MODELING FIRST: A MODELING APPROACH TO TEACHING INTRODUCTORY STATISTICS

James W. Stigler[1] and Ji Y. Son[2]
[1]University of California, Los Angeles
[2]California State University, Los Angeles
stigler@ucla.edu

Students of introductory statistics gain some knowledge of a large assortment of statistical concepts and tests, yet often fail to comprehend the core ideas that link these concepts and procedures together. In order to give students a more coherent view of statistics, and thus a more flexible understanding that can be applied appropriately across a number of situations, we set out to teach the introductory course as modeling, subjugating all other concepts and skills to the General Linear Model. In this paper we report on the initial design of the course and our first experiences implementing these ideas.

INTRODUCTION

Students graduate from our introductory courses armed with a variety of concepts and procedures, but often fail to apply them appropriately to problems that deviate even slightly from those covered in class. The problem, in our view, is our failure to produce coherent understanding of core statistical concepts. Learning of facts and procedures as isolated bits of knowledge not only makes transfer difficult, but also leaves students unprepared for more advanced statistics courses.

Advanced courses in statistics, generally targeted at graduate students, often take a modeling approach, teaching a framework (the General Linear Model) that provides a coherent context for understanding virtually any statistical method. The question that motivates our work is this: Is it possible to start out, from the beginning, teaching statistics as modeling? And would such an approach lead to a more coherent conceptualization of statistics, even for introductory students?

Many would argue that the concepts and methods of modeling are too difficult for the beginning student to understand. As psychologists, our research interest is in how people come to understand complex concepts - things that are hard to learn, that develop slowly over weeks, months, and even years. Statistics, in our view, is such a domain.

In our work on statistics, we are seeking to apply current theories of cognitive and developmental psychology to the design and implementation of an introductory statistics course that starts with modeling. Our goal is to support students' coherent understanding of statistics, to leave them with knowledge they can apply flexibly to novel situations, and to prepare them for more advanced courses in statistics.

We base our approach on the work of others who have envisioned teaching of statistics as modeling, especially Judd, McClelland, and Ryan (2011), Kenny (1987), Rodgers (2010), Wild (2006), and the MOSAIC project (http://mosaic-web.org/).

OVERVIEW OF STATISTICS: WHAT WE WANT OUR STUDENTS TO UNDERSTAND

We frame statistics as the study of variation. If variation didn't exist, then we wouldn't need statistics. If everyone who took a particular drug got well, and everyone who didn't died, we wouldn't need statistics. But that's usually not what happens. Usually some people who take the drug get well, but some don't. Some people who don't take the drug get well anyway. It's not that easy to tell whether the drug really cures the ailment, or if the cure just happened by chance. Statistics is the body of tools and concepts that help us make sense of such situations.

We divide our course into three parts: *exploring variation*, *modeling variation*, and *evaluating models*. These are three main goals statisticians have when analyzing data. Importantly, everything in a

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

traditional introductory statistics course, as well as in more advanced statistics courses, can also be understood in this coherent framework.

Exploring Variation

We start by exploring variation in a bunch of numbers, then discuss where the numbers come from (sampling, measurement, and research design). Using statistical tools such as graphs and frequency tables, we develop the concept of distribution, a core concept and the primary conceptual lens through which we view variation (Wild, 2006). It is a way to see the forest for the trees.

We can learn a lot by examining distributions of data. But our interest usually goes beyond the data, to the Data Generating Process (or DGP). When we examine distributions of data, we do so to help us understand distributions of the DGP. These two kinds of distributions (data and the DGP) make up two-thirds of what we refer to as the Distribution Triad. (Later in the course we will bring in the third kind of distribution, distributions of estimates.)

Because the DGP is unknown and we can't see it directly, we must model it. In this first section of the course we introduce only qualitative models, guessing the shape of the distribution of the DGP, for example. Importantly, we don't introduce concepts such as mean or standard deviation until we formally introduce the concept of a statistical model in the second part of the course.

We do, however, introduce the concept of explaining variation in one variable with another, and spend time discussing the sources of variation in an outcome variable. We also introduce the precursors of statistical notation by having students learn to write "word equations" (e.g., height = sex + other stuff) to represent relationships among variables.

After considerable time building up an informal and intuitive basis for models, we pose the question of how we might quantify models. Quantifying models will help us make predictions and judge the accuracy of those predictions. It also will help us to compare alternative models by quantifying the amount of "other stuff," or unexplained variation, left after explaining variation in an outcome variable.

Modeling Variation

We develop the idea of a statistical model first with a simple model - sometimes called the empty model: the mean of a quantitative variable. The mean that we calculate from our data is a statistic. We use the statistic to estimate a parameter, the mean of the DGP. Statistics are calculated from data; parameters are estimated based on data. Parameters must be estimated because, as we have seen, we have no way to directly measure the DGP.

We explore properties of the mean as a model rather than as a measure of central tendency. Likewise, we approach measures of spread as tools for quantifying error from a statistical model. We can estimate how far off the mean is by calculating the deviation of each data point from the mean, then aggregating these deviations to indicate error from the model (e.g., sum of the absolute deviations, sum of the squared deviations, variance).

In the context of this simple model, we begin to develop in very concrete terms the basic idea behind statistical modeling: DATA = MODEL + ERROR. In the distribution (of data), each score can be expressed as the mean (as a model) and a deviation from the mean (error). We can represent this idea using simple mathematical notation, the notation of the General Linear Model.

The sum of squares as a measure of error is related to the larger goal of statistics: explaining variation in some outcome variable, or, in a complementary fashion, reducing error variation. The mean gives us a place to start because the error has already been reduced as much as possible. Adding more explanatory variables into the model can further reduce this error. The empty model then becomes a point of comparison for evaluating more complex models.

We start by adding a simple grouping variable. Using the notation of the General Linear Model, we can represent this new model either as two group means, or as a grand mean plus the deviation of each group (above or below) the grand mean. The error term is now the sum of squared deviations of scores

from their own group mean. More complex models reduce the error relative to simpler models, something we can quantify as Proportional Reduction in Error (PRE).

Although we can reduce error further by making more and more complex models, we also sacrifice degrees of freedom. Degrees of freedom is, in a sense, the currency of statistical power. We "earn" more degrees of freedom by having a larger sample but "spend" degrees of freedom every time parameters are added to a model. The F ratio corrects for this by quantifying error per degree of freedom.

Once we have developed a model with a grouping variable as the explanatory variable, we follow the same approach to building models that have a quantitative (as opposed to categorical) explanatory variable. We can also begin to imagine building more complex models (e.g., multiple regression) in the same way.

Evaluating Models

From estimating the parameters in our models, we then ask how accurate are our parameter estimates? Clearly, if we had studied a different sample we would have come up with slightly different parameter estimates. Our measures of fit, such as PRE or F, would also be different if we had a different sample. The variation in an estimate is called sampling variation.

Just as interpreting a single score requires us to know about the distribution from which it came, interpreting a statistic (such as a parameter estimate) requires us to know something about the distribution from which it comes. This distribution, which only exists in our imagination, is called a sampling distribution, or the distribution of an estimate.

We can learn something about the variability in samples by simulating sampling distributions given some known DGP. There are predictable patterns that emerge (formalized in the Central Limit Theorem). We also introduce creating sampling distributions through bootstrapping and idealizing them with mathematical models.

A sampling distribution, and more specifically the standard deviation of a sampling distribution (or Standard Error) allows us to reason about our parameter estimates using logic like this: If the DGP has the mean, variance, and shape we assumed in our simulation, then we can calculate the likelihood of getting a random sample with a mean more extreme than a given value. Using similar logic, and working backwards, we can think about possible DGPs that gave rise to a particular sample.

Finally, we apply these concepts to the task of comparing models, ruling out more complex models in favor of simpler ones based on data. We compare models with categorical predictors, quantitative predictors, and also mixed models with a variable of each type.

CONCLUSION

Early implementations of this course at Cal State LA and UCLA suggest that the real benefit of using modeling to teach the coherence of statistics is that it helps students transfer what they have learned to new situations, and extend the models covered in class to more complex situations. For example, an earlier implementation of this course did not teach students how to evaluate the regression model against the empty model, yet students were able to invent ways to do so by the end of the course.

Our experiences convince us that practicing the connections of a small set of core concepts to a variety of situations throughout the course is more successful than the traditional course in which many concepts are introduced, but used less extensively. Further evidence for this view will be presented at the conference.

We are still grappling with a variety of questions as we move forward in developing this course and pedagogy. Is this a complete story for novices? How do we do research and gather evidence on what approaches work better? How do we handle variation in mathematical preparation?

Our current attempts to develop materials have led us to develop learning outcomes that are always nested in a broader goal and to rely on pedagogical materials invented by other statistics educators who emphasize modeling (e.g., Pruim, Kaplan, & Horton, 2017; simulations by Rossman and Chance).

We are also incorporating routines that have been studied by cognitive scientists (e.g., inventing based on contrasting cases, Schwartz et al., 2011).

REFERENCES

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2011). *Data analysis: A model comparison approach.* London, Routledge.

Kenny, D. A. (1987). *Statistics for the social and behavioral sciences.* Boston, Little, Brown.

Pruim, R., Kaplan, D. T., & Horton, N. J. (2017). The Mosaic Package: Helping students to 'think with data' using R. *R Journal*, *9*(1).

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: a quiet methodological revolution. *American Psychologist*, *65*(1), 1.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759.

Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, *5*(2), 10-26.