Going beyond the book: Toward critical reading in statistics teaching*

Andrew Gelman†

January 28, 2007

Abstract

We can improve our teaching of statistical examples from books by collecting further data, reading cited articles, and performing further data analysis. This should not come as a surprise, but what might be new is the realization of how close to the surface these research opportunities are: even influential and celebrated books can have examples where more can be learned with a small amount of additional effort.

We discuss three examples that have arisen in our own teaching: an introductory textbook that motivated us to think more carefully about categorical and continuous variables; a book for the lay reader that misreported a study of menstruation and accidents; and a monograph on the foundations of probability that overinterpreted statistically insignificant fluctuations in sex ratios.

Keywords: categorical and continuous variables, handedness, menstruation, primary sources, secondary sources, sex ratio, teaching, textbooks, traffic accidents

Introduction

This article considers three examples from our own teaching experiences in which much was learned by going to the sources of examples in books. It was surprisingly easy to discover areas where even excellent and well-regarded books had opportunities for debate, and we hope that this article will motivate other instructors to more critically examine the monographs, texts, and general-interest books that they use for teaching. We suspect this will help students in learning general research skills as well as statistical methods.

Categorical or continuous?

The book *Mind on Statistics*, by Jessica Utts and Robert Heckard (2001), is an excellent text that is full of examples for statistics classes at all levels. A fun thing about working

*We thank Ji Meng Loh and Martin Lindquist for helpful comments and the National Science Foundation, National Institutes of Health, and Columbia University Applied Statistics Center for financial support.

†Department of Statistics and Department of Political Science, Columbia University, New York, gelman@stat.columbia.edu, www.stat.columbia.edu/∼gelman

Variable	Possible categories
Dominant hand	Left-handed, right-handed
Regular church attendance	Yes, no
Opinion about marijuana legislation	Yes, no, not sure
Eye color	Brown, blue, green, hazel

Figure 1: Variables listed as "categorical" in Utts and Heckard (2001). I don't know enough about eye color to comment, but the first three variables could also be considered as fundamentally continuous; see Figure 2a for the handedness example. Discussing this in class can give students a deeper understanding of discrete and numerical data.

from a good textbook is that more can be learned by considering its examples in further depth. For example, Figure 1 shows a table from early on in the book, where the concepts of numerical and categorical variables are introduced. From another perspective, though, three out of the four variables that they list as categorical could be also be considered as continuous.

The issue is clearest with handedness, which Utts and Heckard categorize as left or right-handed but can be also described by a continuous variable, as we illustrate with the first histogram in Figure 2, which is based on data we collected from students in a class. (More systematic surveys obtain similar results; see, for example, Oldfield, 1971.) As Figure 2a shows, many people fall between the two extremes of pure left- and pure right-handedness. But, as Figure 2b illustrates, students tend to guess the distribution of handedness to be bimodal and thus essentially discrete. This common misconception would make handedness a particularly effective example of a continuous variable that is often summarized discretely.

Similar issues arise for other variables listed in Figure 1. Church attendance can be measured by a numerical frequency (for example, number of times per year), which would be more informative than simply yes/no, or it can be binned in ordered categories. For example, the American National Elections Study asks, "How often to you attend religious services, not counting weddings or funerals?" and records five sorts of response: more than once per week, once per week, more than once per month, several times per year, and never. Finally, the three options for opinion about marijuana legislation could be coded as 1, 0, and 0.5, and further intermediate preferences could be identified with detailed survey questions, for example asking about medical marijuana, criminal penalties, and so forth.

The point of bringing all this up in class is not to lay down the law and say that church attendance, for example, is inherently discrete or continuous. Rather, we want to lead students to think about the ways in which reality is abstracted by numerical measurements. We also find it empowering that we can learn more about the structure of these variables

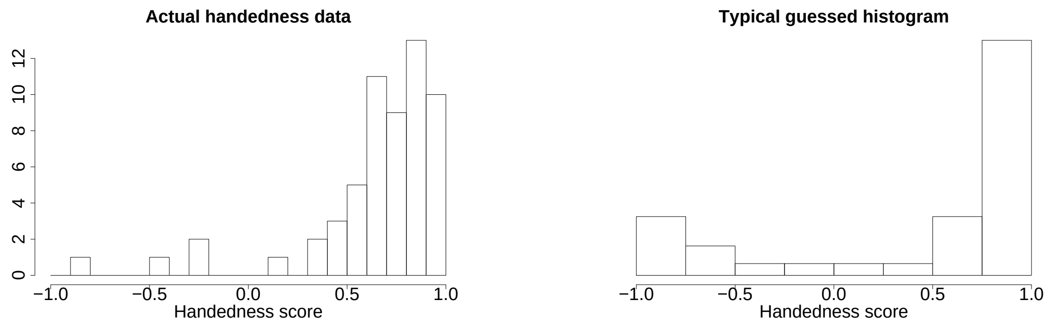

Figure 2: Handedness can be measured by a 10-item questionnaire to yield an essentially continuous score ranging from -1 (pure left-hander) to $+1$ (pure right-hander). We had the students in an introductory statistics class fill out the questionnaire and also asked them to sketch what they thought the histogram of other students' handedness scores would look like. (a) Data from the class; (b) a guess from a group of students of what they thought the histogram would look like (before seeing the actual data). Bimodality was anticipated but did not actually occur.

either by collecting our own data (as illustrated in Figure 2) or with library research (as by looking at the National Elections Study questions).

The graph that wasn't there

About fifteen years ago, when preparing to teach an introductory statistics class, I recalled an enthusiastic review I had read (Sills, 1986) of the sixth edition of Hans Zeisel's book, *Say it With Figures* (1985). I bought the book and, flipping through it to find examples for use in class, came across the two sketches reproduced in Figure 3 here. The curves represent data from hospital admissions of premenopausal women who had been involved in traffic accidents, with the left hump representing accidents that had occurred just before the menstrual period and the right hump showing accidents occurring just after the period.

This seemed like a great example for class. I figured that a graph of the actual data would be even better than a sketch, so I went to the library and found the cited research by Katharina Dalton (1960). The graphs are reproduced in Figure 4, and they look nothing like Zeisel's sketches in Figure 3! For one thing, the sketched densities show almost all the probability mass just before and after menstruation, but the data show only about half the accidents occurring in these periods. Perhaps more seriously, the sketch shows two modes with a gap in the middle, whereas the data show no evidence for such a gap. Similarly, the two bell-shaped pictures in the second sketch of Figure 3 do not match the actual data as shown in the histograms on the right side of Figure 4.

Figure 3: Sketch from an example in Zeisel (1985), who writes, "When the frequency of [driving] accidents is plotted against the time of menstruation a surprisingly shaped curve arises [left graph]. Upon investigation, the curve turned out to be the composite of two easily identified separate curves; one for parous women (those who had given birth) and one for nonparous women). The one group had the accident peak immediately after their period, the other immediately before it." Compare to the actual data shown in Figure 4.

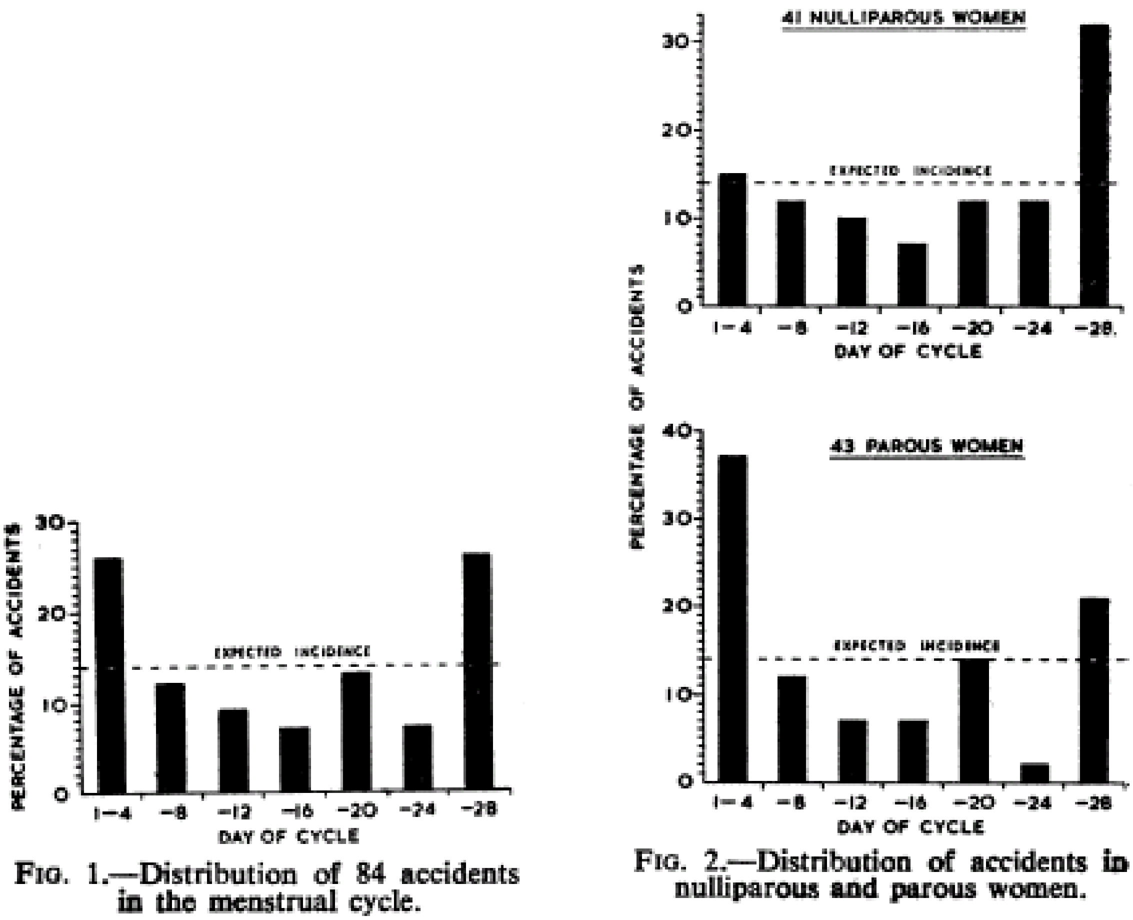

Figure 4: Graphs from Dalton (1960) with the raw data on menstruation and accidents. These histograms look almost nothing like the sketches in Figure 3 taken from Zeisel's book. Many of the accidents fall outside of the days indicated by the modes in Figure 3 and, unlike in those sketches, there is no gap just between the peaks.

Dalton's findings were conveniently summarized by an article in *Time* magazine on November 28, 1960: "In four London general hospitals. Dr. Dalton questioned 84 female accident victims (age range: 15 to 55), all of whom had normal, 28-day menstrual cycles. Her findings: 52% of the accidents occurred to women who were within four days, either way, of the beginning of menstruation. On a purely random basis, the rate would have been only 28.5% for the same eight days. Childless women, noted Dr. Dalton, appear to be abnormally accident-prone just before menstruation, while women who have borne children are vulnerable over the whole premenstrual and menstrual period." What is relevant to our discussion here is that these findings were not accurately described in Zeisel's book. On an unrelated but amusing (from a current perspective) note, the *Time* article quoted Dalton as saying that these findings "cause one to consider the wisdom of administering tranquilizers for premenstrual tension."

I suspect that Zeisel heard about the research (perhaps even by reading *Time* magazine), recognized that it would be a good teaching example, and went to the library to read Dalton's original article. He then could have too hastily summarized the data in a sketch, inadvertently knocking out most of the accidents that did not occur just before or after menstruation and mistakenly inserting a gap in his histogram between the two modes. Or perhaps he was loooking for an example of a mixture model and didn't look too closely at the data. In any case, this is a benefit to our students, who get a lesson in how easy it is to misread a research report. Had Zeisel's book not been so appealing and well written, we would not have been drawn to the example in the first place.

Finding patterns in noise

The book *Probability, Statistics, and Truth* by Richard von Mises (1957) is an important text in the foundations of probability, laying out a derivation of the axioms of probability theory from the concept of infinite random sequences. This work has long been influential in statistics (see, for example, Wald, 1939, and Good, 1958, for classical frequentist and Bayesian reactions) and in philosophy (for example, Gillies, 2000, connects von Mises's ideas to those of Karl Popper and others).

I bought the book several years ago and, in skimming it, alighted on the chapter on "Applications in Statistics," within which von Mises uses the sex ratio of births to illustrate the binomial distribution. He examines the proportion of babies born who were boys in each of the 24 months of 1907–1908 in Vienna and found less variation than expected. In his words: "The average of these 24 values is 0.514 33; the dispersion [variance] ... is

0.000 0533." He computes the expected variance of $p(1–p)/n = 0.000\,0613$ (here, n is about 3900 per month) and then writes, "The actual dispersion is smaller than the theoretical one. In other investigations of the proportion of male births, a value of Lexis's ratio closer to 1 is obtained. We must therefore look for an explanation of the slightly subnormal dispersion found in this special case." He goes on to attribute this lower variance to different sex ratios in different racial or socioeconomic groups.

In fact, though, the variance, though less than expected by chance, is not at all *statistically significantly* less, based on the 23 degrees of freedom available from these data. (Under the $\chi^2_{23}/23$ distribution, the observed ratio of 0.869 has a p-value of 0.36; that is, one would observe a ratio at least this extreme more than a third of the time, just by chance.) Thus, it is unnecessary to search for an explanation for the discrepancy, especially given that, as von Mises notes, sex ratios are among the rare data that actually *do* generally follow the binomial distribution. In addition, irrelevantly for the technical point but of interest when teaching, von Mises makes a presentational lapse by summarizing dispersion with variances rather than standard deviations, which are more interpretable on the original scale of the data.

Von Mises is hardly alone in overinterpreting birth data: there is a long tradition of looking for patterns in birth data, despite that there is no convincing evidence that boys or girls run in families or that sex ratios vary much at all except under extraordinary conditions. (See Freese and Powell, 2001, Das Gupta, 2005, and Gelman, 2007, for more on the overinterpretation of statistical fluctuations in sex ratios.) Thus, in addition to illustrating the important technical point of assessing statistical significance of a variance ratio, this example opens the door to a more general discussion of how and why statistics can be misread.

That this occurred in an influential book merely underscores that even a standard χ^2 test for overdispersion cannot be taken for granted. In a similar vein, finding that the great Francis Galton performed inaccurate calculations with the normal distribution (mistakenly estimating that there were nine-foot-tall men in Britain; see Gelman, 2006, and Wainer, 2007) gives us a new respect for the pioneers who worked out the mathematical property of that model.

Discussion

Individually, these examples are of little importance. After all, one does not go to a statistics textbook to learn about handedness, menstruation, and sex ratios. It is striking, however, that the very first examples I looked at in the Zeisel and von Mises books—the examples with interesting data patterns—collapsed upon further inspection. In the Zeisel example, we went to the secondary source and found that his sketch was not actually a graph of any data, and that he in fact misinterpreted the results of the study. In the von Mises example, we reanalyzed the data and found his result to be not statistically significant, thus casting doubt on his already doubtful story about ethnic differences in sex ratios. In the Utts and Heckard example, we were inspired to collect data on handedness and look at survey questions on religious attendance to find underlying continuous structures.

Teaching activities already exist in which students apply critical reading skills to news reports and scientific articles with statistical content (Gelman and Nolan, 2002); here the recommendation is to have an inquiring eye when reading books that we teach from as well. Much can be learned by redoing analyses and going to the primary and secondary sources to look at data more carefully, and that this can help us teach even from our favorite books.

References

Dalton, K. (1960). Menstruation and accidents. *British Medical Journal* **2**, 1425–1426.

Das Gupta, M. (2005). Explaining Asia's "missing women": a new look at the data. *Population and Development Review* **31**, 529–535.

Freese, J., and Powell, B. (2001). Making love out of nothing at all? Null findings and the Trivers-Willard hypothesis. *American Journal of Sociology* **106**, 1776–1788.

Gelman, A. (2006). Galton was a hero to most. Statistical Modeling, Causal Inference, and Social Science blog, 23 Oct. `www.stat.columbia.edu/ gelman/blog`

Gelman, A. (2007). Letter to the editor regarding some papers of Dr. Satoshi Kanazawa. *Journal of Theoretical Biology* **245**, 597–599.

Gelman, A., and Nolan, D. (2002). *Teaching Statistics: A Bag of Tricks*. Oxford University Press.

Gillies, D. (2000). *Philosophical Theories of Probability*. London: Routledge.

Good, I. J. (1958). Review of "Probability, Statistics and Truth," by Richard von Mises. *Journal of the Royal Statistical Society A* **121**, 238–240.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113.

Sills, D. L. (1986). Review of "Say it with Figures," by Hans Zeisel. *Journal of the American Statistical Association* **81**, 257.

Utts, J. M., and Heckard, R. F. (2002). *Mind on Statistics*. Pacific Grove, Calif.: Duxbury.

Von Mises, R. (1957). *Probability, Statistics, and Truth*, second edition. New York: Dover. Reprint.

Wainer, H. (2007). Galton's normal is too platykurtic. *Chance* **20** (2), 57–58.

Wald, A. (1939). Review of "Probability, Statistics and Truth," by Richard von Mises. *Journal of the American Statistical Association* **34**, 591–592.

Zeisel, H. (1985). *Say it with Figures*, sixth edition. New York: Harper and Row.