ASK, 2007, nr 16, strony 55-66
 ISSN 1234-9224

Katarzyna Wilk
Yale University

HIERARCHICAL MODELING: PROPERTIES AND APPLICATION*

Empirical analyses of hierarchical data[1] are important in various disciplines, but are most common to the social sciences. Until the 1980's, when the method of multilevel modeling was introduced, researchers dealt with the problem of nested data in a variety of ways, none of which was completely effective or accurate. The method of hierarchical modeling, and softwares such as HLM or MLwiN, provide the most appropriate available tools for dealing with the nested data. This article intends to introduce this strategy, as well as provide an empirical example to illustrate the relative advantages of using it to perform analysis.
Key words: hierarchical modeling, hierarchical modeling software, multilevel data, nested data, non-independent units of observation.

The most appropriate method for analyzing nested data, which represents various social, organizational or institutional structures, is hierarchical modeling. This type of modeling allows for a simultaneous analysis of both aggregated data and non-aggregated data (in most cases of an individual type), and avoids statistical pitfalls. Such software programs as HLM (Hierarchical Linear Modeling), ML (Multilevel Modeling), R and STATA are most commonly used in statistical analysis of data with hierarchical structure[2]. The primary advantage of using these softwares is that they automatically take into account and correct

* I am indebted to Ivan Szelenyi for his mentoring. I also thank Henryk Domanski and Kazimierz Slomczynski for their helpful comments. Please direct all correspondence to Katarzyna Wilk, Sociology Department, Yale University, 140 Prospect St, New Haven, CT 06511, USA. E-mail: katarzyna.wilk@yale.edu.

[1] In general, the following terms "hierarchical data", "nested data" and "cross-level data" are interchangeable. Terms "hierarchical modeling" and "multilevel modeling" are also interchangeable.

[2] SPPS and SAS enable the analysis of data with hierarchical structure; however, in this case it is more complicated since it also requires the analysis of residuals.

various errors and biases that are generated when the aggregated data is applied to individuals[3].

Initially, hierarchical models were applied to the analysis of students' achievements and their determinants, such as school location and various characteristics of teachers (Goldstein 1987, 1995, 1997; Gray et al 1995; Bosker 1998). Next, hierarchical modeling was applied in political science to analyses of political attitudes and behaviors (Hurwitz and Peffley 1987, Wilk 2007). Later, in the public health field, its use led to more precise conclusions regarding the impact of social environment on individual health (Greenland 1997). In criminology, it enabled researchers to show various crime rates to be dependent on neighborhood characteristics (Browning, Feinberg, Dietz 2004). In short, hierarchical modeling is becoming more and more popular as a method of dealing with data showing nested structure in such disciplines as sociology, political science and criminology.

The purpose of this article is to introduce the method of hierarchical modeling. In the introductory section, potential problems related to the use of data with nested structure are discussed, with reference to the advantages of hierarchical modeling. This section also presents an example of a hierarchical model depicting the impact of contextual characteristics on students' achievements. The following section provides a brief description of two software programs, HLM and ML that make hierarchical modeling possible. This section is followed by a presentation of three different approaches to analysis of data with hierarchical structure. In two cases, the author uses the regression analysis and SPSS, while in the third example, hierarchical modeling is performed by the HLM software. Using those three examples, the article discusses the various outcomes that result from the use of two different softwares and three different methods, ultimately demonstrating the advantage of analyzing this multilevel data through hierarchical modeling.[4] Conclusions follow in the final section.

[3] More discussion of potential statistical problems related to the use of hierarchical data can be found in the section titled: "Restrictions in Multilevel Data Analysis."

[4] For more information about how to use the HLM software (*Hierarchical Linear Modeling*) see the manual by Raunderbush, S. W., A. S. Bryk, Y. F. Cheong, R. T. Congdon. 2002. *HLM, 5 Hierarchical Linear and Nonlinear Modeling, Scientific Software International.*

Types of Hierarchical Models

In general, there are three categories of analysis for models with hierarchical structure: analysis of multilevel data, of longitudinal data, and data collected by different interviewers.

1) analysis of multilevel data

In this case, there are two or more levels of analysis. Those consist of micro-level, which includes an individual/basic level variables (most often the unit of observation is an individual); mezo-level, including group level data (e.g. school, cohort, neighborhood); and macro-level (e.g. districts, voivodships[5], regions).

2) analysis of longitudinal data

In this case, the data collected from each wave of panel study is viewed as one level of analysis. The number of panel waves thus corresponds to the number of levels of analysis.

3) analysis of data collected by different interviewers

In this case, the data collected by each interviewer is treated as a separate level of analysis. The number of interviewers thus corresponds to the number of the levels of analysis (Goldstein, 1995).

This article focuses on the first type of analysis – multilevel data.

Multilevel Data

In practice, multilevel data is obtained through at least two levels of observation, each defined by the particular social, institutional or organizational structure under observation. The nested structure of data, in other words, represents various levels of analysis. In such models, level 1 data most frequently describes individuals, while level 2, which includes aggregated data, characterizes the phenomenon being analyzed, such as region or neighborhood. Most commonly, the dependent variable is defined on level 1 – an individual level, while the independent variable describes either the analyzed phenomenon (e.g. regional macroeconomic conditions), or an individual (e.g. socio-demographic variables).

For example, in the analysis of determinants for elementary school student achievement, students are nested in the educational system. This structure

[5] Voivodship is a local administrative unit in Poland.

includes: school grade/class, schools, school districts, and voivodships/states. Here, student characteristics are nested within schools while schools are nested within districts, which in turn belong to voivodships/states. In a similar manner, one can view and analyze characteristics of individuals employed in a certain industry; i.e., describing them according to their particular skills, followed by characteristics of the company, and then in turn by characteristics of the voivodship where the job is located. Or, to use another example, individual political opinions can be explained by certain individual characteristics, which are nested, for instance, within regions or countries[6]. These examples serve as a brief overview of the multilevel structure of data. However there are certain restrictions inherent to the use of nested data, which will now be discussed.

Restrictions in Multilevel Data Analysis

The researcher, who is applying empirical analysis to hierarchical data, including both individual and aggregated type of variables, faces some potential pitfalls.

First, the incorrect application of both individual and aggregated data into one statistical model may lead to ecological fallacy. Ecological fallacy means that individual relationships were incorrectly inferred from group-based relationships. In order to avoid this problem, the researcher can limit his/her analysis to individual or aggregated data only. This leads to another concern, however. In the event that the researcher limits analysis to individual data, which only reflects the specific context of each available respondents[7] one should be cautious because the relationship between individual-level variables does not necessarily translate into the relationship between aggregated units of observation. Similarly, when aggregated variables are used alone, one can only make conclusions about the relationship between the aggregated units of observations, not the actual relationship between individuals and their larger environmental or demographic context. In other words, it is incorrect to conclude that the impact of individual level variables on the aggregated level is significant or insignificant (Przeworski and Teune 1970).

[6] For more elaborate examples of how attitudes towards the EU are hierarchically structured, you may see Wilk, K. 2007. Disseration Manuscript, Yale University.

[7] Individual data does not provide objective contextual information.

Another problem when dealing with hierarchical data is dependency of observations. This refers to the fact that individuals belonging to the same group tend to be more similar to each other than to individuals randomly selected from the entire population. Thus, for example, fourth grade students who are not randomly assigned to schools but assigned based on their place of residence, tend to be more similar to each other socio-demographically and psychologically than to fourth grade students who were randomly selected. Moreover, since those students come from a more homogenous population, receive similar experience and knowledge, and are influenced by the same teachers, it is reasonable to assume that their similarities will tend to strengthen over the time. Therefore, observations in this particular example are not truly independent. Such non-independent observations and correlated errors lead to results bearing significant statistical errors.

When data reveals a hierarchical structure, the problem of non-independent units of observation and correlated errors may be present not only on an individual level, but at any level of analysis. For instance, in the case of the fourth-graders, not only does this dependency of observation apply to each individual student, but also to each fourth grade class and voivodship. Therefore, in analyzing multilevel data that includes both an individual level (student characteristics) and macro-level (school grade, school, district and state), it is essential to define precisely what is being analyzed. In this example, then, we must clarify whether the unit of analysis is (one fourth grader) or a group (a fourth grade class, a school) and then, to correct for the non-independence factor.

Hierarchical modeling, and its associated softwares, account for and correct the problems of ecological fallacy and dependency of observations discussed above.. In contrast, the use of OLS regression[8] (*Ordinary Least Square Regression*) ignores the dependence of observations factor, leading to an underestimation of standard error. As a result, the probability of rejection of the null hypothesis increases. Hierarchical modeling, then, would appear to be a superior method of analysis; however, it is possible only if the following conditions are fulfilled.

---

[8] OLS estimation is based on a minimal sum of squared residuals. For more information, please go to www.hec.unil.ch/schmidheiny/sea2/ols2up.pdf

Hierarchical Modeling Conditions

As with standard linear modeling, hierarchical linear modeling requires that the dependent variable be continuous. This is not the case, however, for the non-linear hierarchical model. Non-linear modeling becomes necessary when the dependent variable is of a nominal[9] or dichotomous[10] type. In addition, in the case of both linear and non-linear hierarchical modeling, the dependent variable must have a significant variance in relation to contextual variables.

Another necessary condition is that each data set representing the corresponding levels of analysis needs to include a variable (matching variable) that allows the merging of two or more separate data sets. This would result in a new data set displaying the hierarchical structure. The following section depicts one example of a two-level hierarchical model.

Two-Level Model Example

The two-level model includes two submodels: level 1 and level 2.. For instance, when student achievement is analyzed, the level 1 variables include student characteristics i=1,...n (student), while level 2 variables include information on macro-context – in this case, school characteristics.

The relationship between level 1 variables can be summarized in the following manner:

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

B_{0j} is an intercept of group j, B_{1j} is the regression coefficient of variable X_1 of group j, a and r_{ij} are the residuals for an individual i within group j (random impact of level 1 variables). On the following levels of analysis the level 1 intercept and coefficient (B_{0j}, B_{1j}) become the dependent variables, which are

[9] The nominal variable classifies data into mutually exclusive categories, which involves assigning labels to categories in order of frequency of occurrence (Runyon and Haber 1976). For instance, the place of residence can be represented by a nominal variable.

[10] The dichotomous variable is the simplest type of variable that represents a binominal distribution. The number of values in this case, equals two; most often, it is a zero-one representation of a phenomenon or its quality. For instance, gender can be coded as a dichotomous variable in the following manner: woman=1, man=0. In this case, one needs to apply the Bernoulli regression in hierarchical modeling. For more information regarding the Bernoulli regression, see *Raunderbush*, S. W., A. S. Bryk, Y. F. Cheong, R. T. Congdon. 2002. *HLM, 5 Hierarchical Linear and Nonlinear Modeling, Scientific Software International.*

predicted based on level 2 variables. The level 2 data describes the individual's environment (i.e. the school's impact on a student's achievement), where j=1...j are subsequent units of this environment (i.e. schools). This relationship can be summarized in the following manner:

$$B_{0j} = \gamma_{00} + \gamma_{01}W_1 + ... + \gamma_{0k}W_k + u_{0j}$$
$$B_{1j} = \gamma_{10} + \gamma_{11}W_1 + ... + \gamma_{1k}W_k + u_{1j}$$

$\gamma_{00,}\gamma_{10}$ are intercepts, and $\gamma_{01,}\gamma_{11}$ are coefficients predicting the coefficients B_{0j} and B_{1j} based on variable W1. U_{0j} and U_{1j} are statistical errors generated at level 2.

To summarize, hierarchical modeling enables one to precisely analyze the impact of level 1 and level 2 variables on the dependent variable.

Hierarchical Modeling Softwares: HLM and MLwiN

The following section presents two hierarchical modeling softwares: HLM 6.0[11]and MLwiN[12]. Both programs are user-friendly; however, HLM is especially helpful in guiding the researcher in the step-by step process of building hierarchical models. The major difference between HLM and ML is the fact that HLM analyzes both conditional and non-conditional models, while ML can only be applied to conditional models. One advantage of MLwiN is the interface, which allows one to save subsequent commands. This is particularly important when analyzing more complex multi-level data, for instance, panel data.

HLM 6.0 has some functions for estimating multilevel models, including random and fixed-effect models, as well as mixed models, and multinomial or ordinal models for three level data sets that may or may not include missing values. The software additionally provides for a complex analysis of the latent dependent variable and estimation of the impact of independent variables upon it. It also provides expanded graphics for fitted models, and easier automated input from other software packages such as SPSS, SAS and STATA.

The following section presents an example of an analysis performed with HLM, as well as two other methods.

---

[11] HLM 6.0 is the newest version of the software. There is a free HLM download of the student version. Although the student version does not allow one to save syntax, it is extremely useful for the first step of getting acquainted with the software. In addition, it is possible to buy HLM software for a period of 6 to 12 months, or to rent it for 15 days. For more information, go to http://www.ssicentral.com/hlm/hlm.htm or http://www.ssicentral.com/ rental.htm

[12] The MlwiN manual and training version are available at the Center for Multilevel Modeling website. For more information, go to http://multilevel.ioe.ac.uk.

Analyzing Hierarchical Data: Three Approaches

The three methods of hierarchical data analysis presented here are based on the analysis performed by Jason Osborne (Osborne 2000), in which multilevel data from the 1998 National Education Longitudinal Survey is used. This data set includes a representative sample of US eighth grade students, and includes individual data on students' socio-economic status (individual level data), and aggregated data on teachers and schools (N(students)=28000, N(schools)=995). This analysis attempts to explain students achievement, as measured by math and reading test results (dependent variable), according to the following level 1 and level 2 variables: for level 1, socio-economic status[13] (SES) and a psychological variable, student locus of control (LOCUS); and for level 2, – describing school environment – the proportion of ethnic minority students (MINORITY), and the proportion of students that receive free lunch (LUNCH). The first analysis examines aggregated data: level 1 variables are aggregated and moved to level 2. The second analysis treats disaggregated data; here, level 2 data is moved to level 1 and approached as level 1 data. In the third analysis, the hierarchical modeling method, neither the disaggregation nor aggregation of data is necessary.

Table 1. Regression Analysis of Students Achievements by Individual- and Group- Level Characteristics: Results of Three Statistical Approaches*

		Aggregated			Disaggregated			Hierarchical		
		B	SE	t	B	SE	t	B	SE	t
Individual Variables (Student's Characteristics)	Student locus of control (LOCUS)	4.97	.49	10.22**	2.96	.08	37.71**	2.82	.08	35.74**
	Socio-economic index (SES)	7.28	.26	27.91**	4.97	.08	62.11**	4.07	.10	41.29**
Contextual Variables (School Environment)	% ethnic minority (MINORITY)	-0.40	.06	-8.76**	-.45	.03	-15.53**	-0.59	.07	-8.73**
	% free lunch (LUNCH)	0.03	.05	0.59	-.43	.03	-13.50**	-1.32	.07	-19.17**

* Source: Osborne, Jason. 2000."Advantages of Hierarchical Linear Modeling".

[13] This is a measure of the socio-economic status of a student's family, assigned to each individual student.

Analysis of Aggregated Data (SPSS)

In order to perform the analysis of aggregated data, all level 1 variables, including test results, student locus of control (LOCUS) and socio-economic status (SES) were aggregated and moved to level 2. This model is statistically significant (R=.87, F(4999)=746.41, p<.0001). Results show that two variables – the mean values of LOCUS and SES – positively impact student achievement. Meanwhile, the MINORITY variable has a negative impact on the dependent variable, and the impact of the LUNCH variable is statistically insignificant.

Analysis of Disaggregated Data (SPSS)

In order to perform the analysis of disaggregated data, level 2 variables were disaggregated, moved to level 1, and assigned to units of analysis (in this case, to students). This model is statistically significant (R=.56, F(422899)=2648.54, p<.0001). Results show that all independent variables have significant impact on the dependent variable (test results). Specifically, the variables of socioeconomic status and student locus of control bear a positive relationship to student achievement as measured by test results. The two variables, which describe school environment – the proportion of minority students and the proportion of students receiving free lunch – have a negative relationship with student test results.

Hierarchical Modeling (HLM)

In this case, the level 1 variable was defined as individual level and disaggregated, variables, while and the level 2 variable, aggregated. All variables from individual and aggregated levels were centered to the mean value. This model is statistically significant (Chi squared=4231.39, 5df, p<.0001[14]). Those results lead to the conclusion that the impact of the two level 1 variables (LOCUS and SES) on the dependent variable is positive and statistically significant, while the impact of the aggregated variables (MINORITY and LUNCH) is negative and statistically significant,

[14] For hierarchical modeling, there is no statistical equivalent to R available in the OLS regression.

Discussion

Assuming that the results generated through HLM most closely reflect reality, a comparison of the three methods suggests the following conclusions. First, in the case of the model with aggregated variables of individual type, B coefficients for SES and LOCUS variables are overestimated, while B coefficients for MINORITY and LUNCH variables are underestimated. Next, the analysis of disaggregated data leads one to underestimate the impact of the level 2 variables, overestimate the effect of SES, and underestimate standard errors. Comparing results from the three types of analysis presented above, there is a clear advantage to using hierarchical modeling for the analysis of nested data and hierarchical models.

Conclusions

Hierarchical modeling is more and more commonly applied to empirical analysis of data with hierarchical structure in the social sciences. The increased popularity of this method stems from the fact that it makes possible a more precise analysis of the impact of both disaggregated variables (most commonly of an individual type), and aggregated variables, on a dependent variable. Hierarchical modeling therefore enables researchers to analyze models that include both aggregated and individual types of data, and at the same time, to avoid related statistical pitfalls. HLM, ML, STATA and R are the most commonly used softwares applied to the analysis of nested data.

References

Bates, Douglas. M., and Jose. C. Pinheiro. 1998. *Computational Methods for Multilevel Modeling*. Madison: University of Wisconsin.

Bosker, Roel. 1998. "The Class Size Question in Primary Schools: Policy Issues, Theory, and Empirical Findings from The Netherlands." *International Journal of Educational Development* 29: 763–778.

Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models*. London: Sage Publications.

Browning, Christopher R., Seth L. Feinberg, Robert Dietz. 2004. "The Paradox of Social Organization: Networks, Collective Efficacy, and Violent Crime in Urban Neighborhoods." *Social Forces* 83 (2): 503–34.

Chan, Jennifer and Anthony Kuk. 1997. "Maximum Likelihood Estimation for Probit-Linear Mixed Models with Correlated Random Effects." *Biometrics* 53: 86–97.

Cronbach, Lee J., Robert L. Linn, Robert L. Brennan, and Edward H. Haertel. 1997. "Generalizability Analysis for Performance Assessments of Student Achievement or School Effectiveness." *Educational and Psychological Measurement* 57: 373–99.

Ernste, Huib (ed.). 1996. *Multilevel Analysis with Structural Equation Models*. Zurich, Switzerland: Swiss Federal Institute of Technology (ETH).

Freedman, David. 2001. "Ecological Interference and the Ecological Fallacy." in Neil J. Smelser and Paul B. Baltes (ed.) *International Encyclopedia of the Social & Behavioral Sciences* 6: 4027–30.

Gibson, Alex and Sheena Asthana. 1998. "School Performance, School Effectiveness and the 1997 White Paper." *Oxford Review of Education* 24: 195–210.

Goldstein, Harvey. 1987. *Multilevel Models in Educational and Social Research*. London, Griffin; New York: Oxford University Press.

Goldstein, Harvey. 1995. *Multilevel Statistical Models*. Second Edition. London: Arnold.

Goldstein, Harvey. 1997. "Methods in School Effectiveness Research." *School Effectiveness and School Improvement* 8: 369–95.

Gray, John, David Jesson, Harvey Goldstein, Keith Hedger, Jon Rasbash.1995. "A Multilevel Analysis of School Improvement: Changes in Schools' Performance Over Time." *School Effectiveness and School Improvement* 6: 97–114.

Greenland, Sander. 1997. "Second Stage Least Squares Versus Penalized Quasi-Likelihood for Fitting Hierarchical Models in Epidemiological Analyses." *Statistics in Medicine* 16: 515–26.

Hurwitz, Jon and Mark Peffley. 1987. "How are Foreign Policy Attitudes Structured? A Hierarchical Model." *American Political Science Review* 81 (Dec), 1099–1120.

Kref, Ita and Jan de Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage Publications.

Osborne, Jason. W. 2000. "Advantages of Hierarchical Linear Modeling." *Practical Assessment, Research and Evaluation*, 7(1).

Rasbash, Jon, William Browne, Harvey Goldstein, Min Yang, et al. 2000. *A User's Guide to MLwiN* (Second Edition). London, Institute of Education.

Raunderbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Application and Data Analysis Methods.* Thousand Oaks, CA: Sage Publications.

Raunderbush, Stephen. W., Anthony S. Bryk, Yuk Fai Cheong, Richard T. Congdon. 2002. *HLM, 5. Hierarchical Linear and Nonlinear Modeling.* Chicago, IL: Scientific Software International.

Runyon, Richard P., and Audry Haber. 1976. *Fundamentals of Behavioral Statistics,* MA: Addison-Wesley Publishing Company.

Snijders, Tom and Roel Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.

Yang, Min and Harvey Goldstein. 1996. "Multilevel Models for Longitudinal Data." in Engel, U. and Tanner, J. (Eds), *Analysis of Change, Advanced Techniques in Panel Data Analysis*, p. 191–220. Berlin: Walter de Gruyter.

Wilk, Katarzyna. 2007. "Attitudes to European Integration in Old and New Members. Effects of Class, Generation and Gender." Dissertation Manuscript, Yale University.