

Statistical Literacy for Efficient Citizenship

A Blog About Statistical Literacy and Introductory Statistics Instruction

02.12.07

The Evolution from “Statistical Methods” to “Data Analysis”

Posted in [Intro. Stats.](#), [Data Analysis](#) at 4:14 pm by schafers

Like many other statistical educators, I think the introductory statistics course should evolve to attend even more to statistical literacy and understanding statistical reasoning. I might be more on my own, though, when I argue that data analysis proficiency is an unsuitable objective. There are several parts to this argument. The first has to do with history.

Fisher and Snedecor

While studying crop yield data in the 1920s, Ronald Fisher assembled the core of what we now think of as modern statistics. In 1925, he wrote the book *Statistical Methods for Research Workers* for research biologists with statistical problems similar to the ones he faced at Rothamsted agricultural experiment station. The book included one and two-sample t-tests, correlation, chi-squared tests of goodness of fit and independence, analysis of variance, and regression. These are the main topics that one still finds today in a traditional course on statistical “methods.”

The American, George Snedecor, wrote a textbook, *Statistical Methods*, in 1937, which made Fisher’s methods accessible to a wider audience. (Notice his dropping of “for Research Workers” from Fisher’s title.) Successive generations of statistical textbooks since Fisher’s and Snedecor’s have made statistics accessible to increasingly wider audiences. They are wider in two ways: (1) the range of disciplines that use statistics has expanded and (2) the student audience has changed from advanced biological researchers to prospective researchers to first- and second-year college students.

While the introductory textbooks have evolved to meet the wider and lower-level audience, I think the evolution needs to go further to address two important changes in the field of statistics since Fisher’s time: the different way in which statistical data analysis is done and the greater relevance of statistical literacy to ordinary citizens.

Tukey and Interactive Data Analysis

In 1962, the American statistician, John Tukey, wrote an influential paper titled “The Future of Data Analysis.” In response to the wider range of disciplines using statistics and, more particularly, the broader types of data problems that this caused, Tukey made this point: the traditional methods—such as those in Fisher’s textbook—were based on fairly narrow assumptions (such as normality), which were clearly violated in many data problems. While it would be easy to continue offering these methods without thinking about the correctness of their conclusions, the right path for the field of statistics, he argued, was to find ways to produce appropriate conclusions for all data problems, including those that did not fit neatly into the traditional framework.

Statisticians responded to Tukey’s challenge with studies of robustness and resistance, new techniques, and a new way of going about the statistical analysis of data. One instructional complication in this revolution was a new role for the old methods. Traditional t-tests and F-tests were found to give correct or approximately correct answers for many data problems in which their formal assumptions were not met exactly. That’s not to say they could always be used. In applying the two-sample t-test based on normal distributions, as an example, the data analyst must understand that the population distributions can be quite skewed as long as the skewness is about the same in the two populations; that non-constant variance may not be a problem if the sample sizes are equal; that even though the t-test may be robust against skewness, a more efficient use of the data may be possible by

applying the test after a transformation; and that outlier and influence analysis can be used to cope with sensitivity to outliers.

The traditional methods, therefore, remained in the statistical tool kit, but assumed a new role within a broader process of interactive data analysis. Statistical data analysts, today, use their knowledge of robustness and resistance, and a practical exploratory approach for sizing up a situation to choose a proper course of action. The practical execution of statistical data analysis after Tukey became a craft. As Tukey put it, statistics should no longer be taught as “cookbookery” but, rather, as “the art of cooking.”

This view of statistical data analysis as a process, rather than the selection of a method dictated by data structure, has been embraced by the field of statistics. It has not been embraced by other fields in which statistics is taught, particularly mathematics. There tends to be a “data analysis” dimension on which introductory textbooks differ. Publishers describe this dimension as “traditional” on one side and “data analytic” on the other. Traditional textbooks, which are generally favored in mathematics departments, tend to teach statistical methods in the pre-Tukey style. Many of them now include elements of modern data analysis but do not convey the craft-like process of data analysis. Data analytic books, which are more popular in statistics departments, emphasize real data and exploratory methods, and try to convey a plan of attack for the process of drawing conclusions from data.

The modern data analytic approach better shows the actual approach but, as a topic of an introductory course, has problems of its own. I will argue in my next article that data analysis proficiency is an unrealistic and inappropriate goal (as is quasi-proficiency). Furthermore, the dogged attention to data analysis often comes at the expense of missing the bigger picture of statistical reasoning, which students need for evaluating statistical arguments they encounter as citizens and, if applicable, as the appropriate preparation for a serious course in data analysis.

[Permalink](#)

1 Comment »

1. Mohamed Limam said,

May 9, 2007 at [12:31 am](#)

I appreciate what you’ve said and agree with you. Also, we may extend this reasoning to today’s Data Mining processes.

Bye

Leave a Comment

Name (required)

E-mail (required)

URI