USING HIERARCHICAL MODELS IN ASSESSMENT OF TEACHING METHODS: AN INITIAL ILLUSTRATION

Jillian Lyon, Ulrike Genschel, and Mark Kaiser
Iowa State University, Ames, IA
jdlyon@iastate.edu

As the field of statistics education continues to grow, new methods and approaches to teaching introductory level statistics are being developed with the goal of improving students' understanding of statistical concepts. While there have been a number of studies showing positive effects for particular teaching methods in small case studies, generalizations to larger populations of students are difficult based on inferential procedures that rely on randomization, replication, and a high level of control over relevant factors. For example, assumptions of independence and identical distributions are typically not justified through randomization of students to treatments and instructors to classes of students. In addition, if one correctly identifies experimental units, many studies are faced with a problem of little or no replication. A related set of difficulties is associated with assumptions of control over factors that may be relevant to responses of student performance. Dealing with potential effects due to individual instructors, characteristics of student groups, and settings, such as time of day classes meet, result in either highly conditional analyses (e.g., one class or two classes) or the need to assume such effects do not exist.

We propose that rather than the experimental concept of controlling all relevant factors other than an active treatment, assessment of educational methods might proceed by mixing over as many levels and specific occurrences of those factors as possible. That is, rather than attempting to control factors at fixed values, we treat factors as random variables and integrate over their distributions. This can be accomplished through the use of hierarchical models.

As an illustration of this type of analysis, we fit a hierarchical model to data on test scores in an introductory statistics class at Iowa State University. The data available included 17 sections, each with an average enrollment of 30 students. Sections were grouped into two lecture periods in each of two semesters, all with a common instructor, but each section met individually with a separate instructor for a lab or recitation period. The model used took student as the observational unit and included levels for section, lecture group, and semester. Student response was taken to be final exam score, and score on a previous exam was used as a covariate in a linear expectation function. The model has the flexibility to incorporate non-constant variances in the observational component. We discuss how one can approach determination of appropriate distributional forms for factors at each level of the hierarchy (data model, section, and lecture group), issues that arise in fitting the model via Markov Chain Monte Carlo methods, and procedures that can be used for model assessment. Inference is achieved through interpretation of joint and marginal posterior distributions, and posterior predictive distributions for the random factors included in the model. We discuss some of the issues important in determining what quantities should be used for the purpose of statistical inference.

In this illustration, it is not of primary interest to investigate whether the two semesters at the top level of our hierarchy differ. Rather, we are attempting to demonstrate the potential of the methodology for educational assessment. In a more relevant application the upper level of the hierarchy would be method of instruction rather than semester. We take this analysis as representing a first step. Situations involving the assessment of pedagogical methods often lack a natural nesting of potentially relevant factors. For example, if a number of instructors each handle a class in a number of semesters, one could consider instructor nested within semester, or semester nested within instructor. Work remains in untangling the issues involved in this, and other, potential complications. Nevertheless, we present our example as a first step in developing hierarchical models that offer meaningful alternatives to existing methodology in the analysis of studies involving educational assessment data.