Chapter 8
Principles for the Analysis of Large Complex Secondary Databases in Educational Settings

Yen To
The University of Southern Mississippi, USA

Hansel Burley
Texas Tech University, USA

ABSTRACT

A primary feature of institutional research work is prediction. When statistics are used as the primary analysis tool, much of this work depends upon ordinary least squares regression, which assumes that data have one level. However, much of the data in educational research, in general, and in higher education research, in particular, is multilevel or nested. This chapter explores multilevel data analysis, with a focus on exploring issues associated with sampling, weighting, design effects, and analysis of data. Additionally, it emphasizes the importance of considering contextual effects using as a reference large secondary datasets. The chapter will also explore opportunities and challenges presented by these types of data.

BACKGROUND

Like never before, Big State University has towering expectations. Two decades ago, Big State was a large, but sleepy regional university with expectations for its future indistinguishable from the accomplishments of its past. These expectations focused around agriculture, the health sciences, football, and girls basketball. Its cultural outlook was simultaneously Western, Midwestern, and Southern, a mix that distinguished it from other universities. The university community reveled in this unique identity. Further, the uniqueness

DOI: 10.4018/978-1-60960-857-6.ch008

of the university was solidified by its possession of some unusual characteristics for a regional university: it possessed a comprehensive undergraduate outlook that included 11 colleges, a law school, and an all-inclusive health sciences center. Many programs, particularly those in the health sciences center served parts of 4 states, in addition to the university's home state. Still, the areas served are rural and are experiencing shrinking populations. Additionally, the region is plagued by water concerns, so many new industries are dissuaded from moving to the area. Finally, the university is 300 miles from some of the fastest growing metropolitan areas in the nation, so it has had to work hard to get the attention of those who live in these areas.

In the last decade, new leadership desired first a higher profile for the university, then a national research university, or Tier I status. Wisely, this leadership understood the university's potential for good, but also realized that if the university were to survive the problems of the region, it needed to evolve from a quaint regional university to a university known for its discoveries. They wove these desires into the university strategic plan, and this strategic plan was one that did not sit on a shelf and collect dust. The university community actively set the plan in motion. The results of putting the plan in motion have produced some extraordinary results. The university has seen its fall enrollment grow from 25,000 to 33,000, with most new students coming from metropolitan areas outside the region. In addition, the university is moving faster than expected to its announced goal of 40,000 students by the year 2020 . The highest growth has been in the graduate programs, with strong researchers attracting stronger students. It has greatly expanded its production of Ph.D.s, and it has made a concerted effort to hire top researchers. It now has campuses at sites that are 100, 200, and 300 miles away, respectively, and it is a leader in distance education. The university is close to Tier I status, which will mean at least $40 million in additional state dollars, per biennium, should it be so classified. Still, this growth means more scrutiny and accountability from layers of accrediting bodies and state and federal agencies. It means unending reports and self-studies. From faculty and mid-level administrator, these changes have invoked an insatiable desire for data— simple, complex, cross sectional, and longitudinal.

How can this university assure quality programs and services that keep pace with its growth? Will student learning be sacrificed? Who will assess university outcomes? Can the university support a centralized institutional research system? As technology advances, how can the university make sure that its institutional research efforts keep pace?

Like never before, this university needs an agile institutional research office, one with the capacity to use multiple methods in order to examine the complex problems from multiple perspectives. As perspectives get more complex, analyses need to become more sophisticated. One such recognition is that in such a far-flung university, the data are naturally complex—that is, the unit of analysis (i.e., a student or department) is nested within some larger grouping of the data. For example, students are nested within schools and colleges. When the institutional researcher conducts longitudinal studies, time is an additional nesting variable.

The notion of nesting introduces the importance of context; for example, engineering undergraduate students probably experience their educational experiences at the university in a far different fashion from those students in the college of Visual and Performing Arts. From a statistical standpoint, average achievement levels almost certainly vary from college to college, making problematic direct comparisons across colleges on outcome measures like GPA. Therefore, examining program effectiveness by comparing the GPAs of electrical engineering students to those in early childhood education will also most certainly mislead. As Bickel (2007) asserts, values at the individual level and contextual group level

Related Content

Positive Unlabelled Learning for Document Classification
Xiao-Li Li (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1552-1557).*
www.irma-international.org/chapter/positive-unlabelled-learning-document-classification/11026

Data Mining for Model Identification
Diego Liberati (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 438-444).*
www.irma-international.org/chapter/data-mining-model-identification/10857

Mining Email Data
Steffen Bickel (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1262-1267).*
www.irma-international.org/chapter/mining-email-data/10984

Using Dempster-Shafer Theory in Data Mining
Malcolm J. Beynon (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2011-2018).*
www.irma-international.org/chapter/using-dempster-shafer-theory-data/11095

Learning Bayesian Networks
Marco F. Ramoniand Paola Sebastiani (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1124-1128).*
www.irma-international.org/chapter/learning-bayesian-networks/10962