Series Editor's Introduction to Hierarchical Linear Models

In the social sciences, data structures are often hierarchical in the following sense: We have variables describing individuals, but the individuals also are grouped into larger units, each unit consisting of a number of individuals. We also have variables describing these higher order units.

The leading example is, perhaps, in education. Students are grouped in classes. We have variables describing students and variables describing classes. It is possible that the variables describing classes are aggregated student variables, such as number of students or average socioeconomic status. But the class variables could also describe the teacher (if the class has only one teacher) or the classroom (if the class always meets in the same room). Moreover, in this particular example, further hierarchical structure often occurs quite naturally. Classes are grouped in schools, schools in school districts, and so on. We may have variables describing school districts and variables describing schools (teaching style, school building, neighborhood, and so on).

Once we have discovered this one example of a hierarchical data structure, we see many of them. They occur naturally in geography and (regional) economics. In a sense, one of the basic problems of sociology is to relate properties of individuals and properties of groups and structures in which the individuals function. In the same way, in economics there is the problem of relating the micro and the macro levels. Moreover, many repeated measurements are hierarchical. If we follow individuals over time, then the measurements for any particular individual are a group, in the same way as the school class is a group. If each interviewer interviews a group of interviewees, then the interviewers are the higher level. Thinking about these hierarchical structures a bit longer inevitably leads to the conclusion that many, if not most, social science data have this nested or hierarchical structure.

The next step, after realizing how important hierarchical data are, is to think of ways in which statistical techniques should take this hierarchical structure into account. There are two obvious procedures that have been somewhat discredited. The first is to disaggregate all higher order variables to the individual level. Teacher, class, and school characteristics are all assigned to the individual, and the analysis is done on the individual level. The problem with this approach is that if we know that students are in the same class, then we also know that they have the same value on each of the class variables. Thus we cannot use the assumption of independence of observations that is basic for the classical statistical techniques. The other alternative is to aggregate the individual-level variables to the higher level and do the analysis on the higher level. Thus we aggregate student characteristics over classes and do a class analysis, perhaps weighted with class size. The main problem here is that we throw away all the within-group information, which may be as much as 80% or 90% of the total variation before we start the analysis. As a consequence, relations between aggregated variables are often much stronger, and they can be very different from the relation between the nonaggregate variables. Thus we waste information, and we distort interpretation if we try to interpret the aggregate analysis on the individual level. Thus aggregating and disaggregating are both unsatisfactory.

If we limit ourselves to traditional linear model analysis, we know that the basic assumptions are linearity, normality, homoscedasticity, and independence. We would like to maintain the first two, but the last two (especially the independence assumption) should be adapted. The general idea behind such adaptations is that individuals in the same group are closer or more similiar than individuals in different groups. Thus students in different classes can be independent, but students in the same class share values on many more variables. Some of these variables will not be observed, which means that they vanish into the error term of the linear model, causing correlation between disturbances. This idea can be formalized by using variance component models. The disturbances have a group and an individual component. Individual components are all independent; group components are independent between groups but perfectly correlated within groups. Some groups might be more homogeneous than other groups, which means that the variance of the group components can differ.

There is a slightly different way to formalize this idea. We can suppose that each of the groups has a different regression model, in the simple regression case with its own intercept and its own slope. Because groups are also sampled, we then can make the assumption that the intercepts and slopes are a random sample from a population of group intercepts and slopes. This defines random-coefficient regression models. If we assume this for the intercepts only, and we let all slopes be the same, we are in the variance-component situation discussed in the previous paragraph. If the slopes vary randomly as well, we have a more complicated class of models in which the covariances of the disturbances depend on the values of the individual-level predictors.

In random-coefficient regression models, there is still no possibility to incorporate higher level variables, describing classes or schools. For this we need multilevel models, in which the group-level model is again a linear model. Thus we assume that the slope of the student variable SAT depends linearly on the class variables of class size or teacher philosophy. There are linear models on both levels, and if there are more levels, there are more nested linear models. Thus we arrive at a class of models that takes hierarchical structure into account and that makes it possible to incorporate variables from all levels.

Until about 10 years ago, fitting such models was technically not possible. Then, roughly at the same time, techniques and computer programs were published by Aitkin and Longford, Goldstein and co-workers, and Raudenbush and Bryk. The program HLM, by Bryk and Raudenbush, was the friendliest and most polished of these products, and in rapid succession a number of convincing and interesting examples were published. In this book, Bryk and Raudenbush describe the model, the algorithm, the program, and the examples in great detail. I think such a complete treatment of this class of techniques is both important and timely. Hierarchical linear models, or multilevel models, are certainly not a solution to all the data analysis problems of the social sciences. For this they are far too limited, because they are still based on the assumptions of linearity and normality, and because they still study the relatively simple regression structure in which a single variable depends on a number of others. Nevertheless, technically they are a big step ahead of the aggregation and disaggregation methods, mainly because they are statistically correct and do not waste information.

I think the main gain, illustrated nicely in this book by the extensive analysis of the examples, is conceptual. The models for the various levels are nicely separated, without being completely disjointed. One can think about the possible mechanisms on each of the levels separately and then join the separate models in a joint analysis. In educational research, as well as in geography, sociology, and economics, these techniques will gain in importance in the next few years, until they also run into their natural limitations. To avoid these limitations, they will be extended (and have been extended) to more levels, multivariate data, path-analysis models, latent variables, nominal-dependent variables, generalized linear models, and so on. Social statisticians will be able to do more extensive modeling, and they will be able to choose from a much larger class of models. If they are able to build up the necessary prior information to make a rational choice from the model class, then they can expect more power and precision. It is a good idea to keep this in the back of your mind as you use this book to explore this new exciting class of techniques.

JAN DE LEEUW
SERIES EDITOR