Questions, Answers and Statistics
Biometrics in the CSIR: 1930-1940
Teaching Statistics Theory through Applications

Deborah Nolan
University of California, Berkeley

When Terry Speed arrived in Berkeley in the 1980s, I too was a new arrival. He was coming to Berkeley as a senior hire and I as a junior. It was through our connection to David Pollard that we discovered our mutual interest in teaching statistics. We first collaborated on a small project to introduce computing into the advanced undergraduate theoretical statistics course. The computing exercises we developed were aimed at students uncovering through simulation studies some of the rules of thumb that a practicing statistician regularly uses. This was not as successful as we had hoped because our students didn't see any reason to care about the simulation results. We had fallen into the trap Terry warns against in Speed (1986) - teaching pseudo-applied statistics with context-free numbers. Subsequent attempts led us to connect the work to real applications and then to the template described in Nolan and Speed (1999) and used in *Stat Labs: Mathematical Statistics through Applications* (Nolan and Speed, 2002). It would seem that this template should have been an immediate and obvious result of Speed (1986). It wasn't. While Terry modestly claims to be "no exception - for allowing ourselves to forget the fundamental importance of the interplay of questions, answers and statistics", I dare conjecture that one of his goals in the project was for me to gain experience through trial and error in developing an effective approach to teaching statistics.

Speed (1986) successfully argues that the "whole point of statistics lies in the interplay between context and statistics." Others share this viewpoint as noted in the quotes included in the article from James, Cox, Dawid, and Tukey. However, Terry takes this assertion into the education arena and compels us to reflect this important thesis in our teaching. Following Speed (1986), others have made similar arguments to change statistics education. According to Cobb and Moore (1997), "The focus on variability naturally gives statistics a particular content that sets it apart from mathematics itself and from other mathematical sciences, but there is more than just content that distinguishes statistical thinking from mathematics. Statistics requires a different kind of thinking, because data are not just numbers, they are numbers with a context." Higgins (1999) and Nicholls (2001) echo these statements, e.g., Higgins claims that "for the past 40 years, statistics has been doing a great job of training theoretical statisticians, but we have a more data based society and it is crucial that we identify changes to course content and delivery that need to occur." Similarly, Wild and Pfannkuch (1999) note "the biggest holes in our educational fabric, limiting the ability of graduates to apply statistics, occur where methodology meets context (i.e. the real world)."

One teaching strategy offered in Speed (1986) is to meet people with data, by for example, pairing the statistics teacher with a teacher in an empirical field of inquiry or pairing your statistics students with students who have subject matter knowledge. Anecdotal evidence of the success of this approach appears in Field et al. (1988). There we learn of the preparation of Betty Allan, Mildred Barnard, and Helen Turner for successful careers in biometrics at CSIR in the 1930s. All three women spent significant time in Rothamsted Station where they learned statistics by designing and carrying out experiments under the guidance of Fisher, Wishart, and Yates.

Terry raised and answered in his 1986 paper two common objections to working with real problems: that these problems are too complex and the data too large to be practical in the classroom and that only the most advanced students who have a sufficiently large set of tools can successfully attack real problems. Today, we face a new version of these same concerns. Data are now free and ubiquitous. People with all sorts of backgrounds have ready access to data. This data explosion is an enormous opportunity for us to make better, more informed decisions. However, this opportunity presents challenges because people expect to be able to interact with data in new ways and the role of the statistician is changing.

As I reflected on Terry's call to change how we teach statistics, it was at first disconcerting to see that we are still asking statistics educators to consider this issue. Cobb (2007) explains that "What we teach was developed a little at a time, for reasons that had a lot to do with the need to use available theory to handle problems that were essentially computational." Efron (2003) describes the mathematical statistics course as "caught in a time warp" that "does not attempt to teach what we do and certainly not why we do it." Brown and Kass (2009) examine statistics graduate training and warn us to break away from the view of the statistician's role as "short-term consultant" because that model "relegates the statistician to a subsidiary position, and suggests that applied statistics consists of handling well-formulated questions, so as to match an accepted method to nearly any kind of data." I have since realized that we must periodically revisit this question of how best to teach statistics and that is precisely the point. We are not aiming at a fixed target that once arrived at we will have accomplished our goal.

Efron (2003) suggests starting over by imaging "a universe where computing preceded mathematics in the development of statistics" and advocates "starting more muscularly without worrying about logical order of presentation" and focussing instead on the basic kinds of reasoning and explanations that can be arrived at through randomization-based inference. Cobb (2007) further develops this notion, explaining how randomization-based inference "makes a direct connection between data production and the logic of inference that deserves to be at the core of every introductory course." Cobb further posits that "Technology allows us to do more with less: more ideas, less technique. We need to recognize that the computer revolution in statistics education is far from over." Brown and Kass (2009) advocate taking a "less restrictive view of what constitutes statistical training." They see a blurring of the distinction between people with data and people with statistical expertise and state "some of the most innovative and important new techniques in data analysis have come from researchers who would not identify themselves as statisticians." Brown and Kass recommend we minimize prerequisites to research, require real-world problem solving in our courses, and embrace a deeper commitment to cross-disciplinary training. Efron (2003), Cobb (2007), and Brown and Kass (2009) advocate twenty-first century changes to statistics education that echo Terry's call to include the value of statistics in our training programs.

Terry Speed's advice from twenty-five years ago remains extremely relevant today as computational and data challenges continue to evolve and shape our field.

References

Brown, E. and Kass, R. (2009). What is statistics? *The American Statistician*, 63:105–110.

Cobb, G. (2007). The introductory statistics course: A ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1:1–15.

Cobb, G. and Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104:801–823.

Efron, B. (2003). Is the math stat course obsolete? In *Panel Discussion at the Joint Statistics Meeting, San Francisco, CA*. `www.amstat.org/sections/educ/MathStatObsolete.pdf`.

Field, J., Speed, F., Speed, T., and Williams, J. (1988). Biometrics in the CSIR: 1930–1940. *Australian Journal of Statistics*, 30(B):54–76.

Higgins, J. J. (1999). Nonmathematical statistics: A new direction for the undergraduate discipline. *The American Statistician*, 53:1–6.

Nicholls, D. F. (2001). Future directions for the teaching and training of statistics at the tertiary level. *International Statistical Review*, 69:11–15.

Nolan, D. and Speed, T. (2002). *Stat Labs: Mathematical Statistics Through Applications*. Springer Texts in Statistics. Springer-Verlag New York.

Nolan, D. and Speed, T. P. (1999). Teaching statistics theory through applications. *The American Statistician*, 53:370–375.

Speed, T. (1986). Questions, answers and statistics. In *Proceedings of the ICOTS2, University of Victoria, Canada, IASE*, pages 18–28.

Wild, C. and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistics Review*, 67:223–248.