Integrating Data Analysis with Parametric Inference in Undergraduate Statistics

K.L. Weldon
Department of Statistics and Actuarial Science
Simon Fraser University

Abstract:

A rising interest in graphical methods and resampling seems at odds with traditional parametric inference. The role of graphs is often portrayed as a supplementary technique for checking parametric assumptions, and yet it deserves to be recognized as an independent method of analysis. Computer software availability is still providing new opportunities for curriculum change. The problem for the curriculum designer is to construct an appropriate merger of the two approaches for undergraduate courses. This session suggest ways to ease into these more computationally intensive methods.

Introduction:

Most textbooks in introductory undergraduate statistics courses include topics under the general headings of descriptive statistics including bivariate data, elementary probability, sampling theory, estimation, and hypothesis testing. Some will also include chapters on regression, categorical data analysis, and analysis of variance, and probability models. These topics were present in my courses as a student in 1960, and even in Fisher's 1925 text "Statistical Methods for Research Workers". Since then a revolution in computation and communication has changed the way we teach statistics, but surprisingly, it has not changed the list of topics. The premise of this paper is that the topics should change to match the modern context. The GAISE Reports of the ASA (http://www.amstat.org/education/gaise/) recommend many useful changes but response has been modest. Radical suggestions like those in Weldon(2005) and on www.stat.sfu.ca/~weldon, and Bryce(2005), are suppressed here in favor of a more feasible and gradual shift in emphasis.

Although textbooks in undergraduate statistics have been slow to adopt new topics, it is still true that textbooks have improved. As recent examples, texts by Moore(2004), De Veaux, Velleman and Boch (2008), and Agresti and Agresti (2007) are great improvements on earlier texts: they use real data, they encourage conceptual understanding, and they try to accommodate the computer-age availability of statistical software. What I would like to explore is whether the best modern textbooks have adapted sufficiently to the topics best suited to the modern context.

How are we to subvert the tradition of pre-computer topics in statistics? The suggestion here is to use the new methods in parallel with the old methods. Of course, this requires reducing some detail of the old methods. We argue that this would be an acceptable way to introduce incremental change in the list of required topics.

Factors resisting curriculum evolution:

There are a few obvious reasons why topics in statistics courses have remained so constant in spite of the indications for change. In fact, some are the same reasons that apply to change in instruction in any discipline.

First, there is the fact that it is very convenient if an instructor can follow an existing textbook. Textbooks are sponsored by a publisher only when the publisher is convinced that there is a current market for the text. The publisher relies on peer evaluation for this judgment of the market, and the most influential peers are the academics that are the most successful in the accepted traditions of the field. Consequently, textbooks that shift emphasis to a novel list of topics are rare.

Second, changing course content requires extra time and energy. The institution needs to support the change and the argument to produce this compliance must be provided. The actual courses to convey the new course material must be generated without a precedent as a guide. The reaction of students who are troubled by the lack of coherence between the new course material and available textbooks can be a problem requiring explanation. The academic agent-for-change must have a strong motivation to allocate this extra energy and time to the novel course content.

Third, the main incentives provided by universities are to produce publishable research, and to a lesser degree, to keep students happy with the requirements of the courses taught. Time spent on working out new material, and making it student-friendly, will definitely detract from the research productivity of the academic.

Fourth, many states, provinces and countries, try to have similar curricula in several different institutions of higher education, and so change in one usually requires change in the others. Transfer credit in our shrinking world is convenient for students but makes unilateral change at one institution difficult.

Fifth, our discipline of statistics has evolved out of mathematics, and the change of emphasis from deductive to inductive logic has been difficult. There is still a reluctance to accept the intuitive aspects of statistical analysis instead of entirely objective aspects. This has slowed acceptance of needed changes in statistics curricula.

It seems easier to contemplate change in a course in Electronics or English, where new works have a clear claim to status as new course content, than in a course in statistics, which tries to produce generally applicable methods. A methods course might be expected to evolve more slowly than a facts course, and statistics is often considered to be a methods course.

But we should question whether our early statistics courses should be entirely methods courses. Most students taking a first course in statistics are also taking their last course in statistics! In view of this, it is surely appropriate to include some practical implications of simple statistical strategies as an inherent part of the courses. As examples, portfolio diversification and insurance profits are understood better as a consequence of the sampling distribution of the sample mean; the illusion of quality in some successful sports teams can be seen as a consequence of randomness; the disincentives of lotteries and gambling can be understood with simple probability models. There are many reasons why these things are not explained in our courses, but one that does not seem valid is that these applications are less important than the theory underlying them. R programs for demonstrating various applications are given in Weldon (2007).

An English or Electronics course teaches not only the techniques of analysis but also the current products, and perhaps we could do both theory and real-life application in our statistics courses. Perhaps our reluctance to do this our statistics courses is partly responsible for the loss of control of statistics courses from statistics departments to departments of business and psychology. But the acceptance of application content as a central part of our courses, and not merely as illustration of methods, seems to require a huge shift in attitude, and perhaps this is a sixth reason why our curricula has been slow to adapt to a context outside of mathematics.

Considering these six reasons for resistance to change, it is not too surprising that statistics curricular development has been slow over the last 50-100 years. For the same reasons, we cannot expect any violent revolution over the next 50-100 years. But for those who feel change would be helpful to the discipline, it is worthwhile to consider ways to accelerate acceptance of gradual changes.

Many eminent statistics professionals have contributed to the discussion of topics for a reformed curriculum. A particularly useful one is Garfield etc al (2002). The comments here add some suggestions for content and for implementation, but also reinforce the earlier suggestions.

Suggestions for Gradual Change:

As mentioned earlier, early statistics courses tend to follow the pattern of descriptive statistics including bivariate data, elementary probability, sampling theory, estimation, and hypothesis testing. In the following, I discuss ways to reduce some traditional material in these categories in favor of more modern material.

Descriptive Statistics A major lack in traditional courses is methods for smoothing relationships. Time series are one of the most common data structures in both general media and social or environmental science research, and non-parametric smoothing is an easy-to-understand and useful technique – even the moving average is a start. Of course, the same methods can be used with bivariate (non-time-series) data as well. The substitution of simulation demonstrations for hand-calculations allows smoothing to be described fairly quickly. Both the use and the dangers of smoothed relationships can be described – the important relationship between the degree of smoothing and the preservation of detail can be conveyed. To find time for this, we might reduce the discussion of details of histogram or stem-and-leaf plot construction, and the hand-calculation formula for the standard deviation.

Bivariate Data A descriptive approach to bivariate data usually involves the definition of the correlation coefficient and simple linear regression. Variability phenomena can be described using simulation. Hand-calculation procedures can be omitted although definitions must be explained. Formulas for slope and intercept can be de-emphasized in favor of explaining the use and hazards of the least-squares criterion. The emphasis on linear regression can be reduced slightly with the addition of an explanation of non-parametric smoothes, and that residual plots for both have the same purpose. With this approach, the straight-line model can be described as just one option rather than the only simple method. Methods of smoothing could be illustrated with something simple like a moving average, but software-supplied methods like loess should also be included in the discussion.

Probability and Sampling The trend in recent years has been to reduce the discussion of probability models and to limit discussion to the normal distribution and the central limit theorem. It is true that many students have trouble understanding models like the Binomial and the Hypergeometric, and especially the combinatorics underlying them, and these topics are best left to more advanced courses. But many basic notions of probabilities can be built up from as simple an experiment as the toss of a coin, and these ideas are really essential for understanding variability phenomena. Simulation of a random walk provides a useful demonstration of the convergence of averages and the non-convergence of totals. Many students think "If I toss a fair coin 1000 times, I am likely to get 500 heads and 500 tails"!

The difficulty with the probability so often emphasized in stats courses in the 60s and 70s was the mathematics, not the understanding of variability. By reducing both in an effort to make statistics more easily understood may have been a mistake. The advent of "resampling" may allow replacement of some of the mathematics of probability by the mechanics of simulation. For classroom implementations, see Taffe and Garnham (1996), Christie (2004), Wood (2005), and Arnholt(2007).

The concept of the sampling distribution of the sampling mean is very puzzling to students but it is crucial to an understanding of phenomena like diversification of investments, the insurance business, and the rationale of replicated measurements in science. The role of sample size in survey sampling needs to be understood. We need to spend more time on this section, even if stealing time from estimation and hypothesis testing is required.

Estimation and Hypothesis Testing The emphasis on method of inference in traditional courses has been overdone. The emphasis on parametric modeling of every situation, of being the parametric-inference police instead of the important-information detective, has evolved from our mathematical beginnings in the early 20th century. We need to move in the direction of a more informal and more flexible set of tools to extract information from data. Resampling, simulation and smoothing techniques are methods warranting increased attention in discussion of inference. We can reduce the long list of situations involving small samples, unknown variances, and multiple groups. We can admit that our P-value methods are not decision theoretic but rather credibility assessments. More emphasis on precision and less on decision would seem to be a useful trade-off. The important role of graphical methods in assessing data needs to be recognized, and the limited role of parametric estimation can make room for this. In real-world applications, the entire distribution of measurements is usually more useful than a parametric summary of them, and the ease of graphical display in the modern context makes this a preferred approach.

Discussion The changes suggested above could be accomplished by current instructors in current courses. If a student has lacked drill in choosing the appropriate hypothesis test for the various situations for testing means and proportions, they will have gained a better understanding of when the hypothesis test is meaningful, and what it really means. This would be a useful trade-off. Once a student understands what is required, they can track down the details from texts or experts.

The article mentioned above by Bryce(2005) is more radical in that it suggests that the teaching of statistics should be developing expertise, rather than transmitting knowledge. He quotes Snee(1993):

> "There is a growing consensus that the 'content side' of statistical education should move away from the mathematical and probabilistic approach and place greater emphasis on data collection, understanding and modeling variation, graphical display of data, design of experiments, surveys, problem solving, and process improvement."

An extension of this suggestion is that statistics education should be driven mostly by case studies, to teach the process of learning from data, with theory supplied as needed. However ideal this might be, we need to be practical about what is possible in our current context of higher education. As Bryce(2005) says "These calls for reform have gone largely unheeded, at least in part because of our current paradigm of statistical education." Love (1998) outlines a case study approach but admits the difficulty presented by the lack of a suitable text for such a course. One attempt to fill this gap is Schafer and Ramsey (2003). The text edited by Peck et al (2005) is a possible candidate recently published. However, the incremental change suggested here is a realistic way to head in the right direction, and can be implemented immediately.

One aspect of instruction using simulation and graphics that must be mentioned is the use of them during assessment projects, tests and exams. Unless students believe that the point of the demonstrations is to convey some examinable understanding, they will not pay sufficient attention. For example, if we use portfolio diversification to illustrate the sampling distribution of the sample mean, we need to include an understanding of portfolio diversification on the examination.

The sharing of software-based instructional materials among instructors is much easier now than in past decades. The free software R (2006) is available for instructors and students, and while its generality complicates its use by students in first courses, the investment of time by instructors can provide benefits quite quickly. Understanding statistics will help instructors learn R!

Conclusion A re-examination of the discipline of statistics suggests that the traditional approach emphasizing parametric estimation and hypothesis testing is becoming anachronistic. A practical approach to curriculum reform is to use computer software for simulation and graphics for demonstrations of the traditional material. These demonstrations can be built into courses whose curriculum follows the traditional pattern. This makes the illustration of some new techniques an incremental step, and time for the new techniques can be found by reducing the detail of the traditional topics.

References:

Agresti,A. and Franklin,C. (2007) *The Art and Science of Learning from Data*. Pearson Prentice Hall, Upper Saddle River, NJ.

Arnholt, A.T. (2007) Resampling with R. Teaching Statistics, 29(1) 21-26

Bryce, G.R. (2005) Developing Tomorrow's Statisticians. Journal of Statistics Education. v13n1. www.amstat.org/publications/jse/v13n1/bryce.html

Christie, D. (2004) Resampling with Excel. Teaching Statistics, 26(1),9 -14

De Veaux R.D., Velleman, P.F. and Bock, D.E. (2008*) Stats: Data and Models*.2nd ed. Pearson Addison Wesley. New York.

Garfield, J, Hogg,B., Schau, C, and Whittinghill, D. First Courses in Statistical Science: The Status of Educational Reform Efforts. Journal of Statistics Education, v10, no. 2. www.amstat.org/publications/jse/v10n2/garfield.html

Love, T.E. (1998) A Project-Driven Second Course. Journal of Statistics Education, v6, no. 1. www.amstat.org/publications/jse/v6n1/love.html

Moore, D.S. (2004) *The Basic Practice of Statistics*. 3rd Ed. W.H.Freeman, New York.

Peck, R. et al (2005) *Statistics: a guide to the unknown*. 4th edition. Thomson, Brooks/Cole. Belmont, Ca.

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Schafer,D.W. and Ramsey,F.L. (2003) Teaching the craft of Data Analysis. Journal of Statistics Education, v11, no. 1. www.amstat.org/publications/jse/v11n1/schafer.html

Snee, R.D. (1993) "What's missing in Statistical Education?" The American Statistician, 37, 284-289.

Taffe, J and Garnham,N.(1996) Resampling, the bootstrap and Minitab. Teaching Statistics, 18(1), 24-25.

Weldon, K.L (2005) Less Parametric Methods in Statistics. Metodolski Zvezki, vol 2, July, 2005

Weldon, K.L. (2007) Stat Ed Resources (BCCUPM) in www.stat.sfu.ca/~weldon.

Wood, M. (2005) The Role of Simulation Approaches in Statistics. . Journal of Statistics Education, v13, no. 3. www.amstat.org/publications/jse/v13n3/wood.html