Advantages of Hierarchical Linear Modeling

Jason W. Osborne
University of Oklahoma

Hierarchical, or nested, data structures are common throughout many areas of research. However, until recently there has not been any appropriate technique for analyzing these types of data. Now, with several user-friendly software programs available, and some more readable texts and treatments on the topic, researchers need to be aware of the issue, and how it should be dealt with. The goal of this paper is to introduce the problem, how it is dealt with appropriately, and to provide examples of the pitfalls of not doing appropriate analyses.

What is a Hierarchical Data Structure?

People (and other living creatures, for that matter) tend to exist within organizational structures, such as families, schools, business organizations, churches, towns, states, and countries. In education, students exist within a hierarchical social structure that can include family, peer group, classroom, grade level, school, school district, state, and country. Workers exist within production or skill units, businesses, and sectors of the economy, as well as geographic regions. Health care workers and patients exist within households and families, medical practices and facilities (a doctor's practice, or hospital, e.g.), counties, states, and countries. Many other communities exhibit hierarchical data structures as well.

Bryk and Raudenbush (1992) also discuss two other types of data hierarchies that are less obvious: repeated-measures data and meta-analytic data. Once one begins looking for hierarchies in data, it becomes obvious that data repeatedly gathered on an individual is hierarchical, as all the observations are nested within individuals. While there are other adequate procedures for dealing with this sort of data, the assumptions relating to them are rigorous, whereas procedures relating to hierarchical modeling require fewer assumptions. Also, when researchers are engaged in the task of meta-analysis, or analysis of a large number of existing studies, it should become clear that subjects, results, procedures, and experimenters are nested within experiment. While this paper will not delve into these issues further, readers are encouraged to refer to Bryk and Raudenbush (1992) for further discussion of the advantages of hierarchical analysis for these types of data.

Why is a Hierarchical Data Structure an Issue?

Hierarchical, or nested, data present several problems for analysis. First, people or creatures that exist within hierarchies tend to be more similar to each other than people randomly sampled from the entire population. For example, students in a particular third-grade classroom are more similar to each other than to students randomly sampled from the school district as a whole, or from the national population of third-graders. This is because students are not randomly assigned to classrooms from the population, but rather are assigned to schools based on geographic factors. Thus, students within a particular classroom tend to come from a community or community segment that is more homogeneous in terms of morals and values, family background, socio-economic status, race or ethnicity, religion, and even educational preparation than the population as a whole. Further, students within a particular classroom share the experience of being in the same environment-- the same teacher, physical environment, and similar experiences, which may lead to increased homogeneity over time..

The problem of independence of observations. This discussion could be applied to any level of nesting, such as the family, the school district, county, state, or even country. Based on this discussion, we can assert that individuals who are drawn from an institution, such as a classroom, school, business, or health care unit, will be more homogeneous than if individuals were randomly sampled from a larger population. Herein lies the first issue for analysis of this sort of data. Because these individuals tend to share certain characteristics (environmental, background, experiential, demographic, or otherwise), observations based on these individuals are not fully independent. However, most analytic techniques require independence of observations as a primary assumption for the analysis. Because this assumption is violated in the presence of hierarchical data, ordinary least squares regression produces standard errors that are too small (unless these so-called design effects are incorporated into the analysis). In turn, this leads to a higher probability of rejection of a null hypothesis than if: (a) an appropriate statistical analysis were performed, or (b) the data included truly independent observations.

The problem of how to deal with cross-level data. Going back to the example of our third-grade classroom, it is often the case that a researcher is interested in understanding how environmental variables (e.g., teaching style, teacher behaviors, class size, class composition, district policies or funding, or even state or national variables, etc.) affect individual outcomes (e.g., achievement, attitudes, retention, etc.). But given that outcomes are gathered at the individual level, and other variables at classroom, school, district, state, or nation level, the question arises as to what the unit of analysis should be, and how to deal with the cross-level nature of the data.

One strategy would be to assign classroom or teacher (or school, district, or other) characteristics to all students (i.e., to bring the higher-level variables down to the student level). The problem with this approach, again, is non-independence of observations, as all students within a particular classroom assume identical scores on a variable.

Another way to deal with this issue would be to aggregate up to the level of the classroom, school, district, etc. Thus, we could talk about the effect of teacher or classroom characteristics on average classroom achievement. However, there are several issues with this approach, including: (a) that much (up to 80-90%) of the individual variability on the outcome variable is lost, which can lead to dramatic under- or over-estimation of observed relationships between variables (Bryk & Raudenbush, 1992), and (b) the outcome variable changes significantly and substantively from individual achievement to average classroom achievement.

Aside from these problems, both these strategies prevent the researcher from disentangling individual and group effects on the outcome of interest. As neither one of these approaches is satisfactory, the third approach, that of hierarchical modeling, becomes necessary.

How do Hierarchical Models Work? A Brief Primer

The goal of this paper is to introduce the concept of hierarchical modeling, and explicate the need for the procedure. It cannot fully communicate the nuances and procedures needed to actually perform a hierarchical analysis. The reader is encouraged to refer to Bryk and Raudenbush (1992) and the other suggested readings for a full explanation of the conceptual and methodological details of hierarchical modeling.

The basic concept behind hierarchical modeling is similar to that of OLS regression. On the base level (usually the individual level, referred to here as level 1), the analysis is similar to that of OLS regression: an outcome variable is predicted as a function of a linear combination of one or more level 1 variables, plus an intercept, as so:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_1 + \ldots + \beta_{kj}X_k + r_{ij}$$

where β_{0j} represents the intercept of group j, β_{1j} represents the slope of variable X_1 of group j, and r_{ij} represents the residual for individual i within group j. On subsequent levels, the level 1 slope(s) and intercept become dependent variables being predicted from level 2 variables:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_1 + \ldots + \gamma_{0k}W_k + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_1 + \ldots + \gamma_{1k}W_k + u_{1j}$$

and so forth, where γ_{00} and γ_{10} are intercepts, and γ_{01} and γ_{11} represent slopes predicting β_{0j} and β_{1j} respectively from variable W_1. Through this process, we accurately model the effects of level 1 variables on the outcome, and the effects of level 2 variables on the outcome. In addition, as we are predicting slopes as well as intercepts (means), we can model cross-level interactions, whereby we can attempt to understand what explains differences in the relationship between level 1 variables and the outcome. This will be discussed a bit more below.

An Empirical Comparison of the Three Approaches to Analyzing Hierarchical Data

To illustrate the outcomes achieved by each of the three possible analytic strategies for dealing with hierarchical data, disaggregation (bringing level 2 data down to level 1), aggregation, and multilevel modeling, data were drawn from the National Education Longitudinal Survey of 1988. This data set contains data on a representative sample of approximately 28,000 eighth graders in the United States at a variety of levels, including individual, family, teacher, and school. The analysis we performed predicted composite achievement test scores (math, reading combined) from student socioeconomic status (family SES), student locus of control (LOCUS), the percent of students in the school who are members of racial or ethnic minority groups (%MINORITY), and the percent of students in a school who receive free lunch (%LUNCH). Achievement is our outcome, SES and LOCUS are level 1 predictors, and %MINORITY and %LUNCH are level 2 indicators of school environment. In general, SES and LOCUS are expected to be positively related to achievement, and %MINORITY and %LUNCH are expected to be negatively related to achievement. In these analyses, 995 of a possible 1004 schools were represented (the remaining nine were removed due to insufficient data).

Disaggregated analysis. In order to perform the disaggregated analysis, the level 2 values were assigned to all individual students within a particular school (which is how the NELS data set comes). A standard multiple regression

was performed via SPSS entering all predictor variables simultaneously. The resulting model was significant, with R=.56, R-square=.32, $\underline{F}$ (4,22899)=2648.54, $\underline{p} < .0001$. The individual regression weights and significance tests are presented in Table 1.

Table 1
Comparison of three analytic strategies.

Variable:	Disaggregated			Aggregated			Hierarchical		
	B	SE	t	B	SE	t	B	SE	t
SES	4.97a	.08	62.11**	7.28 b	.26	27.91**	4.07 c	.10	41.29**
LOCUS	2.96 a	.08	37.71**	4.97 b	.49	10.22**	2.82 a	.08	35.74**
%MINORITY	-0.45 a	.03	-15.53**	-0.40 a	.06	-8.76**	-0.59 b	.07	-8.73**
%LUNCH	-0.43 a	.03	-13.50**	0.03 b	.05	0.59	-1.32 c	.07	-19.17**

Note: B refers to an unstandardized regression coefficient, and is used for the HLM analysis to represent the unstandardized regression coefficients produced therein, even though these are commonly labeled as betas and gammas. SE refers to standard error. Bs with different subscripts were found to be significantly different from other Bs within the row at p< .05. ** p < .0001.

All four variables were significant predictors of student achievement. As expected, SES and LOCUS were positively related to achievement, while %MINORITY and %LUNCH were negatively related.

Aggregated analysis. In order to perform the aggregated analysis, all level 1 variables (achievement, LOCUS, SES) were aggregated up to the school level (level 2) by averaging. A standard multiple regression was performed via SPSS entering all predictor variables simultaneously. The resulting model was significant, with R=.87, R-square=.75, $\underline{F}$ (4,999)=746.41, $\underline{p} < .0001$. As seen in Table 1, both average SES and average LOCUS were significantly positively related to achievement, and %MINORITY was negatively related. In this analysis, %LUNCH was not a significant predictor of average achievement.

Multilevel analysis. In order to perform the multilevel analysis, a true multilevel analysis was performed via HLM, in which the respective level 1 and level 2 variables were specified appropriately. Note also that all level 1 predictors were centered at the group mean, and all level 2 predictors were centered at the grand mean. The resulting model demonstrated goodness of fit (Chi-square for change in model fit =4231.39, 5 df, $\underline{p} < .0001$). this analysis reveals significant positive relationships between achievement and the level 1 predictors (SES and LOCUS), and strong negative relationships between achievement and the level 2 predictors (%MINORITY and %LUNCH). Further, the analysis revealed significant interactions between SES and both level 2 predictors, indicating that the slope for SES gets weaker as %LUNCH and as %MINORITY increases. Also, there was an interaction between LOCUS and %MINORITY, indicating that as %MINORITY increases, the slope for LOCUS weakens. There is no clearly equivalent analogue to R and R-square available in HLM.

Comparison of the Three Analytic Strategies and Conclusions

For the purposes of this discussion, we will assume that the third analysis represents the best estimate of what the "true" relationships are between the predictors and the outcome. Unstandardized regression coefficients (Bs in OLS, betas and gammas in HLM) were compared statistically via procedures outlined in Cohen and Cohen (1983).

In examining what is probably the most common analytic strategy for dealing with data such as these, the disaggregated analysis provided the best estimates of the level 1 effects in an OLS analysis. However, it significantly overestimated the effect of SES, and significantly and substantially underestimated the effects of the level 2 effects. The standard errors in this analysis are generally lower than they should be, particularly for the level 2 variables.

In comparison, the aggregated analysis overestimated the multiple correlation by more than 100%, overestimated the the regression slope for SES by 79% and for LOCUS by 76%, and underestimated the slopes for %MINORITY by 32% and for %LUNCH by 98%.

These analyses reveal the need for multilevel analysis of multilevel data. Neither OLS analysis accurately modeled the true relationships between the outcome and the predictors. Additionally, HLM analyses provide other benefits, such as easy modeling of cross-level interactions, which allows for more interesting questions to be asked of the data. With nested and hierarchical data common in the social and other sciences, and with recent developments making HLM software packages more user-friendly and accessible, it is important for researchers in all fields to become acquainted with these procedures.

ADDITIONAL READINGS

Bryk, A.S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, 20 (2), 115-147.

Hoffman, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, 24(5), 623-641.

Nezlek, J. B., & Zyzniewski, L. E. (1998). Using hierarchical linear modeling to analyze grouped data. *Group Dynamics*, 2, 313-320.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research,* (pp. 675-711). Harcourt Brace: Orlando, FL.

Raudenbush, S. W. (1995). Reexamining, reaffirming, and improving application of hierarchical models. *Journal of Educational and Behavioral Statistics*, 20 (2), 210-220.

AUTHOR NOTES