The Statistics Teacher Network

Number 33 — ASA/NCTM Joint Committee on the Curriculum in Statistics and Probability — Spring 1993

Learning Opportunity

CHANCE Workshop Heads for San Francisco in August

You are invited to attend a workshop on a course called CHANCE at the summer meetings of the American Statistical Association in San Francisco. The workshop will be held Sunday, August 8, 1993, from 5:00 to 7:00 p.m. (room to be announced later). There is no charge for the workshop.

CHANCE is a case study course based upon current events in the news that involves statistics and probability such as: the use of DNA fingerprinting in the courts, statistical problems in AIDS, current clinical trials, quality control, SAT scores as predictors of college performance, etc.

CHANCE has been taught as an elementary college course by Tom Moore at Grinnell, Bill Peterson at Middlebury, Chris Thron at King College, and by myself and John Finn at Dartmouth, and Peter Doyle and myself at Princeton and UCSD. John Garfield has worked on teaching and assessment techniques related to the CHANCE course. Members from this group will share their experiences at the workshop.

For more information about CHANCE see "A Course Called CHANCE" in *CHANCE* magazine Vol 5, No 3-4 published by ASA and Springer. Those who use gopher may also obtain additional information by pointing their gopher to chance.dartmouth.edu.

If you would like to attend the workshop or would like more information about the CHANCE course please write to me by June 6, 1993 (e-mail is fine) at: J. Laurie Snell, Department of Mathematics, Dartmouth College, Bradley 6188, Hanover, NH 03755 or at JLSnell@dartmouth.edu

—J. Laurie Snell
Dartmouth College
Hanover, New Hampshire

In this Issue...

Classroom Tools

Simulation Techniques Using the HP 48

Simulation is the process of representing an experiment with a model. The simulation technique has the advantage over the actual experiment in that many identical repetitions can be performed quite efficiently with the aid of a computer or in this case, your calculator. Once simulations are performed, you can compare the outcomes of a large number of trials to the "theoretical" results. Simulation techniques usually involve *random numbers*—numbers chosen in such a way that each one is equally likely to be the one selected. Most statistics texts include a table of random numbers.

The HP 48 has its own built-in program for generating pseudo-random numbers on the

From the Editor

Homework for QL Graduates

I have some summer homework for all of you but especially for QL graduates, i.e., those hundreds of you across the country who have been through a Quantitative Literacy workshop. Your (voluntary) assignment is to write a short article (max 1,200 words) on some neat classroom statistical activity that you and your students have done. Trust me, you DO have something to share with your colleagues. You may not think you do, but we are all learning what works and what doesn't work in the classroom. The last couple issues of STN have had some good articles from California and Ohio. I know that equally exciting things are being done elsewhere—Virginia, Kansas, Florida, Pennsylvania to name only a few. So, come on folks, inundate me with articles to use in future issues! By the way, the deadline for the fall issue is August 15. Please send a hard copy and, if possible, a disk written in standard ASCII code. Thanks for your contribution and sharing.

Have a very enjoyable and educational summer. See you in the fall.

—Jerry Moreno, Editor
Statistics Teacher Network

interval [0,1). A "true" random number generator on the interval [0,1) would select each *real* number in that interval equally as often. Since it is impossible to simulate with the precision of real numbers, random number generators in calculators and computers generate pseudo-random numbers—that is, random numbers that are obtained to a fixed number of decimal places. These random outcomes look and behave for the most part like theoretical random numbers. There are $10^{12} - 1$ pseudo-random numbers on the interval [0,1) which may be obtained with the HP 48's random number generating function RAND. Access this function in the MATH PROB menu. Press **RAND** several times and you will see some calculator-generated random numbers between 0 and 1.

The first random number that is generated is dependent on the value stored in the calculator memory. You may "seed" the random number generator (simulate randomly choosing a position in a table of random numbers) by entering any nonnegative number and pressing **RDZ**. You can repeat a particular series of random numbers by executing RDZ with the same argument.

To simulate the experiment of tossing one fair coin, let the outcome "tails" be represented by 0 and the outcome "heads" be denoted by the number 1. Since RAND yields a random number x such that $0 \le x < 1$, we have $0 \le 2x < 2$. Notice also that whenever $0 \le x < 0.5$, we have $0 \le 2x < 1$ and that whenever $0.5 \le x < 1$, it is true that $1 \le 2x < 2$. A command you will use with the random number generator is IP which is found in the MATH PARTS menu. IP returns the integer portion of a number, and entering RAND 2 * IP on the stack will return either a 0 or a 1 representing an outcome for the experiment of tossing of a fair coin.

Let's now look at a program that will allow the calculator do the coin tossing for you. Store this program with the name COINTOSS.

```
<<1 RES  ERASE  0 → n
  <<{ n } MENU HALT CLΣ  n .8  *  DUP
  -.2  * SWAP  YRNG  1 n  FOR  n
  RAND  2  *   IP  Σ+ 0  2  XRNG  HIS-
  TOGRAM  DRAW    NEXT >>
  GRAPH  0  RES  FUNCTION  0  MENU
  { ΣPAR PPAR  CST }  PURGE >>
```

When this program is executed, you will see a menu containing the letter N. It is important to note that the number of tosses, n, must be entered in the program as a local variable (that is, using a lowercase letter obtained with α [↰] N). A local variable appears on the temporary menu as a capital letter.

To run this program, enter the value of n, the number of times you wish to toss the coin, and press [↰] **N** to store the value as n. Notice the HALT message at the top of the display screen. Whenever HALT appears in this position, you are to press [↰] CONT after the proper input to continue the program. Use program COINTOSS to flip your calculator coin 10 times, 25 times, and then 50 times.For each experiment, record the number of heads (1's – the rightmost bar) and determine the relative frequency of the event (heads). What value do the relative frequencies approach as the value of n increases?

Let's consider another simulation which this author calls the "histogram horse races." Enter the two programs in Table 1.

When you execute each of these programs, you will see a histogram of 6 rectangles is drawn one vertical block at a time. Think of each rectangle in the histogram as the moves for one horse on the track. Thus, you are watching a race for 6 horses. Think of each block in the rectangles as a horse moving one unit. The total number of units moved by all the horses is N.

To execute either of the above programs, press the menu key in whose name you have stored the program. Load the value of N by entering the numerical value on the stack and pressing [↰] N. Press [↰] CONT to continue the program and run the 6 horse race. Good luck on picking the winner!

Exploration Exercises

1. Execute program COINTOSS for several different values of N, each time recording whether you get more tails (0's) or heads (1's). Each time tails "wins", score 1 for T. Each time heads "wins", score 1 for H. If it is a tie, each receives 1/2 of a point. What would you expect as the final result if you play this game many times? Do you think your conjecture depends on N, the number of times the coin is tossed?

2. Refer to programs COINRACE and DIERACE.
   a) How many coins are being tossed in the COINRACE program?
   b) If x is a random variable representing the number of heads obtained in the toss of these coins, what are the possible values for x?
   c) How many dice are being rolled in the DIERACE program?
   d) If y is the random variable representing the number of dots appearing on the upturned face of a die, what are the possible values for y?
   e) What is the relationship of x to y?
   f) Run each of these two programs using N = 10, 25, 40, and 55. Note for each the shape of the histogram as the value of N increases. Repeat if necessary. If you do not see a definite pattern emerging as the value of N gets larger, run the programs for N = 75 and N = 100. (This may take a while!)
   g) What theorem in statistics explains the different shapes of the histograms obtained from these two programs?

Further Reading

Fetta, I., *Calculator Enhancement for Introductory Statistics: A Manual of Applications Using the Sharp EL-5200, HP-28S and HP48S Graphing Calculators*, Saunders College Publishing, 1992, Philadelphia, PA.

Fetta, I., *Explorations in Algebra, Precalculus and Statistics: A Manual for the TI-81 Graphing Calculator*, Saunders College Publishing, Philadelphia, PA, 1992.

Fetta, I., "Explorations in Probability and Statistics," *Explorations with the Texas Instruments TI-85*, Harvey, J. and J. Kenelly (ed.), Academic Press, Boston, MA, 1992.

—Iris Brann Fetta
Clemson University

Table 1. "The Histogram Horse Race" Programs

COINRACE

```
<< 1 RES ERASE 0 → n
<< { n } MENU HALT
      CLΣ n .8 * DUP
      -.2 * SWAP YRNG
      0 6 XRNG 1 n FOR k
      RAND 2 * IP RAND
      2 * IP RAND 2 * IP
      RAND 2 * IP RAND
      2 * IP + + + + Σ+
      HISTOGRAM DRAW
      NEXT >>
      GRAPH 0 RES
      FUNCTION 2 MENU >>
```

DIERACE

```
<< 1 RES ERASE 0 → n
   << { n } MENU HALT
      CLΣ n .6 * DUP
      -.2 * SWAP YRNG
      1 7 XRNG 1 n
      FOR k RAND 6 *
      1 + IP Σ+
      HISTOGRAM DRAW
      NEXT >>
      GRAPH 0 RES
      FUNCTION 2 MENU
      >>
   >>
```

Getting More Mileage Out of Histograms

Many introductory statistics courses begin with an emphasis on graphical displays of data such as histograms, stem and leaf plots, and box plots. I agree that these topics comprise a vital foundation of data exploration skills that can be used throughout a first statistics course, as well as in subsequent courses, but I find that students view these topics as being easy. Many of my students have at least an intuitive understanding of means, medians, and other basic statistics, so the only new material they encounter in the first few weeks of the course involves learning the mechanics of the construction of graphical displays. The students fail to budget enough of their study time for their statistics course, and when they encounter topics such as probability and statistical inference, they become bogged down in material they perceive to be harder. In the middle of a semester it is difficult for students to reallocate their resources to meet the challenge of the more mathematical material. The purpose of this article is to provide an example of the application of histograms that can be used to push students beyond the mechanics of graphical displays and into the processing and interpretation of real data. Incorporating similar examples into lectures and homework makes it clear that the most important concepts are not contained in algorithms for the construction of graphical displays, but rather involve the integrity of data and the interpretation of the results of an analysis.

I teach an introductory statistics course at Villanova University to a class made up primarily of freshmen liberal arts majors. During the first class meeting each semester, I issue a survey of the students. One of the questions I ask is "How far from Villanova is your hometown?" I intentionally do not ask "What is the distance in miles from Villanova to your hometown?" because I want to increase the chance of ambiguities in the data. Questions that might result in similar responses include "How far away is your place of birth?" or "How far away do your grandparents live?"

After lecturing on the construction of graphical displays, and before I begin the formal coverage of measures of central location such as the mean and median, I summarize the results of the survey in a handout that I issue to the students. The handout resulting from my fall 1992 semester class is listed below. Notice that I provide the literal responses provided by the students, complete with ambiguities.

The following data represent the distances (in miles, except where noted) from Villanova to the homes of the students in Introductory Statistics I:

240 195 18 4500 100 150 4500 20 6
101 50 100 1500 140 1500 120 120 100

300-400 About 250 About 400

2 hours 2 hours 15 minutes

2.5 hours = 160 miles?

5 hours so about 300 miles

After issuing the handout at the beginning of a class period, I assign the students into groups of three and ask each group to solve the following problems:

1. Construct a histogram for these data.
2. Report the distance to the hometown of a typical Villanova student.

Immediately after the students glance at the data, at least one of them will raise a hand to ask "What should we do with the ones that are times?" Obviously, my answer is for them to decide what is reasonable. The integrity and uniformity of the data become the first topics of discussion within the groups.

Prior to this assignment, the students have been taught that 5 to 15 groups are usually present in a reasonable histogram. They quickly realize that the extreme values in the data make this guideline impractical. Including the outliers collapses the relatively large number of observations at the low end of the distance scale into one or two cells. The more creative groups find a way to display all of the data in an informative manner, but they must break the "rules" for the construction of a typical histogram.

Once the students have constructed a histogram, they usually calculate an average to represent the typical distance. This puts the hometown of a typical Villanova student somewhere in Indiana, which is clearly unrealistic since most of our students are from Pennsylvania, New Jersey, and the New England states.

General familiarity with the geographical distribution of the Villanova student body provides an introduction to the concept of resistance to extreme values.

There are many other statistical concepts that can be highlighted during discussion of this example. Mentions of measures of location and dispersion arise naturally, but topics such as outliers, transformations, misleading displays, and the communication of results can also be introduced through the construction of histograms for this and other real data sets. Examples such as this one emphasize to students that the basic statistical tools such as histograms and other displays are in fact easy to construct, but that the statistical evidence and resulting judgments are not always as clear nor are they necessarily universally agreed upon. Incorporating real data into class exercises and homework assignments not only helps to maintain student interest, but also emphasizes that the interpretation of statistical evidence involves careful thought and critical examination.

—Tom Short
Department of Mathematical Sciences
Villanova University

Learning Opportunity

TORCH Institutes

TORCH (Teacher OutReaCH) one-week institutes in statistics are administered by the Woodrow Wilson National Fellowship Foundation (WWNFF). During the summer of 1993, there will be three institutes for middle and secondary school teachers. They are conducted by a team of four experienced teachers who participated in a four-week Summer Mathematics Institute held at Princeton University under the direction of WWNFF.

As stated in the brochure: "Participants use calculators, computers and user-friendly software as aids to learning and as tools for teaching statistics. They explore ways to incorporate statistical ideas into the traditional mathematics curriculum at various levels, with an emphasis on methods and materials readily available in the classroom."

To participate in this institute, you need not be an experienced teacher of statistics. The workshop provides a meaningful and substantive experience for all mathematics teachers—whether or not they have a statistics background.

Listed are the "where, when, and who to contact" for the three institutes:

- Western Kentucky University, July 12-16, Dr. Robert Bueker, 502-745-3651.
- Amarillo College (Texas), July 19-23, Ms. Therese Jones, 806-371-5091.
- Stephens Institute of Technology (New Jersey), July 26-30, Dr. Lawrence Levine, 201-216-5448.

For information about the many other WWNFF institutes in science, mathematics, and history, please call the Foundation at (609) 452-7007.

—Janet M. Gnall, Ph.D.
Program Coordinator

Assessment Roundtable

Assessment of Statistical Learning Roundtable Planned

Last November a two-day workshop on Assessment of Statistical Learning was held at the University of Pennsylvania, sponsored by the National Center of Adult Literacy and the National Science Foundation. A dozen participants representing the fields of mathematics education, statistics, science education, and educational assessment, explored the unique aspects of assessing students' statistical knowledge and outlined further work needed in this area. A follow-up roundtable on the same topic is planned for November 1993, to explore these issues in more depth. Several commissioned papers will be presented and discussed, and results of current assessment efforts connected to funded curriculum projects will be presented.

Anyone who is interested in more information on this conference should contact co-organizers Iddo Gal (NCAL, 3910 Chestnut St., Philadelphia, PA 19104-3111; e-mail: gal@a1.relay. upenn.edu) or Joan Garfield (140 Appleby Hall, 128 Pleasant St. S.E., Minneapolis, MN 55455, e-mail: JBG@vx.cis.umn.edu).

Call for Workshop Sites

Apply Now to Host a QL Workshop on the High School Curriculum

During the next academic year, several locations will be given the special opportunity to host a workshop for the Quantitative Literacy Project "A Data-Driven Curriculum Strand for High School Mathematics," which is a joint project of the National Council of Teachers (NCTM) of Mathematics and the American Statistical Association (ASA). This is the third such cooperative effort; the fourth Quantitative Literacy Project (QLP)—on elementary education—is also underway.

The first QLP, funded by the National Science Foundation (NSF), developed four booklets "Exploring Data," Exploring Probability," "The Art and Techniques of Simulation" and "Exploring Samples and Surveys," to teach students about important ideas in statistics. The second Project created a model workshop designed to help teachers implement the ideas presented in those materials in their classrooms and held a series of workshops at various sites around the United States. This third QLP, funded by NSF in 1991, addresses two complementary needs: how to use data analysis to motivate some of the essential topics of a restructured mathematics curriculum and how to identify and teach those data analysis skills that are required for effective participation in society.

The goals of the project include identifying which topics in data analysis are appropriate for a high school curriculum and how to integrate these topics into the teaching of traditional content in mathematics so the content is motivated from real applications of interest to students, and the statistical concepts form a unified presentation of the subject. Technology, an integral part of several of the modules, is used as a tool to solve problems and as a means to investigate and explore situations to advance student understanding. The modules are constructed so the students must assume responsibility for reading, investigating, and interacting with each other and the teacher as they develop strategies to resolve problems, many of which are open ended. There are strong emphases on explaining, justifying, and communicating in both written and oral forms.

Three modules are currently being field tested. "Our Numerical World" provides an introduction to using numerical data to describe phenomena in the world and to the statistical concepts of bias, variability, and correlation as well as simple statistical measures and graphs. In "Symbolic Expressions," students explore the notions of variable, expressions, formulas and function as they look at plots over time, compare data sets, and find expected values. "Linear Relations" includes slope as a rate of change, linear equations and inequalities, systems of equations, and an informal introduction to correlation and residuals. Four more modules are in draft stages: "Exploring Equations" introduces quadratic and absolute value functions through the development of the least squares regression; "Centers" explores centers from both an arithmetic and geometric perspective; "Projects and Surveys" has suggestions for assigning and evaluating student projects; "Nonlinear Models" provides experiences with modeling relationships, prediction, and error analysis. Tentative plans include modules on matrices, probabilistic modeling and sequences and series.

As part of the grant, five regional two- to three-day workshops will be held during the 1993-94 school year to introduce the modules to teachers. If you are interested in arranging for such a workshop to be held in your area, please send an invitation to the project.

Your invitation should include the following:

- suggested location for workshop,
- tentative dates,
- name of person who would be willing to serve as local facilitator,
- general description of possible participants, and
- a short rationale for why your site should be selected.

Please send correspondence to: Gail Burrill, Quantitative Literacy, Whitnall High School, 5000 S. 116 Street, Greenfield, Wisconsin 53228.

—Gail Burrill
Greenfield, Wisconsin

Statistics in the Classroom

Two Exercises for Students

• *An introduction to hypothesis testing*

Mars, Inc, the makers of M&Ms candies, claims that their plain M&Ms are "30% brown, 20% yellow, 20% red, 10% orange, 10% green, and 10% tan...[and] each large production batch is blended precisely to those ratios and mixed thoroughly." I have my students test this claim in the classroom by distributing "fun-sized" bags of plain M&M candies and having students count colors. Once the color counts are obtained for the entire class we evaluate the goodness of fit statistic for our counts. To decide if this measure of deviation from what was expected is too large for the claim of Mars to be reasonably true, I then give a computer demonstration in which a p-value is estimated. In particular, the computer takes repeated samples (of the same size as the students) from a population having the claimed color distribution, each time computing the associated goodness of fit statistic. By comparing the goodness of fit statistic for the M&Ms counted with the sorted listing of those values generated by simulation a p-value is estimated. Each time I have conducted this test in the classroom I have found a very small p-value. Thus, it would appear that the color distribution above is not correct or that production batches are not as thoroughly mixed as claimed (preventing us from getting a random sample).

The claimed color distribution for peanut M&Ms, by the way, is 20% each for brown, yellow, red, orange, green (there are no tan peanut M&Ms). More details on the above classroom example appear in the article "Testing color proportions of M&Ms" in the February 1993 issue of *Teaching Statistics*.

• *A multiple regression example*

The sports section of many newspapers during the NFL football season contains a ranking of quarterbacks passing ability with a rating number. Occasionally, there will be a footnote to this table explaining that rating is computed from the percentage of completions, percentage of touchdown passes, percentage of interceptions, and average gain per pass attempt, but the particular form of this function is not publicized. Upon contacting the NFL office I was provided extensive tables in which one could look up values for each of the four categories above and total to determine the rating. Generally speaking, one can infer from these tables that

$$\text{Rating} = [\, 25 + 10(\text{Completion }\%) + 40(\text{TD }\%) - 50(\text{Interception }\%) + 50(\text{Yards/Attempt}) \,]/12$$

This assumes the completion percentage is between 30% and 77.5%, TD % is between 0% and 11.875%, interception percentage is between 0% and 9.5%, and average gain per attempt is between 3 yards and 12.5 yards.

Using completions, touchdown passes, interceptions, yards gained, attempts and rating statistics for several players then, one can uncover the above relationship using regression. I have found "all-time" quarterback rankings, which are updated annually in almanacs, of special interest to some students.

A similar regression I'm having my students try for the first time this term is to go to the grocery store and record calorie, carbohydrate, protein, and fat information per serving on a number of food items and then, using regression, have them tell me how to predict calories from the other three items. Note that after the new FDA labeling act goes into effect, this won't be as much fun because labels will be required to indicate the number of calories per gram of fat, carbohydrate, and protein.

—Roger Johnson
Dept. of Mathematics
and Computer Science
Carleton College

(Editor's note: The multiple regression example is tentatively planned to appear in the fall issue of *The College Mathematics Journal*.)

Book Review

Statistics for the Twenty-First Century

Sheldon and Florence Gordon, Editors
Available from The MAA, 1529 Eighteenth St. NW, Washington, DC 20036 ($22.00 + $1.90 postage and handling)

Statistics for the Twenty-First Century is required reading for anyone involved in statistical education. This excellent book (Number 26 in the MAA Notes series) is a collection of papers by leading statistics educators who address the problem of how to improve statistical education. The book is divided into four parts: Issues in Statistical Education; Innovative Curricula for Statistical Education; Technology in Statistical Education; and Resources for Statistical Education. The list of papers and authors is quite impressive, and the following brief summary does not do justice to the wealth of information in this book.

In the first section, there are some provocative position papers on what's wrong with statistics courses, what an introductory statistics course should look like, how a statistics course should differ from a mathematics course, the role of assessment in learning statistics, and viewpoints of faculty from different disciplines regarding the teaching of statistics.

In the second section, examples are given of particular approaches to teaching statistics (e.g., a nonparametric approach, or a course based on exploring real data). Other papers focus on activities that have been used successfully in statistics courses (e.g., student projects, and hands-on activities).

The third section explores the use of graphics, simulations, spreadsheets, specific software programs, and graphing calculators as they relate to teaching statistics. The final section on teaching resources includes bibliographies of teaching materials, background readings, sources of real-world data sets, and information on the different e-mail networks used for statistical discussions.

—Reviewed by Joan Garfield
University of Minnesota
Minneapolis, Minnesota

Keep Us Informed...

The Statistics Teacher Network is a newsletter published three times a year by the American Statistical Association—National Council of Teachers of Mathematics Joint Committee on the Curriculum in Statistics and Probability.

We need your letters, announcements, articles, and information about what is happening in statistics education! Please send hard copy, and, if possible, a disk written in standard ASCII text to the editor:

Jerry Moreno, Dept. of Mathematics, John Carroll University, University Heights, OH 44118

or **moreno@jcvaxa.jcu.edu**

or **fax: (216) 397-3033**

Layout & Design: *Alison Stern-Dunyak, American Statistical Association*

To be added to the mailing list or make an address change, please send your name and address to: *Statistics Teacher Network*, c/o American Statistical Association, 1429 Duke St., Alexandria, VA 22314-3402; (703) 684-1221; fax: (703) 684-2036; e-mail: alison@asa.mhs.compuserve.com

Statistics Teacher Network
c/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402

Let us know your Zip +4!

Non-Profit
Organization
U.S. Postage
PAID
Permit No. 361
Alexandria, VA

1 5 30
JERRY MORENO
DEPARTMENT OF MATHEMATICS
JOHN CARROLL UNIVERSITY
UNIVERSITY HTS., OH 44118