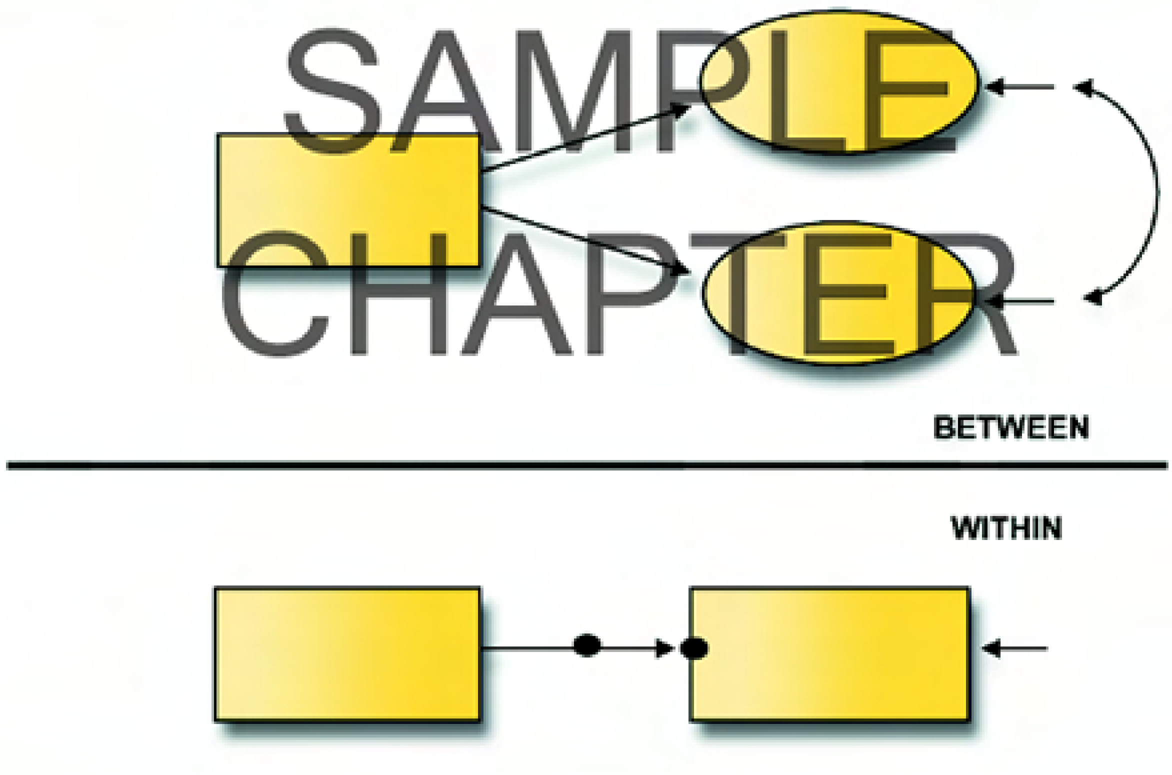
Multilevel and Longitudinal Modeling with IBM SPSS

QUANTITATIVE METHODOLOGY SERIES



Ronald H. Heck | Scott L. Thomas | Lynn N. Tabata

Routledge
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Routledge
Taylor & Francis Group
27 Church Road
Hove, East Sussex BN3 2FA

Routledge is an imprint of Taylor & Francis Group, an Informa business

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-84872-862-2 (Hardback) 978-1-84872-863-9 (Paperback)

Library of Congress Cataloging-in-Publication Data

Heck, Ronald H.
Multilevel and longitudinal modeling with IBM SPSS/ Ronald H. Heck, Scott L. Thomas, Lynn N. Tabata.
p. cm. -- (Quantitative methodology series)
Includes bibliographical references and index.
ISBN 978-1-84872-862-2 (hbk. : alk. paper) -- ISBN 978-1-84872-863-9 (pbk. : alk. paper)
1. Social sciences--Longitudinal studies. 2. Social sciences--Statistical methods. 3. PASW (Computer file) 4. SPSS (Computer file)
I. Thomas, Scott Loring. II. Tabata, Lynn Naomi. III. Title.

HA32.H39 2010
005.5'5--dc22 2009053643

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Psychology Press Web site at
http://www.psypress.com

Contents

Preface

Multilevel modeling has become a mainstream data analysis tool over the past decade, emerging from a somewhat niche technique in the late 1980s to a technique now figuring prominently in a range of social and behavioral science disciplines. As the approach gained popularity over the 1990s specialty software programs began to appear addressing the needs of an ever-widening group of users. Eventually, mainstream statistics packages such as SPSS, SAS, and Stata began to include routines for multilevel modeling in their programs.

While some devotees of the various mainstream packages began making use of this new multilevel modeling functionality, progress toward carefully documenting these routines was slow, thereby hindering meaningful access to the average user. In the meantime, the specialty software packages were becoming increasingly refined and accessible, offering the user a growing number of generalizations of the traditional multilevel model. In some ways the software proved to be both driving and limiting the development of the field.

The various approaches to multilevel modeling represented in these packages have in some ways made it difficult for a clear lingua franca to emerge and have long challenged those interested in teaching these techniques. In addition to the considerable expense of purchasing the better-documented specialty programs, there is also the additional challenge of mastering the new programming logic, syntax, and file structure unique to each program. This is the first book to demonstrate how to use the multilevel and longitudinal modeling techniques available in SPSS (version 18). We have devoted our energy in this book to addressing this problem and to providing both the budding or seasoned multilevel analyst with a set of concepts and programming skills within the SPSS environment. We have designed this book to enable the development, specification, and testing of a range of multilevel models using a statistical program, SPSS, that is standard in many graduate programs and organizations around the world.

Drawing on years of our own teaching and our work explicating the multilevel approach (Heck & Thomas 2009), we have chosen to adopt a workbook format here. Our intent is to help readers set up, run, and interpret a variety of different types of introductory multilevel and longitudinal models using the linear mixed-effects model (Mixed) procedure in SPSS. The routine enables users to fit linear-mixed effects models with continuous outcomes. We provide a concise conceptual treatment of the multilevel approach and then walk readers in a step-by-step fashion through data management, model conceptualization, and model specification issues related to the multilevel model. We offer multiple examples of several different types of multilevel models, carefully showing how to set up each model and how to interpret the output. Most chapters feature an extended example illustrating the logic of model development. These examples show readers the context and rationale of the research questions and the steps around which the analyses are structured. Annotated screen shots from SPSS are provided to help guide users through the program. We also provide an introduction to diagnostic tools, data management, and relevant graphics. Readers can work with the various examples developed in each chapter by using the corresponding data files on the book-specific web site at www.psypress.com/multilevel-modeling-techniques/. The screen shots provided and the supporting syntax statements in the workbook's appendix (also available online) should facilitate learning the various techniques developed sequentially in each chapter.

The workbook begins with an introductory chapter highlighting several relevant conceptual and methodological issues associated with defining and investigating multilevel and longitudinal models, followed by a discussion of SPSS data management techniques which we have found facilitate working with multilevel, longitudinal, and/or cross-classified data sets. In the next two chapters, we detail the basics of multilevel modeling, how to develop a multilevel model, and trouble-shooting techniques for common programming and modeling problems. We develop several models for investigating individual and organizational change in Chapters 5 and 6, followed by an introduction to multilevel models with multivariate outcomes in Chapter 7. Chapter 8 illustrates SPSS's facility for examining models with cross-classified data structures, a type of hierarchical structure that greatly expands the possibilities for following subjects through multiple organizational units or subunits over time. We conclude with thoughts about ways to expand on the various multilevel and longitudinal modeling techniques introduced and issues to keep in mind in conducting multilevel analyses. We hope the workbook becomes a useful guide to readers' efforts to learn more about the basics of multilevel and longitudinal modeling and the expanded range of research problems that can be addressed through their application.

Ideal as a supplementary text for graduate level courses on multilevel, longitudinal, latent variable modeling, multivariate statistics, and/or advanced quantitative techniques taught in departments of psychology, business, education, health, and sociology, we hope the workbook's practical approach will also appeal to researchers in these fields. We believe the workbook provides an excellent supplement to our other multilevel book, *An Introduction to Multilevel Modeling Techniques, 2nd edition*; however, it can also be used with any multilevel and/or longitudinal modeling book or as a stand-alone text.

Several people have played an important role in the development of this workbook. In particular, we wish to thank our reviewers: Karen A. Barrett of Colorado State University, Jason T. Newsom of Portland State University, Debbie L. Hahs-Vaughn of the University of Central Florida, and Dick Carpenter of the University of Colorado, Colorado Springs. Our series editor, George Marcoulides of the University of California, Riverside; Debra Riegert, our Senior Editor; Erin Flaherty, our Senior Editorial Assistant; and the Project Editor at Taylor & Francis, Michael Davidson, have all been incredibly supportive throughout the process. While we remain responsible for any errors remaining in the text, the book is much stronger as a result of their support and encouragement.

Ronald H. Heck
Scott L. Thomas
Lynn N. Tabata

Chapter 3

Defining a Basic Two-Level Multilevel Regression Model

This chapter introduces the basic approach to two-level, multilevel modeling. The material is challenging because the models are more complex than the general linear model, with which most readers will be familiar with from their basic statistics courses. Like everything else, however, one has to start somewhere. The general concepts we present in this chapter become more familiar as one reads more research that makes use of multilevel techniques. After first reviewing some basic concepts of the single-level multiple regression model, we develop the basic steps of conducting a multilevel regression analysis using an extended example. Our intent is to develop the rationale behind the specification of this general class of models in a relatively nontechnical manner and to illustrate its use in an applied research situation. The methods presented in this chapter should provide a basis for the application of these techniques to a wider set of research problems in the chapters that follow.

From Single-Level to Multilevel Analysis

Linear models (e.g., analysis of variance [ANOVA], analysis of covariance, multiple regression, multivariate analysis of variance [MANOVA]) have long been used in the social sciences to analyze data from experimental, quasi-experimental, and nonexperimental designs. Univariate analysis such as multiple regression is concerned with examining the variability in a single outcome (or dependent) variable from information provided by one or more predictor (or independent) variables. Multivariate analysis (e.g., MANOVA, factor analysis) is the more general case of univariate analysis; that is, it facilitates the examination of multiple independent and dependent variables in one simultaneous model. A commonality between these univariate and multivariate approaches, however, is that they are confined to single-level analyses; that is, either individuals are the unit of analysis or groups (aggregates) are the unit of analysis.

Multiple linear regression requires a continuous dependent variable (i.e., measured on an interval or ratio scale) and can handle both continuous and dichotomous (e.g., gender) independent variables. It cannot handle categorical variables (i.e., referred to as factors in analysis of variance terminology) without recoding them in some way. There are two broad conceptual approaches to the regression model, predictive and explanatory. Through the predictive approach, the analyst uses the multiple regression model to optimize predictions about an outcome based on values of a set of independent variables. The linear regression model assumes that a unit increase in the independent variable is related to an expected constant change in the dependent variable. For example, we might wish to predict someone's likely starting salary in a new job if she or he has a certain level of education and experience. For the linear model to hold, it is assumed that an increase in education (or experience) will bring an expected similar change in starting salary, regardless of where someone starts in terms of education. In this type of single-level regression model, the coefficients that describe the prediction equation (i.e., the intercept and slope coefficients for each predictor in the model) are generally considered as fixed values in the population estimated from the sample data. For this type of research purpose, the focus of the analysis is primarily on the efficiency of the prediction and the parsimony of variables included in the prediction equation (in other words, the analyst hopes to make the best predictions using the smallest number of variables).

The second broad approach is explanatory rather than predictive. Through the explanatory approach, the analyst sets out to determine how a set of independent variables affects a dependent variable and to estimate the magnitude of the effects for each independent variable. For example, existing research may suggest that a particular model (for example, consisting of identified market processes, individual background, and perhaps organizational factors) interact in a way that influences beginning salary. The focus in this type of study rests on the correct specification and testing of a theoretical model that is under consideration. In this case, it is important to include in the model a set of variables identified as important by theory and previous research. More specifically, the researcher formulates a model from theory, tests the model against the data, and determines how well the empirical test of the model conforms to theoretical expectations.

Of course, these goals are not mutually exclusive. We distinguish between these two goals, however, because in predictive studies variables might be retained in a model only because they are statistically significant and dropped simply because they are not (Heck & Thomas, 2009). In other words, theory would not enter into decisions about model efficiency. In contrast, in the explanatory approach, the specification of the theoretical model should be carefully considered, and subsequent changes (i.e., whether to add or remove a variable from a model) should be made sparingly with careful attention to theory. Otherwise, it may be difficult to attach any substantive meaning to the final model. This latter point has particular relevance to the investigation of multilevel data (i.e., data on individuals and the groups they define) that tend to go with more complex theories about how processes operate across multiple social groupings.

Although researchers were aware of problems due to the nesting of individuals within higher level units of the data hierarchy in the past, the presence of similarities among individuals in the same groups did not enter directly into single-level analyses. For example, in analyses of large-scale survey data, the analyst typically applied sample weights to address the oversampling of some subgroups in the data set (e.g., by socioeconomic status, by ethnicity). Failure to account for similarities among individuals (due to grouping) within the study, however, can lead to biased estimates of model parameters and therefore erroneous conclusions about the effects of some predictors in the model (Thomas & Heck, 2001).

Multilevel modeling represents a compromise between modeling each unit separately and modeling all unit contexts simultaneously within the same model (Kreft & de Leeuw, 1998). These models obviate the forced choice of conducting either an individual-level analysis or a group-level analysis. We use the term *multilevel model* with respect to two separate statistical objectives described within one model. The first objective concerns inferences made about a model's structural parameters (Morris, 1995), often referred to as the model's fixed effects. The second objective concerns inferences about the unknown variance parameters in the model, referred to as the random parameters (Morris, 1995). Although researchers are generally most interested in the model's structural parameters, the distribution of a model's random parameters (e.g., variances, covariances) is also of interest.

There are several advantages of multilevel analysis over traditional single-level univariate and multivariate approaches (Heck & Thomas, 2009). First, as we have stated, multilevel analysis helps researchers avoid the choice of individuals or groups as the unit of analysis. Second, it allows researchers to deal with more complicated sampling strategies. Single-level analyses are based on the assumption of simple random samples. In many data collection strategies, however, individuals may be sampled within the same neighborhoods or schools, or subgroups of individuals (e.g., by ethnicity or SES) may be oversampled compared with their representation in the population. Such complex sampling strategies create clustering effects that violate the assumptions of simple random sampling (i.e., that every individual has an equal chance of being selected in the sample). Third, where similarities among individuals are present (e.g., clustering effects due to sharing similar circumstances), multilevel models are acknowledged to provide more accurate estimates of model parameters than single-level analyses (Hox, 2002). This is primarily due to their greater accuracy in calculating standard errors associated with parameter estimates. Because hypothesis tests are based on the ratio of the unstandardized estimate to its standard error, ignoring the presence of nested data structures can lead to underestimating standard errors and, therefore, false inferences about the significance of model parameters (Thomas & Heck, 2001). Fourth, multilevel analysis allows the researcher to define variables at their correct theoretical level of the data hierarchy. So, for example, in a two-level hierarchy, a variable such as school size can be determined with respect to the number of schools in the sample, while a variable like gender can be evaluated with respect to the number of individuals in the sample. Finally, multilevel modeling allows researchers to ask more complex questions about the data. One is about the distribution of outcomes (e.g., means or regression slopes) across a sample of groups (such as schools). We may attempt to determine what types of school variables might account for variability in school outcomes, net of the individuals within the school. Examples might be the quality of a school's teachers, its leadership, and its classroom processes.

Building a Two-Level Model

The basics of multilevel modeling involve the investigation of randomly varying outcome parameters. These typically include variation in the levels of the outcome (intercepts) and the strength of within-group relationships indicated by regression coefficients (slopes) across groups. Once we identify that variation exists in the parameters of interest, we can build models to explain this variation. As we suggested in Chapter 1, in some cases, we may have a specific theoretical model in mind that we wish to test; in others, however, we might be trying to explore possible new mechanisms that explain this variation.

Consider an analysis where the researcher wishes to determine whether there is an association between a predictor, X, such as student socioeconomic status (SES) and an outcome, Y, such as a math test score. Because the current educational policy context in the United States demands increasing accountability for student outcomes, schools are accountable for reducing gaps in achievement due to students' social backgrounds. Such concerns are related to the social distribution of learning within schools (e.g., see Lee & Bryk, 1989). Ideally, we may wish to identify school settings where achievement is generally high for all the students in the school and where there is little or no relationship between student SES and outcomes. Such schools would be considered both effective (i.e., producing high achievement outcomes) and equitable (i.e., having little or no social distribution of learning due to students' social backgrounds). In contrast, we may also wish to identify schools where achievement is consistently low for students and where student SES background is consequential for their outcome levels. We might be able to intervene effectively (e.g., increase teacher quality, reallocate resources) if we could identify such settings that are ineffective and inequitable for students.

Research Questions

Our first research question focuses on whether student achievement in math varies across schools. We can ask simply: Does student math achievement vary across schools? We might then investigate the relationship between students' socioeconomic status and their math achievement. Second, we might ask whether the effects of individual SES tend to compound at the school level to influence student math achievement; that is, do both individual-level SES and school-level aggregate (or average) student SES influence math achievement? Third, we investigate whether features of schools' contexts (i.e., student composition, type of school) and their academic environments (i.e., the number of students planning to attend 4-year universities after they graduate from high school) affect the relationship between individual student SES and math achievement. More specifically, we ask: Do features of schools' contexts and academic environments moderate the relationship between individual student SES and math achievement? Our research questions, therefore, provide an illustration of building a two-level model to investigate (1) a randomly varying intercept (math achievement level) and, subsequently, (2) a randomly varying slope (i.e., the individual SES–math achievement relationship).

The Data

The data set used in this example consists of 6,871 secondary students in 419 schools. We will begin with a single-level analysis (i.e., considering only the students and not their nesting within schools) as a starting point. One typical research question for a single-level analysis might be: Is there a relationship between students' SES background and their achievement levels in math? We might hypothesize that socioeconomic status is positively related to the subject's score on the math test. The single-level multiple regression model to explain an individual's (*i*) math achievement outcome would be

$$Y_i = \beta_0 + \beta_1 (\text{SES})_i + \varepsilon_i , \tag{3.1}$$

where β_0 is the intercept, β_1 is a slope parameter, and ε_i represents error in predicting individual outcomes from the equation. The intercept represents the expected math achievement score for a student whose SES is 0. It is often useful to consider the scaling of the independent variable or variables in a model. In this simple case, because SES (which is a continuous variable) was standardized (i.e., rescaled into a standardized score, or *z*-score), the mean is 0 and the standard deviation (*SD*) is 1.0. This is often a convenient scale for continuous variables in multilevel modeling, since a score of 0 on SES, therefore, represents score for a person whose SES background is equal to the SES grand mean for the sample. We discuss these types of "centering" decisions in further detail in the next chapter. The slope coefficient (β_1) represents the expected change in math achievement for a one-unit (in this case 1 *SD*) change in SES.

The key point about a single-level model is that the estimates of the intercept and slope are each fixed to one value that describes the average for the sample. For example, the slope expressing the relationship between SES and math scores will be the same across all cases. This also means that the errors (ε) in estimating the intercept and slope parameters are assumed to be independent, normally distributed, have constant variance, and a mean of zero.

To examine her research question preliminarily, the researcher might first develop a scatterplot of the relationship between student SES and math achievement. To illustrate this in relation to single and multilevel designs, we will develop a scatterplot for the first 80 students in the data set. The resulting graph is summarized in Figure 3.1. The figure suggests that as student SES increases, so do math scores. The goal of the overall analysis is to determine what the best fitting line that describes the relationship between student SES and test scores in this sample. This is accomplished by estimating values for the intercept and slope. Of course, once we estimate the predicted values for each subject on the two variables, there will be a discrepancy between the predicted values (which would lie on the line) and subjects' actual values on the SES and math test score measures. The difference between observed and predicted values is represented as error. The intercept coefficient represents the average level of student scores when SES is zero (which represents a mean adjusted for SES) and the slope represents the average effect of SES on the math score across the sample of students. These values become "fixed" for the entire sample; that is, because individuals are randomly sampled, it is assumed that the value represents population averages.

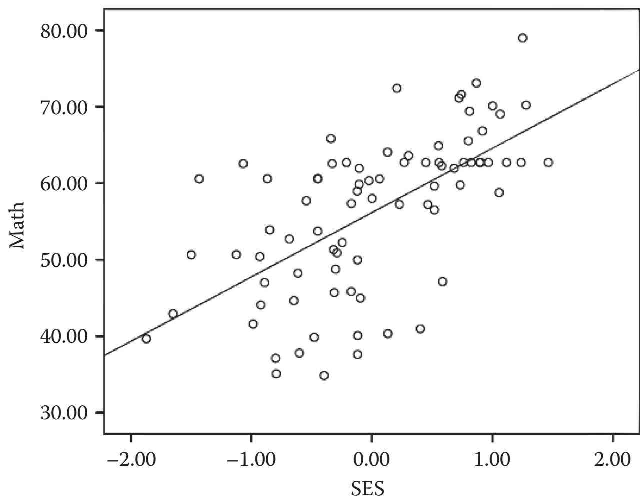


FIGURE 3.1 Regression line describing the fixed intercept and slope for student SES and math achievement.

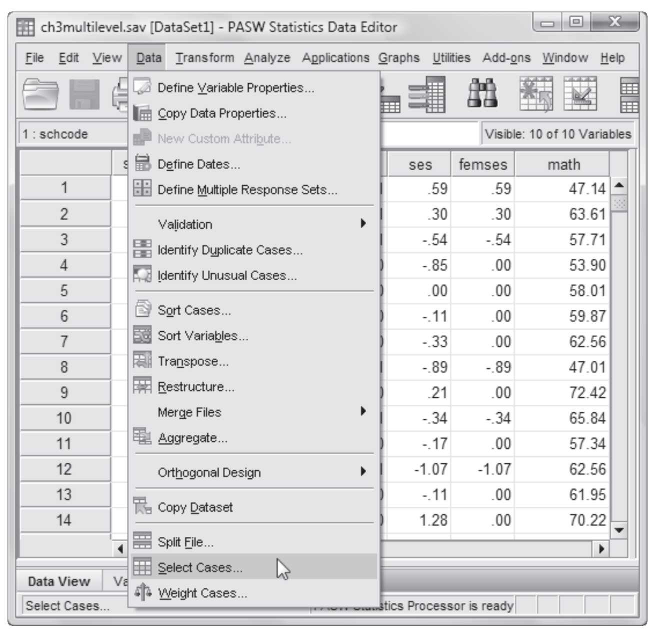
Fixing the values of the intercept and slope results in the regression line in Figure 3.1 that summarizes the relationship between SES and math test scores. The principle of least squares states that the correct regression line is the one that best fits the data points. Model fit is assessed by summing the squared distances of each observed value from the predicted value that rests on the regression line. The line that minimizes the sum of these squared distances (they are squared to cancel out positive and negative errors above or below the line) is said to fit the data best; hence, the term least squares regression (Neter, Kutner, Nachtsheim, & Wasserman, 1996). In the linear regression model, the error term is a random source of variation, which we assume is zero on average and normally distributed, varies independently of *X*, and has constant variance across all levels of *X*. Interested readers can reproduce the scatterplot in Figure 3.1 in SPSS menu commands.

Graphing the Relationship Between SES and Math Test Scores With SPSS Menu Commands

Launch the SPSS application and select the data file *ch3multilevel.sav*.

1. Go to the toolbar and select DATA, SELECT CASES.

This command will open the *Select Cases* dialog box.



of the six schools in our illustrative subsample (Figure 3.2) have positive slope relationships, and one school has a negative slope relationship. One can also notice that the R^2 coefficients (which summarize the strength of the relationship between student SES and math) also vary considerably in the first six schools (i.e., from 0.02 to 0.21). This suggests that there might be schools where there is little gap in math achievement due to students' SES background. These schools could be viewed as more equitable in terms of the social distribution of learning due to students' social backgrounds (e.g., SES or perhaps gender and race/ethnicity if we were to include those variables).

In other schools, students' social background might be very consequential in determining their achievement. A second question that could be posed: Do the slopes (i.e., the strength of relationship between student SES and math achievement) vary across schools? Answering this question would provide information about schools' equity in producing outcomes, given the social backgrounds of their students. We might examine whether the relationship between individual SES and math achievement is stronger or weaker in schools of differing average social composition. Moreover we can investigate whether there is a difference in the strength of association between individual SES and achievement in public and private schools or in schools having a stronger academic focus, after controlling for the mean SES level at the school.

Parameters that are proposed to vary randomly across units are referred to as random effects or random coefficients from various statistical perspectives. In experimental research, for example, a *random effect* describes a situation where the treatment (or levels of a treatment) is assumed to represent a sample drawn from a universe of treatments or treatment levels. Because the effect is considered as randomly varying across a universe of treatments, the intent is to make inferences beyond the specific treatment levels included in the study. The effects, therefore, are not assumed to be constant. In contrast, a *fixed effect* describes the situation where all possible treatments are present in the experiment (Kreft & de Leeuw, 1998). In this latter case, inferences can only be made about the specific treatments used. The effects are considered to be constant and measured without error because all possible cases are included.

Unlike single-level (ordinary least squares [OLS]) regression, where random errors are assumed to be independent, normally distributed, and have constant variance, in multilevel models, error structures are more complex. The individual-level errors are dependent within each unit because they are common to every individual within that unit (Heck & Thomas, 2009). Errors do not have constant variance because the residual components describing intercepts and slopes may also vary across units. The estimation of these unknown random parameters associated with intercepts or slopes may also depend on characteristics of the data (e.g., sample size, degree of imbalance in sample sizes of higher level units, degree of similarity among individuals within groups), the type of analysis conducted, and the measurement scale of the dependent variables (Muthén & Muthén, 1998–2006; Raudenbush & Bryk, 2002). Because the model's random parameters must be estimated with group samples containing differing numbers of individuals, iterative estimation procedures must be used to obtain efficient estimates (Muthén & Muthén, 1998–2006).

Building a Multilevel Model With SPSS Mixed

We can use SPSS Mixed to run a variety of different multilevel cross-sectional and longitudinal (e.g., growth) models. The program is flexible and can be used to estimate a number of different types of models with random intercepts (i.e., means that vary across groups) and random slopes (i.e., within-group regression coefficients that vary across groups). It is also useful in looking at individual change over repeated measurements or in studies of changes of individuals within organizations over time. In addition, it can also be applied to situations where individuals may be cross-classified by higher level groupings (e.g., in several different classrooms within schools or within various high schools and subsequent universities). In the chapters that follow we develop each of these possibilities in more detail. In the remainder of this chapter, we focus on the univariate cross-sectional multilevel model.

There are several ways to develop models using SPSS Mixed. Some users prefer the SPSS graphical user interface (GUI), while others favor using syntax statements to define the model. We will build models in the chapters that follow using the GUI, but also provide some of the key syntax in Appendix A. As we mentioned in Chapter 2, syntax can be very useful on occasions, since one does not have to return to the specific window to make successive changes in a model that is being investigated. Syntax also provides a record of what has been done previously. This record can be saved and used on subsequent occasions without having to set the model up again through the SPSS menu system (i.e., GUI). We have found small differences may result depending on whether models are developed with syntax commands or the GUI (e.g., owing to some default commands that users may not specify with the syntax commands). Also we note that if users have different versions of SPSS (e.g., version 15, 16, 17, or 18), estimates may be slightly different (owing to different default rounding procedures). In this chapter, we take the reader through the steps in the SPSS GUI to develop the basic two-level regression model. An alternative to writing your own syntax is to develop models using the SPSS GUI and then tell SPSS to generate the syntax (which is then pasted into a syntax window). Although we use the GUI interface, we will also demonstrate how to generate syntax through the menu system. We follow this convention throughout the remainder of the book. Note that all screen shots from the SPSS GUI provided here are based on SPSS for Windows version 18.0.

Step 1: Examining Variance Components Using the Null Model

There are three distinct steps in developing the multilevel model. We develop these in this chapter in the following order: (1) specification of the null, or no predictors model; (2) specification of the Level 1 model; and (3) specification of the Level 2 model. This latter step can include the model to explain intercepts and the model or models to explain randomly varying slopes. The first step in a multilevel analysis usually is to develop a null (or no predictors) model to partition the variance in the outcome into its within- and between-groups components. This will help the researcher determine how much of the variance in math achievement lies between the schools in the sample. Notice that in Equation 3.2, we add a subscript for schools (j). The null model for individual i in school j can be represented as

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}, \tag{3.2}$$

where β_{0j} is the intercept and ε_{ij} represents variation in estimating individual achievement within groups. Between groups, variation in intercepts can be represented as

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{3.3}$$

Through substitution, the null model can be written as

$$Y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}. \tag{3.4}$$

The null model therefore provides an estimated mean achievement score for all schools. It also provides a partitioning of the variance between Level 1 (ε_{ij}) and Level 2 (u_{0j}). Altogether there are three effects to estimate: the intercept, the between-school variation in intercepts (u_{0j}), and the variation in individual scores within schools (ε_{ij}).

This model also provides a measure of dependence within each Level 2 unit by way of the intraclass correlation (ρ). The intraclass correlation (or ICC) describes the proportion of variance that is common to each unit, as opposed to variation that is associated with individuals within their units. It can be thought of as the population estimate of the amount of variance in the outcome explained by the grouping structure (Hox, 2002). The proportion of variance found between groups can be calculated in SPSS by using either the Variance Components or Mixed procedures. Both will give the same estimation of within-groups and between-groups variance components. The ICC can be represented as

$$\rho = \sigma_B^2 \Big/ \left(\sigma_B^2 + \sigma_W^2\right), \tag{3.5}$$

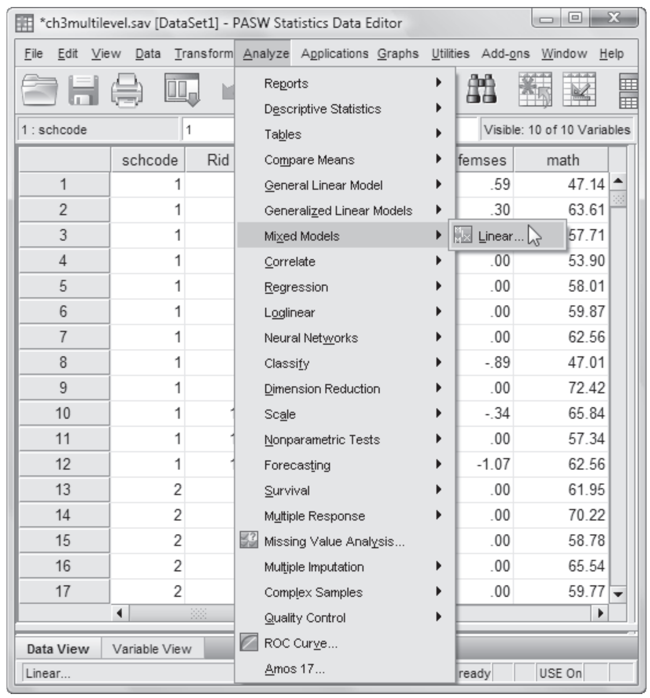
where σ^2 represents the variance and B and W stand for between groups and within groups, respectively. Stated differently, the ICC is the ratio of between-groups variance to the total variance. The higher the ICC, the more homogeneous are the units (i.e., there exists substantial variability between schools). In contrast, if the ICC is quite small (i.e., researchers often use 0.05 as a rough "cutoff" point), then there would be little advantage to conducting a multilevel analysis. Simply put, the higher level grouping does not affect the estimates in any meaningful way. In these cases, a single-level analysis conducted at the individual level would suffice.

Defining the Null Model With SPSS Menu Commands

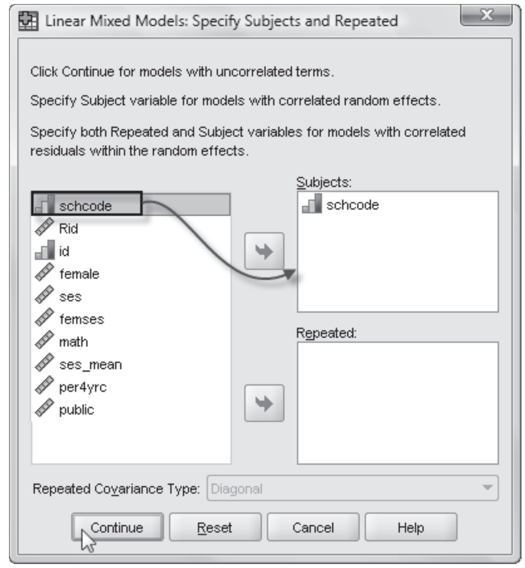
Note: If continuing after creating Figures 3.1 and 3.2, remove the *Select Cases* conditional setting before proceeding. This is achieved by going to the SPSS toolbar, and selecting DATA, SELECT CASES, RESET, OK.

1. Go to the SPSS toolbar and select ANALYZE, MIXED MODELS, LINEAR.

This command opens the *Linear Mixed Models: Specify Subjects and Repeated* dialog box.

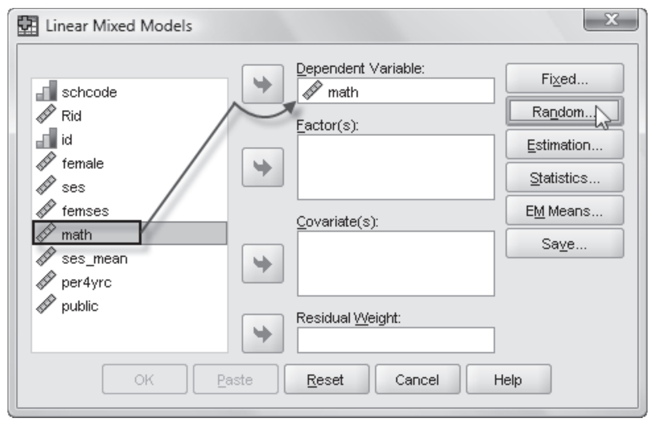


2. From the *Linear Mixed Models: Specify Subjects and Repeated* dialog box, click to select the *schcode* variable from the left column. Then click the right-arrow button to transfer *schcode* into the *Subjects* box.



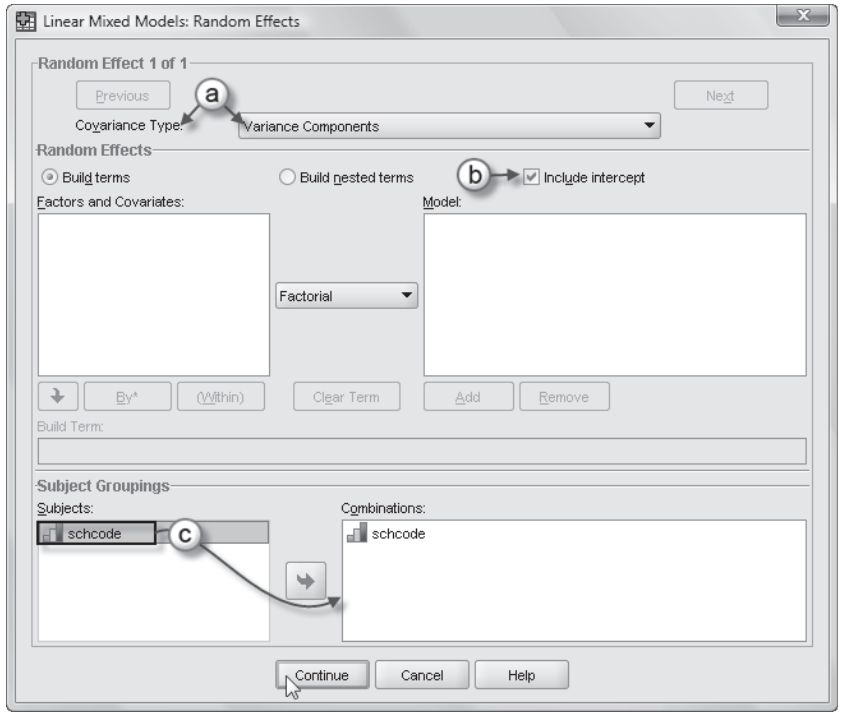
Click the CONTINUE button to display the *Linear Mixed Models* dialog box.

3. The *Linear Mixed Models* dialog box is used to define the dependent variable as well as any factors (categorical variables) or covariates (interval variables) in the analysis. Click to select the *math* variable from the left column. Then click the right-arrow button to transfer *math* into the *Dependent Variable* box.



In this case, there are no predictors in the "null" model. Since there are no fixed effects in the model, skip over the FIXED button and click the RANDOM button to access the *Linear Mixed Models: Random Effects* dialog box.

4. The RANDOM command is used to specify the random effects; that is, which variables are to be treated as randomly varying across groups.
   a. At the top of the *Linear Mixed Models: Random Effect 1 of 1* display screen, *Covariance Type* refers to several different types of covariance matrices that can be used for multilevel analyses. The default is VARIANCE COMPONENTS (VC), which we will use for the first model.



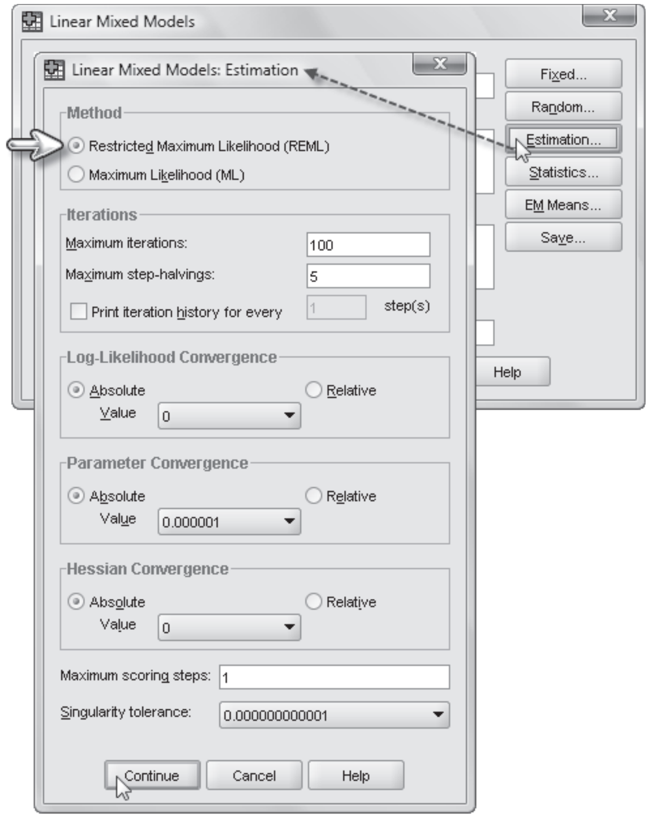
VC is the default covariance structure for random effects. This specifies a diagonal covariance matrix for the random effects; that is, it provides a separate variance estimate for each random effect, but not covariances between random effects.

In this case, there is only one random effect (the intercept), so we can use the default VC. For models with random intercepts and slopes, a common choice is an "unstructured" (UN), or a completely general, covariance matrix, which fits all variances and covariances between random effects.

   b. In the *Random Effect 1 of 1* display screen you can build other random effects for the predictors. In this case, however, only the intercept is going to be randomly varying so click to select the *Include intercept* option.
   c. Click to select the *Subject Groupings* variable *nschcode*. Then click the right-arrow button to transfer the variable into the *Combinations* box.

Click the CONTINUE button to return to the *Linear Mixed Models* dialog box.

5. In the *Linear Mixed Models* dialog box, click the ESTIMATION button to access the *Linear Mixed Models: Estimation* dialog box. The estimation method choices are maximum likelihood (ML) or restricted maximum likelihood (REML). In ML, both regression coefficients and variance components are included in maximizing the likelihood function; that is, the process of minimizing the difference between the sample covariance matrix and the model-implied covariance matrix. In REML, only the variance components are included in estimating the likelihood function; thus, REML is a restricted solution.

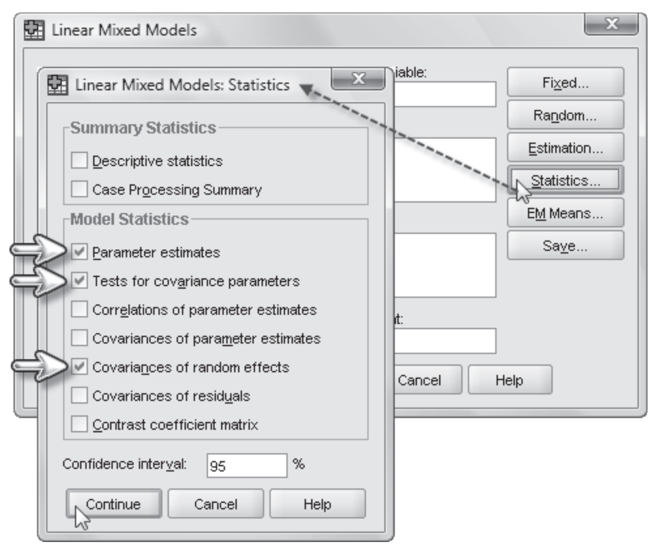


Because in REML the regression coefficients are considered to be unknowns, taking the loss in degrees of freedom due to estimating P + 1 regression coefficients in the model results in unbiased estimates of variance components when there are small numbers of groups (Snijders & Bosker, 1999). With sufficient numbers of groups, the differences in estimation methods are negligible. Restricted maximum likelihood (REML), the default estimation method for both SPSS techniques, which also is the better choice with small data sets, will be used to develop the variance component estimates.

We will use the REML default setting.

Click the CONTINUE button to return to the *Linear Mixed Models* dialog box.

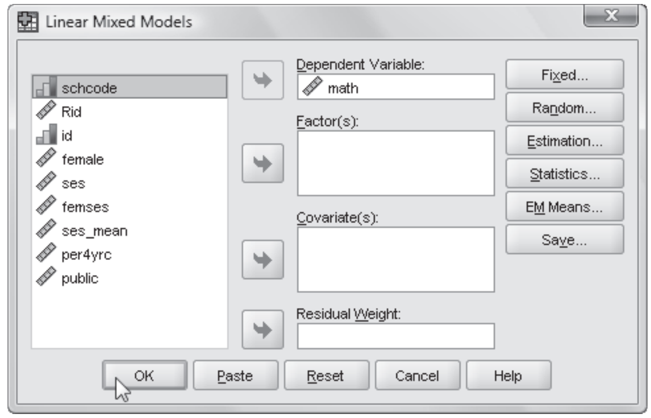
6. In the *Linear Mixed Models* dialog box, click the STATISTICS button to access the *Linear Mixed Models: Statistics* dialog box.



Click and select the following three statistics: Parameter estimates, tests for covariance parameters, and covariances of random effects.

Then click the CONTINUE button to return to the *Linear Mixed Models* dialog box.

7. Finally, in the *Linear Mixed Models* dialog box, click the OK button to run the model.



Interpreting the Output From the Null Model

The first table (Table 3.3) in the resulting output summarizes the total number of parameters being estimated (3). This is the same as we noted in Equation 3.4. The total parameters estimated include the fixed effect value for the intercept, random Level 2 variance, and the Level 1 variance (referred to as "Residual" in the SPSS output).

The column referred to as "Number of Levels" describes the fixed effects (1) and the number of random effects (1). There is one fixed effect to be estimated (the intercept) and one random effect (the randomly varying intercept). The column referred to as "Subject Variables" indicates the number of levels in the analysis (i.e., in this case, schcode [the school identifier] implies a two-level analysis). The covariance structure describes the way the covariance matrix of random effects is dimensionalized at the group level. In this case, we use the default (VC, or variance components), which provides an estimate of the intercept variance (σ_I^2). However, in this first example, at Level 2 there is no random slope variance (σ_S^2) or covariance between the intercept and slope. In this case, the VC covariance matrix will be the same as an identity matrix.

The next piece of SPSS output describes model-fitting criteria (Table 3.4). This can be useful in examining the improvement of model fit when comparing two successive (or nested) models (e.g., using the change in log likelihood). Note that comparing nested models should only be done when using ML estimation (and not REML, unless only the random parameters are compared, see Heck & Thomas, 2009 for a complete treatment of these estimators). This output also includes other type of information about overall model fit. We will discuss model fit in further detail in Chapter 4.

Table 3.5 reports the estimates of fixed effects in the model. The intercept (or grand mean for school outcomes) is estimated as 57.67.

The variance component output (Table 3.6) indicates the proportion of variance in achievement that lies between schools is .138. This can calculated be from Equation 3.5

TABLE 3.3 Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
Random Effects	Intercept[a]	1	Variance Components	1	schcode
Residual				1	
Total		2		3	

[a] As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.
[b] Dependent Variable: math.

TABLE 3.4 Information Criteria[a]

–2 Restricted Log Likelihood	48877.256
Akaike's Information Criterion (AIC)	48881.256
Hurvich and Tsai's Criterion (AICC)	48881.257
Bozdogan's Criterion (CAIC)	48896.925
Schwarz's Bayesian Criterion (BIC)	48894.925

Note: The information criteria are displayed in smaller-is-better forms.
[a] Dependent Variable: math.

TABLE 3.5 Estimate of Fixed Effects[a]

Parameter	Estimate	Std. Error	*df*	*t*	Sig.
Intercept	57.674234	.188266	416.066	306.344	.000

[a] Dependent Variable: math.

TABLE 3.6 Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald *Z*	Sig.
Residual		66.550655	1.171618	56.802	.000
Intercept [subject = nschcode]	Variance	10.642209	1.028666	10.346	.000

[a] Dependent Variable: math.

[10.642/(10.642 + 66.551) = 10.642/77.193], or 13.8%. The intraclass correlation provides a sense of the degree to which differences in the outcome Y exist between Level 2 units; that is, it helps answer the question of the existence or nonexistence of meaningful differences in outcomes between the Level 2 units. The results of the null or no-predictors model (basically a one-way ANOVA analysis) suggest that the development of a multilevel model is warranted. Because intercepts vary significantly across schools (Wald $Z = 10.346$, $p < .001$), and the ICC suggests that about 13.8% of the total variability in math scores lies between schools, we can develop a multilevel model first to explain this variability in intercepts within and between schools.

The reliability of the sample mean for any unit as an estimate for its population mean can also be assessed with information drawn from the variance components. This can provide the analyst with a means by which the assumption of differences in outcomes across units can be checked. Because sample sizes are likely to differ within each unit j, this reliability will vary across the Level 2 units. Reliability within any particular unit can be estimated as

$$\lambda = \frac{\sigma_B^2}{\left[\sigma_B^2 + \left(\sigma_W^2 / n_j\right)\right]}. \qquad (3.6)$$

In this example, the within-group sample sizes range from 12 to 37. Using the within-group sample sizes we can estimate the reliability for the smallest unit as 10.64/[10.64 + (66.55/12)] = 0.657. This is contrasted with the school that has 37 students tested: 10.64/[10.64 + (66.55/37)] = 0.855.

Our first type of multilevel question was: Do the average math scores students receive vary across schools in the sample? We can answer that question from Table 3.6. The Residual parameter describes the variance due to individuals within groups. As the table suggests, there is significant variance to be explained within groups (Wald $Z = 56.802$, $p < .001$). Similarly, the intercept parameter indicates that the intercepts vary significantly across the sample of schools

(Wald $Z = 10.346$, $p < .001$). The Wald Z test provides a Z statistic summarizing the ratio of the estimate to its standard error.

Researchers have noted several problems with this statistic, however, that analysts may want to keep in mind. First, the Wald Z test is a two-tailed test. Because variances cannot be smaller than 0 (i.e., the null hypothesis is that the parameter = 0), the test should be conducted as a one-tailed test (Hox, 2002). This means that when testing variances (as in Table 3.6), the significance level should be divided by 2 to reflect a one-tailed probability level. Of course, in this instance this will not make a difference, since the two-tailed p value is very small. Note that for testing covariances between random effects (which can be positive or negative) the two-tailed significance test can be maintained. Second, for a large estimated variance coefficient, the standard error can be inflated, which lowers the Wald statistic value and, therefore, can make it overly conservative. Third, the Wald Z test can also perform poorly under conditions of extreme multicollinearity and in situations with small sample sizes. For small samples, the likelihood-ratio test (which is also provided as SPSS output) tends to be more reliable than the Wald test.

Step 2: Building the Individual-Level (or Level 1) Random Intercept Model

In a multilevel analysis, we work primarily with three equations: a within-group (or individual-level) equation, a between-groups intercept equation, and a between-groups slope equation. For each individual i in school j a proposed model similar to Equation 3.1 (summarizing the effect of student SES on math achievement) can be expressed as

$$Y_{ij} = \beta_{0j} + \beta_1(\text{SES})_{ij} + \varepsilon_{ij}. \tag{3.7}$$

Equation 3.7 suggests that at the individual level, within-groups student SES is related to achievement levels.

Equation 3.8 (which is the same as Equation 3.4) implies that variation in intercepts can be described by a school-level intercept (γ_{00}), or grand mean, and a random parameter capturing variation in individual school means (u_{0j}) from the grand mean:

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{3.8}$$

Equation 3.9 implies that a within-unit slope (e.g., SES–achievement) can also be examined as randomly varying across units in the sample:

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{3.9}$$

Equation 3.9 suggests variability in slopes can be described by a school-level average slope coefficient (γ_{10}), or grand mean, and a random parameter capturing variation in individual school coefficients (u_{1j}) from the grand mean. Because the slope is considered to be randomly varying across schools, the corresponding test of significance of the parameter will be based on the number of schools in the sample. Often, in building models, we may treat the within-group slopes as fixed in preliminary analyses (i.e., in situations where we are not testing a particular hypothesized relationship). In the case where we wish to treat the within-group slope as fixed (i.e., it does not vary across schools), Equation 3.9 would be rewritten as

$$\beta_{1j} = \gamma_{10}. \tag{3.10}$$

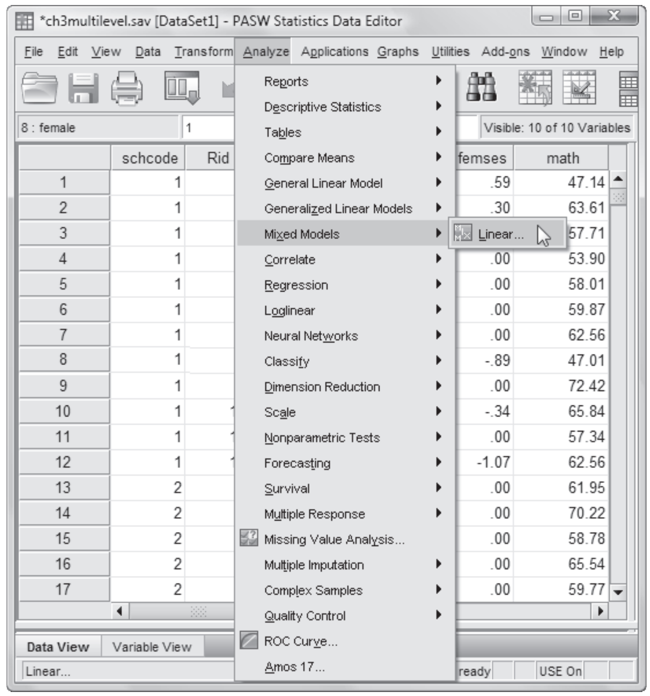
As Equation 3.10 indicates, there is no random component (u_{1j}), so the slope coefficient is fixed to one value for the sample. In the case where the Level 1 slope coefficient is fixed, the significance test for the slope will be based on the number of individuals in the sample. If it turns out that both intercepts and slopes vary randomly across schools, Equations 3.8 and 3.9 suggest that group-level models can subsequently be built to explain variation in the random intercept and slope across groups.

Model 1: Defining the Level 1 Random Intercept Model With SPSS Menu Commands

(SPSS settings will default to those used for the preceding Null Model.)

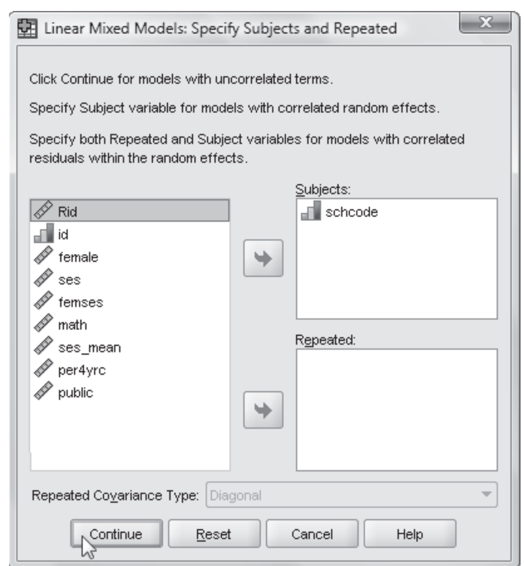
1. Go to the SPSS toolbar and select ANALYZE, MIXED MODELS, LINEAR.

This command opens the *Linear Mixed Models: Specify Subjects and Repeated* dialog box.



2. The *Linear Mixed Models: Specify Subjects and Repeated* dialog box, displays *schcode* in the *Subjects* box.

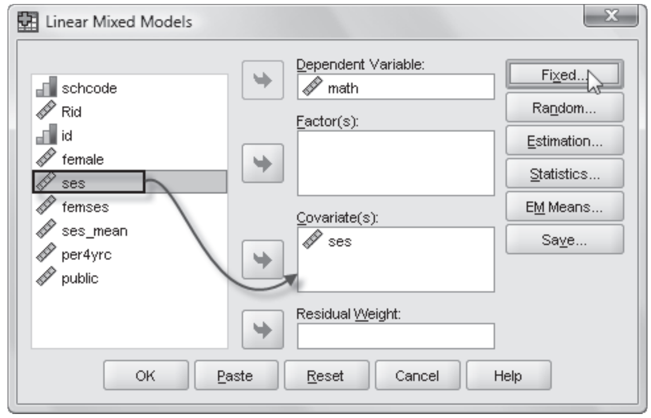
Click the CONTINUE button to display the *Linear Mixed Models* dialog box.



3. The *Linear Mixed Models* dialog box displays *math* in the *Dependent Variable* box.

Locate and click the *ses* variable from the left column. Then click the right-arrow button to transfer *ses* into the *Covariate(s)* box.

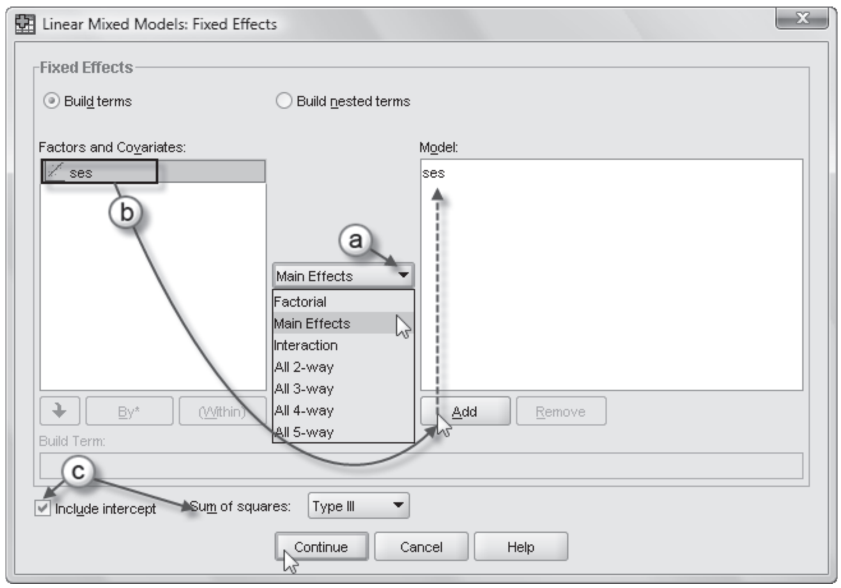
Click the FIXED button to access the *Linear Mixed Models: Fixed Effects* dialog box.



4a. Within the *Linear Mixed Models: Fixed Effects* dialog box, click the pull-down menu to change the factorial setting to *Main Effects*.

b. Click to select *ses* from the *Factors and Covariates* box, then click the ADD button to move the variable into the *Model* box.

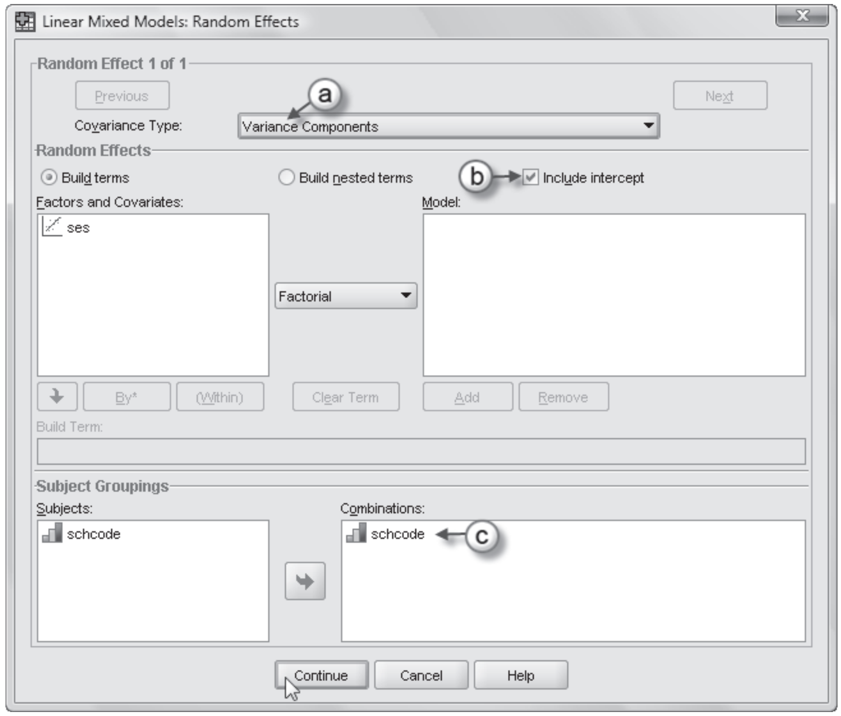
c. Note on the lower left of the screen the intercept is selected and the sum of squares is the default setting.



Click the CONTINUE button to return to the *Linear Mixed Models* dialog box. In the *Linear Mixed Models* dialog box, click the RANDOM button to access the *Linear Mixed Models: Random Effects* dialog box.

5a. The *Random Effect 1 of 1* screen displays the default settings of the prior model. We will stay with variance components (VC) for the covariance structure, since only the intercept will be randomly varying in this within-school model.
b. Confirm *Include Intercept* is selected.
c. Confirm *schcode* appears in the *Combinations* box.

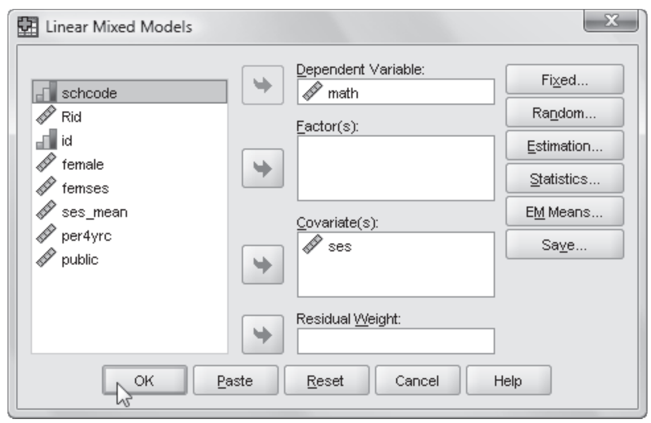
Click the CONTINUE button to return to the *Linear Mixed Models* dialog box.



6. In the *Linear Mixed Models* dialog box, the estimation and statistics settings remain the same as those used for the Null Model so skip over these buttons and click OK to run the model.

Note: It is possible to save predicted estimates and residuals for random and fixed effects using the SAVE button. This would allow us to examine their normality and linearity.

Since no further changes are to be made, click OK to run the model.



We will first build a model to examine variability in intercepts across schools (as in Equations 3.3 and 3.4). The individual-level model is represented as Equation 3.7, which suggests that students' SES background affects math achievement. The school-level model is represented as Equation 3.8, suggesting only that the school-level intercepts vary randomly across schools. We discuss the school-level model in a subsequent section.

Interpreting the Output From Model 1

Following is the SPSS output generated from the Level 1 model. Table 3.7 summarizes the total number of parameters being estimated (4). This fits with Equations 3.8 and 3.9 (suggesting one within-school predictor SES, the intercept, the Level 2 variance, and the residual [or Level 1] variance). The column referred to as "Number of Levels" describes the fixed effects (2) and the number of random effects (1). There are two fixed effects to be estimated (the intercept and SES) and one random effect (the Level 2 variance component describing variability in the intercept across schools in the sample). The covariance structure describes the way the covariance matrix

TABLE 3.7 Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	ses	1		1	
Random Effects	Intercept[a]	1	Variance Components	1	schcode
Residual				1	
Total		3		4	

[a] As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

[b] Dependent Variable: math.

TABLE 3.8 Type III Tests of Fixed Effects[a]

Source	Numerator *df*	Denominator *df*	*F*	Sig.
Intercept	1	375.699	187802.817	.000
ses	1	3914.638	803.954	.000

[a] Dependent Variable: math.

of random effects is dimensionalized at the group level. In this case, we use the default (VC), which provides an estimate of the intercept variance but again no slope variance because we have opted to fix the slope (nor will there be covariance between intercept and slope). This is the same as specifying an identity covariance matrix.

Table 3.8 presents the typical ANOVA table for examining the significance of the fixed-effect parameters in the model. The large *F*-ratio associated with SES in the table suggests that student SES is significantly related to student math scores. The test of the significance of the intercept is generally not of interest (unless perhaps the mean for math has been standardized to 0), as it is merely a test of whether the intercept is 0 in the model. As the table shows, we can reject the null hypothesis that it is 0 (and we already know from the null model and descriptive analysis that it is approximately 57.6).

Table 3.9 provides the estimates of the fixed effect coefficients. First, we can see that the intercept (adjusted for student SES) is 57.596. This represents the average school mean adjusted for student SES. The standard error is 0.133. The degrees of freedom reported for each fixed effect, which reflect the Satterthwaite correction for approximating the denominator degrees of freedom for significance tests of fixed effects in models where there are unequal variances and group sizes, are useful in determining at what level each variable is measured in the model. For example, we know there are 419 schools in the sample. This is consistent with the 375.7 degrees of freedom reported in Table 3.9. In contrast, we know that SES is an individual-level variable. There are 6,871 individuals in the sample, so the degrees of freedom of 3914.6 are consistent with a variable measured at Level 1. Once again, the *t*-test of the significance of the parameter is not really interesting, since it is a test of whether the intercept is equal to 0.

TABLE 3.9 Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	*df*	*t*	Sig.
Intercept	57.595965	.132905	375.699	433.362	.000
ses	3.873861	.136624	3914.638	28.354	.000

[a] Dependent Variable: math.

When we compare the intercepts between this model and the single-level linear regression analysis at the beginning of the chapter, we find they are very similar (57.598 for the OLS model vs. 57.596 for the multilevel model). The difference in standard errors between these two estimates is larger, as we would expect given the clustering and associated intraclass correlation. More specifically, the estimated standard error (*SE*) from the OLS model was 0.098, and the standard error from the Level 1 multilevel model is 0.133, which is 36% larger. In that original single-level OLS analysis, the intercept described the average student achievement in the sample, without regard for students' school settings.

Second, in this case, we are more interested in the slope for SES ($\beta = 3.874$) and the standard error (0.137). When we compare these against the single-level model, we find they are considerably different (i.e., for the single-level analysis the slope was 4.255 and the standard error was 0.126). When we test hypotheses about model estimates (e.g., whether an unstandardized regression coefficient is significantly related to the dependent variable), the hypothesis test (often a *t*-test) is based on the ratio of the unstandardized estimate to its standard error (e.g., for SES this in the single-level model this is 4.255/.126 = 33.858). If the *t*-ratio is significantly large, given the sample under consideration, the parameter is considered statistically significant. In the multilevel analysis, the ratio of the unstandardized estimate of SES to its standard error is smaller (*t*-ratio = 28.344, but still is significant as we might expect).

These different results in our example illustrate the general point that parameter estimates and standard errors can be different in single-level versus multilevel analyses. As this simple multilevel analysis suggests, standard errors are often underestimated in single-level analyses, which can lead to a greater number of significant *t*-values and, hence, support for a proposed model than would be observed if a proposed model were tested with multilevel techniques. Directly as a result, in a multilevel analysis, adjusting for clustering generally results in a reduction of Type I errors (false rejection of the null hypothesis).

The output also provides information about the model's random parameters. The output suggests that the addition of the within-group predictor, SES, reduces the residual (within-group) variability (i.e., from 66.551 in the null model to 62.807 in the Level 1 model). This reduction in variance between the one-way ANOVA (or null) model and the current model can be used to calculate a reduction in variance estimate (or R^2) estimate for the within-school and between-school portions of the model. For each level, it is calculated as follows:

$$\left(\sigma^2_{M1} - \sigma^2_{M2}\right) \Big/ \sigma^2_{M1}, \tag{3.11}$$

where *M*1 refers to the one-way ANOVA (no predictors) Level 1 or Level 2 variance components and *M*2 refers to the current model's variance components. For the within-groups portion, this is calculated as 0.056 (66.551 – 62.807 = 3.744/66.551 = 0.056). This suggests that student SES background accounts for about 6% of the within-school variability in student scores. Notice also that the within-school predictor also affects that residual variability in intercepts at the school level. In particular, the initial variance component for schools, from the null model, was 10.642. After SES is added, however, the between-school variance in math achievement shrinks to 3.469 (see Table 3.10). For the reduction in variance between schools, this would be calculated as 0.674 [(10.642 – 3.469)/10.642]. This suggests that within-group SES accounts for almost

TABLE 3.10 Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald *Z*	Sig.
Residual		62.807187	1.108877	56.640	.000
Intercept [subject = schcode]	Variance	3.469256	.538821	6.439	.000

[a] Dependent Variable: math.

TABLE 3.11 Random Effect Covariance Structure (G)[a]

	Intercept | nschcode
Intercept | schcode	3.469256

Variance Components
[a] Dependent Variable: math.

two-thirds (67.4%) of the between-groups variability in achievement. In other words, a full two-thirds of the variation in means across schools can be attributed to differences in the socio-economic status of students in those schools. Another way of looking at this is that the initial variability in math achievement observed between schools (i.e., the ICC) is reduced considerably after controlling for student SES. From Equation 3.2, we find that ICC is now a little over 5% [3.469/(3.469 + 62.807) = 0.052].

We note, however, this reduction-in-variance type of R^2 statistic should be interpreted cautiously, because it is sometimes possible to obtain negative values for R^2. This is because the variance components may be less accurately estimated when there are no predictors in the model. For example, when an individual variable is sampled through a multilevel sampling process, it may show some between-group variability even if there is no between-group variability present in the population (Hox, 2002). In these cases, Hox (2002) notes that the reduction-in-variance procedures described previously may work as described in our example (i.e., where the within-group variable reduces between-group variability). In other situations, however, there may be variables that have almost no variation at one of the levels in the model. For example, this might occur if we had exactly the same percentage of males and females (50%) in each school. If there were no school-level variation in average gender composition, this would be less variance than expected by simple random sampling and could produce negative explained variance (see Hox, 2002, or Snijders & Bosker, 1999, for further discussion). Hox notes that another problem that can occur is that if random slopes are present, the estimated residual variances are related to the scale of measurement chosen for the explanatory variables (i.e., how they are centered, whether they might be standardized). When the interest is in the size of the variance components produced, it is therefore desirable to center the explanatory variables with random slopes on their grand means, since this will provide estimates for an "average" sampling unit (Hox, 2002). We discuss centering decisions further in subsequent chapters.

The covariance parameters table (Table 3.10) also suggests that after the introduction of SES into the model, there is still significant variability to be explained both within schools (Wald Z = 56.640, $p < .001$) and between schools (Wald Z = 6.439, $p < .001$). The Wald Z test suggests that, even after controlling for student SES within schools, a statistically significant amount of variation in outcomes still remains both within and between schools. This suggests that we could add other predictors (e.g., gender, ethnicity, motivation) within schools and between schools (e.g., student composition, school process indicators) that might explain this residual variability in intercepts.

The final piece of output (Table 3.11) provides the random effect covariance structure for Level 2. In this case there is only one random effect. Because we specified the SES–achievement slope to be fixed within schools, there is no variance component describing variability in slopes.

Step 3: Building the Group-Level (or Level 2) Random Intercept Model

Next, we will add school-level variables to explain the variability in intercepts across schools. In this case, our thesis is that school context variables (i.e., aggregated student SES composition, type of school [public coded 1, other = 0]) and the focus of the school's academic program (i.e., the aggregate percentage of students who intend to study at 4-year universities after high school) will impact the remaining variability in achievement between schools. It is important