From Hierarchical Linear Models, by Anthony S. Bryk and Stephen W. Raudenbush. Book # 1 in the Sage Series on Advanced Quantitative Techniques in the Social Sciences. 1992. Sage Publications Inc: Newbury Park, California.

1 Introduction

- Hierarchical Data Structure: A Common Phenomenon
- Persistent Dilemmas in the Analysis of Hierarchical Data
- A Brief History of the Development of Statistical Theory for Hierarchical Models
- Applications of Hierarchical Linear Models
- Organization of the Book

Hierarchical Data Structure: A Common Phenomenon

Much social research involves hierarchical data structures. In organizational studies, for example, researchers might investigate how workplace characteristics, such as centralization of decision making, influence worker productivity within these contexts. Both workers and firms are units in the analysis; variables are measured at both levels. Such data have a hierarchical structure with individual workers nested within firms.

In cross-national studies, demographers might examine how differences in national economic development interact with adult educational attainment to influence fertility rates (see, e.g., Mason, Wong, & Entwistle, 1983). Such research combines economic indicators collected at the national level with household information on education and fertility. Both households and countries are units in the research with households nested within countries, and the basic data structure is again hierarchical.

Similar kinds of data occur in developmental research where multiple observations are gathered over time on a set of persons. The repeated measures contain information about each individual's growth trajectory. The psychologist is typically interested in how characteristics of the person, including variations in environmental exposure, influence the shape of

these growth trajectories. For example, Huttenlocher, Haight, Bryk, and Seltzer (1991) investigated how differences among children in exposure to language in the home influenced the development of each child's vocabulary over time. When every person is observed at the same fixed number of time points, it is conventional to view the design as occasions crossed by persons. But when the number and spacing of time points vary from person to person, we may view occasions as nested within persons.

The quantitative synthesis of findings from many studies presents another hierarchical data problem. An investigator may wish to discover how differences in treatment implementations, research methods, subject characteristics, and contexts relate to treatment effect estimates within studies. In this case, subjects are nested within studies. Although the development of methodological techniques for meta-analysis has proceeded independently of work on hierarchical linear models, such models in fact provide a very general statistical framework for these research activities (Raudenbush, 1984a; Raudenbush & Bryk, 1985).

Educational research is often especially challenging because studies of student growth often involve a doubly nested structure of repeated observations within individuals, who are in turn nested within organizational settings. Research on instruction, for example, focuses on the interactions between students and a teacher around specific curricular materials. These interactions usually occur within a classroom setting and are bounded within a single academic year. The research problem has three foci: the individual growth of students over the course of the academic year (or segment of a year), the effects of personal characteristics and individual educational experiences on student learning, and how these relations are in turn influenced by classroom organization and the specific behavior and characteristics of the teacher. Correspondingly, the data have a three-level hierarchical structure. The Level-1 units are the repeated observations over time, which are nested within the Level-2 units of persons, who in turn are nested within the Level-3 units of classrooms or schools.

Persistent Dilemmas in the Analysis of Hierarchical Data

Despite the prevalence of hierarchical structures in behavioral and social research, past studies have often failed to address them adequately in the data analysis. In large part, this neglect has reflected limitations in conventional statistical techniques for the estimation of linear models with nested structures. In social research, these limitations have generated concerns about

aggregation bias, misestimated precision, and the "unit of analysis" problem (see Chapter 5). They have also fostered an impoverished conceptualization, discouraging the formulation of explicit multilevel models with hypotheses about effects occurring at each level and across levels. Similarly, studies of human development have been plagued by "measurement of change" problems (see Chapter 6), which at times have seemed intractable, and have even led some analysts to advise against attempts to directly model growth.

Although these problems of unit of analysis and measuring change have distinct, long-standing, and virtually nonoverlapping literatures, they share a common cause: the inadequacy of traditional statistical techniques for modeling hierarchy. In the past, sophisticated analysts have often been able to find ways to cope at least partially in specific instances. With recent developments in the statistical theory for estimating hierarchical linear models, however, an integrated set of methods now exists that permits efficient estimation for a much wider range of applications.

Even more important from our perspective is that the barriers to use of an explicit hierarchical modeling framework have now been removed. We can now readily pose hypotheses about relations occurring at each level and across levels and also assess the amount of variation at each level. From a substantive perspective, the hierarchical linear model is more homologous with the basic phenomena under study in much behavioral and social research. The applications to date have been encouraging in that they have both afforded an exploration of new questions and provided some empirical results that might otherwise have gone undetected.

A Brief History of the Development of Statistical Theory for Hierarchical Models

The models discussed in this book appear in diverse literatures under a variety of titles. In sociological research, they are often referred to as *multilevel linear models* (cf. Goldstein, 1987; Mason et al., 1983). In biometric applications, the terms *mixed-effects models* and *random-effects models* are common (cf. Elston & Grizzle, 1962; Laird & Ware, 1982). They are also called *random-coefficient regression models* in the econometrics literature (cf. Rosenberg, 1973) and in the statistical literature are often referred to as *covariance components models* (cf. Dempster, Rubin, & Tsutakawa, 1981; Longford, 1987).

We have adopted the term *hierarchical linear models* because it conveys an important structural feature of data that is common in a wide variety of

applications, including studies of growth, organizational effects, and research synthesis. This term was introduced by Lindley and Smith (1972) and Smith (1973) as part of their seminal contribution on Bayesian estimation of linear models. Within this context, Lindley and Smith elaborated a general framework for nested data with complex error structures.

Unfortunately, the Lindley and Smith (1972) contribution languished for a period of time because use of the models required estimation of covariance components in the presence of unbalanced data. With the exception of some very simple problems, no general estimation approach was feasible in the early 1970s. Dempster, Laird, and Rubin's (1977) development of the EM algorithm provided the needed breakthrough: a conceptually feasible and broadly applicable approach to covariance component estimation. Dempster et al. (1981) demonstrated applicability of this approach to hierarchical data structures. Laird and Ware (1982) and Strenio, Weisberg, and Bryk (1983) applied this approach to the study of growth, and Mason et al. (1983) applied it to cross-sectional data with a multilevel structure.

Subsequently, other numerical approaches to covariance component estimation were also offered through use of iteratively reweighted generalized least squares (Goldstein, 1986) and a Fisher scoring algorithm (Longford, 1987). Finally, a number of reasonably sophisticated statistical computing programs have become available for fitting these models including GENMOD (Mason, Anderson, & Hayat, 1988), HLM (Bryk, Raudenbush, Seltzer, & Congdon, 1988), ML2 (Rabash, Prosser, & Goldstein, 1989), and VARCL (Longford, 1988). For a technical review of these computers programs see Kreft, de Leeuw, and Kim (1990).

Applications of Hierarchical Linear Models

As noted above, behavioral and social data commonly have a nested structure, including, for example, repeated observations nested within persons. These persons also may be nested within organizational units such as schools. Further, the organizational units themselves may be nested within communities, within states, and even within countries. With hierarchical linear models, each of the levels in this structure is formally represented by its own submodel. These submodels express relationships among variables within a given level, and specify how variables at one level influence relations occurring at another. Although any number of levels can be represented, the essential statistical features are found in the basic two-level models that are the principal focus of this book.

The applications discussed here address three general research purposes: improved estimation of effects within individual units (e.g., developing an improved estimate of a regression model for an individual school by borrowing strength from the fact that similar estimates exist for other schools), the formulation and testing of hypotheses about cross-level effects (e.g., how varying school size might affect the relationship between social class and academic achievement within schools), and the partitioning of variance and covariance components among levels (e.g., decomposing the correlation among set of student-level variables into within- and between-school components). Published examples of each type are summarized briefly below.

Improved Estimation of Individual Effects

Braun, Jones, Rubin, and Thayer (1983) were concerned about the use of standardized test scores for selecting minority applicants to graduate business schools. Many schools base admissions decisions, in part, on equations that use test scores to predict later academic success. However, because most applicants to most business schools are white, their data dominate the estimated prediction equations. As a result, these equations may not produce an adequate ordering for purposes of selecting minority students.

In principle, a separate equation for minority applicants in each school might be fairer, but difficulties often arise in estimating such equations, because most schools have only a few minority students and thus little data on which to develop reliable predictions. In Braun et al.'s (1983) data on 59 graduate business schools, 14 schools had no minorities, and 20 schools had only one to three minority students. Developing prediction equations for minorities in these schools would have been impossible using standard regression methods. Further, even in the 25 schools with sufficient data to sustain a separate estimation, the minority samples were still small, and as a result, the minority coefficients would have been poorly estimated.

Alternatively, the data could be pooled across all schools, ignoring the nesting of students within schools, but this also poses difficulties. Specifically, because minorities are much more likely to be present in some schools than others, a failure to represent these selection artifacts could bias the estimated prediction coefficients.

Braun et al. (1983) used a hierarchical linear model to resolve this dilemma. By borrowing strength from the entire ensemble of data, they were able to efficiently utilize all of the available information to provide each school with separate prediction equations for whites and minorities. The estimator for each school was actually a weighted composite of the information from that school and the relations that exist in the overall sample.

As one might intuitively expect, the relative weights given each component depend on its precision. The estimation procedure is described in Chapter 3 and illustrated in Chapter 4. For a related application see Rubin (1980), and for a statistical review of these developments see Morris (1983).

Modeling Cross-Level Effects

The second general use of a hierarchical model is to formulate and test hypotheses about how variables measured at one level affect relations occurring at another. Because such cross-level effects are common in behavioral and social research, the modeling framework provides a significant advance over traditional methods.

An Example from Social Research. Mason et al. (1983) examined the effects of maternal education and urban versus rural residence on fertility in 15 countries. It has been well known that in many countries high levels of education and urban residence predict low fertility. However, the investigators reasoned that such effects might depend on characteristics of the countries, including level of national economic development, as indicated by gross national product (GNP), and the intensity of family planning efforts.

Mason et al. (1983) found that higher levels of maternal education were indeed associated with lower fertility rates in all countries. Differences in urban versus rural fertility rates, however, varied across countries with the largest gaps occurring in nations with a high GNP and few organized family planning efforts. The identification of *differentiating effects,* such as this rural-urban gap, and the prediction of their variability is taken up in Chapter 5.

An Example from Developmental Research. Investigators of language development hypothesized that word acquisition depends upon two sources: exposure to appropriate speech and innate differences in ability to learn from such exposure. It is widely assumed that innate differences in ability are largely responsible for observed differences in children's vocabulary development. However, the empirical support for this assumption is weak. Heritability studies have found that parent scores on standardized vocabulary tests account for 10% to 20% of the variance in children's scores on the same tests. Exposure studies have not fared much better. Most of the individual variation in vocabulary acquisition has remained unexplained.

In the past, researchers have relied primarily on two-time point designs to study exposure effects. Children's vocabulary might first be assessed at, say, 14 months of age, and information on maternal verbal ability or language use also might be collected at that time. Children's vocabulary size

then would be reassessed at some subsequent time point, say 26 months. The data would be analyzed with a conventional linear model with the focus on estimating the maternal speech effect on 26-month vocabulary size after controlling for children's "initial ability" (i.e., the 14-month vocabulary size).

Huttenlocher et al. (1991) collected longitudinal data with some children observed on as many as seven occasions between 14 and 26 months. This permitted formulation of an individual vocabulary growth trajectory based on the repeated observations for each child. Each child's growth was characterized by a set of parameters. A second model used information about the children, including the child's sex and amount of maternal speech in the home environment, to predict these growth parameters. The actual formulation and estimation of this model are detailed in Chapter 6. The analysis, based on a hierarchical linear model, found that exposure to language during infancy played a much larger role in vocabulary development than had been reported in previous studies. (In fact, the effects were substantially larger than would have been found had a conventional analysis been performed using just the first and last time points.) This application demonstrates the difficulty associated with drawing valid inferences about correlates of growth from conventional approaches.

Partitioning Variance-Covariance Components

A third use of hierarchical linear models draws on the estimation of variance and covariance components with unbalanced, nested data. For example, educational researchers often wish to study the growth of the individual student learner within the organizational context of classrooms and schools. Formal modeling of such phenomena require use of a three-level model.

Bryk and Raudenbush (1988) illustrated this approach with a small subsample of longitudinal data from the Sustaining Effects Study (Carter, 1984). They used mathematics achievement data from 618 students in 86 schools measured on five occasions between Grade 1 and Grade 3. They began with an individual growth (or repeated measures) model of the academic achievement for each child within each school. The three-level approach enabled a decomposition of the variation in these individual growth trajectories into within- and between-school components. The details of this application are discussed in Chapter 8.

The results were startling—83% of the variance in growth rates was between schools. In contrast, only about 14% of the variance in initial status was between schools, which is consistent with results typically encountered in cross-sectional studies of school effects. This analysis identified substantial

differences among schools that conventional models would not have detected, because such analyses do not allow the partitioning of learning-rate variance into within- and between-school components.

Organization of the Book

The book is organized into nine additional chapters. We have attempted to keep the presentation as nontechnical as possible, relying primarily on simple examples to explicate the basic features of the models. Although the book is primarily intended for behavioral and social scientists interested in using these methods in their research, we have also attempted to provide sufficient detail for the more methodologically oriented reader.

Chapter 2 introduces the logic of hierarchical linear models by building on simpler concepts from regression analysis and random-effects analysis of variance. In Chapter 3, we summarize the basic procedures for estimation and inference used with these models. The emphasis is on an intuitive introduction with a minimal level of mathematical sophistication. These procedures are then illustrated in Chapter 4 as we work through the estimation of an increasingly complex set of models with a common data set.

Applications of hierarchical linear models follow in the next four chapters. Chapters 5 and 6 consider the use of two-level models in the study of organizational effects and individual growth, respectively. Chapter 7 discusses use in the context of research synthesis or meta-analysis applications. This actually represents a special class of applications where the Level-1 variances are known. Chapter 8 introduces the three-level model and describes a range of applications including a key educational research problem—how school organization influences student learning over time.

The last two chapters further discuss the logic of statistical inference in hierarchical linear models. Chapter 9 reviews the basic model assumptions, describes procedures for examining the validity of these assumptions, and discusses what is known about sensitivity of inferences to violation of these various assumptions. Chapter 10 is intended for a more methodological audience desiring a formal derivation of the statistical methods used in this book.