

THE CONSEQUENCES OF SAMPLING LEVEL ERROR IN EDUCATIONAL RESEARCH: THE USES AND MISUSES OF RATEMYPROFESSOR.COM

Dennis E. Clayson, University of Northern Iowa

Abstract

There is a distinct difference between the utilization of student-level and instructor-level data when studying student interactions with instructors. Instructor-level data is characterized by using the mean responses of students from an entire class. Student-level data uses a single student's response as a case. Instructor-level data is typically utilized when a researcher is interested in how instructors' behavior influences education, while the latter is appropriate when studying student behavior.

Purpose

This study investigated differences in the conclusions that could be found that are dependent upon the level of data which could be utilized in an analysis. The study used data from RateMyProfessors.com (hereafter called RMP) for three reasons. First, this data set was utilized by other publications establishing comparative norms. Second, it is a large and easily accessible source of student/instructor data. Third, there is research to indicate that sites like RMP are analogous to, and consistent with, the SET instruments used by colleges and universities.

Methodology

Random samples were drawn from RMP in three different formats. (1) Ten instructors in each of 54 schools were selected for a total of 540 student evaluations. This sampling procedure is equivalent to going into hundreds of classrooms and giving the SET instrument to only one randomly selected student per class. (2) Ten instructors in each of 54 schools were selected and the average of all students evaluating that instructor was recorded. (3) A third data set was drawn replicating the actual way that SET is administrated in a typical university or college by looking at student-level data for a distinct group of instructors with one additional caveat, the sample size was standardized. The data from the last 20 students to evaluate a randomly selected instructor were recorded. This resulted in 20 responses for 54 instructors for a total of 1,080 individual evaluations.

Results

Inconsistent conclusions can be drawn from the different samples. (1) The student-level sample would indicate that instructors are generally well-liked. It would be concluded that students evaluate male and female instructors the same. The results are not incapable with a learning hypothesis, suggesting that the evaluations are valid measures of teaching effectiveness. (2) The second sample would find that instructors are not be as well-liked as the student-level data would suggest. Male and female instructors are evaluated differently, with males being more helpful and having greater clarity than female instructors. There is no evidence that the instruments measure learning. (3) Instructors are even less liked. Students again rate male instructors higher than female instructors. If this sample is broken down by individual instructor means, the 54 instructors (or their administrators) could come to almost *any* conclusion from the sample.

Discussion

Different methods of sampling what was assumed to be the same population created different conclusions. All of these samples were large, taken at random, and violated no major statistical assumptions. Yet, conclusions drawn from the samples differ widely. The simplest explanation is that student-level and instructor-level samples did not come from the same population, in that the instructor-level data was modified by “ecological” variables that were not assumed to be related to the purpose of the evaluation. If these intervening variables were related to one but not all of the study variables, or if the intervening variables were related to the grouping of individual responses, then Simpson’s Paradox could explain the different findings related to the learning hypothesis, but would also indicate that grouping procedures that combines student averages is inappropriate.

Implications

1. It was not the purpose of this paper to identify what RMP data is actually measuring, but the conclusions do have wider implications.
2. The choice between individual and grouped data can have dramatic influences on the outcome and conclusions of educational research.