Applied Research Methods

December 2014

A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study

Laura M. O'Dwyer
Boston College

Caroline E. Parker
Education Development Center

Summary

Analyzing data that possess some form of nesting is often challenging for applied researchers or district staff involved in data analysis. This primer for using multilevel regression modeling to resolve some of the challenges includes a research situation with nested data to illustrate how and why to use a multilevel approach, specifically the Advanced Statistics module of SPSS IBM Statistics.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development Center, Inc.

REL 2015–046

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

December 2014

This report was prepared for the Institute of Education Sciences (IES) under contract ED-IES-12-C-0009 by Regional Educational Laboratory Northeast & Islands administered by Education Development Center, Inc. (EDC). The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

O'Dwyer, L. M., and Parker, C. E. (2014). *A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study* (REL 2015–046). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

Researchers often study how students' academic outcomes are associated with the characteristics of their classrooms, schools, and districts. They also study subgroups of students such as English language learner students and students in special education. However, district personnel may not be aware that commonly used analytic methods might give inaccurate readings of the statistical significance of results when individual (student) data are nested within groups (classrooms, programs, schools) or when group-level data are disaggregated to predict individual outcomes. In these cases, it may be necessary to use multilevel regression modeling (also known as hierarchical linear modeling or linear mixed modeling) to analyze data. This primer on conducting multilevel regression analyses to address these issues using the Advanced Statistics module of SPSS IBM Statistics should be useful to applied researchers and district staff engaged in or in charge of data analysis.

A recent study by Regional Educational Laboratory Northeast & Islands focusing on the achievement of a cohort of English language learner students (Parker, O'Dwyer, & Irwin, 2014) asked which characteristics of students, English language learner programs, and schools were most closely related to the students' English proficiency scores. Such research is often addressed using a prediction modeling technique called ordinary least squares (OLS) regression. OLS regression is used to examine the strength and direction of the relationship between two variables in a statistical model while holding other variables constant. It is used extensively as an exploratory, explanatory, and predictive tool. However, a standard OLS model may not be appropriate in situations where individuals are nested in groups because nesting may lead to a statistical dependency among the observations in the sample.[1]

Statistical dependencies can occur for multiple reasons, including situations in which individuals share the same educational context—where individuals are nested in classrooms or schools—and situations in which group-level characteristics are used to predict individual outcomes. An example of the first situation is English language learner students who attend the same school. These students are likely to be more similar to each other than to individuals in other schools (they are drawn from the same communities, use similar school resources, have the same teachers), and so statistical dependency may occur. The second situation occurs when the analysis aims to examine the relationships between group characteristics and individual outcomes. Because each individual in a group is assigned the same value for a group characteristic, a statistical dependency can occur.

In the presence of statistical dependency and assuming that the model is correctly specified, standard OLS regression can produce unbiased estimates of the relationships between variables (the regression coefficients). However, the standard errors associated with the regression coefficients may be biased, leading to incorrect conclusions about the statistical significance of the observed relationships. One of the key assumptions of OLS models (and several other common analysis procedures) is that each individual provides a unique piece of statistical information that is unrelated to the information provided by other individuals in the sample. Statistical dependency violates this assumption and can lead to downwardly biased estimates of the standard errors associated with the regression coefficients and ultimately to incorrect statistical conclusions.

Multilevel regression modeling does not correct bias in the regression coefficient estimates compared with an OLS model; however, it produces unbiased estimates of the standard errors associated with the regression coefficients when the data are nested, and easily allows group characteristics to be included in models of individual outcomes (Snijders & Bosker, 1999; Raudenbush & Bryk, 2001; Bickel, 2007, Gelman & Hill, 2007; Hox, 2010). Although multilevel modeling is an advanced data analysis procedure that requires specialized software and data analysis skills, several readily available statistical packages provide the capability to conduct such analyses, including the Advanced Statistics module of SPSS IBM Statistics, used for the analysis in this primer.

Contents

Why this primer?

A recent study by the Regional Educational Laboratory Northeast & Islands on the achievement of a cohort of English language learner students in a large school district in Connecticut asked which characteristics of students, English language learner programs, and schools were most closely related to the students' English proficiency scores (Parker et al., 2014).

A common statistical method used to address this type of question is ordinary least squares (OLS) regression analysis. OLS regression, which can examine the strength and direction of the relationship between two variables while holding other variables constant, is used extensively as an exploratory, explanatory, and prediction tool. However, a standard OLS model may produce misleading results about the statistical significance of a relationship when it is used to analyze data collected from students in classrooms and schools because of this "nesting" of data—students nested within classrooms.

A standard ordinary least squares regression can produce statistically unbiased estimates of the relationships among variables; however, the nesting of students in schools leads to statistical dependency and the possibility of downwardly biased estimates of the standard errors

A key assumption of OLS models (and several other common analysis procedures) is that each individual in the sample provides a unique piece of statistical information unrelated to the information provided by other individuals in the sample. Because students who attend the same school are likely to be more similar to each other than they are to individuals in other schools (they are drawn from the same communities, use similar school resources, have the same teachers), meeting this assumption can be difficult. A standard OLS regression can produce statistically unbiased estimates of the relationships among variables (regression coefficients); however, the nesting of students in schools leads to correlated observations (a dependency among the data) and the possibility of downwardly biased estimates of the standard errors associated with the regression coefficients. If adjustments are not made to the OLS model to account for the statistical dependency introduced by nesting, analysts can make substantive errors in interpreting the statistical significance of relationships (Raudenbush & Bryk, 2001).[2]

Education researchers also frequently examine how group or organizational characteristics, such as the characteristics of schools, are associated with individual outcomes. Each individual in a group is assigned the same value for the group characteristic. However, a group characteristic that is disaggregated to the individual level cannot explain individual differences within a group. For example, the percentage of students who receive free or reduced-price lunches at a school cannot explain student-to-student differences in test performance within a school. A standard OLS model that includes a disaggregated variable will use the individual-level sample size, which is inaccurate for a variable measured at the group level, for determining statistical significance. This practice can also lead to incorrect conclusions about the significance of the relationships under investigation.

The REL Northeast & Islands study of English language learner students examined the characteristics of students, English language learner programs, and schools most closely related to their English proficiency scores (Parker et al., 2014). The researchers had to consider the possibility of a statistical dependency among the English language learner students who attended the same school while also including school characteristics in models for predicting individual student outcomes.

Over the past two decades, multilevel modeling (also known as hierarchical linear modeling or linear mixed modeling)[3] has become a popular way to analyze data with statistical dependency (see box 1 for definitions of key terms). Multilevel regression modeling does not correct bias in the regression coefficient estimates compared with an OLS model; however, it produces unbiased estimates of the standard errors associated with the regression coefficients when the data are nested and easily allows group characteristics to be included in models of individual outcomes (Snijders & Bosker, 1999; Raudenbush & Bryk, 2001; Bickel, 2007, Gelman & Hill, 2007; Hox, 2010).[4] Although multilevel modeling is an advanced data analysis procedure that may require specialized software and data analysis skills, several readily available statistical packages make it more accessible. For example, Stata and SAS allow analyses that account for statistical dependency in the estimation of standard errors. Likewise, the Advanced Statistics module of SPSS IBM Statistics uses multilevel modeling to analyze data with statistical dependency.

Multilevel regression modeling produces unbiased estimates of the standard errors associated with the regression coefficients when the data are nested and easily allows group characteristics to be included in models of individual outcomes

This primer provides guidance to applied researchers and district staff engaged in or in charge of analyzing data with nested components and who have access to the Advanced Statistics module of SPSS IBM Statistics.[5] Although this discussion focuses on OLS regression analysis, the effects of nesting are relevant for other commonly used inferential tests

Box 1. Key terms

Intercept-only model: A type of multilevel regression model in which only the intercept is allowed to vary from group to group.

Intraclass correlation coefficient: The proportion of variance in the dependent variable that lies between groups. The intraclass correlation coefficient ranges from 0 to 1 and is used to estimate the degree of statistical dependency in the data. If there is no statistical dependency, all of the variance would be expected to lie among individuals, and the intraclass correlation coefficient would be zero or close to zero. Conversely, with highly dependent data, the largest proportion of variance would lie among groups, and so the intraclass correlation coefficient would be closer to 1.

Multilevel regression modeling: A set of statistical models that accounts for the statistical dependency that may result from nested data. When the assumptions of the models are met, they provide unbiased estimates of the standard errors associated with regression coefficients and allow group characteristics to be included in statistical models of individual outcomes (Snijders & Bosker, 1999; Raudenbush & Bryk, 2001; Bickel, 2007, Gelman & Hill, 2007; Hox, 2010). The approach is also known as hierarchical linear modeling or linear mixed modeling.

Nested data: When data are collected from multiple individuals in a group, the individual data are considered nested within that group.

Ordinary least squares (OLS) regression: A statistical model used to describe the relationship between an array of variables (independent variables) and a dependent variable. OLS models assume that the relationship between the dependent and independent variables is linear, that the prediction errors are normally distributed and equally variable across the range of scores, and that the prediction errors for individuals are uncorrelated with each other.

Statistical dependence: When the prediction error for the i^{th} individual is correlated with the prediction error for the i^{th}+1 individual. It occurs when individuals are nested in groups or when group characteristics are disaggregated to the individual level and leads to biased estimates of the standard errors associated with the regression coefficients in an OLS model.

such as *t*-tests, analysis of variance, and correlation. While this primer considers data with two levels, such as individuals nested within groups (students nested within schools), multiple levels of nesting, such as students within classrooms, classrooms within schools, or schools within districts, are also common.[6]

The following sections describe the challenges to OLS regression presented by statistical dependency and the inclusion of group characteristics; how multilevel modeling can accommodate nested data and the inclusion of group characteristics; and the use of multilevel modeling in a study of English language learner students in a large urban school district (Parker et al., 2014).

Challenges in using ordinary least squares regression analysis with nested data

Unless the independent variables in an ordinary least squares regression predict the dependent variable perfectly, there will be error in the prediction model

In OLS regression analysis a statistical model specifies an array of variables (independent variables) and uses them to predict some other variable (the dependent variable). The statistical model generated by OLS analysis shows the strength and direction of the relationship between a particular independent variable and the dependent variable while holding constant all other independent variables in the model. The focus here is on prediction models in which there is a single dependent continuous variable.[7] For example, an OLS regression model for a continuous dependent variable for i individuals, Y_i, and three independent variables X_{1i}, X_{2i}, and X_{3i} measured at the individual level, and one independent variable W_{1j} measured at the group level takes the following form:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4W_{1j} + e_i. \quad (1)$$

The dependent variable for individual i, Y_i, is predicted from a linear combination of variables X_{1i}, X_{2i}, X_{3i}, and W_{1j} plus some amount of error, e_i. For example, in the study used as an illustration in this primer (Parker et al., 2014), English language learner students' English proficiency scores (Y_i) might be predicted using students' school attendance rate (X_{1i}), their dominant language (X_{2i}), the program in which they are enrolled (X_{3i}), and the percentage of students achieving math proficiency in the school (W_{1j}).

The intercept, a, and the regression coefficients, b_1, b_2, b_3, and b_4, are fixed values that are the same for every i individual in the dataset. The model predicts a value of Y_i based on individuals' unique values on X_{1i}, X_{2i}, X_{3i}, and W_{1j}. The value of the intercept, a, represents the predicted value of Y_i when all X_i and W_j variables in the model are zero. Each regression coefficient in the model (b_1, b_2, b_3, and b_4) indicates the predicted change in Y_i for every one unit increase in the value of the associated independent variable, holding constant the other independent variables in the model. Holding all other independent variables in the model constant means that the regression coefficient is the relationship between Y_i and a particular independent variable for individuals who have the same value for all other independent variables in the model.

Unless the independent variables predict Y_i perfectly, there will be error in the prediction model; for some individuals the predicted Y_i value will be higher than their actual Y_i value, while for others it will be lower. The error term in the model, e_i, is of particular interest when analyzing data that have a statistical dependence; it is discussed in more depth later.

Statistical significance tests evaluate the strength of relationships

Significance tests are typically used to evaluate the strength of the relationships represented by the regression coefficients. They show whether the predicted change in Y_i associated with a one-unit change in an independent variable is statistically significantly different from zero. Decisions about statistical significance in a sample rely on the probability of observing a regression coefficient of the size observed or larger by chance if the coefficient is really zero in the population. If the probability of a chance relationship is large, the regression coefficient is not statistically significantly different from zero (X_i is not a significant predictor of Y_i). If the probability of a chance relationship is small, the regression coefficient is statistically significantly different from zero (X_i is a significant predictor of Y_i).

For analysts to be confident in a decision about the significance of the regression coefficient, the data must meet several assumptions.[8] First, the relationship between the dependent and independent variables is assumed to be linear—that is, it can be described using the equation of a straight line that contains only the intercept, *a*, the regression coefficients, *b*s, and some amount of error, *e*. Second, the OLS model assumes that the errors are normally distributed and equally variable across the range of scores. Third, and most pertinent to this discussion, the prediction errors for *i* individuals are assumed to be uncorrelated with each other. When this assumption is met, each piece of data in the sample is unique and statistically independent of all other pieces of data, and the prediction error for the i^{th} individual is not correlated with the prediction error for the $i^{th}+1$ individual.

The assumption of uncorrelated errors is difficult to meet when there is a statistical dependency in the data because individuals are nested in the same group

Danger of false-positive or false-negative errors

However, the assumption of uncorrelated errors is difficult to meet when there is a statistical dependency in the data because individuals are nested in the same group (students in the same school are drawn from the same communities, use similar school resources, and have the same teachers). In that case the estimated probability of observing a relationship of the size observed or larger in the sample, if the relationship is really zero in the population, is artificially reduced. Because of this reduction, a relationship may be considered statistically significant when it is not (Snijders & Bosker, 1999; Raudenbush & Bryk, 2001; Bickel, 2007; Gelman & Hill, 2007; Hox, 2010). This type of incorrect false-positive decision is called a Type I error. It is also possible for an independent variable to be nonsignificant in an OLS model and significant in a multilevel model. In this case the OLS result is considered a false-negative or Type II error.

Type I errors are problematic because they may lead to the conclusion that a particular student characteristic is associated with higher academic performance or that a school- or district-wide program or policy is effective for improving academic outcomes when in reality there is no effect. Type II errors are problematic because they could miss a significant relationship in the data that could guide decisionmaking.

Analyzing nested data with multilevel modeling

Multilevel modeling techniques were developed to help analysts avoid erroneous conclusions from using inappropriate analysis procedures, such as using OLS regression with unadjusted standard errors when analyzing nested data. In a multilevel model the associations between the dependent variable and the independent variables are expressed as regression

coefficients and interpreted in the same way as OLS regression coefficients. However, in a multilevel regression analysis the coefficients refer to specific levels (for example, students within schools) in the nested data. The details of how this is done are described in this section. A step-by-step procedure for running a multilevel regression analysis using the Advanced Statistics module of SPSS IBM Statistics is shown in appendix A.

Accounting for statistical dependency

Multilevel regression modeling accounts for statistical dependency by assigning each level in the data hierarchy its own statistical model that includes an intercept, regression coefficients, and error term. For example, an analysis of students nested within schools would require the use of a model that includes separate statistical expressions for students at level 1 and schools at level 2. Since there are error terms in each level of the data, the result is a more complex error structure than in a standard OLS model.

Multilevel modeling techniques were developed to help analysts avoid drawing erroneous conclusions from inappropriate analysis procedures, such as using ordinary least squares regression with unadjusted standard errors when analyzing nested data

Because statistical models are formulated for each level in the data hierarchy, group characteristics can be included in models of individual outcomes, which can improve estimates of the effects within the groups and allow testing of hypotheses about cross-level effects.[9] When the assumptions of the multilevel model hold, the bias in the standard errors of the regression coefficients is reduced, the estimated probability that researchers rely on to evaluate statistical significance is no longer artificially reduced, and the incidence of Type I, false-positive errors is reduced (Snijders & Bosker, 1999; Raudenbush & Bryk, 2001; Bickel, 2007; Gelman & Hill, 2007; Hox, 2010).

Variance and covariance can be partitioned into within-group and between-group components

Another advantage of multilevel modeling is that it allows the variance and covariance components to be partitioned across the levels of the data. Whereas OLS regression provides a single value to indicate the proportion of variability in the dependent variable explained by the combination of independent variables (referred to as R^2), the multilevel model's more complex error term allows partitioning the total variance in the dependent variable into a within-group component (student-to-student variation within schools, for example) and a between-group component (variation between schools, for example). It also allows calculating the reduction in variance with the addition of independent variables at the individual and group levels in subsequent models.

The available variance in the dependent variable is partitioned into its within-group variance (σ^2) and between-group variance (τ_{00}) components using an unconditional multilevel model that includes only a random group effect. The intraclass correlation coefficient (*ICC*) in equation 2 is used to calculate the portion of variance in the dependent variable that is explained at each level in subsequent models with the addition of individual and group measures:

$$\text{ICC} = \frac{\tau_{00}}{\sigma^2 + \tau_{00}}. \qquad (2)$$

If there is no statistical dependency in the data, all of the variance would be expected to lie among individuals, and the intraclass correlation coefficient would be expected to be zero or close to zero. Conversely, with highly dependent data, the largest proportion of variance would lie among groups, and so the intraclass correlation coefficient would be closer to 1.[10]

Comparing the two statistical models

To draw parallels between the statistical models for OLS and multilevel regression, a two-level model in which individuals are nested in groups is described. For individual i nested in group j, a multilevel regression model analogous to equation 1, with a continuous dependent variable Y_{ij} and three independent variables, X_{1ij}, X_{2ij}, and X_{3ij}, measured at the individual level, would take the following form:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + r_{ij}. \quad (3)$$

This model states that the dependent variable Y_{ij}, for individual i in group j, is predicted from a linear combination of variables X_{1ij}, X_{2ij}, and X_{3ij}, and an individual-level prediction error term, r_{ij}. Despite the different notation, this model is similar to the OLS expression in equation 1 in that it consists of an intercept, β_{0j}; regression coefficients β_{1j}, β_{2j}, and β_{3j}; and an individual-level error term, r_{ij}, that is assumed to be normally distributed with a mean of zero. As in the OLS model, the value of the intercept, β_{0j}, represents the predicted value of Y_{ij} when all X_{ij} in the model are zero and each of the regression coefficients, β_{1j}, β_{2j}, and β_{3j}, indicates the predicted change in Y_{ij} for every one-unit increase in the value of the associated X_{ij} variable, holding constant the other variables in the model.

Whereas ordinary least squares regression provides a single value to indicate the proportion of variability in the dependent variable explained by the independent variables, the multilevel model allows partitioning the total variance into a within-group component and a between-group component

Since i individuals are nested in j groups, a group-level model can be formulated. To illustrate the model, consider the situation in which a characteristic measured at the group level, W_{1j}, is included in the group model to predict the intercept in the individual model, β_{0j}:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + u_{0j}. \quad (4)$$

Again, this model contains an intercept, γ_{00}, a regression coefficient, γ_{01}, and a group-level prediction error term, u_{0j}. The regression coefficient, γ_{01}, represents the predicted increase in Y_{ij} for every one-unit increase in the group-level measure, W_{1j}, holding constant the individual-level variables in the model. Each group has a unique error term, u_{0j}, which is assumed to be normally distributed with a mean of zero.

Models may also be formulated for the level 1 (individuals) regression coefficients. In the simplest case these models are as follows:[11]

$$\beta_{1j} = \gamma_{10}, \beta_{2j} = \gamma_{20}, \text{ and } \beta_{3j} = \gamma_{30}. \quad (5)$$

The similarity between multilevel models and OLS models (equation 1) is evident when the multilevel models (equations 3–5) are combined into a single model, referred to as a linear mixed model:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{1ij} + \gamma_{20} X_{2ij} + \gamma_{30} X_{3ij} + \gamma_{01} W_{1j} + r_{ij} + u_{0j}. \quad (6)$$

As in the OLS model, the value of the intercept, γ_{00}, represents the predicted value of Y_{ij} when all predictors in the model (X_{1ij}, X_{2ij}, X_{3ij}, and W_{1j}) are zero, and each regression coefficient (γ_{10}, γ_{20}, γ_{30}, and γ_{01}) indicates the predicted change in Y_{ij} for every one-unit increase in the value of the associated independent variable, holding constant the other variables in the model. The linear mixed model also shows a more complex error structure

than the OLS model. Instead of the single-error term, e_i, in the OLS model, the linear mixed model indicates a within-group error, r_{ij}, and a between-group error, u_{0j}.

Sample size is important

Multilevel modeling procedures require a minimum sample size to achieve accurate estimates of the regression coefficients and their standard errors, as well as of the variance components and their standard errors (Raudenbush & Bryk, 2001). Typically, the minimum sample size refers to the highest level in the data hierarchy (for analyzing students nested in schools, the number of available schools is the primary sample-size consideration), though the numbers of individuals within groups and the magnitude of the intraclass correlation coefficient may also have an effect on the model estimates (Hox & Maas, 2002). The sample size will also affect the complexity of the models that can be formulated.[12] Guidelines for sample-size requirements and their implications for model complexity, the regression coefficients, variance components, and their standard errors are given in various studies and texts. For example, models with fewer than 20–25 groups may not provide accurate estimates of the regression coefficients and their standard errors, or of the variance components and their standard errors. Simulation studies have shown that analyses with fewer than 50 groups can lead to biased estimates of the standard errors at the group level (Hox & Maas, 2002).[13]

Multilevel modeling procedures require a minimum sample size to achieve accurate estimates of the regression coefficients and their standard errors, as well as of the variance components and their standard errors

An illustration using English language learner student and school data

The recent study by the REL Northeast & Islands on the achievement of a cohort of English language learner students in a large school district in Connecticut is used here to illustrate how multilevel regression modeling was applied to examine the relationships between student, program and school characteristics and English proficiency scores. The study asked which characteristics of students, English language learner programs, and schools were most closely related to the students' English proficiency scores (Parker et al., 2014). Data compiled for English language learner students in grades K–1 are used in the illustration (see appendix F and table G1 in Parker et al., 2014, for the complete results).

Complete data were available for 566 English language learner students in grades K–1 who attended 28 schools (with an average of 20 English language learner students per school). The unconditional model showed that the within-school variance was 1,990.4 (σ^2) and the between-school variance was 579.7 (τ_{00}). Using equation 2, the intraclass correlation coefficient was calculated as 0.23, or 23 percent. Thus 23 percent of the variability in the students' English proficiency scores was due to school-to-school differences (such as differences in school socioeconomic status, communities, and policies or practices) and 77 percent to student-to-student differences within schools (such as differences in student attitudes and socioeconomic status).

Two-level model used to predict English proficiency scores

A two-level intercept-only model was used to predict the students' English proficiency scores using student and school characteristics. At the student level the model included indicators of students' dominant home language (0 = Spanish, 1 = language other than Spanish), special education status (0 = does not have an individualized education program, 1 = has an individualized education program), attendance rate (continuous variable, total

number of days in attendance divided by the possible number of days in attendance), and dichotomous variables for comparing the English language learner programs to the grand mean across all programs (transitional bilingual, dual-language bilingual, eligible for but not attending a program by parent request, and English as a second language services).

At the school level the model included two variables: the percentage of students in the school who achieved proficiency in math and the percentage who achieved proficiency in reading. These school variables were constant for every English language learner student in the same school and had no variation within schools. (See appendix A for a step-by-step guide with screenshots for running this type of analysis using the SPSS menu interface as well as the syntax that these steps produced.)

Interpreting the results of ordinary least squares and multilevel regression models

Table 1 compares the fixed components from a multilevel regression analysis predicting English language learner students' scores on a test of English proficiency (see appendix F and table G1 in Parker et al., 2014) along with the results from an OLS analysis conducted with the same data.

Student-level independent variables. Both analyses produced regression coefficients for the student-level independent variables that were generally in the same direction, with approximately similar magnitudes. English language learner students whose dominant home language was not Spanish, who were not in special education (did not have an

Table 1. Comparison of results for a multilevel model and an ordinary least squares model predicting English language learner students' scores on a test of English proficiency

	Multilevel model			Ordinary least squares model		
Variable	Regression coefficient	Standard error	Significance (*p* value)	Regression coefficient	Standard error	Significance (*p* value)
Constant	418.23	5.02	<.001	420.75	1.88	<.001
Student-level independent variable						
Dominant home language not Spanish	49.95	14.55	.001	55.14	14.70	<.001
Special education status (has an individualized education program)	–20.25	6.18	.001	–20.65	6.41	.001
Attendance rate (10 percentage point increment above the grand mean)	14.90	3.70	<.001	16.50	3.80	<.001
Transitional bilingual[a]	–13.36	2.49	<.001	–9.60	2.20	<.001
Dual-language bilingual[a]	17.26	8.92	.054	–8.81	6.63	.184
Eligible but not served due to parent request[a]	35.16	7.88	<.001	33.48	8.16	<.001
English as a second language services (for students speaking languages other than Spanish)[a]	–28.54	14.19	.045	–24.01	13.57	.077
School-level independent variable						
Math proficiency rate (10 percentage point increment)	–14.09	8.00	.091	–1.65	3.20	.603
Reading proficiency rate (10 percentage point increment)	16.57	7.60	.040	4.38	3.20	.174

Note: Dependent variable: English proficiency scores.

a. Compared with the grand mean.

Source: Authors' analysis based on data from Parker et al. (2014).

individualized education program), and who attended school more frequently were predicted to have higher English proficiency scores, holding all other variables constant. Students who were eligible for a language program but did not attend due to parental request were predicted to have English proficiency test scores that were higher than the grand mean, whereas English language learner students in transitional bilingual or English as a second language services were predicted to have scores lower than the grand mean, holding all other variables in the model constant. For several independent variables the statistical decisions are the same under both approaches. For example, students with a home language other than Spanish and students with higher attendance rates were predicted to have higher English proficiency scores.

For English language learner program indicators the results were also similar under the two approaches. Holding all else constant, students enrolled in transitional bilingual programs were predicted to have English proficiency scores that were significantly lower than the grand mean English proficiency score, while students who were not served due to parental request were predicted to have significantly higher English proficiency scores than the grand mean. Despite the similarities in magnitude, the coefficient for enrollment in English as a second language services was significant only in the multilevel model, likely because of the reweighting that occurs in that model based on the variances and covariances of the error term. For student enrollment in a dual-language bilingual program the direction and magnitude of the associated regression coefficients were different under the two approaches; however, the coefficients were not significant in either model.

The ordinary least squares and multilevel regression approaches provided similar results regarding the relationship between the dependent variable and the independent variables; however, some significant relationships could have been overlooked had the ordinary least squares approach alone been used

The multilevel model produced smaller standard errors for the regression coefficients for four of the seven student-level independent variables than did the OLS model. Although this seems at odds with what was expected, it is not surprising given that the parameters were based on different estimation algorithms. OLS analysis, like other general linear model approaches, relies on least squares estimation to produce the model parameters and gives equal weight to each student but not to each school. Conversely, multilevel analysis uses maximum likelihood or restricted maximum likelihood procedures for estimating model parameters, which tend to downweight the contribution of individuals from larger groups, in this case, students from larger schools. Moreover, the results are reweighted based on the variances and covariances of the error terms in the model, which, when the assumptions are met, produce more precise estimates than the OLS model.[14]

School-level independent variables. The school-level regression coefficients indicating the predicted increase in English proficiency scores associated with a 10 percentage point increase in the percentage of students achieving proficiency in both math and reading differed under the two approaches. The only coefficient significantly associated with English proficiency was the coefficient for reading proficiency in the multilevel model. For the school-level variables the standard errors of the OLS coefficients were smaller than those in the multilevel model. When data are nested, the estimated probability of observing a school-level regression coefficient of the size observed or larger in the sample, if the coefficient is really zero in the population, tends be artificially reduced. Because of this reduction, a relationship may be considered statistically significant when it is not.

Implications of statistical dependency

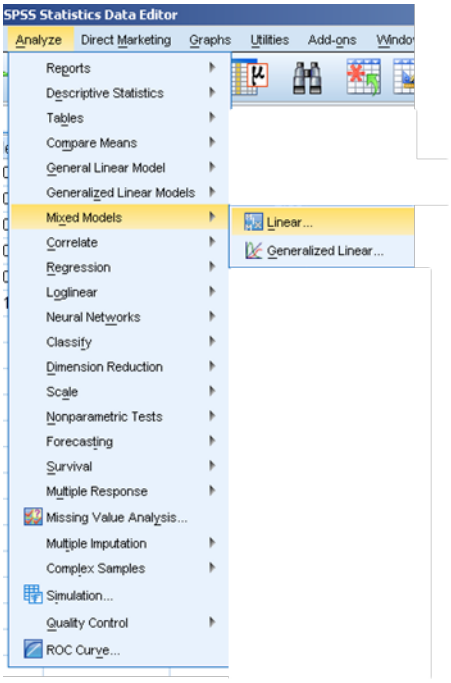
The two approaches provided similar results regarding the relationship between the dependent variable and the independent variables. However, some significant relationships could have been overlooked had the OLS approach alone been used. For example, enrollment in English as a second language services reached statistical significance only in the multilevel model, as did the school-level variable of reading proficiency rate. Although not statistically significant, the different direction and magnitude of the associated regression coefficients of student enrollment in a dual-language bilingual program indicate a negative relationship in the OLS approach but a positive relationship in the multilevel model.

When data possess some degree of statistical dependency such as what was observed for the K–1 data (approximately 23 percent of the variability in English language learner students' English proficiency score was due to school-to-school differences), analysts should use appropriate procedures that provide accurate results for guiding policy and practice. In the example used here, the lack of statistically significant school-level variables in the OLS model could lead administrators to erroneously conclude that there were no school effects predicting English language learner students' English proficiency.

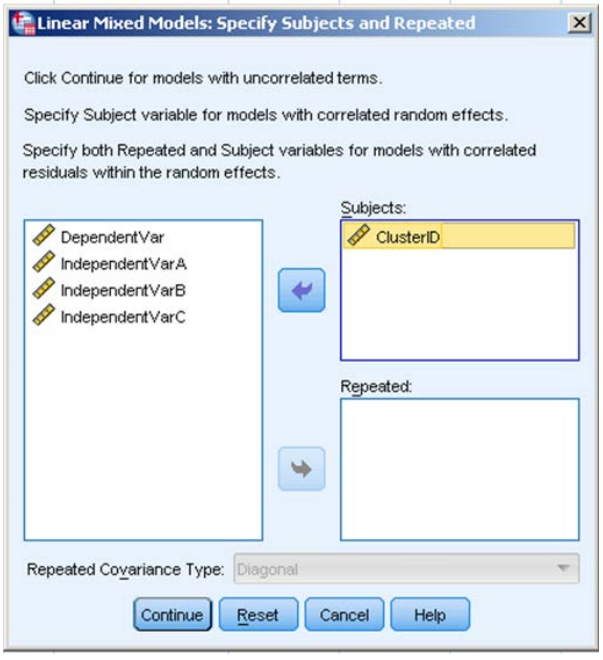
Appendix A. Step by step procedure for using the Advanced Statistics module of SPSS IBM Statistics

Steps for running the type of intercept-only, two-level model described in this primer using the Advanced Statistics module of SPSS IBM Statistics are given with screenshots. The SPSS syntax produced through these steps and syntax for a parallel ordinary least squares (OLS) regression model are shown at the end.

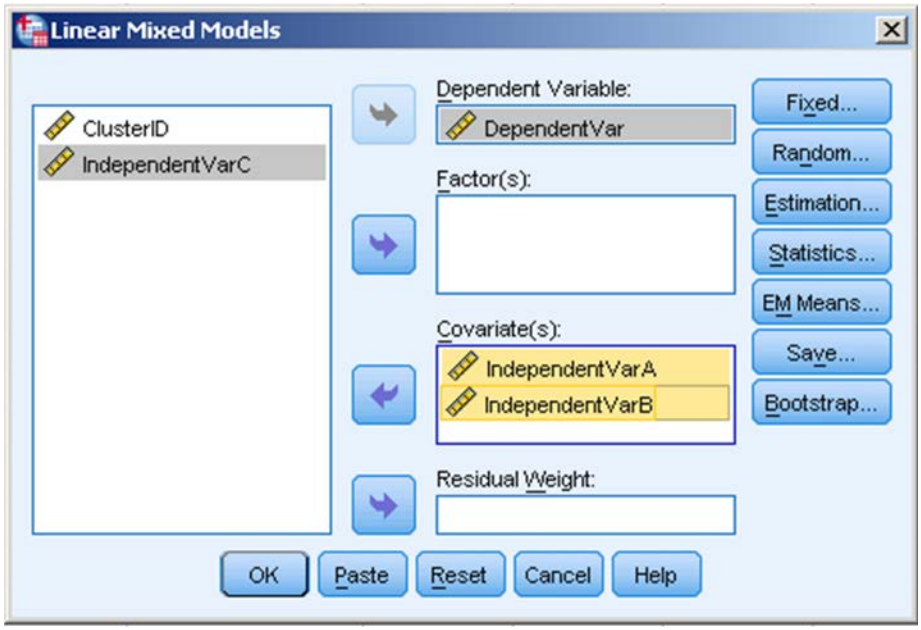
Step 1: Select the Linear option from the Mixed Models menu under Analyze.



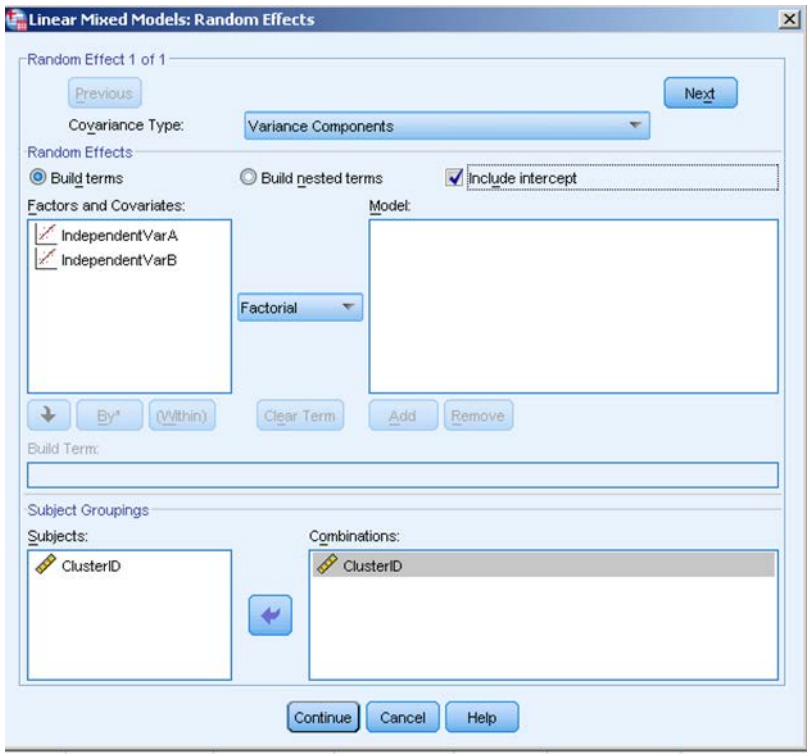
Step 2: Select the variable that identifies the groups within which individuals are nested. Select Continue.



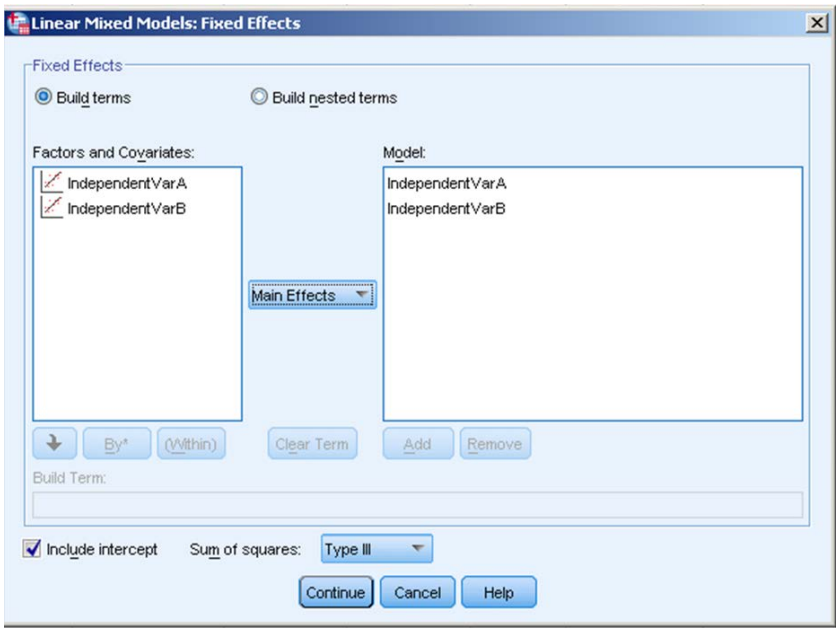
Step 3: Select the Dependent and Independent Variables.



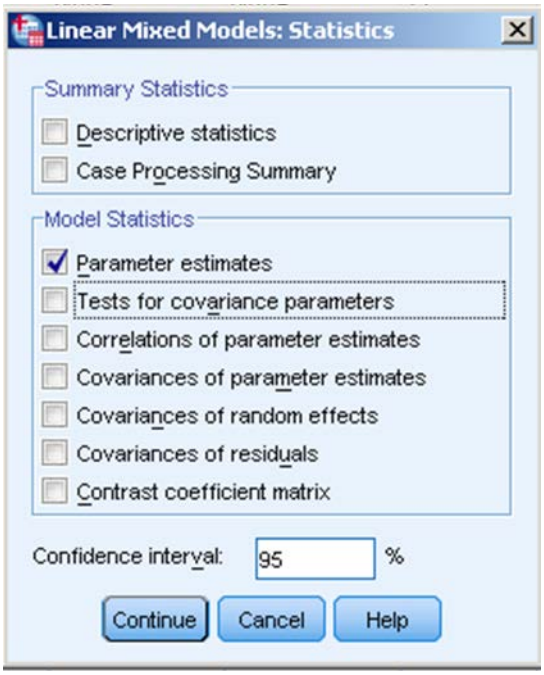
Step 4: Select the Random button on the right to open a new dialogue box. Under Subject Groupings, move the variable that identifies the groups within which individuals are nested into Combinations. Check the box titled Include intercept. Select Continue.



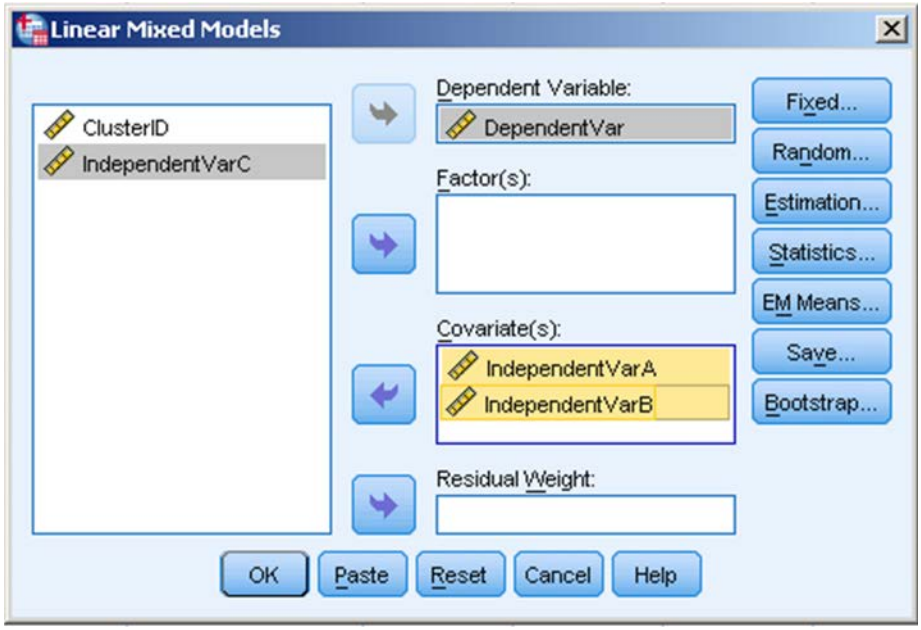
Step 5: Select the Fixed button on the right to open a new dialogue box. Select the Independent Variables in the Factors and Covariates window and move them into the Model window. Select Continue.



Step 6: Select the Statistics button on the right to open a new dialogue box. Check the box next to Parameter estimates. Select Continue.



Step 7: Select OK to run the model that will produce the output.



This seven-step process produced the following syntax for an intercept-only, two-level model with two fixed independent variables:

```
MIXED DependentVar WITH IndependentVarA IndependentVarB
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
 /FIXED=IndependentVarA IndependentVarB | SSTYPE(3)
 /METHOD=REML
 /PRINT=SOLUTION
 /RANDOM=INTERCEPT | SUBJECT(ClusterID) COVTYPE(VC).
```

A parallel OLS model would be formulated using the Linear Regression option available in both the Base version and the Advanced Statistics module of SPSS IBM Statistics. The syntax would be as follows:

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT DependentVar
 /METHOD=ENTER IndependentVarA IndependentVarB.
```

Notes

1. Some statistical packages such as Stata and SAS allow analysts to adjust the standard errors in an OLS regression to account for the effects of nesting.
2. Some statistical packages such as Stata and SAS allow analysts to adjust the standard errors in an OLS regression to account for the effects of nesting.
3. Multilevel regression modeling is a special case of the more general linear mixed model. See Raudenbush (1993).
4. Multilevel modeling is not appropriate in all situations, even in the presence of statistical dependency. For example, multilevel regression models reweight the data based on the variances and covariances of the error terms in the model (Selden, 1994) and tend to downweight the contribution of individuals from larger clusters. In situations where unweighted results are substantively necessary, it may be more appropriate to use an OLS model in which the standard errors are adjusted for the dependency. For additional information, see Raudenbush and Bryk (2001), Gelman and Hill (2007), and Hox (2010). For additional information about the procedures and diagnostic information for correctly specifying prediction models (such as OLS and multilevel regression models), including the specification of models with interaction terms, see applied linear regression modeling texts such as Chatterjee and Haldi (2012) and Gelman and Hill (2007).
5. A poll of school districts in the Northeast & Islands Region showed that SPSS IBM Statistics is the most frequently used statistical software package. Not all school districts have access to the Advanced Statistics module.
6. For more information about analyses with multiple levels of nesting, see Raudenbush and Bryk (2001) and Gelman and Hill (2007).
7. Prediction models that use multiple dependent variables or have noncontinuous (dichotomous or multicategorical) dependent variables are not discussed here. For further information about these types of models, see Snijders and Bosker (1999), Raudenbush and Bryk (2001), and Hox (2010).
8. The assumptions described here are also common to other types of analyses, such as correlation, testing the difference between means, and analysis of variance.
9. Providing that the assumptions of the model are met.
10. The impact of a nonzero intraclass correlation coefficient on the standard errors associated with the regression coefficients depends on the size of the groups in which individuals are nested, and so there are no set guidelines for what constitutes a large or small intraclass correlation coefficient. For additional information about the partitioning of variance in multilevel models, see Snijders and Bosker (1999), Raudenbush and Bryk (2001), Bickel (2007), Gelman and Hill (2007), and Hox (2010).
11. The models described here are intercept-only models. For information about other model formulations, see Snijders and Bosker (1999), Raudenbush and Bryk (2001), and Hox (2010).
12. For example, models that allow the relationships between individual-level measures and the dependent variable to vary across groups and models that include interactions between individual-level and group-level measures generally require larger samples than the simple intercept-only model described here.
13. For additional information about sample size requirements, see Raudenbush and Bryk (2001), Gelman and Hill (2007), and Hox (2010).
14. The weighting in multilevel modeling could produce model parameters that are no longer generalizable to all students, thereby resulting in biased parameters. In situations

where weighting is not desirable, OLS with cluster- corrected standard errors (available in Stata) may be more appropriate. For additional information about the estimation procedures and their consequences, see Raudenbush and Bryk (2001), Gelman and Hill (2007), and Hox (2010).

References

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* (1st ed.). New York: The Guildford Press.

Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* (5th ed.). Hoboken, NJ: Wiley.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Oxford, UK: Routledge Academics.

Hox, J. J., & Maas, C. J. M. (2002). Sample sizes for multilevel modeling. In J. Blasius, J. J. Hox, E. de Leeuw, & P. Schmidt (Eds.), *Social science methodology in the new millennium: Proceedings of the Fifth International Conference on Logic and Methodology* (2nd ed.). Opladen, Germany: Leske + Budrich.

Parker, C. E., O'Dwyer, L. M., & Irwin, C. W. (2014). *The correlates of academic performance for English language learner students in a New England district* (REL 2014–020). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast Islands. http://eric.ed.gov/?id=ED546480

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science* (pp. 459–496). New York: Marcel Dekker. Retrieved May 10, 2013, from http://www.unt.edu/rss/class/Jon/MiscDocs/Raudenbush_1993.pdf

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Selden, T. M. (1994). Weighted generalized least squares estimation for complex survey data. *Economic Letters, 46*(1), 1–6.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage Publications.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research