

Teaching Bits: A Resource for Teachers of Statistics

Journal of Statistics Education v.8, n.1 (2000)

Robert C. delMas
General College
University of Minnesota
354 Appleby Hall
Minneapolis, MN 55455
612-625-2076

delma001@maroon.tc.umn.edu

William P. Peterson
Department of Mathematics and Computer Science
Middlebury College
Middlebury, VT 05753-6145
802-443-5417

wpeterson@middlebury.edu

This column features "bits" of information sampled from a variety of sources that may be of interest to teachers of statistics. Bob abstracts information from the literature on teaching and learning statistics, while Bill summarizes articles from the news and other media that may be used with students to provoke discussions or serve as a basis for classroom activities or student projects. Bill's contributions are derived from Chance News (http://www.dartmouth.edu/~chance/chance_news/news.html). Like Chance News, Bill's contributions are freely redistributable under the terms of the GNU General Public License (<http://gnu.via.ecp.fr/copyleft/gpl.html>), as published by the Free Software Foundation. We realize that due to limitations in the literature we have access to and time to review, we may overlook some potential articles for this column, and therefore encourage you to send us your reviews and suggestions for abstracts.

From the Literature on Teaching and Learning Statistics

"Statistics in Context"

by Jane Watson (2000). *The Mathematics Teacher*, 93(1), 54-58.

Abstract: Judging statistical claims in social contexts is fundamental to statistical literacy. This article uses a particularly contentious newspaper report that makes a cause-and-effect claim as the basis for discussing this important aspect of statistical understanding. The issue's relevance across the school curriculum is shown by extracts from curriculum documents. Teachers need to structure experiences to build ability to question claims made without proper justification. This article suggests a hierarchy to help teachers plan for and assess student learning in this area, and it closes with a plea for teachers to cooperate across subjects to achieve results.

Journal for Research in Mathematics Education

"Students' Probabilistic Thinking in Instruction"

by Graham A. Jones, Cynthia W. Langrall, Carol A. Thornton, and A. Timothy Mogill (1999). *Journal for Research in Mathematics Education*, 30(5), 487-519.

Abstract: In this study we evaluated the thinking of 3rd-grade students in relation to an instructional program in probability. The instructional program was informed by a research-based framework that included a description of students' probabilistic thinking. Both an early- and a delayed-instruction group participated in the program. Qualitative evidence from 4 target students revealed that overcoming a misconception in sample space, applying both part-part and part-whole reasoning, and using invented language to describe probabilities were key patterns in producing growth in probabilistic thinking. Moreover, 51% of the students exhibited the latter 2 learning patterns by the end of instruction, and both groups displayed significant growth in probabilistic thinking following the intervention.

"The Meaning of Randomness for Secondary School Students"

by Carmen Batanero and Luis Serrano (1999). *Journal for Research in Mathematics Education*, 30(5), 4558-4567.

Abstract: In the experimental study reported here we intended to examine possible differences in secondary students' conceptions about randomness before and after instruction in probability, which occurs for the Spanish students between the ages of 14 and 17. To achieve this aim, we gave 277 secondary students a written questionnaire with some items taken from Green (1989, 1991). With our results we extend Green's previous research to 17-year-old students and complement his results with the analysis of students' arguments to support randomness in bidimensional distributions. Our results also indicate that students' subjective understanding of randomness is close to some interpretations of randomness throughout history.

"Developing Concepts of Sampling"

by Jane M. Watson and Jonathan B. Moritz (2000). *Journal for Research in Mathematics Education*, 31(1), 44-70.

Abstract: A key element in developing ideas associated with statistical inference involves developing concepts of sampling. The objective of this research was to understand the characteristics of students' constructions of the concept of sample. Sixty-two students in Grades 3, 6, and 9 were interviewed using open-ended questions related to sampling; written responses to a questionnaire were also analyzed. Responses were characterized in relation to the content, structure, and objectives of statistical literacy. Six categories of construction were identified and described in relation to the sophistication of developing concepts of sampling. These categories illustrate helpful and unhelpful foundations for an appropriate understanding of representativeness and hence will help curriculum developers and teachers plan interventions.

The American Statistician: Teachers Corner

"Use of Course Home Pages in Teaching Statistics"

by Ramón V. León and William C. Parr (2000), *The American Statistician*, 54(1), 44-48.

Abstract: Our focus in this article is on how a course home page can be used to support classroom teaching (not on Web-based education, where the primary teacher-student interaction is by way of the World Wide Web). Our discussion is primarily based on our experiences over the last four years using course home pages to support statistics courses at several levels from introductory to graduate. Over these years we have had the opportunity to try several ideas; seeing some of them work as expected -- or better -- and seeing others fail to produce any benefits or even detract from the classroom experience. Our discussion should be of value for deciding if your course would benefit from having a home page or for improving an existing one. We also give some advice in the Appendix on how to organize the file structure of a course home page.

"Using a Term-Long Project Sequence in Introductory Statistics"

by John P. Holcomb, Jr. and Rochelle L. Ruffer, *The American Statistician*, 54(1), 49-53.

Abstract: We propose a series of projects for introductory data analysis classes using a single, real multivariate dataset. The assignments combine four current trends in statistics education: computers, real data, collaborative learning, and writing. Completing the project sequence allows students to appreciate and understand the interconnectedness of statistical concepts. We present rationale for the assignments including five objectives. We provide institutional and course background, an explanation of the data source and variables, grading criteria, a summary of each project, and an assessment of student learning. We conclude with a Web address to facilitate implementation of this approach.

Topics for Discussion from Current Newspapers and Journals

"Year 2000 Computer Problems May Get an Alibi"

by Barnaby J. Feder, *The New York Times*, 14 December 1999, C1.

With all the warnings as the year 2000 approached, it seemed people would be ready to blame any snafu they encountered on the Y2K bug. To mitigate such perceptions, the President's Council on Year 2000 Conversions set out to document the existing failure rates for a range of government and industry activities. Here are some examples from their report. Ten percent of automated teller transactions fail on the first attempt, usually because of customer errors; also, at any given time, one to two percent of the nation's ATM machines are either out of service or out of money. In each of the last five years, tens of thousands of US or Canadian residents have suffered power outages during late December or early January, primarily resulting from weather. Over the last three years, pipelines carrying hazardous materials have averaged 16 "reportable disruptions" during the period from December 31 to January 3.

The idea behind the report is that actual problems encountered on January 1 could be compared to these documented "baseline" rates. Baselines are notoriously ignored by the average citizen, but should prove helpful to public authorities trying to identify Y2K-specific failures. William Ulrich, a Y2K expert in Soquel, California, is not so sure about that last point. He says that "most people will have experts who know what's normal in their command posts but 90 percent will be doing their assessments based on gut feeling."

"Is Complexity Interlinked with Disaster? Ask on Jan. 1"

by Laurence Zuckerman, *The New York Times*, 11 December 1999, B11.

In his 1984 book, *Normal Accidents: Living With High-Risk Technologies*, Charles Perrow argued that disasters such as the accident at the Three Mile Island nuclear reactor and the explosion of the Space Shuttle Challenger should not be attributed to "human error." He theorized that the complex and highly interconnected nature of modern systems would lead to accidents in spite of best human efforts; hence, we should view them as "normal accidents." Perrow's ideas are in the news now because January 1, 2000, will effectively amount to a large scale test of his theory. Perrow himself expects that Y2K accidents will cause some inconvenient disruptions in services, but no major disasters.

Another view of systems is called the "high-reliability" theory, which argues that by building many backup systems, safety can be enhanced. Normal accident theory would say that the existence of complicated backup systems can actually contribute to the likelihood of an accident. For example, Perrow feels that a technology like nuclear energy cannot be managed safely enough to justify the risks.

Coincidentally, a number of recent high profile stories related to accident rates were in the news this past fall. The Mars Polar Lander disappeared, President Clinton highlighted the National Academy of Sciences' warnings about fatal medical accidents, and the first fatality attributable to experimental gene therapy was reported.

"Point, Counterpoint and the Duration of Everything"

by James Glanz, *The New York Times*, 8 February 2000, F5.

In our last issue we considered the article "How to Predict Everything" (*The New Yorker*, 12 July 1999, pp. 35-39), which describes how physicist John Gott proposes to compute prediction intervals for the future duration of any observed phenomenon. Gott's method hinges on the "Copernican assumption" that there is nothing special about the particular time of your observation, so with 95% confidence it occurs in the middle 95% of the lifetime. If the phenomenon is observed to have started A years ago, Gott infers that A represents between $1/40$ (2.5%) and $39/40$ (97.5%) of the total life. He therefore predicts that the remaining life will extend between $A/39$ and $39A$ years into the future. (Given Gott's assumptions, this is simple algebra: if $A = (1/40)L$ where L is the total life, then the future life is $L - A = 39A$.) Gott has used the method to predict everything from the run of Broadway plays to the survival of the human species!

But can such broad applicability really be justified? Not according to Dr. Carlton Caves, a physicist at the University of New Mexico (and a *New Yorker* reader!) who has put together a systematic critique of Gott's work. His article, "Predicting Future Duration from Present Age: A Critical Assessment," will be published in *Contemporary Physics*.

Caves' ideas are based on Bayesian analysis. He says that Gott errs by ignoring prior information about the lifetimes of the phenomena in question. For example, Gott claims to have invented his method while standing at the Berlin Wall in 1969, eight years after it was erected. With 50% confidence he inferred that those eight years represented between $1/4$ and $3/4$ of its total life, so he predicted that the Wall would last between $2\frac{2}{3}$ years and 24 years into the future. (For the record, twenty years later, the Wall did come down.) But what sense does it make, asks Caves, to ignore historical and political conditions when making such a prediction? Surely, such prior knowledge is relevant, and Bayesian ideas provide a framework for incorporating it. In Caves' view, failing to do so in favor of some "universal rule" is unscientific.

To illustrate the matter more simply, Caves imagines discovering a party in progress, where we learn that the guest of honor is celebrating her 50th birthday. Gott's theory predicts that, with 95% certainty, she will live between 1.28 and 1950 additional years, a range which Caves dismisses as too wide to be useful. Even worse, he points out, would be to predict that with 33% confidence she is in the first third of her lifetime and thus has a 33% chance to live past the age of 150! As a challenge to Gott, Caves has produced a notarized list of 24 dogs owned by people associated with his department. He identified the half dozen who are older than 10 years -- prior information that Gott would presumably ignore. Gott's method would presumably predict that each had a

50% chance of living to twice its current age. Caves is willing to bet Gott \$1000 on each dog, offering 2-to-1 odds that it won't live that long. Caves cites Gott's refusal to bet as evidence that he doesn't believe his own rule.

For more technical details, you can read Caves' paper online at

<http://xxx.lanl.gov/abs/astro-ph/0001414>

and Gott's rebuttal at

<http://www.physicsweb.org/article/news/4/2/6/1/news-04-02-06a>

As for the dogs, Gott thinks his analysis would apply to the whole sample, not to each dog individually hand-picked by Caves. Trying to sort out all of this is an interesting discussion exercise involving the notions of sampling, confidence levels, and prediction.

"Estrogen Study Raises Concerns"

by Judy Foreman, *The Boston Globe*, 15 February 2000, E1.

"Combining Estrogen, Progestin Raises Risk of Cancer, Study Says"

By Judy Foreman, *The Boston Globe*, 26 January 2000, A1.

Postmenopausal women using hormone replacement therapy face a complicated picture of benefits and risks. These articles concern two recent studies on the combined use of estrogen and progestin. This combined treatment has been known to reduce the risk of uterine cancer by a factor of eight compared with estrogen alone. But the new studies -- one published in the January edition of the *Journal of the American Medical Association*, and the later one due to appear in the *Journal of the National Cancer Institute* -- have found that the combined treatment is associated with an increase in breast cancer risk.

The more recent study was directed by researchers at the University of Southern California. They looked at 1897 women with breast cancer and 1637 similar women without breast cancer, comparing the use of postmenopausal hormone treatments. Each five years of estrogen alone was associated with a 6% increase in breast cancer risk, while each five years of combined treatment was associated with a 24% increase. Actually there are two regimes for the combined treatment. In "combined continuous" treatment, both hormones are taken daily; in "sequential" treatment, estrogen is taken every day, but progestin is taken for only part of each month. For sequential treatment, breast cancer risk rose 9% for each five years of use, compared with a 38% increase per five years of continuous treatment. However, the article reports that these differences were not statistically significant.

The earlier study was larger, involving more than 46,000 women. According to the *Globe* article:

...[the study] found that while taking estrogen alone increased a woman's risk of breast cancer by 1 percent for each year of use, taking estrogen plus progestin increased the risk by 8 percent per year. If that's true, according to an editorial accompanying the study, taking the combination therapy for 10 years could increase breast cancer risk by 80 percent.

Milt Eisner provided the additional references below and remarked that it is interesting to compare how different news sources reported the risk.

The New York Times ("Study Backs Hormone Link to Cancer for Women," by Denise Grady, 27 January 2000, A17) stated that

The researchers ... found that women who took the hormone combination for five years had a 40 percent increase in the risk of breast cancer, compared with those who did not take the treatment. Women who took estrogen alone had a 20 percent increase.

These are statements about relative risk. Later in the article, the *Times* quoted lead investigator Catherine Schairer, giving the following interpretation in terms of absolute risk:

... consider a group of 100,000 normal-weight women, ages 60 to 64, none of whom take hormone replacement. During a five-year period, 350 cases of breast cancer would be expected. But if all the women took combined hormone replacement for 5 years, about 560 cases would be expected.

The original research report appears in the January 26 edition of the *Journal of the American Medical Association* (C. Schairer, et al., "Menopausal Estrogen and Estrogen-Progestin Replacement Therapy and Breast Cancer Risk," 283(4), pp. 485-491). The results are summarized by the following relative risk statements, which readers might try to match up with the news reports.

Increases in risk with estrogen only and estrogen-progestin only were restricted to use within the previous 4 years (relative risk [RR], 1.2 [95% confidence interval {CI}, 1.0-1.4] and 1.4 [95% CI, 1.1-1.8], respectively); the relative risk increased by 0.01 (95% CI, 0.002-0.03) with each year of estrogen-only use and by 0.08 (95% CI, 0.02-0.16) with each year of estrogen-progestin-only use among recent users, after adjustment for mammographic screening, age at menopause, body mass index (BMI), education, and age.

You can read the full report online at <http://www.jama.com>.

"The Placebo Prescription"

by Margaret Talbot, *The New York Times Magazine*, 9 January 2000, p. 34.

This is a long essay filled with intriguing examples and insights. It describes the continuing debate within the medical community about how placebos work, and whether it is ethical to try to harness the placebo effect as a viable treatment.

The lead example concerns "placebo surgery." A Houston surgeon had 10 patients scheduled for arthroscopic knee surgery to treat their arthritis. In a double-blind experiment, two of the patients underwent the standard surgery of rinsing and scraping the joint, three had rinsing alone, and five had a sham surgery consisting only of three incisions to the knee with no arthroscopy. Here blinding the surgeon meant that he did not know which surgery to do until he entered the operating room and opened an envelope with instructions; by this time the patient was already under anesthesia. Six months later, the patients -- still blinded to the treatment -- were all reporting less pain.

People are probably more familiar with the placebo effect in the context of drug trials. Here the article gives several examples of cases in which patients' improvement seems to be mainly attributable to the placebo effect. In one case, both the placebo-takers and the drug-takers of a food-allergy drug had a 75% positive response. Another case involved the study of a genetically engineered heart drug, where subjects receiving the placebos actually showed more improvement than did those who received the drug.

These examples may seem dramatic, but evidence for the placebo effect is widespread. The author cites a classic study from 1955, in which Harvard research Harry Beecher estimated that 30 to 40% of any treatment group could be expected to respond favorably to placebo. This all leads to the following question: Why not include placebos among doctors' standard treatments? While this may not seem consistent with the rational stance of western medicine, recent years have seen increasing numbers of patients turning to "alternative" medical procedures that science finds difficult to explain (see, for example, the next article on herbal remedies). However,

the article does raise the disturbing thought that HMOs -- everybody's favorite villains in the health care arena -- might come to favor the use of placebos as a cost-cutting measure.

Perhaps the idea of placebos-as-treatment would be more palatable if we could understand the reasons that they work. A popular theory is that their power derives from the care shown by the doctor for the patient. Several examples are cited to support this notion. Unfortunately, the trend towards managed care seems to be working against the traditional image of the empathetic, caring doctor. Many of today's doctors are discouraged by their diminished role in decision-making and are reluctant to make authoritative statements for fear of lawsuits. But the article concludes on an optimistic note. If it could be shown that patients would require fewer visits to a caring physician than to a harried one, perhaps even the HMOs would pay attention.

"Go the medical route if herb doesn't relieve depression"

by Judy Foreman, *The Boston Globe*, 10 January 2000, C4.

"St. John's wort: Less than meets the eye; Globe analysis shows popular herbal antidepressant varies widely in content, quality"

by Judy Foreman, *The Boston Globe*, 10 January 2000, S4.

"How the Globe did its testing"

by Judy Foreman, *The Boston Globe*, 10 January 2000, C4.

In recent years, the herb St. John's wort has been widely marketed as a non-prescription antidepressant. While many people claim to have experienced benefits, the first article here points out that these could be attributable to the placebo effect. More convincing evidence may be available next year, when a nationwide study led by researchers at Duke University is due to be completed. This study, which is being sponsored by the National Institute of Mental Health, will compare St. John's wort with the antidepressant drug Zoloft and a placebo.

Various preparations of St. John's wort are sold under different brand names. In 1996, German researchers used meta-analysis to combine 23 short-term studies involving different brands. For mild to moderate depression, St. John's wort performed as well as prescription medications, and better than placebo. The Duke study will use Kira, a German brand most studied to date. It will also enforce a consistent protocol. Although the smaller studies have indicated positive results, one member of the Duke team pointed out that none of them had a completely satisfactory methodology.

The last two articles address concerns about what consumers actually get when they purchase St. John's wort. The FDA does not certify health claims of herbal products. Also, natural products are sensitive to factors like temperature, so the potency of herbal preparations can degrade on the shelf, diminishing whatever benefit they might have. A *Globe* investigative team therefore set out to test store-bought samples of six leading brands: Natrol, NatureMade, Nature's Resource, Quanterra, YourLife and the CVS store brand. They also purchased a seventh product, HerbalLife, which is sold only through distributors.

Two laboratories were chosen to analyze the samples: a chemical testing company called PhytoChem Technologies in Chelmsford, Massachusetts, and an herbal testing company called Paracelsian, Inc. in Ithaca, New York. Both laboratories received a complete set of seven bottles of pills. Each set also contained an eighth bottle of "dummy" pills supplied by the Massachusetts College of Pharmacy. The bottles had coded labels to blind the analysts to the brand names.

PhytoChem used spectrophotometry to measure the amount of hypericin in each sample. This chemical is believed to be the herb's "active ingredient." The lab found that only Nature's Resource lived up to its labeling claim of 0.3% hypericin content, which is considered the industry standard. Four others were close: Natrol had 0.28%, NatureMade had 0.27%, and HerbalLife and YourLife both had 0.25%. Quanterra, which makes no claim, turned out to contain almost no hypericin!

Paracelsian compared each sample to Prozac and Zoloft and to the Perika brand of St. John's wort. The first two are prescription drugs that treat depression by inhibiting the absorption of the neurotransmitter serotonin by brain cells. Perika has been shown in rat studies to have similar effects on serotonin and a second neurotransmitter, dopamine. Only two preparations, Quanterra and NatureMade, passed the Paracelsian's "BioFIT" test for their ability to inhibit the absorption of both serotonin and dopamine.

The last article comments on the strengths and weaknesses of the investigation. Cited as strengths were choosing two labs, blinding the labs to the brand names and to the presence of a dummy preparation, and buying the herbal preparations at a store as consumers would. Among the weaknesses cited were not testing the preparations on humans, omitting some brands (Kira, for example), and not confirming the results with other laboratories. Finally, some scientists now think that hyperforin, not hypericin, may be the relevant chemical in the herb.

"Food Surveys Have Wide Margins of Error; Researchers Know that Questionnaire Results Don't Always Reflect Actual Eating Habits"

by Lawrence Lindner, *The Washington Post*, 1 February 2000, Z9.

"In Defense of the Harvard Nurses' Health Study"

by Walter Willett, M.D., Letter to the Editor, *The Washington Post*, 8 February 2000, Z4.

We frequently read stories on the latest link found between diet and disease. The controversy reported here concerns Harvard University's Nurses' Health Study. Dr. Willett was concerned by the tone of the *Post* article which criticized the methodology of food surveys in general and the Harvard study in particular.

The *Post* cited a number of reasons to doubt the quality of data produced by food surveys. One was the complexity of the questions. The following example from the Harvard study was cited:

How often, on average, did you eat a quarter of a cantaloupe during the past year? One to three times a month? Once a week? Two to four times a week? Once a day? Please try to average your seasonal use over the entire year. For example, if cantaloupe is eaten four times a week during the approximately three months it is in season, then the average use would be once a week.

Even when people understand the questions, they may not accurately report their eating habits. Writing in the *American Journal of Clinical Nutrition*, psychologist John Blundell cited a recent study finding that obese men under-reported their calorie intake by 36%. He speculates that as we hear more and more about the evils of fat consumption, for example, respondents will similarly tend to under-report that. Recognizing that the survey data may be flawed, Blundell warns against drawing strong conclusions from them.

These problems are compounded when study results are reported, because the news media do a poor job distinguishing between correlation and causation. A recent *Washington Post* headline announced "Study Links Hot Dogs, Cancer: Ingestion by Children Boosts Leukemia Risk, Report Says." The story began: "Children who eat more than 12 hot dogs per month have nine times the normal risk of developing childhood leukemia..."

Nevertheless, the *Post* does not conclude that food surveys are worthless. It quotes nutrition researcher James Fleet of the University of North Carolina, who says that such research "generates new hypotheses to test in more

controlled settings."

In his letter to the editor, Dr. Willett states that the cantaloupe question was presented out of context. He points out that questionnaire design is a scientific process, with careful consideration given to the question wording and order. Furthermore, researchers do not blindly accept survey data. Instead, they attempt to validate the reports using previously established risk factors. For example, the ratio of polyunsaturated to saturated fat intake is known to be related to cardiovascular disease. The confirmation of this relationship in the Nurses' Health Study gives additional confidence in the reporting. Finally, the 20-year duration of the study strengthens conclusions about long-term effects of diet.

[JSE Homepage](#) | [Subscription Information](#) | [Current Issue](#) | [JSE Archive \(1993-1998\)](#) | [Data Archive](#) | [Index](#) | [Search JSE](#) | [JSE Information Service](#) | [Editorial Board](#) | [Information for Authors](#) | [Contact JSE](#) | [ASA Publications](#)