THE STATISTICS TEACHER NETWORK

No. 28

SEPTEMBER 1991

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.

A CLASSROOM EXAMPLE

As part of our unit on Linear Equations and Graphs, my Honors Algebra I class used some data based activities to discover the meaning and importance of slope, intercepts and the equation of a line. One of our activities had an interesting result. Each boy in the class was asked to measure his height and his father's height. Each girl was asked to measure her height and her mother's height. The data was collected from the students and each student graphed the data using the parent's height for the X coordinate and the student's height for the Y coordinate. The data for the girls were plotted separately from the boys' data.

Each student drew an "eyeball" fit line for the boys and one for the girls. The slope of each line was determined and the Y-intercept was calculated. Although the lines constructed for the girls' data varied somewhat, almost everyone agreed that the slope was close to one. In fact the equation of the best fit line, according to the TI-81 was $Y = 1.1X - 8$. The data for the boys yielded much more divergent results and no agreement could be reached as to the equation of the line. The TI-81 computed an equation of $Y = .09X + 61$.

The class discussed the difference between the two sets of data. I pointed out how the scatter plot for the girls "looked" linear–generally the taller the mother the taller the daughter. But the boys' scatter plot was "a mess" with no apparent pattern. I then pointed out the fallacy of trying to represent data that did not appear to be linear with a straight line.

The interesting result happened when we discussed "why the difference between girls and boys?" Immediately the students decided that most of the girls had completed much of their growth while others had much growing yet to do. They all knew that girls started growing before most boys did, but here the use of linear equations provided "graphical" evidence of this.

The Data

GIRLS

X	61	61	63	63	63	63	65	65	66	66
	66	67								
Y	56	63	57	62	63	63	66	60	67	66
	62	66								

BOYS

X	66	68	68	70	70	70	71	71	72	72
	72	72	73	73	74	74	75			
Y	68	63	69	71	71	68	64	73	68	73
	65	69	70	67	62	71	70			

-- Murray H. Siegel

IT COULD HAVE HAPPENED — SPORTS HEADLINE:
BRAVES MUST LOSE FINAL TWO GAMES TO QUALIFY FOR PLAYOFFS

Conclusions that are wrong or just incomprehensible are often the result of plain old-fashioned blunders. Rates and percentages are the most common cause of crooked arithmetic. Sometimes the matter can be straightened out by some numerical detective work.

— David S. Moore (Statistics: Concepts and Controversies)

This story is true: it took place a few years ago in major league baseball. It was a year when the players went on strike in the middle of the season. It was also the year when the famous Los Angeles Dodger infield of Garvey, Lopes, Russell, and Cey won their first and only World Series, beating the New York Yankees, 4 games to 2.

Because of the strike and the resulting loss of scheduled games, major league administrators decided that each of the four existing divisions would have a split season. The teams leading their respective divisions at the beginning of the strike would be declared champions for the first half of the season. When play resumed after the strike, the "second half season" would begin. The following guidelines were established for each division:

(1) If two different teams won a half-season, they would have a playoff for the division pennant.
(2) If the same team won both half-seasons, then this team would have a playoff for the division pennant with the team (other than itself) having the best overall winning percentage over both half-seasons. (This assured a money-generating playoff for each division pennant.)

These guidelines were accepted and published by the proper authorities. They certainly seemed reasonable — until a hypothetical situation similar to the one to be demonstrated was presented to league officials. Suppose the following situation presented itself:

First half season (Final standings)	W	L	Second half season (Two games remaining)	W	L
Dodgers	50	20	Dodgers	48	20
Braves	49	21	Astros	47	21
Astros	40	30	Braves	45	23
Padres	32	38	Reds	25	43
Giants	20	50	Giants	21	47
Reds	19	51	Padres	18	50

Assume now that the Braves are scheduled to play their final two games against the Dodgers. As the second half standings indicate, the Braves have no chance of being second half season champions since only two games remain to be played. If the Astros win the second half season, they will, according to the stated rules, play the Dodgers for the division championship and the Braves will be totally out of the championship picture. However, if the Dodgers win the second half season, the championship series will be between them and the team with the best overall record for the entire season. Even if the Braves lose their last two games, their overall record of 96 wins and 44 losses will be superior to the 89 wins and 51 losses the Astros will achieve if they win their last two games. The Braves would face this seemingly unbelievable situation: **If they lose their last two games, they automatically qualify for the division championship playoff. If they win one or two games against the Dodgers, they face the distinct possibility of being eliminated from the championship playoff.** Needless to say, major league administrators did an about-face on the original split-season guidelines that they published.

It was the best of times, it was the worst of times; it was the age of wisdom, it was the age of foolishness...

— Dickens (A Tale of Two Cities)

Indeed, it is commonplace to be deceived by rates and percentages. Consider, for example, the batting averages of two players during the split baseball season previously mentioned.

	Batting Average (First half season)	Batting Average (Second half season)
Brown	.333	.300
Jones	.320	.250

Which of the two individuals had the highest batting average for the entire baseball season? The answer would be obvious to most casual observers. But wait! Let's attach a few numbers to these percentages.

	First half season			Second half season		
	At Bat	Hits	Avg.	At Bat	Hits	Avg.
Brown	39	12	.333	300	90	.300
Jones	300	96	.320	40	10	.250

	Entire season		
	At Bat	Hits	Avg.
Brown	339	103	.304
Jones	340	106	.312

Hence, given only the half season batting averages, we cannot conclude that Brown's batting average for the entire season would be greater than Jones'. Indeed, our perceptions frequently do not match up with reality when it comes to dealing with rates and percentages.

> *Our society would be unimaginably different if the average person truly understood basic mathematical concepts.*
> — Douglas (Author of Godel, Escher, Bach)

-- Sanderson M. Smith

- -

We begin a new series in this issue titled "Statistics in My School". We want to show how instruction in statistics is implemented in various school settings.

More articles are needed for this series. If you wish to contribute please contact the editor.

-- Editor

STATISTICS IN MY SCHOOL

Columbus, Indiana, is the county seat of Bartholomew County whose population is around 70,000. One school corporation serves practically the entire county. Columbus is located about fifty miles south of Indianapolis and has a strong industrial base for employment, as well as being home to many fine farms.

Columbus East High School is one of two four-year high schools in the school system. We have about twelve hundred students who are required to pass two years of mathematics for graduation. The faculty consists of ten certified mathematics teachers with masters degrees.

The study of statistics in the various mathematics classes is somewhat hit or miss, depending upon who teaches the course. The general mathematics classes (ninth and tenth grade) do a unit on statistics. The students explore data through the use of various types of graphs including stem-and-leaf plots as well as the more traditional line graphs, circle graphs, and bar graphs. The textbook material is supplemented with the Exploring Data book from the quantitative literacy series published by Dale Seymour. Some of the classes pursue a study of probability using polyhedra dice and other manipulatives. They attempt some simulations on computer using the QL disk. The school has a computer lab of 30 Apple IIe's for the use of the math students. (There are also computer labs of MacIntosh and IBM in the tech prep, English and business departments.)

Advanced algebra classes also do a unit on probability and statistics. In addition to the topics covered by the general math classes, they study permutations and combinations.

Pre-Calculus classes study probability as it relates to the binomial distribution and Pascal's triangle when studying the binomial expansion.

Columbus East High School also offers a one semester course in probability and statistics. The nine period day allows for flexibility in scheduling and many college intending students enroll in two math courses. The statistics course is taught to about twenty (mostly) seniors and is fairly rigorous. The textbook is John Freund's Modern Elementary Statistics published by Prentice Hall. It is supplemented with quantitative literacy materials, especially the computer simulations. More recently, the text has been supplemented with material from Moore and McCabe's Introduction to the Practice of Statistics, published by Freeman, and the video tape series *Against All Odds* aired on PBS in the last couple of years and funded by the Annenberg/CBP Project. (I highly recommend this video tape series and hope to adopt Moore's book in the next textbook adoption.) The topics covered in the course include basic data analysis, (graphs, measures of center and spread, etc.) probability, probability distributions (normal, binomial, exponential, Poisson, geometric, hypergeometric, chi square, student's t) linear regression, correlation, confidence intervals and hypothesis testing. I have found the TI81 graphing calculator to be a valuable tool for this course.

More needs to be done in my school to look at probability in geometry classes, but we have a strong foundation. Incorporating new materials takes time – time to fit in and time to bend people's ways.

-- Linda Young

SOFTWARE REVIEW

Title: *Data Insights*
Publisher: Sumburst Communications, Inc. 101 Castleton St., Pleasantville, NY 10570
Price: $99.00 — This includes the program disk and a backup, a teacher's guide, and a quick reference card. There is an 8% charge for shipping and handling.
System Requirements: IBM PC with 256K and a color graphics adapter; IBM PS/2 (models 25 and 30) 256K RAM; Tandy 1000 with 256K RAM; Apple II with 128K RAM; DOS version 2.0 or later.

Data Insights is a tool for teaching and applying graphic techniques to analyze data. It includes six types of plots for viewing data sets: line plot, histogram, stem-and-leaf plot, box plot, scatterplot, and line graph. Descriptive statistics can be obtained for up to six columns and 500 rows of data. Data files may be saved on disk, and printed output is available for the data itself, the statistics, and the plots. Students can build their own data sets or use the sample data sets included in the program. These provide excellent examples of real world applications which are interesting and relevant to students' daily lives.

The grade level is 6 through 12 with some applications which can be used in grades 4 and 5. Students must be familiar with number lines and have elementary level graphing skills as a prerequisite to introducing these techniques. Use of this tool reinforces problem solving skills which include evaluation, pattern recognition, trial and error, analysis, generalization, and an introduction to the idea of hypothesis testing.

Data Insights is a user-friendly application which is flexible, menu driven, and includes a help option which provides the user with general descriptions of each command. It includes an instruction box with details about the current command and incorporates the safety net approach of "Esc"aping from all levels of the command structure.

This package is an effective mechanism for helping students visualize and interpret data and could be used as a demonstration tool, as a laboratory tool, or as a computer station in the classroom. In a demonstration setting, it may be used with an LCD display device for purposes of illustration during class discussions and to clarify student assignments and projects. In addition, it is adaptable to the computer lab setting where students are conducting data analysis individually or in teams. The program could also function effectively if available from a computer station in the classroom for use in long-term student projects.

Data Insights accomplished the objectives of teaching students about interpreting plots and characterizing data by the use of descriptive statistics without their having to spend a great deal of time on the mechanics of graphing and calculation. It enables students to:

1) view a graph of data quickly
2) use different types of plots to view the same data
3) visualize patterns and trends
4) obtain descriptive statistics without time consuming number crunching
5) plot two sets of data for purposes of comparison
6) change the scale of a graph

Some of the more noteworthy and useful features include flexibility of choosing quick or customized plots, emphasis on data exploration, and ease of data entry and editing. The program allows the students to customize graphs with titles, labels, and user determined scaling and interval choice and also to dictate which statistics will be displayed and how outliers will be handled.

In summary, *Data Insights* is an excellent tool for increasing student's skills in data analysis and plotting. It is particularly valuable because it reduces the classroom time needed to produce graphs, calculate descriptive statistics, and handle large data sets. It comes highly recommended by the large cadre of teachers who have been involved in the classroom testing and review.

— — Beth Bryan

Leslie V. Glickman — A Tribute

Readers will be saddened to hear that Leslie Glickman died suddenly on July 28th, just two days after we three had submitted the finished manuscript of *Teaching Statistical Concepts* to our publishers.

We shall miss Leslie both as a colleague and as a friend, and it is undoubtedly the case that the area of statistical education will also feel the loss. With the manuscript completed, he was full of plans for future projects; writing materials on probability for a new SMP (School Mathematics Project) 16-19 syllabus, researching and developing a new idea in statistical graphics, and further work on statistical concepts and intuitions with us.

He leaves a wife and an 18-year old daughter and *Teaching Statistical Concepts* will be dedicated to them in his memory.

– – Anne Hawkins & Flavia Jolliffe

INTRODUCTION TO PROBABILITY AND STATISTICS

Eighth edition by William Mendenhall and Robert J. Beaver, 1991. PWS-Kent Publishing Co.

This recently revised book is described by its authors as providing a "modern approach to the subject" and being aimed at "today's readers who require a knowledge of beginning statistical techniques."

Considering their first point: stem-and-leaf displays and box plots are introduced briefly, but not encouraged elsewhere. None of the other equally useful new exploratory data analysis techniques are given. The computer is dealt with a little more thoroughly. Some output from the best-known statistical software packages (MINITAB, SAS, SPSS) is given, and where appropriate MINITAB commands are quoted. These packages are expensive and not often used by a statistical novice. But a calculator is, so reference to both its statistical capabilities and shortcomings would have been helpful. There is the, by now obligatory, large data set — blood pressure readings for nearly two thousand 15 to 20 year-olds, but little use is made of it.

However modernity is amply evident in the "superb exercises" accompanying each chapter. A symbol code is used for quick recognition of application questions covering 15 areas, including environmental studies, agriculture, medicine, sports, and law. Full reference is made to the original source, and the questions using this "real-world" material are well thought out. All fourteen chapters contain an opening statement, which refers to the interesting case-studies to be found at the end of the chapter, and to the general objectives and specific topics contained within the chapter. "Tips on Problem Solving" are found throughout the text, and phrased in language well suited to its readers. Chapters conclude with a brief summary, and a helpful listing of current references for more detailed study.

The authors claim an "overall streamlining" and "more concise coverage" in this edition. Nonetheless the book runs to 635 pages of text, followed by three appendixes on mathematical notations, and fourteen tables of statistical functions. Mendenhall's original 1964 Introduction to Statistics was a mere 261 pages long, but great progress has been made since then, both in the attractive style of presentation and readability. As a text for a one-semester survey course for pre-calculus college students, or as a resource for high schools where there is growing emphasis on statistics and probability, this latest edition can be recommended. Even admitting that some of the authors' selling points are sales hype, there is no doubt that their exercise sets do qualify for the superlative "superb".

The publisher provides both a Partial Solutions Manual (relevant questions color-coded) and a Study Guide (partially programmed to help the reader master the more difficult concepts), but neither were available to this reviewer.

-- Michael Cooke

ELEMENTARY STATISTICS

Elementary Statistics by Leon Marziller (Wm. C. Brown Publishers, 1990) is a new textbook primarily for introductory college statistics course, and particularly those in community colleges. However, I will be reviewing this text for possible use in secondary school classes.

The stated goal of the author in writing yet another statistics text is to bring real understanding of statistics to students and to provide an alternative to two categories of traditional texts: those written by mathematicians, which tend to be overly mathematical, and those written by behavioral scientists which may give a nice overview but avoid some of the mathematics this author views as necessary for learning statistics. He has attempted to combine these two approaches.

The text covers the basic material of an introductory course, beginning with presentation of data, measures of central tendency and dispersion, and then moving to probability and the normal distribution. Inferential statistics follow beginning with confidence intervals and hypothesis tests and finishing with analysis of variance. The book concludes with a section on correlation and linear regression. It is apparent that this is much more material than needs to be covered in a high school statistics course. Less than one-fourth of the book is spent on descriptive statistics, leaving room for much technical detail in the areas of probability and inferential statistics. Even though scatterplots could be introduced earlier in the book along with other methods of graphing data, they are introduced at the end of the book very briefly, before the more detailed sections on correlation and regression.

The sections of the book on graphical techniques seem somewhat dated and insufficient. Old fashioned grouped frequency distributions are described but not stem-and-leaf displays. There is no real motivation developed for even looking at graphics as a way to explore and understand data. Even the sections on descriptive statistics fail to provide a rationale for calculating summary measures. For example, the author states that it is desirable to obtain summary statistics for data but does not give a reason why it is desirable to do this. While the traditional measures of central tendency and dispersion are presented, new and easier measures such as trimmed averages and interquartile ranges are left out.

On the positive side, there are several good data sets included in the chapters, such as a breakdown of "what the world eats" in terms of calories per day for 14 types of foods and 41 different countries. There are also excerpts from comic strips that relate to probability and statistics which help lighten the reading material. At the end of each chapter is a suggested hands-on student project which teachers might find useful.

In summary, Elementary Statistics appears to be a fairly traditional college level statistics text which does not include the newer methods of exploring and displaying data. It would probably not be suitable for a high school statistics class and it does not appear to be a suitable resource to use in helping to achieve the NCTM standards for learning probability and statistics.

-- Joan Garfield

TRAINING TEACHERS TO TEACH STATISTICS

Proceedings of the International Statistical Institute Round Table Conference, Budapest Hungary 23-27 July, 1988, edited by Anne Hawkins; published by International Statistical Institute, Voorburg, Netherlands.

This book contains the keynote papers from the conference. Also included are the comments and discussions following the presentation of each paper. When reading this material, it seems that one is actually at the conference during the presentations. The presenters include teachers, curriculum experts and researchers involved in a number of pilot projects for training teachers in teaching statistics. Several countries are represented.

The book is divided into four sections: The Changing Nature of Statistics, Evaluation of Training Needs and Some Solutions, Defining Problem Areas, and Evaluation of Existing Programs.

The speakers in the first section presented a variety of thoughts on the nature of statistics that present problems for teachers. Along with these possible problems, potential strategies for overcoming these difficulties were presented.

The presenters in the second section spoke about difficulties encountered by teachers in teaching statistics to elementary and secondary students. The teachers background experiences and the misconceptions that they bring to teaching were explored. Once again, potential solutions for these problems were presented.

The speakers in the third section attempted to define problem areas. Included were a lack of well-trained or suitable teachers of statistics and probability, inappropriate curricula, lack of training resources, and the influence of cultural diversity. These were contrasted to the problems that changes in the Australian secondary education presented to teacher educators.

The last section was devoted to describing and evaluating existing programs. This is the largest section of the book and, for me, the most interesting. The speakers described programs that are conventional and others that are innovative. A great variety of countries were represented. the majority of the programs were for in-service programs rather than pre-service programs.

I enjoyed reading the book. I believe that Anne Hawkins has captured the spirit of each of the presentations. I cannot say that I am a better teacher of statistics due to my reading of this book, but I am a better informed person.

-- Ken Sherrick

PROBABILITY JOBCARDS

The *Intermediate Probability Jobcards* packet contains twenty activity cards. Each card lists needed supplies, poses a question about probability or an occurrence, and instructs the student to make a prediction. Instructions are then given for a related experiment and for the recording of the results. Student results are displayed either in charts, tables, lists or graphs. The method of recording for each activity is clearly illustrated. Each activity ends with a "think and write" segment, requiring the student to carefully consider the results of the experiment.

Also included are directions for playing various probability games, for which students need to work in pairs. Upon completion of each game, students are instructed to write about why they think the game is fair or unfair.

Necessary materials for the activities include: regular dice — red and green, octahedra dice, transparent spinners, plastic coins, and Rainbow Cubes (all available from Creative Publications). Also utilized are such everyday items as paper and paper bags, as well as spinner faces, which are included with the "Jobcards".

The *Junior High Probability Jobcards* follow precisely the same design as the "Intermediate" with little variation. Matrices are included as a method of recording results. The "think and write" sections require more critical thinking on the student's part, as do the writing assignments following the games.

The *Intermediate Probability Jobcards* are targeted toward grades 3 through 6; the *Junior High Probability Jobcards*, toward 6 through 9. Both sets are very well suited to their designated levels. However, both sets are appropriate for older students with little experience with probability.

The experiments, in most cases, can be done individually. It is recommended, though, that students work in pairs, as with the games, to encourage communication about the activities and the results. The cards should be done in the order presented, as terms are defined and concepts presented sequentially throughout the activities.

The activity cards are 8 1/2 inches by 11 inches, laminated on one side. Each set comes packaged in a reclosable plastic bag. Information and directions are presented clearly and concisely. The probability definitions and concept explanations are accurate and easy to understand, as is the mathematics involved. As an aid to the instructor, a brief description of and solutions to each activity are included.

These materials lend themselves to various extensions: The activities are appropriate for small cooperative learning groups and for learning centers. The moderate number of trials for each activity can be increased. Students' results can be combined for class discussion. Students can be encouraged to design their own experiments and games.

The *Intermediate Probability Jobcards* and the *Junior High Probability Jobcards*, both by Judy Goodnow and Shirley Hoogeboom, are available from Creative Publications, 5040 West 111th Street, Oak Lawn, IL 60453-9941, (800) 624-0822. The cost of each set is $15.50.

– – Rebecca Maggard

RESAMPLING STATS

Resampling Stats (also known as "Bootstrap Stats") is a computer software program which allows the user to simulate a variety of probability experiments. I have reviewed the IBM version of this program. An Apple version is also available, which I have not seen. My overall impression of this program is that it is of high quality, an excellent step forward for those teachers of probability and statistics who desire a new approach to introducing students to concepts of probability.

Julian Simon has extensive experience in the probability and statistics classroom, using simulation and randomization concepts to give students appropriate intuition in solving probability problems. This experience and expertise is quite evident in *Resampling Stats*. The software is well designed and easy to use. It actually includes its own programming language. The commands for this language are easy to use and well documented in the *Resampling Stats Manual*. There is also a Help facility built into the software to give assistance to beginning users of the package. There are a number of examples illustrating the important concepts in various types of simulation experiments. By reading the descriptions and programs code provided for these examples, a user can quickly become comfortable with the software.

The commands for the package, though few in number and easy to use, actually form a powerful tool for quickly performing simulation and randomization experiments. The notion of the significance level of a statistic can be illustrated through the resampling "Monte Carlo" capabilities of the software.

I wrote a sample program after only a brief exposure (about two hours) to the software. This program simulates a baseball player with a .300 batting average, giving the player four at-bats per game for each of 162 games (a typical major league season). The program computes the maximum length hitting streak (consecutive games in which the player has gotten at least one hit) for each season, and repeats the simulation for 100 seasons. There are few tools which could have been used to program this type of experimental situation so easily. The program took me about 10 minutes to write, and then another 10 minutes for it to run on my IBM AT computer. The output presents a frequency distribution for the maximum length hitting streak for the player in a 162 game season.

– – David Robinson