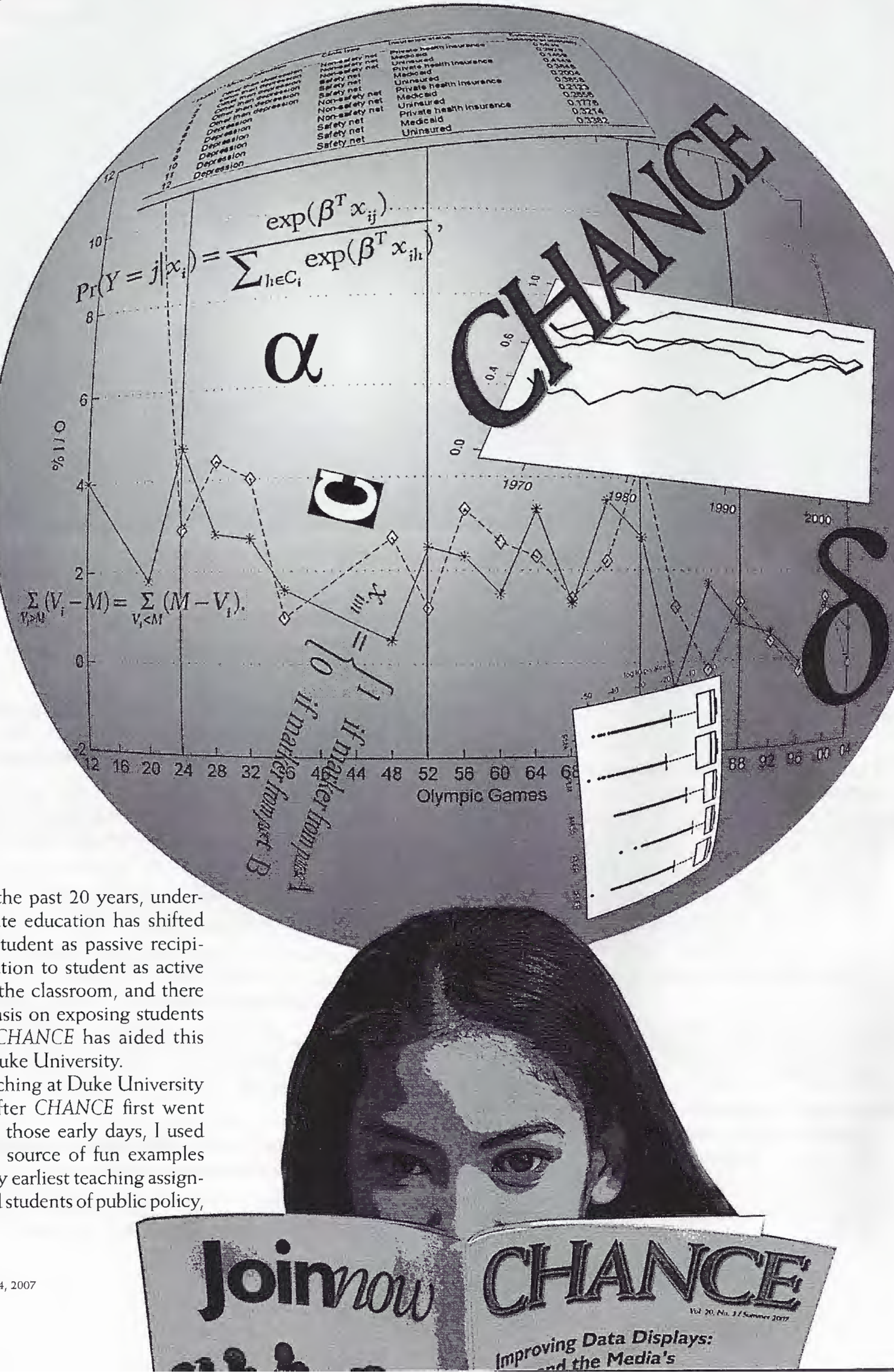
Using *CHANCE* To Engage Undergraduates in the Study of Statistics

Dalene Stangl



During the past 20 years, undergraduate education has shifted from student as passive recipient of information to student as active participant in the classroom, and there is more emphasis on exposing students to real data. *CHANCE* has aided this transition at Duke University.

I began teaching at Duke University a few years after *CHANCE* first went to print. From those early days, I used *CHANCE* as a source of fun examples for lectures. My earliest teaching assignments included students of public policy, so lectures included examples focusing on social issues and government interventions. These and other articles showed students the importance of careful statistical analysis and presentation of statistics for research and decision-making.

As my teaching assignments broadened to include students interested in fields ranging from the social sciences to markets and management, the examples I used from *CHANCE* broadened to include examples covering marketing, political science, sociology, psychology, health, education, and pop culture. These articles provided fun, current examples of the use of graphics, descriptive statistics, probability, estimation, prediction, and modeling. Each year, *CHANCE* provided new examples that engaged students and kept my teaching fresh.

As the 1990s progressed, undergraduate education shifted. Educational researchers began promoting more interaction, more participation, and more active engagement of students. My first attempt to follow this path involved supplemental readings from *CHANCE*. Instead of just interweaving examples from *CHANCE* into my lectures, I asked students to pick *CHANCE* articles of interest to read and present to their classmates. Students were graded on how well they related the article to the material covered in lectures and the textbook.

Throughout the latter half of the 1990s, students grew more sophisticated in their presentations. These included PowerPoint slides, relevant images, applets, animated graphics, and output from data analyses that replicated those presented in *CHANCE* articles. Students blended the *CHANCE* articles with emerging IT tools. As the '90s ended and a new decade began, the shift continued. Students no longer just gave presentations, but supplemented their presentations with classroom exercises and hands-on data collection and analysis. One student presenting Allan Rossman's "'Techno-Thriller' Statistics: *CHANCE* in the Fiction of Michael Crichton" pieced together relevant film clips from movies adapted from Crichton's books, supplementing those with film clips from other movies that made reference to statistical ideas. Another student presenting Jeremy Taylor's "Betting Strategies in Final Jeopardy" created a Jeopardy game using categories related to statistics and tried out betting strategies with the class.

Group Exercises and Computing Labs

One of the most consistently selected *CHANCE* articles is "How Birth Order Influences Individual Characteristics," by Kris Moore, Jonathan Trower, and Kent Borowick. Rarely does a semester pass when at least one student does not choose to present this article. The paper uses data from the General Social Survey conducted annually for the National Data Program for the Social Sciences by the National Opinion Research Center at the University of Chicago. This is a data set we make use of in our computing labs, so students are familiar with its format.

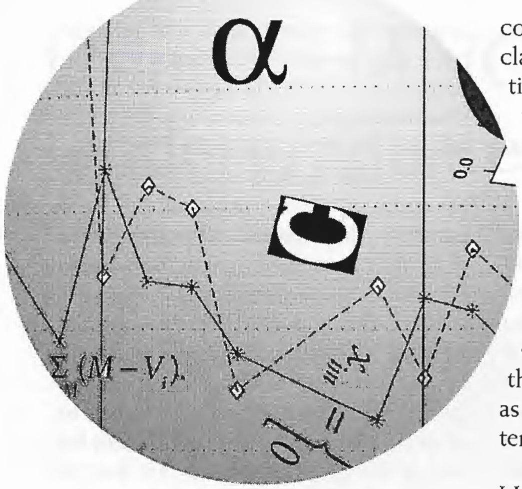
In discussing the paper, students make the exercise more hands-on by asking fellow classmates to fill out a survey about their own birth order and their attitudes toward career, education, work, social status, financial position, religion, family, politics, abortion, relationships, and leisure. We make sure some questions have qualitative responses, while others are quantitative, discrete, or continuous. Students record the survey responses and use mosaic plots, contingency tables, and chi-square statistics to examine the relationship between birth order and qualitative responses. They also use side-by-side boxplots and differences in means or analysis of variance to examine the relationship between birth order and quantitative continuous responses. In analyzing results, students not only compare first-, middle-, and last-born children, but also check the claim that only-born children resemble first-born children. In discussing the hypothesis testing presented in the paper and those we conduct in class, the issues of type I and type II errors and of multiple testing are addressed.

CHANCE is chocked full of sports-related articles that demonstrate a wide variety of statistical methods, but I purposely limit the number of sports articles presented so they do not dominate the course. One exception is "The Cold-Foot Effect," by Scott Berry. I let this football article through because the class activity students created involves paper football, a game for which the women are equally enthusiastic as the men to play. In this activity, students design an experiment, collect data, and analyze data using analysis of variance techniques. The outcome of interest is the student's accuracy at scoring in the game of paper football.

Materials needed include tables or flat surfaces, a ruler, masking tape, and paper footballs (small triangular pieces of folded paper). The goal of the game as described at *www.paperfootballzone.com* is to score touchdowns. To score, you must flick the ball with your fingers and have it stop with part of it sticking over the table's edge. The ruler and tape is used to mark off consistent kicking distances for each player. We used 6, 12, 18, and 24 inches as our blocks. We also blocked on dominant and nondominant hand used for kicking. Finally, the intervention of primary interest was whether the student calling out orders to "kick" hesitated between the time he/she said "Ready, set" and "kick." This simulates football's icing the kicker, where a team delays the game in hopes it will play with the psyche of the field goal kicker and result in a poor kick.

Students kick four times at each distance with each hand. For each of the 32 kicks, one student calls out in a randomized fashion the distance and the hand followed by instructions to kick. Delays were randomized into the design, but balanced across distance and hand of the kicks. Students recorded the outcome of each kick as a success or failure. Data was entered on a laptop and analyzed in real time by the class. Typically, hand has the largest effect, followed by distance, with delay having no effect.

In "Surprises from Self Experimentation," by Seth Roberts with comments from Robert Rosenthal and Donald Rubin, Roberts conducts a series of small self-experiments—experiments with himself as the subject. His experiments examined the relationship between sleep and breakfast, morning faces and mood, standing and sleep, and sugar water and weight. When students read this article, they are both fascinated and critical, so their assignment is to come up with and carry out their own self-experiment. Students have designed experiments that examined diet, sleep, and study strategies. Designing and discussing these experiments develops their critical thinking skills and teaches them to be more understanding of the difficulties in developing and implementing good design. This is always a good time to



discuss faked data and the detection of faked data in both the self-experimentation and more broadly. Two *CHANCE* articles that cover this are "The Aftermath of Falsified Data in Breast Cancer Trials," by David Harrington, and "Reflections on the NSABP Affair," by Judith Rich O'Fallon.

In presenting the *CHANCE* articles "Inferences About Testosterone Abuse Among Athletes," by Donald A. Berry and LeeAnn Chastain, and "Does Drug-Testing Deter Participation in Athletic Events? EPO and the Sydney Olympics," by Heather Mitchell and Mark F. Stewart, students designed an in-class simulation that examined how well caffeine users can be detected by examining changes in heart rate. All students abstained from caffeine consumption during the four hours before class. Upon arrival, they were trained to a protocol for taking heart rate, and then baseline measures of heart rate were taken. Students were randomized (in a double-blind manner) to two groups, one that drank 12 ounces of coffee (about 200 milligrams of caffeine) and one that drank 12 ounces of decaffeinated coffee (about 10 milligrams of caffeine). The chance of being assigned to each group was 10% and 90%, respectively; however, this information was not disclosed to subjects. After 10, 15, and 20 minutes, heart rates were again measured and differences from baseline calculated, recorded, and graphed (separately for men and women) using JMP statistical software.

The class proceeded through steps of creating decision rules (using the magnitude and consistency of increase in heart rates) for classification of "drug users" and "nonusers" by thinking through the costs of false positive and false negative classifications. Students discussed sensitivity, specificity, predictive value positive, and predictive value negative. After results were disclosed, receiver operating characteristic curves were drawn to further discuss the error rates and better define decision rules that balanced the cost of each error. After the simulation, the class discussed the testing of athletes for steroid use and then generalized what they had learned to other contexts, such as exam cheating, disease detection, and terrorist profiling.

The first student to select "Does Having Boys or Girls Run in the Family?" by Joseph Lee Rodgers and Debby Doughty initially thought it was too difficult to present to his peers. Showing him how to simulate data analogous to that in the paper by rolling dice and tossing coins changed his mind and led to a fun class activity. The question in the article was whether there is a genetic factor that produces a bias in favor of one gender for some families. The article made use of data from the National Longitudinal Survey of Youth (NLSY), while we collected information about the families of the students in the class. We also simulated data using several of the models discussed in the paper. Some models assumed the likelihood of having a girl (or boy) did not depend on the gender composition of previous children, while other models defined this probability as being conditional on having had a previous number of girls (or boys) because of parental stopping patterns.

First, students simulated families with a fixed probability of boys/girls and fixed probabilities for stopping after 1, 2, 3, and 4 children without regard to the gender of any child. The second simulation accounted for the gender of previous children by defining separate probabilities for boys/girls depending on whether previous children defined a boy-biased family, girl-biased family, or a gender-neutral family. As described in the paper, we used several parental stopping strategies by setting parameters for the chances of stopping after the first, second, or third child, depending on whether the child was a boy or girl. The probability of stopping after four children was always set to one. The last type of stopping parameter allowed a parental "balance preference," that is, a desire to have at least one child of each gender. This required additional parameters for stopping when at least one child of each gender was achieved.

These schemes and the models presented in the paper seemed complicated to students until they simulated data using coin tosses to represent the birth of a boy or girl, and using dice rolls to determine stopping. Working backward from the simulation to estimating parameters based on observed data gave students insight about how estimation and inferences were made from the NLSY data.

Inspired by the article "The Mysterious Case of the Blue m&ms," by Ronald Fricker, I created my own group exercise to teach the difference between Frequentist and Bayesian hypothesis testing. While the *CHANCE* article looked at percentages of blue m&ms and compared the observed percentages to those claimed by the manufacturer, we looked at the proportion of yellow m&ms in a glass jar. I knew the percentage of yellow m&ms in the jar was 20%; however, students knew only that the percentage was 10% or 20%. Students were asked to test the null hypothesis that the percentage of yellow m&ms is 10% against the alternative that it is 20%. The loss table is such that if they decide incorrectly, they must buy me a $6 lunch, and if they decide correctly, I buy them a $4 lunch. They also must pay me $1 for every five m&ms they sample.

Under the Frequentist paradigm, they must select a significance level at which to reject and a sample size before any sampling is done. Once they make these decisions, I sample and announce the color of a sequence of 5, then 10, then 15, and then 20 m&ms. Each student writes down the data for the sample size they purchased, and then they calculate the test statistic and p-value for the data. Rarely does a student set a high enough significance level or buy enough m&ms to reject the null hypothesis. Knowing this, I make sure the sample has 20% yellows for each sample size. Despite many warnings about the arbitrariness of a 5% significance level, students still rely on it. Nearly 100% of the students make decisions that pay me $3 to $4 for a few m&ms and results in them buying me a $6 lunch.

After examining the m&m problem from a Frequentist perspective, we flip the problem and make decisions using Bayesian methods. I tell the students I flipped a coin to determine the percentage of yellows placed in the jar. This suggests equal prior probabilities on the null and alternative hypotheses. This time, students are allowed to sample sequentially, deciding on whether to buy an additional five m&ms after observing a sequence of five. We use the same data sequence already observed, except the sample size is not predetermined. After each sequence of five, students calculate the posterior probabilities of the null and alternative hypotheses. Starting with equal prior probabilities, the posterior quickly moves to favor the alternative hypothesis. Usually, students continue to buy all 20 m&ms, even though the posterior for the alternative is 1.5 times greater after only 10 m&ms. While the students end up spending $4 to buy 20 m&ms, most make the correct decision that there are 20% yellow m&ms in the jar, and hence they pay for their own $4 lunch.

Beyond group exercises, *CHANCE* also comes into the classroom through computing labs, in which students analyze the data sets that sometimes accompany articles. In "Uncounted Votes: Does Voting Equipment Matter?" by Mary Meyer, the author examines whether percentages of uncounted votes vary by whether the voting equipment is a punch card ballot, a lever machine, or and an optical scan system. Meyer made the data available through the *CHANCE* web site, and I turned the article and data into a computing case study. Students had to replicate some of the analysis presented in the paper and think through how the analysis helps answer the research question. This replication and thinking through the analyses helps them begin building skills formulating statistical arguments relevant to a research question. Issues of weighting and confounding are front and center in this case study.

The data from the Florida elections in the presidential election of 2000 are widely circulated and the topic of several articles in *CHANCE*, including "Voting Irregularities in Palm Beach Florida," by Greg Adams, "Bush v. Gore: Two Neglected Lessons from a Statistical Perspective," by Michael O. Finkelstein and Bruce Levin, and "Discussion of 'Bush v. Gore …'" by Richard L. Smith. The data make a good case study useful in teaching concepts in probability, regression, prediction, confounding, and the link between data analysis and decisionmaking.

Courses for Specialized Audiences

At Duke, I teach several other courses for students with more specialized interests. One is a first-year student seminar limited to no more than 18 students. Typically, students are a fifty-fifty split of students who took and loved AP statistics and student athletes looking for a way to meet the Duke quantitative studies requirement. The course is titled Statistics, Science, and Society. The course description reads as follows:

> Students will read popular-press accounts of self-selected topics from newspapers and magazines such as *The New York Times*, *Science News*, *Discover*, *Nature*, *CHANCE*, and *Scientific American*. Popular-press readings will be supplemented by readings from scientific journals, as well as readings on basic probability and statistics related to the topic. Topic possibilities include, but are not limited to the safety of cell phones, grade inflation at elite universities, statistical evidence in court cases on discrimination, trademark and patent violations, global warming, undercount errors in the U.S. Census, environmental justice, DNA evidence in the courts, the AIDS epidemic, detecting exam cheating and drug use in sports, the reliability of political polls, etc. For each topic chosen, we will examine in depth the interplay of scientific advancements and advances in statistical computing that provide the technologies necessary to answer these complex questions. We will also examine how these solutions can be used to impact decisionmaking, and hence the political, economic, ecological, and sociological aspects of our daily lives.

CHANCE plays a larger role in this course than the other sources combined. This is because while the questions raised in the other sources are timely and fascinating, they rarely provide enough detail for students to understand the research design or how the analysis was done.

The second course I teach for a more specialized audience is Introduction to Biostatistics. Enrollees are a mix of students in pre-medicine, students getting certificates in health policy, and students in environmental studies. For this course, I use *CHANCE* in all the ways I discussed earlier: as a source of examples for my lectures, as supplemental readings for topics we covered in class and the text, and as foci for class exercises. Also, while editor of *CHANCE*, I initiated a new column, "Here's to Your Health," edited by Mark Glickman and Cindy Christiansen. Now, every issue of *CHANCE* provides a fun example for any introductory biostatistics course.

I began teaching a third course for a more specialized audience in Spring 2006, Statistics in the Courtroom, in which I incorporated *CHANCE* readings. During the late 1980s and into the 1990s, there was rarely an issue without an article about census adjustment, so I typically assigned a couple to read, such as "An Adjusted Census in 1990? The Supreme Court Decides" and "Census 2000 Controversies," by Margo Anderson and Stephen E. Fienberg. In 2003, a law-related column, "*CHANCE* at the Bar" was begun. It provides excellent examples linking statistical analysis and court cases.

Finally, a course I will teach for a specialized audience this spring is one requested by a group of English literature majors who took one of my introductory statistics courses. It is a seminar course on stylometry and other topics that bridge statistics and literature. The core of the reading list will center on *CHANCE* articles.

CHANCE has been at the core of my undergraduate teaching for nearly 15 years. I owe the magazine a great deal of credit for the success of my teaching. It helps me engage students and keep my teaching fresh. It does this in much the way *Statistics: A Guide to the Unknown* does, except *CHANCE* updates my cache of examples four times every year. I encourage more use of *CHANCE* by AP and undergraduate teachers, and I thank the ASA, the ASA staff, Springer, and the long list of volunteer editors and authors who have made this happen for 20 years.

Subject	Title	Author	Issue
Social Issues and Government Interventions	The Employment Situation: How Do We Know About It?	Janet L. Norwood	2(3), 1989
	Scientific Inferences and Environmental Health Problems	John C. Bailar, III	4(2), 1991
	An Adjusted Census in 1990? The Trial	Stephen Fienberg	5(3–4), 1992
	Expanding the Role of the Bar Chart in Representing Crime Data	Terry Allen and Glen Buckner	5(3–4), 1992
	Should Pregnant Women Move? Linking Risks for Birth Defects with Proximity to Toxic Waste Sites	Sandra Geschwind	5(3–4), 1992
	Like a New Drug, Social Programs Are Put to the Test	Peter Passell	6(4), 1993
	What Happened to HIV Transmission Among Drug Injectors in New Haven?	Edward H. Kaplan and Robert Heimer	6(2), 1993
Marketing, Political Science, Sociology, Psychology, Health, Education, and Pop Culture	Misuses and Abuses of Statistical Techniques in Market Research Surveys	Solomon Dutka and Lester R. Frankel	3(4), 1990
	A Statistical History of the AIDS Epidemic	Ron Brookmeyer and Mitchell H. Gail	3(3), 1990
	Graphs in the Presidential Campaign: Why Weren't They Used by More Than One Candidate?	Howard Wainer	6(1), 1993
	Statistical Evidence of Cheating on Multiple-Choice Tests	Stephen P. Klein	5(3–4), 1992
	Catching Drug Couriers	Daniel D. Reidpath, Mark R. Diamond, and K. Vijayan	6(4), 1993

Subject	Title	Author	Issue
... Marketing, Political Science, Sociology, Psychology, Health, Education, and Pop Culture *continued*	Monty Hall's Probability Puzzle	Eduardo Engel and Achilles Venetoulias	4(2) 1991
	Wringing the Bell Curve: A Cautionary Tale About the Relationships Among Race, Genes, and IQ	Bernie Devlin, et al.	8(3), 1995
	Cost-Effectiveness in Public Education	Jay Bennett	8(1), 1994
Student Favorites	How Birth Order Influences Individual Characteristics	Kris Moore, Jonathan Trower, and Kent Borowick	8(4), 1995
	'Techno-Thriller' Statistics: *CHANCE* in the Fiction of Michael Crichton	Allan J. Rossman	7(1), 1994
	Betting Strategies in Final Jeopardy	Jeremy M. G. Taylor	7(2), 1994
	Small Cars, Big Cars: What Is the Safety Difference?	Leonard Evans	7(3), 1994
	DNA, Statistics, and the Simpson Case	Donald A. Berry	7(4), 1994
	Propensity to Abuse—Propensity to Murder?	Jon F. Merz and Jonathan P. Caulkins	8(2), 1995
	How Will the NCAA's New Standards Affect Minority Student Athletes?	Stephen P. Klein and Robert M. Bell	8(3), 1995
Health	Does Exposure to Second-Hand Smoke Increase Lung Cancer Risk?	Alan J. Gross	6(4), 1993
	What Evidence Is Needed To Link Lung Cancer and Second-Hand Smoke?	Howard E. Rockette	6(4), 1993
	Does Vasectomy Cause Prostrate Cancer?	Rebecca DerSimonian and John Clemens	10(1), 1997
	The Aftermath of Falsified Data in Breast Cancer Trials	David Harrington	10(2), 1997

Subject	Title	Author	Issue
... Health *continued*	Reflections on the NSABP Affair	Judith rich O'Fallon	10(2), 1997
	Random Ranking of Hospitals Is Unsound	Jonas Andersson, Kenneth Carling, and Stefan Mattson	11(3), 1998
	Developing an AIDS Vaccine by Sieving	Peter Gilbert	13(4), 2000
	Bayesian Clinical Trial Design in a Cancer Center	Peter Thall	14(3), 2001
	The Unpleasant Placebo?	Stephen Senn	15(2), 2002
	Stopping Rules in Clinical Trials	Michael O. Finkelstein and Bruce Levin	17(4), 2004
	Reconsidering Smallpox	Sarah Nichalak and Greg Wilson	18(1), 2005
	Statistical Sleuthing During Epidemics: Maternal Influenza and Schizophrenia	Nicholas J. Horton and Emily C. Shapiro	18(1), 2005
Law	Statistics, Expert Witnesses, and the Courts	Stephen E. Fienberg, Samuel Krislov, and Miron Straf	1(4), 1988
	Flawed Reasoning in Court	Ian Evett and Bruce Weir	4(4), 1991
	Ecological Regression in Voting Rights Cases	Stephen P. Klein and David A. Freedman	6(3), 1993
	Improving the Odds on Justice	Robert Mathews	7(3), 1994
	Simpson's Paradox: An Example in a New Zealand Survey of Jury Composition	Ian Westbrooke	11(2), 1998
	Counting Cars in a Legal Case Involving Differential Enforcement	Norma Terrin and Joseph B. Kadane	11(3), 1998
	How To Catch a Thief	Bryan F. J. Manly and Richard Thomson	11(4), 1998
	Statistics, Expert Witnesses, and the Courts	Stephen Fienberg, Samuel Krislov, and Miron Straf	11(4), 1998
	An Allegation of Examination Copying	Joseph B. Kadane	12(3), 1999

Subject	Title	Author	Issue
... Law *continued*	Voting Irregularities in Palm Beach Florida	Greg Adams	14(1), 2001
	Anatomy of a Jury Challenge	Joseph B. Kadane	15(2), 2002
	Discussion of Bush v. Gore	Richard L. Smith	16(4), 2003
	Alleged Racial Steering in an Apartment Complex	Jason T. Connor and Joseph B. Kadane	14(2), 2001
	Crossing Lines in a Patent Case	Joseph B. Kadane	15(4), 2002
	Inferences About Testosterone Abuse Among Athletes	Donald A. Berry and LeeAnn Chastain	17(2), 2004
Literature	Survival on the Bestseller List	M. A. Grove	4(2), 1991
	Who Was Shakespeare?	Ward Elliott and Robert Valenza	4 (3), 1991
	Digit Preference in the Bible	David Salsburg	10(2), 1997
	The Torah Codes: Puzzle and Solution	Maya Bar-Hillel, Dror Bar-Natan, and Brendan McKay	11(2), 1998
	Searching for the 'Real' Davy Crockett	David Salsburg and Dena Salsburg	12(2), 1999
	Math Chaps and Other Oddities: Statisticians in Literature	Nicole Lazar and David Sidore	16(2), 2003
	Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution	José Nilo G. Binongo	16(2), 2003
	Who Was the Author? An Introduction to Stylometry	David Holmes and Judit Kardos	16(2), 2003
	Stylometry and the Civil War: The Case of the Pickett Letters	David Holmes	16(2), 2003
	Cherry Picking in Nontraditional Authorship Attribution Studies	Joseph Rudman	16(2), 2003