Apt C200, Huntsville, TX 77340, *youfeng@shsu.edu*; Cecil Hallum, Sam Houston State University

***Key Words:*** multivariate, inference, Hotelling

This paper investigates the impact on various multivariate statistical entities including Hotelling's T-square, the Mahalanobis distance function, etc. due to the erroneous inclusion or erroneous exclusion of variables. For example, this impact will occur in multivariate testing hypotheses on the population multivariate mean vector and in constructing confidence ellipsoids for the p-dimensional mean under multivariate normality. Suppose X is a k by 1 vector and suppose p-k additional variables (p>k) are added to X, then the question is how the Hotelling's T-square and other multivariate statistics change? The new entity is to be detailed and simplified to a function of the original entity. Initially, the one-sample and two-sample multivariate statistical entities (e.g. such as Hotelling's T-square) will be examined in this study followed by a generalization to the multivariate case.

## On Testing about a Construct Mean for Likert-Scale Data

❖ Cherng Ding, National Chiao Tung University, 118 ChungHsiao W., Road Sec. 1, Taipei, 100 Taiwan, *cding@mail.nctu.edu.tw*; Hsiu-Yu Lee, National Chiao Tung University

***Key Words:*** construct, Likert scale, Monte-Carlo simulation, power, t test

In marketing research, the Likert-scale is often used to measure a construct through several indicators. Under satisfactory internal consistency and the assumption that the construct has a normal distribution with unknown mean and variance 1, if it is of interest to test if the construct mean achieves a specified level, the traditional t test, based on factor-based scores, may be used. However, its power depends on unknown variances and covariances of indicators, and is difficult to evaluate. To improve, an alternative t test, based on factor scores, is proposed. Its power depends only on sample size and the construct mean. It has been demonstrated by extensive Monte-Carlo simulation that the theoretical power of the test about a normal construct mean can be well approximated by the power of the proposed t test. Some limitations about its use have been discussed.

## Caution When Using Covariate Adjustment in Mixed Effect ANOVA

❖ Zhenxu Ma, Battelle, 5005 King Ave., Room 11 7 002, Columbus, OH 43201-2693, *maj@battelle.org*; Paul Feder, Battelle

***Key Words:*** mixed effect model, covariate adjustment

One of the reasons to include a covariate into an analysis of variance model is that the covariate may be responsible for some of the variation in the dependent variable. Therefore the use of covariate adjustment removes this variation from the error or random variance. The result is to improve sensitivity of the tests for treatment effects. Often the covariate is simply included into the analysis without any transformation. In this presentation, we show that if the slope of the relationship between the dependent variable and the covariate interacts with a random main effect, the variance for the error or random variance may increase with an extra variance being artificially added, and the results are then misleading. A proper way to center the covariate will effectively remove this extra variance.

## Estimating a Population Median from a Small Sample

❖ Boris Shulkin, Christy Industries/Magna International, 18320 Malyn Blvd., Fraser, MI 48026, *bshulkin@comcast.net*; Shlomo Sawilowsky, Wayne State University

***Key Words:*** median, Harrell-Davis, Maritz-Jarrett, l-estimator, k-estimator

The sample median suffers from several limitations: (1) its sampling distribution is intractable, and (2) it is not the best unbiased estimate of the population median. A popular competitor is the Harrell-Davis (1982) estimator, which is based on Maritz and Jarrett (1978). Its beta deviates are an approximation of the probability that the i-th order statistic is the population median. However, that fact is irrelevant to finding the best estimate of the population median for a small sample. A new competitor is proposed that is based on a modified double exponential distribution. It has the desirable properties of producing a smaller root mean square error from the population median, and it is closer to the population median more frequently, as determined by a Monte Carlo study on a variety of theoretical distributions and real data sets, for samples of n < 30.

## On the Use of Heywood Cases for Specification Testing in SEM

❖ Stanislav Kolenikov, University of Missouri-Columbia, 146 Middlebush Hall, Department of Statistics, Columbia, MO 65211, *kolenikovs@missouri.edu*; Kenneth A. Bollen, The University of North Carolina at Chapel Hill

***Key Words:*** latent variables, structural equation models, misspecification, sandwich estimator, estimation on the boundary

The situations when an error or a latent variable variance is estimated to be negative are referred to as Heywood cases in structural equation modeling. We shall discuss the reasons why Heywood cases emerge, and provide asymptotic theory framework for testing whether a Heywood case is indicative of gross model misspecification, as well as simulation results and empirical examples.

## S-PLUS and R Package for Least Angle Regression

❖ Tim C. Hesterberg, Insightful Corporation, 1700 Westlake Ave., Suite 500, Seattle, WA 98109-3044, *timh@insightful.com*

***Key Words:*** regression, feature selection, variable selection, regularization, L1 penalty

Least Angle Regression is a promising new technique for variable selection applications, offering a nice alternative to stepwise regression. It provides an explanation for the similar behavior of Lasso (L1-penalized regression) and forward stagewise regression, and provides a fast implementation of both. I'll demonstrate a prototype open-source S-PLUS/R package "glars" for generalized least angle regression, extending the "lars" package of Efron and Hastie and "glmpath" of Park and Hastie. We invite outside collaboration, and plan for future versions of the package to provide a framework others can build on.

## Partially Repeated Measurements

❖ Mitchell Watnik, California State University, East Bay, Department of Statistics, 25800 Carlos Bee Blvd, Hayward, CA 94542, *mitchell.watnik@csueastbay.edu*; Erica Wong, California State University, East Bay; David Schlessinger, California State University, East Bay

***Key Words:*** mixed models, repeated measures, incomplete data

A problem that occasionally arises in practice occurs when individuals are observed for a particular trait, but some individuals are measured only once, while others are seen multiple times. We refer to this situation as "partially repeated measures". We present the problem formally and develop a solution to it, along with SAS and R code to obtain estimates for parameters. This scenario will be motivated by an example from our consulting.

## Selecting the Best Confidence Interval for a Variance Ratio (or Heritability)

❖ Brent Burch, Northern Arizona University, Flagstaff, AZ 86011-5717, *brent.burch@nau.edu*

***Key Words:*** expected length, probability of false coverage, unbiasedness, mixed linear models

For mixed linear models having two variance components, inferences concerning the ratio of variances are often of primary interest. The variance ratio is related to the proportion of the overall variance due to a source of variation, and the proportion serves as a parameter which measures the importance of the source. Using normality assumptions, computing exact confidence intervals for the parameter under study is a straightforward process. However, difficulties arise when trying to choose an exact interval from a large collection of exact intervals. To select a particular confidence interval procedure, the author considers a quantity that is related to both the expected length and bias of an interval. Applications related to animal breeding studies are presented.

## Calculating Power for Generalized Linear Models Using the Wald Test

❖ Jonathan Mahnken, The University of Kansas Medical Center, 3901 Rainbow Blvd., MSN 1008, Kansas City, KS 66160, *jmahnken@kumc.edu*

***Key Words:*** power, generalized linear model, sample size, Wald test

Wald tests are commonly used for analysis, particularly when the regression model is a generalized linear model. When Wald tests are to be used for analysis it is important to also estimate the power during the design phase using this same test. Also, complex designs are often encountered (e.g., designs with interaction terms). A method of calculating power based on the Wald test is presented. This approach transforms random multivariate normal vector draws to calculate the power of the test statistic under the alternative hypothesis empirically. This method estimates power for more complicated composite hypotheses. Results are presented for data from normal, binomial, and Poisson distributions, which are among the most common distributions assumed in the medical literature. Results are compared to simulations and to other available methods for calculating power.

## Confidence Interval Coverage for Four Effect Sizes for Predictor Variables in a Multiple Linear Regression Model

❖ Todd Bodner, Portland State University, P.O. Box 751 PSY, Portland, OR 97207, *tbodner@pdx.edu*

***Key Words:*** multiple linear regression, effect sizes, confidence intervals, coverage probabilities

Bodner and McCune (2006) compared four effect sizes for multiple regression predictors and reviewed their confidence interval (CI) procedures. The present study evaluates the coverage probability of those CI procedures using a simulation experiment. Varied were the degree of association among the variables and the sample size. Results indicate that CI procedures for unstandardized and standardized partial regression coefficients exhibit excellent coverage probabilities across the conditions studied. However, CI procedures for the semi-partial correlation and change-in-R-squared statistic exhibit varying and often poor coverage probabilities. Although Bodner and McCune (2006) favor the latter two effect sizes for effect size description, these results suggest that CIs based on these statistics should be used for the heuristic purpose of quantifying precision rather than formal inference.

## Large-Cluster Asymptotics for GEE: Working Correlation Models

❖ Hyoju Chung, University of Washington, Box 357232, Seattle, WA 98195-7232, *hyojuch@u.washington.edu*; Thomas Lumley, University of Washington

***Key Words:*** generalized estimating equations, working correlation model, empirical process theory

We consider generalized estimating equations models for correlated data and present large cluster asymptotics. The complexity of working correlation models, which is characterized in terms of the dimension of working correlation parameter, may increase with the number of independent clusters. In our two-index asymptotic setting we show the existence, weak consistency and asymptotic normality of marginal regression parameter estimators using results of empirical process the-

ory. We present sufficient conditions for the increasing complexity of working correlation models using maximal inequalities.

## Free SAS/IML (r) Software for Computing Confidence Limits for Power in the Univariate and Multivariate Approaches to Repeated Measures

❖ Jacqueline Johnson, The University of North Carolina at Chapel Hill, 700 Bolinwood Drive, Apt. 24G, Chapel Hill, NC 27514, *jjohnson@bios.unc.edu*; Matthew Gribbin, The University of North Carolina at Chapel Hill; Sola Park, The University of North Carolina at Chapel Hill; Keith E. Muller, The University of North Carolina at Chapel Hill

***Key Words:*** linear models, repeated measures, power

We describe the latest version of free software which provides convenient power calculations for a wide range of multivariate linear models with Gaussian errors, including the "univariate" (UNIREP) and "multivariate" (MULTIREP) approaches to repeated measures, as well as a limited but useful range of mixed models. New in this version are computations for approximate confidence limits for power for the Box, Geisser-Greenhouse, Huynh-Feldt, and uncorrected UNIREP tests and exact confidence limits for power for the Wilks, Pillai Bartlett, and Hotelling-Lawley MULTIREP tests. These confidence limits may be requested for any power value to reflect the uncertainty due to using any combination of estimated variances and means. We give several examples will to illustrate use.

## Goodness-of-Fit Tests for Proportional Odds Model with GEE for Ordinal Categorical Responses

❖ Junxiang Luo, University of Cincinnati, Department of Environmental Health, College of Medicine, Cincinnati, OH 45267, *luojxgl@gmail.com*; Rakesh Shukla, University of Cincinnati; Qi Zhang, University of Cincinnati

***Key Words:*** goodness of fit, cumulative logit, GEE, proportional odds, ordinal

In 1994, generalized estimating equations approach was extended by Lipsitz, Kim, and Zhao to fit the proportional odds model for analyzing repeated (or clustered) ordinal categorical data. However, few methods exist to assess the goodness-of-fit (GOF) of the fitted models. A U statistic is proposed in this study to do the GOF test, and its performance is evaluated with respect to type I error rates and powers for detecting various model departures by simulation studies and an example illustration. We also compare the proposed U statistic with a kappa-like classification statistic. Results show the proposed U statistic displays much lower type I error rates and appropriate powers compared with the kappa-like classification statistic.

## Latent Class Growth Models: an Application

❖ Maragatha Kuchibhatla, Duke University Medical Center, Box 3003, Durham, NC 27110, *mnk@geri.duke.edu*; Gerda Fillenbaum, Duke University Medical Center

***Key Words:*** latent class growth model, longitudinal data, latent classes, epidemiological

Most growth modeling techniques assume a single growth model to describe the change. Latent class growth models (LCGM) on the other hand, tests if more than one distinct class can be used to describe the sample. The number of latent classes is determined by sequentially increasing the number of classes and examining fit statistics (Akaike information criteria (AIC), Bayesian information criteria (BIC), and sample size adjusted BIC (SSABIC)). LCGM is used here to determine the number of classes to describe change in activities of daily living (ADL) and in cognition (Short Portable Mental Status Questionnaire) using 10-year longitudinal sample of Duke Established Populations for Epidemiologic Studies of the Elderly. LCGM derived 3 classes each in ADL (low declining, moderate declining and high improving groups) and SPMSQ (low stable, low declining and high stable groups).

## Model Selection for the Impact Evaluation of Energy Efficiency Programs

❖ Kathryn Parlin, West Hill Energy and Computing, Inc., 23 Williamstown Road, Chelsea, VT 05038, *kathryn@westhillenergy.com*; Larry Haugh, University of Vermont

***Key Words:*** pooled cross-section time series, model selection, information theoretic approach, impact evaluation

The impact evaluation of conservation programs is designed to determine the savings that can be attributed to a program during a specific period. Linear regression is commonly applied to consumption records for estimating the impacts from energy efficiency measures and programs. Specific modeling decisions can have a major impact on the results of the analysis and variations in methodology can be nearly as numerous as the individuals performing the analysis. This research provides a comprehensive discussion of the types of models that can be applied to the impact analyses of energy conservation programs, some strategies for selecting among the candidate models and an illustration of the range of models and selection process through the application to a specific impact evaluation targeted to the low income population in Vermont. The information-theoretic approach is used to rank mod

## An Empirical Power Analysis of Hierarchical Multivariate Linear Model under Three Covariance Structures in Longitudinal Data Analysis

❖ Hua Fang, Ohio University, 22 First Street, Athens, OH 45701, *hf101702@ohio.edu*; Gordon P. Brooks, Ohio University; Maria L. Rizzo, Ohio University; Robert S. Barcikowski, Ohio University

***Key Words:*** empirical power, hierarchical, multivariate, longitudinal data analysis, bootstrap

This paper examines the empirical power of hierarchical multivariate linear model (HMLM) under three covariance structures in longitudinal data analysis. The three covariance structures are called random slope with homogeneous level-1 variance, unstructured and first-order autoregressive. A stacked SAS macro is written to generate standard hierarchical multivariate data and to compute power under each covariance structure. The power is examined by varying correlation, reliability, effect size, and ratio of group sample size to time points. The bootstrap estimates for the fixed treatment effect are calculated under each covariance structure. Power patterns and bootstrap estimates under each covariance structure are compared through tables and figures. The conclusion discusses importance of covariance selection in the application of HMLM to the longitudinal data analysis.

## A General Probability Distribution Using B¸rmann Power Series

❖ Pali Sen, University of North Florida, Department of Math and Statistics, 4567 St. Johns Bluff Road, Jacksonville, FL 32224, *psen@unf.edu*; Richard F. Patterson, University of North Florida

***Key Words:*** binomial distribution, moment generating function, maximum likelihood estimators, Poisson distribution

The goal of this paper is to present a general power series distribution that exhibits the properties of some well known distributions. To accomplish this goal we examine an infinite sequence of independent random variables having a B¸rmann’s power series distribution. Consequently, we derive moment generating function of the distribution and establish the maximum likelihood estimate of the component that can be attributed to the parameter of the distribution. Using the results mentioned above we verify our conjecture on two known parametric discrete distributions, the Poisson and the Binomial.

### A Modified Asymmetric Simes Procedure for Multiple Tests of Significance

❖ Li Deng, New England College of Optometry, 424 Beacon Street, Boston, MA 02115, *DengL@neco.edu*

***Key Words:*** multiple tests, Simes procedure, asymmetric, type I error, power

Simes (1986) gave a rejection procedure for the overall null hypothesis of multiple tests. Combining with Hommel’s individual test rejection rule, the S-H procedure can yield higher power than Bonferroni approach. But such procedure can produce different conclusions when some p values change slightly. For instance, in the two scenarios: 1) p1=0.0251, p2= 0.0501 and 2) p1=0.0251, p2=0.0499, S-H procedure rejects H1 and H2 in 2) at level 0.05 and accepts both in 1). Here I propose a new procedure. It controls overall type I error as well as increases the chance of claiming significance for tests with small p-values via adjusting that chance for tests with large p-values. This procedure will reject H1 in 1) and reject both hypotheses in 2). Theoretical derivation of the procedure for n=2 and simulations results for type I error and power will be presented.

### An Examination of the Utility of Bonferroni Adjustments for Tests of Regression Coefficients

❖ Daniel Mundfrom, University of Northern Colorado, 501 20th Street, Campus Box 124, Greeley, CO 80639, *daniel.mundfrom@unco.edu*; Jamis Perrett, University of Northern Colorado; Jay Schaffer, University of Northern Colorado; Adam Piccone, University of Northern Colorado

***Key Words:*** multiple linear regression, simulation, Bonferroni adjustments

A common research question for which a MLR analysis would be used involves the determination of which, if any, of the predictor variables display a unique ability to explain variation in the response. Often a test of the adequacy of the overall model is followed with separate tests of each predictor to determine which variables are needed in the model to explain the variation in the response. It is not uncommon for each of these tests to be conducted at the nominal 5% level. This research investigated whether conducting each test with a Bonferroni-type adjusted alpha-level is useful for controlling the overall alpha-level closer to the nominal level. In each case, the Bonferroni-adjusted tests more consistently maintained the overall alpha-level below the nominal level. As the number of regressors increased, the utility of making the adjustment to the overall alpha-level also increased.

### The Brave New Hockey World: a Statistical Assessment of NHL Rules Changes

❖ Paramjit Gill, The University of British Columbia, IK Barber School of Arts and Sciences, 3333 University Way, Kelowna, BC V1V 1V7 Canada, *paramjit.gill@ubc.ca*

***Key Words:*** ice hockey, Bayesian model, dyadic data, NHL, Poisson distribution, sports statistics

The National Hockey League (NHL) last year approved the most radical set of rule changes in many years. The new rules have resulted in more penalties being rewarded and thus leading to more “power play” goals. Traditionally, the Poisson distribution has been used to model the number of goals in NHL. But with power play goals being so prominent, more appropriate distribution will be a mixture of two distributions accounting for non-power and power goals. We present a Bayesian dyadic model for the number of home-team and away-team goals in a game. The model incorporates home-ice advantages in the offensive and defensive abilities of the teams. The teams can be ranked by these abilities and the effect of power play scoring and penalty killing abilities on the final ranking can be studied statistically. Data from the last two seasons are used for modeling the effects of rule changes.

### Statistical Approches To Analyze Censored Data with Multiple Detection Limits

❖ Wei Zhong, ICON Clinical Research, 1700 Pennbrook Parkway, North Wales, PA 19454, *zhongw@iconus.com*; Linda Levin, University of Cincinnati; Paul Succop, University of Cincinnati; Rakesh Shukla, University of Cincinnati; Jeffrey Welge, University of Cincinnati

***Key Words:*** multiply censored data, multiple detection limits, EM algorithm, meta-analysis method, lognormal distribution

Censored data with multiple lower detection limits frequently arise when data are measured by different procedures or are combined from multiple laboratories. The substitution method which replaces nondetects with certain functions of the detection limit is a common approach to compute summary statistics for censored data. However, this method lacks a theoretical basis and results differ depending on the substituted value. MLE with the EM algorithm integrated method and the meta-analysis method were introduced for censored data with lognormal distribution and their properties were evaluated through simulation studies. Compared to the substitution methods, simulation results consistently showed that the proposed methods provided the most accurate and efficient summary statistics for multiply censored data and they were able to incorporate the sample collection process into the estimation.

## 482 COPSS Awards and Fisher Lecture

Committee of Presidents of Statistics Societies (COPSS), The ASA, ENAR, WNAR, IMS, SSC
**Wednesday, August 9, 4:00 pm–5:50 pm**

### Recombination and Linkage

❖ Terence P. Speed, University of California, Berkeley, Department of Statistics, Berkeley, CA , *terry@stat.berkeley.edu*

***Key Words:*** meiosis, crossing-over, recombination, linkage, interference

Genetical recombination is a fundamental biological process, taking place in diploid cells during the formation of haploid gametes. The phenomenon of linkage was discovered in the second decade of the 20th century, and our understanding of its biological basis---recombination---has been developing ever since. R. A. Fisher made several contributions to the theory and practice of linkage and the stochastic modeling of recombination. In this lecture, I’ll summarize the development of such models, from Fisher’s to the present. The methods used range from renewal processes to stochastic point processes to Markov and hidden Markov chains, and the results are elegant models that don't quite fit the data. Interestingly, this has only marginal impact on their use.

## 483 Collaborative Research in Statistics ●

General Methodology, Section on Physical and Engineering Sciences
**Thursday, August 10, 8:30 am–10:20 am**

### Sensor Analytics: Radioactive Gas Quantity Estimation and Error Propagation

❖ Dale N. Anderson, Pacific Northwest National Laboratory, Richland, WA 99352, *dale.anderson@pnl.gov*; Justin I. McIntyre, Pacific Northwest National Laboratory; Deborah K. Carlson, Pacific Northwest National Laboratory; Reynold Suarez, Pacific Northwest National Laboratory; James C. Hayes, Pacific Northwest National Laboratory

***Key Words:*** nuclear gas, probability model, error propagation

This paper develops the mathematical statistics of a radioactive gas quantity measurement and associated error propagation. The probabilistic development is a different approach to deriving attenuation equations and offers easy extensions to more complex gas analysis components through simulation. The mathematical development assumes a sequential process of three components: the collection of an environmental sample, component gas extraction from the sample through the application of gas separation chemistry, and the estimation of radioactivity of component gases.

### Using Informative Bayesian Priors in a Sales Forecasting System

❖ Phillip M. Yelland, Sun Microsystems Laboratories, 3203 Longfellow Drive, Belmont, CA 94002, *phillip.yelland@sun.com*

***Key Words:*** forecasting, Bayesian, prior, elicitation, sales

The use of informative priors for Bayesian forecasting models offers a potentially effective means of combining judgmental and statistical information, which is seen commonly as a prerequisite of consistently superior performance in practical forecasting applications. This presentation centers on a sales forecasting system developed at computer systems vendor Sun Microsystems that relies on the provision of such informative Bayesian priors by sales and supply management personnel. The discussion will recount how---in consultation and collaboration with company staff---the functional forms of the priors were derived and how elicitation procedures were devised for them. The role played by the priors in the forecasting models also will be discussed, as will the performance of the resulting system.

### Formulation Prediction for Derivative Product Development

❖ Martha Gardner, GE Global Research, 4031 Buckingham Drive, Schenectady, NY 12304-2472, *martha.gardner@ge.com*

***Key Words:*** robust design, experimental design, materials, optimization

When new materials are being designed, it may not be obvious which combination of ingredients will produce a material that best meets the key properties of interest. Thus, experimentation may be required to find the best combination of ingredients that will yield such a material. Such experimentation may be time consuming and expensive, but can have valuable results by showing relationships between the inputs and outputs. When requirements for a new material are defined, it is possible that experimental work done in the past could be useful in the current assessment, but the results of such studies may not be readily available. This paper describes a formulation prediction system based on robust platform design space development that was developed to solve this problem and dramatically decrease the cycle time of supplying derivative product samples to customers.

## 484 Statistical Methodology for Environmental Applications ✪

International Indian Statistical Association
**Thursday, August 10, 8:30 am–10:20 am**

### Bayesian Melding: an Application and Critical Assessment

❖ James Zidek, The University of British Columbia, Department of Statistics, 6356 Agriculture Rd, Vancouver, BC V6T 1Z2 Canada, *jim@stat.ubc.ca*; Zhong Liu, The University of British Columbia; Nhu Le, BC Cancer Agency/University of British Columbia

***Key Words:*** ozone, Bayesian melding, physical modeling, spatial prediction

Adrian Raftery and his coinvestigators developed Bayesian melding as a method for combining grid cell outputs from a deterministic model with point measurements from a random spatial field made at monitoring sites. This paper will describe the method and present the results obtained by applying it to the hourly and weekly ozone field over the eastern United States. The data involves hourly grid cell outputs from the MAQSIP model and monitoring data from the AIRS database. Through simulation studies, we investigate topics such as how well spatial predictors based on the method work and whether the grid cell data helps to make the predictor more accurate (as well as better calibrated). We demonstrate the method on the data and describe extensions of the method to, for example, space-time fields.

### Fixed Rank Kriging for Massive Datasets

❖ Noel Cressie, The Ohio State University, Department of Statistics, 1958 Neil Avenue, Columbus, OH 43210-1247, *ncressie@stat.ohio-state.edu*; Gardar Johannesson, Lawrence Livermore National Laboratory

***Key Words:*** BLUP, covariance function, geostatistics, kriging variance, nonstationarity, total column ozone

Environmental datasets obtained from satellites are typically massive in size. The massiveness causes problems in computing optimal spatial predictors, such as kriging. In this paper, a flexible family of nonstationary covariance functions is constructed using a set of basis functions fixed in number. This results in computational simplications in deriving the kriging predictor and its kriging variance. We call the methodology fixed rank kriging (FRK), and we apply it to a large dataset of remotely sensed Total Column Ozone (TCO) data, observed over the entire globe.

### Minimum Distance Inference in Unilateral Autoregressive Lattice Processes

❖ Marc G. Genton, Texas A&M University, Department of Statistics, College Station, TX 77843-3143, *genton@stat.tamu.edu*; Hira L. Koul, Michigan State University

***Key Words:*** autoregressive, lattice, least squares, minimum distance, spatial, time series

We discuss two classes of minimum distance estimators of the underlying parameters and their robust variants in unilateral autoregressive lattice models. We present an asymptotically distribution free test for testing the symmetry of the error distribution, a goodness-of-fit test for fitting an error distribution, and a lack-of-fit test for the hypothesis that the given process is doubly geometric based on the least absolute deviation residuals. A simulation study investigates some small sample properties of the estimators and their robustness. It shows that some of the proposed estimators are more efficient than the least squares estimator at non-normal error distributions. We also study the empirical level and power of the test of a doubly geometric process at various error distributions. Our methodology is then applied to a real data set of yields from an agricultural experiment.

### Computation and Modeling for Large Space-Time Environmental Datasets

❖ Michael L. Stein, The University of Chicago, Department of Statistics, 5734 University Ave., Chicago, IL 60637, *stein@galton.uchicago.edu*

***Key Words:*** Gaussian processes, stratospheric ozone

Our ability to analyze large space-time environmental datasets often is constrained by computational considerations. Most environmental monitoring datasets have regular patterns in their space-time locations, such as a fixed monitoring network with evenly spaced observation times. We will consider space-time models under which exact or highly accurate approximations to the likelihood can be computed for such datasets, paying particular attention to the common situation in which there is a modest fraction of missing observations.

## 485 Statistical Genetics and Nonparametric Methods: in Memory of Howard Levene

Memorial, Section on Nonparametric Statistics

**Thursday, August 10, 8:30 am–10:20 am**

### The Influence of Professor Levene's Test of the Equality of the Variances of k Groups on the Development of Robust Procedures and Their Application

❖ Joseph Gastwirth, The George Washington University, Department of Statistics, 2140 Pennsylvania Ave., N.W., Washington, DC 20052, *jlgast@gwu.edu*

***Key Words:*** equality of variances, robust tests, preliminary tests

After reviewing the statistical test proposed by Professor Levene in 1960, several major subsequent modifications and applications will be described. Because the original paper has been cited in more than 500 scientific articles, this survey cannot be comprehensive, but will describe several important developments and uses of robust methods concerning the testing and estimation of variances that followed Levene's original, insightful contribution.

### The Influence of Levene's Paper on Polymorphism in Subdivided Populations

❖ Edward Pollak, Iowa State University, Department of Statistics, Ames, IA 50011-1210, *pllk@iastate.edu*

***Key Words:*** selection, subdivided population, fixation probabilities

In 1953, Howard Levene wrote a short paper that was published as a letter to the editor of The American Naturalist. In it, he showed that if more than one ecological niche is available to a population and one allele is favored over another in some niches and is at a selective disadvantage elsewhere, it is possible to have a stable equilibrium with both alleles present in substantial proportions everywhere. This paper still appears in reference lists of publications in which ideas contained in it are elaborated and generalized. I shall discuss lines of research to which Levene's note has given rise.

### Howard Levene: Statistician, Educator, Colleague, and Friend

❖ Theodore W. Anderson, Stanford University, Department of Statistics, Sequoia Hall, Stanford, CA 94305, *twa@stanford.edu*

Howard Levene had essentially all of his career at Columbia University as a professor of mathematical statistics and zoology. In statistics, his major interest was in nonparametric inference; in zoology, he taught biometrics and pursued research in genetics, often in collaboration with professors Theodosius Dobzhansky and K. C. Dunn. His PhD was one of the first two granted by the Department of Mathematical Statistics (1947). He immediately became a faculty member and continued until retirement. He bore a considerable responsibility for the success of the department through more than half of its existence. Many of his students are active in departments of statistics, biostatistics, and genetics.

### Testing Multivariate Scale Difference by Depth Rank Tests

❖ Regina Liu, Rutgers University, Department of Statistics, 501 Hill Center, Piscataway, NJ 08854-8019, *rliu@stat.rutgers.edu*; Kesar Singh, Rutgers University

***Key Words:*** data depth, Kruskal-Wallis test, Wilcoxon rank sum test, multivariate scale difference

Professor Levene contributed much to the subject of scale difference detection. We use data depth to develop new rank tests for testing scale difference in multiple multivariate samples. Consider two samples from two multivariate distributions which are identical except for a possible scale difference. The sample with the larger scale would tend to be more outlying in the combined sample. Since a data depth is a measure of outlyingness, its center-outward ranking is well-suited for forming multivariate rank tests for testing scale difference in two samples. These depth rank tests can be carried out the same way as the Wilcoxon rank sum test for univariate locations. We also generalize depth rank tests to Kruskal-Wallis--type tests for testing scale homogeneity of multiple multivariate samples. Finally, as an application, we apply our tests to compare performance stability of airlines.

# 486 Experimental Design for Nonlinear Modeling ●

Biopharmaceutical Section
**Thursday, August 10, 8:30 am–10:20 am**

### Curvature, Robustness, and Optimal Design in Applied Nonlinear Regression Modeling

❖ Timothy E. O'Brien, Loyola University Chicago, Loyola Math Department, 6525 N. Sheridan Road, Chicago, IL 60626, *teobrien@gmail.com*

***Key Words:*** experimental design, lack of fit, differential geometry, optimality

Researchers often find that nonlinear regression models are more applicable for modeling their processes than are linear ones. These researchers are thus often in a position of requiring optimal or near-optimal designs for a given nonlinear model. A common shortcoming of most optimal designs for nonlinear models used in practical settings, however, is that these designs typically focus on only (first-order) parameter variance or predicted variance and ignore the inherent nonlinearity of the assumed model function. Another shortcoming of optimal designs is that they often have only p support points, where p is the number of model parameters. This talk examines the reliability of Clarke's marginal curvature measures in practical settings and introduces a design criterion that combines variance minimization with nonlinearity minimization. Numerous illustrations will be provided.

### Locally D-Optimal Designs Based on Models Combining Emax and One-Compartment Models

❖ Sam Hedayat, University of Illinois at Chicago, 851 S. Morgan Street, Dept. of Math., Stat. & Comp. Sciences (M/C 249), Chicago, IL 60607-7045, *hedayat@uic.edu*

***Key Words:*** pharmacokinetic, pharmacodynamic, compartmental model, non-linear model, Emax model, d-optimal design

Four- and five-parameter models combining a pharmacokinetic compartmental model and a pharmacodynamic Emax model are studied here. Locally D-optimal (LD) designs and two classes of robust designs are investigated based on these two models. The required support size of a locally D-optimal design is found to be between four and six, based on the four-parameter combined model. Further, a sufficient condition of when the support requires the minimum number of design points is given based on the four-parameter model. Examples of the LD designs are presented for a certain set of nominal parameter values. Simulation results show LD designs and robust LD designs cannot be traded in terms of coefficient of variation and point estimation if the nominal values of the unknown parameter aren't close to the true values. It's demonstrated that LD design should be employed instead of robust LD designs.

### Optimal Designs in Nonlinear Mixed-Effects Models: Applications for Population Pharmacokinetics/Pharmacodynamics Studies

❖ Sylvie Retout, INSERM U738, Biostatistics Unit, Bichat - Claude Bernard Hospital, 46 rue Henri Huchard, University Paris 7, Paris, 75018 France, *sylvie.retout@bch.ap-hop-paris.fr*; Emmanuelle Comets, INSERM U738; France Mentre, INSERM U738

***Key Words:*** optimal population designs, Fisher information matrix, nonlinear mixed effects models, population pharmacokinetics pharmacodynamics

We focus on the evaluation and optimization of designs in the context of nonlinear mixed-effects models. We present our methodology based on an approximation of the expression of the Fisher information matrix for these models, using a first-order linearization of the model around the random effects. Models including additional parameters for covariates and/or inter-occasions variability also are considered. Based on this methodology, we present PFIM and PFIMOPT, R/SPlus functions for population designs evaluation and optimization, respectively. Optimization uses the D-optimality criterion. Two algorithms are implemented, a general Simplex algorithm and a more specific one, the Fedorov-Wynn algorithm. Models can be supplied either analytically or under differential equations form. We illustrate the usefulness of PFIM and PFIMOPT on population pharmacokinetics/pharmacodynamics examples.

### Modeling and Clinical Trial Simulation in the Post--Critical Path Drug

❖ Michael D. Hale, Amgen Inc., One Amgen Center Drive, MS#38-3-B, Thousand Oaks, CA 91320-1799, *mdhale@amgen.com*

***Key Words:*** simulation, modeling, critical path

Clinical trial simulation was promoted a decade ago as a tool to better understand relationships among dosing schedules, patient populations, trial design, and analysis. While much skepticism was present, many of the ideas have become ubiquitous and are clearly present in high-visibility efforts such as the FDA's Critical Path program. Questions remain about how such methods can become part of mainstream drug development, rather than a specialist activity applied to particularly difficult projects, and how new biomarker imaging techniques and genomic data can be integrated into these efforts. This talk will consider the complexities of incorporating these new sources and implications.

# 487 Women Pioneering Solutions to New Global Challenges ● ✪

Committee on Women in Statistics
**Thursday, August 10, 8:30 am–10:20 am**

### Aggregation, Lasso, and Sparsity

❖ Florentina Bunea, Florida State University, FL 32304, *bunea@stat.fsu.edu*

***Key Words:*** aggregation, LASSO, high dimensions, oracle inequalities, optimal risk bounds, sparsity

In this talk we show that penalized least squares, with a Lasso-type penalty, yields risk optimal aggregates (data dependent linear combinations) of arbitrary estimators of a conditional mean function. A new notion of risk optimality is introduced and discussed. Moreover, if the target mean function has a sparse representation in terms of the given estimates, our aggregate will adapt to this unknown sparsity. This has immediate applications to high dimensional/low sample size regression problems.

## A Bayesian Hierarchical Model with Nest-Specific Covariates in Nest Survival Study

❖ Jing Cao, Southern Methodist University, 3225 Daniel Ave., P O Box 750332, Dallas, TX 75275-0332, *jcao@smu.edu*; Chong He, Virginia Polytechnic Institute and State University

***Key Words:*** Bayesian hierarchical model, nest survival, irregular visiting schedule, intrinsic auto-regressive prior, nest-specific covariates

Recently, logistic nest survival models have been developed to incorporate biological covariates with some restrictive assumptions. Also, the nest curve is assumed to be a parametric function (linear or quadratic) of nest age. We propose a Bayesian hierarchical model with nest-specific covariates to estimate age-specific daily survival rates. The model has the least restrictive assumptions, and it can handle irregular visiting schedules. The typical features of nest survival data---truncation and censoring---are accounted for by the likelihood function and latent variables. An intrinsic auto-regressive (IAR(2)) prior is employed for the nest age effect. This nonparametric prior provides a much more flexible and parsimonious alternative to the parametric specification. A simulation study and an analysis of a Missouri dickcissel dataset also are presented.

## Joint Statistical Models for Genome-Wide Tiling Array and Sequence Data

❖ Sunduz Keles, University of Wisconsin-Madison, Departments of Statistics and Biostatistics, 1300 University Avenue 1245 MSC, Madison, WI 53710, *keles@stat.wisc.edu*; Heejung Shim, University of Wisconsin-Madison

***Key Words:*** genome-wide tiling array, transcription factor, mixture regression models, motif discovery

Tiling arrays that interrogate the genome in a systematic, unbiased fashion are becoming instrumental for the genome-wide identification and characterization of functional elements. We propose a class of statistical models to analyze data from these arrays in combination with sequence data to effectively identify transcription factors binding sites. These joint models carefully address the cases where tiling array data show enrichment in regions that do not correspond to transcription factor binding sites. We discuss inference procedures for such models and illustrate their utility with several real data examples.

## A Markov-Hidden Markov Model for Genetic Admixture

❖ Hua Tang, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., M2-B500, P.O. Box 19024, Seattle, WA 98109, *huatang@fhcrc.org*; Pei Wang, Fred Hutchinson Cancer Research Center; Marc Coram, The University of Chicago

We consider the problem of inferring ancestry along chromosomes in an admixed population using high-density genotyping data. The Markov-Hidden Markov Model (MHMM) we propose accounts for background linkage disequilibrium (LD) that exists in ancestry populations. We also extend the existing approaches to allow each ancestral population to admix at a different time in history. We use simulation to illustrate the importance of modeling the background LD. The MHMM opens up the possibility to infer ancestry and perform admixture mapping without genotyping any special ancestry-informative marker panels.

# 488 Bayesian Methods in Bioinformatics ● ✪

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

## Bayesian Mixture Models and Application to High-Throughput Data

❖ Kim-Anh Do, M. D. Anderson Cancer Center, Department of Biostatistics and Applied Mathematics, 1515 Holcombe Blvd, Houston, TX 77030-4095, *kim@mdanderson.org*

***Key Words:*** mixture models, gene expression, mass spectrometry, Dirichlet process, beta distribution, cancer experiments

The scientific world recently witnessed an explosion in the development of comprehensive, high-throughput methods for molecular biology experimentation. Our talk will focus on the development of Bayesian nonparametric mixture models with applications to two main platforms: microarray gene expression and mass spectrometry (MS) proteomic profiles. First, model-based inference is proposed for differential gene expression using a variation of the traditional Dirichlet process (DP) mixture models for the distribution of gene intensities under different conditions. Second, the unknown distribution of mass/charge ratios in MS data can be represented as a density estimation problem via a mixture of betas. We will present simulation studies and cancer-related experiments, contrasting the intrinsic differences in the technology and posterior inference that can incorporate multiplicities automatically.

## Bayesian Modeling of Complex Traits

❖ Paola Sebastiani, Boston University, Department of Biostatistics, 715 Albany Street, Boston, MA 02118, *sebas@bu.edu*

***Key Words:*** Bayesian networks, complex traits, gene-environment

The advances in genotyping technology are producing massive amounts of genetic and genomic data that, together with information collected in large epidemiology studies, hold the promise of discovering the genetic basis of many common diseases. This talk will describe recent Bayesian models that have been used to discover the genetic basis of some complex traits and to develop prognostic models that can identify subjects with higher susceptibility to disease. Issues covered will include how to deal with missing data, how to model the complex web of interaction between genetic and epidemiology variables with relatively small sample sizes using Bayesian networks, and how to define phenotype from data precisely.

## Functional Clustering by Bayesian Wavelet Methods

❖ Bani K. Mallick, Texas A&M University, Department of Statistics, 3143 TAMU, College Station, TX 77845, *bmallick@stat.tamu.edu*

***Key Words:*** functional clustering, wavelets, mixture of Dirichlet, microarray gene expression data

We propose a nonparametric Bayes wavelet model for clustering of functional data. The wavelet-based methodology is aimed at the resolution of generic global and local features during clustering and is suitable for clustering high-dimensional data. Based on the Dirichlet process (DP), the nonparametric Bayes model extends the scope of traditional Bayes wavelet methods to functional clustering and allows the elicitation of prior belief about the regularity of the functions and the number of clusters by suitably mixing the DP. Posterior inference is carried out by Gibbs sampling with conjugate priors, which makes the computation straightforward. The models have been used successfully to analyze time course microarray gene expression profiles.

### Multiclass Cancer Diagnosis with Bayesian Kernel Machine Models

❖ Sounak Chakraborty, University of Missouri-Columbia, 146 Middlebush Hall, Department of Statistics, Columbia, MO 65211-6100, *chakrabortys@missouri.edu*

***Key Words:*** support vector machine, reproducing kernel Hilbert space, glioma cancer, classification, Bayesian analysis

In modern clinical neuro-oncology, the diagnosis and classification of malignant gliomas remains problematic, and effective therapies are still elusive. In recent years, there has been a move toward the use of cDNA microarrays for tumor classification. The classification becomes more difficult when there are more than two cancer types, as with glioma. This paper considers several Bayesian classification methods for the analysis of the glioma cancer with microarray data based on reproducing kernel Hilbert space under the multiclass setup. We consider the multinomial logit likelihood and the likelihood related to the multiclass Support Vector Machine (SVM) model. It is shown that our proposed Bayesian classification models with multiple shrinkage parameters can produce accurate classification scheme for the glioma cancer compared to the existing classical methods.

## 489 Bayesian Inference for Population Survey Data ●

Section on Survey Research Methods, Section on Bayesian Statistical Science

**Thursday, August 10, 8:30 am–10:20 am**

### The Bayesian Approach to Inference for Finite Population Samples

❖ Roderick J. Little, University of Michigan, Department of Biostatistics, 1420 Washington Heights, Ann Arbor, MI 48109, *rlittle@umich.edu*

***Key Words:*** sample survey inference, complex sample designs, Bayesian methods

The Bayesian approach to statistical inference has become popular in many areas of science, but the approach is not embraced widely for sample survey inference, where the design-based paradigm remains predominant. Reasons include concerns about the effects of model misspecification, the need for simple procedures for extensive survey analysis, and antipathy toward the subjective elements of Bayes when it comes to 'objective' reporting of survey results. The author argues that these objections can be overcome and the Bayesian paradigm is an excellent tool for survey inference, provided attention is paid to incorporating features of the survey design into the models and strong parametric assumptions are avoided when the sample size is large. Some applications of the Bayesian approach are offered in support of this claim.

### Weight Trimming via Bayesian Variable Selection Methods

❖ Michael R. Elliott, University of Michigan, M4041 SPH II, 1420 Washington Heights, Ann Arbor, MI 48109, *mrelliot@umich.edu*

***Key Words:*** sample survey inference, sampling weights, regression estimators, variable selection

In unequal-probability-of-inclusion samples, correlations between probability of selection and sampled data can induce bias. Weights equal to the inverse of the probability of inclusion are used to reduce this bias. Highly disproportional sample designs have large weights, which often introduces a high degree of variability in inference. Weight trimming reduces large weights to a fixed cutpoint value, reducing variability but introducing bias. Standard approaches typically do not use the data to make bias-variance tradeoffs. This presentation will develop Bayesian variable selection methods for weight trimming to supplement standard, ad-hoc, design-based methods in disproportional probability-of-inclusion designs where variances due to sample weights exceed bias correction. These methods will be used to estimate linear and generalized linear regression model population parameters.

### Bayesian Approach for Combining Information from Multiple Surveys

❖ Trivellore E. Raghunathan, University of Michigan, Department of Biostatistics, SPH-II, 1420 Washington Heights, Ann Arbor, MI 48109-2029, *teraghu@umich.edu*

***Key Words:*** complex surveys, Markov chain Monte Carlo, nonresponse, measurement errors, post-stratification, probability surveys

In many survey organizations, a large number of primary sampling units are drawn and their dwelling units listed close to decennial census years. Over the intercensus years, different surveys select dwelling units, households, subjects from this pool, and any new dwelling units constructed since the last listing through a well-defined probability selection scheme. Furthermore, the same questions may be asked in multiple surveys. Such sampling schemes provide a unique opportunity to combine information from multiple surveys to enhance the estimates of the population quantities and the temporal changes in the population quantities. This paper explores a Bayesian framework for the design and analysis that combines information from multiple multistage complex surveys. The methodology also will be used to inform on designs to increase precision and assess nonsampling errors.

### Influence Functions and Robust Bayes Estimation

❖ Malay Ghosh, University of Florida, Department of Statistics, Box 118545, Gainesville, FL 32611-8545, *ghoshm@stat.ufl.edu*; Tapabrata Maiti, Iowa State University; Ananya Roy, University of Florida

***Key Words:*** influence functions, small area estimation, robust Bayes, robust empirical Bayes

We introduce new robust Bayes and empirical Bayes small-area estimation procedures based on area level models. We have derived an expression for the predictive influence function and proposed new small-area estimators based on standardized versions of the same. Mean-squared errors and estimated mean-squared errors of these estimators also are found. A small simulation study shows the superior performance of the proposed estimators over the standard estimators.

# 490 Graphical Tools for Spatial Econometrics ●

Section on Statistical Graphics
**Thursday, August 10, 8:30 am–10:20 am**

## Using ArcView Mapping Files in Statistical Software Programs

❖ James P. LeSage, The University of Toledo, Economics Department, 2801 W. Bancroft St., Toledo, OH 43606, *james.lesage@utoledo.edu*

***Key Words:*** shapelib, MATLAB, map polygons

Use of statistical functionality for spatial modeling and analysis in conjunction with a mapping interface has received a great deal of attention in the spatial analysis literature. The feasibility of extracting map polygon and database information from ESRI's ArcView shape files for use in statistical software environments is demonstrated here. Specifically, we show that information containing map polygons can be used in most statistical software environments to produce high-quality mapping functionality. Recent improvements in computer graphics hardware allow rapid rendering of map polygons in statistical software that runs on desktop and laptop computers. This allows the basic plotting functionality of statistical software environments to produce mapping functionality based on the high-quality ArcView map polygons.

## GeoXp: an R Package for Interactive Exploratory Spatial Data Analysis

❖ Anne Ruiz-Gazen, University Toulouse I, Gremaq, 21, allee de Brienne, Toulouse, 31042 France, *ruiz@cict.fr*; Christine Thomas-Agnan, University of Toulouse

***Key Words:*** spatial exploratory data analysis, R package, spatial econometrics and statistics, interactive graphical tools

GeoXp is a tool for researchers and experienced users in spatial statistics, spatial econometrics, geography, ecology, epidemiology, etc allowing to link dynamically statistical plots with elementary maps. GeoXp includes tools from different areas of spatial statistics including geostatistics as well as spatial econometrics. Besides elementary plots like boxplots, histograms or scatterplots, GeoXp also couples maps with Moran scatterplots, variogram clouds, angle plots, neighbor plots, etc. Among the unique features of GeoXp, we will emphasize non parametric density estimators, weighted Lorentz curves and generalized principle components analysis. As an open source code, it is flexible and easily adaptable and benefits from the incredible expansion of the R environment. We illustrate the use of GeoXp with a data basis from education in the French Midi-Pyrenees region.

## Exploring Spatial Data with openGeoDa and PySAL

❖ Luc Anselin, University of Illinois, Urbana, IL 61801, *anselin@uiuc.edu*; Sergio J. Rey, San Diego State University

***Key Words:*** spatial analysis, ESDA, spatial econometrics, GIS, geovisualization, open source software

This paper presents two open-source software tools to visualize and explore spatial data and carry out spatial econometric analysis. openGeoDa is the cross-platform, open-source successor to the GeoDa package, written in C++, and primarily aimed at data exploration. PySAL is a library of modules for spatial data analysis written in Python. It is intended to provide a comprehensive and flexible toolbox for a range of spatial data analyses, including visualization, rate smoothing, clustering, space-time exploration, and spatial regression. The library can be combined with different graphical front ends, such as STARS (using tK) and PySpace (using wxPython). We review the different architectures of the two approaches and their respective strengths and weaknesses and provide illustrative examples.

# 491 Bayesian Nonparametric Methods ●

ENAR, WNAR, Section on Bayesian Statistical Science, Section on Nonparametric Statistics
**Thursday, August 10, 8:30 am–10:20 am**

## Split-Merge Markov Chain Monte Carlo for a Nonconjugate Dirichlet Process Mixture Model

❖ Sonia Jain, University of California, San Diego, Division of Biostatistics and Bioinformatics, 9500 Gilman Drive, MC 0717, La Jolla, CA 92093-0717, *sojain@ucsd.edu*; Radford Neal, University of Toronto

***Key Words:*** Bayesian mixture model, Markov chain Monte Carlo, split-merge moves, nonconjugate prior

A major impediment in designing Markov chain Monte Carlo algorithms for nonconjugate models is the computational difficulty that arises when the model is no longer analytically tractable. We propose a new nonincremental Markov chain sampling technique that efficiently clusters heterogeneous data by splitting and merging mixture components of a nonconjugate Dirichlet Process mixture model. Our method, which is a generalization of our conjugate, split-merge, Metropolis-Hastings procedure, will accommodate models with a specific type of nonconjugate prior---the conditionally conjugate family of priors. Appropriate Metropolis-Hastings split-merge proposal distributions are obtained by utilizing properties of a restricted Gibbs sampling scan. Highlights from a simulation study will be shown.

## Bayesian Semiparametric Inferences for Disease Risk, ROC Curves, and Prevalence

❖ Wesley O. Johnson, University of California, Irvine, Donald Bren School of Information and Computer Science, Department of Statistics, Irvine, CA 92697-1250, *wjohnson@uci.edu*; Adam Branscum, University of Kentucky

***Key Words:*** sensitivity, specificity, receiver operating characteristic curve, diagnosis, mixture model

We discuss the application of mixtures of Polya Trees to the semiparametric estimation of mixture distributions in the context of assessing disease risk and assessing the quality of a diagnostic procedure. Sampled individuals are either diseased or not, and their status is unknown. Diagnostic procedures result in a score from a corresponding mixture distribution where the mixing parameter is the prevalence of the disease. Bayesian nonparametric and semiparametric inferences are provided for Receiver Operating Characteristic curves and areas under them, as well as for prevalences. Covariates are modeled for the purpose of classification of individuals with unknown status as either "diseased" or "nondiseased" and incorporated as factors that might affect the quality of a diagnostic test.

### Nonparametric Estimation of Copulas via Mixtures

❖ Peter Hoff, University of Washington, Seattle, WA 98195-4322, *hoff@stat.washington.edu*

***Key Words:*** extreme points, multivariate estimation, mixture model

Every multivariate probability distribution can be decomposed into its univariate marginal densities and a copula, a joint density on the square with uniform marginals. In this talk we discuss mixture models for nonparametric estimation of copulas. In particular, we present a Dirichlet process mixture prior based on Bernstein polynomials whose support is the entire space of multivariate copulas. Estimation proceeds via a straightforward MCMC algorithm. Example analyses of mixed discrete and continuous data are presented, and model simplification strategies are discussed.

### Bayesian Nonparametric Spatial and Spatio-Temporal Models for Disease Incidence Data

❖ Athanasios Kottas, University of California, Santa Cruz, School of Engineering, MS: SOE2, 1156 High Street, Santa Cruz, CA 95064, *thanos@ams.ucsc.edu*; Jason Duan, Duke University; Alan E. Gelfand, Duke University

***Key Words:*** areal unit spatial data, Dirichlet process mixture models, disease mapping, dynamic spatial process model, Gaussian process, spatial Dirichlet process

We propose Bayesian nonparametric approaches to modeling disease incidence data, which are typically available as rates or counts for specified regions and collected over time. Within a hierarchical framework, we use a spatial Dirichlet process as the model for the surface of spatial random effects. We then block average the spatial Dirichlet process to the areal units determined by the regions in the study to obtain a prior model for the finite dimensional distribution of the spatial random effects. Moreover, we employ a dynamic formulation for the spatial random effects to extend the model to spatio-temporal settings. We illustrate the methodology with simulated data and a dataset on lung cancer incidences for all 88 counties in Ohio over an observation period of 21 years.

## 492 Fusing Environmental Data with Numerical Models ●

Section on Statistics and the Environment, Section on Bayesian Statistical Science

**Thursday, August 10, 8:30 am–10:20 am**

### Statistical Data Assimilation To Improve Hurricane Forecasting

❖ Montserrat Fuentes, North Carolina State University, 8203 Campus Box, Statistics Department, Raleigh, NC 27695, *fuentes@stat.ncsu.edu*; Kristen M. Foley, North Carolina State University

***Key Words:*** Bayesian statistics, coastal ocean models, covariance functions, multivariate analysis, spatial statistics

Estimating the spatial and temporal variation of surface wind stress fields plays an important role in hurricane forecasting. According to NOAA Hurricane Research Division, more than 85% of storm surge is caused by winds pushing the ocean surface ahead of the storm. Estimated storm surge values using coastal ocean deterministic models are used for assessments of warnings and evacuation notices and for recovery operations. We propose a Bayesian spatial-temporal statistical framework to obtain more accurate prediction of wind fields by combining information from buoys, ships, satellites, and physical models. A spatial-temporal nonstationary linear model of coregionalization is used for wind fields. We present the improvement in the estimated storm surge using coastal ocean deterministic models by having our predicted wind fields as input fields.

### Stochastic Parameterizations in Numerical Weather Forecasting Models

❖ Christopher K. Wikle, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211, *wiklec@missouri.edu*; Yong Song, University of Missouri-Columbia; Christopher Anderson, National Oceanic & Atmospheric Administration

***Key Words:*** weather, climate, Bayesian, assimilation, precipitation

The numerical solution of the large systems of differential equations that govern the atmosphere are critical for weather and climate forecasting. Such models require parameterizations for physical processes, such as convection, that occur on scales smaller than the model grid resolution. These subgrid-scale processes typically are based on quasi-empirical/physical relationships and considered fixed in the numerical model. As such, these parameterizations are the source of much of the model uncertainty that leads to forecast error. We develop a stochastic version of a common convective initiation scheme and train model parameters in a Bayesian setting based on radar observations of convective precipitation. The resulting posterior distribution is then used in the numerical model in a forecasting mode. Case studies demonstrate a remarkable improvement in precipitation forecasts.

### Ensemble Smoothing for Understanding Geophysical Processes

❖ Douglas W. Nychka, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, *nychka@ucar.edu*

***Key Words:*** data assimilation, Kalman filter, spatial statistics

Much of the understanding and prediction of geophysical processes comes from the construction and use of large numerical models that simulate natural phenomenon. A key area for statistical science is when these models are confronted with data. Part of the challenge is computational, and straight forward maximum-likelihood or Bayesian approaches are not feasible. The Ensemble Kalman Filter (EKF) is a Monte Carlo--based approximation to a Gaussian-Bayesian posterior that can handle high-dimensional state vectors and processes that evolve nonlinearly over time. Examples will be given using the EKF in the Data Assimilation Research Testbed and NCAR community atmospheric model to tune model parameters and estimate the state based on sparse observations.

## 493 Federal Data-Sharing Requirements: Issues, Ethical Challenges, and Roles for the Statistical Community ● ✪

Section on Government Statistics

**Thursday, August 10, 8:30 am–10:20 am**

### Federal Data Sharing Requirements: Issues, Ethical Challenges, and Roles for the Statistical Community

❖ Joan E. Sieber, California State University, East Bay, 2060 Quail Canyon Court, Hayward, CA 94542, *joan.sieber@sbcglobal.net*; ❖ Karen M. Hansen, Fred Hutchinson Cancer Research Center, J6-110, PO Box 19024, Seattle, WA 98109-1024, *khansen@fhcrc.org*; ❖ Helen McGough, University of Washington, 3935 University Way NE, Office of Research, Human Subjects Division, Seattle, WA 98105-6613, *hmcgough@u.washington.edu*; ❖ Alan M. Zaslavsky, Harvard Medical School, Department of Health Care Policy, 180 Longwood Avenue, Boston, MA 02115-5821, *zaslavsk@hcp.med.harvard.edu*

***Key Words:*** data sharing, disclosure avoidance, research ethics committees, HIPAA

In this panel discussion, we will review briefly current data sharing practices and relevant requirements of selected federal agencies in the United States that fund human subjects research. We also will summarize the Privacy Rule of the Health Insurance Portability and Accountability Act of 1996. Panel members will address important questions: How can statisticians work with principal investigators (PIs) and research ethics committees as they seek to comply with data sharing requirements? How can statisticians better educate research administrators about their responsibilities as their institutions participate in data sharing agreements with funders and secondary data users? Through what programs can the statistical community better educate PIs, research ethics committees, and research administrators about suitable methods for protecting confidentiality?

## 494 Survival Analysis Issues for Medical Devices ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

### Confidence Levels of Confidence Intervals in the Survival Analysis

❖ Xiaolong Shih, Boston Scientific Corporation, One Scimed Place, Maple Grove, MN 55311, *alex.shih@bsci.com*

***Key Words:*** survival analysis, variance estimate, confidence interval and simulation

Survival analysis has been used widely in clinical research. In many single-arm trials with objective performance criteria, the one-sided upper or lower confidence limits have been chosen as the only standard in the testing procedure on whether the trial objectives are met. In this presentation, simulation examples will illustrate the performances of several commonly used methods---such as Greenwood, Peto, Dorey, and Korn---in computing the confidence intervals. Also, recommendations of which methods should be used will be made.

### Futility Analysis for a Randomized Trial with a Time-to-Event Endpoint

❖ Corina M. Sirbu, Boston Scientific Corporation, 100 Boston Scientific Way, Marlborough, MA 01752-1242, *sirbuc@bsci.com*; Kay M. Larholt, Boston Scientific Corporation

***Key Words:*** futility analysis, clinical trials, conditional power, stochastic curtailing, interim analysis

Consider a double blind, randomized study with a primary endpoint of subject's time to intervention. The endpoint usually is reached within 10 days of device placement. An interim analysis is planned after approximately 25% of patients have been enrolled in the study and reached the study end time point. With a time-to-event endpoint, the essential information is gained through the number of events in addition to the number of patients enrolled. The purpose of the proposed interim analysis is to determine if the trial is unlikely to achieve success (futility analysis) using conditional power. Conditional power is the probability that the reference test will reject the null hypothesis, given the data observed at the interim point. We will present, through simulations, various methods for futility stopping based on conditional power for the primary endpoint.

### General Statistical Issues and Applications for Survival Analysis in Medical Device Clinical Studies

❖ Chang S. Lao, U.S. Food and Drug Administration, 1350 Piccard Drive, CDHR, Rockville, MD 20850, *Chang.Lao@FDA.HHS.GOV*

***Key Words:*** censoring, sample size, prediction

Survival analysis is used frequently to analyze data from medical device clinical trials, radiation research, or post-approval studies, particularly for long-term follow-up implant studies, such as orthopedic (hip and knee implants), plastic surgery (breast implants), or cardiovascular (pacing lead and cardiovascular stents) devices. Practical applications, rather than theoretical approaches, are discussed in this paper. We discuss sample-size selections based on censoring mechanisms, examples of correlated survival in medical device trials, recurrent events of same or different types, random-effects frailty models, identification of prognostic covariates by Cox proportional hazard regression model, and prediction of future survival with validation based on parametric distributions (Weibull or Gompertz) using actual radiation research data from long-term beagle dog studies.

### Group Sequential Design of Device Trials Using Time-to-Event Analysis with Highly Unbalanced, Recurrent Events

❖ Zengri Wang, Medtronic, Inc., 1015 Gramsie Road, Shoreview, MN 55126, *zengri@gmail.com*; Andrew Mugglin, University of Minnesota

***Key Words:*** correlated data, group sequential design, interim analysis, survival analysis, time-varying effect

In the setting of cardiac rhythm management, events of interest can occur multiple times (e.g., hospitalizations or arrhythmic episodes), and it is of interest to decrease the event density---such as number, rate, hazard, or intensity. In some settings, time-to-event analysis seems to be the most appropriate, but in the setting of trial design, it presents challenges. In this talk, we will discuss a group sequential design for medical device trials for which the primary endpoint is time to all recurrent events. We will describe a procedure for obtaining the sample size while adjusting for imbalance of event density across subjects. Statistical considerations are given to model robustness, time-varying covariate effects, and patient heterogeneity. Examples of design parameters and monitoring plans for efficacy/harm or futility will be provided.

## 495 Industry Applications in Defense and National Security ✪

Section on Statisticians in Defense and National Security, Section on Physical and Engineering Sciences
**Thursday, August 10, 8:30 am–10:20 am**

### Exploiting Ontologies To Enhance Situation Awareness

❖ Steven A. Lien, The Boeing Company, P.O. Box 516, M/C S064-2374, St. Louis, MO 63166-0516, *steven.a.lien@boeing.com*; Matt Easley, Rockwell Scientific

***Key Words:*** ontologies, interaction-module, Bayesian, machine usable

This project creates the first instantiation of an ontology interaction module that will convert ontological knowledge into a form that is machine useable. We propose to make a software tool that creates, modifies, or prunes an "executable ontology." Not only will the module support reasoning in the traditional forward direction of knowledge engineering, where domain knowledge is converted into a form usable by situation awareness systems, but it will support techniques that will learn trends and recognize anomalies from the current operating environment.

### Multi-Sensor Fusion Using Nontraditional Sensors

❖ Edward Wright, Information Extraction and Transport, Inc., 1911 N. Ft. Myer Drive, Suite 600, Arlington, VA 22209, *ewright@iet.com*

***Key Words:*** multi-sensor fusion, Bayesian inference, Bayesian networks, credibility models

Traditional fusion algorithms do a good job of estimating the physical state of entities. Unfortunately, in current and future operational environments, physical characteristics of entities are not sufficient to identify threats and adversaries reliably. Today, it is necessary to extend the fusion process to reason about the cognitive state of the human adversaries. Much of the information available will be from nontraditional sources, including HUMINT from agents and civilians, traditional and web-based news reports, and other open sources. This paper presents an approach for information fusion to reason about human cognitive states from diverse sources. Fusion is accomplished using hierarchical Bayesian inference. Nontraditional sources are integrated through probabilistic credibility models that extend the performance metrics used in traditional sensor models.

### Design of Experiments for Wind Tunnel Testing: an Application on Load Investigation

❖ Julio Peixoto, The Boeing Company, P.O. Box 3707, MC 7L-22, Seattle, WA 98124-2207, *julio.l.peixoto@boeing.com*; Winson Taam, The Boeing Company; I-Li Lu, The Boeing Company

***Key Words:*** optimal design, nonstandard design, airplane, vibration, sample size, linear model

Experiments in wind tunnel have unique features that cannot be addressed with standard statistical design of experiments (DOE) methods. An aspect of DOE that relies on specification of an empirical model of the response is especially useful in the design of wind tunnel experiments. Load of an aircraft is a special type of measurement that relates to duration of each experimental run and is a function of vibration. This presentation addresses challenges in wind tunnel experimentation and analysis of such data. It addresses the design issues related to wind tunnel test, in particular with a load investigation study. We will discuss aspect of modeling in terms of response variable and its relationship to design. We provide a solution to address the sample size determination problem and conclude with remarks regarding this type of experiment.

### Parametric Analysis of Advanced Feature-Aided Tracking with Dynamic Resource Allocation and Management

❖ Thomas C. Bradley, Boeing Phantom Works, P.O. Box 3707, Seattle, WI 98124-2207, *Thomas.C.Bradley@Boeing.com*; Dwight Rousu, Boeing Phantom Works; Jeffery D. King, Boeing Phantom Works; Jeffery D. Musiak, Boeing Phantom Works

***Key Words:*** tracking, kinematic, non-kinematic, allocation, dynamic, parametric

This paper presents performance data from parametric analyses of complete kinematic and nonkinematic state estimation---including ID and Feature-Aided Tracking---of an advanced fielded missile defense tracking application integrated with the AMRFS Resource Allocation Manager (RAM) and applied to both ground and space objects using high-fidelity scenario modeling and simulation. The analyses include performance of track initiation and range convergence in combination with observability and probability of correct association and identification issues while using RAM-coordinated, spatially separated ranging and angle-only sensors.

### Weighted Least Squares Approach To Assess Flight Test with Unbalanced Samples

❖ Winson Taam, The Boeing Company, Applied Statistics, Boeing Math Group, Phantom Works, P.O. Box 3707, MC 7L-22, Seattle, WA 98124-2207, *winson.taam@boeing.com*

***Key Words:*** weighted least squares, vibration data, linear regression

Flight test vibration data is used to assess structural stress of air vehicles. Prespecified maneuvers are designed to represent the flight envelop in which an air vehicle would experience during normal operation. Aeronautic and s vibration measurements are taken over time to characterize and compare improvement on vehicle design. However, certain segments of a maneuver could be more difficult to maintain in flight. Hence, the amount of data in these segments are under-sampled, while other segments are well-sampled. These under-sampled segments could be critical to a vehicle's structural integrity and should be weighted more importantly in any assessment of structural stress. Empirical models based on a weighted least squares method are used to assess design changes on a flight test program. This method is compared with traditional models based on ordinary least squares.

## 496 Statistical Developments in Cancer Surveillance Research ●

Section on Statistics in Epidemiology, Section on Health Policy Statistics, Biometrics Section, ENAR
**Thursday, August 10, 8:30 am–10:20 am**

### Learning from Cancer Incidence: Secular Trend, Lead Time, and Overdiagnosis in Prostate Cancer Screening

❖ Ruth Etzioni, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., Seattle, WA 98109, *retzioni@fhcrc.org*; Donatello Telesca, University of Washington

*Key Words:* lead time, maximum likelihood, surveillance data, PSA screening

Introduction of a cancer screening test leads to a rise in disease incidence followed by a decline. The size and duration of the rise inform about the lead time associated with the test. The lead time is critical in predicting overdiagnosis, which occurs when screening detects disease that would not otherwise have been diagnosed. Previously, we used simulation to make informal inferences about lead time and overdiagnosis in prostate cancer due to PSA screening. In this paper, we develop a likelihood approach that uses surveillance data to formally estimate the lead time due to PSA screening. A key model input is the secular trend in disease incidence, which is the incidence that would be have been expected in the absence of PSA. We use the model to estimate lead time given a projected secular trend, but also to estimate age-specific secular trends given a range of lead times.

### Screening with Virtual Colonoscopy: Should Small Polyps Be Referred for Removal?

❖ Iris Vogelaar, University Medical Center Rotterdam, P.O. Box 2040, Room AE-103, Erasmus MC, Rotterdam, 3000 CA The Netherlands, *i.vogelaar@erasmusmc.nl*; Marjolein van Ballegooijen, University Medical Center Rotterdam; Ann Zauber, Memorial Sloan-Kettering Cancer Center; J. Dik F. Habbema, University Medical Center Rotterdam

*Key Words:* colorectal cancer, virtual colonoscopy, cost-effectiveness, micro-simulation modeling

There has been considerable debate as to whether patients with small- or medium-sized polyps detected at virtual colonoscopy (VC) should be referred for immediate diagnostic colonoscopy. We used the MISCAN-Colon microsimulation model to compare the cost-effectiveness of three regimes of referral for diagnostic colonoscopy after screening with VC: persons with any polyp, polyps >= 6 mm, and polyps >= 10 mm detected. Following up on all persons with polyps detected by VC, independent of size, generates the lowest costs per life year saved compared with restricting follow-up to patients with polyps of 6 mm or 10 mm and larger. This result is robust for most changes in screening test characteristics and natural history assumptions. It therefore seems worthwhile to follow-up immediately all patients with polyps detected by VC with diagnostic colonoscopy.

### Modeling the Impact of Customizing Colorectal Screening by Age, Sex, and Race

❖ Ann Zauber, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street, Room 357, New York, NY 10012, *zaubera@mskcc.org*; Iris Vogelaar, University Medical Center Rotterdam; Marjolein van Ballegooijen, University Medical Center Rotterdam; J. Dik F. Habbema, University Medical Center Rotterdam; Sidney Winawer, Memorial Sloan-Kettering Cancer Center

*Key Words:* colorectal cancer, microsimulation, screening guidelines, customization

Current colorectal cancer (CRC) screening guidelines recommend screening all asymptomatic men and women from age 50 onward. We explored whether customizing guidelines by sex and race would be beneficial. We used a microsimulation model for CRC (MISCAN-COLON) to estimate the number of colonoscopies and life-years gained with different screening ages and intervals. We then recommended the number of screenings and compared the required resources and benefits with customized and noncustomized guidelines by sex and race. Customization increased life-years gained from 169.3 to 187.3 per 1,000 in black men and from 203.6 to 219.4 in black women. The redistribution of colonoscopies from whites to higher-risk blacks slightly decreased the life-years gained in white men from 142.6 to 140.4 life-years per 1,000. The actual benefit would depend on the effect of customization on adherence.

### Short-Term Prediction of Time Series Using Semiparametric Bayesian Techniques

❖ Kaushik Ghosh, New Jersey Institute of Technology, Department of Mathematical Sciences, Room 606 Cullimore Hall, 323 Martin Luther King Blvd., Newark, NJ 07102, *ghosh1@verizon.net*; Ram Tiwari, National Institutes of Health

We present Bayesian local-linear and local-quadratic models for short-term prediction of time series. Dirichlet process priors are assumed in the distribution of the slope and acceleration terms, respectively, to make the model flexible and to accommodate various shapes of the series. Markov chain Monte Carlo techniques are used to obtain predicted values along with prediction intervals. We illustrate the proposed models using mortality data on common cancers in the United States and discuss how one would choose prior parameters to balance model smoothness with flexibility.

## 497 Bayesian Methods and Computation in Finance ✪

Section on Bayesian Statistical Science

**Thursday, August 10, 8:30 am–10:20 am**

### Estimating Macroeconomic Models: a Likelihood Approach

❖ Juan Rubio, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, NE, Research Department, Atlanta, GA 30309-4470, *juan.rubio@atl.frb.org*; Jesus Fernandez-Villaverde, University of Pennsylvania

This paper shows how particle filtering allows us to undertake likelihood-based inference in dynamic macroeconomic models. The models can be nonlinear and/or non-normal. We describe how to use the output from the particle filter to estimate the structural parameters of the model---those characterizing preferences and technology---and to compare different economies. Both tasks can be implemented from either a classical or Bayesian perspective. We illustrate the technique by estimating a business cycle model with investment-specific technological change, preference shocks, and stochastic volatility.

### A Full Information Bayesian Approach to the Evaluation and Estimation of DSGE Models

❖ John Landon-Lane, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, *lane@econ.rutgers.edu*

*Key Words:* DSGE models, MCMC, full information, segmented markets

In recent years, the estimation of DSGE models using Bayesian methods has become commonplace. One important issue in this literature is that

DSGE models are typically degenerate in that there are fewer stochastic terms than the number of endogenous variables in the model. In this paper, I show how all endogenous variables can be used in the calculation of the likelihood function without the need to add shocks to the model. I use a dynamic factor model to approximate a DSGE model and use this to estimate the structural parameters of a segmented markets cash-in-advance monetary DSGE model using MCMC techniques. I compare these estimates to the case of using only a subset of the endogenous variables and to the case of adding nonstructural shocks to the model and find that the impact on the structural estimates is considerable in both a statistical and economic sense.

### Bayesian Selection of Multivariate Stochastic Volatility Models

❖ Antonello Loddo, University of Missouri-Columbia, Department of Statistics, 146 Middlebush Hall, Columbia, MO 65211, *ALoddo@mizzou.edu*

***Key Words:*** multivariate stochastic volatility models, stochastic search model selection, Gilks-Wild adaptive sampler, particle-filters, modified Cholasky decomposition

We consider a multivariate regression model with exogenous and endogenous explanatory variables and time-varying volatilities in the error term. Volatilities are of unknown nature and may be deterministic or stochastic. We propose Bayesian stochastic search Markov chain Monte Carlo (MCMC) algorithms for restrictions on the regression and volatility equations. Efficient parameterization of the time-varying covariance matrices is obtained through modified Cholesky decomposition. We compare two algorithms for volatility simulation, the Gilks-Wild algorithm, and particle filters. We propose a hierarchal approach for selection of the volatility equation's variance components. Numerical simulations show the proposed methods are effective and that they improve the forecasting performance of the model. Empirical applications shed new light on two macroeconomic problems.

### Bayesian Smoothing in Generalized Additive Models

❖ Shawn Ni, University of Missouri-Columbia, Department of Economics, Columbia, MO 65211, *nix@missouri.edu*; Dongchu Sun, Virginia Polytechnic Institute and State University/University of Missouri-Columbia

***Key Words:*** Bayesian, smoothing, vector, state space

We consider a vector-generalized additive model. The smoothing problem can be written as a state-space model for which we give full Bayesian treatment. We discuss the issues of prior elicitation and posterior simulation of the state-space model.

### Leveling the Playing Field: Prior Choice and DSGE Model Comparisons

❖ Marco Del Negro, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, Atlanta, GA 30309, *marcodelnegro@frbatlanta.org*

***Key Words:*** Bayesian analysis, DSGE models

What frictions are important in a DSGE model? In the Bayesian DSGE literature (e.g., Smets and Wouters 2003), this question is answered by computing the posterior odds of the model with and without the friction of interest. The prior distribution for the deep parameters plays a key role in these model comparisons. For some parameters, such as the autocorrelations and standard deviations of the structural shocks, the choice of the prior is not straightforward and can make a difference for model comparison. We provide an approach for choosing the prior that is aimed at leveling the playing field for DSGE model comparisons.

## 498 New Directions in Comparative and Cross-Cultural Survey Methods ● ✪

Section on Survey Research Methods

**Thursday, August 10, 8:30 am–10:20 am**

### Quality Assurance in Comparative Studies

❖ Lars Lyberg, Statistics Sweden, Box 24300, 10451 Stockholm, Sweden, *Lars.Lyberg@scb.se*

***Key Words:*** international surveys, quality control, supervision, continuous improvement, multicultural surveys, survey coordination

In any survey, the design and data collection and processing procedures require control to ensure they are conducted as intended. Survey organizations have mechanisms in place to perform this quality assurance (QA). These mechanisms can be complicated, and, in international and other comparative studies, special control procedures are needed, as the complexity of such studies is so great. Examples of issues include organization of QA activities, choosing among various models for the coordination and control of the survey implementation in various cultures and settings, and choosing QA activities based on varying national levels of funding and administrative and methodological resources. The QA process results should feed back into the survey process, and the feedback operation is likely to involve multiple agencies and survey teams.

### Harmonization by Design

❖ Beth-Ellen Pennell, University of Michigan, Institute for Social Research, 426 Thomspon Street, 402 EP, Ann Arbor, MI 48106-1248, *bpennell@umich.edu*

***Key Words:*** cross-cultural, cross-national, documentation, ex ante harmonization

This presentation will discuss the challenges of ex ante harmonization and documentation in a large-scale, cross-national epidemiological study of mental health conducted in 28 countries, in 35 languages, and with more than 200,000 respondents. Cross-cultural, cross-national survey research is inherently more difficult in both planning and execution than research conducted in one culture or one nation. Layers of complexity are added with variations in sample design; survey content and concept comparability; translation approaches; human subject and ethics review and oversight; interviewer staffing and training; quality control processes and procedures; and local conditions, customs, and context. Each of these issues will be explored within the framework of balancing standardization with appropriate and necessary local adaptations.

### The Effects of Acculturation on Survey Question Comprehension among Latino Respondents in the United States

❖ Timothy P. Johnson, University of Illinois at Chicago, 412 S. Peoria Street, Chicago, IL 60607, *timj@uic.edu*; Allyson L. Holbrook, University of Illinois at Chicago; Young I. Cho, University of Illinois at Chicago

***Key Words:*** cross-cultural, behavior coding, acculturation

Research has documented cultural variability in respondent comprehension and interpretation of survey questions. Less information is available regarding culture-based mechanisms that may underlie these differences. We examine the potential mediating effects of acculturation to host culture on respondent comprehension of survey questions among Mexican Americans and Puerto Ricans. Our hypothesis is that high levels of acculturation to a host culture will minimize disparities in question comprehension across groups. This question is investigated by examining comprehension-related respondent behaviors coded from 273 interviews conducted with Mexican American, Puerto Rican, and non-Hispanic Whites in Chicago. HLM is used to examine whether the level of acculturation among Latino respondents accounts for group differences in comprehension.

### Use of Interpreters in Telephone Interviews

❖ Laura Branden, Westat, 1650 Research Blvd., Rockville, MD 20850, *Laurabranden@westat.com*

***Key Words:*** telephone surveys, interpreters, behavior coding

While the use of interpreters to conduct interviews increases the participation of individuals, concern exists about how their use may affect data quality. As a result, most non-English, non-Spanish speakers are excluded from survey participation. Interpreters were used to help conduct telephone interviews with child care providers for the Early Childhood Longitudinal Study, Birth Cohort. Building on work done by Edwards (2004), 29 telephone interviews conducted using interpreters were recorded. Behavior coding was conducted on the three-way interaction between the interviewer, interpreter, and respondent. Results of the behavior coding show the interpreter successfully interpreted key concepts of the questions 89% of the time and correctly interpreted answers 96% of the time. The use of interpreters reduced the number of cases finalized as nonresponse.

### Cross-Cultural, Cross-National Questionnaire Design

❖ Sue Ellen Hansen, University of Michigan, Survey Research Center, Institute for Social Research, Ann Arbor, MI 48106-1248, *sehansen@umich.edu*

***Key Words:*** questionnaire design, standards, cross-national, cross-cultural

This presentation will discuss issues specific to the design of questionnaires used in cross-cultural and cross-national research. The goals are to provide a theoretical framework for the discussion, drawing on research in communication, linguistics, and other disciplines; provide an overview of what is known about design of cross-cultural/national questionnaires; identify gaps in current theory and research to provide guidance for future research; and establish an international dialogue that will lead toward the development of guidelines and standards for cross-cultural/national questionnaire development.

## 499 Explorations in Behavior and Outcomes

Social Statistics Section

**Thursday, August 10, 8:30 am–10:20 am**

### Shift Workers and Their Spouses' Daily Activities

❖ Anne Polivka, Bureau of Labor Statistics, 1005 Jennifer Lane, Falls Church, VA 22046, *Polivka.anne@bls.gov*

***Key Words:*** shift work schedules, dual income households, time use

Social scientists have long been concerned that working something other than a regular daytime schedule could affect the quality of life of workers. However, relatively little is actually known about how working an alternative schedule affects other aspects of individuals' lives. This paper combines information from the U.S. Bureau of Labor Statistics May 2004 Current Population Survey supplement on work schedules with information from the American Time Use Survey (ATUS) to explore how working an alternative shift affects the amount of time individuals eat, sleep, watch television, attend events, volunteer, and interact with family members or friends both on days that individuals work and on days that they do not. The paper concludes with an examination of whether spouses of non-day workers utilize their time differently than do spouses of day workers.

### A Group-Based Approach to Causal Analysis Using Longitudinal Data with Applications to Gang Membership and Teen Violence

❖ Amelia Haviland, RAND Corporation, 201 N. Craig Street, Pittsburgh, PA 15206, *haviland@rand.org*; Daniel Nagin, Carnegie Mellon University; Paul Rosenbaum, University of Pennsylvania

***Key Words:*** trajectory groups, mixture models, propensity score, optimal matching, covariance adjustment, observational study

Using data from the MontrÈal Study of Boys, the effects on subsequent violence of joining a gang at age 14 are studied controlling for characteristics of boys prior to age 14. The boys are divided into trajectory groups based on violence from ages 11 to 13, and within trajectory groups, joiners are optimally matched to a variable number of controls using propensity scores, Mahalanobis distances, and a combinatorial optimization algorithm. The trajectory groups define meaningful subpopulations where effects may be different, while propensity scores and optimal matching tend to balance twelve baseline covariates. By using between 1 and 7 controls for each joiner, greater efficiency is obtained than by pair matching, with greater bias reduction than is available by matching in a fixed ratio. We develop new efficiency bounds to guide decisions about the structure of the matching.

### Toward Redistricting Rules Based on Housing Density and Indivisible Census Tracts

❖ Tom Belin, University of California, Los Angeles, Department of Biostatistics, 51267 Ctr for Hlth Sciences, Los Angeles, CA 90095-1772, *tbelin@ucla.edu*

***Key Words:*** census, compactness, congressional districts, one person one vote, partisanship, voting rights act

Recent experience shows the political impact of different approaches to drawing congressional district lines. Redistricting necessarily involves statistical considerations, but legal criteria for judging redistricting plans have remained ambiguous. This presentation outlines a framework for redistricting rules that would: (1) treat census tracts (rather than census blocks) as indivisible units; (2) strive for balance between districts not solely in terms of population size but also in terms of housing density; and (3) incorporate geometric compactness into a criterion summarizing the quality of the redistricting plan. While requiring changes in the legal environment of redistricting, such an approach, it is argued, would enhance the competitiveness of congressional elections and would reduce the potential for partisan games in the redistricting process.

### Test Equating for Mixed-IRT Models: an Illustration with 3PL and Generalized Partial Credit Model Using the Stacking-Lord Approach

❖ Yann-Yann Shieh, 28 Adrain Way, Fredericksburg, VA 22405, *yyshieh@cox.net*

***Key Words:*** test equating, IRT

Test equating is essential to reporting progress on standardized and criterion-driven assessments. In many situations, using mixed-IRT models to calibrate the test items is either desirable or inevitable. Currently, few references have demonstrated how mixed-model equating can be accomplished. In this paper, we use three-parameter logistic and generalized partial credit models as an example for test equating with mixed-models. We demonstrate how the Stocking-Lord characteristic curve approach can be use to perform the IRT scale transformation when the common linking items consist of different test types and are calibrated with different IRT models. Data are from a state's grade 3 Reading achievement tests administered in 2003 and 2004. The goal is to equate the 2004 tests to the 2003 tests.

### Oral Translations in Surveys: Issues of Data Quality and Question Design

❖ Janet Harkness, University of Nebraska-Lincoln, Gallup Research Center, 200 North 11th Street, Lincoln, NE , *jharkness2@unl.edu*

***Key Words:*** cross-cultural question design, oral translation, survey error

The paper presents findings from a systematic 2-country study on the effects that using oral translations instead of written translations has on interview interaction, participant burden and on survey error. Research conducted in Germany and Switzerland indicates that validity, reliability and interview communication can be severely impaired. The findings also reveal regular strategies of simplification and omission in orally translated interviews. Oral translation cannot always simply be avoided-some languages do not have a written form and so oral translation or interpreting are the only options to interview in these languages. Thus the findings raise important questions about questionnaire design and source question formulation if these are to be transformed through translation. Key words

### Comparing Internet Users and Nonusers' Behavior toward Market Products

❖ Nuria Diaz-Tena, Mathematica Policy Research, Inc., 600 Alexander Park, Princeton, NJ 08543, *NDiaz-Tena@mathematica-mpr.com*; Robert Benford, TNS-Global

***Key Words:*** web survey, internet user, bias, segmentation

Web surveys are used for obtaining information about consumer habits, so that companies can identify segments of the population with different behavioral attitudes towards market products. Only 70 percent of the American population has access to the internet, therefore the samples are weighted to mimic the US population. We used the Sprint telephone survey conducted by TNS which includes a question to identify which sample cases have internet connections at home or at work. We compared the bias of representing the entire US population with the internet users only. We tested for significant differences among internet users and non internet users towards market trends and compared the results.

### A Policy Tool for Assessing Alcohol Intervention Strategies

Edward Wegman, George Mason University; ❖ Yasmin H. Said, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, *ysaid99@hotmail.com*

***Key Words:*** acute outcomes, DALY, alcohol intervention, stochastic digraph

This paper establishes a modeling framework for alcohol abuse that allows evaluation of interventions meant to reduce adverse effects of alcohol overuse without the financial, social and other costs of imposing interventions that are ultimately ineffective (or even simply not cost effective). The framework is ecological (individual agents and their activities are represented), stochastic (neither individual behavior nor consequences of interventions are certain) and flexible. Constructing the framework involves interactions among the domain science of alcohol studies, statistics, and computer science.

## 500 Imputation of Sample Surveys ●

Section on Survey Research Methods

**Thursday, August 10, 8:30 am–10:20 am**

### Fractional Imputation for Missing Values in Linear Regression Models

❖ Minhui Paik, Iowa State University, 1425 Coconino Road, 308, Ames, IA 50014, *minhui@iastate.edu*; Michael D. Larsen, Iowa State University; Shin-Soo Kang, Iowa State University

***Key Words:*** nonresponse, variance estimation, jackknife, response probability, survey sampling, replication

Sample surveys typically gather information on a sample of units from a finite population. They frequently have missing values for some variables for some units. Fractional imputation using a regression imputation model selects multiple values from the set of observed residuals and produces multiple predicted values for each missing value by adding residuals to a regression prediction. The method assigns fractional survey weights to the imputed values. In this paper, we consider the situation in which some covariate information is missing. In particular, attention is paid to the performance of fractional imputation using a regression imputation model for estimating regression coefficients when missing covariate data is either missing completely at random or missing at random.

### 2010 Census Count Imputation: Research Results Using Spatial Modeling

❖ Robert D. Sands, U.S. Census Bureau, Washington, DC , *robert.d.sands@census.gov*

***Key Words:*** count imputation, spatial modeling, census

After the completion of all Census data collection operations a small number of households remain for which either the count of the number of persons resident, or the status as occupied, vacant or non-existent is not known. In order to produce an accurate count of persons in the U.S., it is necessary to impute this missing data. An investigation of a Spatial Modeling approach suggested by Thibaudeau (2002) for the imputation of missing household size, occupancy or housing unit status was undertaken using 2000 Census household data. The paper will describe aspects of the selection of model variables, the model selection process, and implementation of the models. In addition, the results from this approach for count imputation are compared to the results from a traditional Nearest Neighbor Hot Deck method relative to a Truth Deck of simulated missing household sizes.

### Bias-Reduced Multivariate Imputation: Use of the Locally Adjusted Predictive Mean Matching Method

❖ Masato Okamoto, Statistical Research and Training Institute, 19-1 Wakamatsu-cho, Shinjuku-ku, Tokyo, 162-8668 Japan, *PFC02714@nifty.com*

***Key Words:*** predictive mean matching, nearest neighbor imputation, multivariate imputation, predictive mean neighborhoods, fractional imputation, regression imputation

Donor imputation by the Predictive Mean Matching (PMM) method tends to yield biased estimations in multivariate cases. To improve the PMM imputation, I propose adjustment of imputed value by offsetting difference of (re-estimated) predicted means between the donor and the donee. The re-estimation of predicted means is performed in the enlarged neighborhood of the donee. This locally adjusted Predictive Mean Matching (laPMM) method is not a complete donor imputation method anymore, being in between donor and regression imputation in a sense. Empirical results based on a simulation study show a significant reduction of bias, which can be fully utilized by the fractional method for reducing MSE.

### Imputation of Economic Data Subject to Linear Restrictions Using a Sequential Regression Approach

❖ Caren Tempelman, University of Groningen, Korte Poten 36 A, Den Haag, 2511 EE The Netherlands, *carentempelman@hotmail.com*

***Key Words:*** imputation, truncated regression, conditional distributions

Economic data are subject to many logical linear restrictions, such as for example the fact that most variables need to be non-negative. Standard imputation techniques do not take these restrictions into account and are therefore likely to provide imputations that do not satisfy all restrictions on the data. Additionally, it is difficult to find an appropriate multivariate model as economic survey data usually consist of a large number of variables, which may have several distributional forms and specific properties such as the linear bounds. In this paper the data are modeled through conditionally specified distributions. A sequential regression approach is used to univariately impute each variable, where the regression models depend on the type of variable that is imputed. The performance of this method is assessed by comparing it to other common imputation methods.

### Imputations via Triangular Regression-Based Hot Deck: Methods for Rapid Development of an Allocation Scheme and Preservation of the Covariance Matrix

❖ Scott Susin, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20002, *scott.j.susin@census.gov*

***Key Words:*** imputation, allocation, predictive mean, hot deck

In principle, hot deck imputation methods preserve means and variances, and can also preserve covariances with other variables in the allocation matrix. In practice, dimensionality problems arise quickly as predictive variables are added and allocation matrix cells become small, undermining the hot deck's theoretical advantages. Predictive-mean nearest-neighbor imputation avoids dimensionality problems, but does not preserve the variance. A combination method is described: using the predicted values from a set of triangular regressions to form hot deck matrices. Triangularity allows the inclusion of predictive variables that are themselves subject to non-response. The method enables the rapid development of allocation schemes, eliminates dimensionality problems, and aids in predictor selection. I evaluate the method's ability to preserve the covariance matrix of survey income data.

### Creating Imputation Classes Using Nonparametric Classification Trees

❖ Darryl Creel, RTI International, 312 Trotter Farm Drive, Rockville, MD 20850, *dcreel@rti.org*; Stephen Black, RTI International; Karol Krotki, RTI International; Jeremy Porter, RTI International

***Key Words:*** nonresponse, imputation, CHAID, CART

Virtually all surveys encounter some level of item nonresponse. To address this potential source of bias, practitioners often use imputation to replace missing values with valid values through some form of stochastic modeling. In order to improve the reliabilities of such models, imputation classes are formed to produce homogenous groups of respondents, where homogeneity is measured with respect to the item that is being imputed. A common method used to form imputation classes is CHAID where the splitting rule is based on Chi-squared tests. This paper examines an alternative methodology used to form imputation classes, nonparametric classification trees where the splitting rules are based on the Gini index of impurity. In addition to a brief description of the nonparametric classification tree methodology, comparative examples are provided.

### 2010 Census Count Imputation: Research Results for Alternative Methods

❖ Andrew Kilmer, U.S. Census Bureau, 804 N Wayne St #302, Arlington, VA 22201, *adk129@yahoo.com*

***Key Words:*** imputation, hot-deck, truth deck, spatial modeling, administrative records

The U.S. Census Bureau conducts count imputation to estimate the population of housing units that do not provide such information. In 2000, a hot-deck methodology was used to do count imputation. Since then, the Bureau has tested several alternatives to the traditional hot deck imputation method, including variations on the hot-deck, spatial models, direct assignment from administrative records, and a model that uses data from both administrative records and the census. All methods were tested by running them on a truth deck. Several statistics were computed from these imputation results. Some of these statistics measured accuracy at an aggregate level, while others measured accuracy for individual housing units. This paper examines the results and interpretation of these various statistics.

## 501 Theory and Application of Survival Analysis ● ✪

Section on Statistics in Epidemiology, Biometrics Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

## Proportional Hazards Model with Empirically Estimated Weights

❖ Qing Pan, University of Michigan, Biostat, 1518 Gilbert Court F21, Ann Arbor, MI 48105, *qingpan@umich.edu*; Douglas E. Schaubel, University of Michigan

***Key Words:*** survival analysis, selection bias, inverse probability weighting, weighted Cox model, empirical weights, logistic regression

In observational studies, the study population is often a biased sample of the underlying target population. Weighted Cox models have been developed for settings in which weights are fixed (e.g., chosen by the investigator). However, in many practical settings, the weights are estimated and must be treated as such for accurate inference. We propose a two-stage procedure: weights are estimated through a parametric model fitted to a representative sample from the target population at the first stage; with a weighted proportional hazards model fitted to the biased sample at the second stage. Estimators for the regression parameter and cumulative baseline hazard are proposed. Asymptotic properties of the parameter estimators are derived, with their applicability to finite samples evaluated through simulation. The proposed method is applied to transplant survival data.

## Power Estimation for Clearance of Live Vaccine Virus Using Weibull Survival Models with Interval Censoring

❖ Daniel Zaccaro, Rho, Inc., 6330 Quadrangle Drive, SUITE 500, CHAPEL HILL, NC 27517, *Daniel_Zaccaro@RhoWorld.com*; Barry Eggleston, Rho, Inc.; Susan Lieff, Rho, Inc.; Dennis Wallace, Rho, Inc.; Jon Hanifin, Oregon Health & Science University; Donald Leung, National Jewish Medical and Research Center; Marshall Plaut, National Institute of Allergy and Infectious Diseases; Erica Brittain, National Institute of Allergy and Infectious Diseases; Robert Holliday, Rho, Inc.; Mark Slifka, Oregon Health & Science University

***Key Words:*** vaccine, Weibull model, survival analysis, atopic

A clinical study of an attenuated live vaccine is planned to evaluate immunity of atopic dermatitis (AD) and nonatopic subjects (40 subjects per group), with 5 assessments of viremia up to 14 days postvaccination. Due to interval censoring, comparison of duration of viremia between groups is planned using a Weibull survival model. Power for group comparisons was estimated using simulation studies, assuming Weibull and Cox Proportional Hazards models. A hazard ratio of 0.5 was assumed, but varying sample sizes and extra time points were considered. With 52 subjects per group, an extra postvaccination assessment increased anticipated power to 90%. Without the extra assessment, 70 subjects per group are required for 90% power. The Weibull models had greater power and estimates of the true hazard ratio were unbiased when compared to Cox models. (NIH/NIAID Contract HHSN266200400033C)

## Comparison of Survival Methods and Polytomous Logistic Regression with Competing Risks

❖ Robert Glynn, Brigham and Women's Hospital, Division of Preventive Medicine, 900 Commonwealth Ave, Boston, MA 02215, *rglynn@rics.bwh.harvard.edu*; Bernard Rosner, Harvard Medical School

***Key Words:*** competing risks, survival analysis, proportional hazards, time on study, epidemiologic methods

Both polytomous logistic regression and extensions of the proportional hazards model offer the ability to compare the relationships of risk factors with the development of multiple competing outcomes. We compared the performance of these two approaches in the setting of prospective cohort studies examining risk factors for competing outcomes in several different real datasets. The two approaches give similar answers in the setting of rare diseases, with moderate risk factors, and no relationship of censoring to covariates. Polytomous logistic regression has the advantages of a full likelihood and readily interpretable measures of discrimination and goodness-of-fit. Extensions of the proportional hazards model can better accommodate tied failure times, variable follow-up, and varying underlying hazards for components of the outcome. Both approaches can use time-varying covariates.

## Software for Survival Analysis of Studies Nested within Cohorts To Estimate Relative, Absolute, and Attributable Risks

❖ Hormuzd Katki, National Cancer Institute, 3526 Nimitz Road, Kensington, MD 20895, *hkatki@gmail.com*; Steven D. Mark, University of Colorado Health Sciences Center

***Key Words:*** case-cohort, nested case-control, two-stage design, two-phase design, validation study, epidemiology

In studies nested within cohorts, censored survival outcomes and easy to obtain covariates are observed on everyone, but hard to obtain covariates are observed only on a subsample. Examples include two-stage, case-cohort, and nested case-control designs. Current methodology in these studies focuses on relative risks, but absolute and attributable risks are needed to assess clinical and public health impact. Our software estimates all three risks (standardized for confounders) using Cox models or Kaplan-Meier in studies nested within cohorts. Subsampling must be MAR, so can stratify on any variable observed on everyone, can subsample cases, and can frequency-match. We estimate efficient weights that could exploit surrogates for exposure. Our R software (based on Mark & Katki, JASA, in press) is freely available. We analyze a study of cancer and zinc where only 25% of cohort had zinc measured.

## Population Attributable Risk in the Presence of Cure/Immune Individuals: a Simulation-Based Study

❖ Jayawant Mandrekar, Mayo Clinic College of Medicine, 2665 Boulder Ridge Drive, NW, Division of Biostatistics, Rochester, MN 55901, *mandrekar.jay@mayo.edu*; Melvin L. Moeschberger, The Ohio State University

***Key Words:*** cure models, Cox ph models, PAR, simulations

An important question in the assessment of a clinical trial is to determine whether treatment cures (some) patients. On a similar note, in the public health field, it is important to assess the amount of disease burden that could be eliminated or reduced if risk factors were eliminated or reduced. Cure/immune models, used when long term survivors are present, consist of two parts; an accelerated failure time model for survival time of susceptible individuals and a logistic model for estimating the cure/immune proportion. Our simulations demonstrated that the true PAR is underestimated by a Cox PH model in the presence of cured/immune individuals. In addition, PAR estimates increase (or decrease) with the increasing (or decreasing) levels of the risk factor, but remain approximately unchanged (as expected) with increase or decrease in the cured/immune proportion.

## Comparison of Methods to Model Mortality in the Presence of Time-Dependent Confounders

❖ Ouhong Wang, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, *owang@amgen.com*; Trevor McMullan, Amgen Inc.

***Key Words:*** time dependent confounding, propensity scores, restriction analysis, marginal structural models

In observational studies frequently confounding-by-indication complicates analysis and interpretation of relationship between treatments and outcomes. For example, patients with a more severe form of a disease are at a higher risk of mortality, and they are often exposed to higher doses of certain treatment as well. In order to study the independent relationship between dose and mortality, the researcher must consider the possibility of time dependent confounding: Physicians use previous markers for severity to decide on subsequent drug dose, and drug dose last administered affects the next markers assessment. Here we consider different methods to approach the time dependent confounding problem, including propensity scores, time dependent Cox regression, restriction analysis, marginal structural models, structural nested models, Instrumental Variables, etc.

### Estimating Lifetime Risk Accounting for Baseline Prevalence: High Total Cholesterol in Framingham Heart Study

❖ Michael Pencina, Boston University, 111 Cummington Street, Boston, MA 02215, *mpencina@bu.edu*; Ralph B. D'Agostino, Boston University; Ramachandran S. Vasan, Framingham Heart Study; Alexa Beiser, Boston University; Mark R. Cobain, Unilever Corporate Research

***Key Words:*** disease-free survival, incidence, prevalence, hypercholesterolemia, risk adjustment

The lifetime risk statistic has been successfully applied to estimate and highlight the risk of numerous diseases. It can be estimated as cumulative incidence of new onset adjusted for the competing risk of death. We propose an extension of this methodology which additionally adjusts for baseline prevalence of the disease. The lifetime risk is most commonly calculated based on a follow-up of disease-free individuals which poses no problems when applied to the conditions rare in younger ages. However, hypercholesterolemia is present at substantial prevalence regardless of how early we start the follow-up. Adjusting for baseline prevalence is necessary to avoid serious underestimation of actual risk. We apply this methodology to estimate the long-term risks of high total cholesterol level based on incidence from the Framingham Heart Study and prevalence from NHANES III.

## 502 Survey-Based Estimation V ●

Section on Survey Research Methods

**Thursday, August 10, 8:30 am–10:20 am**

### Weighting an Internet Panel Survey on Drug Use and Abuse

❖ Harpe Gordek, RTI International, 303 Hill Building, 3040 Cornwallis Rd, Research Triangle Park, NC 27709, *gordek@rti.org*; Rick L. Williams, RTI International; Lanting Dai, RTI International

***Key Words:*** internet panels, propensity scoring, weighting

There has been much debate over the growing use of internet panels in survey research. Most panels are not probability based and may produce a sample that is not representative of the population. Weighting by demographic characteristics can account for some differences between the panel and the population, but bias may remain. The propensity scoring method tries to deal with bias by using the likelihood of response to a telephone survey as compared to an internet survey to produce representative results. A recent internet survey on drug use and abuse included questions similar to those of the National Household Survey on Drug Use and Health (NSDUH). Using these questions and the propensity scores in the weighting process, we hoped to produce estimates that would compare to this dwelling unit survey. We found that the weighted internet survey produced very similar results to NSDUH.

### A New Method for Weighting Survey Respondents

❖ Feiming Chen, Spectra Marketing Systems, 9371 Notre Dame Drive, Apt. C, Indianapolis, IN 46240-4172, *feimingchen@yahoo.com*

***Key Words:*** sample survey, poststratification, raking, demographic weighting, singular value decomposition

In sample survey, we often need to adjust the weights of survey respondents via poststratification. Raking (Iterative Proportional Fitting) is typically used to align marginal distributions of the sample to those of the population. However it may converge slowly when there are many categories or empty cells. In this paper we present a non-iterative weighting method based on Singular Value Decomposition. It is computationally fast and may be useful in complex alignment problems. We provide a free R library for interested users.

### Nonparametric Regression with Complex Survey Data

❖ Torsten Harms, Freie Universit‰t Berlin, Ebersstrasse 69, Berlin, 10827 Germany, *torsten.harms@gmx.com*; Pierre Duchesne, UniversitÈ de MontrÈal

***Key Words:*** nonparametric regression, survey sampling, design weights, bandwidth

Nonparametric regression has found widespread application for data that is generated from independent and identically distributed variables. The resulting methods - particularly for bandwidth selection - do however not apply in the context of survey sampling. We will present a modification of the Nadaraya-Watson estimator that does correct for complex survey designs. Using a general superpopulation model, we can show consistency as well as derive analytical expressions for the AMSE which also depends on the sampling design. This allows for the construction of adjustment factors for the optimal bandwidth in order to correct for different inclusion probabilities of elements as well as varying sample size. An empirical study will evaluate this new estimator under different sampling designs and for different populations.

### Combining Survey and Population Data in Inference

❖ Raymond Chambers, University of Wollongong, Centre for Statistical and Survey Methods, The University of Wollongong, Wollongong, 2522 Australia, *rc6@soton.ac.uk*

***Key Words:*** analysis of survey data, analysis of linked data, maximum likelihood

Analysis of survey data does not happen in a vacuum. We typically know more about the target population than just the data observed in the survey. In some cases this extra information can be incorporated via calibration of survey weights. However, model fitting using weights often leads to increased standard errors. Also, weights are usually calibrated to a relatively small set of variables, while population data may be known for many more variables. Here we use the general approach to maximum likelihood estimation for complex surveys described in Breckling et. al. (1994) to develop methods for efficiently incorporating external population information into model fitting using survey data.

In particular, we focus on two simple, but very popular, models fitted to survey data. These are the linear regression model and the logistic regression model.

### Dual Frame Estimation in the National Survey of College Graduates

❖ John Finamore, U.S. Census Bureau, 3905 Stone Mansion Court, Alexandria, VA 22306, *john.m.finamore@census.gov*; David W. Hall, U.S. Census Bureau; Ronald S. Fecso, National Science Foundation

***Key Words:*** dual frame sample design, dual frame estimation, nonsampling error, SESTAT, NSCG

The Scientists and Engineers Statistical Data System (SESTAT) is a system of surveys that provides information about the science and engineering population in the U.S. The largest of the three surveys in the SESTAT system is the National Survey of College Graduates (NSCG). In the 2003 version of the NSCG, the sample design included two sampling frames. One frame included a longitudinal sample that originated in 1993. The second frame was based on Census 2000 long form information. By design, the two frames included an overlap in target population. The dual frame design of the NSCG was incorporated to allow the analysis of nonsampling errors that exist in the target population overlap. This analysis will include defining the overlapping target population, investigating the sources of differences between the frame estimates, and evaluating the methodology and value of integrated estimates.

### Use Sampling Weights in Hierarchical Modeling

❖ Yue Jia, Southern Methodist University, Department of Statistical Science, 144 Heroy Science Hall, 3225 Daniel Avenue SMU PO Box 750332, Dallas, TX 75275-0332, *yjia@smu.edu*; S. Lynne Stokes, Southern Methodist University

***Key Words:*** sampling weights, hierarchical models, NAEP, bias

Large-scale survey data often arise from complex multi-stage designs with known but unequal selection probabilities at each level. Hierarchical model (HM) is widely used to analyze the correlation structure induced by cluster designs. In addition, Sampling weights, the inverse of the selection probabilities, are used to produce unbiased parameter estimators. This paper is concerned with the question of incorporating sampling weights in HM, with the objective of predicting when ignoring the weights will bias results. We develop expressions for the bias of estimators from one-way random effects model. Our study shows that the bias is related to not only the sample size and the population size, but also the covariance between sampling weights and some functions of the random effects. The study is applied to National Assessment of Educational Progress (NAEP) 2003 4th Grade Reading data.

### Anomaly of Survey Weights for American Indians and Alaska Natives

❖ Sunghee Lee, University of California, Los Angeles, 10911 Weyburn Ave., Suite 300, Los Angeles, CA 90024, *slee9@ucla.edu*; John H. Kurata, University of California, Los Angeles; Yan Xiong, University of California, Los Angeles

***Key Words:*** weighting, race classification, American Indians and Alaska Natives

Weighting is a standard practice in probability sample surveys. One of its constraints is that the population distributions come from external data, such as Census. What is available in those data determines the choice of weighting variables. Race and ethnicity are common weighting variables but complex concepts with many ways to be classified. Their classification may alter weighting effectiveness. Its influence may be minimal for large racial/ethnic groups but substantial for small ones. This study examines the effect of using different race/ethnicity variables focusing on the American Indians Alaska Natives (AIAN) in the California Health Interview Survey. First, two types of weighting variables will be examined. Second, a composite method will be proposed to correct for the discrepancy in weighted AIAN population totals. Third, results of three weighting methods will be compared.

## 503 Broad Quality Issues in Organizations

Section on Government Statistics

**Thursday, August 10, 8:30 am–10:20 am**

### Business Process Improvement in the Economic Programs Directorate at the U.S. Census Bureau

❖ Deborah Stempowski, U.S. Census Bureau, 128 Seahawk Lane, Edgewater, MD 21037, *deborah.m.stempowski@census.gov*; Shirin A. Ahmed, U.S. Census Bureau

***Key Words:*** improvements, process, economic, cost, savings

Formed in July of 2004 by senior managers in the Economic Directorate of the U.S. Census Bureau, the purpose of the Business Process Improvement Team (BPIT) was to review business processes associated with Economic Directorate programs to identify processes where improvements can be made that will reduce costs or improve efficiencies, while maintaining quality and timeliness. We began our work identifying potential areas for improvement then, based on these improvement ideas, we did the analysis of existing costs. The BPIT identified the top five major areas for improvement and made final recommendations to management in January of 2005. This paper covers the process to identify potential areas for improvement, recommendations for next steps after areas were identified and an update on the changes that have been implemented since recommendations were made to senior management.

### Assessing the Improvement of Data Quality: Surveys Revised Using Cognitive Techniques

❖ Kara Norman, Energy Information Administration, 1000 Independence Ave., SW, Washington, DC 20585, *kara.norman@eia.doe.gov*; Carrie Hughes-Cromwick, Energy Information Administration

***Key Words:*** cognitive techniques, survey redesign, data quality

When a form is revised using cognitive techniques, does data quality actually improve? It would seem logical, and the common belief is that, of course the data are more accurate following an improvement to the data collection tool. But how can we measure this improvement not only qualitatively, but most importantly, quantitatively? The purpose of this study is to assess the improvement in data quality for a questionnaire which was redesigned based on the results of cognitive interviews. The rather simplistic approach of paired differences is used to determine whether there is a statistically significant change in the data reported before the revision and after the revision. If there is a statistically significant change in the data series, then there is, in fact, a quantitative correlation between questionnaire improvement and ensuing data quality.

### Revisions and Other Changes in Official Statistics: Swedish Observations on Concepts and Terminology

❖ Eva Elvers, Statistics Sweden, Department for Research and Development, Box 24300, Stockholm, 10451 Sweden, *eva.elvers@scb.se*

***Key Words:*** preliminary statistics, final statistics, revision, correction, recalculation

The term revision is often used broadly for any change in a value of released official statistics. Current Swedish work on a more precise meaning of preliminary statistics has been extended to some related concepts and terms. Different types of changes should be distinguished for the benefit of users and producers. Corrections are necessary but unplanned changes, e.g. when influential mistakes are detected. Revisions are planned actions. Final statistics are the only release or the last one in a sequence. Furthermore, new knowledge or new data sources may motivate new estimation procedures and releases, even after the final statistics. Such statistics are called recalculations. Environment statistics and national accounts are typical examples. Preliminary and final statistics can both be corrected and recalculated. Preliminary statistics but not final statistics can be revised.

### Improving the Job Opening and Labor Turnover Survey

❖ Darrell Greene, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Rm 4985, Washington, DC 20212, *greene_d@bls.gov*

***Key Words:*** Winsorization method, interquartile method, outlier

The Bureau of Labor Statistics embarked on the Job Openings and Labor Turnover Survey (JOLTS) in 1999. The JOLTS collects total employment, job openings, hires, quits, layoffs and discharges, other separations, and total separations data. JOLTS initially used the interquartile method to detect outliers for each characteristic. The list of schedules with potential outliers was reviewed by analysts who determined which schedules were to be treated as atypicals because the value of one or more characteristics was not considered representative of the population. There are several statistical issues associated with the initial method. We discuss a new procedure to address these issues. The new approach uses the Winsorization method.

### Data Quality Review at a Small Agency

❖ John Blachere, U.S. Consumer Product Safety Commission, 4330 East West Hwy., Room 708, Bethesda, MD 20814, *jblachere@cpsc.gov*

***Key Words:*** data quality, utility, information value, project management

One of my agency's strategic management goals over the next several years is to improve the accuracy, consistency and integrity of the data contained in several of our important databases. This effort, from initiation through improvement identification to management presentation of Benefits and Costs, is presented and discussed in this paper. I show how I modified standard Data Quality improvement processes used by industry to fit the needs and resources of a small government agency. This adaptation will be presented step by step, with discussion on decisions made and progress to date.

### Tying Web Site Performance to Mission Achievement in the Federal Government

❖ Diane Milleville, U.S. Internal Revenue Service, P.O. Box 2608, Washington, DC 20013, *diane.m.milleville@irs.gov*

***Key Words:*** website performance measures, mission achievement, website metrics

As people rely on websites as their main resource for information gathering, the World Wide Web is increasingly becoming the public face of the federal government. This upsurge in website usage has placed an emphasis on justifying federal websites and showing that web strategies are tied to mission achievement. While this is intuitive for some web managers, it is a difficult task for others, due to the complexity of agency missions. To support this effort, the Web Managers Advisory Council, comprised of web managers throughout the federal government, created a task group to develop guidelines that provide a methodology for tying website performance measures to mission achievement. Through research and interviews of web managers, this group developed an online guide (available at webcontent.gov) that illustrates how organizations can use specific metrics to show mission achievement.

### Results from the International Survey of ASA Members

❖ Amy Luo, Ernst & Young LLP, 1225 Connecticut Ave., NW, Washington, DC 20036, *amy.luo@ey.com*; Roberta Sangster, Bureau of Labor Statistics

***Key Words:*** ASA members, membership survey, international

Earlier this year, the non-USA members of the ASA completed a survey regarding their participation and interests in the ASA. This paper will examine different regions of the world. It will focus on satisfaction based on whether ASA is their primary membership, factors that influenced them to join the ASA, and ratings of different aspects of the ASA (e.g., representation of international members on ASA boards and committees, ASA journals and publications reflects international issues of interest). It will also take a special look at members from developing nations. The internet survey was a census of the non-USA ASA membership (N=2031; 84 did not have email addresses) and had a XX % response rate (n=XXXX). {The study is currently in the field so I will update this abstract later.}

## 504 Bayesian Multiple Testing and Decision Theory

Section on Bayesian Statistical Science
**Thursday, August 10, 8:30 am–10:20 am**

### Capturing Uncertainty When the Event Probability Is Subject to Uncertainty

❖ Ehsan S. Soofi, University of Wisconsin-Milwaukee, School of Business Administration, P O Box 742, Milwaukee, WI 53201, *esoofi@uwm.edu*; Paul C. Nystrom, University of Wisconsin-Milwaukee; Masoud Yasai-Ardekani, George Mason University

***Key Words:*** maximum entropy, prior, posterior, incomplete information

The problem of quantifying uncertainty when an event and its probability are under uncertainty arises in various fields of applications. We present various information theoretic solutions for the problem when the data include a point estimate and/or an interval estimate with or without a confidence level for the event probability. The information theoretic procedures provide measures that enable us to quantify and compare uncertainties for the intuitively clear cases in a precise manner. The entropy analysis also takes into account the trade-offs between confidence level, interval length, symmetry/asymmetry of the interval

about the point estimate, and the closeness of the point estimate to certainty points 0 or 1, as well.

## A New Bayes Rule

❖ James R. Meginniss, Vista Research, Inc., 755 North Mary Avenue, CC101, Sunnyvale, CA 94085, *JRMeginniss@acm.org*

***Key Words:*** utility, gamble, Bayes, prior, posterior, likelihood

A new Bayes' Rule is derived from the author's remarkable symmetric utility rule for gambles (Bus. & Econ. Stat. Section Proceedings of ASA, 1976) -- an expected utility with probabilities^c + Renyi entropy (using same c), c = constant, where p^c means p raised to power c; the limiting case, c-->1, is the Expected Utility + Shannon Entropy Rule. The link between utility (U) and probability (P) is the recognition that U[x] = P[Success | x]. Substitute this into the utility-of-a-gamble rule, allow multiple kinds of Success, and normalize, to get a new marginal probability rule. Standard relationships of joint and marginal probabilities then yield an elegant new Bayes' Rule: P[xi | yj] = P[yj | xi]*(P[xi] / Sum[P[xk, {k, 1, n}]])^c / (P[yj] / Sum[P[yt, {t, 1, m}]] )^c , i=1,., n, j=1, ., m. Becomes usual Bayes' Rule as c-->1. Examples illustrate effects of c upon Bayesian inference.

## Estimating the Positive False Nondiscovery Rate and False Nondiscovery Rate in Multiple Hypothesis Testing

❖ Changzheng Chen, K & L Consulting Services Inc., 3 Birkdale Court, Westampton, NJ 08060, *richard.chen@klconsultingservices.com*; Burt Holland, Temple University

***Key Words:*** multiple testing, FDR, FNR, PFNR, n-value, type II error

We present new results concerning Type II error concepts termed False Non-discovery rate (FNR) and positive False Non-discovery rate (pFNR) appropriate for multiple testing using FDR or pFDR for Type I error control. The application of these concepts depends on the nature of the scientific investigation under study and the characteristics of the family of hypotheses being tested, particularly its proportion of false hypotheses. We specify scenarios in which FNR and pFNR are more appropriately used than other Type II error concepts. For a fixed rejection region we provide point and bootstrap interval estimators of the FNR and pFNR that are applicable for large families. Conditions under which these estimators are conservative are discussed. We introduce the n-value, a pFNR analogue of observed Type II error. The new methodology is applied to gene expression data.

## Decision Theoretic Bayesian Hypothesis Testing with the Selection Goal

❖ Naveen Bansal, Marquette University, 1313 W. Wisconsin Ave., Milwaukee, WI 53201, *naveen.bansal@marquette.edu*

***Key Words:*** selection of populations, Bayes factor, microarray, multiple comparisons, decreasing in transposition property

Consider a probability model with T-parameters(?π,?≤,...,?_{k}). The problem of testing null hypothesis H0:?π=?≤=...=?_{k} against selecting one of k-alternative hypotheses H_{i}:?_{i}=?_{[k]}, i=1,2,...,k, where ?_{[k]}=max{?π,?≤,...,?_{k}}, is formulated from a Bayesian decision theoretic point of view. This problem can be viewed as selecting a component with the largest parameter value if the null hypothesis is rejected. Some interesting loss functions are considered, and Bayes rules are obtained for the k-normal populations under these loss functions. It is demonstrated through this example that the classical hypothesis testing yields unnecessarily large acceptance region and thus unnecessarily fails to reject H0 when the purpose is to find the component with the largest parameter value. Consequences of this for the high dimensional data such as microarray data is pointed out.

## Bayes Approach to Dependent Multiple Comparisons

❖ Lemuel Moye, The University of Texas Health Science Center at Houston, School of Public Health, RAS Building E815, 1200 Herman Pressler, Houston, TX 77025, *moyelaptop@msn.com*

***Key Words:*** Bayes, multiple comparisons, type I error level, dependency

The statistician working in the setting of hypothesis testing has several useful tools. The approach to dependent hypothesis testing that models the conditional probability of the type I error of one hypothesis test in a set, given knowledge of a type I error on another hypothesis test in the same set, is a useful approach. This procedure permits investigators who wish to conserve alpha-level errors in a dependency setting to retain control of the type I error levels for the hypotheses they wish to test. However, this latter approach is hampered by requiring investigators to select the dependency parameter D (0< =D< =1). A probability distribution for D is provided. A closed-form solution for the type I error levels for dependent statistical hypothesis tests is available. An example of the procedure is offered.

## A Bayesian Multiple Comparison Procedure for Order-Restricted Mixed Models

❖ Junfeng Shang, Bowling Green State University, 450 Mathematical Sciences Building, Bowling Green, OH 43403, *jshang@bgsu.edu*; Farroll T. Wright, University of Missouri; Joseph E. Cavanaugh, The University of Iowa

***Key Words:*** hierarchical model, simple order restriction, Gibbs sampling, posterior probability, hypothesis testing, two-way ANOVA mixed model

A Bayesian hierarchical mixed model is developed for multiple comparisons under a simple-order restriction. The model makes inferences on the successive differences of the population means, for which we choose independent prior distributions that are mixtures of an exponential distribution and a discrete distribution with its entire mass at 0. We employ Markov Chain Monte Carlo (MCMC) methods to estimate parameters and to obtain estimates of the posterior probabilities that any two of the means are equal, which allow one both to determine if these two means are significantly different and to test the homogeneity of all of the means. The simulation and application results exhibit that the proposed hierarchical model can effectively unify parameter estimation, tests of hypotheses, and multiple comparisons in one setting.

## Empirical Bayes Analysis on the Power Law Process

❖ Zhao Chen, Florida Gulf Coast University, Department of Physics and Math CAS, 10501 FGCU BLVD S, Fort Myers, FL 33965-6565, *kchen@fgcu.edu*

***Key Words:*** empirical Bayes, power law process, natural conjugate prior

The power law process has been used extensively in software reliability growth models. In our study, we focus on Empirical Bayes (EB) analysis on the power law process by employing parametric EB priors and nonparametric EB priors. For the parametric priors, we apply two-hyperparameter natural conjugate prior and a more generalized three-hyperparameter natural conjugate prior. We also make comparison with Bayesian approach. Since we have past experience about the parameters of the model, we can employ data to estimate the hyperparameters of priors, hence estimate the priors in parametric empirical Bayes.

## 505 Simulation

Section on Statistical Computing, Section on Physical and Engineering Sciences
**Thursday, August 10, 8:30 am–10:20 am**

### Importance Sampling with the Huber and Elliptical Multivariate Logistic Distributions

❖ Stella W. Karuri, North Carolina State University, 2501 Founders Drive, Raleigh, NC 27695-8203, *karuri@stat.ncsu.edu*; Buffy Hudson-Curtis, North Carolina State University; John Monahan, North Carolina State University

***Key Words:*** integration, importance sampling, Huber distribution, multivariate logistic distribution, radial-spherical distributions

Importance sampling in Monte Carlo integration is one method employed in numerical integration of the posterior. We propose two radial-spherical distributions as importance functions, the Huber and the Elliptical Multivariate Logistic (EMVL) distributions. The motivation of the study was to find good importance functions, functions whose tail behavior matches that of the posterior. The appeal of the Huber and the EMVL as importance functions is that it is possible to generate random variables from them and their densities can be evaluated up to a constant. In addition they have tail behavior that decays exponentially. They can also be constructed in such a way that their tail behavior matches to the 'worst' direction of the posterior in spherical co-ordinates. We outline the construction of these distributions and present several importance sampling applications.

### Application of Randomized Spherical-Radial Integration to the E-step of the Monte Carlo Expectation Maximization Algorithm for Generalized Linear Mixed Models

❖ Vadim Zipunnikov, Cornell University, Department of Statistical Science, 301 Malott Hall, Ithaca, NY 14853, *vvz2@cornell.edu*; James Booth, Cornell University

***Key Words:*** MCEM, IRSL, randomized quadratures, spherical radial rule, salamander data

Very often in practice the E-step of the EM algorithm applied to GLMM involves analytically intractable integrals. We show how the randomized spherical-radial (SR) integration can be used in such cases. After a standardizing transformation, a change to polar coordinates allows to represent an integral as a double integral consisting of a one dimensional integral on the real line and a multivariate integral on the surface of a unit sphere. Randomized quadratures are used to approximate both of them. An attractive feature of the randomized SR rule is that its implementation requires only generating from standard probability distributions. Another advantage of the randomized rules is a direct way to construct MC error estimates. The resulting approximation has a form of a fixed effects GLM likelihood and a standard iteratively reweighted least squares procedure may be utilized for M-step.

### An Implementation of Constrained Sequential Monte Carlo by Particle Resizing

❖ Lixin Lang, The Ohio State University, 1958 Neil Ave., Cockins Hall Room 404, columbus, OH 43210, *langl@stat.ohio-state.edu*; Prem K. Goel, The Ohio State University; Bhavik R. Bakshi, The Ohio State University

***Key Words:*** sequential Monte Carlo, particle filter, effective sample size

The Sequential Monte Carlo (SMC) method, also known as particle filtering, has been successfully applied to perform Bayesian analysis in engineering problems. As a recursive simulation method, it allows rapid online estimation for high dimensional nonlinear dynamic models. However, it can give invalid results if ignoring some constraints, which are quite common in many chemical processes, say, a limited range of concentration degree. In the paper, we propose a particle resizing method to make an efficient implementation of SMC in such a problem. Simulation results are also given to demonstrate its performance.

### Continuous Contour Monte Carlo for Marginal Density Estimation

❖ Faming Liang, Texas A&M University, Department of Statistics, College Station, TX 77843, *fliang@stat.tamu.edu*

***Key Words:*** marginal density estimation, contour Monte Carlo, stochastic approximation, Wang-Landau algorithm, autologistic model, kernel density estimation

In this work, we proposed the continuous contour Monte Carlo (CCMC) algorithm as a general algorithm for estimating the marginal density function. CCMC can be regarded as a continuous version of the contour Monte Carlo (CMC) algorithm, proposed in the literature recently. It abandons the concept of sample space partition, on which CMC is based, and makes use of kernel density estimation techniques to improve its efficiency. CCMC was applied to estimate the normalizing constant function for an autologistic model. The estimate was then used in place of the true normalizing constant function in a Bayesian analysis for the model. The superiority of the Bayesian approach for this model justifies the value of CCMC in Bayesian computation.

### Annealing Contour Monte Carlo Algorithm for Real Protein Folding

❖ Sooyoung Cheon, Texas A&M University, Department of Statistics, 447 Blocker Building, 3143 TAMU, College Station, TX 77843-3143, *scheon@stat.tamu.edu*

***Key Words:*** annealing contour Monte Carlo, Markov chain Monte Carlo, protein folding, simulated annealing

Recently, contour Monte Carlo has been proposed by Liang (2004) as a general simulation and optimization algorithm. The algorithm was tested on real protein folding problems. The numerical results indicate that it outperforms simulated annealing, Metropolis Monte Carlo, and some other protein structure prediction approaches proposed in the literature.

### Simulation-Based Estimation

❖ Weijie Cai, George Mason University, 10720 West Drive, Apt. 304, Fairfax, VA 22030, *wcai@gmu.edu*; James Gentle, George Mason University

***Key Words:*** simulation-based estimation, empirical CDF, goodness of fit, random number generator, dynamic systems

Simple distributional models, perhaps based on a PDF, may not be sufficiently complex to adequately model the underlying mechanisms of many systems of interest. Simulation models are often more useful for studying these complex systems. A simulation model, just as a PDF, has parameters, and the statistical task in either case is to estimate those parameters using observed data. We propose estimators that are based on comparisons of the empirical CDF of the observed data and the simulated CDF, which is a function of the parameters of the simulation model. We consider various measures of discrepancy between the ECDF and the simulated CDF. The estimators of the model parameters are defined as the values that minimize the discrepancy. Simulation studies on a Normal random number generator are provided to illustrate our method.

### Solving Complex Network Optimization Problems Using Response Surface Methodology

❖ Ying-Chao Hung, National Central University, Graduate Institute of Statistics, Jhongli, 32049 Taiwan, *hungy@stat.ncu.edu.tw*

***Key Words:*** network optimization, simulation, performance measure, response surface methodology

In designing and analyzing complex network systems, one is particularly interested in how to minimize the service cost by controlling a large number of input factors. This service cost usually corresponds to performance measures (e.g., delays and backlogs) for which analytical solutions are invalid due to the system's complexity. Therefore, one has to rely on simulation. Simulations of large systems are expensive in terms of CPU time and use of available resources. Thus, it is important to select the inputs of simulation carefully in order to minimize the cost function of interest and the required number of simulation runs (i.e., experimental trials). In this study, response surface methodology (RSM) is used to achieve these goals. We briefly summarize the steps of RSM and show it is a powerful tool for solving various network optimization problems.

## 506 Bayesian Statistics and Methods

IMS, Section on Bayesian Statistical Science

**Thursday, August 10, 8:30 am–10:20 am**

### Robust Prior Bayes Estimation on Infinite Dimensional Normal Mean and Spectral Densities

❖ Herman Rubin, Purdue University, Department of Statistics, Purdue University, West Lafayette, IN 47907, *hrubin@stat.purdue.edu*; Hui Xu, Purdue University

***Key Words:*** prior Bayes robustness, empirical Bayes, Bayes risk, nonparametric estimation

We consider the infinite dimensional normal mean problem from the prior Bayes robustness viewpoint. Assumptions are made that the prior variances are decreasing, with some extensions. Under squared error loss, empirical Bayes estimation of the mean vector is derived without knowing the posterior. The effects of various types of decreasing rates are investigated. It is shown that although we cannot estimate the individual prior parameters consistently, the overall risk of using this kind of estimate is still asymptotically optimal compared to the true prior Bayes risk. Similar methods applies to spectral density estimation.

### Geometric Ergodicity of the Gibbs Sampler for the Probit Model

❖ Vivekananda Roy, University of Florida, Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611-8545, *vroy@stat.ufl.edu*

***Key Words:*** data augmentation, convergence rate, drift condition, CLT

Let $\pi$ be the posterior distribution of the regression parameter $\beta$ that results when a probit regression model is combined with a uniform prior on $\beta$. Albert and Chib (1993)proposed a data augmentation algorithm for sampling from $\pi$. We show that the Markov Chain driving this algorithm is geometrically ergodic. This result has important practical ramifications concerning the calculation of standard errors.

### Markov Chain Monte Carlo Approximation of the Posterior in Hierarchical Linear Models

❖ Galin Jones, University of Minnesota, 224 Church Street, SE, Minneapolis, MN 55455, *galin@stat.umn.edu*

***Key Words:*** Markov chain, Monte Carlo, hierarchical model

We consider constructing a Monte Carlo-based approximation to the posterior for a general Bayesian hierarchical linear model. In particular, we provide a method of using simulations from the Markov chain to construct a statistical estimate of the posterior from which it is straightforward to sample. We show that this estimate is strongly consistent in the sense that the total variation distance between the estimate and the posterior converges to 0 almost surely as the number of simulations grows. Moreover, we use some recently developed asymptotic results to provide guidance as to how much simulation is necessary. Draws from the estimate can be used to approximate features of the posterior or as intelligent starting values for the original Markov chain.

### On the Bayesian Detection of a Change in the Arrival Rate of a Poisson Process

❖ Marlo Brown, Niagara University, 343 Dunleavy Hall, Niagara University, NY 14109, *mbrown@niagara.edu*

***Key Words:*** Bayesian, change-point, risk, dynamic programming

In the Bayesian framework, the epoch at which the arrival rate of a Poisson process is assumed is a random variable having a mixed exponential distribution. We consider the case where monitoring the arrival process is at discrete time points and the decision process is based on a sequence of independent Poisson random variables, representing the number of arrivals in the periods between observation points. For a loss function consisting of the cost of late detection and a penalty for early stopping, we develop---using dynamic programming---the one and two steps look ahead stopping rules. We give conditions under which the myopic stopping rule is optima and provide numerical results to illustrate the effectiveness of the detection procedures.

### The Bayesian Lasso

❖ Trevor Park, University of Florida, Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611, *tpark@stat.ufl.edu*; George Casella, University of Florida

***Key Words:*** Gibbs sampler, inverse Gaussian

The Lasso estimate for linear regression parameters can be interpreted as a Bayesian posterior mode estimate when the priors on the regression parameters are independent double-exponential (Laplace) distributions. This posterior can also be accessed through a Gibbs sampler using conjugate normal priors for the regression parameters, with independent exponential hyperpriors on their variances. This leads to tractable full conditional distributions through a connection with the inverse Gaussian distribution. Although the Bayesian Lasso does not automatically perform variable selection, it does provide standard errors and Bayesian credible intervals. Moreover, the structure of the hierarchical model provides both Bayesian and likelihood methods for selecting the Lasso parameter.

### Robust Estimators and Influence Measures of Extremal Dependence

❖ Yu-Ling Tsai, University of Western Ontario, Department of Statistical and Actuarial Sciences, London, ON N6A 5B7 Canada, *ytsai@stats.uwo.ca*; Duncan Murdoch, University of Western Ontario; Debbie Dupuis, HEC MontrÈal

***Key Words:*** asymptotic dependence, Bayesian robustness, contamination, influence measure, multivariate extremes

We achieve robust Bayesian estimation of bivariate extremal dependence through the use of weighted log-likelihood techniques. The new estimator is easy to implement. It can be used for complicated likelihoods such as those involved in multivariate extreme value problems. In the course of assessing whether a Bayesian estimator is robust, we develop a simple influence measure that can be used as a first step in judging the robustness. The proposed measure is easy to compute and successfully captures both the effect of contamination and Monte Carlo uncertainty. We demonstrate our techniques on simulated data.

## 507 Teaching and Technology ●

Section on Statistical Education

**Thursday, August 10, 8:30 am–10:20 am**

### Real-Time Classroom Data Collection

❖ Jeffrey Lidicker, Temple University, Center for Statistical and Information Science, 623 Jones Hall, 3307 North Broad Street, Philadelphia, PA 19140, *lidicker@temple.edu*; Alicia Graziosi, Temple University

***Key Words:*** data collection, survey, chi-squared, regression, descriptive statistics, hypothesis testing

Use of real-time classroom data gives students hands-on experience and personal relevance to statistical analysis procedures. Real-time classroom data is created by using an anonymous survey where students are asked for a short list of personal nonthreatening information. This survey includes both continuous and categorical data so both regression and chi-squared tests can be performed. All the statistical concepts and issues a text covers can be motivated, investigated, and explored using this data. Students gain experience with concepts not covered well by a text, such as false responses, missing values, and data entry error. To maximize interest and motivation, it can be stressed that inferences and estimates gained from the data are all original and that the students actually may know more about the university student body (in these respects) than anyone else.

### Visualizing Hypothesis Testing Concepts through the Power of PowerPoint

❖ Edward Mansfield, The University of Alabama, Box 870226, ISM, Tuscaloosa, AL 35487-0226, *emansfie@cba.ua.edu*

***Key Words:*** introductory statistics, teaching, animation, pedagogy

Hypothesis testing is a difficult concept to teach in a first course in Statistics. Performing the mechanics without developing a full comprehension of the logic is counterproductive to lifelong learning. Computer animation can be a valuable asset in explaining the logic behind many concepts and techniques. As presentation tools become more powerful, instructors have greater flexibility for communicating with students. In this presentation hypothesis testing is demonstrated via confidence intervals; the explanations are not unusual, but the visual impact can help drive home the underlining concepts. Several animated illustrations for in-class use will be demonstrated. Our animations work well in large lecture classes as well as small sections. PowerPoint files will be made available to those who participate at JSM.

### Ways To Increase the Use of Graphs throughout the Introductory Applied Statistics Course

❖ John McKenzie, Babson College, Math Science Division, Babson College, Babson Park, MA 02457-0310, *mckenzie@babson.edu*

***Key Words:*** GAISE, graphical communication, statistical displays, tables, textbooks

The GAISE College Report states that students in an introductory applied statistics course should know “how to graph the data as a first step in analyzing data, and how to know when that’s enough to answer the question of interest” and “how to interpret graphical displays of data - both to answer questions and to check conditions”. It maintains that students should be able to communicate the results of these analyses, which often is done through graphs. Sadly, many textbooks have limited graphical (and tabular) displays after their introduction in a second or third chapter and hence such displays are not reinforced in many introductory courses, even though with today’s technology it is quite easy to generate quality displays and modify them. In this paper some ways to increase the use of graphs and tables for both analysis and presentation in these courses will be presented.

### I Spy (Assessing the Reliability of Eye-Witness Testimony)

❖ Mary Richardson, Grand Valley State University, Department of Statistics, 1 Campus Drive, Allendale, MI 49401, *richamar@gvsu.edu*; Paul Stephenson, Grand Valley State University

***Key Words:*** active learning, descriptive statistics, inferential statistics

We will discuss an interactive activity that allows students to utilize introductory statistical techniques to assess the reliability of eye-witness testimony. Students are shown a video of a crime and are asked to recall specific details pertaining to the crime. Students then apply appropriate statistical techniques on the data collected in order to ascertain whether or not eye-witness testimony can be viewed as being reliable.

### Using Online Videos To Supplement Classroom Instruction

❖ Monnie McGee, Southern Methodist University, 3225 Daniel Ave., Room 144 Heroy, PO Box 750332, Dallas, TX 75275-0332, *mmcgee@smu.edu*; Jing Cao, Southern Methodist University; Robert Skinner, Southern Methodist University; Ian K. Aberle, Southern Methodist University

***Key Words:*** introductory statistics, hybrid instruction, course management system, streaming video

Students taking introductory statistics courses often complain that there are not enough examples given in class. In lieu of working homework problems in class (which is what they really want), we created digital videos of course instructors working problems from the text. The videos were placed on a course website. Half of the students in each of two sections of the course (taught by different instructors) were randomized to access the videos. Although ability to view the videos did not affect the students’ final course grades or exam grades, students who could see the videos said that they were very helpful in doing homework and preparing for exams.

### Using Computer-Aided Learning To Teach Statistics

❖ Tristan Denley, University of Mississippi, 335 Hume Hall, University, MS 38677, *tdenley@olemiss.edu*; Kim Denley, University of Mississippi

Over the last few years there has been much interest in the use of computer aided learning as part of the educational experience in mathematics classes. At the University of Mississippi we have been using technology in a variety of different ways. In this presentation we will present the results of a recent study that compares the educational outcomes for students in Elementary Statistics classes that employed software in two different learning models.

### A Probability Problem on the Number of Loops Formed When Ends of Strings Are Tied

❖ Marepalli Rao, University of Cincinnati, Department of Environmental Health, 3223 Eden Avenue, Cincinnati, OH 45267, *marepalli.rao@uc.edu*; Subramanyam Kasala, The University of North Carolina at Wilmington

***Key Words:*** discrete probability, expected value, variance, distribution, central limit theorem, senior project

We present a problem from our collection which can be used as a senior project. There are n strings. Each string has two ends. Out of 2n ends, two ends are selected at random and tied. If the two ends come from the same string, we will have a loop right away. If not, we will have (n-1) strings with 2(n-1) ends. We now select two ends at random from these 2(n-1) strings. The process continues until we have one string and its ends are tied. Let X be the total number of loops formed. The possible values of X are 1, 2, ... , n. The project is to calculate the expected value and variance of X and determine the distribution of X.

## 508 Robust Statistical Methods ●

Section on Nonparametric Statistics

**Thursday, August 10, 8:30 am–10:20 am**

### Multivariate Spatial Median for Clustered Data

❖ Denis Larocque, HEC MontrÈal, 3000 Chemin de la Cote Sainte Catherine, Montreal, H3T 2A7 Canada, *denis.larocque@hec.ca*; Jaakko Nevalainen, University of Tampere; Hannu Oja, University of Tampere

***Key Words:*** spatial median, multivariate location problem, cluster correlated data, intracluster correlation

We develop the spatial median and its affine equivariant version as companion estimates to the affine invariant sign test proposed recently by Larocque (2003). The asymptotics of the proposed estimates are extended to dependent data and their limiting as well as finite sample efficiencies for multivariate t-distributions are explored. We learn the efficiency of the spatial median suffers less from intracluster dependencies than the mean vector.

### Multivariate Robust Regression Based on General Depth Function

❖ Weihua Zhou, The University of North Carolina at Charlotte, Department of Mathematics and Statistics, 9201 University Blvd, Charlotte, NC 28223, *wzhou2@uncc.edu*

***Key Words:*** multivariate regression, robust, depth function, influence function, breakdown point

A new estimator of the regression parameters is introduced in a multivariate regression model. The affine equivariant estimate is based on the general depth-weighted mean and scatter estimators. The influence function, finite breakdown and asymptotic theorem are developed to consider robustness and limiting efficiencies of this new estimate. The estimate is shown to be fisher consistent with a limiting multivariate normal distribution. The influence function, as a function of the length of the contaminated vector, is shown to be bounded in elliptic cases. The new estimate is highly efficient in the multivariate normal case and it is also highly robust. Simulations are used to consider finite sample efficiencies with similar results.

### Robust Efficient Identification of an Outlying Cell in a Two-Way Layout with Replicates

❖ Nathalie Malo, Genome Quebec Innovation Centre/McGill University, 740 Avenue du Docteur Penfield, Montreal, QC H3A 1A4 Canada, *nathalie.malo@mail.mcgill.ca*

***Key Words:*** robustness, two-way layout, outlier detection, Tukey's median polish, L1 solution, m-estimators

The increasing amount of two-way data in diverse areas has lead to renewed interest in robust methods for detecting outlying cells. One such method, Tukey's median polish, was introduced in the 1970s as a general statistical tool. To overcome the potential lack of uniqueness of the L1 solution, Terbeck and Davies (1998) have developed methods based on M-estimators. The statistical literature provides no technical guidance, however, on how these algorithms should be adapted to handle replicates. A simulation study was performed to compare four options for handling replicates: median polish applied to individual values, median polish applied to cell medians, and two modifications of Terbeck and Davies (1998) methods also applied to cell medians. Results show that median polish applied to individual values had the best performance in detecting an outlying cell in a two-way layout.

### Robust Testing of the Nonparametric Behrens-Fisher Hypothesis Using a Density Ratio Model

❖ James Troendle, National Institute of Child Health & Human Development, Building 6100, Room 7305, Bethesda, MD 20892, *jt3t@nih.gov*; Kostas Fokianos, University of Cyprus

***Key Words:*** Box-Cox transformation, empirical likelihood, likelihood ratio test, Mann-Whitney-Wilcoxon test, power, simulation

Consider the problem of comparing two populations based on a random sample from each distribution, without making any distributional assumptions. A common formulation of this problem is known as the nonparametric Behrens--Fisher problem. To obtain a robust test, we assume a density ratio model holds for a suitable Box--Cox transformation of the data. The transformation, along with the density ratio model, is estimated by maximum empirical likelihood. A new test procedure is developed specifically for testing the nonparametric Behrens-Fisher hypothesis within this framework and its performance is examined by simulations, and on birthweight data of Nulliparous women in spontaneous labor. The test is compared with some nonparametric competitors, and it is found to have relatively high power across a wide variety of distributions including those outside of the density ratio family.

### Robust Nonparametric Confidence Intervals and Tests for the Median in the Presence of (c,r)-Contamination

❖ Masakazu Ando, Japan Society for the Promotion of Science, Nagoya City University, 1 Yamanobata, Mizuho-cho, Mizuho-ku, Nagoya, 467-8501 Japan, *andomasa@econ.nagoya-cu.ac.jp*; Itsuro Kakiuchi, Kobe University; Miyoshi Kimura, Nanzan University

***Key Words:*** robust nonparametric inference, median, confidence interval, sign test, coverage probability, (c,r)-contamination

The problem of constructing robust nonparametric confidence intervals and tests for the median is considered when the data distribution is unknown and the data may be contaminated. A new form of the (c,r)-neighborhood is proposed and it is used in order to describe the contamination of the data. The (c,r)-neighborhood is a generalization of the neighborhoods defined in terms of epsilon-contamination and total variation distance. A modification of the sign test and its associated confidence intervals are proposed, and their robustness and efficiency are studied under the (c,r)-neighborhood of a continuous distribution. The derived results are natural extensions of those in the case of epsilon-contamination. Some tables and figures of coverage probability and expected length for the confidence intervals are given. Further possible generalizations are also discussed.

### Inference in a Simple Random Effects Model with Low Replication and Nonnormal Distributions

❖ Hongjuan Liu, University of California, Riverside, Department of Statistics, 3420 Kentucky Street, Riverside, CA 92507, *liuh06@student.ucr.edu*; Xinping Cui, University of California, Riverside

***Key Words:*** components of variance, nonparametric

We studied the nonparametric inference on the group effect in a random effect components of variance model when the number of groups diverges without bound, but the replications remain fixed as small as 2.With normality assumption, the exact F-test can infer the existence of group effect. When normality assumption is not valid, it has been shown that the F-statistic is robust only in a balanced design and the asymptotic distribution of the F-statistic requires the existence of the fourth order moment. In this work, we pushed the frontier to a very general setting. Our new method only requires the existence of the second order moment and can be applied to balanced or unbalanced data under skewed or heavy-tailed distribution. Our simulations show that the proposed method is very powerful and computationally efficient. We also show its application in microarray-based heritability analysis.

### On the Estimation of Disability-Free Life Expectancy

❖ Kosuke Imai, Princeton University, Department of Politics, Princeton University, Princeton, NJ 08544, *kimai@princeton.edu*; Samir Soneji, Princeton University

***Key Words:*** demography, morbidity, mortality, stationarity, Sullivan's method, life tables

A rapidly aging population is characterized by the increased prevalence of chronic impairment among the elderly. An important question is whether additional years of life are spent in good health. Robust estimation of disability-free life expectancy (DFLE) is essential for addressing this question. Thirty years after publication, Sullivan's method still remains the most widely used method for estimating DFLE when large-scale longitudinal data are not available. Therefore, it is surprising to note that Sullivan did not provide any formal justification of his method. In this paper, we prove that under stationarity assumptions, Sullivan's estimator is unbiased and consistent. This resolves the debate in the literature which has generally concluded that additional assumptions are necessary. We also show that Sullivan's method can be extended to estimate DFLE without stationarity assumptions.

## 509 New Methods and Practical Application ●

Section on Quality and Productivity, Section on Physical and Engineering Sciences

**Thursday, August 10, 8:30 am–10:20 am**

### Robust Designs for One-Way Random Effects Model Using Q-estimator

❖ Julie Zhou, University of Victoria, Department of Math and Statistics, University of Victoria, Victoria, BC V8W 3P4 Canada, *jzhou@uvic.ca*; Xiaolong Yang, University of Victoria

***Key Words:*** robust design, robust estimation, breakdown point, one-way random effects model, q-estimator

Optimal designs for one-way random effects model usually are studied based on ANOVA estimation, maximum likelihood estimation, or modified maximum likelihood estimation. Because those estimation methods are sensitive to outlier observations in the experiment, the optimal designs are not robust. We introduce a robust estimator, Q-estimator, to estimate variance components for one-way random effects model and study its robustness properties. In particular, finite sample breakdown points and efficiency are investigated. Based on Q-estimator, we propose a robust design criterion to construct optimal designs. These designs provide reliable and efficient estimates for variance components with possible outliers generated from the experiment. These designs also are compared with robust designs derived from M-estimator. Guidelines will be given to construct robust designs for applications.

### Beta-Geometric Distribution in Survival Modeling

❖ Alfred Akinsete, Marshall University, Department of Mathematics, 1 John Marshall Drive, Huntington, WV 25755, *akinsete@marshall.edu*

***Key Words:*** probability, distribution, manpower, fecundability, beta-geometric, modeling

We discuss the conditional distribution of a geometric random variable, where the parameter itself has a beta distribution. The mixture of these distributions is referred to as the beta-geometric distribution (BGD). We obtain a generalized expression for the moments, with a recursion noted in the probability function. The application of the distribution in fecundability and manpower or contractual modeling is investigated. We obtain estimates of the survival and hazard functions of the fitted model. The method of the maximum likelihood estimates is discussed, and estimates of the parameters of a set of data are obtained.

### Optimal Targeting under an Asymmetric Loss

❖ Yvonne Zubovic, Indiana University Purdue University Fort Wayne, 2101 Coliseum Blvd., E., Fort Wayne, IN 46805-1445, *zubovic@ipfw.edu*; Chand K. Chauhan, Indiana University Purdue University Fort Wayne

***Key Words:*** loss function, optimal value

Traditionally a symmetric quadratic loss function is used in measuring the quality of a product. In such a case the desired target value of a characteristic of the product is equal to the optimal value. In this paper the authors consider an asymmetric loss function. The loss associated with a deviation from the optimal value in one direction is equal to ? times the loss associated with a deviation in the other direction. In such a situation it is necessary to adjust the target value to achieve the minimum loss. The authors examine the relationship between the amount of adjustment and other factors such as the underlying distribution, the unknown parameters, the form of the loss function, and the value of ?.

### Adaptively Trimmed L-Moments with Applications to Heavy-Tailed Distributions

❖ Jonathan Hosking, IBM Research, P.O. Box 218, 1101 Kitchawan Road, Yorktown Heights, NY 10598, *hosking@watson.ibm.com*

***Key Words:*** estimation, heavy-tailed distributions, l-moments, order statistics

L-moments are statistics that enable summarization of data samples and parameter estimation of probability distributions using linear combinations of order statistics. Elamir and Seheult (2004) introduced trimmed L-moments, which exclude one or more extreme order statistics and can be used for inference about heavy-tailed distributions for which the ordinary L-moments may not exist. However, the appropriate degree of trimming is usually not clear a priori. Here I define adaptively trimmed L-moments, in which the degree of trimming is suggested by the data values themselves. In particular for the generalized Pareto distribution, a simple relation between expectations of fractional order statistics can be used to derive estimators that remain valid no matter how heavy the tail of the distribution. I define these estimators and explore some of their theoretical and practical properties.

### Goodness-of-Fit Testing and Pareto-Tail Estimation

❖ Yuri Goegebeur, University of Southern Denmark, JB Winslows Vej 9B, Department of Statistics, Second floor, Odense, DK5000 Denmark, *yuri.goegebeur@stat.sdu.dk*; Jan Beirlant, K.U. Leuven; Tertius de Wet, University of Stellenbosch

***Key Words:*** extreme value index, quantile-quantile plot, kernel statistic, goodness-of-fit

In this contribution, a general kernel goodness-of-fit test statistic for assessing whether a sample is consistent with the Pareto-type model is introduced. The derivation of the class of statistics is based on the close link between the strict Pareto and the exponential distribution and puts some of the available goodness-of-fit procedures for the latter in a broader perspective. The limiting distribution for this general kernel statistic will be derived under mild regularity conditions and important special cases will be investigated in greater depth. The relation between goodness-of-fit testing and the optimal selection of the sample fraction for tail estimation (e.g., using Hill's estimator) is examined. The methodology is illustrated with a practical study.

### New Methods Using Levene-Type Tests for Hypotheses about Dispersion Differences

❖ Xiaoni Liu, North Carolina State University, Department of Statistics, 601 Jones Ferry Road, Apt F5, Carrboro, NC 27510, *xliu@ncsu.edu*; Dennis A. Boos, North Carolina State University; Cavell Brownie, North Carolina State University

***Key Words:*** Levene type test, RCB designs, bootstrap, robustness, dispersion differences, Huber m-estimation

Levene type tests are well known to be robust tests for equality of scale in one-way designs. Currently, the available Levene type tests for two-way designs with one observation per cell (i.e., RCB designs) employ either standard ANOVA F tests on the absolute values of least squares (OLS) residuals, or weighted least squares (WLS) ANOVA F tests on the OLS residuals. In this paper, we develop new methods based on Levene type tests for two-way RCB designs that can assess equality of variances across blocks and/or treatments. We use OLS ANOVA and WLS ANOVA F tests on the absolute values of residuals obtained from models fit by Huber M-estimation. We also apply bootstrap methods to these Levene type tests and compare these Levene type tests in terms of robustness and power by simulation. In result, the new methods compare favorably with older methods.

### On Goodman and Kruskal's G and Stuart's Measure of Association

❖ Jeffrey Green, Ball State University, Economics Department, Ball State University, Muncie, IN 47306, *jgreen@bsu.edu*

***Key Words:*** contingency table, measures of association, cross classification

Stuart's measure of association for a two-way contingency table in which the rows and columns have a natural ordering is compared to Goodman and Kruskal's G. It is shown that the absolute value of G is always greater than or equal to the absolute value of Stuart's measure. The asymptotic variance of Stuart's 1953 measure is derived along with an upper bound for it that is stricter than the one originally presented by Stuart. Finally, it is seen that either measure can have the larger estimated upper bound for its asymptotic variance. However, the estimated upper bound for the asymptotic coefficient of variation of G is shown to never be larger in absolute value than the estimated upper bound for the asymptotic coefficient of variation of the measure suggested by Stuart.

## 510 Microarrays ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

### Validation of Biomarkers Identified by Gene Expression Profiles

❖ Boris Zaslavsky, U.S. Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852, *Boris.Zaslavsky@fda.hhs.gov*; Jing Han, U.S. Food and Drug Administration; Jawahar Tiwari, U.S. Food and Drug Administration; Raj K. Puri, U.S. Food and Drug Administration

***Key Words:*** logistic regression, gene expression, perturbation, leave-one-out cross-validation

Two hundred ninety three human embryonic kidney (HEK) cells were cultured under different confluence states and RNA isolated and labeled targets were hybridized with high quality ~10K cDNA microarrays. The logistic regression method of SAS software was applied to identify a set of biomarkers by gene expression profiles. Two-group classification (i.e., over confluence vs. 90% confluence) was used to select the subset of differentially expressed genes. To evaluate the robustness or stability of the logistic classifier, we generated simulated datasets by perturbation of the raw dataset with various error levels. We also used leave-one-out cross-validation procedure and perturbed datasets to estimate the subset of most informative genes. Results confirmed that simulated datasets based on the perturbation of the raw data can be used for the validation of biomarkers.

## Gene expression data analysis using the Gene Ontology

❖ Jiajun Liu, North Carolina State University, 3009 Dorner Circle, Apt. C, Raleigh, NC 27606, *jliu6@stat.ncsu.edu*; Jacqueline Hughes-Oliver, North Carolina State University; Alan J. Menius, GlaxoSmithKline

***Key Words:*** gene expression data, gene aggregation, gene ontology, data mining, interpretability

New technologies for biological systems give scientists the ability to measure thousands of gene expression for bio-molecule, including genes, proteins, and lipids. Our goal is to use domain knowledge (e.g., Gene Ontology) to guide analysis of such data. By focusing on domain-aggregated results at the molecular function level, increased interpretability is available to biological scientists beyond what is possible if results are presented at the gene level. We use a "top-down" approach to perform domain aggregation by combining gene expression before testing for differentially expressed patterns. This is contrast to the more standard "bottom-up" approach, where genes are tested individually, then aggregated by domain knowledge. Our method is assessed and compared to other methods using a series of simulation studies. Implications from analysis of a real dataset are presented.

## Microarray, PCR, and Northern Blot: a Comparison

Yongzeng Ding, Northwestern University; ❖ Borko Jovanovic, Northwestern University, 680 N. Lake Shore Drive, Preventive Medicine, Suite 1102, Chicago, IL 60611, *borko@northwestern.edu*; Raymond Bergan, Northwestern University; Irene Helenowski, Northwestern University

***Key Words:*** prostate, cancer, reproducibility, lab techniques, gene expression

We provide a comparison of gene expression measurements (from native RNA) among Microarray (MA), quantitative polymerase chain reaction (PCR) and Northern blot (NB) measurements on two cell lines (PC3, PC3M) with two measurements each. Measurements are made on twenty-six "randomly selected" genes. In addition, we have two MA and PCR measurements of same genes from amplified RNA. Using NB as gold standard we describe and discuss variability and co-relation observed in these measurements.

## Technical, Longitudinal, and Genetic Variation in Proteomic Analysis of Human Plasma

❖ Imola K. Fodor, Lawrence Livermore National Laboratory, 7000 East Ave., MS L448, Livermore, CA 94551, *fodor1@llnl.gov*; Todd H. Corzett, Lawrence Livermore National Laboratory; Megan Choi, Lawrence Livermore National Laboratory; Vicki L. Walsworth, Lawrence Livermore National Laboratory; Kenneth W. Turteltaub, Lawrence Livermore National Laboratory; Sandra L. McCutchen-Maloney, Lawrence Livermore National Laboratory

***Key Words:*** proteomics, gel electrophoresis, technical variation, genetic variation, biomarker, plasma

While human plasma is an ideal sample for the discovery of disease biomarkers, its complexity presents a number of challenges. To date, no results have been published on the technical variation of two-dimensional difference gel electrophoresis (2D DIGE), or on the longitudinal and genetic variation of human plasma. We address this gap by presenting: 1) a technical variation study where one human plasma sample was processed and analyzed on 12 replicate gels, and 2) a longitudinal and genetic variation study where plasma from 11 subjects collected at 3 time points were analyzed in triplicate on 100 gels. Estimates of the spot-wise protein variance (overall, as well as its components such as dye, subject, and random error effects) quantify the sources of variation in 2D DIGE, and thus lay the foundation of future disease biomarker detection studies with human plasma. UCRL-ABS-218476.

## A Valid False Discovery Rate Procedure in Presence of Biased Null P-values

❖ Hoa Phuong Nguyen, Johnson & Johnson, ALZA Plaza, 1950 Charleston Road, M11 3B, Mountain View, CA 94043, *hnguye19@alzus.jnj.com*

***Key Words:*** FDR, SNPS, microarray, genome wide disease association

In recent years, genome wide disease association effort has been focused on (1)identifying gene expression differentials using microarray technology and (2)searching through millions of single nucleotide polymorphisms (SNPs) for liability alleles affecting complex diseases. Due to a large number of hypotheses tested to assess association, the FDR procedure introduced by Benjamini and Hochberg (BH) has become mainstream in controlling the combined type-I error rate. The authors showed validity of the BH procedure assuming uniform p-values of null hypotheses. In this talk, I discuss genome wide disease association scenarios in which biased (non-uniform) null p-values can arise and how biased p-values degrade performance of the BH procedure. I propose the tcFDR procedure which improves upon the performance of the BH procedure in presence of biased null p-values.

## A Sequential Monte Carlo EM Solution to the Transcription Factor Binding Site Identification Problem

❖ Edmund Jackson, Cambridge University, Trinity College, Cambridge, CB21TQ UK, *ej230@cam.ac.uk*; William Fitzgerald, Cambridge University

***Key Words:*** sequential Monte Carlo, transcription factor binding sites, expectation maximization, sequence analysis

A significant and stubbornly intractable problem in genome sequence analysis has been the identification of novel transcription factor binding sites in promoter regions. Probabilistic methods have faced difficulties from prior ignorance and poor models of the biological sequence. In addition, this inference problem occurs in an extremely irregular, high dimensional space. In order to overcome this problem we propose utilizing an iterated particle optimization in place of the standard Gibbs sampling approach. We derive and demonstrate this novel method and show improved convergence to the global mode.

## Statistical Issues in High-Throughput Screening

❖ E. Venkatraman, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., Box 44, New York, NY 10021, *venkatre@mskcc.org*; Hakim Djaballah, Memorial Sloan-Kettering Cancer Center

***Key Words:*** HTS, drug development

High throughput screening is the modern approach to speed up drug development. Once an assay developed by a scientist is adapted to the high throughput process, there are several steps that eventually lead to the identification of potential drug targets. We will present the data generated in this process and the methods for their analysis.

# 511 Power and Sample-Size Calculations ●

Biometrics Section

Thursday, August 10, 8:30 am–10:20 am

## Quick Calculation for Sample Size While Controlling False Discovery Rate with Application to Microarray Analysis

❖ Peng Liu, Cornell University, 2605 Hasbrouck Apt., Ithaca, NY 14850, *PL61@cornell.edu*; J. T. Gene Hwang, Cornell University

***Key Words:*** FDR, sample size calculation, microarray, q-value

Sample size calculation is important in microarray or proteomic experiments since only a few repetitions can be afforded. In the multiple testing problems involving these experiments, it is more powerful and more reasonable to control false discovery rate (FDR) or positive FDR (pFDR) instead of type I error, e.g., family-wise error rate (FWER). When controlling FDR, the traditional approach of estimating sample size by controlling type I error is no longer applicable. Our proposed method applies to controlling FDR. The sample size calculation is straightforward and requires minimal computation, as illustrated with two sample t-tests and F-tests. Based on simulation with the resultant sample size, the power is shown to be achievable by the q-value procedure of Storey, Taylor and Siegmund (2004).

## Sample Size for FDR-Control in DNA Microarray Studies

❖ Yongzhao Shao, New York University, Division of Biostatistics, 650 First Ave 5th Floor, New York, NY 10016, *shaoy01@med.nyu.edu*; Chi-Hong Tseng, New York University

***Key Words:*** bioinformatics, genomic studies, experimental design, false positive rates, DNA microarray, false discovery rate

DNA microarrays have been widely used for the purpose of monitoring expression levels of thousands of genes simultaneously and identifying those genes that are differentially expressed. The probability that a false identification is committed can increase sharply when the number of tested genes gets large. Correlation between the test statistics attributed to gene co-regulation and dependency in the measurement errors of the gene expression levels further complicates the problem. This paper introduces a general approach to estimate the sample size needed to ensure adequate overall power while controlling the false discovery rates (FDR) to avoid an abundance of false positive results. The proposed approaches allow dependence among test statistics.

## Sample-Size Determination for Multiple Comparisons

❖ Chi-Hong Tseng, New York University, Division of Biostatistics, 650 First Ave, Fifth Floor, New York, NY 10016, *ch.tseng@med.nyu.edu*; Yongzhao Shao, New York University

***Key Words:*** false discovery rate, family-wise error rate

Multiple comparisons are now commonly made in genetic research and many other studies. An appropriate sample size is crucial for the success of such studies, however, it typically depends on other design parameters that have to be estimated based on limited preliminary data. In this short note, we present a simple sample size formula for simultaneous testing of multiple hypotheses. This formula can be used to estimate the sample sizes required to provide adequate power while effectively controlling the family-wise error rate (FWER) or the false discovery rate (FDR). This simple formula also elucidates the relationship between the required sample sizes and other design parameters, such as the number of comparisons to be made.

## Power Calculations for Linear Mixed Effects Models Using SAS/PROC MIXED

❖ Andrzej Galecki, University of Michigan, 300 North Ingalls Building, Geriatrics Center and Institute of Gerontology, Ann Arbor, MI 48109, *agalecki@umich.edu*; Tomasz Burzykowski, Hasselt University

***Key Words:*** linear mixed effects models, fixed effects, approximate f test, power calculations, statistical software

Contrasts of the fixed effects in the linear mixed effects models can be tested using approximate F test involving REML estimation of regression coefficients and of covariance parameters. Helms (1992) developed an analytical method for computing approximate power for the approximate F test. The non-null distribution of the test statistics is approximated by a non-central F distribution with appropriate noncentrality parameter. The aim of this presentation is to illustrate how to implement power calculations using standard statistical software for linear mixed models, such as PROC MIXED in SAS. The advantages of using PROC MIXED include: 1.performing power calculations and analysis itself using the same software, and 2. substantial flexibility in performing power calculations under various missing data mechanisms.

## Power Analysis for Longitudinal Study Designs

❖ Xin Tu, University of Rochester, Department of Biostatistics and Computational Biology, Rochester, NY 14623, *xin_tu@urmc.rochester.edu*; Wan Tang, University of Rochester

***Key Words:*** asymptotic distribution, correlation analysis, generalized estimating equations, linear mixed-effects models, Markov process, missing data

In recent years, longitudinal study designs have became increasingly popular in biomedical, pharmacologic and psychosocial research. However, existing methods for power analysis for longitudinal and other clustered study designs are very limited and unable to sufficiently address the key issues associated with such study designs. In this talk, we review existing methods as well as present some recent development on this topic, particularly in the area of models for regression and correlation analyses. For regression analysis, we focus on the two most popular approaches for clustered data analysis, the generalized estimating equations and the linear mixed-effects models. For correlation modeling, we focus on the product-moment correlation. The discussion is illustrated with examples that are motivated by real study designs.

## Accuracy of P-Values and Sample Size in Comparing Skewed Clinical Trial Data

❖ Jun Zhao, Organon, 56 Livingston Ave., Roseland, NJ 07068, *j.zhao@organonusa.com*; Gang Li, Johnson & Johnson

***Key Words:*** p-value, skewness, sample size, type-I error, tail probability, edgeworth expansion

In drug development, one regulatory requirement is to have the Type I error controlled to evaluate the effectiveness of a study drug. For relatively large samples, the p-values are often obtained by using normal approximation in comparing two treatment groups. Theoretically the

skewness along with kurtosis of data hinders the accuracy of p-values, even though the accuracy can be improved by increasing sample size. Practically a relatively small sample sizes needed when apply a balance design even for skewed non-normal data. A question is how accurate of p-values and how large sample sizes needed to get certain accuracy level when an un-balanced design applied. Theoretical approximation and simulation are used to assess the accuracy of p-values from normal approximation and the corresponding sample needed is also examined for both one-sided and two-sided tail probabilities.

### Does Pair-Matching on Baseline Measures Improve Power in a Pre-Post Cluster Randomized Trial?

❖ Misook Park, Virginia Commonwealth University, 1101 E Marshall Street, Dept of Biostatistics, Richmond, VA 23298-0032, *mipark@vcu.edu*; Robert E. Johnson, Virginia Commonwealth University

***Key Words:*** cluster randomized design, blocking, pre-post design, order statistics

The variance between cluster units may be controlled in part by matching prior to randomization. Clusters may be placed into blocks by matching on known characteristics such as demographics, cluster size, etc. The baseline measure of the study's primary outcome is oft recommended for matching. One matching scheme is to first sort clusters on their baseline means and match the two clusters with the smallest means into a block, match the two clusters with the next smallest means, etc. Does this matching strategy decrease the variance and improve the power to detect treatment effects? Standard methods involve using the baseline mean as a covariable or analyzing pre-post differences. Does the addition of matching improve these methods? We investigate this question in light of the variance components, ICC, pre-post correlation, and allocation of within-cluster sample size across time points.

## 512 Methodology and Applications Based on Mixed Models ●

Biometrics Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

### A Two-Step Logistic Regression-Linear Mixed Model Method to Calculate CpG Island Methylator Phenotype (CIMP) Scores in Colorectal Cancer Patients

❖ Wei Wei, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *wwei@mdanderson.org*; Jeffrey S. Morris, M. D. Anderson Cancer Center; Jean-Pierre Issa, M. D. Anderson Cancer Center

***Key Words:*** mixture distribution, logistic regression, linear mixed model, CIMP score

CpG island regions of multiple cancer-specific genes are methylated in CIMP+ patients. An Arizona colorectal cancer methylation study yielded methylation data for up to 4 genes for over 800 patients. Raw methylation scores for each gene consisted of zeros, percentages and missing data, with several different platforms used to measure methylation for some genes. In order to assess overall CIMP status, summarizing propensity for methylation, a single score for each patient is needed. The naive aggregation methods used in existing literature ignore gene-gene and platform-platform variation in methylation patterns, and fail to appropriately handle missing data. We propose a model-based CIMP score constructed based on a two-step, logistic regression-linear mixed model method. These model-based CIMP scores were more strongly correlated with known clinical factors than naive CIMP scores.

### A Mixed Model Analysis of Errors in Radiotherapy

❖ Alai Tan, The University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-1148, *altan@utmb.edu*; Giuseppe Sanguineti, The University of Texas Medical Branch; Daniel H. Freeman, The University of Texas Medical Branch

***Key Words:*** mixed model, error estimation, radiotherapy

The accuracy of radiotherapy is limited by various error sources introduced during treatment preparation and execution. The estimation of errors is critical to generate a safety margin to improve the effectiveness of radiotherapy. Although methods for quantifying the errors were well established, to our knowledge, mixed model methods have never been reported. The present study used a mixed model to estimate errors in radiotherapy of a sample of patients with prostate cancer. We found that the mixed model analysis could match the systematic test almost exactly. This suggests the mixed model approach provides a standard methodology, which is well understood and has the advantage of accounting for within-patient correlation, in estimating errors in radiotherapy.

### Flexible Random Intercept Models for Binary Outcomes Using Mixtures of Normals

❖ Ming-Wen An, Johns Hopkins University, 3016 Saint Paul Street, Apt. 2, Baltimore, MD 21218, *man@jhsph.edu*; Brian Caffo, Johns Hopkins University; Charles Rohde, Johns Hopkins University

***Key Words:*** probit-normal, logit-normal, marginalized multilevel models

Random intercept models for binary data are useful for addressing between-subject heterogeneity. Unlike linear models, the non-linearity of link functions used for binary data force a distinction between marginal and conditional random intercept models. This distinction is blurred in probit models with a normally distributed random intercept because the resulting model implies a probit marginal link as well. We explore another family of random intercept models with this "closure" property. In particular, we consider instances when the distributions associated with the conditional and marginal links and the random effect distribution are mixtures of normals. We relate this flexible family of models to others in the literature and show the associated computational benefits. A diverse series of examples illustrates the wide applicability of the approach.

### Generalized Linear Mixed Models with Sparse Binary Outcome Data: Comparing Estimation Methods

❖ Marie-Eve Beauchamp, McGill University, 1020 Pine Ave., W., Room 16-C, Montreal, PQ H3A 1A2 Canada, *marie-eve.beauchamp@mail.mcgill.ca*; Robert W. Platt, McGill University; James A. Hanley, McGill University

***Key Words:*** generalized linear mixed models, sparse data, estimation methods, penalized quasi-likelihood

The limits of estimation methods for generalized linear mixed models that are available in commercial software packages have not been studied systematically. The penalized quasi-likelihood (PQL) is among the

most commonly used. However, PQL may produce severely biased variance component estimates for binary outcome data where the number of occurrences is low within each cluster. We compare via simulations the performance of PQL and other estimation methods, e.g. adaptive Gauss-Hermite quadrature, that are available in commercial packages and that presumably produce more accurate results than PQL. We focus on binary outcomes with a low number of occurrences. The bias in the parameter estimates is quantified in several settings. Data and model characteristics influencing the bias are also studied. Results to date will be presented.

### Baseline Adjustment: Issues for Mixed-Effects Regression Models in Clinical Trials

❖ Ronald Thisted, The University of Chicago, Department of Health Studies, MC2007, 5841 S Maryland Ave, Chicago, IL 60637, *thisted@health.bsd.uchicago.edu*

***Key Words:*** random effects, longitudinal analysis, panel data

In randomized clinical trials, the outcome variable is usually measured prior to randomization (baseline). Eligibility for study participation sometimes requires this baseline value to exceed a threshold. The outcome variable is measured at prespecified intervals over the course of the study, so that each subject has a panel of measurements over time. A common approach is to model within-subject correlation using random effects, and to assess treatment differences by focusing on a model for treatment by time interaction. When the degree of response depends upon the level of pathology at baseline, primary interest generally focuses on treatment differences net of baseline effects. We examine common approaches for baseline adjustment, and demonstrate how popular methods can sacrifice power, validity, and/or numerical stability. We also examine alternative approaches.

### Performance of Pseudo-Rsquare Statistics in the Linear Mixed Model

❖ Jean Orelien, SciMetrika LLC, 100 Capitola, Durham, NC 27713, *jorelien@scimetrika.com*; Lloyd Edwards, The University of North Carolina at Chapel Hill

***Key Words:*** linear mixed model, Rsquare, goodness-of-fit, simulation, longitudinal data, repeated measures

In the linear mixed model (LMM), several pseudo-Rsquare statistics have been proposed. Vonesh et al. (1996) suggested the concordance correlation coefficient be used to measure the percent agreement between the observed and predicted values. Vonesh and Chinchilli (1997) offered a formula for computing the proportion of reduction in residual variation. Zheng (2000) introduced a statistic to measure the proportional reduction in penalized quasi-likelihood. Xu (2003) proposed three statistics to measure the proportion of explained variation. However, the performance of these different Rsquare statistics has not been sufficiently demonstrated. We will present results that show that they do not perform adequately. Explanations for the shortcomings are given. We propose a new Rsquare for the LMM that performed well in simulations and has intuitive interpretation.

## 513 Health, Resources, Energy, and Ranking ● ✪

Business and Economics Statistics Section, Section on Health Policy Statistics

**Thursday, August 10, 8:30 am–10:20 am**

### Diseased-Based Price Index: a Cure Worse Than the Disease

❖ Ralph Bradley, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Room 3105, Washington, DC 20212, *bradley.ralph@bls.gov*

***Key Words:*** price indexes, hedonic adjustments

Many economists recommend the Bureau of Labor Statistics change its Medical CPI (MCPI) so it indexes the cost of treating disease, rather than the use of medical inputs. This study shows this proposed solution introduces more errors than it reduces. Many of the assumptions behind this recommendation are wrong, and the inability to correctly match all treatments with a diagnosis and the incidence of diagnostic error make it inferior.

### Switching from Retrospective to Current-Year Data Collection in the Medical Expenditure Panel Survey-Insurance Component (MEPS-IC)

❖ Anne T. Kearney, U.S. Census Bureau, 15710 Peach Walker Drive, Bowie, MD 20716-1650, *anne.theresa.kearney@census.gov*; John P. Sommers, Agency for Healthcare Research and Quality

***Key Words:*** retrospective design, nonresponse adjustment, poststratification

The MEPS-IC is a national survey of approximately 44,000 establishments and governments that produces national and state-level estimates of insurance availability and costs through employers. Under the current design, MEPS-IC collects data retrospectively, i.e., the respondents are asked to report data for the year prior to the current calendar year. The authors researched the ramifications of asking the respondents to report for the current calendar year making the resulting estimates more timely. In order to switch to current-year data collection, we have to address frame and reweighting issues since the data used in these operations will be one year older than under the retrospective design. In this paper we will present the methodology we used to address and overcome some of these obstacles and make recommendations on the feasibility of switching to a current-year data design.

### Nonparametric Inferences on Conditional Quantile Processes

❖ Chuan Goh, University of Toronto, Department of Economics, 150 St George St, Toronto, ON M5S 3G7 Canada, *goh@economics.utoronto.ca*

***Key Words:*** quantile regression, quantile treatment effect, local polynomials, nonparametric inference, minimax rate, generalized likelihood ratio

A variety of useful distributional hypotheses can be formulated in terms of restrictions on the sample paths of conditional quantile processes. This paper extends tests of the generalized likelihood ratio (GLR) type as introduced by Fan et al (2001) to nonparametric inferences on conditional quantile processes. Members of a class of GLR statistics based on locally polynomial quantile regressions are shown to have asymptotic null distributions that are free of relevant nuisance parameters. A resampling strategy is proposed for the consistent estimation of the asymptotic null distribution of the GLR statistics. These tests are also shown to be asymptotically rate-optimal for nonparametric hypothesis testing according to the formulations of Ingster (1993) and Spokoiny (1996). These tests present viable alternatives to existing methods based on the Khmaladze (1981) martingale transformation.

### Risk Sharing by Local Governments and Private Institutions in Environmental Transformation

❖ Neela Manage, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33434, *manage@fau.edu*

***Key Words:*** time-to-event data, environment, risk-sharing

The process of transforming environmentally distressed areas in urban communities into economically productive ones has received increasing attention in recent years. The successful, timely completion of this process is a function of several factors including local economic conditions, risk, and the role of local government agencies and private financial institutions. This paper obtains statistical estimates of the impacts of alternative economic, political, and financial variables that are relevant in models of environmental transformation.

### Nonparametric Transfer Function Model

❖ Jun Liu, Georgia Southern University, P.O. Box 8151, Dept of Finance and Quantitative Analysis, Statesboro, GA 30460, *jliu@georgiasouthern.edu*; Qiwei Yao, London School of Economics; Rong Chen, University of Illinois at Chicago

***Key Words:*** nonparametric regression, local polynomial regression, regression splines, nonlinear time series analysis, transfer function model, ARIMA

A new approach is proposed to model the relationship between an output and some input time series, perturbed by correlated noise. The functional form of this relationship (the transfer function) is unknown but assumed to be smooth. We propose to model the transfer function by nonparametric smoothing methods and model the noise as an ARIMA process. By using nonparametric smoothing, the model is very flexible and can be used to model highly nonlinear relationships of unknown form; by modeling the noise, the correlation in the data is removed so the transfer function can be estimated more efficiently; additionally, the estimated ARIMA structure can be used to improve the forecasting performance. The estimation procedures are introduced and the asymptotic properties of the estimators are studied. The finite-sample properties of the estimators are studied by simulation and real-life examples.

### Rank-Based Estimation for Autoregressive Moving Average Time Series Models

❖ Beth Andrews, Northwestern University, Department of Statistics, 2006 Sheridan Road, Evanston, IL 60208, *bandrews@northwestern.edu*

***Key Words:*** time series, autoregressive moving average models, rank estimation

A rank-based technique is used to estimate the parameters of autoregressive moving average (ARMA) time series models. The estimators minimize the sum of mean-corrected model residuals weighted by a function of residual rank. They are shown to be consistent and asymptotically normal under very mild conditions on the noise distribution, and so the estimation technique is robust. Because the weight function can be chosen so that rank estimation has the same asymptotic efficiency as maximum likelihood estimation, the estimators are also relatively efficient. The relative efficiency of the estimators extends to the unknown noise distribution case since rank estimation with the Wilcoxon weight function (a linear weight function) is nearly as efficient as maximum likelihood for a large class of noise distributions.

### Bayesian Seemingly Unrelated Regression in Spatial Regional Model: Economics of Agglomeration in Japan from 1991--2000

❖ Kazuhiko Kakamu, Institute of Advanced Studies, Stumpergasse 56, Vienna, 1060 Austria, *kakamu@ihs.ac.at*; Wolfgang Polasek, Institute for Advanced Studies; Hajime Wago, Nagoya University

***Key Words:*** economics of agglomeration, Markov chain Monte Carlo (MCMC), panel data, seemingly unrelated regression (SUR), spatial autoregressive model (SAR), spatial error model (SEM)

This paper considers the seemingly unrelated regression (SUR) model with spatial dependency from a Bayesian point of view. We consider Markov chain Monte Carlo methods to estimate the parameters of the model. Our approach is illustrated with simulated and real data sets. In the real data example, we analyze the economics of agglomeration in Japan during the period 1991 to 2000.

## 514 Regression, Time Series, and Adjustments ●

**Business and Economics Statistics Section**
**Thursday, August 10, 8:30 am–10:20 am**

### Using Firm Optimization To Evaluate and Estimate Returns to Scale

❖ Yuriy Gorodnichenko, University of Michigan, 125 W. Hoover, Apt. 11B, Ann Arbor, MI 48103, *ygorodni@umich.edu*

***Key Words:*** production function, identification, returns to scale, covariance structures

Because of data limitations, available estimates of returns to scale at the firm level are for the revenue function, not production function. Given this observation, the paper argues that micro-level estimates of returns to scale are often inconsistent with profit maximization or imply implausibly large profits. The puzzle arises because popular estimators ignore heterogeneity/endogeneity in factor/product prices, assume perfectly elastic factor supplies or neglect the restrictions imposed by profit maximization so that estimators are inconsistent/poorly identified. To address these problems, the paper proposes a simple structural estimator that models the cost and the revenue functions simultaneously and accounts for unobserved heterogeneity in productivity and factor prices symmetrically. The strength of the proposed estimator is illustrated by simulations and an empirical application.

### How Far to Asymptopia? Errors and Regressors with Realistically Asymmetric Distributions Generate Unreliable t-Statistics in Large Finite Samples

❖ Robert McClelland, Bureau of Labor Statistics, Suite 3105, 2 Massachusetts Ave NE, Washington, DC 20212, *mcclelland.robert@bls.gov*; Elliot Williams, Bureau of Labor Statistics

***Key Words:*** central limit theorem, Berry Esseen

Various central limit theorems guarantee that the distribution of t-statistics for significance of regression coefficients is asymptotically Normal under regularity conditions. Without violating the assumptions of the CLTs, however, characteristics of the data can dramatically slow the convergence of the finite-sample distributions to their Normal limiting distribution. Here we show that asymmetrically distributed regressors and errors can significantly slow convergence of the tail quantiles of the finite sample distribution of the t-statistic to their Normal limiting values. Particularly, we find that the asymptotic Normality assumption is misleading for inference at the 5% level with sample sizes from 10,000 to 100,000 when applied to data with skewnesses typical of economic wage and income data.

## Effectiveness of Two-Stage Least Squares in Correcting Endogeneity Bias: a Monte Carlo Study

❖ V. A. R. Samaranayake, University of Missouri-Rolla, 202 Rolla Building, Rolla, MO 65401, *rrbryant@umr.edu*; Xujun Wang, University of Missouri-Rolla

***Key Words:*** endogeneity, simultaneity bias, instrumental variables, ordinary least squares, structural equations

A common solution to the problem of endogeneity bias in regression is the use of instrumental variables to predict the endogenous variable and the using the predicted values in place of the original variable values in the regression model. It has been pointed out that this two-stage least squares (2LS) method may not work satisfactorily when the correlation between the endogenous variable and its predicted value is weak and one or more of the instrumental variables employed is correlated, even weakly, with the error term associated with the dependent variable. We conduct a Monte Carlo study to investigate the bias of the 2SLS estimators under various conditions. The results show that the 2SLS approach does not eliminate or reduce the bias substantially in many situations.

## Tuning Parameter Selectors for SCAD

❖ Hansheng Wang, Peking University, Guanghua School of Management, Beijing, 100871 China, *hansheng@gsm.pku.edu.cn*; Runze Li, The Pennsylvania State University; Chih-Ling Tsai, University of California, Davis

***Key Words:*** AIC, BIC, GCV, LASSO, SCAD

The penalized least squares approach with smoothly clipped absolute deviation (SCAD) penalty has been consistently demonstrated to be an attractive regression shrinkage and selection method. It not only automatically and consistently selects the important variables, but also produces estimators which are as efficient as the oracle estimator. However, these attractive features depend on appropriately choosing the tuning parameter. In this article, we show that the most commonly used GCV cannot select a tuning parameter satisfactorily with a non-ignorable overfitting effect in the resulting model. In addition, we propose a BIC tuning parameter selector, which is shown to be able to identify the true model consistently. Simulation studies are presented to support theoretical findings, and an empirical example is given to illustrate its use in a Female Labor Supply data set.

## A Stepwise SPA Test for Data Snooping and Its Application on Fund Performance Evaluation

❖ Yu-Chin Hsu, The University of Texas at Austin, 3600 N. Hills Drive, Apartment 135, Austin, TX 78731, *yh428@eco.utexas.edu*; Po-Hsuan Hsu, Columbia University

***Key Words:*** data snooping, reality check, stepwise test, SPA test

In this study, we propose a Stepwise SPA Test which is powerful in searching for predictive models or profitable investment targets with appropriate family-wise error control. Our testing method, built on White's Reality Check (2000), Hansen's SPA test (2005), and Romano and Wolf's stepwise procedure (2005), aims to conduct large-scale joint hypothesis testing in a given data set. Based on the Monte Carlo simulations, we show that the proposed test is more powerful than the other three tests in limited samples. We then apply the test to examine the performance of mutual funds and hedge funds, and obtain some interesting results. We find that only eight mutual funds are found to beat the S$\&$P 500 index, and only few hedge funds outperform the risk-free rate. With these two empirical cases, we substantiate the empirical value of our test in fund performance evaluation.

## An Alternative Framework for Univariate and Multivariate Seasonal Adjustment

❖ Stéphane Gregoir, CREST/INSEE, 15 Blvd. Gabriel Peri, Malakoff, 92245 France, *gregoir@ensae.fr*

***Key Words:*** seasonal adjustment, seasonal unit root, seasonal cointegration, persistence

To describe the seasonal patterns in economic data, we use a weak linear representation of integrated processes. We assume that the polynomial associated to the Wold representation of the process appropriately differenced satisfies a not too restrictive summability condition. Relying on a simple additive decomposition of these integrated processes, the statistical treatment consists of removing from raw data the pure seasonal random walk components present at each seasonal frequencies. These random walk components are separately derived from a Beveridge-Nelson type decomposition of the demodulated process at each seasonal frequency. This framework allows for a simultaneous seasonal adjustment of a set of variables, the use of robust estimation procedures to limit the influence of outlier definition, and gives a rationale for the computation of asymmetric filters at the ends of the sample

## Nonlinear Properties of Conditional Returns under Scale Mixtures

❖ Venkata Jandhyala, Washington State University, Department of Mathematics, Pullman, WA 99164, *jandhyala@wsu.edu*; Stergios B. Fotopoulos, Washington State University

***Key Words:*** financial log returns, APT model, scale mixtures, regression equation, conditional variance, GIG family of distributions

Analytical expressions are derived for the non-linear regression and its prediction error by modeling the log returns of financial assets as scale mixtures of the multivariate normal distribution. The expressions involve conditional moments of the mixing variable. These conditional moments are explicitly derived when the mixing variable belongs to the generalized inverse Gaussian family, of which gamma, inverse gamma and the inverse Gaussian distributions are members. The effectiveness of the nonlinear model over the usual linear model is captured through simulations for the above three families of distributions. The proposed scale mixture models extend the well-known arbitrage pricing theory (APT) in financial modeling to non-Gaussian cases. The methodology is applied to analyze the log returns intra-day data for DELL, COKE and S&P500 for the years 1998-2000.

# 515 Psychiatric, Cardiovascular, and Infectious Disease Trials ●

Biopharmaceutical Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

## Comparison of Methods for Handling Missing Data in NonInferiority Psychiatric Trials

❖ Isaac Nuamah, Johnson & Johnson Pharmaceutical R&D, 5 Orly Way, Burlington, NJ 08016, *inuamah@prdus.jnj.com*; Guohua Pan, Johnson & Johnson Pharmaceutical R&D; Kwang-Shi Shu, Johnson & Johnson Pharmaceutical R&D; Pilar Lim, Johnson & Johnson Pharmaceutical R&D

***Key Words:*** non-inferiority, per-protocol population, psychiatric trials, intent-to-treat, missing data

Noninferiority (NI) analysis is used to demonstrate a new treatment is no worse than a known effective treatment by some predefined margin. While the intent-to-treat (ITT) principle is accepted as an approach to demonstrate superiority, an NI comparison requires consideration of the analysis population (AP). The inclusion of subjects with protocol deviations usually leads to reduced treatment differences, therefore resulting in an anti-conservative bias that would favor the NI conclusion. Analyses based on both the per-protocol (PP) population and the ITT population may need to be consistent to support a claim of NI. Regardless of the AP, missing data and methods to address them would have to be addressed during analysis. We compare the type I error rate and power of several methods for handling missing data in NI analysis based on ITT or PP via simulation.

## Differentiating Major Depressive Disorder in Youths with Attention Deficit Hyperactivity Disorder

❖ Adriana Lopez, University of Pittsburgh, 2717 Cathedral of Learning, Pittsburgh, PA 15260, *adl5@pitt.edu*; Satish Iyengar, University of Pittsburgh; Rasim Diler, University of Pittsburgh; William Daviss, University of Pittsburgh

***Key Words:*** attention-deficit/hyperactivity disorder, discriminant analysis, model selection, comorbidity, depression

Accurate diagnosis of depression in children and adolescents with Attention Deficit Hyperactivity Disorder (ADHD) is very important for early intervention. In this paper, we examine which depressive symptoms best discriminate Major Depressive Disorder (MDD) based on a large clinical sample of youth with ADHD who completed the Mood and Feelings Questionnaire. Best predictors of MDD in ADHD youth were found by means of stepwise linear discriminant analysis, stepwise logistic regression and classification trees. Based on crossvalidated misclassification rates and expert opinion the stepwise logistic regression gave the best predictors. Our findings suggest the particular importance of anhedonia, social withdrawal, psychomotor retardation, negative views of self and future, and suicidal thoughts as symptoms that distinguish MDD in youths who have ADHD.

## Impact of Reference Range Selection on Long Term Growth Evaluations

❖ Dustin Ruff, Eli Lilly and Company, 1238 Burr Oak Court, Greenwood, IN 46143, *ruffestimates@sbcglobal.net*

***Key Words:*** clinical trials, growth curve

Because of ethical concerns in conducting long-term placebo controlled trials, the effects of any drug on growth frequently have to be studied within open-label clinical trials using reference growth standards. In studies of this type, it is crucial to understand just how well the reference standards being used truly model the population of interest. Additionally, when the choice of an ill-fitting reference range is mandated by regulatory precedent, it is important to understand how its use can affect the results seen. These questions were examined using a large database of growth data from ADHD patients. Various tests were used to evaluate reference range fit. The effects of noticed departures from the standard were further explored by comparing it with new references fit for this population. Simulations were conducted examining the expected effect on eventual clinical trial results.

## Is There an Optimal Sample in Depression Studies?

❖ Kenneth Liu, Merck & Co., Inc., BL 3 2, PO Box 4, West Point, PA 19486, *Kenneth_Liu@merck.com*; Duane Snavely, Merck & Co., Inc.

***Key Words:*** sample size, patient drift, clinical trials

Four parallel-group, placebo-controlled phase III studies were conducted to evaluate the antidepressant effect of a test compound compared to placebo; an approved antidepressant, paroxetine, was included as an active control. Paroxetine consistently outperformed placebo in reducing the 17-item total of the Hamilton Depression Rating Scale after eight weeks of treatment. During the design of these studies, questions existed as to whether there may be a “negative” return on investment with respect to the sample sizes; that is, larger sample sizes may not necessarily increase the chance for a positive study due to patient drift. The results of a post-hoc analysis will be discussed.

## Placebo Effect Adjusted Assessment of Quality of Life in Placebo-Controlled Clinical Trials

❖ Jens Eickhoff, University of Wisconsin-Madison, 250 WARF Building, Madison, WI 53726-2397, *eickhoff@biostat.wisc.edu*

***Key Words:*** latent variable modeling, Monte Carlo EM, misclassification, placebo effect

Quality of life (QoL) has become an accepted and widely used endpoint in clinical trials. The analytical tools used for QoL evaluations in clinical trials differ from those used for the more traditional endpoints, such as response to disease, overall survival or progression-free survival. Since QoL assessments are generally performed on self-administered questionnaires, QoL endpoints are more prone to a placebo effect than traditional clinical endpoints. The placebo effect is a well-documented phenomenon in clinical trials which has led to dramatic consequences on the clinical development of new therapeutic agents. In order to account for the placebo effect, a multivariate latent variable model is proposed which allows for misclassification in the QoL responses. The approach is illustrated with analysis of data from a cardiovascular phase III clinical trial.

## Clinical Trial Designs To Study Rare Infectious Diseases

❖ Yufeng Li, The University of Alabama at Birmingham, 1824 6th Ave., S., WTI 153, Birmingham, AL 35294, *yufengli@uab.edu*; Seng-jaw Soong, The University of Alabama at Birmingham

***Key Words:*** clinical trial design, rare infectious disease, Bayesian method, frequentist method

Due to the small number of patients affected by some rare infectious diseases, clinical trials studying these diseases are either underpowered because of the difficulty of recruiting adequate numbers of subjects, or the study recruiting period is prolonged, resulting in great expense. An appropriately designed trial is imperative. Currently, however, the methodologies for studying rare infectious diseases are not significantly different from the methodologies for studying other diseases. In this research, we provide a review of the methodologies for studying rare infectious diseases with comparisons of the statistical properties. We provide an example of designing a rare infection trial with both frequentist and Bayesian approaches. The sample size and power under each approach are presented.

## 516 Environmental Models and Assessing Gene-Environment Interactions

Biometrics Section, WNAR, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

### Statistical Issues Related to Development of a New Model of the Life Cycle of Salmonid Fishes in the Pacific Northwest

❖ Steven G. Smith, U.S. Department of Commerce, NOAA Fisheries, 2725 Montlake Blvd E, Seattle, WA 98112, *steven.g.smith@noaa.gov*; James R. Faulkner, U.S. Department of Commerce; Richard W. Zabel, U.S. Department of Commerce/NOAA Fisheries

***Key Words:*** salmonids, tag-recapture, survival, endangered species, stochastic model

Scientists representing federal, state, and tribal agencies are collaborating to develop a new stochastic model of the life cycle of endangered anadromous salmonid fishes of the Pacific Northwest. Components of the model include tributary spawning and rearing, downstream migration of juveniles through a complex river system with many hydroelectric dams, estuary and ocean residence, upstream migration of adults, and return to spawning streams. I will present analyses supporting the development of this complex model. The principal data are from millions of juvenile salmon marked with passive electronic tags over the past decade. I will emphasize analyses of environmental influences on survival of juvenile migrants; methods for implementing stochasticity in all components; and of potential latent mortality caused by the hydroelectric system but expressed outside the migration corridor.

### Breeding Value Estimation in Forest Genetics

❖ Anne M. Millar, Dahousie University, Math and Statistics Department, Chase Building, Halifax, NS B3H 3J5 Canada, *millar@eastlink.ca*

***Key Words:*** linear mixed effects model, bootstrap, pseudo-likelihood, breeding values, indirect estimation, forest genetics

In forest tree improvement programs, the goal is to identify and select parents who will produce offspring with desired characteristics to improve the next generation. Our new model for offspring height accounts for spatial correlation due to micro-environmental differences, and for differing within family genetic variances. Since the likelihood for such a model is intractable we approximate the maximum likelihood estimates, using a weighted linear mixed effects model for estimating the within family variances along with pseudo-likelihood and indirect estimation to account for the correlation. Bootstrap methods provide inference on the potential gain from selection and for the selection process itself.

### A Model for Ordinal Data with Spatial and Temporal Dependency

❖ Graciela Gonzalez-Farias, CIMAT, Jalisco, Guanajuato, 36240 Mexico, *farias@cimat.mx*; Rogelio Ramos-Quiroga, CIMAT; Felipe Peraza-Garay, Universidad Autonoma de Sinaloa

***Key Words:*** spatio-temporal, likelihood, ordinal response

We propose a model to handle both spatial an temporal dependencies to study ordinal data using a Markovian type time structure through transition probabilities, which contain both structures of dependencies. These probabilities are based on a logistic distribution. In terms of inference we use the maximum likelihood approach and discuss some of the asymptotic issues. An example on agave plants is presented and prediction maps are constructed in order to understand the spread of the illness in time and space.

### New Method for Assessing Gene-Environment Interaction in Case-Only Studies

❖ Shizue Izumi, Oita University, Department of CIS, Faculty of Engineering, 700 Dannoharu, Oita, 870-1192 Japan, *shizue@csis.oita-u.ac.jp*; John Cologne, Radiation Effects Research Foundation

***Key Words:*** bias, cancer, causation, environmental exposure, molecular epidemiology

We introduce a new method for assessing gene-environment interaction in case-only studies, adjusting the case-only odds ratio by the population attributable fraction for environmental exposure. Binomial regression models are used to estimate the relative marker frequency (a measure of gene-environment interaction) in the causally related cases compared to that in the non-causally related cases. The authors illustrate the proposed approach for multiple exposure levels and compare it with the traditional approach, using example data on a CYP1A1 gene polymorphism and years smoked among lung cancer patients. We conclude that the proposed approach can be used straightforwardly in epidemiologic research to examine whether non-causally related exposed cases might reduce the extent of gene-environment interaction estimated by the traditional approach.

### An Evaluation of the Benefit of Genetic Information in Discovering the Effect of an Environmental Factor on Disease

❖ Abhijit Dasgupta, Thomas Jefferson University, 211 S. 9th Street, Suite 602, Philadelphia, PA 19107, *aikidasgupta@gmail.com*; Nilanjan Chatterjee, National Cancer Institute; Sholom Wacholder, National Cancer Institute

***Key Words:*** gene-environment, case-control studies, design of epidemiologic studies, omnibus test, statistical efficiency

The study of gene-environment interaction is often touted as a way to use genes to discover the environmental agents that are the preventable causes of disease. However, there are no strategies for using genetics to identify which environmental factors are causal, except for extreme situations where an environmental factor X causes disease D. In the omnibus strategy, the effect of X is evaluated based on the sum of K stratum-specific test statistics, where we have K alleles at a locus G. We show for dichotomous G and X, the omnibus strategy performs well regardless of the joint effects of X and G, running a close second to efficient tests in extreme situations where either there is no G-effect or the effect is restricted to only one G-stratum. With no prior knowledge about the risk model, collecting information on G and using the omnibus test likely is the best strategy for detecting X.

### Analysis of an Interaction Threshold in a Mixture of Drugs and/or Chemicals

❖ Adam Hamm, Rho, Inc., 6330 Quadrangle Drive, Suite 500, Chapel Hill, NC 27517, *adam_hamm@rhoworld.com*

*Key Words:* drug combinations, chemical mixtures, segmented regions, synergism, dose-dependent interactions, threshold model

Increasingly, humans are exposed to drug/chemical mixtures. These exposures can result from therapeutic interventions or environmental sources. Of interest is the interaction that may occur among the components of these mixtures. Since interaction can be dose-dependent, it is important to determine exposure levels to either exploit the benefits of the interaction in a therapeutic application or to avoid the effect of the interaction in the case of an environmental risk assessment. We propose generalized linear models that permit the estimation of interaction threshold boundaries. The methods developed are applied to the combination of ethanol and chloral hydrate.

## 517 Statistical Methods in HIV Research ●

Biometrics Section, ENAR

**Thursday, August 10, 8:30 am–10:20 am**

### Combining Retrospective and Prospective Data To Improve Markov Transition Parameter Estimation for Characterizing the Accumulation of HIV-1 Drug Resistance Mutations

❖ Brian Healy, Harvard School of Public Health, 655 Huntington Ave., SPH 1 Room 412B, Boston, MA 02115, *bhealy@hsph.harvard.edu*; Victor DeGruttola, Harvard School of Public Health; Marcello Pagano, Harvard School of Public Health

*Key Words:* hidden Markov models, branching trees, HIV resistance mutations, constrained optimization, genetic pathways

Prospective studies of HIV-infected patients permit investigation of acquisition of new HIV resistance mutations, but patients often have many such mutations at study entry. We propose methods for modeling genetic pathways that combine Markov models for prospective data and branching trees for cross-sectional baseline data. Most links between the two sets of model parameters are functions of the time on treatment before study entry---information that may not be available. Nonetheless, some link functions eliminate the dependence on this time under certain assumptions. Approaches to fitting models to the combined information that use these functions include constrained maximization and weighted least squares. The latter easily accommodates error in the link function due to uncertainty in the necessary assumptions. Using both sources of information improves precision of the estimation.

### Inference for Multiple Kappas with Nested and Clustered Study Designs: application to HIV Prevention and Sexual Abuse Research

❖ Yan Ma, University of Rochester, Department of Biostatistics, 1310 Valley Lake Drive, Apt# 523, Schaumburg, IL 60195, *yan_ma@urmc.rochester.edu*; Xin Tu, University of Rochester

*Key Words:* asymptotic distribution, longitudinal study, missing data, u-statistics

Measurement in mental health and psychosocial research hinges critically on the reliability of instruments (questionnaires) designed to capture the often latent constructs that define a disorder or a condition. As such instruments are rated by independent evaluators or by subjects themselves, instrument/rater reliability has been a major concern for deriving scientific evidence from such a "soft" outcome based research discipline, affecting scientific and research integrity in many psychosocial interventions. One of the key indices for assessing outcome reliability is the kappa class of statistics. However, existing methods on such a class of statistics do not sufficiently address data clustering and missing data, two most common issues in longitudinal data analysis. In this talk, I introduce a new class of U-statistics based estimates to facilitate inference under clustered study designs.

### On Sensitivity Analysis of HIV-Free Survival

❖ Peng Zhang, Harvard University, 655 Huntington Ave., Sph1 412H, Boston, MA 02115, *pzhang@hsph.harvard.edu*; Stephen W. Lagakos, Harvard University

*Key Words:* composite ending points, HIV vertical transmission, Monte Carlo, panel data, sensitivity analysis

We consider estimation of the distribution of HIV-free survival in infants of HIV-infected mothers. Because HIV infection is a silent event, HIV infection status is assessed periodically using a diagnostic test, which is imperfect. Thus neither the occurrence of infection nor the time of occurrence can be known. The competing cause of failure--death--is observed continuously. The problem is further complicated by missing data, which includes non-informative missingness due to skipping pre-scheduled visits and possibly informative missingness due to death or lost follow-up. Because of the data structure, aspects of the underlying process are non-identifiable. We apply nonparametric methods to estimate the estimable aspects of the process and supplement these with non-parametric sensitivity analyses to make the inference about the distributions of HIV survival and overall survival time.

### Hierarchical Poisson Regression Models for HIV Vaccine Studies

❖ Xin Huang, University of California, Los Angeles, 3241 S. Sepulveda Blvd., Apt. 302, Los Angeles, CA 90034, *xinhuang@ucla.edu*; W. John Boscardin, University of California, Los Angeles; Elissa Schwartz, Harvey Mudd College

*Key Words:* nonlinear regression, Markov chain Monte Carlo

The safety and immunogenicity of two potential HIV-1 vaccines were evaluated by NIH/NIAID randomized blinded studies (AI 50467) comparing mucosal and blood responses of deltoid versus inguinal vaccination. CD8+ cellular immune (CTL) responses were measured longitudinally in the blood versus mucosal lymphocytes, and assayed for frequencies of cells reacting to 53 different peptide pools of the HIV-1 genome (by measuring the frequency of spot-forming cells in the ELISpot assay). The rate of any potential vaccine activity was low compared to the background noise, and it was thus essential to make efficient use of the data. We present a hierarchical Poisson regression model for assessing the de-noised rate of HIV-1-specific CTL responses by immunization site, peptide pool, and time point. Log linear mean structures are used, but the background is assumed additive on the original scale.

### Hypothesis Testing of Treatment Policies in Two-Stage Randomization Designs in Clinical Trials

❖ Xiang Guo, sanofi-aventis, 35 River Drive, S., APT 1904, Jersey City, NJ 07310, *xiang.guo@sanofi-aventis.com*; Anastasios A. Tsiatis, North Carolina State University

*Key Words:* counting process, induction therapy, inverse weighting, maintenance therapy, potential outcomes, outcomes

In many clinical trials related to diseases such as cancers and HIV, patients are treated by different combinations of therapies. This leads to two-stage designs, where patients are randomized initially to a primary therapy and then, depending on the disease remission and patients consent, a maintenance therapy is assigned randomly. In this paper, we propose a method for hypothesis testing of treatment policies in such two-stage studies with right censoring by using a weighted version of log-rank test. The method is demonstrated through simulations.

### Design and Analysis of Neutralizing Antibody Assays in HIV-1 Vaccine Trials

❖ Yunda Huang, Fred Hutchinson Cancer Research Center, Statistical Center for HIV/AIDS Research and Prevention, 16605 SE 31st Street, Bellevue, 98008, *yunda@scharp.org*; Peter Gilbert, Fred Hutchinson Cancer Research Center/University of Washington; David Montefiori, Duke University Medical Center; Steve Self, University of Washington

***Key Words:*** AIDS, cross-reactive, neutralizing antibody, potency, M-B curves

An effective HIV vaccine must induce potent neutralizing antibody (Nab) responses against a wide range of circulating viruses. Magnitude and breadth (M&B) of neutralization is the main endpoint for comparing antibody-based HIV-1 vaccine candidates in Phase I and II trials and for advancing candidates to Phase III trials. In Phase III trials, M&B of neutralization is also a key marker to evaluate as a putative surrogate endpoint for protection against HIV infection. In this paper, we discuss different approaches to integrating M&B in the design of HIV vaccine trials and analysis of Nab data. We provide suggestions on the summarization and group comparison of multivariate Nab data when left censoring often occurs. Applications to Nab response data from the recent first HIV vaccine efficacy trial will be discussed. Sample size calculations based on simulations will also be presented.

### Antigen Scanning Methods for Identifying Peptide Signatures

❖ Allan deCamp, Fred Hutchinson Cancer Research Center, Seattle, WA , *adecamp@u.washington.edu*; Peter Gilbert, Fred Hutchinson Cancer Research Center/University of Washington

***Key Words:*** false discovery rate, genetic data, high-dimensional data, human immunodeficiency virus, Kullback-Leibler

We develop an antigen scanning method in order to evaluate the results from HIV vaccine efficacy trials. An HIV amino acid sequence is measured from each volunteer who acquires HIV during the trial. These sequences are multiply aligned, along with the prototype HIV sequence represented in the vaccine, to form a high-dimensional genomic dataset. Based on these data, we identify short, 8-12 amino acid long, peptide regions at which the peptide sequences from infected vaccinees tends to be more divergent from the prototype vaccine sequence than the sequences from infected placebo recipients. We approach this problem by extending earlier methods for comparing positions, using Euclidean and Kullback-Leibler-type metrics for measuring peptide distances. The methods are illustrated by a simulation study and with data from an HIV vaccine efficacy trial.

## 518 Releasing Tabular Data Collected under a Confidentiality Pledge: Going beyond Cell Suppression

Business and Economics Statistics Section

**Thursday, August 10, 10:30 am–12:20 pm**

### A Comparison on Data Utility between Publishing Fixed Intervals versus Traditional Cell Suppression on Tabular Employment Data

❖ Steve H. Cohen, Bureau of Labor Statistics, 2 Massachusettes Ave., NE, Washington, DC 20212-0001, *cohen.steve@bls.gov*; Bogong Li, Bureau of Labor Statistics

***Key Words:*** disclosure, fixed-intervals, data utility

The Bureau of Labor Statistics' Quarterly Census of Employment and Wages series publishes employment and wage estimates classified by county and six-digit North American Industry Classification System level. Unfortunately, due to data disclosure issues, nationally sixty percent of the county-level detail is suppressed using traditional cell suppression techniques. Alternative disclosure-protecting techniques have been proposed, specifically replacing these tabulations with fixed-interval for the suppressed cells. This paper will focus on comparing the data utility of these disclosure avoidance techniques. The positives and negatives of each will be discussed in addition to utility tradeoffs between the two methodologies. Data simulations will be presented.

### Quality-Preserving Controlled Tabular Adjustment: an Alternative to Cell Suppression for Disclosure Limitation of Tabular Magnitude Data

❖ Lawrence H. Cox, National Center for Health Statistics, Room 3211, 3311 Toledo Road, Hyattsville, MD 20782-2003, *lcox@cdc.gov*

***Key Words:*** linear programming, distribution, confidentiality

Cell suppression thwarts standard methods of data analysis and is not describable probabilistically nor amenable to statistical methods for missing data. An alternative is controlled tabular adjustment (CTA). CTA replaces disclosure cells by safe values, adjusts remaining values to rebalance tabular equations, and performs optimally over a range of measures of 'closeness.' Recent research extends CTA to quality-preserving CTA by augmenting the linear constraint system with linear constraints that mimic important distributional parameters and statistics. In most cases, this can mitigate significant changes to the underlying distribution, in both univariate and multivariate settings. We control changes to means, variances, correlations, covariances, and regression coefficients. Based on linear programming, these formulations are easy to implement or enhance.

### Combining Synthetic Data and Noise Infusion for Confidentiality Protection of the Quarterly Workforce Indicators

❖ John Abowd, Cornell University, CISER, 391Pine Tree Road, Ithaca, NY 14850, *john.abowd@cornell.edu*; Lars Vilhuber, Cornell University

***Key Words:*** confidentiality, noise infusion, synthetic data, quarterly workforce indicators, analytic validity

The Quarterly Workforce Indicators (QWI) statistics currently are protected by a combination of noise infusion and suppression. Dynamically consistent noise infusion is used for all measures. Suppression is used for counts from sparsely populated cells. The Longitudinal Household Employer Dynamics (LHED) Program, which produces the QWIs, has experimented with replacing the suppressed QWIs with synthetic values. The synthetic values are generated by sampling from the posterior predictive distribution of the indicator, given its history and the rules that cause the suppression. The synthesis is done from the underlying confidential QWI data, not from the released data. The use of synthetic data in this application improves the analytic validity of the QWIs without compromising the protection in the noise infusion system.

### Protecting the Confidentiality of Commodity Flow Survey Tables by Adding Noise to the Underlying Microdata

❖ Paul B. Massell, U.S. Census Bureau, Room 3209, Building 4, Stop 9100, 4700 Silver Hill Road, Washington, DC 20233-9100, *paul.b.massell@census.gov*; J. Neil Russell, National Center for Education Statistics

***Key Words:*** cell suppression, adding noise to microdata, commodity flow survey

The Commodity Flow Survey (CFS) produces data on the movement of goods in the United States. Data from the CFS are used by analysts for transportation planning, decisionmaking, and modeling transportation facilities and services demand. Complementary cell suppression has been used throughout the years to protect establishment identity in CFS data. Data users, especially transportation modelers, have indicated their need for access to data tables that do not have missing data due to suppression. Agency research now has focused on adding noise to CFS microdata to protect public-use CFS tabular data. Initial research findings of these efforts have been quite positive. This paper will present initial findings of this research and future research steps for a more detailed assessment of CFS data quality after application of this noise methodology.

## 519 Monte Carlo Methods for Computationally Intensive Problems ●

Section on Statistical Computing
**Thursday, August 10, 10:30 am–12:20 pm**

### Espousing Modern Computation with Classical Statistics: Sufficiency, Ancillarity, and a New Generation of MCMC

❖ Xiao-Li Meng, Harvard University, 1 Oxford Street, 7th Floor, Dept. of Statistics, Harvard University, Cambridge, MA 02138, *meng@stat.harvard.edu*

***Key Words:*** data augmentation, working parameter

Reparametrizations, or variable transformations, are known to be key to efficient implementation of MCMC algorithms. However, the majority of proposals to date focus on either a single transformation or straightforward combinations of several transformations. Here, we demonstrate that by interweaving two specific kinds of (one-to-one) transformations, we can gain considerable speed in convergence and simplicity in construction. Simplicity is due to choosing transformations via sufficiency and ancillarity, two familiar classical concepts. In addition, by using conditional sufficiency and ancillarity, we can interweave different transformations for different steps (e.g., Gibbs steps) within each iteration, and thereby the proposed strategy provides a fairly general recipe for constructing a new generation of efficient algorithms for complicated applications.

### Conditional Inference in Log-Linear Models: Exact Calculation versus Monte Carlo Approximation

❖ James Booth, Cornell University, Department of Statistical Sciences, Ithaca, NY 14853, *jb383@cornell.edu*

***Key Words:*** sufficient statistics, reference set, lack-of-fit

It is well-known that standard chi-squared asymptotic approximations for assessing the fit of a log-linear model to a sparse contingency table can be very misleading. An alternative is to assess the fit based on the conditional distribution of the table, given the sufficient statistics. Because this distribution does not involve any unknowns, it is possible---in principle---to construct an exact test for lack-of-fit. In practice, the support of the conditional distribution can be so large or complicated as to make exact computation infeasible. In many of these cases, Monte Carlo approximation algorithms have been developed that render 'essentially' exact inferences. In this paper, we show exact calculation is possible for some models that were previously thought to be intractable. Moreover, the exact calculations are sometimes actually faster than competing Monte Carlo approximations.

### Constrained Sequential Monte Carlo (CSMC)

❖ Rong Chen, National Science Foundation, Division of Mathematical Sciences, 4201 Wilson Blvd, Arlington, VA 22230, *rchen@nsf.gov*

The sequential Monte Carlo (SMC) methodologies are powerful tools in solving very high-dimensional and complex problems often encountered in applications. The key to a successful SMC implementation is efficiency, which is related directly to the design of the key components of SMC. Many problems in applications share a common feature: The target distribution is constrained highly. That is, the target distribution is a truncated distribution on an ill-shaped subspace of a high-dimensional space. The constraints, without careful treatments, are a main source of obstacles in successful implementations of SMC. In this talk, we develop a set of algorithms categorized as Constrained Sequential Monte Carlo (CSMC) for solving such problems, including strategies in designing the intermediate distributions, the trial distributions, the resampling steps, and Markov moves with CSMC.

### Sequential Monte Carlo for Estimating Ratio of Normalizing Constants

❖ Arnaud Doucet, The University of British Columbia, 333-6356 Agricultural Road, Department of Statistics, Vancouver, BC V6T 1Z2 Canada, *arnaud@stat.ubc.ca*

***Key Words:*** sequential Monte Carlo, normalizing constants, Bayes factors, importance sampling, Markov chain Monte Carlo

We propose an original sequential Monte Carlo method for estimating the ratio of normalizing constants. This method requires sampling from a sequence of simple auxiliary distributions, determined automatically by the algorithm. We illustrate the performance of this method on various Bayesian inference problems.

# 520 Classification of Data with a Large Number of Polychotomous Variables ● ✪

Classification Society of North America, Section on Physical and Engineering Sciences
**Thursday, August 10, 10:30 am–12:20 pm**

## Similarity Index for Polychotomous Variables

❖ Ranjan K. Paul, Boeing Math Group, P.O. Box 3707, MC 7L-21, Seattle, WA 98124-2207, *ranjan.k.paul@boeing.com*; I-Li Lu, The Boeing Company

***Key Words:*** ordinal, categorical, similarity, dissimilarity, clustering, classification

In this paper, we extend the concept of similarity measure and apply it in order to develop a classification index for data with strings of polychotomous responses. This index generalizes the development of a modified version Jaccard-Tanimoto coefficient from binary to multicategory settings. Asymptotic theory from probability is employed with the stability principle to derive the index weights. Maximum likelihood estimators of the probabilities of occurrence for categories are derived to estimate the index weights. When structures of the probabilities of occurrence are specified by the Dirichlet priors, estimates based on the admissible minimax principle are computed and compared with those estimated by the maximum likelihood procedures. Asymptotic distributions of these indexes are derived.

## Learning Accurate Probability Estimates: Why and How?

❖ Dragos D. Margineantu, The Boeing Company, M&CT, Adaptive Systems, P.O. Box 3707, M/S 7L-66, Seattle, WA 98124-2207, *dragos.d.margineantu@boeing.com*; Roman D. Fresnedo, The Boeing Company

***Key Words:*** class probability estimation, cost-sensitive learning, ranking

All classification algorithms can be modified to compute a class membership score that, in turn, can be transformed into class probability estimates P(y|x) by normalization. Practice has shown, however, that learning classification models that compute accurate probability estimates is, in general, a difficult task. Hence, several research studies have addressed improving the accuracy of the estimates computed by different algorithms. We show that for some types of problems (different types of ranking, classification decisions) accurate probability estimates are needed only in some critical regions of the input space and we propose solutions for dealing with poor estimates. We also discuss different measures of assessing probability estimates and the decisions derived from these estimates. The experimental results give new insights into cost-sensitive decision making and ranking.

## Assessing the Risk of Classification Decisions

❖ Roman D. Fresnedo, The Boeing Company, M&CT, Applied Statistics, P.O. Box 3707, M/S 7L-22, Seattle, WA 98124-2207, *roman.d.fresnedo@boeing.com*; Dragos D. Margineantu, The Boeing Company

***Key Words:*** classification, risk, non-uniform cost, hypothesis testing, bootstrap, Dirichlet priors

Most learning algorithms are not amenable to complete mathematical analysis. In spite of proven performance, proofs and constructive theories on how to choose classifier parameters or assess performance are rare. We present and compare bootstrap and Bayesian methods for assessing the expected risk of single classifiers' decisions, and analyze the estimation and correction of small values in the confusion matrix. We extend our techniques to the paired comparison of classifiers, of importance in industrial applications. How much evidence is there in favor of one classifier over another? What are the right criteria and metrics to compare classification decisions? Preliminary experiments show that the correction of small entries with uniform priors can lead to large errors in the estimated risk. We propose and study alternatives.

# 521 Semiparametric and Nonparametric Modeling and Goodness-of-Fit Tests for Longitudinal Data ● ✪

WNAR, Biometrics Section, Section on Nonparametric Statistics
**Thursday, August 10, 10:30 am–12:20 pm**

## Analysis of Longitudinal Data with Semiparametric Estimation of Covariance Function

❖ Runze Li, The Pennsylvania State University, Department of Statistics, University Park, PA 16802-2111, *rli@stat.psu.edu*; Jianqing Fan, Princeton University; Tao Huang, Yale University

***Key Words:*** kernel regression, local linear regression, profile weighted least squares, semiparametric varying coefficient model

Two important aspects in analysis of longitudinal data are improving efficiency for regression coefficients and predicting trajectories of individuals. Both involve estimation of the covariance function. A class of semiparametric models for the covariance function is proposed by imposing parametric correlation structure while allowing nonparametric variance function. We introduce semiparametric-varying coefficient partially linear models for longitudinal data and propose an estimation procedure for their regression coefficients by using a profile-weighted least squares approach and an estimation procedure for variance function and parameters in correlation. We will study the sampling properties of the proposed procedures and assess their finite sample performance by Monte Carlo simulation. A real data example is used to illustrate the proposed methodology.

## Bayesian Model Assessment Using Pivotal Quantities

❖ Valen Johnson, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 448, Houston, TX 77030-4009, *vejohnson@mdanderson.org*

***Key Words:*** prior-predictive density, posterior-predictive density, Bayesian chi-squared test

Suppose that S(Y,theta)$ is a function of data Y and a model parameter theta, and suppose that the sampling distribution of S(Y,theta) is invariant when evaluated at theta(0), the true (i.e., data-generating) value of theta. Then S is defined to be a pivotal quantity, and I show that the distribution of S(Y,theta) is identical to the distribution of S(Y,theta(Y)), where theta(Y) is a value of theta drawn from the posterior distribution given Y. This fact is used to investigate the properties of a number of Bayesian model diagnostic and assessment tools.

### Consistent Model Selection and Goodness-of-Fit Test for Marginal Regression Analysis of Longitudinal Data

❖ Lan Wang, University of Minnesota, 385 Ford Hall, 224 Church Street SE, Minneapolis, MN 55347, *lan@stat.umn.edu*; Annie Qu, Oregon State University

***Key Words:*** GEE, marginal model, lack-of-fit, model selection

An important approach to analyze longitudinal data is to model the marginal expectation and treat the correlation as a nuisance, such as the method of GEE. The validity of this approach depends on the correct specification of the marginal regression model. We proposes a BIC type procedure to select a marginal regression model from a group of candidate models. The procedure uses the quadratic inference function (QIF) recently proposed by Qu, Lindsay and Li (2000) and does not need to specify the full likelihood. We establish that with probability approaching to one the proposed procedure selects the most parsimonious correct model. Furthermore, a data-driven smooth test is constructed to check the adequacy of a proposed marginal model. Simulation results and data analysis are provided.

### Semiparametric Modeling in Applications

❖ Naisyin Wang, Texas A&M University, Department of Statistics, College Station, TX 77843-3143, *nwang@stat.tamu.edu*

Due to their flexibility and easy implementation, various semiparametric models have been used more often in recent biological and medical studies. These models allow underlying trends of the responses to be unspecified and nonparametric. In this talk, I will discuss recent applications of non- and semiparametric modeling (e.g., a membrane protein clustering tendency investigation). Theoretical support behind the methods will be discussed briefly. I also will use examples and simulations to illustrate the connections between the theoretical findings and their implication in applications.

## 522 What Makes a Successful Career in Statistics? Ruminations and Advice from Veteran Statisticians who have Made Major Contributions Over the Long Haul ●

Social Statistics Section, Section on Statistical Education

**Thursday, August 10, 10:30 am–12:20 pm**

### Good (?) Advice for Young Statisticians

❖ Robert V. Hogg, The University of Iowa, 241 SH, Department of Statistics, Univ. of Iowa, Iowa City, IA 52242, *bhogg@mccoymail.net*

***Key Words:*** fun, opportunities, kiss, DO, perseverance, thank

During 55 years of teaching, I have found the following points---which will be illustrated from my own experience---to be of great importance: (1) Do something you enjoy; that is, have fun. (2) Look for (even create) opportunities. (3) "Vague it up," remembering KISS: Keep It Simple Statistician. (4) "Anything worth doing is worth doing poorly at first." Thus, you must follow through with point 5. (5) Have patience and perseverance. (6) In most cases, you cannot do it alone and must recognize those who have helped you.

### Why Some Statisticians Never Die or Fade Away: Reflections on the Career of Joe Waksberg

❖ David Morganstein, Westat, 1650 Research Blvd., RE492, Rockville, MD 20850, *davidmorganstein@westat.com*; Daniel Levine, Westat; Ed Bryant, ; Graham Kalton, Westat

***Key Words:*** survey research, survey design, retired, satisfaction

This paper presents selected recollections from the career of Joseph Waksberg and describes why he and some of the various 'veteran' contributing authors continued working in the field of survey research and survey design long after a 'normal' retirement age.

### A Personal "10 Commandments" for a Successful Career in Biostatistics

❖ Theodore Colton, Boston University, 715 Albany Street, School of Public Health, Boston, MA 02118, *tcolton@bu.edu*

***Key Words:*** biostatistician, successful career

Over my half-century career, I have evolved a set of 10 personal suggestions---or a '10 Commandments'---for success in biostatistics. Most are not specific for biostatistics and apply to any field; several, however, are particular for biostatistics. The 10 are gain objective knowledge of your strengths and weaknesses; seek guidance from trusted peers and mentors; savor the joy of achievements and weather the bitterness of disappointments; have faith in serendipity; learn how to seize opportunities; know how to deal with prima donnas; hone your interpersonal skills; avoid those who seek to exploit, abuse, and misuse statisticians; be wary of those with "a gold mine of data just waiting for you to mine them"; and know when to let go. For each 'Commandment,' I cite from my own career instances when I did and did not follow the suggestion and the resulting consequences.

## 523 Statistical Issues in Emerging Areas in Cancer Research ● ✪

Biometrics Section, WNAR, ENAR

**Thursday, August 10, 10:30 am–12:20 pm**

### DNA Methylation, Aging, and Cancer

❖ Kimberly Siegmund, University of Southern California, Los Angeles, CA 90089, *kims@usc.edu*; Paul Marjoram, University of Southern California; Darryl Shibata, University of Southern California

***Key Words:*** population genetics, most recent common ancestor, simulation, rejection methods

Researchers are familiar with Knudsen's two-hit model for carcinogenesis, but may not realize one of these hits may be via DNA methylation, not mutation. DNA methylation is an epigenetic modification of DNA that does not involve a change in DNA sequence and is inherited during cell division. At certain genomic sequences, epigenetic errors accumulate at much faster rates than mutations. Such errors primarily accumulate in stem cells, the only long-lived cells in dividing tissues. Little is known about human stem cells due to experimental limitations. We explore a population genetic approach to reconstruct human stem cell histories from methylation errors that occur during genome replication, using methylation as a somatic cell "molecular clock." We simulate data under different population models and apply rejection methods to determine which models fit real data the best.

### Molecular Classification of Prostate Tumors

❖ Jaya M. Satagopan, Memorial Sloan-Kettering Cancer Center, 307 E. 63rd Street, New York, NY 10021, *satagopj@mskcc.org*

***Key Words:*** logistic regression, risk prediction, nomogram, prostate cancer

Approximately 25%--40% of prostate cancer patients undergoing radical prostatectomy experience disease recurrence. Predicting which patients are likely to recur can be valuable for proper disease management. The nomogram estimates the probability of recurrence in a defined time period given the clinical characteristics of a patient. However, the predictive performance of the nomogram needs improvement. Disease outcome, such as prostate cancer recurrence, is a complex combination of various genomic and exposure factors. We develop a prediction tool based on penalized logistic regression to determine five-year disease recurrence among prostate cancer patients using genomic risk factors and other clinical variables. The predictive performance of the resulting model is compared to a model based solely on genomic factors and a model based on clinical variables alone.

### Composite MicroRNA Target Predictions and Comparisons of Several Prediction Algorithms

❖ Jin Zhou, The Ohio State University, Mathematical Biosciences Institute, Columbus, OH 43210, *jzhou@mbi.osu.edu*; Vincent Melfi, Michigan State University; Joe Verducci, The Ohio State University; Shili Lin, The Ohio State University

***Key Words:*** microRNA, target (gene) predictions, composite statistics, gene ontology, Kendall's distance, Spearman's footrule distance

MicroRNAs (or miRs) are noncoding RNAs whose role is to repress translation by regulating gene expression through binding to mRNA targets. There are computational algorithms for miR target predictions, but their results vary. Thus, it would be useful to consolidate these results to have a greater degree of certainty before engaging in costly experiments. We studied three popular algorithms, miRanda, TargetScan, and PicTar, systematically through the use of three measures of similarity and a statistical test on the gene ontology categories. Two composite statistics were also devised to combine and rank the composite target list. We applied these methods to all human miRs. Our results indicate that TargetScan and PicTar tend to have a greater degree of similarity. We also demonstrate that our composite statistics can be useful tools for short listing genes for downstream experiments.

## 524 Statistical Models of Natural Language Text

IMS, Section on Statisticians in Defense and National Security

**Thursday, August 10, 10:30 am–12:20 pm**

### Structured Prediction Problems in NLP

❖ Michael Collins, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Stata Center, Bldg 32-G484, Cambridge, MA 02139, *mcollins@csail.mit.edu*

Natural language processing (NLP) is an area of computer science concerned with the application of computational methods to linguistic data. Examples of application areas within NLP include automatic (machine) translation between languages, dialogue systems, and information extraction. NLP problems offer a rich problem domain for statistical approaches; they often require the modeling of complex, discrete structures such as strings, labeled sequences, or trees. Generative statistical models, such as hidden Markov models or probabilistic context-free grammars, are a very common approach for this kind of problem. In this talk, I'll focus on recently proposed alternatives to generative statistical models. In particular, I'll describe generalizations of discriminative linear models, such as support vector machines or the perceptron algorithm, to structured problems found in NLP.

### Statistical Approaches for Machine Translation

❖ Dan Klein, University of California, Berkeley, Computer Science Division, 775 Soda Hall, Berkeley, Berkeley, CA 94720-1776, *klein@cs.berkeley.edu*

Statistical approaches have dominated recent work on automatic (machine) translation. Parameters of a statistical model are inferred from a large corpus of example translations. Developing statistical models for translation presents several challenges. First, the task involves substantial latent structure: Our data may tell us that one sentence translates to another, but it does not tell us what the specific word- or phrase-level correspondences are. Second, the combinatorial structures involved, such as matchings between sequences and trees, create additional difficulties for parameter estimation and inference. Finally, many alternative translations may be acceptable for a given sentence.

## 525 Population-Based Genetic Association Studies: Using Genetics/Genomics To Advance Public Health ●

Section on Statistics in Epidemiology, Section on Health Policy Statistics, WNAR, ENAR

**Thursday, August 10, 10:30 am–12:20 pm**

### Gene-Environment Interaction: Implications for Public Health and Methodological Challenges

❖ Ramal Moonesinghe, Centers for Disease Control and Prevention, Office of Genomics and Disease Prevention Coordinating Center for Health Promotion, MS-K89 4770 Buford Hwy NE, Atlanta, GA 30341-3717, *Rmoonesinghe@cdc.gov*; Muin Khoury, Centers for Disease Control and Prevention

Concerns have been raised about the value of genomic research for disease prevention and public health, especially involving complex diseases with risk factors amenable to environmental modification. The identification and characterization of genes influencing the risk of multifactorial disease through interactions with other genes and environmental factors remains a challenge. Multifactor dimensionality reduction was developed to detect high-order interactions among genetic and environmental variables associated with disease susceptibility. Neural network methods have been used to improve the identification of gene combinations associated with disease risk. These applied methods are important because they could help stratify disease risks, target interventions to achieve overall population health benefits, and help identify new environmental risk factors for disease.

### Candidate Gene Association Studies: Power and Sample Size Requirements

❖ Bingshu E. Chen, National Cancer Institute, 6120 Executive Blvd., Room 8033, Rockville, MD 20852, *cheneric@mail.nih.gov*; Philip S. Rosenberg, National Cancer Institute

***Key Words:*** case control study, genetic epidemiology, haplotype frequency test, power and sample size, single nucleotide polymorphisms

Power and sample size requirements are a critical design issue in genetic case control association studies. This problem is complicated by correlations among nearby SNPs (LD), and different statistical methods are available to test for association (i.e., SNP-based tests that adjust for multiple comparisons and haplotype frequency tests). In this study, we propose guidelines for power and sample size using the omnibus test that combines the min P test for SNP analysis and the 'directed' haplotype frequency test; the former is sensitive to scenarios where disease susceptibility is conferred by a SNP and the latter to scenarios where disease susceptibility is conferred by a common haplotype. In simulation studies to determine guidelines, we incorporate empirical LD patterns of SNPs and SNP haplotypes using genetic data on candidate genes obtained by the International HapMap Project.

### Diversity of Haplotype Configurations in the Human Genome: Implications for Genome-Wide Association Studies

❖ Shuying S. Li, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave., N., Seattle, WA 98109, *sli@fhcrc.org*

The International HapMap Consortium and Perlegen, Inc., independently completed more than 1 million SNP genotypes on individuals with African, European, and Asian ancestors. The availability of both sources of genotype data allows the research community to investigate aspects of human genome variations and aim to genome-wide association studies. In this talk, our focus is on assessing diversity of haplotype configurations among African, European, and Asian populations. Using the chi-square test, we tested the significance of the diversity. The diversity of haplotype configuration in the human genome will implicate the study design and analysis planning in future association studies of human populations, especially in admixture population association studies. We hope our method and results will provide guidance for those studies' design and analysis planning.

### Risk Estimation in Persons at Genetic Risk of Cancer

❖ Sining Chen, Johns Hopkins Bloomberg School of Public Health, MD 21205, *sichen@jhsph.edu*; Edwin S. Iversen, Jr., Duke University; Giovanni Parmigiani, Johns Hopkins University

***Key Words:*** gene characterization, risk estimation, cancer epidemiology, BRCA, Bayesian hierarchical model

Once a causal gene is identified, effective characterization of that gene becomes immediately important. We will present two approaches for characterizing rare genetic variants using family data ascertained based on family history. The retrospective likelihood approach uses the likelihood of the genetic test result conditioning on the family history. It takes full advantage of the Mendelian transmission of mutated alleles and provides unbiased estimates. The ascertainment-conditional likelihood approach conditions on the event of ascertainment and models the probability of ascertainment through a family history summary variable. This approach provides improved efficiency over the retrospective likelihood and remains unbiased under appropriate modeling of the ascertainment event. We will present a metaanalysis of BRCA penetrance studies using a Bayesian hierarchical approach.

## 526 Survival Analysis with Medical Applications: Parametric and Nonparametric Adjustment for Survival Function

Biopharmaceutical Section, Biometrics Section, ENAR, WNAR

**Thursday, August 10, 10:30 am–12:20 pm**

### Evaluating the Added Value of an Expensive Marker in Predicting Survival

❖ Tianxi Cai, Harvard University, 655 Huntington Ave., Boston, MA 02115, *tcai@hsph.harvard.edu*

***Key Words:*** survival prediction, accuracy measures, transformation model, model selection

Suppose an expensive marker Y may be obtained in addition to a set of routine predictors X for predicting survival T. We are interested in evaluating the added value of Y in predicting the failure status at t, $D=I(T < t)$ and determining when and whether Y is needed to improve the prediction. To quantify the added value of Y, we propose to first consider working models, $E(D|X)=g(a'X)$ and $E(D|W)=g(b'W)$, to construct prediction rules. Then we evaluate the added value of Y for given $a'X = x$, by the reduction in the prediction error, $V(x) = E\{|D-g(a'X)| \mid a'X=x\}-E\{|D-g(b'W)| \mid a'X = x\}$. We develop procedures for making inference about V(x) and for selecting subsets of population that would benefit from having the Y measurement.

### Adjusted Nelson-Aalen Estimator with Inverse Probability of Treatment Weighting

❖ Jun Xie, Purdue University, Department of Statistics, 150 N. University Street, West Lafayette, IN 47907, *junxie@stat.purdue.edu*; Chaofeng Liu, Eli Lilly and Company

***Key Words:*** adjusted Nelson-Aalen estimator, hazards rate, inverse probability of treatment weighting, weighted log-rank test

Estimation and group comparison of survival rates are two very common issues in survival analysis. In nonrandomized clinical trials or observational studies, the Nelson-Aalen estimator of hazards rate function may be biased due to unbalanced distributions of confounding variables. We develop an Adjusted Nelson-Aalen Estimator (ANAE) with Inverse Probability of Treatment Weighting (IPTW). Each observation is weighted by its inverse probability of being in a certain group. The ANAE estimates marginal hazards rate function without the assumption of proportionality. It is a consistent estimate. A weighted log-rank test is developed for comparing group difference of hazards rates. Simulation and real data studies illustrate the ANAE and weighted log-rank test outperform the Nelson-Aalen estimate, and it does better than or as well as other estimators based on stratification.

### Split Point Estimation for Logistic Regression

❖ Ian McKeague, Columbia University, 455 Central Park W., Apt 8-D, New York, NY 10025, *im2131@columbia.edu*

***Key Words:*** nonparametric regression, cube-root asymptotics, maximum likelihood, changepoints, misspecified, non-normal limit

We investigate the problem of finding confidence sets for split points in a general class of nonparametric regression models. In particular, we study a nonparametric logistic regression model of the form logit(p)=f(X), where X is the univariate predictor and a binary decision tree for the regression function, f, is used as a working model. We are interested in the behavior of the maximum likelihood estimator of the split point under misspecification of the working model by a smooth regression function, f. The asymptotic distribution of such estimators of the split point is determined and used to construct confidence sets. Cube-root asymptotics with non-normal limit distributions are involved. This is joint work with Mouli Banerjee.

## 527 Misuse of Statistical Methodology ● ✪

Committee on Applied Statisticians, Section on Statistical Education, Section on Statistical Consulting, Section on Teaching Statistics in the Health Sciences
**Thursday, August 10, 10:30 am–12:20 pm**

### Users of Statistics Beware!

❖ Gipsie B. Ranney, Belmont University, 1040 Beech Tree Lane, Brentwood, TN 37027-7642, *gbranney@comcast.net*; ❖ Michael Tveite, Guidant Corporation, 4100 Hamline Ave., N., St. Paul, MN 55112, *tveite.michael@guidant.com*; ❖ Ian Stuart Bradbury, Peaker Services, Inc., 8080 Kensington Court, Brighton, MI 48116-8591

***Key Words:*** statistical inference, pitfalls, training practitioners, enumerative and analytic studies

With the “democratization” of statistics, statistical methods are being used more extensively by nonstatisticians. We favor this trend; however, it makes it increasingly important that we impress upon practitioners the assumptions and limitations of formal statistical inference methods and make these methods maximally robust to potential misuse. In this talk, we will discuss and illustrate concepts we need to impress upon users of statistical inference, including the differentiation between enumerative and analytic studies, the dangers of extrapolation, and the misuse of statistical significance.

## 528 Assessing Treatment Effect in Cardiovascular Medical Device Clinical Trials ●

Biopharmaceutical Section, Biometrics Section, ENAR
**Thursday, August 10, 10:30 am–12:20 pm**

### Using Different Multiple Imputation Strategies in Estimating Treatment Effects in Drug-Eluting Stent Trials

❖ Liang Li, Boston Scientific Corporation, 108 Francis Ave., Shrewsbury, MA 01545, *liang.li@bsci.com*; Helen Chmiel, Boston Scientific Corporation; Hong Wang, Boston Scientific Corporation

***Key Words:*** missing data, multiple imputations, missing dependent variable

Drug-eluting stents (DES) have been shown to be more effective in reducing target lesion revascularization than bare metal stents. Intravascular ultrasound (IVUS) data have confirmed the consistent performance of DES in suppressing neointimal hyperplasia in patients with analyzable IVUS data. The attrition rate in IVUS follow-up data typically has been in the range of 20--30%. This investigation is to impute missing values in IVUS data using different multiple imputation strategies. Each imputed dataset is analyzed separately using mixed linear model to handle missing dependent variable, and results of separate analyses are combined to account for variation within and between the imputed datasets. This results in valid statistical inferences that reflect the uncertainty introduced by the missing values.

### Survival Analysis of Repeated Events in Drug-Eluting Stent Trials

❖ Yongyi Yu, Boston Scientific Corporation, 15620 76th Place, N., Maple Grove, MN 55311, *yua@bsci.com*; Jian Huang, Boston Scientific Corporation; Brian Johnson, Boston Scientific Corporation; Hong Wang, Boston Scientific Corporation

***Key Words:*** Kaplan-Meier estimator, recurrent events, independent censoring, survival distribution

In a series of TAXUS drug-eluting stent trials, there was overwhelming evidence showing that TAXUS Stent can improve the event-free rate on the first TLR at one year. However, repeated events in both the TAXUS Stent and Bare Metal Stent (BMS) arms are observed. This paper tries to explore different methods to estimate and interpret the survival distribution and effect of TAXUS Stent on the survival time between first and second events.

### Bootstrap Generalized Estimating Equation for Treatment Differences for Randomization Trials with Correlated Data

❖ Duo Zhou, Medtronic, Inc., 1015 Gramsie Road, MS Z240, Saint Paul, MN 55126, *duo.d.zhou@medtronic.com*; Zengri Wang, Medtronic, Inc.

***Key Words:*** bootstrap, correlated data, generalized estimating equation, multiple events

Generalized estimating equations are used commonly to estimate differences in treatment success rates differences in randomization trials with events occurring multiple times within patients. A naÔve estimation of treatment effects without considering within-patient correlation may lead to biased estimates, especially when the event distribution is highly skewed. We consider a bootstrap-based estimating equation method for estimating treatment differences between groups. Results from several commonly used approaches are compared with respect to point estimates and confidence intervals. Simulation studies are carried out to assess the sensitivity of the methodology. An example involving a device trial with an implantable cardioverter defibrillator will be given to illustrate the details of calculation.

### Gender by Treatment Interaction on Coronary Lesion Revascularization in Drug-Eluting Stent Trials

❖ Zheng Zhou, Boston Scientific Corporation, 1 Boston Scientific Place, CP4, Natick, MA 01760, *zhouf@bsci.com*; Liang Li, Boston Scientific Corporation; Hong Wang, Boston Scientific Corporation

***Key Words:*** treatment-by-gender interaction, logistic regression, drug-eluting stent

The use of drug-eluting stents has reduced target lesion revascularization (TLR) after coronary angioplasty significantly. It has been of clinical interest whether there is a "gender effect" in outcomes associated with this treatment. Using data from large-scale randomized trials of drug-eluting stent, the magnitude of treatment effect of drug-eluting stent versus bare-metal stent was examined in men and women. Multiple logistic regression models were used to analyze gender by treatment interaction with and without controlling for patient baseline and lesion characteristics. The results showed that after adjustment, there was no significant difference in treatment outcomes between the groups. Men and women appeared to enjoy equally the clinical benefit associated with drug-eluting stent treatment.

## 529 New Technology for Data Collection

Section on Government Statistics
**Thursday, August 10, 10:30 am–12:20 pm**

### The Impact of Using Hand-Held Computers To Reduce Count Imputation in the 2004 Census Test

❖ Jamie Burnham, U.S. Census Bureau, , *Jamie.L.Burnham@census.gov*; Rosemary Byrne, U.S. Census Bureau

***Key Words:*** monitoring nonresponse, nonresponse followup, computer assisted interviewing, mobile computing environments

The 2004 Census Test used hand-held computers (instead of paper questionnaires as in Census 2000) to collect data for personal visit interviews for the Nonresponse Followup (NRFU) Operation. During the NRFU Operation, enumerators followed up households that did not return their 2004 Census Test questionnaires. In 2000, hot deck imputation was used to impute housing unit status, household occupancy status, and/or household size when not reported or when there were data inconsistencies. The 2004 plan to reduce the need for count imputation consisted of using automation to more closely monitor addresses missing this information, using built-in edits in the questionnaire, and collecting status and population count for refusal situations. This paper presents the results of these changes on 2004 count imputation, compared to the Census 2000 rates in the corresponding NRFU areas.

### Staffing and Infrastructure Issues Related to Hand-Held Computer-Based Data Collection in the 2004 Nonresponse Follow-up Operation

❖ Darlene Moul, U.S. Census Bureau, 2206 Traies Court, Alexandria, VA 22306, *darlene.a.moul@census.gov*; Geraldine Burt, U.S. Census Bureau

***Key Words:*** hand-held computer, infrastructure, staffing, training, space

The 2004 Census Test was the first time the U.S. Census Bureau attempted to use hand-held computing devices for interviewing in the Nonresponse Followup operation. We were concerned that the use of hand-held computers would change both field staffing and data collection infrastructure requirements. Bureau staff hypothesized that the introduction of the hand-held computer would not only affect the job of the field enumerator, but also substantially change the nature of local census office-based operations and tasks involving management and support of the hand-held computer. This report addresses the lessons learned about the impact of using a hand-held computer for nonresponse follow up data collection activities on field staffing, training requirements, and space usage in the 2004 Census Test.

### Electronic Wage Reporting Customer Satisfaction: a Good News Case Study

❖ Stephen Dienstfrey, Schulman, Ronca & Bucuvalas, Inc., 8403 Colesville Road, Suite 820, Silver Spring, MD 20910, *s.dienstfrey@srbi.com*; Michael Greenberg, U.S. Social Security Administration

***Key Words:*** customer satisfaction, GPRA, Social Security Administration

In January 2003, the Social Security Administration unveiled an online service allowing employers to file the annual wage report of employees electronically. While this service, W-2 Online, is available to all employers with less than 250 employees who are not required to file electronically, it was aimed at the large number of businesses with 20 or fewer employees. Of the 5 million employers with 20 or fewer employees, only 35,000 used W-2 Online in 2004. Further, almost 3 in 5 of the paper W-2 forms were prepared using electronic software. Ten-minute telephone interviews were completed with random samples of paper and web filers. The results pointed out steps SSA could take to increase electronic filing by targeting nonusers of W-2 Online and by partnering with major accounting and tax preparation software developers for automatic filing of W-2 data.

### Evaluating the Collection of Global Positioning System Coordinates with Hand-Held Computers in the 2004 Census Test

❖ Diane Barrett, U.S. Census Bureau, 4701 Silver Hill Road, Building #2, Room 2410, Washington, DC 20233, *Diane.F.Barrett@census.gov*; Doug Olson, U.S. Census Bureau

***Key Words:*** GPS, geocode, hand held computers, non response followup

During the 2004 Census Test, enumerators were equipped with hand-held computers to follow up with households that did not return their census questionnaires. Enumerators used these computers to find their assigned work areas, obtain questionnaire data, and collect Global Positioning System (GPS) coordinates for each structure they visited. The primary goal of this evaluation was to determine if enumerators could correctly and consistently collect coordinates using newly developed field procedures. The evaluation also assessed if enumerator-collected GPS coordinates could be used for geocoding structures to census blocks and if the coordinates (displayed as map spots on the hand-held computer) enhanced the enumerators ability to return to or relocate the structure.

## 530 Statistics of Extremes ✪

IMS
**Thursday, August 10, 10:30 am–12:20 pm**

### Dependence Estimation and Prediction in Max-Stable Random Fields

❖ Daniel Cooley, Colorado State University, Department of Statistics, Fort Collins, CO 80523-1877, *cooleyd@stat.colostate.edu*; Philippe Naveau, Laboratoire des Sciences du Climat et de l'Environnement; Richard Davis, Colorado State University

***Key Words:*** extreme value theory, spatial statistics, geostatistics, variogram, madogram, kriging

Spatial data gives rise to two immediate questions: how to measure spatial dependence and how to perform spatial prediction. The field of geostatistics answers these questions using the variogram and kriging. However, geostatistical methods are not well-suited for extreme observations. To measure spatial dependence in max-stable random fields, we propose the madogram---a first-order variogram similar to geostatistical methods that benefits from a convenient relationship with multivariate extreme- value distributions and the extremal coefficient, an existing measure of dependence for extremes. We are working on the problem of spatial prediction and have investigated several predictors.

### Analyzing the Extreme Behavior of Large-Scale Meteorlogical Variables Found To Have Influence on Severe Storms and Tornadic Events Using Global Reanalysis Data

❖ Eric Gilleland, National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80301-2260, *ericg@ucar.edu*; Matt Pocernich, National Center for Atmospheric Research

***Key Words:*** extreme values, climate change, reanalysis data, spatial, temporal, GEV

Concurrently high values of convective available potential energy (cape) and 0--6 km vertical gradient of horizontal weight (shear) are found to occur with severe storms and tornadic events, which occur at fine scales that cannot be resolved by climate models, unlike cape and shear. Difficulties arise in analyzing these events (e.g., cape is typically zero, high amounts of both do not occur often at the same time, the spatial structure is highly nonstationary). Initial analysis is on global reanalysis data on a 1.9 by 1.9 degree grid (approx 17,000 points) every 6 hours from June 1, 1957, through December 31, 1999. The relationship of cape and shear is inspected and significant changes over time are investigated, accounting for the false discovery rate. Finally, data are fit to the GEV distribution individually at grid points using constrained MLE.

### Variance Reduction in Multiparameter Likelihood Models

❖ Liang Peng, Georgia Institute of Technology, School of Mathematics, Atlanta, GA 30332-0160, *peng@math.gatech.edu*; Ming-Yen Cheng, National Taiwan University

***Key Words:*** bootstrap, extreme value distribution, local likelihood, local linear MLE, variance reduction

There is an increasing interest in employing multivariate likelihood models to investigate trends of sample extremes in environmental statistics. When sample maxima are modeled by a generalized extreme value distribution, the sample size is not large in general and local likelihood estimation exhibits a large variation. In this paper, variance reduction techniques are employed to improve the efficiency of inference. An application to annual maximum temperature shows our methods are effective.

### Bootstrap Confidence Band for Dependence Functions in Extreme-Value Statistics

❖ Yongcheng Qi, University of Minnesota, Duluth, Department of Mathematics and Statistics, 1117 University Drive, Duluth, MN 55812, *yqi@d.umn.edu*; Liang Peng, Georgia Institute of Technology

***Key Words:*** confidence band, dependence function, bootstrap, extreme value

Bivariate extreme-value theory was used to estimate a rare event (de Haan and Ronde, 1998). This procedure involves estimating a tail dependence function. There are several estimators for the tail dependence function in the literature, but their limiting distributions depend on partial derivatives of the tail dependence function. In this paper, we employ bootstrapping to construct the confidence band for the tail dependence function based on Huang's (1992) estimator. A simulation study is conducted to assess the accuracy of the bootstrap approach.

### Asymptotically (In)dependent Multivariate Maxima of Moving Maxima Processes

❖ Zhengjun Zhang, University of Wisconsin-Madison, Department of Statistics, 1300 University Ave, Madison, WI 53706, *zjz@stat.wisc.edu*

***Key Words:*** asymptotic (in)dependence, multivariate time series

Smith and Weissman introduced an M4 class of processes, which are flexible models for temporally dependent multivariate extreme-value processes. However, all variables in these models are asymptotically dependent. This paper extends this M4 class in a number of ways to produce classes of models that are also asymptotically independent. We will study properties of the proposed models. In particular, asymptotic dependence indexes, coefficients of tail dependence, and extremal indexes are derived for each case.

## 531 Bayesian Biostatistical Modeling ●

Section on Bayesian Statistical Science, Biometrics Section, WNAR, ENAR

**Thursday, August 10, 10:30 am–12:20 pm**

### A Bayesian Adaptive Design for Two Drug Combination Phase I Clinical Trials

❖ Lu-May Chiang, University of California, Los Angeles, 11130 Rose Ave., Apt 302, Los Angeles, CA 90034, *lmchiang@ucla.edu*; William G. Cumberland, University of California, Los Angeles

***Key Words:*** DLT, MCMC, multivariate probit model

Most cancer therapies invoke multiple drugs given to patients simultaneously. However, most traditional methods are designed to search for the MTD for one drug with a fixed dose of other drugs present. Shortcomings of the existing designs include combining the observed toxicities into a dichotomous outcome, ignoring the information on the severity of the toxicities, focusing on a single drug, and not taking into account the correlation between toxicities. The methods we propose are motivated by these limitations. The prespecified dose levels of the two drugs under investigation form a dose combination grid. Our goal is to estimate the MTD boundary of the dose combination grid, taking into account how two drugs interact in terms of toxicities. We assume these drugs have synergistic effects in terms of toxicity.

### A Bayesian Model Selection Approach for Simultaneous Clustering of Treatment Intercepts and Slopes

❖ Susan Alber, University of California, Los Angeles, 6081Pickford Place 2B, Los Angeles, CA 90035, *alber@ucla.edu*; Robert Weiss, University of California, Los Angeles

***Key Words:*** model selection, partitioning, clustering

We present a model selection approach to clustering treatment groups. Each partition of the groups into subsets with equal means is viewed as a separate model, and all possible partitions are incorporated into a single hierarchical model. The model will be expanded to include interaction with a continuous covariate by adding treatment-specific slope parameters. Each model now contains two separate partitions of the treatment groups: one into treatment clusters with equal intercepts and one into treatment clusters with equal slopes. The method will be demonstrated using a four-group categorical treatment from a study on orthodontia fixtures.

### Hierarchical Zero-Inflated Poisson Models

❖ Warren Comulada, Center for Community Health, 10920 Wilshire Blvd., Suite 350, Los Angeles, CA 90024, *scomulad@ucla.edu*; Robert Weiss, University of California, Los Angeles

***Key Words:*** zero-inflated Poisson data, longitudinal count data

Substance use is measured in HIV behavioral interventions. Count measures for the number of times of use often are zero-inflated with more zeros than can be explained by a Poisson distribution. The ZIP regression is a model for zero-inflated count data. A zero is allowed to come from two processes: With probability p one process---the zero state---has zeros as the only possibility, and with probability 1 - p---the other process---has Poisson-distributed counts. In longitudinal data, observation may be in the zero-state across all time points or at individual time points. We incorporate the zero hierarchy into hierarchical ZIP (HZIP) models for ZIP data and extend the HZIP models to multivariate ZIP data; outcomes are correlated through random effects. We illustrate the HZIP models on substance use measures from an HIV behavioral intervention study. Models are fit into a Bayesian framework.

### Using a Bayesian Phylogenetic Dual Change-Point Model To Detect HIV-1 Intrahost Recombination

❖ Erik Bloomquist, University of California, Los Angeles, 1547 S. Westgate Ave., Apt 4, Los Angeles, CA 90025, *ebloomqu@ucla.edu*; Christina Kitchen, University of California, Los Angeles; Marc A. Suchard, University of California, Los Angeles; Vladimir N. Minin, University of California, Los Angeles; Hua Guo, University of California, Los Angeles

***Key Words:*** Bayesian, phylogenetics, recombination, HIV

Recent advances in Bayesian phylogenetics have utilized dual change-point models to detect changes in evolutionary pressure and tree topology along an alignment. We applied a model (Minin et al. 2005) to HIV-1 sequence data to test for intra-host recombination. The data involve multiple HIV-1 clones from three subjects, sequenced at specific times during disease. Maximum likelihood analysis showed differing tree topologies for different gene regions, suggesting possible recombination. Maximum likelihood techniques were used subsequently to identify possible recombinant progeny sequences. Each progeny strain, along with a set of possible parental strains, was then entered into the Bayesian change-point model to identify if recombination took place and the location of the breakpoints. The Bayesian model also gave evidence for intra-host recombination.

### Hierarchical Post-Processing of Longitudinal Bayesian Phylogenetic Analyses Using Semiparametric Regression Models

❖ Li-Jung Liang, University of California, Los Angeles, 3161 Sepulveda Blvd., Biostatistics, Apt 208, Los Angeles, CA 90034, *liangl@ucla.edu*; Robert Weiss, University of California, Los Angeles; Marc A. Suchard, University of California, Los Angeles

***Key Words:*** longitudinal phylogenetic analysis, MCMC, importance reweighting, HIV-1

We collected multiple HIV gene sequences from HIV+ subjects. For each patient, a Bayesian phylogenetic model was used to estimate the phylogenetic parameters. This was repeated over time with different numbers of visits per subject. We applied semiparametric Bayesian hierarchical random effects regression models to combine the individual subject-visit phylogenetic analyses to better estimate the parameters of interest within and across patients. These models allow for covariate and time effects on the phylogenetic parameters of interest. We use the existing posterior samples from each analysis, applying an iterative reweighting/resampling MCMC algorithm.

## 532 Applications of Modeling and Simulation for Homeland Security ✪

Section on Statisticians in Defense and National Security

**Thursday, August 10, 10:30 am–12:20 pm**

### Modeling the Interaction between Intelligent Site Selection and Other Stochastic Processes with Applications to Terrorism

❖ Michael Porter, University of Virginia, 1038 Earls Road, Goochland, VA 23063, *mdp2u@virginia.edu*; Charles D. Robinson, University of Virginia; Donald E. Brown, University of Virginia

***Key Words:*** intelligent site selection, space-time Cox point process, terrorism, improvised explosive device (IED)

This work seeks to combine an intelligent site selection (ISS) point process with other space-time stochastic processes and model their interactions. The rationale for this modeling effort is to study the placement and subsequent detonation or discovery of improvised explosive devices (IED) commonly used by terrorists. The ISS process, in this situation, is one in which the terrorists judiciously select the locations and times to place IEDs. For an IED attack to be successful, the device must interact with a potential target prior to being detected and cleared. Thus, the target movements and detection efforts are the additional stochastic processes considered. The ISS process is modeled as a space-time Cox point process with the interaction resulting in a thinned point process. A result of this modeling effort is to observe the emergent behavior of agents acting within this framework.

### Issues in the Analysis of Combat Data

❖ Michael Last, National Institute of Statistical Sciences, 19 TW Alexander Drive, Research Triangle Park, NC 27709, *mlast@niss.org*

***Key Words:*** military history, classification, national defense, homeland security

Large datasets of historical battles containing both their outcomes (victor and casualty levels) and a number of possible predictors (weather, nationality, force size, year, surprise) are available for analysis, yet surprisingly little work on them has been done in the statistical community. The current state-of-the-art in the military science community involves using linear regression, and, by implication, linear discriminant analysis. I will discuss background information and my efforts to better these analyses with modern classification techniques.

### Strategies for Validating Complex Simulations

❖ David Banks, Duke University, ISDS Box 90251, Duke University, Durham, NC 27708, *banks@stat.duke.edu*

***Key Words:*** simulation, agent-based models, validation, parsimony

Many military offices (e.g., DMSO) rely upon complex simulations for planning. These simulations may involve agent-based models, systems of differential equations, or complex hybrids. Also, these simulations typically involve large numbers of parameters and are highly nonlinear. This means traditional use may be unreliable; even sensitivity analysis is hard to do. This talk will consider methods for understanding and assessing such models based on the intrinsic parameterization, local low-dimensional behavior, and parsimony. No method is perfect in such situations, but there are strategies for testing face-validity and obtaining partial corroboration of results.

### The Brave New World of Designing Simulation Experiments for Defense and Homeland Security Applications

❖ Thomas W. Lucas, Naval Postgraduate School, 1411 Cunningham Road, Bldg 302 Room 236, Monterey, CA 93943, *twlucas@nps.edu*; Susan M. Sanchez, Naval Postgraduate School

***Key Words:*** simulation, design of experiments, validation, exploratory modeling, Latin hypercubes

The U.S. Department of Defense relies heavily on simulation models in its decisionmaking. Simulations assist in developing war plans, deciding what equipment to acquire, and determining how to organize and employ forces. Most of these simulations are extremely large and have many uncertain input variables. Also, their veracity is difficult to assess. This talk will examine our approaches to generating useful information from such models. We also will introduce a portfolio of designs we are developing and applying to efficiently obtain insights into a variety of defense and homeland security issues.

## 533 Record Linkage ✪

Section on Survey Research Methods, Section on Statisticians in Defense and National Security

**Thursday, August 10, 10:30 am–12:20 pm**

### Automatically Estimating Record Linkage False Match Rates

❖ William E. Winkler, U.S. Census Bureau, Statistical Research Division, Washington, DC 20233-9100, *william.e.winkler@census.gov*; William E. Yancey, U.S. Census Bureau

***Key Words:*** EM algorithm, unsupervised learning, semi-supervised learning

This paper provides a mechanism for automatically estimating record linkage false match rates in situations where the subset of the true matches is reasonably well-separated from other pairs. The method provides an alternative to the method of Belin and Rubin (1995) and is applicable in more situations. We provide examples demonstrating why the general problem of error rate estimation (both false match and false nonmatch rates) is likely impossible in situations without training data and exceptionally difficult in the extremely rare situations where training data are available.

### Record Linkage and Automatic Maintenance Activities

❖ Holly Smith, U.S. Department of Agriculture, 1400 Independence Ave., SW, Room 6344A South Building, Washington, DC 20250-2023, *holly_smith@nass.usda.gov*; Denise Abreu, National Agricultural Statistics Service; Kara Daniel, National Agricultural Statistics Service; Stan Hoge, National Agricultural Statistics Service; Bill Iwig, National Agricultural Statistics Service

***Key Words:*** maintenance updates, record linkage, list frame

The National Agricultural Statistics Service's mission is to provide timely, accurate, and useful statistics in service to U.S. agriculture. To achieve this goal, NASS needs a reliable and efficient list sampling frame of agricultural producers in the United States. In 1997, responsibility for the Census of Agriculture was transferred to NASS from the U.S. Census Bureau. NASS's annual survey program has grown to include economic and environmental statistics. These responsibilities created the need to handle list maintenance activities more efficiently. This paper provides an overview of NASS and its procedures for utilizing record linkage for list-building and maintenance. It describes how NASS reviews records and highlights features of its RecLink application and discusses current record linkage projects utilized for maintaining demographic and control data updates.

### A Study of String Comparator Performance on Census Name Data

❖ William E. Yancey, U.S. Census Bureau, Statistical Research Division, Mail Stop 9100, Washington, DC 20233, *william.e.yancey@census.gov*

***Key Words:*** record linkage, string comparator, edit distance, ROC curve

We compare the performance of several string comparators on first and last name data from the clerically reviewed census and accuracy follow-up files from 2000 and 1990. We compare the Jaro-Winkler string comparator with and without optional enhancements and several edit-distance--based string comparators. We also consider a string comparator that combines the Jaro-Winkler and edit-distance approach. The main statistical comparison is based on areas under portions of the ROC curve (sensitivity vs. selectivity) for each of the comparators on each of the datasets of name pairs from the data files that have been judged to be from matching records but are not spelled identically. We consider the effect of the choice of string comparators with differing ROC-based scores on actual record linkage results.

### A Comparison of Big Match and the Current NDI Record Selection System Using Artificial Data

❖ Bryan Sayer, Social & Scientific Systems, Inc., 921 Chatham Lane, STE 100, Columbus, OH 43221-2418, *Bsayer@s-3.com*

***Key Words:*** National Death Index, record linkage, error rate

Using a frequency of sex-specific first names and middle initials and nonsex-specific last names from 22 years of the NDI (about 48 million people), we generate an artificial database that replicates the NDI structure and an artificial matching file that contains both decedents and nondecedents. The matching file also contains "altered" submission records that represent problems found in survey records matched to the NDI. We compare the results using a variety of relaxed blocking assignments in Big Match to estimated results under the current system for those records selected. We show which correct records are selected by one, the other, or both of the systems and which system works the best with which types of altered records.

### Using Test Databases To Evaluate Record Linkage Models and Train Linkage Practitioners

❖ Michael H. McGlincy, Strategic Matching, Inc., PO Box 334,
Morrisonville, NY 12962, *mcglincym@strategicmatching.com*

***Key Words:*** codes, crash, outcomes, probabilistic, record, linkage

Traffic safety research grantees develop Crash Outcome Data Evaluation Systems (CODES) by linking police and medical records. The process is complex because most records lack unique identifiers and exhibit high levels of misclassification and nonresponse. Grantees learn to develop Bayesian models that compare quasi-identifiers in order to estimate the probability that a record pair is a true link. To help teach effective modeling techniques, the CODES program uses test databases for which true link status is known. A data generator creates test databases for each grantee by simulating crash and medical events. Quasi-identifiers specific to each grantee are drawn from multinomial distributions. Case duplication, item misclassification, and item nonresponse are simulated as Bernoulli events. The generator captures clustering of data for vehicle occupants that occurs in real life.

## 534 Survey-Based Estimation VI ●

Section on Survey Research Methods
**Thursday, August 10, 10:30 am–12:20 pm**

### Detection Method for the Sources of Change in Estimations

❖ Serge Godbout, Statistics Canada, 120 Parkdale Ave., RHC
Building 11N, Ottawa, ON K1A 0T6 Canada,
*serge.godbout@statcan.ca*

***Key Words:*** outlier detection, derivatives, estimator

In a survey context, we often need to know and quantify the main sources of an observed change between two estimations across time. This allows us to explain changes or to point out outlying units with large impact on the change. Depending on the estimator, the type of data, or the number of units, this can be complex. Using implicit derivatives, we will describe a method to find the sources of change, usable in many contexts. Knowing the estimator formula, we need to consider the variables as functions of time, calculate the derivative of the estimator, and approximate the differentials using the differences of the variables. This will help quantify the contribution to the change by variables and units. The advantage of this simple method is to identify easily and precisely which micro-changes have a real impact among all the data.

### Model-Based Approach in Two-Stage Sampling of Audit Data

❖ Yan Liu, National Opinion Research Center, 1350 Connecticut
Ave., NW, Suite 500, Washington, DC 20036, *yliu00@yahoo.com*;
Mary Batcher, Ernst & Young LLP

***Key Words:*** jackknife, model-based approach, prediction approach, ratio estimation, stratified sampling, two-stage sampling

In an audit data setting, two-stage stratified sampling is typically done on relatively small populations and small samples. In the paper "Empirical Study on the Second-stage Sample Size" (Liu, Batcher, Petska and Luo, 2004) where a sampling design at both stages was used, the impact of the second stage sample size on the overall estimation was investigated empirically in terms of the bias and variance estimation. Different methods of the classical design-based approach and resampling approach were compared. In this paper, we look at this issue from the model-based approach. We construct models at the overall level and at the stratum level and compare their performances to design-based and resampling approaches. We also check the assumptions of error terms and investigate the gain from the stratification with PSU.

### Confidence Interval Coverage in Stratified Sample Design

❖ Jinhee Yang, Ernst & Young LLP, 1225 Connecticut Ave., NW,
Washington, DC 20036, *jinhee.yang@ey.com*; Wendy Rotz, Ernst &
Young LLP

***Key Words:*** confidence interval, model-based estimation, deep stratification, goodness-of-fit

In some business settings, the cost of reviewing a sample is extremely high, giving rise to the need for an accurate estimate and narrow confidence interval using a very small sample size. When there is a strongly related auxiliary variable, model-based estimation with deep stratification may be used. Generally, deep stratification tends to reduce the sampling error and is expected to produce conservative confidence intervals. This paper explores whether there are situations when the confidence interval coverage is too conservative, or not as robust as believed. We will study the effect of population distribution, model, model fit, and degrees of freedom on the performance of confidence interval coverage.

### Evaluating a Calibration Weighting Scheme for Elementary Indexes for Commodities and Services in the U.S. Consumer Price Index

❖ Sylvia G. Leaver, Bureau of Labor Statistics, 2 Massachusetts
Ave., NE, Room 3655, Washington, DC 20212, *Leaver_S@bls.gov*;
Robert A. Cage, Bureau of Labor Statistics; Darin T. Solk, Bureau of
Labor Statistics

***Key Words:*** consumer expenditures, simulation, ratio adjustment

The estimate of price change for an elementary cell of the U.S. CPI is a weighted average of quote-level price changes, where the quote-level weight is a function of inverse selection probabilities at each stage of selection. This paper evaluates, by means of retrospective estimation and sample simulation, an alternative weighting scheme employing calibration. The theory behind this approach is that the sum of the weights of each "useable" quote in each elementary cell should yield an estimate of the expenditure of the elementary cell. The CPI produces two distinct estimates of elementary item-area expenditure: (i) that derived from the sum of commodities and services quote weights, and (ii) that derived from Consumer Expenditure Survey (CEX) biennial data. The new formula forces quote weights in any given month to sum to biennial expenditure estimates from the CEX.

### Comparing Procedures To Estimate the Measure of Heteroscedasticity and Their Effect on Estimating Totals and Their Variances

❖ Kimberly Henry, Internal Revenue Service, P.O. Box 2608,
Statistics of Income Division, Washington, DC 20013, *Kimberly.
A.Henry@irs.gov*; Richard Valliant, University of Michigan

***Key Words:*** survey sampling, prediction approach, general linear models, simulation

Estimation of totals is often an objective in survey sampling. With a model-based approach, one factor that can affect the variance and bias of estimated totals is the superpopulation structure. We consider the case when a dependent variable's variance is proportional to some power, called the measure of heteroscedasticity (moh), of the independent variable. For various sample designs, sizes, and estimators, alternative strategies for estimating moh are compared. The strategies' effect on estimates of totals and their variances is then evaluated. Results are extended to a subset of 1040 tax returns taken from the Statistics of Income's Tax Year 2001 Individual Public Use File.

### Power-Shrink Excessive Weights Preferred to Trimming

Naihua Duan, University of California, Los Angeles; Xiao-Li Meng, Harvard University; ❖ Chihnan Chen, Boston University, 1137 Commonwealth Ave., Apartment 1, Allston, MA 02134, *cnchen@bu.edu*; Margarita Alegria, Cambridge Health Alliance

Large-scale surveys typically produce raw weights with very large variations. A standard approach is to trim, a way to reduce potentially huge variances of various survey estimators. The trimming method is popular mainly because of its simplicity. In this paper we investigate a more principled method by shrinking the variance of the log of the weights, which is equivalent to introducing a power parameter p in [0,1] for the weights. Our investigation suggests that p=1/2, that is, taking square root is a good general-purpose choice according to the mean-squared error criterion. This power shrinking method, besides preserving the ranking of the raw weights, provides a natural way to deal with measurement errors and outliers in the raw weights. We demonstrate the use of this method with the National Latino and Asian American Study (NLAAS) on mental health.

### Confidence Interval Coverage in Complex Model-Based Estimation with Deep Stratification

❖ Archana Joshee, Ernst & Young LLP, 1225 Connecticut Ave., NW, Washington, DC 20036, *archana.joshee@ey.com*; Wendy Rotz, Ernst & Young LLP

***Key Words:*** deep stratification, model based estimation, degrees of freedom

This paper studies confidence interval coverage in model based estimation under complex designs with two layers of stratification: a major stratum designation and deep stratification within each major stratum. The paper explores overall degrees of freedom (df) and confidence interval coverage in different settings. It is known in design-based stratified estimation, under specific assumptions, the Satterthwaite approximation estimates the df to lie between the df of the smallest stratum and the sum of the strata df. This approximation is not robust in presence of positive kurtosis. Yet little research has been done to explore df and the robustness of confidence interval coverage in complex model based estimation. This paper explores the df and robustness of confidence interval coverage in complex sample designs with deep stratification within major strata.

## 535 Bayesian and Empirical Bayesian Models

Section on Bayesian Statistical Science

**Thursday, August 10, 10:30 am–12:20 pm**

### The Analysis of Fluctuating Asymmetry as a Measure of Developmental Instability

❖ Stefan Van Dongen, University of Antwerp, Groenenborgerlaan 171, Antwerp, B-2020 Belgium, *stefan.vandongen@ua.ac.be*

***Key Words:*** Bayesian, asymmetry, developmental instability, hierarchical model, prior

Any phenotype deviates from its expected value, resulting in asymmetric development. Any deviation from perfect symmetry reflects the outcome of developmental noise and/or stability (i.e., of developmental instability). Fluctuating asymmetry reflects variation in development such that, at the individual level, we attempt to estimate a variance with two data points with high sampling variation. Theoretic models of morphological development predict either log-normal or gamma distributions, while traditional analysis tools have assumed normality. We present the analysis of simulated datasets to explore the robustness of results under the different alternative distributions. Results from normal and gamma models appeared comparable, while data generated assuming log-normality of developmental noise yielded biased results if analyzed assuming a normal or gamma distribution.

### Empirical Bayes Distribution Rule for Classification Using Unsupervised Learning

❖ Shui-Ching Chang, The Oversea Chinese Institute of Technology, Department of Business Administration, Taichung, 407 Taiwan, *monet@ocit.edu.tw*; Tze-Fen Li, National Chung Hsing University

***Key Words:*** Bayes decision rule, classification, empirical Bayes, quality control, stochastic approximation, unsupervised learning

A set of unlabelled items is used to establish a decision rule to classify defective items. We suppose the lifetime of an item has a Weibull distribution. The Bayes decision rule with estimated parameters is an empirical Bayes (EB) decision rule. A stochastic approximation procedure using the set of unidentified samples is established to estimate these unknown parameters. When the size of unlabelled items increases, the estimates computed by the procedure converge to the real parameters and gradually adapt our EB decision rule to be a better classifier until it becomes the Bayes decision rule. The results of a Monte Carlo simulation study are presented to demonstrate the convergence of the correct classification rates made by the EB decision rule to the highest correct classification rates given by the Bayes decision rule.

### Bayesian Parameter Estimation of Stable Distributions

❖ Ece Oral, Hacettepe University, Institute for Graduate Studies in Science and Engineering, Department of Statistics, 06532, Ankara, Turkey, *ece.oral@tcmb.gov.tr*; Cenap Erdemir, Hacettepe University

***Key Words:*** stable distribution, characteristic exponent, distribution functions, parameter estimation, Metropolis-Hastings algorithm, posterior distribution

Stable distributions are widely used in modeling heavy tailed data. There are many estimation methods for the parameters of stable distributions. In this study, the Bayesian approaches using Metropolis random walk chain (Tsionas, 2000) are employed and the posterior distributions of the parameters are produced. The likelihood function required for Bayesian estimation is computed by the program described by Nolan (1997). The performance of the method is investigated by simulation and a numerical example illustrated on real life data.

### Bayesian Analysis and Model Selection in Closed-Population, Capture-Recapture Models

❖ Ross Gosky, Bucknell University, 361 Olin Science Center, Lewisburg, PA 17837, *rgosky@bucknell.edu*

***Key Words:*** Bayesian inference, model selection, capture-recapture, AIC, DIC

Capture-Recapture models are used to estimate the unknown sizes of animal populations. For closed populations, those with constant size during the study, eight standard models exist for estimating population size. These models allow for variation in animal capture probabilities due to time effects, heterogeneity among animals, and behavioral effects after the first capture. Using Bayesian statistical modeling, we present versions of these eight models. We explore the use of Akaike's Information Criterion (AIC), and the Deviance Information Criterion (DIC) as tools for selecting the appropriate model for a given dataset. Through simulation, we show that AIC performs well in selecting the true model, and whether a particular model is robust when fit to data with different effects. Finally, we illustrate our methods by analyzing a population of cottontail rabbits.

### Estimation of Parameter Covariance for a Penalized Likelihood Approach to Estimation of Kinetic Models in PET (Positron Emission Tomography)

❖ Michelle Byrtek, Western Washington University, Department of Mathematics, Western Washington University, Bellingham, WA 98225-9063, *michelle.byrtek@wwu.edu*; Finbarr O'Sullivan, University College Cork

***Key Words:*** penalized likelihood, components of variance, mean square error, Bayes, kinetic parameters

Extraction of kinetic parameters from PET image data involves the use of non-linear weighted least squares regression. Current estimation techniques often have poor mean square error properties. Penalized likelihood techniques have been shown to improve parameter estimation accuracy in this context. The effectiveness of penalized likelihood, however, depends on the appropriate specification of a penalty term. A Bayesian formulation for the penalty, based on a variance component model using a rudimentarily estimated subject-to-subject parameter covariance matrix, was shown to improve parameter loss by as much as 19%. In this study, well-established methods from mixed effects modeling are used to obtain a more robust estimate of subject-to-subject variability. The error in estimation of this covariance matrix, and corresponding error in parameter estimation is also evaluated.

### Empirical Bayes Estimation for Additive Hazards Regression Models

❖ M. Brent McHenry, Bristol-Myers Squibb Company, 5 Research Parkway, GBS F209A, Wallingford, CT 06492, *brent.mchenry@bms.com*; Debajyoti Sinha, Medical University of South Carolina; Stuart Lipsitz, Brigham and Women's Hospital; Malay Ghosh, University of Florida

***Key Words:*** gamma process, marginal likelihood, posterior process

We develop a novel empirical Bayesian framework for the semiparametric additive hazards regression model of Aalen (1980) by using a gamma-process prior on the unknown baseline cumulative hazard. The marginal likelihood obtained via integrating the prior process can be maximized using standard statistical software and the empirical Bayes estimates of regression parameters, survival curves and their corresponding standard errors have easy to compute closed form expressions. This marginal likelihood based methodology, an alternative to currently available methods based on ordinary least squares and method of moments, has monotone estimate of survival function and can accommodate time-varying regression coefficients and covariates. We also present the asymptotic properties of the semiparametric empirical Bayes estimates. We illustrate our methodology using existing statistical software (SAS).

## 536 Data Collection and Missing Data Issues in Health Surveys

Section on Health Policy Statistics

**Thursday, August 10, 10:30 am–12:20 pm**

### A New Method To Evaluate the Completeness of Case Ascertainment by a Cancer Registry

❖ Barnali Das, National Cancer Institute, 6116 Executive Blvd., Suite 504, Bethesda, MD 20892-8317, *dasba@mail.nih.gov*; Limin X. Clegg, National Cancer Institute; Eric J. Feuer, National Cancer Institute; Linda Pickle, National Cancer Institute

***Key Words:*** cancer data, registry operations

Information about newly diagnosed cancer cases is collected by cancer registries across North America, certified by the North American Association of Central Cancer Registries (NAACCR). The index of completeness of incident cancer case ascertainment is a key measure of registry quality. NAACCR currently computes this index using only mortality data and does not account for the variability of this index during certification. Usual confidence intervals overlap preset certification intervals making inclusion of variability confusing in the certification process. We propose a new method for calculating the index based on a model to predict expected incidence using demographic, lifestyle and mortality data. We adjust the new index for reporting delay and data corrections. We calculate the variance of this index and propose a simple method of incorporating this into the certification process.

### The Early Release Program of the U.S. National Health Interview Survey

❖ Jeannine Schiller, National Center for Health Statistics, 3311 Toledo Road, Room 2334, Hyattsville, MD 20782, *jdv2@cdc.gov*; Jane Gentleman, National Center for Health Statistics; Eve Powell-Griner, National Center for Health Statistics

***Key Words:*** health surveys, sentinel surveillance, statistical computing, data reporting, health status indicators, health insurance

The National Center for Health Statistic's (NCHS) National Health Interview Survey (NHIS) has been releasing quarterly estimates for 15 key health indicators through its Early Release (ER) Program since data year 2000. Estimates are released on the NCHS Web site 6 months after data collection. For each indicator, a graph of the trend since 1997 is presented, followed by figures and tables showing age-specific, sex-specific, and race/ethnicity-specific estimates for the new data quarter. Key findings are highlighted. A separate in-depth report on health insurance is also released as part of the ER Program. This presentation will describe how the excellent timeliness of the ER Program has been achieved using expedited data editing, streamlined data analysis, automated report updating, and internet dissemination. Examples from the most recent ER will be used to demonstrate the ER process.

### Providing Health Information to the Public: Analysis of National Audience Data for Public Health Communication Planning

❖ William Pollard, Centers for Disease Control and Prevention, 2554 Circlewood Road, NE, Atlanta, GA 30345, *bdp4@cdc.gov*

***Key Words:*** health surveys, audience segmentation methods, health communication

A number of new health challenges have emerged at the beginning of 21st century including: new infectious diseases with global impact, bioterrorism, environmental threats, an aging population, and serious lifestyle-related illnesses. Providing information on these issues to the general public has become an increasingly important function of the Centers for Disease Control and Prevention (CDC). This presentation focuses on analysis of factors related to public awareness and use of the CDC as a source of information based on the Porter Novelli Styles Surveys with national samples of over 4,000 adults annually 1999 - 2005. The surveys address health topics and other topics such as media habits, lifestyle activities, and internet use. Techniques for identifying and characterizing different audience segments are described and implications for health communication are discussed.

### Measuring Menses: Issues with Patient-Reported Menstrual Cycle Assessment

❖ Stephanie Land, University of Pittsburgh, 201 N. Craig Street, Suite 350, Pittsburgh, PA 15213, *land@nsabp.pitt.edu*; Gina Sucato, Children's Hospital; Marcie Ritter, University of Pittsburgh; Reena Cecchini, University of Pittsburgh; Sandra Swain, National Cancer Institute; Patricia Ganz, University of California, Los Angeles

***Key Words:*** patient-reported outcomes, clinical trials, questionnaire, quality of life

The assessment of the timing and duration of menstrual cycles arises in several areas of medical research, but an ideal approach to collecting this information from patients has yet to be developed. In this presentation, I will discuss measurement issues that have arisen in two studies. In the first, a multi-center breast cancer clinical trial, menstrual history was collected with questionnaires in order to assess the effect of chemotherapy on menopausal status, and the effect of menopausal status on survival. In the second, an observational study regarding adolescent birth control use, menses were recorded on patient calendars. Both designs resulted in intermittent missingness and uncertainty in the measures of interest (such as time to resumption of menses). I will describe alternative approaches presented in medical literature, and make recommendations for future studies.

### Childhood Obesity: What's the Mother's Marital Status Got to Do with It?

❖ Soma Roy, The Ohio State University, 1958 Neil Ave., 404 Cockins Hall, Columbus, OH 43210, *somaroy@stat.ohio-state.edu*; Elizabeth Stasny, The Ohio State University

***Key Words:*** longitudinal study, missing data, NLSY79

In this paper, we study the effect a mother's marital status might have on the weight of her children. We use the National Longitudinal Survey of Youth 1979.Early exploratory data analysis shows that there are 6,283 women in the study; we have complete marital status records on only 3,215 of them for the years 1979 through 2002.There are 3,068 women who have marital status records missing at some point of time. We first explore the data to investigate the nature of this missingness. Next, we build models to use the longitudinal data available, to explain the relationship between mother's marital status and obesity in children. It is possible that missingness in the mother's marital status is related to an unstable relationship and a stressful home environment, resulting in an improper diet for the child. We also investigate other variables such as child's self-esteem, to see how weight is affected.

### When Standard Regression Techniques Fail: the Role of Propensity Score Methods

❖ Michael Posner, Villanova University, Department of Math Sciences, 800 E Lancaster St, Villanova, PA 19085, *michael.posner@villanova.edu*; Arlene Ash, Boston University

***Key Words:*** propensity scores, selection bias, model misspecification, causal inference

When can propensity score methods solve problems that standard covariate adjustment cannot? While some view propensity score methods primarily as a tool for dimension reduction of a complex covariate space (with all important covariates measured), the problem of biased estimates of treatment effect can arise with a single covariate. We examine a simple model, conceptually and via simulations, to show that standard regression can appropriately handle differing covariate distributions between treatment groups, but when model misspecification is also present, estimates from standard regression are biased. We compare analyses using crude estimates (no adjustment), regression, and propensity score methods. When covariate distributions differ by treatment group and models are misspecified, only the propensity score method provides an unbiased estimate of treatment effect.

## 537 Small-Area Estimation ●

Section on Survey Research Methods
**Thursday, August 10, 10:30 am–12:20 pm**

### Accounting for Interviewer Variability in Small-Area Estimation

❖ Benmei Liu, Westat, 1650 Research Blvd., RE 473, Rockville, MD 20850, *BenmeiLiu@westat.com*; Partha Lahiri, University of Maryland

***Key Words:*** EBLUP, mixed linear model, hierarchical Bayes, mean squared error

Variation in interviewers' conduct could impact the data and become a major source of measurement error in interviewer administered surveys. This important source of variability has been largely overlooked in the small area literature. In this paper, we develop alternative small area models that incorporate interviewer variability. We then discuss both empirical Bayes and hierarchical Bayes implementations of these models. A Monte Carlo simulation study is undertaken to assess the effect of interviewer variability on small area estimates and their corresponding measures of uncertainty.

### Using Administrative Records with Model-Assisted Estimation for the American Community Survey

❖ Robert Fay, U.S. Census Bureau, Office of the Director, Stop 9001, Washington, DC 20233-9001, *robert.e.fay.iii@census.gov*

***Key Words:*** ACS, survey estimation, Greg, calibration estimators, generalized regression estimation

Full implementation of the American Community Survey (ACS) began in 2005. Among other purposes, the ACS will replace the decennial census long-form data, enabling a short-form census in 2010. A test implementation of the ACS in 36 test counties during 1999-2001 suggested that the initial ACS estimation procedure, while adequate at the county level, yields higher variances at the small-area levels of tract and block group relative to the decennial long form. Previously reported research argued the likely success of an approach combining administrative record data with a generalized regression estimator. The case for likely success was based on the R-square of the underlying regression. This paper reports detailed results on a full implementation in 34 test counties, showing the degree to which the predictions based on regression diagnostics have been confirmed.

### Small-Area Prevalence Estimates Using two Surveys

❖ William Davis, National Cancer Institute, 6116 Executive Blvd., Suite 504 MSC 8317, Bethesda, MD 20892, *davisbi@mail.nih.gov*; Charuta Soman, IMS Health; Zhaohui Zou, Information Management Services, Inc.

***Key Words:*** MCEM, model-based, BRFSS, NHIS, simulation

We develop a model based approach to estimate small area cancer risk factor prevalence using information from two surveys. We demonstrate the ability to estimate parameters generated from the assumed model using a Monte-Carlo EM algorithm. The method is investigated using simulation of datasets that are similar to those that would be obtained from the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS) surveys. For the simulated data, we assess the accuracy of the parameter estimates and also the accuracy of the county, state, and national estimates.

### Interval Estimation in Small-Area Problems

❖ Snigdhansu Chatterjee, University of Minnesota, School of Statistics, 313 Ford Hall 224 Church St SE, Minneapolis, MN 55455, *snigdhansu@rediffmail.com*; Partha Lahiri, University of Maryland; Huilin Li, University of Maryland

***Key Words:*** small area, interval estimation, mixed effects, parametric bootstrap, coverage accuracy

In typical small area problems, variability is most often characterized by mean squared error. The naive interval estimate constructed using the mean squared error has several shortcomings, and many attempts have been made to improve its coverage accuracy. We propose a different interval estimation technique that does not involve the mean squared error, and has high coverage accuracy.

### Nonparametric M-Quantile Small-Area Estimation via Penalized Splines

❖ Monica Pratesi, Universit‡ di Pisa, Via Ridolfi 10, Pisa, 56124 Italy, *m.pratesi@ec.unipi.it*; Nicola Salvati, University of Pisa; Maria G. Ranalli, University of Perugia

***Key Words:*** small area estimation, m-quantile regression, robust estimation, penalized least squares

In the paper we extend the M-quantile small area estimation approach (Chambers and Tzavidis, 2006) to the setting in which the functional form of the relationship between the variable of interest and the covariates is left unspecified, introducing p-splines in the general M-estimator. The estimation algorithm is based on iteratively reweighted penalized least-squares; generalized cross validation is used for model selection of the smoothing parameter (Ruppert et al., 2003). The method is tested on simulated datasets in which the nonparametrically specified covariates are both univariate and bivariate. The method is then applied to the estimation of the mean Acid Neutralizing Capacity for each 6-digit Hydrologic Unit Codes in the Northeastern states of the U.S., where a survey of 334 lakes in a population of 21,026 has been conducted between the years 1991 and 1996 (Opsomer et al., 2005).

### Using Regression To Combine Information from Multiple Surveys for Small-Domain Estimation

❖ Takis Merkouris, Statistics Canada, R.H. Coats Building, 16th Floor, Tunneys Pasture, Ottawa, ON K1A 0T6 Canada, *takis.merkouris@statcan.ca*

***Key Words:*** small area, rare characteristics, generalized regression estimator, calibration, composite estimator

The possibility of enhancing the efficiency of domain estimators by combining comparable information collected in multiple surveys of the same population has been pointed out in recent literature, but it has not been explored to date. We propose a regression method of estimation that is essentially an extended calibration procedure whereby comparable domain estimates from the various surveys are calibrated to each other. We show through analytic results and an empirical study that this method may greatly improve the efficiency of domain estimators for the variables that are common to these surveys, as these estimators make effective use of increased sample size for the common survey items. The proposed approach is equally suitable for small geographic and non-geographic domains. It is also highly effective in handling the closely related problem of estimation for rare characteristics.

### Using the t-Distribution in Small-Area Estimation: an Application to SAIPE State Poverty Models

❖ Elizabeth Huang, U.S. Census Bureau, 7005 Petunia Street, Springfield, VA 22152, *elizabeth.t.huang@census.gov*; William R. Bell, U.S. Census Bureau

***Key Words:*** Fay-Herriot model, robust models, current population survey, American Community Survey

The Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces state age-group poverty estimates from Bayesian treatment of a Fay-Herriot model applied to direct state estimates from the Current Population Survey (CPS). In recent years supplementary surveys for the American Community Survey (ACS) have also produced state poverty estimates. Huang and Bell (2004) compared posterior variances from the bivariate and univariate models for the CPS and ACS data. While we found some improvements with bivariate models under certain assumptions, we also found occasional large posterior variance increases. These corresponded to outliers or near outliers in the ACS equation. As a means of dealing with this "problem" we examine letting either the model errors or sampling errors in either the ACS or CPS equation follow a t-distribution.

## 538 Model Diagnostics ● ✪

Section on Statistics in Epidemiology

**Thursday, August 10, 10:30 am–12:20 pm**

### Semiparametric Estimation of the ROC Surface

❖ Zheng Zhang, University of Washington, 4311 11th Ave., NE, 300, Seattle, WA 98105, *zhangz@u.washington.edu*

***Key Words:*** ROC surface, ROC curve, diagnostic tests, semiparametric, infectious disease

ROC surface is a natural extension of ROC curve in dealing with ordered multi-class diagnostic tests and other similar type of applications. For large numbers of infectious disease and many chronic diseases, their progression are best measured by stages, hence multi-class diagnostic tests that can be used to assess those diseases would be advantageous over their dichotomous counterpart. An ROC surface can be constructed in those situations when the test results are continuous. Non-parametric estimation of the ROC surface has been proposed in the recent years. Here we propose a semi-parametric definition of the ROC surface and the corresponding estimating procedure. Simulation studies and real data analysis are also included.

## Model Diagnostics for Generalized Linear Mixed Models Using Chi-Square Test

❖ Zhonghua Gu, University of California, Davis, 2900 Solano Park Circle, Apt2723, Davis, CA 95616, *zhgu@ucdavis.edu*

***Key Words:*** longitudinal data, goodness of fit, GLMM, model checking

A simple goodness-of-fit test is developed for checking the distributional assumptions in a generalized linear mixed model. We focus on problems of longitudinal data analysis, where GLMMs are extensively used. Longitudinal data are often encountered in many research fields, while there is no straightforward diagnostic method for checking the distributional assumptions under a GLMM in situations such as the Binary or Poisson observations. A non-standard Chi-square test is developed and proved to be asymptotically correct under regularity conditions. We use simulations to study the empirical performance of the proposed test. In particular, we use numerical integration to simplify the parameter estimation and test statistic calculation, and Monte Carlo method to approximate the critical value of the test. Finally, the method is applied to the analysis of a real-life longitudinal data set.

## Single-Sample Predictive Model Validation via Variance Components Estimated through Resampling and Cross-Validation

❖ Michael Jones, University of Sydney, Room 128A, Edward Ford Building, Sydney, 2006 Australia, *mikej@health.usyd.edu.au*; Petra Macaskill, University of Sydney

***Key Words:*** predictive model validation, resampling, cross-validation, variance components

Background: Predictive models developed on a training sample are prone to over-optimism. An independent validation sample is not always feasible. Aim: To develop a measure of model stability based on a training sample alone. Method: Ratio of within- and between-subject variance components (sw/sb) from cross-validated predicted probabilities across bootstrapped binary logistic models is derived as an index of model stability. In a simulation study sw/sb is compared with differences between independent training and validation samples (?). Results: Validation sample sw/sb correlates clearly but imperfectly with several measures of ? (=0.5) based on AIC and concordance between observed and predicted classification. Conclusion: When independent sample validation is not feasible, useful insight into model stability can be gained from variance components derived from the training sample alone

## Investigating the Categories for Cholesterol and Blood Pressure for Risk Assessment of Death Due to Coronary Heart Disease

❖ Billy Franks, Jr., Astellas Pharma Inc., 3 Parkway North, Biostatistics, Deerfield, IL 60015, *billy.franks@us.astellas.com*; Daniel McGee, Sr., Florida State University

***Key Words:*** CART, relative risk trees, kernel density estimation, coronary heart disease

Many characteristics for predicting death due to coronary heart disease are measured on a continuous scale. These characteristics, however, are often categorized for clinical use. We suggest a systematic approach to determine the best categorizations of systolic blood pressure and cholesterol level for use in identifying individuals who are at high risk for death due to coronary heart disease. We also compare these data derived categories to those in common usage. A version of Classification And Regression Trees (CART) that can be applied to censored survival data will be used to identify categories in multiple data sets. The collection of categories will then be used to identify major cutpoints common in all of the data sets by using kernel density estimation.

## Survival and Classification Tree Analyses in Medicine: Usefulness in Identifying the Predisposing and Precipitating Health Conditions and Events Associated with Death

❖ Gail McAvay, Yale University, Department of Internal Medicine, 333 Cedar Street, School of Medicine, New Haven, CT 06510, *Gail.McAvay@yale.edu*; Mary Tinetti, Yale University School of Medicine; Thomas Gill, Yale University; Heather G. Allore, Yale University

***Key Words:*** classification trees, survival models, mortality rates

Epidemiological studies of mortality often use survival methods to identify contributing conditions. While these methods are helpful in identifying the main effects of individual conditions, interactions between conditions are rarely addressed or are limited to the interaction between two conditions. Classification trees provide an approach to identify contributing conditions for mortality that can reveal important combinations of conditions resulting in homogeneous groupings with respect to mortality and cut-points for continuous risk factors. We present results from these methods to identify the health conditions and events associated with death in a cohort of 754 community dwelling older adults. A multivariable Cox regression model was estimated to determine the conditions associated with death and a classification tree analysis divided these conditions into homogeneous subgroups.

## Adjusting ROC Curves Estimation for Covariates without Gold Standard Tests

❖ Chong Wang, Cornell University, 301 Malott Hall, Department of Statistical Science, Ithaca, NY 14853, *cw245@cornell.edu*; Bruce W. Turnbull, Cornell University; Yrjö Gröhn, Cornell University; Søren Nielsen, The Royal Veterinary and Agricultural University

***Key Words:*** Bayesian, Johne's disease, milk yield, ROC curves

We propose a method for estimating ROC curves based on statistical models to adjust for covariate effects when the true disease states of test animals are unknown. Covariates may be correlated with the disease process, diagnostic testing procedure, or both. We propose a two-part Bayesian model. First, a logistic regression model for disease prevalence is used to fit the covariates. Second, a linear model is used to fit the covariates to the distribution of test scores. We use Markov chain Monte Carlo methods to compute the posterior estimates of the sensitivities and specificities that provide the groundwork for inference concerning the diagnostic procedure's accuracy. We apply the methodology to ELISA scores from several dairy cow herds for the diagnostic testing of paratuberculosis.

### Combination of Multiple Tests by Likelihood and Bayesian Approaches

❖ Carol Lin, Bristol-Myers Squibb Company, 311 Pennington-Rocky Hill Road, Mailstop 19203, Pennington, NJ 08534, *carol.lin@bms.com*; Lance Waller, Emory University

***Key Words:*** Bayesian, likelihood, combined decision, system performance

A combination of various tests or sensors (e.g., a set of biomarkers or combination of biomarkers and screening tests or a group of air pollution monitors) often is used to improve diagnostic performance to better detect an underlying phenomenon. Many decision criteria are based on setting a decision threshold for the likelihood ratio. However, the probability of underlying phenomenon (prevalence) which is required for the likelihood approach, often is not available. A new Bayesian approach developed by Chen and Chair (2002) combines individual tests when no information is available to assess the initial probability of having the underlying phenomenon (the prior). We review the likelihood and Bayesian approaches for combining tests and present results from a simulation study comparing the system performance of each approach when the prior probability is not available.

## 539 Climate, Weather, and Health: Novel Applications of Spatial and Temporal Analyses ●

Section on Statistics and the Environment
**Thursday, August 10, 10:30 am–12:20 pm**

### Predicting Spatial Exceedance Regions

❖ Jian Zhang, The Ohio State University, 1958 Neil Ave., Cockins Hall, Room 304E, Columbus, OH 43210, *zhang@stat.ohio-state.edu*; Noel Cressie, The Ohio State University; Peter F. Craigmile, The Ohio State University

***Key Words:*** spatial exceedance regions, Baddeley's loss, simulated annealing

A common problem in spatial statistics is to predict exceedances and the regions of the exceedances. This has application in environmental sciences, natural resources, and agriculture, since rare events tend to have a strong impact on environment. We develop a method involving a loss function based on Baddeley's (1992) metric for binary images. The predictor based on the loss function is obtained by minimizing the posterior expected loss when the spatial trend, noise, and spatial-covariance parameters are estimated; in practice, minimization is achieved by simulated annealing. Through simulation, we calibrate the choice of a tuning parameter in the loss function, and investigate the predictor's performance. We then apply our methodology to a spatial dataset of decadal temperature changes over the Americas.

### A Geostatistical Method for Texas NexRad Data Calibration

❖ Bo Li, Texas A&M University, 1100 Hensel Drive, Apt U2F, College Station, 77840, *boli@stat.tamu.edu*; Marian Eriksson, Texas A&M University; Raghavan Srinivasan, Texas A&M University; Michael Sherman, Texas A&M University

***Key Words:*** NEXRAD data, threshold, linear regression, variogram estimation

Rainfall is one of the most important hydrologic model inputs and is recognized as a random process in time and space. Rain gauges generally provide good quality data, however they are usually too sparse to capture the spatial variability. Radar estimates provide a better spatial representation of rainfall patterns, but they are subject to substantial biases. Our calibration of radar estimates using gauge data takes season, rainfall type, and rainfall amount into account, and is accomplished via a combination of threshold estimation, bias reduction, regression techniques, and geostatistical procedures. We employ the varying-coefficient model to adapt to the temporal variability of rainfall. The methods are illustrated using Texas rainfall data in 2003, which includes WSR-88D radar-reflectivity data and the corresponding rain gauge measurements.

### Identifying Effect Modifiers in Air Pollution

❖ Sandrah P. Eckel, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe St Rm E3527, Baltimore, MD 21205, *seckel@jhsph.edu*; Thomas A. Louis, Johns Hopkins University

***Key Words:*** air pollution, time series, multi-level, CART, effect modification

Studies of the health effects of air pollution such as the National Morbidity and Mortality Air Pollution Study (NMMAPS) relate changes in daily pollution to daily deaths in a sample of cities and calendar years. Generally, location-specific estimates are combined over locations using a two-stage model. Our analysis relates single-day lagged PM10 to daily mortality in people over 65 from the 50 largest NMMAPS cities. We build on the standard approach by "fractionating" the city-specific analysis to produce month/city/year specific estimated air pollution effects (slopes). We identify potential effect modifiers via linear regression and regression trees with the estimated slopes as dependent variables and predictors such as dew point temperature, temperature, CO, NO2, O3, SO2, season, year and region. We report on our methods and findings.

### Identifying Environmental Risk Factors of Childhood Obesity

❖ Rong Qi, Indiana University, Division of Biostatistics, 1050 Wishard Blvd RG4101, Indianapolis, IN 46202-2872, *rqi@iupui.edu*; Gilbert Liu, Indiana University; Jeffrey Wilson, Indiana University Purdue University Indianapolis; Jun Ying, Indiana University

***Key Words:*** childhood obesity, environmental risk, GIS, spatial logistic model

Body mass index (BMI) and demographics were obtained for children ages 3-16 years old using the Regenstrief Electronic Medical Record System; patient-level environmental variables were developed from U.S. Census data and remote sensing image using Geographic Information Systems (GIS). The association between childhood obesity and environmental risks was investigated in a spatial logistic regression model. Our results showed that neighborhood greenness (measured by the Normalized Difference Vegetation Index) and distance to supermarkets were significant predictors of childhood obesity and these associations differed between children living in high population density region versus low population density region.

### Prediction of State Cancer Case Counts and Rates by Hierarchical Spatio-Temporal Models

❖ Linda Pickle, National Cancer Institute, 6116 Executive Blvd., Suite 504, Bethesda, MD 20892-8317, *picklel@mail.nih.gov*

***Key Words:*** spatio-temporal models, Poisson regression, hierarchical models

Although not all U.S. states provide high quality cancer data annually, estimates of the numbers of incident cancer cases by state are useful, e.g., for measuring cancer burden and monitoring trends. We present results of a hierarchical spatio-temporal Poisson regression model that predicts the annual number of incident cancer cases expected in each U.S. state, which are then projected ahead 4 years, based on age- and county-specific incidence data available from high quality registries and the mortality/sociodemographic/lifestyle profiles of all counties. The model was assessed by usual goodness of fit statistics plus comparison of predictions to state-reported data and projections from the American Cancer Society. Methodologic issues discussed include a comparison of full Bayesian vs. approximate likelihood methods (SAS GLIMMIX) and parametric vs. robust estimates of spatial covariance.

### Testing for Spatial Dependence Based on the SAR Model

❖ Hongfei Li, The Ohio State University, 1958 Neil Ave., Room 412, Columbus, OH 43210, *li.hongfei@gmail.com*; Catherine A. Calder, The Ohio State University; Noel Cressie, The Ohio State University

***Key Words:*** spatial statistics, Moran's I, profile likelihood, spatial autoregressive (SAR) model

Moran's I is a statistic used widely to test for the presence of spatial dependence in lattice data and, less formally, in exploratory analyses of spatial data. We develop an alternative statistic, APLE, an approximate profile likelihood estimator of the spatial dependence parameter in the spatial autoregressive (SAR) model, and demonstrate that it provides a better estimate of the spatial dependence parameter than Moran's I. In addition, we propose the APLE scatterplot, a visual tool for assessing the strength of spatial dependence in lattice data. Finally, we include an analysis of the well-known Mercer and Hall wheat-yield data, based on a SAR model to illustrate the difference between APLE and Moran's I.

### Space Time Analyses of Adverse Pregnancy Outcomes in Allegheny County, Pennsylvania

❖ M. Abdus Sattar, University of Pittsburgh, 226 Shady Ave., Apt 504, Pittsburgh, PA 15206, *mas196@pitt.edu*

***Key Words:*** empirical Bayesian smoothing, disparity statistics, local and global clustering, null spatial model

The main objective of this study is to understand the space-time distribution of adverse pregnancy outcome data obtained from Allegheny County, Pennsylvania. Birth data of the years 1994 to 2001 were obtained from the Allegheny County Health Department. Low Birth Weight (LBW) and Preterm Delivery (PTD) are the key indicators of the adverse pregnancy outcome. In this study, it will be revealed how these indicators have been changing in relation to time and space. Empirical Bayesian Smoothing will be used to reduce the variability in the data, essentially reducing the extreme values more towards the mean of the data set. Disparity statistics will be used to measure the significant spatial differences in the health outcomes. To evaluate the spatial patter, we will use the local and global clustering statistics. Statistical methods will be used for testing the null spatial model.

## 540 Forecasting and Database Modeling in Marketing Applications

Section on Statistics and Marketing

**Thursday, August 10, 10:30 am–12:20 pm**

### Some Challenges in Sales Forecasting in a Global Automotive Environment

❖ Lynn Truss, GM Research & Development, MC 480-106 359, 30500 Mound Rd, Warren, MI 48090-9055, *lynn.t.truss@gm.com*; Satish Sehgal, GM India Science Lab; Atul Saroop, GM India Science Lab; Balaji Jidugu, GM India Science Lab; Peiling Wu, GM Research & Development; Diane Gibbons, GM Research & Development

***Key Words:*** forecasting, automotive, global

There is a tremendous opportunity for automotive growth globally, especially in the Asia-Pacific region. There's also a recognized need for improved short-term and longer-term forecasting models for the different vehicle brands and the variants that are produced. One example of the inherent complexity in this industry relates to the supply chain - kits come from other countries and require extensive lead times. We have explored the use of various models for forecasting including time series, regression, growth models and a combination of these, and will provide a summary of our learnings to date for a particular vehicle in India.

### An Empirical Analysis of Customized and Dynamic Cross-Selling Campaigns

❖ Alan Montgomery, Carnegie Mellon University, 5000 Forbes Ave., 255A Posner Hall, Pittsburgh, PA 15213, *alanmontgomery@cmu.edu*; Baohong Sun, Carnegie Mellon University; Shibo Li, Indiana University

***Key Words:*** marketing, customer relationship management, multivariate probit, MCMC, Bayesian, customer lifetime value

The existing cross-selling literature has focused on developing methodologies to better predict purchase probabilities for the next product to be purchased. The usual goal is to find the best customers for a scheduled campaign. We formulate cross-selling campaigns as a stochastic dynamic programming problem that explicitly accounts for the company's long-term profit goal while taking into account the development of customer demand over time. The model yields optimal cross-selling strategies that are a multi-step, multi-segment and multi-channel cross-selling campaign process about when to target which consumer with what product using what campaign channel. Using cross-selling campaigns and transaction data provided by a national bank, we demonstrate the dynamic and state-dependent nature of the optimal cross-selling campaign decisions.

### A New Methodology to Modeling Prospects of Corporate Database

❖ Kangkang Xu, Experian, 955 American Lane, Schaumburg, IL 60173, *kangkang.xu@experian.com*

***Key Words:*** corporate database marketing, profile model, panel data

Profile models are built to identify good prospects for clients in huge corporate database. To modeling prospects, there is a common modeling bias in this industry because of corporate database structure. We propose a new modeling design to profile prospects. In stand of using cross sectional data regression, we structure panel data to profile new custom by modeling time series and cross sectional data. We also structure custom purchase behavior by utilizing affinity modeling for purchases that provides a better panel data for modeling. We also make universe pre-selection/clustering to reduce bias that is caused by corporate database structure.

### Estimating Factor Effects with Data from Heterogeneous Experimental Units

❖ Sam Weerahandi, Time Warner, 135 W. 50th Street, 751B, New York, NY 10020, *Weerahandi@aol.com*; Martin Koschat, Time Warner

***Key Words:*** weights in WLSE, repeated measures

We address a serious inference problem that arises in marketing experiments especially, a problem involving longitudinal data that has been overlooked by practitioners and researchers. The problem concerns the poor performance and bias of widely used methods of estimating parameters of linear models when the measurements from underlying experimental units could vary substantially from one unit to another. By means of a dataset from a marketing experiment, we demonstrate the drawbacks of conventional methods of making inferences and provide a simple approach to overcoming the problem. For the benefit of those practitioners who insist on conventional methods of modeling, we also will outline a Bayesian solution.

### Hybrid GA-Based Constrained Optimization and Search for High-Potential Product Configurations

❖ Brandon L. Paris, General Mills, Inc., One General Mills Blvd., MC BT09A, Minneapolis, MN 55426, *brandon.paris@genmills.com*; Lynd D. Bacon, Sighthound Solutions, Inc.; Jeff Hunter, General Mills, Inc.; Ashwin Sridhar, Sighthound Solutions, Inc.

***Key Words:*** genetic algorithms, fitness function, optimization, new product development

Manufacturers and service providers collect millions of product and concept evaluations every year. The data typically come from survey responses, and they undoubtedly contain valuable and unanticipated information about appealing new product ideas. Given the growing pressure on firms to innovate, it has become increasingly important to identify these ideas. We describe a system for searching for new product configurations using a hybrid genetic algorithm to perform constrained search and optimization. Our GA implementation is an extensible system that includes a configurable fitness function and a predictive modeling capability that allows it to “learn” parts of the fitness function with either parametric or nonparametric procedures. The GA is embedded in a multiuser environment that provides project management and collaboration functionality behind simple user interfaces.

### Falling into the Data Mine: Nontechnical Errors for Database Model Builders

❖ Sam Koslowsky, Harte Hanks CRM Analytics, 55 Fifth Ave., New York, NY 10003, *sam_koslowsky@harte-hanks.com*

***Key Words:*** CRM, data mining, data base marketing, marketing response models

Most managers agree that data mining and modeling play an important role in assuring success of a marketing program. However, errors can creep into the data mining process---problems that both novice managers and experienced analysts can confront. And these errors are more often nontechnical in nature. This paper will cover the author’s many years of experience as both a marketer and statistician as they relate to model building and the associated nontechnical problems that typically arise, specifically in the database marketing arena. We will discuss what these error-prone situations are and simple solutions to confront them. Areas prone to error include defining objectives correctly, using appropriate predictors, quality control procedures, and many other seemingly mundane activities. The problems discussed will be framed through actual marketing problems.

### Churn Modeling in a Noncontractual Setting

❖ Enping Mai, Syracuse University, 311 Whitman School of Management, Syracuse University, Syracuse, NY 13244, *emai@syr.edu*; Raja Velu, Syracuse University

***Key Words:*** churn, profitable customers, non-contractual setting

With customers churning at an average monthly rate of 2% across the United States, many firms are losing substantial amounts of money. However, not all customers are profitable. Therefore, identifying when a profitable customer is churning is very important to firms and database analysis is a key strategic tool to accomplish that goal. In this paper, we review various models of churn and develop a predictive model of the churn time of profitable customers in a non-contractual setting that incorporates customer heterogeneity. Transactional data along with customer demographics are used to demonstrate the approach.

## 541 Repeated Measures and Longitudinal Data ●

Biopharmaceutical Section, Biometrics Section, ENAR

**Thursday, August 10, 10:30 am–12:20 pm**

### A Multivariate Truncated Model Combined with Multiple Imputation for Longitudinal Data with Nonignorable Missing

❖ Rong Liu, Merck & Co., Inc., 770 Sumneytown Pike, West Point, PA 19468, *rong_liu@merck.com*; Viswanathan Ramakrishnan, Virginia Commonwealth University

***Key Words:*** multivariate truncated normal, multiple imputation, treatment related dropout

In longitudinal clinical trials, the data sets often are incomplete. In some cases, patients drop out due to treatment-related reasons, which lead the distribution of the observed data to resemble a truncated normal distribution. Ramakrishnan and Wang (2005) proposed a method under a truncated multivariate normal distribution to analyze such data. Although majority of multiple imputation (MI) procedures involve the use of ignorable missing models, MI could also be used for nonignorable missing data. A MI procedure under the MDT method to accommodate the uncertainty in imputation is proposed. The combination of MDT method with Rubin’s multiple imputation will be presented. A data set will be used to illustrate the application of the presented method.

### Analysis of Longitudinal Trials with non-MCAR Dropouts and Potentially Non-Normal Data: Is Weighted GEE the Solution?

❖ Robin Mogg, Merck Research Laboratories, 785 Jolly Road, Bldg C UNA102, Blue Bell, PA 19422, *robin_mogg@merck.com*; Devan V. Mehrotra, Merck Research Laboratories

***Key Words:*** missing data, weighted estimating equations, multiple imputation

In a typical longitudinal comparative clinical trial, some subjects drop out because of "treatment failure", some are lost to follow-up, etc. The incomplete longitudinal data are often analyzed using REML, the SAS PROC MIXED default, which assumes multivariate normality of the responses. If the normality assumption is untenable and the missing data are "missing at random", the weighted generalized estimating equations (WGEE) method of Robins et al. (1995) can be used as an alternative to REML. In this talk, we will use simulations to compare WGEE with both REML and a new method that we have developed, multiple imputation of the missing values followed by application of the Wei-Lachin (1984) approach with Wilcoxon scores (MI -> WL). The simulations reveal that, with respect to type I error rate control and power, the MI -> WL method is the best, while WGEE is generally the worst.

## Analysis of Repeated Measures Random Length Data

❖ Ana-Maria Iosif, University of Pittsburgh, 2717 Cathedral of Learning, Pittsburgh, PA 15260, *ani2@pitt.edu*; Allan R. Sampson, University of Pittsburgh

***Key Words:*** random length data, informative length, longitudinal

Random length data occurs in experiments in which data on both the frequency of an event and its severity level are gathered and both are important. We model such data when subjects are measured repeatedly over time. In order to evaluate a treatment, one needs to jointly model the number of events and their correlated severity measures, as well as their relationship over time. For example, in a clinical trial of a new migraine drug, both the number of migraine headaches and their level of pain are recorded for each subject. If the drug is efficacious, the subjects in that treatment group are expected to improve: over time they will have a smaller number of headaches and lower pain levels. We model the vectors of severities with random lengths over time and provide methods to estimate the treatment effect.

## Jointly Modeling the Dropout Process and Repeated Categorical Outcome

❖ Yuting Zhang, MedFocus, 923 Shell Blvd., 203, Foster City, CA 94404, *ytkzhang@gmail.com*; Brent J. Shelton, University of Kentucky

***Key Words:*** missing data, EM algorithm, repeated categorical outcome

For clinical trials with multiple endpoints, analyzing categorical response data measured over time is a big challenge due to possible survival event or lost to follow-up. We proposed a pattern-mixture approach to jointly model the dropout process and the repeated categorical response. Given the distribution of dropout process, the missing data can be treated as missing at random for the categorical outcomes. The estimates are obtained by the method of the weighted Expectation-Maximization (EM) algorithm. The proposed approach is illustrated by an example. The robustness and validity of this approach is investigated by sensitivity analyses

## Inference for Skew-Normal Data Using Stochastic Hellinger Distance Method and Stochastic MLE

❖ Serena Chan, Cornell University, 202 Williams Street, Apt 5, Ithaca, NY 14850, *ssc35@cornell.edu*; Anand Vidyashankar, Cornell University

***Key Words:*** skew normal, confidence coverage, minimum hellinger distance estimators, stochastic search

Let (xn, n>=1) be a random sample from a skew-normal population with parameters lambda1, lambda2, and lambda. The parameter lambda is related to the skewness parameter and called the index of skewness. We are developing two new methodologies, stochastic minimum Hellinger distance estimation (MHDE) and stochastic MLE (SMLE), to perform inferences for (lambda1, lambda2, lambda). We study the coverage properties and establish the asymptotic validity of the proposed methodologies. We also compare our proposed methodologies with the conventional MLE and semiparametric estimation method. We find that MHDE and SMLE perform well for modestly large values of the skewness parameter. Extensions to random effects models and other models related to longitudinal data are under investigation.

## Identifying Worsening MS Patients Using a Generalized Mahalanobis Distance Metric

❖ Daniel Bonzo, Serono, Inc., One Technology Place, Rockland, MA 02370, *daniel.bonzo@serono.com*

***Key Words:*** RRMS, worsening MS, longitudinal clustering, generalized Mahalanobis distance, GMDM

Recent interest in the treatment of relapsing remitting multiple sclerosis (RRMS) has focused on the need to identify subjects who are breaking through in their current treatment. It is argued that the soonest that a subject's MS is identified as progressing and treated the better will be the subject's long-run prognosis. Of import is the identification of key markers that would identify a subject as having a worsening MS condition. In this paper we propose a longitudinal clustering approach to segment a RRMS patient database. The clustering approach will use a generalized Mahalanobis distance derived from a general mixed data model (GMDM). Patients identified as having a worsening MS will be characterized using efficacy and safety parameters isolated by the technique as pivotal in the resulting segmentation.

## Distributions of Patterns in Markov Chains with Stopping Rules

❖ Francis Hsuan, Temple University, Philadelphia, PA 19122, *francish@temple.edu*; Milton Parnes, Temple University; Thomas E. Bradstreet, Merck Research Laboratories

***Key Words:*** Mathematica, embedded Markov chain, potential function

Markov-chain potential theory provides a simple and unified way to find the multivariate distributions of patterns in repeated binary data with a stopping rule. One example is the T-maze swim test in behavioral teratology experiments. Here, each rat swims up the entry channel (the base of the T) and escapes successfully only if it turns in the direction towards an a-priori chosen end of the crossbar. Each rat in the experiment is tested repeatedly until three consecutive successes are observed. The experiment results in repeated binary data, which ends in three consecutive 1's. One question is to find the joint multivariate distribution of sufficient statistics N00 (the number of two consecutive 0's in a string), N01, N10 and N11. We use potential theory to derive explicitly the joint probability generating function of minimal sufficient statistics.

# 542 Principal Components, Predication, and Optimization ●

Business and Economics Statistics Section
**Thursday, August 10, 10:30 am–12:20 pm**

## Likelihood-Based Estimate for the Number of Principal Components

❖ Lifang Hsu, Le Moyne College, 1419 Salt Springs Road, Dept of Mathematics and Computer Science, Syracuse, NY 13214, *hsu@lemoyne.edu*; Pinyuen Chen, Syracuse University

***Key Words:*** principal component analysis, ranking and selection, likelihood-based estimate

Traditional principal components analysis uses ad-hoc criteria to determine the number of components retained for further study and interpretation. The "proportion of the total variance explained by the first several principal components" is a typical such criterion. Based on the "plug-in" likelihood-based sample estimate of the above criterion and the statistical ranking and selection methodology, we propose an upper confidence limit and a lower confidence limit, respectively, for the number of principal components to be retained. We implement our theory on two financial examples. The results illustrate that the upper and lower limits obtained by our procedure are consistent with those obtained by ad-hoc criteria and our results are always accompanied with desired confidence level.

## Long Run Canonical Correlations: Estimation, Inference, and Usefulness in Econometric Analysis of Time Series

❖ Kalidas Jana, Trinity University, Department of Economics, One Trinity Place, San Antonio, TX 78212, *kjana@trinity.edu*; Alastair R. Hall, North Carolina State University

***Key Words:*** long run canonical correlations, generalized method of moments, canonical coherences at frequency zero, structural stability testing, noncentrality parameter, exogeneity testing

Recently, a new kind of canonical correlations, called Long Run Canonical Correlations (LRCCs), has emerged in econometrics literature, though methods of their estimation and inference have not been developed. The objective of this paper is to fill that gap. In addition, this paper illustrates usefulness of LRCCs in econometric analysis of time series beyond their usefulness in moment selection for generalized method of moments models. In particular, it demonstrates how LRCCs can be used to develop tests that play important roles in structural stability testing and exogeneity testing of regressors in time-series models where the regressors are nonstationary.

## Ex Post and Ex Ante Prediction of Unobserved Multivariate Time Series: a Structural Model-Based Approach

❖ Fabio Nieto, National University of Colombia, AA 72157, Bogota, 0000 Colombia, *fhnietos@unal.edu.co*

***Key Words:*** benchmarking, disaggregation, extrapolation, interpolation, multivariate time series, signal extraction

A methodology for estimating high-frequency values of an unobserved multivariate time series from low-frequency values of and related information to it is presented in this paper. This is an optimal solution, in the multivariate setting, to the problem of ex post prediction, disaggregation, benchmarking, or signal extraction of an unobservable stochastic process. Also, the problem of extrapolation or ex ante prediction is optimally solved and, in this context, statistical tests are developed for checking on-line extreme values of the unobserved time series and consistency of future benchmarks with the present and past observed information. The procedure is based on structural or component models, whose assumptions and specification are validated with the data alone.

## Model-Based Sequence Clustering with Application to Resource Pattern Identification in B2B Services Engagements

❖ Bonnie K. Ray, IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, *bonnier@us.ibm.com*; Jianying Hu, IBM Research

***Key Words:*** clustering, HMM, time series, workforce management

Motivated by identifying groups of consulting projects requiring similar skill sets at similar project stages, we introduce an approach for model-based sequence clustering that addresses several drawbacks of existing algorithms. The approach uses a combination of hidden Markov models for sequence estimation and dynamic time warping for hierarchical clustering, with interlocking steps of model selection, estimation, and sequence grouping. We demonstrate experimentally that the algorithm can handle sequences of varying lengths, unbalanced cluster sizes, and outliers effectively.

## Applications of Direct Search Optimization Methods

❖ Otto Schwalb, 2315 Marion Walk Drive, Smyrna, GA 30080, *trey.schwalb.iii@citigroup.com*; Timothy H. Lee,

***Key Words:*** Torczon, direct search methods, grid evaluation, Nelder Mead

For many business applications, it is commonly the case that derivative information for any objective function of interest is not available. In the absence of derivative information, one must rely on optimization algorithms which only require that function evaluations be available. These algorithms are known as direct search methods. We implement the direct search algorithm of Torczon (1989, 1991) in S-plus and explore its utility in several interesting applications. Torczon has proven the superiority of her algorithm to the widely referenced Nelder Mead algorithm. Among many interesting features of the Torczon algorithm, it lends itself well to problems where the constraints are quite complicated.

## Preliminary Micro Data Results from the Business List Comparison Project

❖ Lucia Foster, U.S. Census Bureau, WPII 211, washington, DC 20233, *Lucia.S.Foster@census.gov*; Joel Elvery, Bureau of Labor Statistics; C. J. Krizan, U.S. Census Bureau; David Talan, Bureau of Labor Statistics

***Key Words:*** business register, business list, micro data

The Bureau of Labor Statistics (BLS) and the Bureau of the Census (BOC) each hold separate business list files that are the foundation of some of the most important economic statistics in the United States. These business lists serve a number of critical functions including: creating sampling frames from which surveys and censuses are drawn, providing the data for BOC's County Business Patterns and BLS' Quarterly Census of Employment and Wages, and providing benchmarks for survey data. This paper extends the analysis of our earlier work comparing the lists at the aggregate (published) level by comparing the lists at the micro level. We compare industry and geographic classifications, number of establishments, payroll, and employment over the two lists. While there is some evidence of important differences, we find that the two lists exhibit a high degree of consistency.

### Establishment Survival Using the BLS Longitudinal Database

❖ Merissa Piazza, Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212, *piazza.merissa@bls.gov*

***Key Words:*** establishment survival, longitudinal database

In a May 2005 Monthly Labor Review article, “Survival and Longevity in the Business Employment Dynamics Data,” Amy Knaup shows that about 60% of business establishments that opened in second quarter of 1998 were still in existence 24 months later. Her article also examined other interesting aspects of firm survival, including the characteristics of opening establishments according to two-digit NAICS sector, location, employment levels, ownership status, and survival rates. It is the goal of this paper to re-examine the characteristics of establishment survival for 1998--2004 using the BLS Longitudinal Database (LDB). The LDB is a relational database of 8.4 million business establishments linked longitudinally based on microdata submitted quarterly.

## 543 Inference for Parameters of Continuous Distribution

Biometrics Section

**Thursday, August 10, 10:30 am–12:20 pm**

### Bias-Corrected Point and Interval Estimates for Lognormal Mean

❖ Abu Minhajuddin, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Room E1401, Dallas, TX 75206, *Abu.Minhajuddin@UTSouthwestern.edu*; Xian-Jin Xie, The University of Texas Southwestern Medical Center at Dallas

***Key Words:*** bias correction, confidence interval, coverage probability, lognormal, likelihood principle

Bias-corrected point estimates for the mean of the lognormal distribution is derived based on the likelihood principle. Two different bias correction approaches are discussed. The derived estimate is free of first order bias for small to moderate sample sizes and is unbiased for large samples. A confidence interval for the mean is also derived. Simulation results show the new confidence interval to have good coverage estimates.

### Preliminary Test Estimator for the Mean of Bivariate Normal

❖ Xiaoping Zhu, Novartis Pharmaceuticals Corporation, 180 Park Ave., Oncology, Florham Park, NJ 09832, *zxp72@yahoo.com*; Chien-Pai Han, The University of Texas at Arlington

***Key Words:*** preliminary test estimator, bivariate normal distribution, bias, risk, relative efficiency

This paper considers the estimation of the population mean of a bivariate normal distribution with unknown covariance matrix when prior information about the mean is available. The prior information is that the mean vector is equal to a given value, but this information is uncertain. The experimenter should then test the validity of the prior information by bivariate test or univariate tests (i.e. testing each component mean individually). Hence, preliminary test estimator (PTE) can be constructed to estimate the population mean. The biases and relative efficiencies of the PTE using the bivariate test and the PTE using the univariate test are discussed. The behavior of the bias for both tests is studied. The risk of the PTE is analyzed using relative efficiency with respect to the usual estimator. It is shown that, for certain sample sizes and significance levels, the PTE performs better.

### Inferences on Correlation Coefficients: One-Sample, Independent, and Correlated Cases

Kalimuthu Krishnamoorthy, University of Louisiana at Lafayette; ❖ Yanping Xia, University of Louisiana at Lafayette, 211 Liberty Ave., Apt. 1131, Lafayette, LA 70508, *xia@louisiana.edu*

***Key Words:*** coverage probabilities, Fisher’s z transformation, sizes, UMP test, uniformly most accurate confidence interval

This article concerns inference on the correlation coefficients of a multivariate normal distribution. Inferential procedures based on the concepts of generalized variables and generalized p-values are proposed for elements of a correlation matrix. For simple correlation coefficient, the merits of the generalized confidence limits and other approximate methods are evaluated using a numerical study. The study indicates the proposed generalized confidence limits are uniformly most accurate, even for samples as small as three. The results are extended for comparing two independent correlations, correlated correlations, and non-overlapping correlated correlations. For each problem, the properties of the generalized variable approach and other asymptotic methods are evaluated using Monte Carlo simulation. The generalized variable approach produces satisfactory results.

### A New Approach for a Linear Combination of K Multinormal Mean Vectors

❖ Shu-Hui Lin, National Taichung Institute of Technology, 7F5 390 Wen Shin South 2 Road, Taichung, 408 Taiwan, *suelin@ntit.edu.tw*; Jack C. Lee, National Chiao Tung University

***Key Words:*** coverage probability, generalized confidence region, generalized pivotal quantity, heteroscedasticity

We consider the problem of constructing a confidence region for a linear combination of K multivariate normal populations when covariance matrices are completely unknown and unequal. A new generalized pivotal quantity for constructing a generalized confidence region is derived. If only two populations are considered and b1=1, b2=-1, then our model is reduced to the multivariate Behrens-Fisher problem. The generalized confidence region is illustrated with a numerical example and the merits of the proposed method are numerically compared with those of the existing methods with respect to their expected hyper-volumes, coverage probabilities under different scenarios in simulation studies.

### Estimation of Multivariate Normal Secondary Parameters in Group Sequential Trials

❖ Kai Fun Yu, National Institute of Child Health & Human Development, Bldg 6100 Room 7B05, Bethesda, MD 20892-7510, *yukf@exchange.nih.gov*; Chengqing Wu, National Institute of Child Health & Human Development; Aiyi Liu, National Institute of Child Health & Human Development

***Key Words:*** multivariate normal, group sequential trial, secondary parameter, primary parameter, bias, mean squared error

We study the estimation problems concerning the secondary parameters in multivariate normal distribution in a group sequential trial designed for testing hypotheses about the primary parameters. Relations between the properties of the primary parameter estimators and those of the secondary parameter estimators are established. The biases and mean-squared errors of the estimators are compared.

### The Median Test for Cluster-Correlated Data

❖ Rick L. Williams, RTI International, 2816 Stone Gap Court, Raleigh, NC 27162, *rwilliams@rti.org*; Angela Pitts, RTI International

***Key Words:*** simulation, Taylor series, jackknife, power

A simple method for conducting a large sample median test is demonstrated when the data are observed in clusters such that observations within clusters are correlated and observations between clusters are uncorrelated. In such situations, the standard assumptions underlying the median test are violated. The median test can be recast as a test of proportions which can then be analyzed using methods for proportions estimated from cluster correlated data. A simulation study demonstrates the methods ability to properly maintain the Type I error rate, measures its power for various alternatives and shows that ignoring the correlation of the data greatly misrepresents the Type I error rate. When comparing Taylor series linearization and jackknife variance estimators, it is found that tests based upon the jackknife estimator perform better than those based upon the Taylor series estimator.

### Bivariate Growth Charts

❖ Ying Wei, Columbia University, 722 W. 168th Street, Room 629, New York, NY 10032, *ying.wei@columbia.edu*

***Key Words:*** growth charts, multivariate, quantile regression

Growth charts have been widely used in clinics and medical centers to monitor an individual subject's growth status in context of population values. Current growth charts consider only one measurement at a time. Health evaluation very often evolves multiple measurements, height and weight for example. More informative readings can be obtained by screening multiple measurements simultaneously (jointly). We propose new framework of bivariate growth chart that are mathematically formal but clinically sensible. The charts consist of a nested sequence of bivariate reference quantile contours that are time/age dependent and incorporate with other potential covariate effects. Estimation of such charts based upon quantile regression was also provided. Their performance was demonstrated by a Monte-Carlo simulation study, as well as the applications with real height-weight data set.

## 544 Applications of Statistical Graphics ●

Section on Statistical Graphics

**Thursday, August 10, 10:30 am–12:20 pm**

### Importing Graphics for Statistical Plots

❖ Paul Murrell, The University of Auckland, Department of Statistics, Private Bag 92019, Auckland, New Zealand, *paul@stat.auckland.ac.nz*

***Key Words:*** plots, graphics formats, R, importing graphics, XML

A typical use of statistical graphics software is to produce a plot and export it in a particular format so that it can be included as part of a report. However, for some tasks, such as adding a company logo, or providing a watermark background image, it is useful to be able to go in the opposite direction and import graphics files into statistical software. This talk will discuss issues involved with importing graphics formats, with a focus on importing vector graphics, and an example implementation for the R statistical software system.

### An Implemetation of the Grammar of Graphics in R: ggplot

❖ Hadley Wickham, Iowa State University, 126 University Village, Apt B, Ames, IA 50010, *h.wickham@gmail.com*

***Key Words:*** graphics, R, lattice

The R package ggplot is an implementation of the grammar of graphics in R. It combines the advantages of both base and lattice graphics: conditioning and shared axes are handled automatically, while maintaining the ability to build up a plot step by step from multiple data sources. It also implements a more sophisticated multidimensional conditioning system and a consistent interface to map data to aesthetic attributes. ggplot is built up of four basic building blocks: aesthetic mapping functions, grob functions, scale objects and the plot object. Aesthetic mapping functions transform data into aesthetics, grob functions turn lists of aesthetics into grobs (graphical objects), and scale objects ensure that each mapping function operates on a common domain, as well as producing plot guides. The plot object combines these blocks with data to produce the final graphic.

### Visualization Challenges in Internet Traffic Research

❖ Cheolwoo Park, University of Georgia, Department of Statistics, University of Georgia, Athens, GA 30602-1952, *cpark@stat.uga.edu*

***Key Words:*** Box-Cox transformation, heavy-tailed distribution, internet traffic

This is an overview of some recent research, and of some open problems, in the visualization of internet traffic data. One challenge comes from the sheer scale of the data, where millions (and far more if desired) of observations are frequently available. Another challenge comes from ubiquitous heavy tail distributions, which render standard ideas such as "random sampling will give a representative sample" obsolete. Some alternate sampling approaches are suggested and studied.

### What Can Match 'a Daisy a Day'?

❖ Rachel Graham, Iowa State University, 202 Snedecor Hall, Ames, IA 50011, *gdragon@iastate.edu*; Heike Hofmann, Iowa State University

***Key Words:*** bar charts, categorical data

The idea of this paper is based on the task posed in the KDD Cup 2005, where the challenge was to tag a user given query with up to five categories from a larger set of predefined categories. An interesting generalization of this problem, with many applications in other areas of data analysis, is to allow arbitrary many tags for a single query. The challenge then changes from picking a specified top "X" number of best matches to finding a data driven cut-off point. In this talk we are going to present our approach of defining the best matches based on statistics of the underlying distribution, such as modes, kurtosis and skewness.

### Design Strategies for Sampling in Graphs

❖ James Rosenberger, The Pennsylvania State University, 326 Thomas Building, Dept of Statistics, University Park, PA 16802-2111, *JLR@stat.psu.edu*; Hong Xu, The Pennsylvania State University; Steve Thompson, Simon Fraser University

***Key Words:*** networks

Sampling graphs is an important challenge not adequately addressed in the literature. Inference about graphs often confuses the sample data with the entire population. The data graph is usually a single realization, not representative of the population graph as a whole. Designs

for sampling graphs are important for efficient estimation. Thompson (2004) constructed the Active Set Adaptive Sampling (ASAS) designs and associated estimators, which can apply to populations with network structures. The main advantage is that the sample size can be predefined and fixed prior to implementing the sampling strategy. In this paper, we incorporate some practical restrictions, such as budget or time constraints that determine which units should be selected at each step and when to stop the sampling process. We want to minimize the cost and maximize the sampling procedure information.

### Advanced Sequential Sampling Methods and Their Implementation within a Portable Computing Environment

❖ Arunava Chakravartty, University of California, Riverside, 200 W. Big Springs Road, Apt 4, Riverside, CA 92507, *achak001@ucr.edu*; Daniel R. Jeske, University of California

***Key Words:*** sequential methods, GUI, statistical tools

Sequential inference methods have the potential to greatly reduce the time and costs associated with decision making processes. As a result, they are applied to a variety of different fields such as Entomology, where the average pest density is the parameter of interest, and to the study of clinical trials. The mathematics associated with designing and utilizing sequential procedures can often discourage practitioners from implementing them in the field. We have developed a software tool that facilitates an easier use of the methods, and enables a convenient platform for carrying out sequential inference. The tool hides the complex mathematics of sequential designs, and provides a user-friendly interface to implement a sequential sampling scheme, understand its important statistical properties, and implement the scheme on real data.

## 545 Semi- and Nonparametric Methods ●

Biometrics Section, Section on Nonparametric Statistics
**Thursday, August 10, 10:30 am–12:20 pm**

### Improving the Effciency of the Log-Rank Test Using Auxiliary Covariates

❖ Xiaomin Lu, North Carolina State University, 2800 Avent Ferry Road, Apt. 202, Department of Statistics, Raleigh, NC 27606, *xlu2@ncsu.edu*; Anastasios A. Tsiatis, North Carolina State University

***Key Words:*** log-rank test, proportional hazard model, nuisance tangent space, influence function, RAL estimators, efficient estimators

The log-rank test is widely used in many clinical trials for comparing censored survival times between two treatments. Under the assumption of proportional hazards, that is, the logarithm of hazard ratio between two survival times is a constant fl, it is optimal for testing the null hypothesis of H0 : fl = 0. In practice, additional auxiliary covariates are collected together with the survival times and treatment assignment. If the covariates correlate with survival times, making use of their information will increase the efficiency of log-rank test. In this paper, we apply the theory of semiparametrics to characterize a class of regular and asymptotic linear (RAL) estimators for fl when auxiliary covariates are incorporated into the model, and derive estimators that are more efficient. The Wald tests induced by these estimators are shown to be more powerful than the log-rank test.

### Fitting Density Function with Exponential Polynomials

❖ Eugene Demidenko, Dartmouth Medical School, 7927 Rubin, DHMC, Lebanon, NH 03756, *eugened@dartmouth.edu*

***Key Words:*** maximum likelihood, density estimation, nonparametric statistic

A possibly multimodal density is modeled via exponential orthogonal polynomials. An efficient Fisher scoring algorithm for maximum likelihood density estimation is described. Statistical hypothesis testing is emphasized, such as test of normality and density multimodality. The density estimation is illustrated with two biomedical examples: brain oxygen distribution and toenail arsenic distribution among New Hampshire residents.

### U-Statistics for Right-Censored Data

❖ Dipankar Bandyopadhyay, University of Georgia, 204 Statistics Building, Dept. of Statistics, University of Georgia, Athens, GA 30602, *dban@stat.uga.edu*

***Key Words:*** u-statistics, right censoring, Kaplan-Meier, Kendall's tau

A censored version of U-statistics with a general kernel of size m is obtained. An expression of its asymptotic normality is obtained through regular martingale arguments and validated using probability-probability plots. Using two different kernels, we study its performance measures by simulation. Considering a Kendall's tau kernel, we apply our test for testing independence of time to failure and cause of failure in a competing risk model using simulation and study its performance through size and power values. Its functionality is also assessed by applying it on a real data set.

### Extension of the Rank Sum Test for Clustered Data: Two Group Comparisons with Group Membership Defined at the Subunit Level

❖ Bernard Rosner, Harvard Medical School, 181 Longwood Ave., Channing Laboratory, Boston, MA 02115, *stbar@channing.harvard.edu*; Robert Glynn, Brigham and Women's Hospital; Mei-Ling Ting Lee, Harvard Medical School

***Key Words:*** nonparametric statistics, parallel design, contralateral design, two-stage designs

The Wilcoxon rank sum test is used widely for two-group comparisons for non-normal data. An assumption of this test is independence of sampling units between and within groups. For some clustered data designs, group membership may be defined at the subunit level. In general, binary eye-specific covariates may be present (scored as exposed or unexposed) and one wishes to compare non-normally distributed outcomes between exposed and unexposed eyes using the Wilcoxon rank sum test while accounting for the clustering. In this paper, we present a corrected variance formula for the Wilcoxon rank sum statistic in the setting of eye (subunit)-specific covariates. We apply it to an ocular allergy clinical trial, where some patients receive the same eye drop in both eyes, while other patients receive different eye drops in fellow eyes.

### Outputation Permutation: Exact Inference for Complex Clustered Data

❖ Dean A. Follmann, National Institute of Allergy and Infectious Diseases, Bethesda, MD , *dfollmann@niaid.nih.gov*; Michael P. Fay, National Institute of Allergy and Infectious Diseases

***Key Words:*** correlated data, clustering, permutation test, within cluster resampling

This talk introduces exact permutation methods for use when there are independent clusters of data with arbitrary within cluster correlation. To eliminate the problem of clustering, we randomly select a data point from each cluster and for this now independent data, calculate our test statistic and the associated support points for all possible permutations. While clearly valid, this is also inefficient. We repeat this process until all possible independent data sets have been created and use the averaged (over each unique permutation) support points as our reference distribution for the averaged test statistic. This approach is similar in spirit to the within cluster resampling of Hoffman, Sen, \& Weinberg (2001), or multiple outputation of Follmann, Proschan \& Leifer (2003). We both exact and Monte Carlo version of this idea.

### A New Semiparametric Regression Approach for Modeling Group Differences

❖ Qin Yu, University of Rochester, 601 Elmwood Ave., Box 630, Department of Biostatistics, Rochester, NY 14642, *qin_yu@urmc.rochester.edu*; Xin Tu, University of Rochester

***Key Words:*** non-parametric models, ANOVA, u-statistics, generalized estimating equation (GEE), asymptotic properties, Mann-Whitney-Wilcoxon (MWW) rank-sum test

The most popular approach for modeling group differences is the one-way analysis of variance (ANOVA) model, but it only compares the mean response of different groups. In many real data applications, differences also often exist in variances. Comparing such second-order variability is of great interest for effectiveness studies which are typically conducted under a community setting. In addition, it is critical to ensure homogeneous variance for valid inference for group differences in ANOVA using F tests. In this talk, we introduce a new semi-parametric regression model to jointly model the mean and variance with inference based on U-Statistics based estimating equations. This new model can also used to model and extend other non-parametric statistics such as the two Mann-Whitney-Wilcoxon (MWW) rank-sum test to more then two groups.

## 546 Multiple Comparisons and Testing

Biometrics Section

**Thursday, August 10, 10:30 am–12:20 pm**

### The Combined p-value for Detecting Differentially Expressed Genes from High Density Oligonucleotide Arrays

❖ Ann Hess, Colorado State University, Department of Statistics, Fort Collins, CO 80523, *hess@stat.colostate.edu*; Hari Iyer, Colorado State University

***Key Words:*** microarrays, differential gene expression

Currently, most tests of differential expression are performed using expression summary values representing each probe set on a microarray. We propose using Fisher's combined p-value method to combine probe level tests of significance. The combined p method and other competing methods were compared using three spike-in data sets where the differentially expressed genes are known. Based on power and false discovery rates computed for the spike-in data sets, we demonstrate that the combined p method performs well. In addition, probe level tests allow us to develop useful diagnostic procedures for exploratory analysis of microarray data.

### A Two-Step Multiple Comparison Procedure for a Large Number of Tests and Multiple Treatments

❖ Hongmei Jiang, Northwestern University, 2006 Sheridan Road, Department of Statistics, Evanston, IL 60208, *hongmei@northwestern.edu*; Rebecca W. Doerge, Purdue University

***Key Words:*** false discovery rate, multiple comparisons, testing differential expression

For situations where the number of tested hypotheses is increasingly large, the power to detect statistically significant multiple treatment effects decreases. As is the case with microarray technology, researchers often are interested in identifying differentially expressed genes for more than two types of cells or treatments. A two-step procedure is proposed for the purpose of increasing power to detect significant effects (i.e., to identify differentially expressed genes) while controlling the overall false discovery rate (FDR). The procedure is presented for both a fixed rejection region and a prespecified significance level. When compared via simulation, the two-step approach has increased power over a one-step procedure and controls the FDR at a desired significance level.

### Optimality Results for the Bonferroni Method with Large m

❖ Yonggang Lu, Texas Tech University, MS2101 ISQS COBA, Texas Tech University, Lubbock, TX 79409, *yonggang.lu@ttu.edu*; Peter Westfall, Texas Tech University

***Key Words:*** multiple comparisons, Bonferroni method, asymptotic optimality, loss function

Modern methods of multiple comparisons, particularly those based on controlling the false discovery rate, are lax relative to the Bonferroni method in their assignment of significances; they are relatively more lax as m, the number of tests, increases. We point out that this laxness is based on an assumption concerning the size of the loss due to Type I errors relative to loss due to Type II errors, and challenge the generality of this assumption, providing alternative types of loss function for which the Bonferroni method is asymptotically (as m8) optimal.

### Control of the Family-Wise Error Rate for Multiple Correlated Test Outcomes: the Effect of Stress on Cytokine Production Prior to Spaceflight

❖ Alan Feiveson, National Aeronautics and Space Administration, Johnson Space Center, Mail Code SK3, Houston, TX 77058, *alan.h.feiveson@nasa.gov*; Satish Mehta, Enterprise Advisory Services, Inc.; Duane L. Pierson, NASA Johnson Space Center

***Key Words:*** multiple testing, family-wise error rate, cytokines, spaceflight, correlated p-values

This study compared levels of cytokine production measured in astronauts' blood samples at time periods shortly before and after NASA space missions with corresponding levels under normal conditions. The problem was how to choose a reasonable cutoff p-value for deciding which cytokines showed a significant change, taking into account the highly-correlated cytokine measurements and extremely unbalanced design of the study. The approach was to construct a pseudosample of p-values representative of the case where all 21 null hypotheses (corresponding to zero logistic regression coefficients) were true, while maintaining the dependence of p-values induced by the original measurements. From this pseudo-sample obtained by bootstrapping, it was possible to estimate the distribution of the minimum p-value. The 5-th percentile of this distribution was then used as the p-value cutoff.

### Effects of Dependencies in High-Dimensional Multiple Testing Problems

❖ Kyung In Kim, Eindhoven University of Technology, P.O. Box 513, Department of Mathematics, Eindhoven, 5600 MB The Netherlands, *k.i.kim@tue.nl*; Mark A. van de Wiel, Eindhoven University of Technology

***Key Words:*** gene expression data, multiple testing, conditional independence

We consider effects of dependencies among variables of high-dimensional data in multiple testing problems. Recent simulation studies considered only simple correlation structure among variables, which was hardly inspired by real data features. Our aim is to describe dependencies as a network and systematically study effects of several network features like sparsity and correlation strength. We discuss a new method for efficient guided simulation of dependent data, which satisfy the imposed network constraints. We use random correlation matrices and perform extensive simulations under nested conditional independence structures. We check the robustness against dependency of several resampling-based methods. Powers computed from popular methodologies such as Benjamini-Hochberg FDR, SAM and mixture models are compared. Finally some applications to gene expression data are illustrated.

### Hypothesis Testing of High-Dimensional Data with Applicaitons to Medical Image Analysis

❖ Kun Nie, Boehringer Ingelheim, 124 Coalpit Hill Road, #111, Danbury, CT 06810, *kun.nie@gmail.com*

***Key Words:*** functional linear regression, high dimensional data, Fourier transform, wavelet transform, medical image analysis

High dimensional data sets can be easily obtained in the forms of time series, images and shapes with modern technologies. An important statistical issues is high dimensional hypothesis testing. The power of conventional methods is seriously eroded by the high dimensionality and spatial dependence of the data. We study high dimensional hypothesis testing in a general functional linear regression model. To address the problem of high dimensionality and spatial dependence, we propose two approaches: the Fouriour-based adaptive Neyman (FBAN) test and the wavelet based thresholding (WBTH) test. The proposed methods are extensions of Fan and Lin who considered the problem of testing the differences between two groups of curves. Both simulation study and applications to real data of the proposed methods have shown that FBAN and WBTH have much higher sensitivities than traditional methods.

## 547 Hypothesis Testing in Genetics ●

Biometrics Section, ENAR

**Thursday, August 10, 10:30 am–12:20 pm**

### Testing Hardy-Weinberg Equilibrium for Loci on the X Chromosome

Gang Zheng, National Heart, Lung, and Blood Institute; ❖ Jungnam Joo, National Heart, Lung, and Blood Institute, 6701 Rockledge Drive, MSC 7938, Bethesda, MD 20892, *zhengg@nhlbi.nih.gov*; Chun Zhang, Roche Palo Alto LLC; Nancy L. Geller, National Heart, Lung, and Blood Institute

***Key Words:*** Fisher's combination, HWE, SNP, sex-linked genes

Testing Hardy-Weinberg equilibrium (HWE) is an important step in quality control before genetic data analysis. With 100,000 or more single-nucleotide polymorphisms (SNPs) available for association studies, thousands come from the X chromosome. To study these sex-linked SNPs, testing departure from HWE is important not only for quality control but for determining whether an allele-based analysis is valid. Many statistical methods for testing HWE for loci on autosomal chromosomes have been discussed. For males, a marker on the X chromosome has only one allele. Testing HWE using females is usually less powerful. We will discuss HWE test for loci on the X chromosome.

### A Conditional Test for Finding the Variants that Explain the Evidence for Association

❖ Baoguan Ke, The University of Chicago, 5734 S. University Ave., Department of Statistics, Chicago, IL 60637, *ke@galton.uchicago.edu*

***Key Words:*** association studies, SNPS, haplotypes, linkage disequilibrium (LD), TDT

Association studies are applied commonly to find the genetic markers that contribute to susceptibility to a particular disease. One of the challenges in mapping genes for complex diseases comes from the existence of the multiple risk variants in the same region and the complicated linkage disequilibrium patterns. In the context of family-based studies, we develop a conditional transmission disequilibrium test for haplotype data. The test can be used to infer whether any set of markers will explain all the association in the region. We also will show extensions of our method to the genotype data. We use likelihood calculations to deal with the uncertainty in the haplotype phase. The procedure we propose will result in a large number of tests, and we will discuss how we deal with the multiple testing problem.

### Improved Association Analyses of Disease Subtypes in Case-Parent Trio Studies

❖ Glen Satten, Centers for Disease Control and Prevention, Mailstop K23, 1600 Clifton Road, Atlanta, GA 30333, *gsatten@cdc.gov*; Michael Epstein, Emory University; Irwin Waldman, Emory University

***Key Words:*** association, case-parent trio, ADHD, genetic epidemiology, case-parent triad

When using case-parent trios for genetic association analysis of complex diseases, one may wish to identify genetic variants associated with a disease subtype. Rather than just analyzing only trios with the subtype of interest, we show a new method that includes data from trios without the subtype of interest, and that increases power to detect association with the subtype of interest. We use a likelihood-based framework that permits flexible estimation of allelic effects on disease subtypes and allows for missing parental data. By simulation we show that our proposed association test outperforms association tests that only analyze trios with the subtype of interest. We apply our method to a study of attention-deficit hyperactivity disorder (ADHD) and identify a genetic variant in the dopamine transporter gene associated with a hyperactive-impulsive ADHD subtype.

### Analysis of DNA Gene Sequences with the Smith-Waterman Algorithm

❖ William Owen, University of Richmond, Department of Math and CS, Richmond, VA 23173, *wowen@richmond.edu*

*Key Words:* estimation, extreme value distribution, percentile, scoring function

A gene sequence comparison algorithm provides researchers a means for judging the similarity of DNA sequences. Evolutionary changes are tracked by considering the relative similarity of different DNA sequences, whose differences may be the result of possible insertions, deletions, and/or substitutions of nucleotides that can occur over time. The renowned Smith-Waterman algorithm uses a dynamic programming approach along with a tunable scoring function to quantify the goodness-of-fit for alignments of sequences. Scoring-function values using randomly-generated sequences are known to follow an extreme value distribution; further, empirical evidence suggests that scoring-function values using actual DNA sequences that are not "well-matched" also tend to be extreme value. Here, we will investigate the distributions of scores using both artificially-generated and actual DNA sequence data.

### A New Approach To Performing Segregation Analysis To Detect Imprinting

❖ Sanjay Shete, M. D. Anderson Cancer Center, Department of Epidemiology, 1155 Pressler Street - Unit 1340, Houston, TX 77030, *sshete@mdanderson.org*

*Key Words:* likelihood ratio test, segregation analyses, linkage, family data

Many complex heterogeneous diseases do not exhibit a simple Mendelian transmission pattern. Genomic imprinting is a mechanism by which only one copy of a gene pair is expressed, and this expression is determined by the parental origin of the copy. The Imprinted Gene Catalogue now has more than 200 genes listed, and estimates based on mouse models suggest many more may exist in humans. The deregulation of imprinted genes has been implicated in a number of human diseases. Currently there are no methods that guide investigators to test for imprinting in the segregation phase of the analyses. Here we present a novel coding system in segregation analyses that allows for testing for imprinting using likelihood ratio test. We present simulation results and real data analyses to exhibit utility of our approach.

### Missing Genotypes in TDT

❖ Gulhan Alpargu, California State University, Fullerton, CSUF Mathematics Department, 800 N State College McCarhty Hall 154, Fullerton, CA 92834, *galpargu@fullerton.edu*

*Key Words:* missing genotype, Mendelian inheritance, linkage disequilibrium, TDT

Transmission disequilibrium test (TDT) attempts to detect markers in linkage disequilibrium with a disease in the presence of association. TDT uses complete genotype information from trios. Genotyping errors or disease with late onset may cause missing genotypes. It is common to exclude families when at least one of the genotypes is missing. Several approaches have been proposed to handle missing genotypes of parents, but not much attention has been given to that of children. We present a robust TDT (rTDT) that handles missing genotypes on any trios. The rTDT produces minimum and maximum values of the TDT statistics, consistent with all possible completions of the missing data. We apply rTDT to identify markers of susceptibility to Crohn disease. We show that only two of the 11 markers originally associated to phenotype do not depend on the assumptions about the missing data mechanism.

### Analysis of a Probe-Level Linear Mixed Model for Oligonucleotide Arrays

❖ Alexander Cambon, University of Louisville, 3802 Fallen Timber Drive, Louisville, KY 40241, *accamb01@louisville.edu*; Caryn Thompson, University of Louisville; Brian Wattenberg, University of Louisville

*Key Words:* probe, mixed models, microarrays, oligonucleotide, affymetrix, gene expression

As the number of replicates and treatments in microarray experiments increases, mixed models are becoming more widely applicable as a method for assessing statistical significance of gene expression data. In this paper, a randomized complete block design consisting of 21 hgu133plus2 microarrays, 7 treatments and three replicates is analyzed using the methods outlined in Chu et al.'s 2002 paper. This method includes probe as a fixed factor and array as a random effect. The results are compared with those obtained using quantile normalization, median polish, and two-way ANOVA. One advantage of the probe level model is that probe variation within probe set is accounted for by including probe as a factor. This increases the sensitivity of the test for treatment effect. The probe level model also increases the sample size for residual analysis and outlier detection of influential arrays.

## 548 Step Up, Step Down Multiple Comparisons ●

Biopharmaceutical Section

**Thursday, August 10, 10:30 am–12:20 pm**

### A Non-Bonferroni Step-Up Rejection Procedure

❖ Jianjun Li, Merck Research Laboratories, 785 Jolly Road, Blue Bell, 19422, *jianjun_li@merck.com*

*Key Words:* Bonferroni inequality, multiple tests, family-wise error rate

Common characteristics of the well-known p-value--based stage-wise procedures for multiple tests are that the critical values of the procedures are derived based on the Bonferroni inequality or its modifications and that the critical values at late stages are irrelevant to the magnitude of the p-values used at early stages. Noting the p-values at the early stages can be informative for decisions at the late stages, this paper derives a non-Bonferroni inequality and develops a step-up procedure. The new procedure is as simple as Hochberg's, but remarkably powerful if the p-values are likely to be small (<0.625). In confirmatory clinical trials, the p-values of multiple tests are anticipated to be below or around 0.05 (type-I error rate). The use of the proposed procedure will increase the probability of the trial success greatly.

### Gatekeeping Procedures for Dose-Finding Problems with Multiple Endpoints

❖ Xin Wang, Northwestern University, Department of Statistics, 2006 Sheridan Road, Evanston, IL 60208, *xinbeng@northwestern.edu*

*Key Words:* gatekeeping, stepwise procedure, dose finding

In many dose finding studies there are hierarchically ordered endpoints and a given dose is compared with a control on any endpoint conditional on the tests on the higher-ordered endpoints being significant (serial gatekeeping). It is required to control the familywise error rate at a designated level taking into account multiplicity of tests. We give a closed procedure (Marcus, Pertiz and Gabriel 1976) for this problem by applying the general and flexible tree-structured testing approach to gatekeeping problems developed in Dmitrienko, Wiens, Tamhane and Wang (2006). The proposed procedure uses weighted Bonferroni tests for testing intersection hypotheses. For an easier implementation of this closed procedure, we give a stepwise procedure that uses penalized Bonferroni tests for all endpoints except the last, for which it uses a penalized Holm's test.

### Comparison of Two Step-Down Linear Trend Tests

❖ Kaifeng Lu, Merck & Co., Inc., 126 E. Lincoln Avenue, Merck Research Laboratories, Rahway, NJ 07065, *kaifeng_lu@merck.com*

***Key Words:*** closed testing principle, power, step-down

In dose ranging studies, treatments are usually composed of placebo and a series of increasing doses of active drug. Researchers are often interested in testing for linear trend in the response curve and identifying the maximum dose concentration of the active drug which is not significantly different from placebo. This talk discusses a class of step-down testing procedures which conform to the closed testing principle. The power and type 1 error rate for two commonly used step-down linear trend tests are compared. The alternative hypotheses at which the two step-down testing procedures are targeted are identified. Simulation studies are carried out to assess the performance of the two step-down testing procedures. Extensions to discrete responses are also discussed.

### A New Gatekeeping Strategy for Hierarchical-Structured Hypotheses in Clinical Trials

❖ Junyuan Wang, The Medicines Company, 8 Campus Drive, Parsippany, PA 07054, *jerry.wang@themedco.com*; Yan Zheng, University of Minnesota; Guanghan Liu, Merck Research Laboratories

***Key Words:*** multiplicity adjustment, weighted Simes, Hochberg method

Gatekeeping strategies have been proposed to exploit the hierarchical structure of primary and secondary hypotheses when there are multiple primary hypotheses. Dmitrienko et al. (2003) suggested a weighted Simes method that provides high power compared to other gatekeeping methods. However, it satisfies neither of the two properties suggested in their paper. We propose two two-stage gatekeeping strategies that are simpler, with comparable power for the primary hypotheses and higher power for the secondary hypotheses. The proposed two-stage methods satisfy (1) overall significance of the primary family will not be affected by the significance of the secondary hypotheses and (2) only if at least one of the primary hypotheses is significant, can one proceed to test the secondary family. The properties of the new gatekeeping methods are demonstrated through simulation.

### Hochberg's Step-Up Method: Cutting Corners off Holm's Step-Down Method

❖ Yifan Huang, H. Lee Moffitt Cancer Center & Research Institute, 12902 Magnolia Drive, MRC, Tampa, FL 33612, *huangy@moffitt.usf.edu*; Jason Hsu, The Ohio State University

***Key Words:*** Hochberg's method, Holm's method, multiple testing, step-up test, step-down test, partition testing

Holm's method and Hochberg's methods are popular multiple tests. They are viewed as step-down and step-up versions of the Bonferroni test. We will present our insight that both are special cases of partition testing, and the difference is while Holm's method tests each partition hypothesis using the largest order statistic setting critical value based on the Bonferroni inequality, Hochberg's method tests each partition hypothesis using all the order statistics setting a series of critical values based on Simes' inequality. Geometrically, Hochberg's step-up method "cuts corners" off the acceptance regions of Holm's step-down method by making assumption on the joint distribution of the test statistics. Partition testing making use of the joint distribution of the test statistics is more powerful than partition testing using probabilistic inequalities, thus is recommended.

### Testing the Assumption in Several Amalagamation-Based Tests for Dose Response

❖ Arthur Roth, Pfizer Inc., 2800 Plymouth Road, Ann Arbor, MI 48105, *arthur.j.roth@pfizer.com*

***Key Words:*** dose response, monotonicity, isotonic regression, amalgamation, trend tests, order restriction

Before performing any statistical test that makes an assumption, it is often desirable to do a preliminary test to check the assumption. Several known tests for dose response employ isotonic regression and amalgamated means, including the chibar square test (nonparametric), the ebar square test (parametric with equal variances), and the Brown-Forsythe trend test (parametric with unequal variances, Roth 1983); all of them assume that the means are a monotonic function of dose. For one-sided alternatives, each of them has a known corresponding test for assessing the validity of the monotonicity assumption (one due to Roth 2005) that has no obvious two-sided analogue. We modify all three of the above one-sided tests for monotonicity to obtain appropriate two-sided tests for monotonicity in each setting (nonparametric, parametric with equal variances, and parametric with unequal variances).

### Exploring Dose Response in Flexible-Dose Clinical Trials Using Marginal Structural Models: a Clinical Trial Example

❖ Ilya Lipkovich, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, *lipkovichia@lilly.com*; Craig Mallinckrodt, Eli Lilly and Company; Douglas Faries, Eli Lilly and Company

***Key Words:*** marginal structural models, inverse probability of treatment weighting, dose-response analysis, selection bias, causal inference

Assessing dose response from flexible-dose clinical trials is problematic. The true dose effect may be obscured and even reversed in observed data because dose is related to both previous and subsequent outcomes. To remove selection bias we propose marginal structural models, inverse probability of treatment weighting methodology. Clinical outcomes are compared across dose groups using weighted ANCOVA with weights estimated from the same data (via logistic regression) and determined as products of (i) inverse probability of receiving dose assignments that were actually received and (ii) inverse probability of remaining on treatment by this time. In simulations this method led to almost unbiased estimates of true dose effect under various scenarios. Results were compared with those by unweighted analysis. Clinical trial data are used to illustrate strengths and limitations of the method.

## 549 Better Statistical Intervals with Applications

Section on Physical and Engineering Sciences
**Thursday, August 10, 10:30 am–12:20 pm**

## Adjusting Likelihood Ratio Confidence Intervals for Parameters Near Boundaries Applied to the Binomial

❖ Sundar Dorai-Raj, PDF Solutions, Inc., 101 W. Renner Road, Suite 325, Richardson, TX 75082, *sundar.dorai-raj@pdf.com*; Spencer Graves, PDF Solutions, Inc.

***Key Words:*** probability, minimax, coverage

This research focuses on improving likelihood ratio confidence intervals for parameters at boundaries as applied to binomial confidence intervals. We compute the "inferred confidence coefficient" required to minimize the maximum absolute deviation from the nominal confidence level over all points of discontinuity of the actual coverage probability. Generalizations of these results can help improve confidence interval procedures for many application with parameters near boundaries.

## Parametric 95%--95% Upper Tolerance Limits for Left-Censored Lognormal Data

❖ Charles Davis, EnviroStat, 3468 Misty Court, Las Vegas, NV 89120, *charles.davis@envirostat-nv.com*

***Key Words:*** tolerance limits, censored data, nondetects, lognormal data

Nine Upper Tolerance Limits (UTL) methods are compared: two simple heuristics; two MLE methods; two probability plot methods using PLEs; three simpler prob-plot methods. The first six accommodate multiple censoring points. Factors are (a) distribution (CV = 0.1 to 4.0), (b) censoring proportion (10% to 70%, with five multiple censoring patterns), (c) rounding (fine to coarse), and (d) sample size (10 to 100). Log-scale UTLs are (mean + C * K * sd), where K is the usual multiplier and C = C(N) is a calibration factor intended to approximate 95% confidence over all factor combinations. Criteria are (alpha) how well the empirical confidence maintains 95% and (beta) a power evaluation using "Median Excess" (median UTL divided by actual 95% point). Conclusions include (A) always use fine rounding, (B) for 10 < = N < 20, if one can believe CV < = 2 the uniform heuristic is good, (C) otherwise MLE is preferred.

## Estimating the Variance of the Graybill-Deal Estimator of a Common Mean

❖ Nien Fan Zhang, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8980, *nien-fan.zhang@nist.gov*

***Key Words:*** coverage probability, inter-laboratory comparison, mean square error, unbias, weighted mean

Statistical inference of the common mean of several normal populations with unknown and possibly unequally variances has a long history. In applied statistics, the problem is of combining several estimators of an unknown quantity to obtain an estimator with good statistical properties. For this, weighted mean estimators such as the Graybill-Deal estimator have been used widely. However, the traditional estimator for the variance used in the Graybill-Deal estimator underestimates the true variance. It is well known that the variance of an estimator is as important as the estimator itself. In this paper, some properties of the variance estimators of the Graybill-Deal estimator are discussed. Two new variance estimators are proposed with smaller biases and the correspondingly formed intervals have much better coverage probabilities based on simulation for the true mean.

## Revisiting Beal's Confidence Intervals for the Difference of Two Binomial Proportions

❖ Joshua M. Tebbs, University of South Carolina, Department of Statistics, 209F LeConte, Columbia, SC 29208, *tebbs@stat.sc.edu*; Scott Roths, The Pennsylvania State University

***Key Words:*** adaptive estimation, Agresti-Caffo intervals, empirical Bayes methods, maximum likelihood, plant disease

Confidence interval construction for the difference of two independent binomial proportions is a well-known problem with a full panoply of proposed solutions. In this talk, we largely focus on the family of intervals proposed by Beal (1987). This family, which includes the Haldane and Jeffreys-Perks intervals as special cases, assumes a symmetric prior distribution for the population proportions. We propose new methods that allow the currently observed data to set the prior distribution by taking a parametric empirical-Bayes approach. In addition, we provide an investigation of the new intervals' behaviors in small-sample situations. Unlike other solutions, our intervals can be used adaptively for experiments conducted in multiple stages over time. We illustrate this notion using data from a study involving the "Mal de Rio Cuarto" virus and its transmission to maize.

## Normal Approximations for Computing Confidence Intervals for Log-Location-Scale Distribution Probabilities

❖ Yili Hong, Iowa State University, 114 University Village, Apt. D, Department of Statistics, Ames, IA 50010, *hong@iastate.edu*; William Q. Meeker, Jr., Iowa State University; Luis A. Escobar, Louisiana State University

***Key Words:*** censored data, maximum likelihood, quantile

Normal approximation confidence intervals (CIs) are used in most commercial statistical package because they are easy to compute. However, the performance of such procedures could be poor when the sample size is not large or when there is heavy censoring. A transformation can be applied to improve the approximation, but the degree of improvement depends on the chosen function because some seemingly useful transformation functions will cause the estimated variance blow-up in extrapolation. This article reviews statistical methods to construct CIs for distribution probabilities based on a normal approximation and studies the properties of these CI procedures. Our results suggest that a normal approximation CI procedure based on a studentized statistic, which we call the zhat procedure, has desirable properties. We also illustrate how to apply the zhat method to more general situations.

## Simulation Results To Assess Upper Tolerance Interval Methodology That Adjusts for Multiple Nuisance Uncertainties

Greg Piepel, Battelle-PNNL; ❖ Scott Cooley, Battelle, 902 Battelle Blvd., PO Box 999, Richland, WA 99352, *scott.cooley@pnl.gov*; Matthew Paul, Western Washington University

***Key Words:*** analytical uncertainty, model uncertainty, Monte Carlo simulation, noncentral t-distribution, sampling uncertainty, Satterthwaite's method

A statistical statement with X% confidence that at least Y% of a particular population is below a specified value can be made using an X%/Y% upper tolerance interval (X%/Y% UTI). This presentation provides methodology and formulas for calculating X%/Y% UTIs when multiple sources of nuisance uncertainty are present in addition to the source of variation of interest. The X%/Y% UTI methodology is developed and illustrated using an example with sampling, analytical, and regression model nuisance uncertainties plus the source of variation for which the tolerance interval is desired. The design and results of a simulation study conducted to assess the performance of the X%/Y% UTI methods are described. The method that adjusts for nuisance uncertainties is shown to achieve the intended X% and Y% levels, and yields significantly narrower UTIs than not adjusting for nuisance uncertainties.

## 550 Anxiety, Ambiguity, and Multiculturalism in Statistical Education ● ✪

Section on Statistical Education

**Thursday, August 10, 10:30 am–12:20 pm**

### Ambiguity Intolerance: an Impediment to Inferential Reasoning?

❖ Robert Carver, Stonehill College, 39 Woodland Street, Sharon, MA 02067, *rcarver@stonehill.edu*

***Key Words:*** tolerance of ambiguity, learning styles, teaching inference

In an introductory statistics course, undergraduate students often struggle with the concepts and techniques of statistical inference. At the heart of inference is the inconvenient fact that we often need to make decisions or draw conclusions without benefit of all the relevant facts in ambiguous situations. There is reason to think that students vary in their attitudes and openness to ambiguity in general, and that an individual's discomfort with or intolerance of ambiguity could impede one's learning of inferential reasoning. Yet, little research has considered ambiguity tolerance as an explanatory or moderating factor in learning to apply the techniques of inference directly. This paper reports on empirical classroom research to investigate the extent to which intolerance of ambiguity is an impediment to learning about statistical inference.

### Stochastics Learning Environment: Can a Professor Identify and Displace Student Statistical Anxiety?

❖ Daniel Edelman, Illinois Institute of Technology, 3424 S. State Street, Room 4007, Chicago, IL 60616, *Edeldan@iit.edu*

Statistics anxiety is experienced by a significant number of students at all levels of education. It is reported that statistics is one of the most dreaded course in many curriculums. This project studied the teaching methodology of an experienced higher education instructor and the ability to identify and confront student statistical anxiety. Interviews with students and statistical anxiety assessment instrument were utilized to determine anxiety. Students' and teachers' statistical anxiety perceptions were compared. Methods to identify and displace statistics anxiety are presented.

### Investigating Communication within a Multitiered Instructional Team

❖ Marian Frazier, The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210, *frazier.149@osu.edu*; Jackie Miller, The Ohio State University

***Key Words:*** communication, education, large lecture courses, instructional teams

When dealing with large lecture introductory statistics courses, it is often the case that there is a course coordinator, lecturers, and teaching assistants involved in educating the students. These people, in conjunction with the students taking the course, form a multi-tiered instructional team. The communication needed to make such a team work (and work well) was investigated in Spring 2005. This research concentrated on two major issues: 1) Do those involved in the course make an effort to communicate with those with whom they don't usually interact? 2) Would the teaching/learning experience be improved for all involved if the lines of communication were more open?

### Can You Explain That in Plain English? Making Statistics Group Projects Work in a Multicultural Setting

❖ Michelle R. Sisto, International University of Monaco, 2 Ave. Prince Hereditaire Albert, Monte Carlo, 98000 Principality of Monaco, *msisto@monaco.edu*

***Key Words:*** education, projects, multicultural, assessment, communication

Students increasingly need to learn to communicate statistical results clearly and effectively, as well as to become competent consumers of statistical information. I discuss here how group projects and peer assessment have helped us to meet these objectives at the International University of Monaco. I generally have ten to twenty different nationalities in my undergraduate and MBA statistics courses, and I address here specific language and software challenges I've encountered teaching and directing group work in this multicultural context. I also focus on the learning benefits of having students work collaboratively to discuss, write, present, and assess statistics projects in English. Projects consist of a memo, an appendix, an oral PowerPoint presentation, and self and peer assessment. Some examples of student work, student assessments, and feedback will be available.

### Balancing Cultural Differences in Teaching Statistics

❖ Huizhen Guo, Xavier University, 9214 Deercross Parkway, Apt. 2B, Cincinnati, OH 45236, *guoh@xavier.edu*

***Key Words:*** teaching statistics, sole statistician, cultural differences, evaluation

With more than ten years of teaching experience in Mathematics in China and three years of teaching experience in statistics as a graduate student in US, teaching statistical courses at all levels as a sole statistician in a liberal art school is still quite challenging and therefore exciting for me. Part of the challenge is the gap between students and I that keeps me from understanding students and keeps students from accepting me. In this paper I would like to address how the cultural differences play a big role for the misunderstanding and how the gap can be filled by balancing the cultural differences and adjusting my teaching style. Hope this can be helpful to new faculty members who have different cultural backgrounds.

### A Window on Statistical Education at the University of Lagos, Lagos, Nigeria

❖ Raymond Okafor, University of Lagos, 169 Crown Point Drive, Hype Park, MA 02136, *okaforray@yahoo.com*

***Key Words:*** UNILAG statistics, degree programs, enrollment, degrees awarded, staffing, curricula teaching

We focus on the teaching of statistics at the University of Lagos (UNILAG, for short) Lagos, Nigeria. It is common to find journal articles dealing with programs in statistics at universities and research institutions located in statistically advanced countries such as USA and UK. But it is not quite as common to come across similar articles from under-developed countries in Africa. Yet there is much to be gained on the part of statisticians from the latter countries if they can avail themselves of opportunities to interact with their international or should we say transcontinental colleagues. We showcase what is on offer statistically at UNILAG, highlighting critical areas such as available degree programs, enrollment, degrees awarded, staffing, curricula, organization of teaching and teaching infrastructures, present challenges and future prospects.

## 551 Classification

Section on Statistical Computing

**Thursday, August 10, 10:30 am–12:20 pm**

### Evidence Contrary to the Statistical View of Boosting

❖ David Mease, San Jose State University, Marketing Department, San Jose, CA 95192-0069, *dmease@umich.edu*

***Key Words:*** boosting, logitboost, adaboost

The statistical perspective on boosting algorithms focuses on optimization, drawing parallels with maximum likelihood estimation for logistic regression. In this talk we present empirical evidence that raises questions about this view. Although the statistical perspective provides a theoretical framework within which it is possible to derive theorems and create new algorithms for general contexts, we show that there remain many unanswered important questions. Furthermore, we provide examples that reveal crucial flaws in the many practical suggestions and new algorithms that are derived from the statistical view. We examine experiments using simple simulation models to illustrate some of these flaws and their practical consequences. This is joint work with Abraham Wyner at the University of Pennsylvania.

### Adaptive Learning Rate in Stochastic Boosting

❖ Mark Culp, University of Michigan, 1843 Pointe Crossing Street, 201, Ann Arbor, MI 48105, *culpm@umich.edu*; George Michailidis, University of Michigan; Kjell Johnson, Pfizer Inc.

***Key Words:*** stochastic gradient boosting, learning rate, shrinkage, exponential loss

We present the Dynamic Ensemble Machine as an extension of stochastic gradient boosting that relies on the adaptive learning rate and local cross-validated estimates of regularization. Specifically, the adaptive learning rate is formed by the local ratio of the current penalized boosting stageweight over the unpenalized version of that stageweight. We show that the original learning rate can be expressed equivalently by a local ridge penalty under squared error loss (classification). Using the simple form of this parameter, one can obtain k-fold cross-validated estimates for shrinkage quickly while the ensemble is constructed. We further present two specific versions of the adaptive learning rate designed for exponential loss. These penalties provide flexible versions of the adaptive learning rate, which depend on the current selected learner.

### Additive Regression Trees and Smoothing Splines - Predictive Modeling and Inference in Hybrid Learning

❖ Bin Li, The Ohio State University, 631 Cuyahoga Court, Columbus, OH 43210, *bli@stat.ohio-state.edu*; Prem K. Goel, The Ohio State University

***Key Words:*** boosting, marginal effect, problems of extrapolation, variable importance, non-additive effect

We propose a new predictive modeling strategy that emphasizes prediction as well as model interpretation for both regression and classification problems. Simulations show that it has comparable prediction performance with tree-boosting method. However, when the target function is additive and/or smooth, it tends to achieve better prediction performance. In addition, it allows us to (1) estimate marginal effects without extrapolation; (2) measure the variable importance on main and interaction effects separately; (3) select variables based on their importance; (4) test the significance of non-additive effects. We applied our method to two real data sets.

### Use of Projection Pursuit in Classification

❖ Li Li, George Mason University, 4400 University Drive, MSN 5C3, Fairfax, VA 22030, *lli1@gmu.edu*; James Gentle, George Mason University

Projection pursuit (PP) is often used in exploratory data analysis to find interesting projections of data. Sometimes these projections can indicate clusters in the data, and so PP has been proposed as a tool in cluster analysis. Various projection indices have also been proposed so that PP can be used in classification. Because the projections are linear, PP is most effective when there are linear separators between the classes. Some indices are effective only for binary classification, while some can be extended for use with multiple groups. In this paper we consider some practical issues in using PP for classification. We study various extensions and evaluate their effectiveness using Monte Carlo simulations.

### Tuned and Guided Adaptive Regression by Mixing

❖ Panayotis Giannakouros, University of Missouri-Kansas City, 2051 Brookdale Road, Toledo, OH 43606, *lists@potis.org*; Lihua Chen, The University of Toledo

***Key Words:*** model combining, adaptive regression by mixing, model averaging

The development of Adaptive Regression by Mixing (ARM) has provided a theoretical justification for prediction-based model combining methods and demonstrated they can have superior performance under many statistical settings. However, ARM and its implementations in various statistical settings leave potential for improvement. We systematically explore the properties of prediction-based model combining, pursuing development of a superior tuned and guided prediction-based combining algorithm. We use simulations and visualization to explore and optimize key steps of the algorithm and assess the performance of the tuned and guided algorithm relative to ARM in several settings.

# 552 Rank-Based Statistical Methods

●

Section on Nonparametric Statistics
**Thursday, August 10, 10:30 am–12:20 pm**

## One-Sample and Multisample Tests for Repeated Measurement

❖ Chien-Hua Wu, Chung-Yuan Christian University, 200 Chung Pei Road, Chung Li, 32023 Taiwan, *cwu@cycu.edu.tw*; Shu-Mei Wan, Lunghwa University of Science and Technology; Hon-Ron Lin, Chung-Yuan Christian University

***Key Words:*** multivariate central limit theory, delta method, weighted least squares

The underlying model for the weighted least squares approach is the multinomial distribution. Although the distribution assumptions are much weaker, one still must make basic assumptions concerning the marginal distributions at each time point. In addition, the assumptions of specific ordinal data methods, such as the proportional odds model, may be inappropriate. In all these situations, nonparametric methods for analyzing repeated measurements may be of use. The proposed method is to assign ranks to repeated measurements from the smallest value to the largest for each subject. The vector of rank means can be computed by the linear transformation of these ranks. Then, the multivariate central limit theory and delta method are applied to obtain the test statistics. The methods make no assumptions concerning the distribution of the response variable.

## Test-Based Classification Rule: a Linkage between Classification and Statistical Testing

❖ Shu-Min Liao, The Pennsylvania State University, 326 Thomas Building, Department of Statistics, University Park, PA 16802, *sxl340@psu.edu*; Michael G. Akritas, The Pennsylvania State University

***Key Words:*** classification, testing, rank tests, LDA, QDA

The purpose of this talk is to introduce a new classification methodology. The methodology uses a connection, which we uncover, between classification and testing, and is called Test-Based Classification Rule. In particular, when a rank-based testing method is used, we call it Rank Test-Based Classification. We start from binary classification in terms of univariate and multivariate data, and then move on to the multiclass case. Several simulated data sets and real data sets are covered in this talk. We argue that our new idea is competitive with the Linear and Quadratic Classification Rules when the observed data are normally distributed, but it can be a better alternative than the Linear and Quadratic Rules in most of cases when data are not normal. Lanchenbruch's holdout misclassification rate is used as a tool to judge the performance of classification.

## Bayesian Hypothesis Testing Using Nonparametric Statistics

❖ Ying Yuan, M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 447, Houston, TX 77030, *w4ying@yahoo.com*; Valen Johnson, M. D. Anderson Cancer Center

***Key Words:*** Bayes factor, nonparametric hypothesis test, Wilcoxon signed rank test, Mann-Whitney-Wilcoxon test, Kruskal-Wallis test, log-rank test

Traditionally, the application of Bayes testing procedures to classical nonparametric setting has been restricted by difficulties associated with prior specification, prohibitively expensive computation, and the absence of sampling densities for data. We propose methodology for overcoming these difficulties by using nonparametric test statistics to define the Bayes factors required for Bayesian hypothesis testing. We present the methodology for constructing Bayes factors for a wide class of nonparametric test statistics having limiting normal or chi-square distributions. In addition, we demonstrate how this methodology can be extended to simplify meta-analysis in which only p values or the values of test statistics are reported.

## Asymptotic Efficiency of the Majority Rule Relative to Rank-Sum Method for Selecting Best Population

❖ Samuel Wu, University of Florida, Department of Biostatistics, 1329 SW 16th Street, P.O. Box 100212, Gainesville, FL 32610, *samwu@biostat.ufl.edu*; David Annis, Naval Postgraduate School

***Key Words:*** majority rule, rank-sum method, asymptotic efficiency, Plackett-Luce model, translative strength models

The ranking and selection problem has been well-studied in the case of continuous responses. In this paper, we investigate two common decision rules when individuals in the population provide exhaustive rank-orderings of the alternatives. Both the Pitman efficiencies and the lower bounds on Bahadur efficiencies of the majority rule relative to the rank-sum rule are derived, assuming that the rank data are generated from either the Plackett-Luce or the translative strengths models. In addition, finite sample properties of the two methods are compared with the MLE approach through simulation studies. Our results suggest two things. First, when it is substantially more difficult to obtain a complete rank-ordering than simply the top choice, the majority rule performs adequately. Second, the rank-sum rule compares favorably to, and is substantially more robust than, the MLE approach.

## Nonparametric Methods in Multivariate Factorial Designs

❖ Arne Bathke, University of Kentucky, 875 Patterson Office Tower, Lexington, KY 40506-0027, *arne@ms.uky.edu*; Solomon W. Harrar, South Dakota State University

***Key Words:*** nonparametric statistics, multivariate, rank statistics, factorial designs, asymptotics, MANOVA

We present three different multivariate nonparametric tests: ANOVA-Dempster-Type, Lawley-Hotelling, and Bartlett-Nanda-Pillai. Asymptotic theory has been derived for two different asymptotic frameworks: the situation where the number of replications is limited, whereas the number of treatments goes to infinity; and the regular setup with large sample sizes. The nonparametric tests are based on separate rankings for the different variables. For these tests, we also discuss different approximation procedures. The finite performance of the tests is investigated through simulations. It turns out that the proposed nonparametric tests perform very well as compared to their parametric competitors, especially in the presence of outliers. Application using SAS standard procedures is demonstrated with examples corresponding to both asymptotic frameworks.

## Sequential Monitoring of Randomization Tests: Theory and Calculation Discussion

❖ Yanqiong Zhang, Merck & Co., Inc., 81 Shelley Circle, East Windsor, NJ 08520, *yanqiong_zhang@merck.com*; William Rosenberger, George Mason University

***Key Words:*** linear rank test, unconditional test, information fraction

In many clinical trials, it is desirable to establish a sequential monitoring plan, where the test statistic is computed at an interim point or points in the trial and a decision is made on whether or not to stop the trial early due to evidence of treatment efficacy. In this talk, the joint asymptotic property for greater than one inspection during the trial will be explored for the most generally used score functions and randomization sequences. The focus

will be on how to perform sequential monitoring using unconditional randomization tests in practice. Also, the application of the sequential monitoring involves calculations of the multivariate normal probability, thus calculation details will be discussed with respect to using R and SAS packages. A parallel information fraction concept will be proposed, which involves a development of an analog of Fisher's information.

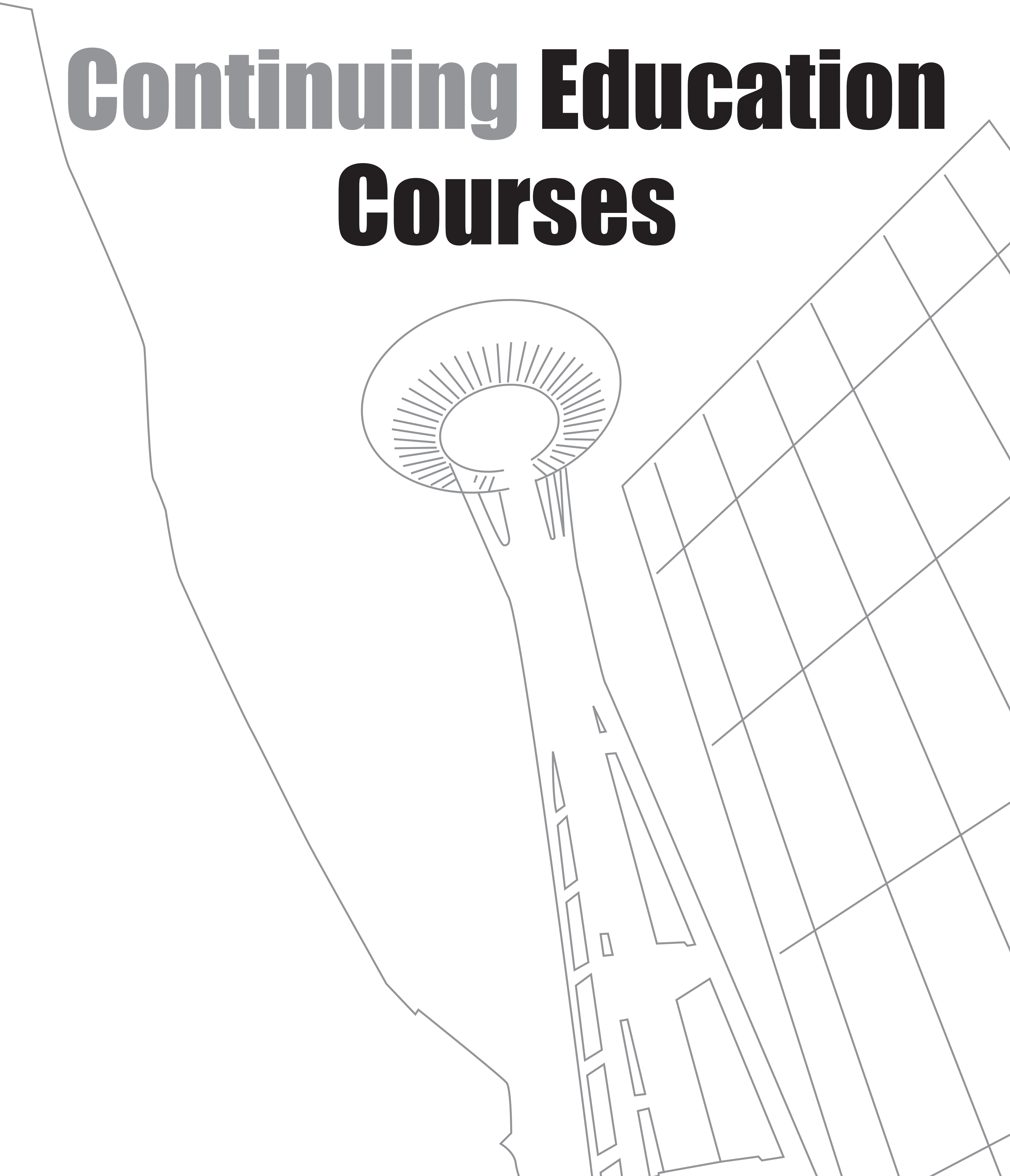
## On the Relationship between Spearman's Rho and Kendall's Tau for Extreme Order Statistics

❖ Yung-Pin Chen, Lewis & Clark College, Mathematics Department, Portland, OR 97219, *ychen@lclark.edu*

***Key Words:*** copula, order statistic, Spearman's rho, Kendall's tau

We will study the relationship between Spearman's rho and Kendall's tau for the two extreme order statistics, minimum and maximum, of independent and identically distributed continuous random variables. We present some new formulas for computing Spearman's rho. One of the formulas leads to a recursion relation. We use this recursion relation to establish inequality relationships between Spearman's rho and Kendall's tau. The recursion relation also provides an alternative proof of the result that the sequence of ratios rho/tau converges to 3/2 as the sample size goes to infinity.

# Continuing Education Courses



# August 5, 2006

**CE_01C 8:30 am–5:00 pm**

## Categorical Data Analysis (two-day course)

The ASA

***Instructor(s):*** Gary Koch, The University of North Carolina at Chapel Hill; Todd Schwartz, The University of North Carolina at Chapel Hill; Rebekkah Dann, The University of North Carolina at Chapel Hill

This course will present the conceptual background and considerations for the application of statistical methods for categorical data analysis. Attention will be given to nonparametric (randomization) methods for testing hypotheses of no association under minimal assumptions and regression methods for fitting statistical models to describe multivariate relationships. Discussion of these strategies will be motivated through examples from clinical trials, observational studies, and sample surveys. Specific topics will include logistic regression; Mantel-Haenszel procedures for sets of 2x2 contingency tables; proportional odds model extension of logistic regression for ordinal data; extensions of Mantel-Haenszel procedures for stratified ordinal data; weighted least squares methods for ordinal data; Poisson (incidence density) regression methods for categorized times to event; methods for studies with repeated measures and/or matching, including generalized estimating equations and conditional logistic regression; and computing procedures for implementing methods.

**CE_02C 8:30 am–5:00 pm**

## Small-Area Estimation

The Section on Survey Research Methods

***Instructor(s):*** Partha Lahiri, University of Maryland

This course will begin with a history of small-area estimation and different uses of small-area statistics in both public and private sectors. It will provide an introduction to important concepts in small-area estimation and outline various approaches for estimating small-area parameters. The prerequisite is knowledge of multiple linear regression and analysis of variance techniques. Topics will include standard design-based methods, traditional indirect methods, and state-of-the-art small-area estimation methods that use mixed models. The course is not designed to provide an in-depth study of any topic, but to provide an overview of small-area estimation. Formulas will be presented wherever necessary to explain some of the advanced topics, but without any derivations. Data analyses using a few real-life examples will be presented. The course is intended for statistics practitioners.

**CE_03C 8:30 am–5:00 pm**

## Modern Approaches to Nonstationary Models of Spatial and Space-Time Processes with Air Quality Applications

The ASA

***Instructor(s):*** Peter Guttorp, University of Washington; Paul D. Sampson, University of Washington

In this course, we will describe a variety of tools-originating in geostatistics but having found their own direction in environmetrics-suitable for estimating the air quality at unmonitored locations with realistic estimates of uncertainty. In particular, we will adapt the geostatistical kriging method to nonstationary spatial fields and nonseparable space-time processes and describe how one can even include deterministic model output in the estimation. The methods will be illustrated with examples out of the lecturers' experience, using freely available software tools.

**CE_04C 8:30 am–5:00 pm**

## Bayesian Inference

The Section on Bayesian Statistical Science

***Instructor(s):*** Bruno Sanso, University of California, Santa Cruz

This course will review the bases of Bayesian inference. It will start by presenting the basic elements of statistical inference that use likelihood functions. We will then consider specifying prior distributions, describe tools for both pointwise and interval estimation and prediction, present the Bayesian theory of hypothesis testing and model comparison, and review the elements of modern computational methods used in the applications of Bayesian models. The course will target students or professionals with a good knowledge of statistics who want to learn or refresh their knowledge of basic Bayesian inference. The level of mathematical sophistication will be kept as low as possible, but calculus and basic probability theory are considered prerequisite. RECOMMENDED TEXTBOOK: Migon, H.S. and Gamerman, D. (1999). Statistical Inference: An Integrated Approach. Oxford University Press. ISBN: 0340740590.

**CE_05C 8:30 am–5:00 pm**

## Practical Data Mining

The ASA

***Instructor(s):*** Richard De Veaux, Williams College

This course will introduce data mining, the exploration and analysis of large datasets by automatic or semiautomatic means with the purpose of discovering meaningful patterns. The knowledge learned from theses patterns can be used for decisionmaking via a process known as "knowledge discovery." Much of exploratory data analysis and inferential statistics concern the same type of problems, so what is different about data mining? What is similar? We will attempt to answer these questions by providing a survey of the problems that motivate data mining and the approaches used to solve them. The course will be case-study and dataset-based and cover many of the algorithms used in data mining from an applications perspective. The applications will come from a variety of industries, and attendees will learn to identify appropriate problems for data mining, explore and prepare data for mining, apply two-stage models, use techniques such as decision trees and neural nets to build accurate predictive models, evaluate the quality of models, and select the appropriate data mining tools for applications.

**CE_06C 8:30 am–5:00 pm**

## Applied Longitudinal Analysis

The Biometrics Section

***Instructor(s):*** Garrett Fitzmaurice, Harvard University

This course will provide an introduction to statistical methods for analyzing longitudinal data. It will emphasize practical aspects of longitudinal analysis, beginning with a review of established methods for analyzing longitudinal data when the response of interest is continuous. We will present an overview of marginal models and generalized linear mixed models. We will highlight the main distinctions between these types of models and discuss the scientific questions addressed by each. Attendees should have a strong background in linear regression and minimal exposure to generalized linear models (e.g., logistic regression). RECOMMENDED TEXTBOOK: Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2004). Applied Longitudinal Analysis. John Wiley and Sons. ISBN: 0-471-21487-6.

## August 6, 2006

### CE_01C 8:30 am–5:00 pm

### Categorical Data Analysis (continuation of two-day course)

The ASA

***Instructor(s):*** Gary Koch, The University of North Carolina at Chapel Hill; Todd Schwartz, The University of North Carolina at Chapel Hill; Rebekkah Dann, The University of North Carolina at Chapel Hill

This course will present the conceptual background and considerations for the application of statistical methods for categorical data analysis. Attention will be given to nonparametric (randomization) methods for testing hypotheses of no association under minimal assumptions and regression methods for fitting statistical models to describe multivariate relationships. Discussion of these strategies will be motivated through examples from clinical trials, observational studies, and sample surveys. Specific topics will include logistic regression; Mantel-Haenszel procedures for sets of 2x2 contingency tables; proportional odds model extension of logistic regression for ordinal data; extensions of Mantel-Haenszel procedures for stratified ordinal data; weighted least squares methods for ordinal data; Poisson (incidence density) regression methods for categorized times to event; methods for studies with repeated measures and/or matching, including generalized estimating equations and conditional logistic regression; and computing procedures for implementing methods.

### CE_07C 8:30 am–5:00 pm

### Text Mining

The Section on Statistical Computing

***Instructor(s):*** David Madigan, Rutgers University; David D. Lewis, David D. Lewis Consulting LLC

This course will provide an overview of the arsenal of statistical and computational advances available, examining both statistical methods that pertain to textual data and language processing techniques that play a vital role in successful text mining. The course will present examples and case studies while attempting to provide an overarching framework that focuses on the connection between textual artifacts and real-world entities. We will describe “mainstream” text mining applications, but will focus on concepts and algorithms, rather than specific commercial software. Prerequisites include knowledge of basic statistical techniques and elementary probability. Familiarity with Introduction to the Practice of Statistics by Moore and McCabe or a similar text would suffice.

### CE_08C 8:30 am–5:00 am

### The Psychology of Survey Response

The Section on Survey Research Methods

***Instructor(s):*** Roger Tourangeau, University of Maryland

This course examines survey questions from a psychological perspective. Its objectives are to cover what is known about how respondents answer survey questions and examine how problems in each component of the response process can produce reporting errors. The class will focus on behavioral questions and include examples of applying findings from the literature on cognitive aspects of survey methods to improve questions. It is intended for researchers who develop survey questionnaires or use survey data and want to understand potential problems with it. It will describe major psychological components of the response process, including comprehension of the questions, retrieval of information from memory, combining and supplementing information from memory through judgment and inference, and reporting an answer. The course has no prerequisites, though familiarity with survey methodology or questionnaire design will be helpful. RECOMMENDED TEXTBOOK: Tourangeau, R., Rips, L.J. and Rasinski, K. (2000). The Psychology of Survey Response. Cambridge University Press. ISBN: 13:9780521576291.

### CE_09C 8:30 am–5:00 pm

### Statistical Methods for the Confirmatory Analysis of Equivalence/Noninferiority Studies

The ASA

***Instructor(s):*** Stefan Wellek, University of Heidelberg

This course will provide working statisticians with a rich repertoire of equivalence testing procedures for use in the confirmatory analysis of studies that aim to prove absence of relevant effects. In addition to equivalence problems in the strict (i.e., two-sided) sense, noninferiority problems will be considered. Throughout, emphasis will be placed on a meaningful choice of the target parameter. Asymptotic techniques applicable to semiparametric and nonparametric models will be presented. For all testing procedures discussed, the description of the decision rule will be supplemented by algorithms for power and sample-size calculation. Each method will be illustrated by a real example. Last, due attention will be given to explaining the use of special computer programs that enable easy implementation of the procedures. RECOMMENDED TEXTBOOK: Wellek, S. (2003). Testing Statistical Hypotheses of Equivalence. CRC Press/Taylor & Francis Group. ISBN: 1584881607.

### CE_10C 8:30 am–5:00 pm

### Computational Statistics: Methods for Monte Carlo Integration and Optimization

The Section on Statistical Computing

***Instructor(s):*** Jennifer A. Hoeting, Colorado State University; Geof H. Givens, Colorado State University

This course will consist of a morning session on Monte Carlo integration strategies and an afternoon session on optimization methods. We will survey a variety of techniques, ranging from classic to state-of-the-art. The course will be based on Computational Statistics. We will seek to give students a practical understanding of how and why existing methods work, enabling students to use modern statistical methods effectively. We will focus on methodological concepts, not details of computer programming. Examples will be drawn from diverse fields, including bioinformatics, ecology, and medicine. The course is targeted for quantitative scientists and statisticians who are unfamiliar with these methods. Upper-division undergraduate mathematical literacy is recommended. RECOMMENDED TEXTBOOK: Givens, G.H. and Hoeting, J.A. (2005): Computational Statistics. John Wiley and Sons. ISBN: 0-471-46124-5.

## CE_11C 8:30 am–5:00 pm

### Regression Modeling Strategies

The ASA

***Instructor(s):*** Frank E. Harrell, Jr, Vanderbilt University School of Medicine

Of the principal assumptions (linearity, additivity, distributional), this course will emphasize methods for assessing and satisfying the first two. Practical, but powerful, tools will be presented for validating model assumptions and presenting model results. This course will provide methods for estimating the shape of the relationship between predictors and response by augmenting the design matrix using restricted cubic splines. Methods for data reduction will be introduced to deal with the number of potential predictors being large in comparison with the number of observations. Methods of model validation will be covered, as will auxiliary topics such as modeling interaction surfaces, efficiently utilizing partial covariable data by using multiple imputation, variable selection, overly influential observations, collinearity, and shrinkage. The methods covered will apply to almost any regression model, including ordinary least squares, logistic regression models, and survival models. RECOMMENDED TEXTBOOK: Harrell, F.E. (2001): Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer. ISBN: 0-387-95232-2.

## CE_12C 8:30 am–5:00 pm

### Hierarchical Bayes Methods and Software for Data Analysis

The Section on Bayesian Statistical Science

***Instructor(s):*** Bradley P. Carlin, University of Minnesota; Thomas A. Louis, Johns Hopkins University

This course introduces hierarchical and empirical Bayes methods, demonstrates their usefulness in challenging applied settings, and shows how they can be implemented using modern Markov chain Monte Carlo (MCMC) computational methods. We also will provide an introduction to WinBUGS. Use of the methods will be demonstrated in advanced, high-dimensional model settings. Participants should have an MS (or advanced undergraduate) understanding of mathematical statistics. Basic familiarity with common statistical models and computing will be assumed, but we will not assume any significant previous exposure to Bayesian methods or computing. The course generally is aimed at students and practicing statisticians who are intrigued by Bayes and Gibbs, but who may mistrust the approach as theoretically mysterious and practically cumbersome. RECOMMENDED TEXTBOOK: Carlin, B.P. and Louis, T.A. (2000). Bayes and Empirical Bayes Methods for Data Analysis, 2nd ed. Chapman and Hall/CRC Press. ISBN: 1584881704.

## CE_13C 8:30 am–5:00 pm

### Generalized Linear Mixed Models: Theory and Applications

The ASA

***Instructor(s):*** Oliver Schabenberger, SAS Institute, Inc.

This course will discuss the theory of generalized linear mixed models and their application. It will cover the requisite theory of generalized linear mixed models and examples and applications. Additional theoretical detail will be added as needed. The first part of the workshop will make the connection between linear models (LM), generalized linear models (GLM), linear mixed models (LMM), and generalized linear mixed models (GLMM) in terms of model formulation, distributional properties, and approaches to estimation. We will discuss the pros and cons of various estimation approaches and describe their implementation with SAS/STATÆ software. The first part will end with a discussion of over-arching issues the analyst must confront when working with correlated, nonnormal data. The second part will use a variety of examples to revisit the theory taught in the first part, develop new insights, and present applications from different disciplines. Applications will range from modeling rates, proportions, and counts with random effects to GEE-type marginal models for nonnormal data to mixed model smoothing. A final section will describe inferential procedures following parameter estimation. Computations will be based on the mixed model tools in SAS/STATÆ software, primarily the GLIMMIX and NLMIXED procedures.

# August 7, 2006

## CE_14C 8:00 am–12:00 pm

### Spatial Survey Design with a Focus on Natural Resources

The Section on Statistics and the Environment

***Instructor(s):*** Anthony R. Olsen, U.S. Environmental Protection Agency; Donald L. Stevens, Jr, Oregon State University

This course will present a unified strategy for selecting spatially balanced probability samples of natural resources and other populations that exist in space. It will show how the strategy accommodates unequal probability sampling and problems that arise from frame imperfections and nonresponse due to inaccessibility. A library for the R statistical language will be used to illustrate survey designs applied to finite, linear, and real populations. An accompanying local variance estimator will be introduced and its computation illustrated.

## CE_15C 8:00 am–12:00 pm

### Analysis of Multivariate Failure Time Data

The Section on Nonparametric Statistics

***Instructor(s):*** Danyu Lin, The University of North Carolina at Chapel Hill

Multivariate failure time data arise when each study subject can experience multiple events or when there exists clustering of subjects such that failure times within the same cluster are correlated. Major complications in analyzing such data include the dependence among related failure times and censoring due to limited follow-up or competing events. This course presents a variety of statistical models and methods for the analysis of these data. We discuss marginal and frailty models, paying primary attention to semiparametric regression methods. Relevant software will be described and a number of clinical and epidemiologic studies will be provided.

**CE_16C 8:30 am–5:00 pm**

## Multiple Comparisons and Multiple Tests

The ASA

***Instructor(s):*** Peter Westfall, Texas Tech University

**2005 Excellence-in-CE Award Winner**

This course will cover multiple comparisons and multiple testing methods, from basic and state-of-the practice to advanced and state-of-the-art. Topics include "family" choice, combined versus component tests, interval versus test-based inference, Intersection-Union versus Union-Intersection tests, False Discovery Rate versus Familywise Error Rate, and Bayes/frequentist views. Models include standard ANOVA and multivariate analysis, as well as more general models used in survival analysis, mixed models, and nonparametrics. Emphasis will be given to closed and stepwise testing procedures, interval-based inference, and power analysis; with specific presentation of Holm, Hochberg, Hommel, Fisher combination, Westfall-Young, Dunnett, Lehmacher, fixed sequence, and gatekeeping methods. Applications include clinical trials, agriculture, business, and genetics.

**CE_17C 8:30 am–5:00 pm**

## Statistical Demography with Applications

The Social Statistics Section, Section on Government Statistics, Section on Survey Research Methods

***Instructor(s):*** Bruce D. Spencer, Northwestern University; Juha M. Alho, University of Joensuu

The course will discuss central concepts of demography from the viewpoint of statistical theory. Topics covered will include data sources, classical mathematical demography and its relation to pension funding and aging, stochastic demographic forecasting and its applications, and census error and its effects. Technical details will be avoided, but material has been selected so heuristic arguments can be presented. Particular attention will be paid to variability in populations and demographic processes. The course will show how stochastic forecasts can be produced using simulation methods, and the mechanics will be illustrated using free software developed by the instructors. The course will demonstrate that the volatility of demographic processes is commonly underestimated. No prior knowledge of demography is required. Familiarity with basic statistical modeling is assumed. RECOMMENDED TEXTBOOK: Alho, J.M. and Spencer, B.D. (2005). Statistical Demography and Forecasting. Springer. ISBN: 0-387-22538-2.

**CE_18C 8:30 am–5:00 pm**

## Analysis of Clinical Trials: Theory and Applications

The ASA

***Instructor(s):*** Christy Chuang-Stein, Pfizer Inc.; Alex Dmitrienko, Eli Lilly and Company; Geert Molenberghs, Limburgs Universitair Centrum

***2005 Excellence-in-CE Award Winner*** This course will cover analysis of stratified data, multiple comparisons and multiple endpoints, interim analysis and interim data monitoring, and analysis of incomplete data, offering a balanced mix of theory and applications. It will present practical advice from experts and discuss regulatory considerations. The discussed statistical methods will be implemented using SAS software, and clinical trial examples will be used to illustrate statistical methods. The course was designed for statisticians working in the pharmaceutical or biotechnology industries and contract research organizations. It is equally beneficial to statisticians working in institutions that deliver health care and government branches that conduct health care-related research. Attendees are required to have basic knowledge of clinical trials; familiarity with drug development is highly desirable. RECOMMENDED TEXTBOOK: Dmitrienko, Alex, Molenberghs, Geert, Chuang-Stein, Christy, and Offen, Warren (2005). Analysis of Clinical Trials Using SAS: A Practical Guide. SAS Publishing. ISBN: 1-59047-504-6.

**CE_19C 8:30 am–5:00 pm**

## Applied Bayesian Nonparametric Modeling

The Section on Bayesian Statistical Science

***Instructor(s):*** Alan E. Gelfand, Duke University; Athanasios Kottas, University of California, Santa Cruz

This course will provide an introduction to Bayesian nonparametric methods with emphasis on modeling approaches employing nonparametric mixtures and with a focus on applications. The course will start by motivating Bayesian nonparametric modeling and providing an overview of nonparametric prior models for spaces of random functions. The focus will be on models based on the Dirichlet process. Particular emphasis will be placed on Dirichlet process mixtures, which provide a flexible framework for nonparametric modeling. We will discuss methodological details, computational techniques for posterior inference, recent extensions to modeling for dependent distributions, and applications of Dirichlet process mixture models. Examples will be drawn from fields such as density estimation, quantile regression, hierarchical generalized linear models, survival analysis, and spatial statistics. The course targets students or professionals with background in Bayesian modeling and inference. Sufficient preparation will include statistics training to the MS level and some exposure to Bayesian hierarchical modeling and computation.

**CE_20C 8:30 am–5:00 pm**

## Generalized Linear and Latent Mixed Models

The Biometrics Section

***Instructor(s):*** Sophia Rabe-Hesketh, University of California, Berkeley; Anders Skrondal, London School of Economics

Generalized linear mixed (or multilevel) models (GLMMs) are useful for longitudinal data, cluster-randomized trials, surveys with cluster-sampling, genetic studies, metaanalysis, etc. The random effects in GLMMs are latent variables representing between-cluster variability and inducing within-cluster dependence. Latent variables also are used often to represent true values of variables measured with error. Measurement models specifying the relationship between measured and latent variables can form part of regression models, giving structural equation models (SEMs). SEMs also can be used to model dependence between processes. Taking a unified view of these models is beneficial because developments for one model-type are applicable to other model-types and the same software can be used for seemingly different models. The course will be structured in three parts: GLMMs, measurement models, and SEMs. Ideas will be developed by starting from simple versions of the models and motivating extensions through examples. Methods of estimation and prediction will be surveyed. Real applications will be considered from different disciplines. RECOMMENDED TEXTBOOK: Skrondal, A. and Rabe-Hasketh, S. (2004): Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Chapman and Hall/CRC Press. ISBN: 1-58488-000-7.

**CE_21C 1:00 pm–5:00 pm**

### Creating More Effective Graphs

The Section on Statistical Graphics, Section on Statistical Education

***Instructor(s):*** Naomi B. Robbins, NBR

This course will describe how to draw clear, concise, accurate graphs that are easier to understand than many of the graphs one sees today. It also will help readers of graphs to read more critically and analytically. The course will emphasize common mistakes that produce confusing-even misleading-graphs and how to avoid them. Graphs for one, two, three, and many variables will be covered as well as general principles for creating effective graphs. We will include a detailed discussion of scales, whether zero must be included, and when to use logarithmic scales. The importance of knowing your audience and choosing appropriate presentations for that audience also will be stressed. This course will cover principles and strategies for selecting the most appropriate graphical displays for communicating quantitative information, but not the specifics for implementing these strategies in various software packages. RECOMMENDED TEXTBOOK: Robbins, N.B. (2004): Creating More Effective Graphs. John Wiley and Sons. ISBN: 0-471-27402-X.

## August 8, 2006

**CE_23C 8:00 am–12:00 pm**

### Adaptive Tests of Significance and Confidence Intervals

The ASA

***Instructor(s):*** Thomas W. O'Gorman, Northern Illinois University

This session will present several methods of performing adaptive tests of significance and adaptive confidence intervals. For each adaptive test, we will compare its performance to the traditional method and show how to perform the test using a SAS macro. An adaptive test for paired data also will be presented. In addition, we will describe a method of computing adaptive confidence intervals. Attendees should be familiar with basic statistical modeling, including multiple regression and the analysis of variance, at the level of Applied Regression Analysis (1998) by Draper and Smith. RECOMMENDED TEXTBOOK: O'Gorman, T.W. (2003): Applied Adaptive Statistical Methods: Tests of Significance. ASA-SIAM. ISBN 0-89871-553-9.

**CE_24C 8:30 am–5:00 pm**

### Bayesian Analysis of Case-Control Data

The Section on Bayesian Statistical Science

***Instructor(s):*** Malay Ghosh, University of Florida; Bhramar Mukherjee, University of Florida; Samiran Sinha, Texas A&M University

In this course, we will try to present the current state of art in Bayesian methods for analyzing case-control data and illustrate each method with diverse applications. We will begin with simple models, and subsequently delve into more complex models involving missingness, measurement errors, and stratification issues.

**CE_25C 8:30 am–5:00 pm**

### Analysis of Environmental Data with Nondetects

The Section on Statistics and the Environment

***Instructor(s):*** Dennis R. Helsel, U.S. Geological Survey; Lopaka Lee, U.S. Geological Survey

This one-day course will survey the use of survival/reliability analysis and nonparametric techniques for the analysis of censored environmental data. Nonparametric and maximum likelihood methods for calculating descriptive statistics, testing hypothesis, and computing regression equations will be illustrated using the freely-available R statistical environment. Methods include two versions of estimating a Kendall-Theil-Sen line for censored data, applicable (unlike MLE) to doubly-censored data. The course covers material from the NADA contributed package for R. RECOMMENDED TEXTBOOK: Helsel, D.R. (2005): Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley and Sons. ISBN: 0-471-67173-8.

**CE_26C 8:30 am–5:00 pm**

### Models for Discrete Repeated Measures

The Section on Statistics in Epidemiology

***Instructor(s):*** Geert Verbeke, K.U. Leuven; Geert Molenberghs, Limburgs Universitair Centrum

Starting from a brief introduction on the linear mixed model for continuous longitudinal data, extensions will be formulated to model outcomes of a categorical nature, including counts and binary data. Several families of models will be discussed and compared, from both an interpretational and computational point of view. Models will be discussed for the full marginal distribution of the outcome vector. Then, random-effects models and semi-parametric marginal models with be discussed, with specification of the first moments only or the first and second moments only. Estimation and inference will be discussed and illustrated in detail, and it will be argued that both approaches yield parameters with different interpretations. Finally, it will be shown that, if no appropriate measures are taken, missing data can cause seriously biased results and interpretational difficulties. Methods to analyze incomplete data properly, under flexible assumptions, will be presented and key concepts of sensitivity analysis will be introduced. RECOMMENDED TEXTBOOK: Molenberghs, G. and Verbeke, G. (2005). Models for Discrete Longitudinal Data. Springer. ISBN 0-387-25144-8.

**CE_27C 8:30 am–5:00 pm**

### Effective Scientific Writing

The ASA

***Instructor(s):*** Judith A. Swan, Princeton University

This course will explore scientific writing from its readers' perspective. Research in many fields has demonstrated that readers of English derive most of their clues for interpretation from the structural locations of those words. This course will develop the readers' perspective on writing at all levels of structure, from sentences to full documents, identifying in the process how many distinctive aspects of scientific writing function to present the detailed and persuasive arguments of scientific thinking. The results are not merely cosmetic, but substantive. Better writing leads to more successful statistics and science.

## CE_28C 8:30 am–5:00 pm

### Bioequivalence and Statistics in Clinical Pharmacology

The ASA

***Instructor(s):*** Scott Patterson, GlaxoSmithKline; Byron Jones, Pfizer Inc.

This course will cover the application and basic elements of the theory of statistical methods in clinical pharmacology. The morning will deal with techniques used in the assessment of bioequivalence---the study of a drug formulation to confirm its equivalence to another. The afternoon will focus on the use of statistics in clinical pharmacology studies of safety, ECG monitoring, efficacy, and population pharmacokinetics. The emphasis will be on study design, analysis, and interpretation of data using real-data examples from the authors' experiences. RECOMMENDED TEXTBOOK: Patterson, S. and Jones, B. (2005): Bioequivalence and Statistics in Clinical Pharmacology. CRC Press. ISBN: 1584885300.

## CE_29C 8:30 am–5:00 pm

### Methods and Computational Tools for the Screening and Classification of Microarray Gene Expression Data

The Biopharmaceutical Section

***Instructor(s):*** Geoff McLachlan, University of Queensland; Kim-Anh Do, M. D. Anderson Cancer Center

This course is aimed at statisticians/biostatisticians/bioinformaticians and investigators working in areas in which relevant use can be made of microarray gene expression data. Although it is anticipated many in the audience will have majored in statistics, only a working knowledge of statistics is expected. Key features include a demonstration of methods (some with a Bayesian perspective) through the reporting of several case studies and newly developed methodology---in particular, that of the presenters---which may not be available in publication form at the time of the course. RECOMMENDED TEXTBOOK: McLachlan, G.J., Do, K. and Ambroise, C. (2004). Analyzing Microarray Gene Expression Data. John Wiley and Sons. ISBN: 0-471-22616-5.

## CE_30C 1:00 pm–5:00 pm

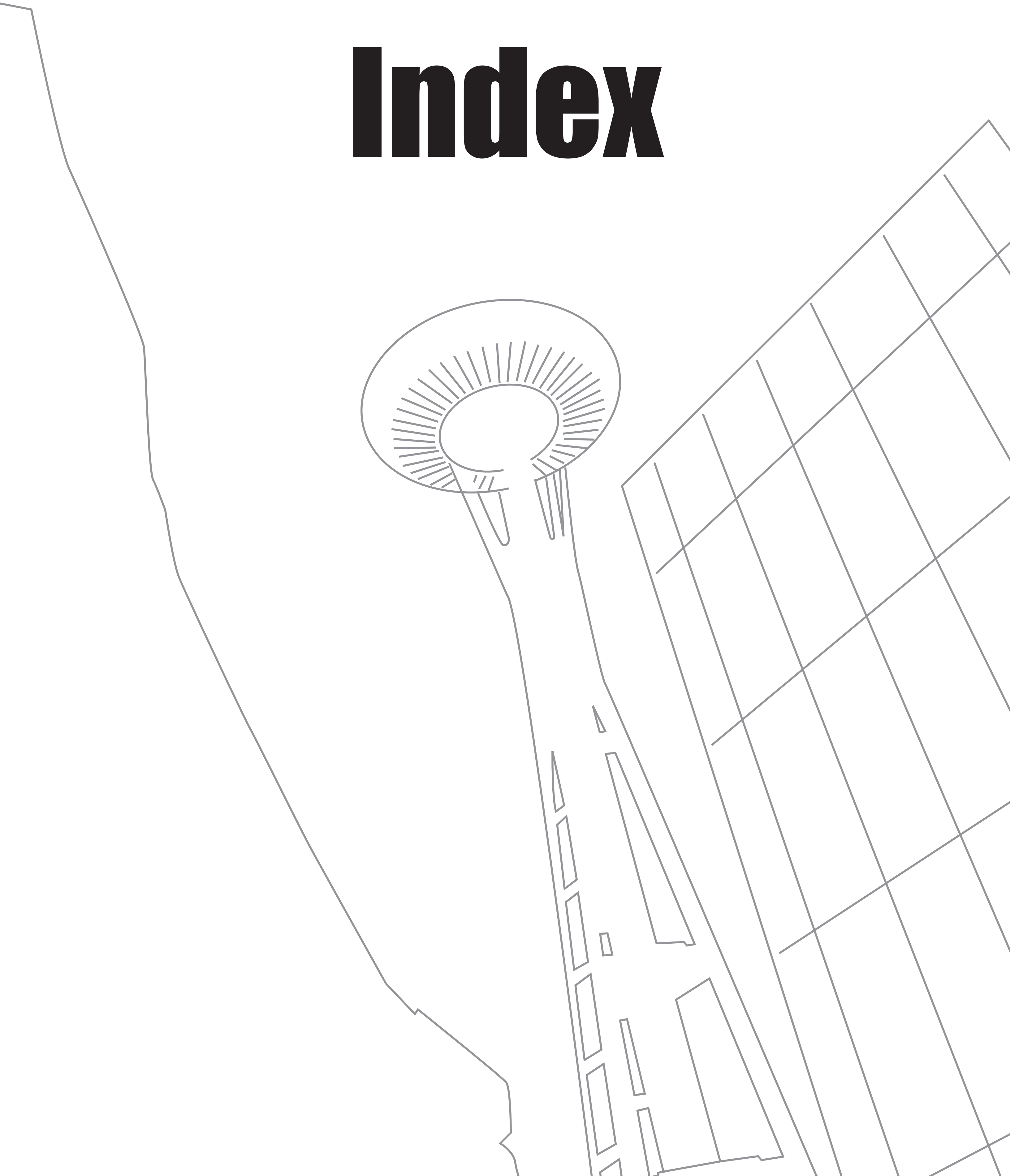
### Meta-analysis: Statistical Methods for Combining the Results of Independent Studies

The ASA

***Instructor(s):*** Ingram Olkin, Stanford University

In this course, we will provide a historical perspective of metaanalysis and discuss issues such as types of bias and the effects of heterogeneity. The statistical methodology will include discussions of nonparametric and parametric models, effect sizes for proportions, fixed versus random effects, and regression and ANOVA models.

# Index



# INDEX OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# INDEX OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# **INDEX** OF PARTICIPANTS

# **INDEX** OF CE INSTRUCTORS

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes

# Notes