Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series

Multilevel Modeling Using R

W. Holmes Finch
Jocelyn E. Bolin
Ken Kelley

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Multilevel Modeling Using R

Chapman & Hall/CRC

Statistics in the Social and Behavioral Sciences Series

Aims and scope

Large and complex datasets are becoming prevalent in the social and behavioral sciences and statistical methods are crucial for the analysis and interpretation of such data. This series aims to capture new developments in statistical methodology with particular relevance to applications in the social and behavioral sciences. It seeks to promote appropriate use of statistical, econometric and psychometric methods in these applied sciences by publishing a broad range of reference works, textbooks and handbooks.

The scope of the series is wide, including applications of statistical methodology in sociology, psychology, economics, education, marketing research, political science, criminology, public policy, demography, survey methodology and official statistics. The titles included in the series are designed to appeal to applied statisticians, as well as students, researchers and practitioners from the above disciplines. The inclusion of real examples and case studies is therefore essential.

Published Titles

Analyzing Spatial Models of Choice and Judgment with R
David A. Armstrong II, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal

Analysis of Multivariate Social Science Data, Second Edition
David J. Bartholomew, Fiona Steele, Irini Moustaki, and Jane I. Galbraith

Latent Markov Models for Longitudinal Data
Francesco Bartolucci, Alessio Farcomeni, and Fulvia Pennoni

Statistical Test Theory for the Behavioral Sciences
Dato N. M. de Gruijter and Leo J. Th. van der Kamp

Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences
Brian S. Everitt

Multilevel Modeling Using R
W. Holmes Finch, Jocelyn E. Bolin, and Ken Kelley

Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition
Jeff Gill

Multiple Correspondence Analysis and Related Methods
Michael Greenacre and Jorg Blasius

Applied Survey Data Analysis
Steven G. Heeringa, Brady T. West, and Patricia A. Berglund

Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists
Herbert Hoijtink

Foundations of Factor Analysis, Second Edition
Stanley A. Mulaik

Linear Causal Modeling with Structural Equations
Stanley A. Mulaik

Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis
Leslie Rutkowski, Matthias von Davier, and David Rutkowski

Generalized Linear Models for Categorical and Continuous Limited Dependent Variables
Michael Smithson and Edgar C. Merkle

Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys
Guo-Liang Tian and Man-Lai Tang

Computerized Multistage Testing: Theory and Applications
Duanli Yan, Alina A. von Davier, and Charles Lewis

Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series

Multilevel Modeling Using R

W. Holmes Finch
Ball State University
Muncie, Indiana, USA

Jocelyn E. Bolin
Ball State University
Muncie, Indiana, USA

Ken Kelley
University of Notre Dame
Notre Dame, Indiana, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140312

International Standard Book Number-13: 978-1-4665-1586-4 (eBook - PDF)

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

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

The goal of this book is to provide you, the reader, with a comprehensive resource for the conduct of multilevel modeling using the R software package. Multilevel modeling, sometimes referred to as hierarchical modeling, is a powerful tool that allows a researcher to account for data collected at multiple levels. For example, an educational researcher may gather test scores and measures of socioeconomic status (SES) for students who attend a number of different schools. The students would be considered level-1 sampling units, and the schools would be referred to as level-2 units.

Ignoring the structure inherent in this type of data collection can, as we discuss in Chapter 2, lead to incorrect parameter and standard error estimates. In addition to modeling the data structure correctly, we will see in the following chapters that the use of multilevel models can also provide insights into the nature of relationships in our data that might otherwise not be detected.

After reviewing standard linear models in Chapter 1, we will turn our attention to the basics of multilevel models in Chapter 2, before learning how to fit these models using the R software package in Chapters 3 and 4. Chapter 5 focuses on the use of multilevel modeling in the case of longitudinal data, and Chapter 6 demonstrates the very useful graphical options available in R, particularly those most appropriate for multilevel data. Chapters 7 and 8 describe models for categorical dependent variables, first for single-level data, and then in the multilevel context. Finally, we conclude in Chapter 9 with Bayesian fitting of multilevel models.

We hope that you find this book to be helpful as you work with multilevel data. Our goal is to provide you with a guidebook that will serve as the launching point for your own investigations in multilevel modeling. The R code and discussion of its interpretation contained in this text should provide you with the tools necessary to gain insights into your own research, in whatever field it may be. We appreciate your taking the time to read our work and hope that you find it as enjoyable and informative to read as it was for us to write.

About the Authors

W. Holmes Finch is a professor in the Department of Educational Psychology at Ball State University where he has been since 2003. He earned a PhD from the University of South Carolina in 2002. Dr. Finch teaches courses in factor analysis, structural equation modeling, categorical data analysis, regression, multivariate statistics, and measurement to graduate students in psychology and education. His research interests are in the areas of multilevel models, latent variable modeling, methods of prediction and classification, and nonparametric multivariate statistics. Holmes is also an Accredited Professional Statistician (PStat®).

Jocelyn E. Bolin earned a PhD in educational psychology from Indiana University Bloomington in 2009. Her dissertation consisted of a comparison of statistical classification analyses under situations of training data misclassification. She is an assistant professor in the Department of Educational Psychology at Ball State University, where she has been since 2010. Dr. Bolin teaches courses on introductory and intermediate statistics, multiple regression analysis, and multilevel modeling for graduate students in social science disciplines. Her research interests include statistical methods for classification and clustering and use of multilevel modeling in the social sciences. She is a member of the American Psychological Association, the American Educational Research Association, and the American Statistical Association and is also an Accredited Professional Statistician (PStat®).

Ken Kelley is the Viola D. Hank Associate Professor of Management in the Mendoza College of Business at the University of Notre Dame. Dr. Kelley's research involves the development, improvement, and evaluation of quantitative methods, especially as they relate to statistical and measurement issues in applied research. Dr. Kelley's most notable contributions have been on research design, especially with regard to sample size planning. Dr. Kelley is the developer of the MBESS package for the R statistical language and environment. He is also an Accredited Professional Statistician (PStat®) and associate editor of *Psychological Methods*.

1

Linear Models

Statistical models provide powerful tools to researchers in a wide array of disciplines. Such models allow for the examination of relationships among multiple variables, which in turn can lead to a better understanding of the world. For example, sociologists use linear regression to gain insights into how factors such as ethnicity, gender, and level of education are related to an individual's income. Biologists can use the same type of model to understand the interplay between sunlight, rainfall, industrial runoff, and biodiversity in a rain forest. And using linear regression, educational researchers can develop powerful tools for understanding the role that different instructional strategies have on student achievement. In addition to providing a path by which various phenomena can be better understood, statistical models can also be used as predictive tools. For example, econometricians might develop models to predict labor market participation given a set of economic inputs. Higher education administrators may use similar types of models to predict grade point averages for prospective incoming freshmen to identify those who might need academic assistance during their first year of college.

As can be seen from these few examples, statistical modeling is very important across a wide range of fields, providing researchers with tools for both explanation and prediction. Certainly, the most popular of such models over the last 100 years of statistical practice has been the general linear model (GLM). The GLM links a dependent or outcome variable to one or more independent variables and can take the form of such popular tools as analysis of variance (ANOVA) and regression.

Based on GLM's popularity and utility and its ability to serve as the foundation for many other models including the multilevel types featured in this book, we will start with a brief review of the linear model, focusing on regression. This review starts with a short technical discussion of linear regression models, followed by a description of how they can be estimated using the R language and environment (R Core Team, 2013).

The technical aspects of this discussion are intentionally not highly detailed as we focus on the model from a conceptual perspective. However, sufficient detail is presented so that a reader having only limited familiarity with the linear regression model will be provided with a basis for moving forward to multilevel models so that specific features of these more complex models that are shared with linear models can be explicated.

Readers familiar with linear regression and using R to conduct such analyses may elect to skip this chapter with no loss of understanding of future chapters.

1.1 Simple Linear Regression

As noted above, the GLM framework serves as the basis for the multilevel models that we describe in subsequent chapters. Thus, in order to provide a foundation for the rest of the book, we will focus in this chapter on the linear regression model, although its form and function can easily be translated to ANOVA as well. The simple linear regression model in population form is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1.1}$$

where y_i is the dependent variable for individual i in the data set and x_i is the independent variable for subject i ($i = 1, \ldots, N$). The terms β_0 and β_1, are the intercept and slope of the model, respectively. In a graphical sense, the intercept is the point at which the line in Equation (1.1) crosses the y axis at $x = 0$. It is also the mean, specifically the conditional mean, of y for individuals with values of 0 on x. This latter definition will be most useful in actual practice. The slope β_1 expresses the relationship between y and x. Positive slope values indicate that larger values of x are associated with correspondingly larger values of y, while negative slopes mean that larger x values are associated with smaller y values. Holding everything else constant, larger values of β_1 (positive or negative) indicate a stronger linear relationship between y and x. Finally, ε_i represents the random error inherent in any statistical model, including regression. It expresses the fact that for any individual, i, the model will not generally provide a perfect predicted value of y_i, denoted $\hat{y}_i$ and obtained by applying the regression model as

$$\hat{y}_i = \beta_0 + \beta_1 x_i \tag{1.2}$$

Conceptually, this random error is representative of all factors that may influence the dependent variable other than x.

1.1.1 Estimating Regression Models with Ordinary Least Squares

In virtually all real-world contexts, the population is unavailable to the researcher. Therefore, β_0 and β_1 must be estimated using sample data taken from the population. The statistical literature describes several methods for obtaining estimated values of the regression model parameters (b_0 and b_1, respectively) given a set of x and y. By far, the most popular and widely used

of these methods is ordinary least squares (OLS). The vast majority of other approaches are useful in special cases involving small samples or data that fail to conform to the distributional assumptions undergirding OLS.

The goal of OLS is to minimize the sum of the squared differences between the observed values of y and the model predicted values of y across the sample. This difference, known as the residual, is written as

$$e_i = y_i - \hat{y}_i \tag{1.3}$$

Therefore, the method of OLS seeks to minimize

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{1.4}$$

The actual mechanism for finding the linear equation that minimizes the sum of squared residuals involves the partial derivatives of the sum of squared function with respect to the model coefficients β_0 and β_1. We will leave these mathematical details to excellent references such as Fox (2008). Note that in the context of simple linear regression, the OLS criteria reduce to the following equations that can be used to obtain b_0 and b_1 as

$$b_1 = r\left(\frac{s_y}{s_x}\right) \tag{1.5}$$

and

$$b_0 = \overline{y} - b_1\overline{x} \tag{1.6}$$

where, r is the Pearson product moment correlation coefficient between x and y, s_y is the sample standard deviation of y, s_x is the sample standard deviation of x, $\overline{y}$ is the sample mean of y, and $\overline{x}$ is the sample mean of x.

1.2 Distributional Assumptions Underlying Regression

The linear regression model rests upon several assumptions about the distribution of the residuals in the broader population. Although a researcher typically is never able to collect data from an entire population, it is possible to assess empirically whether the assumptions are likely to hold true based on sample data.

The first assumption that must hold true for linear models to function optimally is that the relationship between y_i and x_i is linear. If the relationship

is not linear, then clearly an equation for a line will not provide adequate fit and the model is thus misspecified. A second assumption is that the variance in the residuals is constant regardless of the value of x_i. This assumption is typically referred to as homoscedasticity and is a generalization of the homogeneity of error variance assumption in ANOVA. Homoscedasticity implies that the variance of y_i is constant across values of x_i. The distribution of the dependent variables around the regression line is literally the distribution of the residuals, thus making clear the connection of homoscedasticity of errors with the distribution of y_i around the regression line. The third assumption is that the residuals are normally distributed in a population. Fourth is the assumption that the independent variable x is measured without error and that it is unrelated to the model error term ε. It should be noted that the assumption of x measured without error is not as strenuous as one might first assume. In fact, for most real-world problems, the model will work well even when the independent variable is not error free (Fox, 2008). Fifth and finally, the residuals for any two individuals in a population are assumed to be independent of one another. This independence assumption implies that the unmeasured factors influencing y are not related from one individual to another and addressed directly with the use of multilevel models, as we will see in Chapter 2.

In many research situations, individuals are sampled in clusters, such that we cannot assume that individuals from the same cluster will have uncorrelated residuals. For example, if samples are obtained from multiple neighborhoods, individuals within the same neighborhoods may tend to be more like one another than they are like individuals from other neighborhoods. A prototypical example of this is children in schools. Due to a variety of factors, children attending the same school often have more in common with one another than they do with children from other schools. These common factors may include neighborhood socioeconomic status, school administration policies, and school learning environment, to name just a few.

Ignoring this clustering or not even realizing it is a problem can be detrimental to the results of statistical modeling. We explore this issue in great detail later in the book, but for now we simply want to mention that a failure to satisfy the assumption of independent errors is (1) a major problem and (2) often a problem that may be overcome with appropriate models, such as multilevel models that explicitly consider the nesting of data.

1.3 Coefficient of Determination

When a linear regression model has been estimated, researchers generally want to measure the relative magnitude of the relationships of the variables. One useful tool for ascertaining the strength of the relationship between

x and y is the coefficient of determination, which is the squared multiple correlation coefficient denoted R^2 in Equation (1.7). R^2 reflects the proportion of variation in the dependent variable that is explained by the independent variable. Mathematically, R^2 is calculated as

$$R^2 = \frac{SS_R}{SS_T} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{SS_E}{SS_T} \tag{1.7}$$

The terms in Equation (1.7) are as defined previously. The value of this statistic always lies between 0 and 1, with larger numbers indicating a stronger linear relationship between x and y, implying that the independent variable accounts for more variance in the dependent. R^2 is a very commonly used measure of the overall fit of a regression model. Along with the parameter inference discussed below, it serves as the primary mechanism by which the relationship between the two variables is quantified.

1.4 Inference for Regression Parameters

A second method for understanding the nature of the relationship between x and y involves making inferences about the relationship in the population given the sample regression equation. Because b_0 and b_1 are sample estimates of the population parameters β_0 and β_1, respectively, they are subject to sampling error as is any sample estimate. This means that although the estimates are unbiased if the aforementioned assumptions hold, they are not precisely equal to the population parameter values. Furthermore, were we to draw multiple samples from the population and estimate the intercept and slope for each, the values of b_0 and b_1 would differ across samples even though they would estimate the same population parameter values for β_0 and β_1. The magnitude of this variation in parameter estimates across samples can be estimated from our single sample using a statistic known as the standard error.

The standard error of the slope, denoted as σ_{b1} in a population, can be thought of as the standard deviation of slope values obtained from all possible samples of size n taken from the population. Similarly, the standard error of the intercept σ_{b0} is the standard deviation of the intercept values obtained from all such samples. Clearly, it is not possible to obtain census data from a population in an applied research context. Therefore, we must estimate the standard errors of both the slope (s_{b1}) and intercept (s_{b0}) using

data from a single sample, much as we did with b_0 and b_1. To obtain s_{b1}, we must first calculate the variance of the residuals,

$$S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} \tag{1.8}$$

where e_i is the residual value for individual i, N is the sample size, and p is the number of independent variables (one in the case of simple regression). Then

$$S_{b_1} = \frac{1}{\sqrt{1-R^2}}\left[\frac{S_e}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right] \tag{1.9}$$

The standard error of the intercept is calculated as

$$S_{b_0} = S_{b_1}\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \tag{1.10}$$

Because the sample intercept and slope are only estimates of the population parameters, researchers often are interested in testing hypotheses to infer whether the data represent a departure from what would be expected in what is commonly referred to as the null case (the null value holding true in the population can be rejected). Usually (but not always), the inference of interest concerns testing that the population parameter is 0. In particular, a non-0 slope in a population means that x is linearly related to y. Therefore, researchers typically are interested in using the sample to make inference about whether the population slope is 0 or not. Inference can also be made regarding the intercept, and again the typical focus is on whether the value is 0 in the population.

Inference about regression parameters can be made using confidence intervals and hypothesis tests. Much as with the confidence interval of the mean, the confidence interval of the regression coefficient yields a range of values within which we have some level of confidence (e.g., 95%) that the population parameter value resides. If our particular interest is in whether x is linearly related to y, then we would simply determine whether 0 is in the interval for β_1. If so, then we could not conclude that the population value differs from 0.

The absence of a statistically significant result (i.e., an interval not containing 0) does not imply that the null hypothesis is true. Rather it means that the sample data contains insufficient evidence to reject the null. Similarly, we can construct a confidence interval for the intercept, and if 0 is within the interval, we would conclude that the value of y for an individual with $x = 0$ could plausibly be but is not necessarily 0. The confidence intervals for the slope and intercept take the following forms:

$$b_1 \pm t_{cv} s_{b_1} \tag{1.11}$$

and

$$b_0 \pm t_{cv} s_{b_0} \tag{1.12}$$

Here the parameter estimates and their standard errors are as described previously, while t_{cv} is the critical value of the t distribution for $1 - \alpha/2$ (e.g., the 0.975 quantile if $\alpha = 0.05$) with $n - p - 1$ degrees of freedom. The value of α is equal to 1 minus the desired level of confidence. Thus, for a 95% confidence interval (0.95 level of confidence), α would be 0.05.

In addition to confidence intervals, inference about the regression parameters can also be made using hypothesis tests. In general, the forms of this test for the slope and intercept, respectively, are

$$t_{b_1} = \frac{b_1 - \beta_1}{s_{b_1}} \tag{1.13}$$

$$t_{b_0} = \frac{b_0 - \beta_0}{s_{b_0}} \tag{1.14}$$

The terms β_1 and β_0 are the parameter values under the null hypothesis. Again, most often the null hypothesis posits that there is no linear relationship between x and y ($\beta_1 = 0$) and that the value of $y = 0$ when $x = 0$ ($\beta_0 = 0$). For simple regression, each of these tests is conducted with $n - 2$ degrees of freedom.

1.5 Multiple Regression

The linear regression model can be extended very easily to accommodate multiple independent variables at once. In the case of two regressors, the model takes the form

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{1.15}$$

In many ways, this model is interpreted like the one for simple linear regression. The only major difference between simple and multiple regression interpretation is that each coefficient is interpreted in turn *holding constant* the value of the other regression coefficient. In particular, the parameters are estimated by b_0, b_1, and b_2, and inferences about these parameters are made in the same fashion for both confidence intervals and hypothesis tests.

The assumptions underlying this model are also the same as those described for the simple regression model. Despite these similarities, three additional topics regarding multiple regression need to be considered here. These are inference for the set of model slopes as a whole, an adjusted measure of the coefficient of determination, and collinearity among the independent variables. Because these issues will be important in the context of multilevel modeling as well, we will address them in detail.

With respect to model inference, for simple linear regression, the most important parameter is generally the slope, so that inference for it will be of primary concern. When a model has multiple x variables, the researcher may want to know whether the independent variables taken as a whole are related to y. Therefore, some overall test of model significance is desirable. The null hypothesis for this test is that all of the slopes are equal to 0 in the population; i.e., none of the regressors is linearly related to the dependent variable. The test statistic for this hypothesis is calculated as

$$F = \frac{SS_R/p}{SS_E/(n-p-1)} = \left(\frac{n-p-1}{p}\right)\left(\frac{R^2}{1-R^2}\right) \tag{1.16}$$

Here, terms are as defined in Equation (1.7). This test statistic is distributed as an F with p and $n - p - 1$ degrees of freedom. A statistically significant result would indicate that one or more of the regression coefficients are not equal to 0 in the population. Typically, the researcher would then refer to the tests of individual regression parameters described above in order to identify which parameters were not equal to 0.

A second issue to be considered by researchers in the context of multiple regression is the notion of adjusted R^2. Stated simply, the inclusion of additional independent variables in the regression model will always yield higher values of R^2, even when these variables are not statistically significantly related to the dependent variable. In other words, there is a capitalization on chance that occurs in the calculation of R^2.

As a consequence, models including many regressors with negligible relationships with y may produce an R^2 that would suggest the model explains a great deal of variance in y. An option for measuring the variance explained in the dependent variable that accounts for this additional model complexity would be helpful to a researcher seeking to understand the true nature of the relationship between the set of independent

variables and the dependent. Such a measure exists in the form of the adjusted R^2 value, which is commonly calculated as

$$R_A^2 = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-p-1}\right) \tag{1.17}$$

R_A^2 only increases with the addition of an x if that x explains more variance than would be expected by chance. R_A^2 will always be less than or equal to the standard R^2. It is generally recommended to use this statistic in practice when models containing many independent variables are used.

A final important issue specific to multiple regression is collinearity, which occurs when one independent variable is a linear combination of one or more of the other independent variables. In such a case, regression coefficients and their corresponding standard errors can be quite unstable, resulting in poor inference. It is possible to investigate the presence of collinearity using a statistic known as the variance inflation factor (VIF). To calculate the VIF for x_j, we would first regress all the other independent variables onto x_j and obtain an R_{xi}^2 value. We then calculate

$$VIF = \frac{1}{1 - R_x^2} \tag{1.18}$$

The VIF will become large when R_{xj}^2 is near 1, indicating that x_j has very little unique variation when the other independent variables in the model are considered. That is, if the other $p - 1$ regressors can explain a high proportion of x_j, then x_j does not add much to the model above and beyond the other $p - 1$ regression. Collinearity in turn leads to high sampling variation in b_j, resulting in large standard errors and unstable parameter estimates. Conventional rules of thumb have been proposed for determining when an independent variable is highly collinear with the set of other $p - 1$ regressors. Thus, the researcher may consider collinearity a problem if VIF > 5 or 10 (Fox, 2008). The typical response to collinearity is to remove the offending variable(s) or use an alternative approach to conducting the regression analysis such as ridge regression or regression following a principal components analysis.

1.6 Example of Simple Manual Linear Regression

To demonstrate the principles of linear regression discussed above, let us consider a simple scenario in which a researcher collected data on college grade point averages (GPAs) and test anxiety using a standard measure by

TABLE 1.1

Descriptive Statistics and Correlation of GPA and Test Anxiety

Variable	Mean	Standard Deviation	Correlation
GPA	3.12	0.51	–0.30
Anxiety	35.14	10.83	

which higher scores indicate greater anxiety when taking a test. The sample consisted of 440 college students who were measured on both variables. The researcher is interested in the extent to which test anxiety is related to college GPA, so that GPA is the dependent variable and anxiety is the independent variable. The descriptive statistics for each variable and the correlations between them appear in Table 1.1.

We can use this information to obtain estimates for both the slope and intercept of the regression model using Equations (1.4) and (1.5). First, the slope is calculated as

$$b_1 = -0.30\left(\frac{0.51}{10.83}\right) = -0.014$$

indicating that individuals with higher test anxiety scores will generally have lower GPAs. Next, we can use this value and information in the table to calculate the intercept estimate:

$$b_0 = 3.12 - (-0.014)\,(35.14) = 3.63$$

The resulting estimated regression equation is then

$$\hat{GPA} = 3.63 - 0.014\,(anxiety)$$

Thus, this model would predict that for a one-point increase in the anxiety assessment score, the GPA would decrease by −0.014 points.

To better understand the strength of the relationship between test anxiety and GPA, we will want to calculate the coefficient of determination. To do this, we need both the SS_R and SS_T, which take the values 10.65 and 115.36, yielding

$$R^2 = \frac{10.65}{115.36} = 0.09$$

This result suggests that approximately 9% of the variation in GPA is explained by variation in test anxiety scores. Using this R^2 value and Equation (1.14),

we can calculate the *F* statistic t-test for whether any of the model slopes (in this case only one) are different from 0 in the population:

$$F = \left(\frac{440-1-1}{1}\right)\left(\frac{0.09}{1-0.09}\right) = 438(0.10) = 43.8$$

This test has p and $n - p - 1$ degrees of freedom, or 1 and 438 in this situation. The p value of this test is less than 0.001, leading us to conclude that the slope in the population is indeed significantly different from 0 because the p value is less than the Type I error rate specified. Thus, test anxiety is linearly related to GPA. The same inference could be conducted using the t-test for the slope. First we must calculate the standard error of the slope estimate:

$$S_{b_1} = \frac{1}{\sqrt{1-R^2}}\left(\frac{S_E}{\sqrt{\sum(x_i - \bar{x})^2}}\right)$$

For these data,

$$S_E = \sqrt{\frac{104.71}{440-1-1}} = \sqrt{0.24} = 0.49$$

In turn, the sum of squared deviations for x (anxiety) was 53743.64, and we previously calculated $R^2 = 0.09$. Thus, the standard error for the slope is

$$S_{b1} = \frac{1}{\sqrt{1-0.09}}\left(\frac{0.49}{\sqrt{53743.64}}\right) = 1.05(0.002) = 0.002$$

The test statistic for the null hypothesis that $\beta_1 = 0$ is calculated as

$$t = \frac{b_1 - 0}{s_{b1}} = \frac{-0.014}{0.002} = -7.00$$

with $n - p - 1$ or 438 degrees of freedom. The p value for this test statistic value is less than 0.001 and thus we can probabilistically infer that the value of the slope in the population is not zero, with the best sample point estimate being –0.014.

Finally, we can also draw inference about β_1 through a 95% confidence interval, as shown in Equation (1.9). For this calculation, we must determine the value of the t distribution with 438 degrees of freedom that correspond to the 1 – 0.05/2 or 0.975 point in the distribution. We can do so by using a t table in the back of a textbook or with standard computer software

such as SPSS. In either case, the critical value for this example is 1.97. The confidence interval can then be calculated as

$$(-0.014 - 1.97\ (0.002), -0.014 + 1.97\ (0.002))$$
$$(-0.014 - 0.004, -0.104 + 0.004)$$
$$(-0.018, -0.010)$$

The fact that 0 is not in the 95% confidence interval simply supports the conclusion we reached using the *p* value as described above. Also, given this interval, we can infer that the actual population slope value lies between –0.018 and –0.010. Thus, anxiety could plausibly have an effect as small as –0.010 or as large as –0.018.

1.7 Regression in R

In R, the function call for fitting linear regression is `lm`, which is part of the `stats` library that is loaded by default each time R is started. The basic form for a linear regression model using `lm` is:

```
lm(formula, data)
```

where `formula` defines the linear regression form and `data` indicates the data set used in the analysis, examples of which appear below. Returning to the previous example, predicting GPA from measures of physical (`BStotal`) and cognitive academic anxiety (`CTA.tot`), the model is defined in R as

```
Model1.1 <- lm(GPA ~ CTA.tot + BStotal, Cassidy)
```

This line of R code is referred to as a function call and defines the regression equation. The dependent variable `GPA` is followed by the independent variables `CTA.tot` and `BStotal`, separated by `~`. The data set `Cassidy` is also given here, after the regression equation has been defined. Finally, the output from this analysis is stored in the object `Model1.1`. To view this output, we can type the name of this object in R, and hit return to obtain the following:

```
Call:
lm(formula = GPA ~ CTA.tot + BStotal, data = Cassidy)

Coefficients:
(Intercept)    CTA.tot   BStotal
    3.61892   -0.02007   0.01347
```

The output obtained from the basic function call will return only values for the intercept and slope coefficients, lacking information regarding

model fit (e.g., R^2) and significance of model parameters. Further information on our model can be obtained by requesting a summary of the model.

```
summary(Model1.1)
```

Using this call, R will produce the following:

```
Call:
lm(formula = GPA ~ CTA.tot + BStotal, data = Cassidy)

Residuals:
     Min           1Q       Median           3Q          Max
-2.99239     -0.29138      0.01516      0.36849      0.93941

Coefficients:
             Estimate   Std. Error   t value    Pr(>|t|)
(Intercept)  3.618924     0.079305    45.633     < 2e-16 ***
CTA.tot     -0.020068     0.003065    -6.547    1.69e-10 ***
BStotal      0.013469     0.005077     2.653     0.00828 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4852 on 426 degrees of freedom
  (57 observations deleted due to missingness)
Multiple R-squared: 0.1066,       Adjusted R-squared: 0.1024
F-statistic: 25.43 on 2 and 426 DF, p-value: 3.706e-11
```

From the model summary we can obtain information on model fit (overall F test for significance, R^2, and standard error of the estimate), parameter significance tests, and a summary of residual statistics. As the F test for the overall model is somewhat abbreviated in this output, we can request the entire ANOVA result, including sums of squares and mean squares by using the `anova(Model1.1)` function call.

```
Analysis of Variance Table

Response: GPA
            Df   Sum Sq   Mean Sq  F value      Pr(>F)
CTA.tot      1   10.316   10.3159  43.8125  1.089e-10 ***
BStotal      1    1.657    1.6570   7.0376    0.00828 **
Residuals  426  100.304    0.2355
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Often in a regression model, we are interested in additional information that the model produces such as predicted values and residuals. Using the R call `attributes()`, we can obtain a list of the additional information available for the `lm` function.

```
attributes(Model1.1)
$names
 [1] "coefficients"  "residuals"  "effects"       "rank"        "fitted.values"
 [6] "assign"        "qr"         "df.residual"   "na.action"   "xlevels"
[11] "call"          "terms"      "model"


$class
[1] "lm"
```

This is a list of attributes or information that may be pulled from the fitted regression model. To obtain this information, we can call for the particular attribute. For example, if we want to obtain the predicted GPA for each individual in the sample, we would simply type the following followed by the enter key:

```
Model1.1$fitted.values

1        3        4        5        8        9       10       11       12
2.964641 3.125996 3.039668 3.125454 2.852730 3.152391 3.412460 3.011917 2.611103
      13       14       15       16       17       19       23       25       26
3.158448 3.298923 3.312121 2.959938 3.205183 2.945928 2.904979 3.226064 3.245318
      27       28       29       30       31       34       35       37       38
2.944573 3.171646 2.917635 3.198584 3.206267 3.073204 3.258787 3.118584 2.972594
      39       41       42       43       44       45       46       48       50
2.870630 3.144980 3.285454 3.386064 2.871713 2.911849 3.166131 3.051511 3.251917
```

Thus for example, the predicted GPA for subject 1 based on the prediction equation would be 2.96. By the same token, we can obtain the regression residuals with the following command:

```
Model1.1$residuals

 1            3            4            5            8            9
-0.4646405061 -0.3259956916 -0.7896675749 -0.0254537419 0.4492704297 -0.0283914353
          10           11           12           13           14           15
-0.1124596847  -0.5119169570  0.0888967457 -0.6584484215 -0.7989228998 -0.4221207716
          16           17           19           23           25           26
-0.5799383942 -0.3051829226 -0.1459275978 -0.8649791080 0.0989363702 -0.2453184879
          27           28           29           30           31           34
-0.4445727235  0.7783537067 -0.8176350301  0.1014160133 0.3937331779 -0.1232042042
          35           37           38           39           41           42
 0.3412126654  0.4814161689  0.9394056837 -0.6706295541 -0.5449795748 -0.4194540531
          43           44           45           46           48           50
-0.4960639410 -0.0717134535 -0.4118490187  0.4338687432 0.7484894275  0.4480825762
```

From this output, we can see that the predicted GPA for the first individual in the sample was approximately 0.465 points below the actual GPA.

1.7.1 Interaction Terms in Regression

More complicated regression relationships can also be easily modeled using the `lm()` function. Let us consider a moderation analysis involving the anxiety measures. In this example, an interaction between cognitive test anxiety and physical anxiety is modeled in addition to the main effects for the two variables. An interaction is simply computed as the product

of the interacting variables, so that the moderation model using `lm()` is defined as:

```
Model1.2 <- lm(GPA ~ CTA.tot + BStotal + CTA.tot*BStotal,
  Cassidy)

Model1.2

Call:
lm(formula = GPA ~ CTA.tot + BStotal + CTA.tot * BStotal, data
  = Cassidy)

Residuals:
     Min       1Q   Median       3Q      Max
-2.98711  -0.29737  0.01801  0.36340  0.95016

Coefficients:
                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)      3.8977792   0.2307491   16.892   < 2e-16 ***
CTA.tot         -0.0267935   0.0060581   -4.423  1.24e-05 ***
BStotal         -0.0057595   0.0157812   -0.365     0.715
CTA.tot:BStotal  0.0004328   0.0003364    1.287     0.199
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4849 on 425 degrees of freedom
  (57 observations deleted due to missingness)
Multiple R-squared: 0.1101,      Adjusted R-squared: 0.1038
F-statistic: 17.53 on 3 and 425 DF, p-value: 9.558e-11
```

Here the slope for the interaction is denoted `CTA.tot:BStotal`, takes the value 0.0004, and is nonsignificant (t = 1.287, p = 0.199), indicating that the level of physical anxiety symptoms (`BStotal`) does not change or moderate the relationship between cognitive test anxiety (`CTA.tot`) and GPA.

1.7.2 Categorical Independent Variables

The `lm` function is also easily capable of incorporating categorical variables into regression. Let us consider an analysis for predicting GPA from cognitive test anxiety (`CTA.tot`) and the categorical variable gender. To incorporate gender into the model, it must be dummy coded such that one category (e.g., male) takes the value of 1 and the other category (e.g., female) takes the value of 0. In this example, we named the variable `Male`, where 1 = male and 0 = not male (female). Defining a model using a dummy variable with the `lm` function then becomes no different from using continuous predictor variables.

```
Model1.3 <- lm(GPA~CTA.tot + Male, Acad)

summary(Model1.3)

Call:
lm(formula = GPA ~ CTA.tot + Male, data = Acad)

Residuals:
     Min       1Q   Median       3Q      Max
-3.01149 -0.29005  0.03038  0.35374  0.96294

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  3.740318    0.080940   46.211   < 2e-16 ***
CTA.tot     -0.015184    0.002117   -7.173  3.16e-12 ***
Male        -0.222594    0.047152   -4.721  3.17e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4775 on 437 degrees of freedom
  (46 observations deleted due to missingness)
Multiple R-squared: 0.1364,      Adjusted R-squared: 0.1324
F-statistic: 34.51 on 2 and 437 DF, p-value: 1.215e-14
```

In this example, the slope for the dummy variable `Male` is negative and significant ($\beta = -0.223$, $p < 0.001$), indicating that males have significantly lower mean GPAs than females.

Depending on the format in which the data are stored, the `lm` function is capable of dummy coding categorical variables. If a variable has been designated as categorical (as often happens if you read data in from an SPSS file in which the variable is designated as such) and is used in the `lm` function, it will automatically dummy code the variable in your results. For example, if instead of using the `Male` variable as described above, we used `Gender` as a categorical variable coded as female and male, we would obtain the following results from the model specification and summary commands.

```
Model1.4 <- lm(GPA~CTA.tot + Gender, Acad)

summary(Model1.4)

Call:
lm(formula = GPA ~ CTA.tot + Gender, data = Acad)

Residuals:
     Min       1Q   Median       3Q      Max
-3.01149 -0.29005  0.03038  0.35374  0.96294

Coefficients:
                Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)     3.740318    0.080940   46.211   < 2e-16 ***
CTA.tot        -0.015184    0.002117   -7.173  3.16e-12 ***
Gender[T.male] -0.222594    0.047152   -4.721  3.17e-06 ***
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4775 on 437 degrees of freedom
  (46 observations deleted due to missingness)
Multiple R-squared: 0.1364,       Adjusted R-squared: 0.1324
F-statistic: 34.51 on 2 and 437 DF, p-value: 1.215e-14
```

A comparison of results between `models Model1.3` and `Model1.4` reveals identical coefficient estimates, *p* values, and model fit statistics. The only difference between the two sets of results is that for `Model1.4` R reported the slope as `Gender[t.male]`, indicating that the variable was dummy coded automatically so that male is 1 and not male is 0.

In the same manner, categorical variables consisting of more than two categories can also be incorporated easily into a regression model, either through direct use of the categorical variable or dummy coding prior to analysis. In the following example, the variable `Ethnicity` includes three possible groups (African American, Caucasian, and Other). By including this variable in the model call, we are implicitly requesting that R automatically dummy code it for us.

```
GPAmodel1.5 <- lm(GPA~CTA.tot + Ethnicity, Acad)

summary(GPAmodel1.5)

Call:
lm(formula = GPA ~ CTA.tot + Ethnicity, data = Acad)

Residuals:
     Min       1Q   Median       3Q      Max
-2.95019  -0.30021  0.01845  0.37825  1.00682

Coefficients:
                                Estimate  Std. Error t value Pr(>|t|)
(Intercept)                     3.670308   0.079101  46.400  < 2e-16 ***
CTA.tot                        -0.015002   0.002147  -6.989 1.04e-11 ***
Ethnicity[T.African American] -0.482377   0.131589  -3.666 0.000277 ***
Ethnicity[T.Other]            -0.151748   0.136150  -1.115 0.265652
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4821 on 436 degrees of freedom
  (46 observations deleted due to missingness)
Multiple R-squared: 0.1215,       Adjusted R-squared: 0.1155
F-statistic: 20.11 on 3 and 436 DF, p-value: 3.182e-12
```

Since we have slopes for African American and Other, we know that Caucasian serves as the reference category, which is coded as 0. Results indicate

a significant positive slope for African American ($\beta = -0.482$, $p < 0.001$), and a nonsignificant slope for Other ($\beta = 0.152$, $p > 0.05$), indicating that African Americans have significantly lower GPAs than Caucasians but the GPA result for the Other ethnicity category was not significantly different from those for Caucasians.

Finally, let us consider some issues associated with allowing R to dummy code categorical variables automatically. First, R will always automatically dummy code the first category listed as the reference category. If a more theoretically suitable dummy coding scheme is desired, it will be necessary to order the categories so that the desired reference category is first or simply recode dummy variables manually.

Also, it is important to remember that automatic dummy coding occurs only when a variable is labeled in a system as categorical. This will occur automatically if the categories are coded as letters. However, if a categorical variable is coded 1, 2 or 1, 2, 3 but not specifically designated as categorical, the system will view it as continuous and treat it as such. To ensure that a variable is treated as categorical when that is what we desire, we simply use the `as.factor` command. For the `Male` variable in which males are coded as 1 and females as 0, we would type

```
Male<-as.factor(Male)
```

We would then be able to assume the `Male` variable is categorical. In addition, if the dummy variable has only two levels, as is the case with `Male`, then it need not be converted to a categorical factor because the results from the regression analysis will be identical either way.

1.7.3 Checking Regression Assumptions with R

When checking assumptions for linear regression models, it is often desirable to create a plot of the residuals. Diagnostic residual plots can be easily obtained by using the `residualPlots` function from the `car` R package that we would need to install in our R workspace as explained in the appendix at the end of this book that introduces working with R. Let us again return to `Model1.1` predicting GPA from cognitive test anxiety and physical anxiety symptoms. After the regression model is created (`Model1.1`), we can easily obtain diagnostic residual scatterplots using the following command:

```
Library(car)
residualPlots(Model1.1)
```

This command will produce scatterplots of the Pearson residuals against each predictor variable as well as against the fitted values. In addition,

the `residualPlots` command will provide lack-of-fit tests in which a t-test for the predictor squared is computed and a fit line added to the plot to help check for nonlinear patterns in the data. A Tukey's test for non-additivity is also computed for the plot of residuals against the fitted values to acquire further information about the adequacy of model fit along with a lack-of-fit test for each predictor. Tukey's statistic is obtained by adding the squares of the fitted values to the original regression model. It tests the null hypothesis that the model is additive and that no interactions exist among the independent variables (Tukey, 1949). A nonsignificant result, such as that found for this example, indicates that no interaction is required in the model.

The other tests included here are for the squared term of each independent variable. For example, given that the `Test stat` results for `CTA.tot` and `BStotal` are not significant, we can conclude that neither of these variables has a quadratic relationship with GPA. See Figure 1.1.

```
residualPlots(Model1.1)

            Test stat  Pr(>|t|)
CTA.tot         0.607     0.544
BStotal         0.762     0.447
Tukey test      0.301     0.764
```

The `residualPlots` command provides plots with the residuals on the y axes of the graphs, the values of each independent variable, respectively, on the x axes for the first two graphs, and the fitted values on x for the last graph. In addition, curves were fit linking the x and y axes for each graph.

The researcher would examine these graphs to assess two assumptions about the data. First, the assumption of homogeneity of variance can be checked through an examination of the residual by fitted plot. If the assumption holds, this plot should display a formless cloud of data points with no discernible shapes that are equally spaced across all values of x. In addition, the linearity of the relationships between each independent variable and the dependent variable is assessed by an examination of the plots involving them. For example, it is appropriate to assume linearity for `BStotal` if the residual plots show no discernible pattern. This may be further explained by an examination of the fitted line. If this line is essentially flat, as is the case here, we can conclude that any relationship between `BStotal` and `GPA` is only linear.

In addition to linearity and homogeneity of variance, it is also important to determine whether the residuals follow a normal distribution as assumed in regression analysis. To check the normality of residual assumptions, QQ plots (quantile–quantile plots) are typically used.

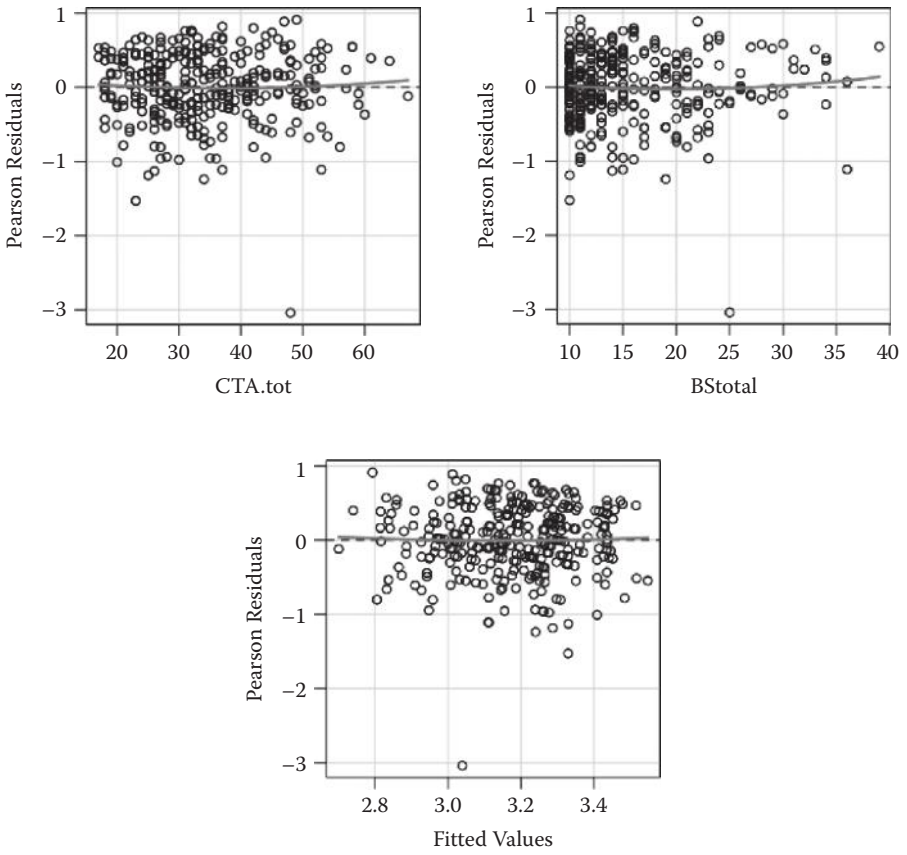
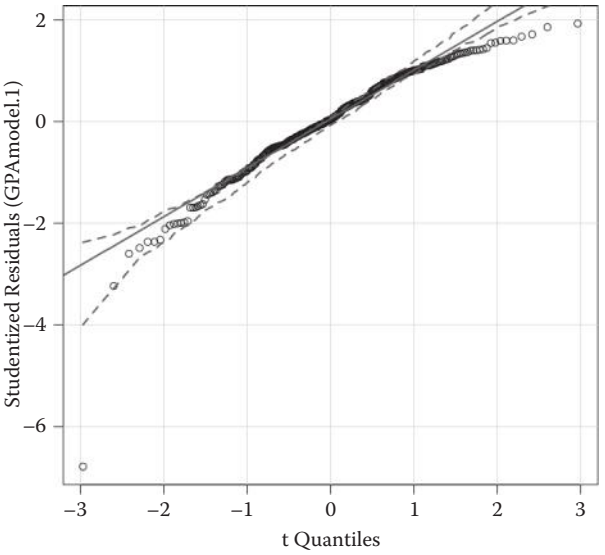


FIGURE 1.1
Diagnostic residuals plots for regression model predicting GPA from `CTA.tot` and `BStotal`.

The `qqPlot` function from the `car` package may be used to easily create QQ plots of run regression models. Interpretation of the QQ plot is quite simple. Essentially, the graph displays the data as it actually is on the x axis and as it would be if normally distributed on the y axis. The individual data points are represented in R by black circles. The solid line represents the data conforming perfectly to the normal distribution. Therefore, the closer the observed data (circles) are to the solid line, the more closely the data conforms to the normal distribution. In addition, R provides a 95% confidence interval for the line, so that when the data points fall within it they are deemed to conform to the normal distribution. In this example, the data appear to follow the normal distribution fairly closely.

```
qqPlot(Model1.1)
```



Summary

Chapter 1 introduced readers to the basics of linear modeling using R. This treatment was purposely limited, as a number of good texts cover linear modeling and it is not the main focus of this book. However, many of the core concepts presented here for the GLM apply to multilevel modeling as well, and thus are of key importance as we move into more complex analyses. In addition, much of the syntactical framework presented here will reappear in subsequent chapters. In particular, readers should leave this chapter comfortable with interpretation of coefficients in linear models and the concept of variance in outcome variables. We would encourage you to return to this chapter frequently as needed to reinforce these basic concepts. In addition, we would recommend that you also refer to the appendix dealing with the basics of using R when questions about data management and installation of specific R libraries arise. In Chapter 2, we will turn our attention to the conceptual underpinnings of multilevel modeling before delving into estimation in Chapters 3 and 4.

2

Introduction to Multilevel Data Structure

2.1 Nested Data and Cluster Sampling Designs

In Chapter 1, we considered the standard linear model that underlies such common statistical methods as regression and analysis of variance (ANOVA; the general linear model). As noted, this model rests on several primary assumptions about the nature of the data in a population. Of particular importance in the context of multilevel modeling is the assumption of independently distributed error terms for the individual observations within a sample. This assumption essentially means that there are no relationships among individuals in the sample for the dependent variable *once the independent variables in the analysis are accounted for*. In the example described in Chapter 1, this assumption was indeed met, as the individuals in the sample were selected randomly from the general population. Therefore, nothing linked their dependent variable values other than the independent variables included in the linear model. However, in many cases the method used for selecting the sample does create correlated responses among individuals. For example, a researcher interested in the impact of a new teaching method on student achievement may randomly select schools for placement in treatment or control groups. If school A is placed into the treatment condition, all students within the school will also be in the treatment condition. This is a cluster randomized design in that the clusters (and not the individuals) are assigned to a specific group. Furthermore, it would be reasonable to assume that the school itself, above and beyond the treatment condition, would have an impact on the performances of the students. This impact would manifest as correlations in achievement test scores among individuals attending the school. Thus, if we were to use a simple one-way ANOVA to compare the achievement test means for the treatment and control groups with such cluster sampled data, we would likely violate the assumption of independent errors because a factor beyond treatment condition (in this case the school) would exert an additional impact on the outcome variable.

We typically refer to the data structure described above as nested, meaning that individual data points at one level (e.g., student) appear in only one level

of a higher level variable such as school. Thus, students are nested within school. Such designs can be contrasted with crossed data structures whereby individuals at the first level appear in multiple levels of the second variable. In our example, students may be crossed with after-school activities if they are allowed to participate in more than one. For example, a student might be on the basketball team and a member of the band.

The focus of this book is almost exclusively on nested designs that give rise to multilevel data. Another example of a nested design is a survey of job satisfaction levels of employees from multiple departments within a large business organization. In this case, each employee works within only a single division in the company, making possible a nested design. It seems reasonable to assume that employees working in the same division will have correlated responses on the satisfaction survey, because much of their views of their jobs will be based exclusively upon experiences within their divisions. For a third such example, consider the situation in which clients of several psychotherapists working in a clinic are asked to rate the quality of each therapy session. In this instance, three levels of data exist: (1) time in the form of an individual session, (2) client, and (3) therapist. Thus, session is nested in client, which in turn is nested in therapist. This data structure would be expected to lead to correlated scores on a therapy rating instrument.

2.2 Intraclass Correlation

In cases where individuals are clustered or nested within a higher level unit (e.g., classroom, school, school district), it is possible to estimate the correlation among individuals' scores within the cluster or nested structure using the intraclass correlation (ICC, denoted ρ_I in the population). The ρ_I is a measure of the proportion of variation in the outcome variable that occurs between groups versus the total variation present. It ranges from 0 (no variance among clusters) to 1 (variance among clusters but no within-cluster variance). ρ_I can also be conceptualized as the correlation for the dependent measure for two individuals randomly selected from the same cluster. It can be expressed as

$$\rho_I = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{2.1}$$

where τ^2 denotes population variance between clusters and σ^2 indicates population variance within clusters. Higher values of ρ_I indicate that a greater share of the total variation in the outcome measure is associated with cluster membership; i.e., a relatively strong relationship among the

scores for two individuals from the same cluster. Another way to frame this issue is that individuals within the same cluster (e.g., school) are more alike on the measured variable than they are like individuals in other clusters.

It is possible to estimate τ^2 and σ^2 using sample data, and thus it is also possible to estimate ρ_I. Those familiar with ANOVA will recognize these estimates as related (though not identical) to the sum of squared terms. The sample estimate for variation within clusters is simply

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{C}(n_j - 1)S_j^2}{N - C} \tag{2.2}$$

where S_j^2 is the variance within cluster

$$j = \frac{\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)}{(n_j - 1)}$$

n_j is the sample size of cluster j, N is the total sample size, and C is the total number of clusters. In other words, σ^2 is simply the weighted average of within-cluster variances.

Estimation of τ^2 involves a few more steps, but is not much more complex than what we have seen for σ^2. To obtain the sample estimate for variation between clusters $\hat{\tau}^2$, we must first calculate the weighted between-cluster variance:

$$\hat{S}_B^2 = \frac{\sum_{j=1}^{C} n_j(\bar{y}_j - \bar{y})^2}{\tilde{n}(C-1)} \tag{2.3}$$

where $\bar{y}_j$ is the mean on response variables for cluster j and $\bar{y}$ is the overall mean on the response variable

$$\tilde{n} = \frac{1}{C-1}\left[N - \frac{\sum_{j=1}^{C} n_j^2}{N}\right]$$

We cannot use as S_B^2 a direct estimate of τ^2 because it is impacted by the random variation among subjects within the same clusters. Therefore, in

order to remove this random fluctuation we will estimate the population between-cluster variance as

$$\hat{\tau}^2 = S_B^2 - \frac{\hat{\sigma}^2}{\tilde{n}} \tag{2.4}$$

Using these variance estimates, we can in turn calculate the sample estimate of ρ_I:

$$\hat{\rho}_I = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2} \tag{2.5}$$

Note that Equation (2.5) assumes that the clusters are of equal size. Clearly, that will not always be the case, in which case this equation will not hold. However, the purpose for its inclusion here is to demonstrate the principle underlying the estimation of ρ_I, which holds even as the equation changes.

To illustrate estimation of ρ_I, let us consider the following data set. Achievement test data were collected from 10,903 third grade students nested within 160 schools. School enrollment sizes ranged from 11 to 143, with a mean size of 68.14. In this case, we will focus on the reading achievement test scores and use data from only five of the schools to make manual calculations easy to follow. First we will estimate $\hat{\sigma}^2$. To do so, we must estimate the variance in scores within each school. These values appear in Table 2.1. Using these variances and sample sizes, we can calculate $\hat{\sigma}^2$ as

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{j=1}^{c}(n_j - 1)S_j^2}{N - C} \\ &= \frac{(58-1)5.3 + (29-1)1.5 + (64-1)2.9 + (39-1)6.1 + (88-1)3.4}{278-5} \\ &= \frac{302.1 + 42 + 182.7 + 231.8 + 295.8}{273} = \frac{1054.4}{273} = 3.9\end{aligned}$$

TABLE 2.1

School Size, Mean, and Variance of Reading Achievement Test

School	*N*	Mean	Variance
767	58	3.952	5.298
785	29	3.331	1.524
789	64	4.363	2.957
815	39	4.500	6.088
981	88	4.236	3.362
Total	278	4.149	3.916

The school means that are required for calculating S_B^2, appear in Table 2.1 as well. First we must calculate $\tilde{n}$:

$$\tilde{n} = \frac{1}{C-1}\left(N - \frac{\sum_{j=1}^{C} n_j^2}{N}\right) = \frac{1}{5-1}\left(278 - \frac{58^2 + 29^2 + 64^2 + 39^2 + 88^2}{278}\right)$$

$$= \frac{1}{4}(278 - 63.2) = 53.7$$

Using this value, we can then calculate S_B^2 for the five schools in our small sample using Equation (2.3):

$$\frac{58(3.952 - 4.149)^2 + 29(3.331 - 4.149)^2 + 64(4.363 - 4.149)^2 + 39(4.500 - 4.149)^2 + 88(4.236 - 4.149)^2}{53.7(5-1)}$$

$$= \frac{2.251 + 19.405 + 2.931 + 4.805 + 0.666}{214.8} = \frac{30.057}{214.800} = 0.140$$

We can now estimate the population between-cluster variance τ^2 using Equation (2.4):

$$0.140 - \frac{3.9}{53.7} = 0.140 - 0.073 = 0.067$$

We have now calculated all the parts needed to estimate ρ_I for the population,

$$\hat{\rho}_I = \frac{0.067}{0.067 + 3.9} = 0.017$$

This result indicates very little correlation of test scores within the schools. We can also interpret this value as the proportion of variation in the test scores accounted for by the schools. Since $\hat{\rho}_I$ is a sample estimate, we know that it is subject to sampling variation, which can be estimated with a standard error as in Equation (2.6):

$$s_{\rho_I} = (1 - \rho_I)(1 + (n-1)\rho_I)\sqrt{\frac{2}{n(n-1)(N-1)}} \qquad (2.6)$$

The terms in Equation (2.6) are as defined previously, and the assumption is that all clusters are of equal size. As noted earlier, this latter condition is not a requirement, however, and an alternative formulation exists for cases in which it does not hold. However, Equation (2.6) provides sufficient insight for our purposes into the estimation of the standard error of the ICC.

The ICC is an important tool in multilevel modeling, in large part because it indicates the degree to which a multilevel data structure may impact the outcome variable of interest. Larger ICC values are indicative of a greater impact of clustering. Thus, as the ICC increases in value, we must be more cognizant of employing multilevel modeling strategies in data analysis. In the next section, we will discuss the problems associated with ignoring this multilevel structure, before we turn our attention to methods for dealing with it directly.

2.3 Pitfalls of Ignoring Multilevel Data Structure

When researchers apply standard statistical methods to multilevel data such as the regression model described in Chapter 1, the assumption of independent errors is violated. For example, if we have achievement test scores from a sample of students who attend several different schools, it would be reasonable to believe that those attending the same school will have scores that are more highly correlated with one another than they are with scores from students at other schools. This within-school correlation would be due, for example, to a community, a common set of teachers, a common teaching curriculum, a single set of administrative policies, and other factors. The within-school correlation will in turn result in an inappropriate estimate of the of the standard errors for the model parameters, which will lead to errors of statistical inference, such as p-values smaller than they should be and the resulting rejection of null effects above the stated Type I error rate for the parameters.

Recalling our discussion in Chapter 1, the test statistic for the null hypothesis of no relationship between the independent and dependent variable is simply the regression coefficient divided by the standard error. An underestimation of the standard error will cause an overestimation of the test statistic, and thus the statistical significance for the parameter in cases where it should not be, that is, Type I errors at a higher rate than specified. Indeed, the underestimation of the standard error will occur unless τ^2 is equal to 0.

In addition to the underestimation of the standard error, another problem with ignoring the multilevel structure of data is that we may miss important relationships involving each level in the data. Recall our example of two levels of sampling: students (level 1) are nested in schools (level 2). Specifically, by *not* including information about the school, for example,

we may well miss important variables at the school level that may help explain performance at student level. Therefore, beyond the known problem with misestimating standard errors, we also develop an incorrect model for understanding the outcome variable of interest. In the context of multilevel linear models (MLMs), inclusion of variables at each level is relatively simple, as are interactions among variables at different levels. This greater model complexity in turn may lead to greater understanding of the phenomenon under study.

2.4 Multilevel Linear Models

In the following section we will review some of the core ideas that underlie MLMs. Our goal is to familiarize readers with terms that will repeat throughout the book and explain them in a relatively nontechnical fashion. We will first focus on the difference between random and fixed effects, after which we will discuss the basics of parameter estimation, focusing on the two most commonly used methods, maximum likelihood and restricted maximum likelihood, and conclude with a review of assumptions underlying MLMs, and overview of how they are most frequently used, with examples. In this section, we will also address the issue of centering, and explain why it is an important concept in MLM. After reading the rest of this chapter, the reader will have sufficient technical background on MLMs to begin using the R software package for fitting MLMs of various types.

2.4.1 Random Intercept

As we transition from the one-level regression framework of Chapter 1 to the MLM context, let us first revisit the basic simple linear regression model of Equation (1.1)

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Here, the dependent variable y is expressed as a function of an independent variable x, multiplied by a slope coefficient β_1, an intercept β_0, and random variation from subject to subject ε. We defined the intercept as the conditional mean of y when the value of x is 0.

In the context of a single-level regression model such as this, one intercept is common to all individuals in the population of interest. However, when individuals are clustered together in some fashion (e.g., students in classrooms and schools, organizational units within a company), there will potentially be a separate intercept for each cluster, that is, different means may exist for the dependent variable for $x = 0$ across the different clusters.

We say *potentially* here because the single intercept model of Equation (1.1) will suffice if there is no cluster effect. In practice, assessing the existence of different means across clusters is an empirical question described below. It should also be noted that in this discussion we consider only the case where the intercept is cluster specific. It is also possible for β_1 to vary by group or even other coefficients from more complicated models.

Allowing for group-specific intercepts and slopes leads to the following notation commonly used for the level 1 (micro) model in multilevel modeling

$$y_{ij} = \beta_{0j} + \beta_{1j}x + \varepsilon_{ij} \tag{2.7}$$

where the *ij* subscript refers to the *i*th individual in the *j*th cluster. We will begin our discussion of MLM notation and structure with the most basic multilevel model: predicting the outcome from only an intercept that we will allow to vary randomly for each group.

$$y_{ij} = \beta_{0j} + \varepsilon_{ij} \tag{2.8}$$

Allowing the intercept to differ across clusters, as in Equation (2.8), leads to the random intercept that we express as

$$\beta_{0j} = \gamma_{00} + U_{0j} \tag{2.9}$$

In this framework, γ_{00} represents an average or general intercept value that holds across clusters, whereas U_{0j} is a group-specific effect on the intercept. We can think of γ_{00} as a fixed effect because it remains constant across all clusters, and U_{0j} is a random effect because it varies from cluster to cluster. Therefore, for a MLM we are interested not only in some general mean value for y when x is 0 for all individuals in the population (γ_{00}), but also the deviation between the overall mean and the cluster-specific effects for the intercept (U_{0j}).

If we go on to assume that the clusters constitute a random sample from the population of all such clusters, we can treat U_{0j} as a kind of residual effect on y_{ij}, very similar to how we think of ε. In that case, U_{0j} is assumed to be drawn randomly from a population with a mean of 0 (recall that U_{0j} is a deviation from the fixed effect) and a variance τ^2. Furthermore, we assume that τ^2 and σ^2, the variance of ε, are uncorrelated. We have already discussed τ^2 and its role in calculating $\hat{\rho}_I$. In addition, τ^2 can also be viewed as the impact of the cluster on the dependent variable, and therefore testing it for statistical significance is equivalent to testing the null hypothesis that cluster (e.g., school) has no impact on the dependent variable. If we substitute the two components of the random intercept into the regression model, we get

$$y = \gamma_{00} + U_{0j} + \beta_1 x + \varepsilon \tag{2.10}$$

Equation (2.10) is termed the full or composite model in which the multiple levels are combined into a unified equation. Often in MLM, we begin our analysis of a data set with this simple random intercept model known as the null model that takes the form

$$y_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \tag{2.11}$$

While the null model does not provide information about the impacts of specific independent variables on the dependent, it does yield important information regarding how variation in y is partitioned between variance among the individual σ^2 values and variance among the clusters τ^2. The total variance of y is simply the sum of σ^2 and τ^2. In addition, as we have already seen, these values can be used to estimate ρ_I. The null model, as will be seen in later sections, is also used as a baseline for model building and comparison.

2.4.2 Random Slopes

It is a simple matter to expand the random intercept model in Equation (2.9) to accommodate one or more independent predictor variables. As an example, if we add a single predictor (x_{ij}) at the individual level (Level 1) to the model, we obtain

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + \varepsilon_{ij} \tag{2.12}$$

This model can also be expressed in two separate levels:

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}x + \varepsilon_{ij} \tag{2.13}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + U_{0j} \tag{2.14}$$

$$\beta_{1j} = \gamma_{10} \tag{2.15}$$

The model now includes the predictor and the slope relating it to the dependent variable γ_{10}, which we acknowledge as being at Level 1 by the subscript 10. We interpret γ_{10} in the same way as β_1 in the linear regression model, i.e., as a measure of the impact on y of a one-unit change in x. In addition, we can estimate ρ_I exactly as earlier although now it reflects the correlation between individuals from the same cluster after controlling for the independent variable, x. In this model, both γ_{10} and γ_{00} are fixed effects, while σ^2 and τ^2 remain random.

One implication of the model in Equation (2.12) is that the dependent variable is impacted by variations among individuals (σ^2), variations among clusters (τ^2), an overall mean common to all clusters (γ_{00}), and the impact of the independent variable as measured by γ_{10}, which is also common to all clusters.

In practice, however, there is no reason that the impact of x on y must be common for all clusters. In other words, it is entirely possible that rather than having a single γ_{10} common to all clusters, there is actually a unique effect for the cluster of $\gamma_{10} + U_{1j}$, where γ_{10} is the average relationship of x with y across clusters, and U_{1j} is the cluster-specific variation of the relationship between the two variables. This cluster-specific effect is assumed to have a mean of 0 and vary randomly around γ_{10}. The random slopes model is

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + U_{1j}x_{ij} + \varepsilon_{ij} \tag{2.16}$$

Written in this way, we have separated the model into its fixed ($\gamma_{00} + \gamma_{10}x_{ij}$) and random ($U_{0j} + U_{1j}x_{ij} + \varepsilon_{ij}$) components. The Equation (2.16) model simply indicates an interaction between cluster and x, such that the relationship of x and y is not constant across clusters.

Heretofore we discussed only one source of between-group variation, expressed as τ^2, that serves as the variation among clusters in the intercept. However, Equation (2.16) adds a second such source of between-group variance in the form of U_{1j}, which indicates cluster variation on the slope relating the independent and dependent variables. To differentiate these two sources of between-group variance, we now denote the variance of U_{0j} as τ_0^2 and the variance of U_{1j} as τ_1^2. Furthermore, within clusters we expect U_{1j} and U_{0j} to have a covariance of τ_{0_1}. However, across different clusters, these terms should be independent of one another, and in all cases it is assumed that ε remains independent of all other model terms. In practice, if we find that τ_1^2 is not 0, we must be careful in describing the relationship between the independent and dependent variables, as it is not the same for all clusters.

We will revisit this idea in subsequent chapters. For the moment, however, it is most important to recognize that variation in the dependent variable y can be explained by several sources, some fixed and others random. In practice, we will most likely be interested in estimating all of these sources of variability in a single model.

As a means for further understanding the MLM, let us consider a simple example using the five schools described above. In this context, we are interested in treating a reading achievement test score as the dependent variable and a vocabulary achievement test score as the independent variable. Remember that students are nested within schools so that a simple regression analysis is not appropriate. To understand the issue being estimated in the context of MLM, we can obtain separate intercept and slope estimates for each school as shown in Table 2.2.

Since the schools are of the same sample size, the estimate of γ_{00}, the average intercept value is 2.359, and the estimate of the average slope value γ_{10} is 0.375. Notice that for both parameters, the school values deviate from these means. For example, the intercept for school 1 is 1.230. The –1.129 difference between this value and 2.359 is U_{0j} for that school. Similarly, the

TABLE 2.2

Intercept and Slope Estimates of Multilevel Linear Model

School	Intercept	U_{0_j}	Slope	U_{1_j}
1	1.230	–1.129	0.552	0.177
2	2.673	0.314	0.199	–0.176
3	2.707	0.348	0.376	0.001
4	2.867	0.508	0.336	–0.039
5	2.319	–0.040	0.411	0.036
Overall	2.359		0.375	

difference between the average slope value of 0.375 and the slope for school 1, 0.552 is 0.177, which is U_{1j} for the school. Table 2.2 includes U_{0j} and U_{1j} values for each school. The differences in slopes also provide information about the relationship between vocabulary and reading test scores. This relationship was positive for all schools, meaning that students who scored higher on vocabulary also scored higher on reading. However, the strength of this relationship was weaker for school 2 than for school 1, as an example.

Based on the values in Table 2.2, it is also possible to estimate the variances associated with U_{1j} and U_{0j}, τ_1^2 and τ_0^2, respectively. Again, because the schools in this example had the same numbers of students, the calculation of these variances is a straightforward matter, using

$$\frac{\sum\left(U_{1j}-\bar{U}_1\right)^2}{J-1} \tag{2.17}$$

for the slopes and an analogous equation for the intercept random variance. We obtain $\tau_0^2 = 0.439$ and $\tau_1^2 = 0.016$. In other words, much more of the variance in the dependent variable is accounted for by variation in the intercepts at school level than is accounted for by variation in the slopes. Another way to think of this result is that the schools exhibited greater differences among one another in the mean level of achievement as compared to differences in the impacts of x on y.

The practice of obtaining these variance estimates using the R environment for statistical computing and graphics and interpreting their meaning are subjects for upcoming chapters. Before discussing the practical "nuts and bolts" of conducting this analysis, we first examine the basics for estimating parameters in the MLM framework using maximum likelihood and restricted maximum likelihood algorithms. While similar in spirit to the simple calculations demonstrated above, they are different in practice and will yield somewhat different results from those obtained using least squares as above. First, one more issue warrants our attention as we consider the use of MLM, namely variable centering.

2.4.3 Centering

Centering is simply the practice of subtracting the mean of a variable from each individual value. This implies the mean for the sample of the centered variables is 0 and also that each individual's (centered) score represents a deviation from the mean rather than representing the meaning of its raw value. In the context of regression, centering is commonly used, for example, to reduce collinearity caused by including an interaction term in a regression model. If the raw scores of the independent variables are used to calculate the interaction and both the main effects and interaction terms are included in the subsequent analysis, it is very likely that collinearity will cause problems in the standard errors of the model parameters. Centering is a way to help avoid such problems (Iversen, 1991).

Such issues are also important to consider in MLM, in which interactions are frequently employed. In addition, centering is also a useful tool for avoiding collinearity caused by highly correlated random intercepts and slopes in MLMs (Wooldridge, 2004). Finally, centering provides a potential advantage in terms of interpretation of results. Remember from our discussion in Chapter 1 that the intercept is the value of the dependent variable when the independent variable is set to 0. In many applications (e.g., a measure of vocabulary), the independent variable cannot reasonably be 0. This essentially renders the intercept as a necessary value for fitting the regression line but not one that has a readily interpretable value. However, when x has been centered, the intercept takes on the value of the dependent variable when the independent is at its mean. This is a much more useful interpretation for researchers in many situations, and yet another reason why centering is an important aspect of modeling, particularly in the multilevel context.

Probably the most common approach to centering is to calculate the difference between each individual's score and the overall, or grand mean across the entire sample. This *grand mean centering* is certainly the most commonly used method in practice (Bickel, 2007). It is not, however, the only manner of centering data. An alternative approach known as *group mean centering* involves calculating the difference between each individual score and the mean of the cluster to which it belongs. In our school example, grand mean centering would involve calculating the difference between each score and the overall mean across schools, while group mean centering would lead the researcher to calculate the difference between each score and the mean for the school.

While the literature indicates some disagreement regarding which approach may be best for reducing the harmful effects of collinearity (Bryk & Raudenbush, 2002; Snijders & Bosker, 1999), researchers demonstrated that either technique will work well in most cases (Kreft, de Leeuw, & Aiken, 1995). Therefore, the choice of which approach to use must be made on substantive grounds regarding the nature of the relationship between x and y. By using grand mean centering, we implicitly compare individuals to one another (in the form of the overall mean) across an entire sample.

On the other hand, group mean centering places each individual in relative position on *x* within his or her cluster. In our school example, using the group mean centered values of vocabulary in the analysis would mean that we are investigating the relationship between a student's relative vocabulary score in his or her school and his or her reading score. In contrast, the use of grand mean centering would examine the relationship between a student's relative standing in the sample as a whole on vocabulary and the reading score. This latter interpretation would be equivalent conceptually (but not mathematically) to using the raw score, while the group mean centering would not.

Throughout the rest of this book, we will use grand mean centering by default based on recommendations by Hox (2002), among others. At times, however, we will also demonstrate the use of group mean centering to illustrate how it provides different results and for applications in which interpretation of the impact of an individual's relative standing in his or her cluster may be more useful than the individual's relative standing in the sample as a whole.

2.5 Basics of Parameter Estimation with MLMs

Heretofore, our discussions of estimation of model parameters have been in the context of least squares—a technique that provides underpinnings of ordinary least squares (OLS) and related linear models. However, as we move from these fairly simple applications to more complex models, OLS is not typically the optimal approach for parameter estimation. Instead, we will rely on maximum likelihood estimation (MLE) and restricted maximum likelihood (REML). In the following sections, we review these approaches to estimation from a conceptual view, focusing generally on how they work, what they assume about the data, and how they differ from one another. For the technical details we refer interested readers to Bryk and Raudenbush (2002) and de Leeuw and Meijer (2008), both of which are excellent resources for those desiring more in-depth coverage of these methods. Our purpose here is to provide readers with a conceptual understanding that will aid their application of MLM techniques in practice.

2.5.1 Maximum Likelihood Estimation

MLE has as its primary goal the estimation of population model parameters that maximize the likelihood of obtaining the sample that we in fact obtained. In other words, the estimated parameter values should maximize the likelihood of our particular sample. From a practical perspective, identifying such sample values takes place by a comparison of the observed data with data predicted by the model associated with the parameter values. The closer the observed and predicted values are to one another, the greater the likelihood

that the observed data arose from a population with parameters close to those used to generate the predicted values. In practice, MLE is an iterative methodology in which the algorithm searches for parameter values that will maximize the likelihood of the observed data (i.e., produce predicted values that are as close as possible to observed values). MLE may be computationally intensive, particularly for complex models and large samples.

2.5.2 Restricted Maximum Likelihood Estimation

A variant of MLE known as restricted maximum likelihood estimation (REML) has proven more accurate than MLE for estimating variance parameters (Kreft & De Leeuw, 1998). In particular, the two methods differ with respect to calculating degrees of freedom in estimating variances. As a simple example, a sample variance is calculated typically by dividing the sum of squared differences between individual values and the mean by the number of observations minus 1 to yield an unbiased estimate. This is a REML estimate of variance.

In contrast, the MLE variance is calculated by dividing the sum of squared differences by the total sample size, leading to a smaller variance estimate than REML and, in fact, one biased in finite samples. In the context of multilevel modeling, REML accounts for the number of parameters being estimated in a model when determining the appropriate degrees of freedom for the estimation of the random components such as the parameter variances described above. In contrast, MLE does not account for these, leading to an underestimate of the variances that does not occur with REML. For this reason, REML is generally the preferred method for estimating multilevel models, although for testing variance parameters (or any random effect), it is necessary to use MLE (Snijders & Bosker, 1999). We should note that as the number of Level 2 clusters increases, the difference in value for MLE and REML estimates becomes very small (Snijders & Bosker, 1999).

2.6 Assumptions Underlying MLMs

As with any statistical model, the appropriate use of MLMs requires that several assumptions about the data hold true. If these assumptions are not met, the model parameter estimates may not be trustworthy, as would be the case with standard linear regression reviewed in Chapter 1. Indeed, while the assumptions for MLM differ somewhat from those for single-level models, the assumptions underlying MLM are akin to those for the simpler models. This section introduces these assumptions and their implications for researchers using MLMs. In subsequent chapters, we describe methods for checking the validity of these assumptions for given sets of data.

First, we assume that the Level 2 residuals are independent between clusters. In other words, the assumption is that the random intercept and slope(s) at Level 2 are independent of one another across clusters. Second, the Level 2 intercepts and coefficients are assumed to be independent of the Level 1 residuals, i.e., errors for the cluster-level estimates are unrelated to errors at the individual level. Third, the Level 1 residuals are normally distributed and have constant variances. This assumption is very similar to the one we make about residuals in the standard linear regression model. Fourth, the Level 2 intercept and slope(s) have a multivariate normal distribution with a constant covariance matrix. Each of these assumptions can be directly assessed for a sample, as we shall see in forthcoming chapters. Indeed, the methods for checking the MLM assumptions are similar to those for checking the regression model that we used in Chapter 1.

2.7 Overview of Two-Level MLMs

We have described the specific terms of MLM, including the Level 1 and Level 2 random effects and residuals. We will close this chapter about MLMs by considering examples of two- and three-level MLMs and the use of MLMs with longitudinal data. This discussion should prepare the reader for subsequent chapters covering applications of R to the estimations of specific MLMs.

First, we consider the two-level MLM, parts of which we described earlier in this chapter. In Equation (2.16), we considered the random slopes model

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + U_{1j}x_{ij} + \varepsilon_{ij}$$

in which the dependent variable y_{ij} (reading achievement) was a function of an independent variable x_{ij} (vocabulary test score) and also random error at both the student and school levels. We can extend this model a bit further by including multiple independent variables at both Level 1 (student) and Level 2 (school). Thus, for example, in addition to ascertaining the relationship between an individual's vocabulary and reading scores, we can also determine the degree to which the average vocabulary score at the school as a whole is related to an individual's reading score. This model essentially has two parts: (1) one explaining the relationship between the individual level vocabulary (x_{ij}) and reading and (2) one explaining the coefficients at Level 1 as a function of the Level 2 predictor or average vocabulary score (z_j). The two parts of this model are expressed as

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij} \tag{2.18}$$

$$\text{Level 2: } \beta_{hj} = \gamma_{h0} + \gamma_{h1}z_j + U_{hj} \tag{2.19}$$

The additional piece of Equation (2.19) is $\gamma_{h1}z_j$, which represents the slope for (γ_{h1}), and value of the average vocabulary score for the school (z_j). In other words, the mean school performance is related directly to the coefficient linking the individual vocabulary score to the individual reading score. For our specific example, we can combine Equations (2.18) and (2.19) to yield a single equation for the two-level MLM.

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_j + \gamma_{1001}x_{ij}z_j + U_{0j} + U_{1j}x_{ij} + \varepsilon_{ij} \tag{2.20}$$

Each of these model terms has been defined previously in this chapter: γ_{00} is the intercept or grand mean for the model, γ_{10} is the fixed effect of variable x (vocabulary) on the outcome, U_{0j} represents the random variation for the intercept across groups, and U_{1j} represents the random variation for the slope across groups.

The additional pieces of Equation (2.13) are γ_{01} and γ_{11}. The γ_{01} represents the fixed effect of Level 2 variable z (average vocabulary) on the outcome and γ_{11} represents the slope for and value of the average vocabulary score for the school. The new term in Equation (2.20) is the cross-level interaction $\gamma_{1001}x_{ij}z_j$. As the name implies, the cross-level interaction is simply an interaction of Level 1 and Level 2 predictors. In this context, it represents the interaction between an individual's vocabulary score and the mean vocabulary score for his or her school. The coefficient for this interaction term, γ_{1001}, assesses the extent to which the relationship between a student's vocabulary score is moderated by the mean for the school attended. A large significant value for this coefficient would indicate that the relationship between an individual's vocabulary test score and overall reading achievement is dependent on the level of vocabulary achievement at his or her school.

2.8 Overview of Three-Level MLMs

It is entirely possible to utilize three or more levels of data structures with MLMs. We should note, however, that four-level and larger models are rare in practice. For our reading achievement data in which the second level was school, a possible third level might be the district in which the school is located. In that case, we would have multiple equations to consider when expressing the relationship between vocabulary and reading achievement scores, starting at the individual level:

$$y_{ijk} = \beta_{0jk} + \beta_{1jk}x_{ijk} + \varepsilon_{ijk} \tag{2.21}$$

The subscript k represents the Level 3 cluster to which the individual belongs.

Before formulating the rest of the model, we must evaluate whether the slopes and intercepts are random at both Levels 2 and 3 or only at Level 1, for example. This decision should always be based on the theory surrounding the research questions, what is expected in the population, and what is revealed in the empirical data. We will proceed with the remainder of this discussion under the assumption that the Level 1 intercepts and slopes are random for both Levels 2 and 3 in order to provide a complete description of the most complex model possible when three levels of data structure are present. When the Level 1 coefficients are not random at both levels, the terms in the following models for which this randomness is not present would simply be removed. We will address this issue more specifically in Chapter 4 when we discuss the fitting of three-level models using R. The Level 2 and Level 3 contributions to the MLM described in Equation (2.13) appear below.

$$\begin{aligned} \text{Level 2: } \beta_{0jk} &= \gamma_{00k} + U_{0jk} \\ \beta_{1jk} &= \gamma_{10k} + U_{1jk} \\ \text{Level 3: } \gamma_{00k} &= \delta_{000} + V_{00k} \\ \gamma_{10k} &= \delta_{100} + V_{10k} \end{aligned} \tag{2.22}$$

We can then use simple substitution to obtain the expression for the Level 1 intercept and slope in terms of both Level 2 and Level 3 parameters.

$$\begin{aligned} \beta_{0jk} &= \delta_{000} + V_{00k} + U_{0jk} \\ \beta_{1jk} &= \delta_{100} + V_{10k} + U_{1jk} \end{aligned} \tag{2.23}$$

In turn, these terms may be substituted into Equation (2.15) to provide the full three-level MLM.

$$y_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + (\delta_{100} + V_{10k} + U_{1jk})x_{ijk} + \varepsilon_{ijk} \tag{2.24}$$

There is an implicit assumption in this expression of Equation (2.24) that there are no cross-level interactions, although they certainly may be modeled across all three levels or for any pair of levels. Equation (2.24) expresses individuals' scores on the reading achievement test as a function of random and fixed components from the school they attend, the district in which the school is located, and their own vocabulary test scores and random variations associated only with them. Although not included in Equation (2.24), it is also possible to include variables at both Levels 2 and 3, similar to what we described for the two-level model structure.

2.9 Overview of Longitudinal Designs and Their Relationship to MLMs

Finally, we will briefly explain how longitudinal designs can be expressed as MLMs. Longitudinal research designs simply involve the collection of data from the same individuals at multiple points in time. For example, we may have reading achievement scores for students tested in the fall and spring of the school year. With such a design, we would be able to investigate aspects of growth scores and changes in achievements over time. Such models can be placed in the context of an MLM where the student represents the Level 2 (cluster) variable, and the individual test administration is at Level 1. We would then simply apply the two-level model described above, including student-level variables that are appropriate for explaining reading achievement. Similarly, if students are nested within schools, we would have a three-level model, with school serving as the third level. We could apply Equation (2.24) again with whichever student- or school-level variables were pertinent to the research question.

One unique aspect of fitting longitudinal data into the MLM context is that the error terms can potentially take specific forms that are not common in other applications of multilevel analysis. These error terms reflect the way in which measurements made over time relate to one another and are typically more complex than the basic error structure described thus far. In Chapter 5, we will consider examples of fitting such longitudinal models with R and focus our attention on these error structures—when each is appropriate and how they are interpreted. In addition, such MLMs need not take linear forms. They may be adapted to fit quadratic, cubic, or other nonlinear trends over time. These issues will be discussed further in Chapter 5.

Summary

The goal of this chapter was to introduce the basic theoretical underpinnings of multilevel modeling, but not to provide an exhaustive technical discussion of these issues. A number of useful resources can provide comprehensive details and are listed in the references at the end of the book. However, the information in this chapter should be adequate as we move forward with multilevel modeling using R software. We recommend that you make liberal use of the information provided here while reading subsequent chapters. This should provide you with a complete understanding of the output generated by R that we will be examining. In particular, when interpreting output from R, it may be helpful for you to return to this chapter to review precisely what each model parameter means.

In the next two chapters, we will take the theoretical information from this chapter and apply it to real data sets using two different R libraries, `nlme` and `lme4`, both of which were developed for conducting multilevel analyses with continuous outcome variables. In Chapter 5, we will examine how these ideas can be applied to longitudinal data. Chapters 7 and 8 will discuss multilevel modeling for categorical dependent variables. In Chapter 9, we will diverge from the likelihood-based approaches described here and explain multilevel modeling within the Bayesian framework, focusing on applications and learning when this method may be appropriate and when it may not.

3

Fitting Two-Level Models in R

In the previous chapter, the multilevel modeling approach to analysis of nested data was introduced along with relevant notations and definitions of random intercepts and coefficients. We will devote this chapter to the introduction of the R packages for fitting multilevel models. In Chapter 1, we provided an overview of the `lm()` function for linear regression models. As will become apparent, the estimation of multilevel models in R is very similar to estimating single-level linear models. After providing a brief discussion of the two primary R packages for fitting multilevel models for continuous data, we will devote the remainder of the chapter to extended examples applying the principles introduced in Chapter 2 using R.

3.1 Packages and Functions for Multilevel Modeling in R

Currently, the two main R libraries for devising multilevel models are `nlme` and `lme4`, both of which can be used for fitting basic and advanced multilevel models. The `lme4` package is slightly newer and provides a more concise syntax and more flexibility. Using the `nlme` package, the function call for continuous outcome multilevel models that are linear in their parameters is `lme()`, whereas the function call in `lme4` is `lmer()`.

In the following sections of this chapter, we will demonstrate and provide examples of using these two packages to run basic multilevel models in R. Following is the basic syntax for these two functions. Details regarding their use and various options will be provided in the examples.

```
lme(fixed, data, random, correlation, weights, subset, method,
    na.action, control, contrasts = NULL, keep.data = TRUE)

lmer(formula, data, family = NULL, REML = TRUE,
     control = list(), start = NULL, verbose = FALSE,
     doFit = TRUE, subset, weights, na.action, offset,
     contrasts = NULL, model = TRUE, x = TRUE,...)
```

For simple linear multilevel models, the only necessary R subcommands for the functions are the formula (consisting of fixed and random effects)

and data. The remaining subcommands can be used to customize models and to provide additional output. This chapter focuses first on defining simple multilevel models and then demonstrates options for model customization and assumption checking.

3.2 The `nlme` Package

3.2.1 Simple (Intercept Only) Multilevel Models Using `nlme`

To demonstrate the use of R for fitting multilevel models, we return to the example introduced in Chapter 2. Specifically, a researcher wants to determine the extent to which vocabulary scores can be used to predict general reading achievement. Since students were nested within schools, standard linear regression models are not appropriate. In this case, school is a random effect and vocabulary scores are fixed. The first model that we will fit is the null model that has no independent variable. This model is useful for obtaining estimates of the residual and intercept variance when only the clustering by school is considered, as in Equation (2.11). The `lme` syntax necessary for estimating the null model appears below.

```
Model3.0 <- lme(fixed = gerad~1, random = ~1|school, data =
                Achieve)
```

We can obtain output from this model by typing `summary(Model3.0)`.

```
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC       BIC     logLik
  46274.31  46296.03  -23134.15

Random effects:
 Formula: ~1 | school
        (Intercept)  Residual
StdDev:   0.6257119   2.24611

Fixed effects: gerad ~ 1
               Value   Std.Error      DF    t-value    p-value
(Intercept) 4.306753  0.05497501   10160    78.3402          0

Standardized Within-Group Residuals:
       Min          Q1         Med          Q3         Max
-2.3229469  -0.6377948 -0.2137753   0.2849664  3.8811630

Number of Observations: 10320
Number of Groups: 160
```

Although this is a null model in which there is no independent variable, it provides some useful information that will help us understand the structure of the data. In particular, the AIC and BIC values that are of primary interest in this case will be useful in comparing this model with others that include one or more independent variables, as we will see below. In addition, the null model also provides estimates of the variance among the individuals σ^2 and among the clusters τ^2. In turn, these values can be used to estimate ρ_I (ICC), as in Equation (2.5). Here, the value would be

$$\hat{\rho}_I = \frac{0.6257119}{0.6257119 + 2.24611} = 0.2178797$$

We interpret this value to mean that the correlation of reading test scores among students within the same schools is 0.22 if we round our result. To fit the model with vocabulary as the independent variable using `lme`, we submit the following syntax in R.

```
Model3.1 <- lme(fixed = geread~gevocab, random = ~1|school,
                data = Achieve)
```

In the first part of the function call, we define the formula for the model fixed effects, very similar to model definition of linear regression using `lm()`. The statement `fixed = geread~gevocab` essentially says that the reading score is predicted with the vocabulary score fixed effect. The `random` part of the function call defines the random effects and the nesting structure. If only a random intercept is desired, the syntax for the intercept is `1`. In this example, `random = ~1|school` indicates that only a random intercepts model will be used and that the random intercept varies within school. This corresponds to the data structure of students nested within schools. Fitting this model, which is saved in the output object `Model3.1`, we obtain the following output by inputting the name of the output object.

```
Model3.1
Linear mixed-effects model fit by REML
 Data: Achieve
 Log-restricted-likelihood: -21568.6
 Fixed: geread ~ gevocab
(Intercept)     gevocab
  2.0233559   0.5128977

Random effects:
 Formula: ~1 | school
       (Intercept)   Residual
StdDev: 0.3158785   1.940740

Number of Observations: 10320
Number of Groups: 160
```

Output from the `lme()` function provides parameter estimates for the fixed effects and standard deviations for the random effects along with a summary of the number of Level 1 and Level 2 units in the sample. As with the output from the `lm()` function, however, the output from the `lme()` function provides limited information. If we desire more detailed information about the model, including significance tests for parameter estimates and model fit statistics, we can request a model summary. The `summary()` command will provide the following:

```
summary(Model3.1)
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC       BIC    logLik
   43145.2  43174.17  -21568.6

Random effects:
 Formula: ~1 | school
        (Intercept) Residual
StdDev:   0.3158785 1.940740

Fixed effects: geread ~ gevocab
                Value   Std.Error    DF   t-value p-value
(Intercept) 2.0233559 0.04930868 10159  41.03447       0
gevocab     0.5128977 0.00837268 10159  61.25850       0
  Correlation:
         (Intr)
gevocab  -0.758

Standardized Within-Group Residuals:
       Min          Q1         Med          Q3         Max
-3.0822506  -0.5734728 -0.2103488   0.3206692  4.4334337

Number of Observations: 10320
Number of Groups: 160
```

From this summary we obtain AIC, BIC, and log likelihood information that can be used for model comparisons in addition to parameter significance tests. We can also obtain a correlation between the fixed effect slope and the fixed effect intercept as well as a brief summary of the model residuals including the minimum, maximum, and first, second (median, denoted Med), and third quartiles.

The correlation of the fixed effects represents the estimated correlation if we had repeated samples of the two fixed effects (i.e., the intercept and slope for `gevocab`). Often this correlation is not particularly interesting. From this output, we can see that `gevocab` is a significant predictor of `geread` ($t = 61.258$, $p < 0.05$), and that as vocabulary score increases by 1 point, reading ability increases by 0.513 points. We can compare the fit

for this model with that of the null model by referring to the AIC and BIC statistics. Recall that smaller values reflect better model fit. For Model 3.1, the AIC and BIC are 43145.2 and 43174.17, respectively. For Model 3.0, the AIC and BIC were 46274.31 and 46296.03. Because the values for both statistics are smaller for Model 3.1, we would conclude that it provides a better fit to the data. Substantively, this means that we should include the predictor variable `gread`, which the results of the hypothesis test also supported.

In addition to the fixed effects in Model 3.1, we can also ascertain how much variation in `gread` is present across schools. Specifically, the output shows that after accounting for the impact of `gevocab`, the estimate of variation in intercepts across schools is 0.3158785, while the within-school variation is estimated as 1.940740. We can tie these numbers directly back to our discussion in Chapter 2 where $\tau_0^2 = 0.3158785$ and $\sigma^2 = 1.940740$. In addition, the overall fixed intercept denoted as γ_{00} in Chapter 2 is 2.0233559, which is the mean of `gread` when the `gevocab` score is 0.

Finally, it is possible to estimate the proportion of variance in the outcome variable accounted for at each level of the model. In Chapter 1, we saw that with single-level OLS regression models, the proportion of response variable variance accounted for by the model is expressed as R^2. In the context of multilevel modeling, R^2 values can be estimated for each level of the model (Snijders & Bosker, 1999). For Level 1, we can calculate

$$\begin{aligned} R_1^2 &= 1 - \frac{\sigma_{M1}^2 + \tau_{M1}^2}{\sigma_{M1}^2 + \tau_{M1}^2} \\ &= 1 - \frac{1.940740 + 0.3158785}{2.24611 + 0.6257119} \\ &= 1 - \frac{2.2566185}{2.8718219} = 1 - 0.7857794 = 0.2142206 \end{aligned}$$

This result tells us that Level 1 of Model 3.1 explains approximately 21% of the variance in the reading score above and beyond that accounted for in the null model. We can also calculate a Level 2 R^2 value:

$$R_2^2 = 1 - \frac{\sigma_{M1}^2 / B + \tau_{M1}^2}{\sigma_{M0}^2 / B + \tau_{M0}^2}$$

where B is the average size of the Level 2 units (schools in this case). R provides the number of individuals in the sample (10320) and the number of schools (160) so that we can calculate B as 10320/160 = 64.5. We can now estimate

$$R_2^2 = 1 - \frac{\sigma_{M1}^2 / B + \tau_{M1}^2}{\sigma_{M0}^2 / B + \tau_{M0}^2} =$$

$$R_1^2 = 1 - \frac{\sigma_{M1}^2 + \tau_{M1}^2}{\sigma_{M0}^2 + \tau_{M0}^2}$$

$$= 1 - \frac{1.940760 + 0.3167654}{2.24611 + 0.6257119}$$

$$= 1 - \frac{2.2575254}{2.8718219} = 1 - 0.7860952 = 0.2139048$$

The model in the previous example was quite simple and incorporated only a single Level 1 predictor. In many applications, researchers utilize predictor variables at both Level 1 (student) and Level 2 (school). Incorporation of predictors at higher levels of analysis is straightforward in R and is handled in exactly the same manner as incorporation of Level 1 predictors. For example, let us assume that in addition to a student's vocabulary test performance, a researcher also wants to determine whether school enrollment size (`senroll`) also produces a statistically significant impact on overall reading score. In that instance, adding the school enrollment Level 2 predictor would result in the following R syntax:

```
Model3.2 <- lme(fixed = geread~gevocab + senroll, random =
                ~1|school, data = Achieve)

summary(Model3.2)
Linear mixed-effects model fit by REML
 Data: Achieve
      AIC       BIC    logLik
  43162.1  43198.31 -21576.05

Random effects:
Formula: ~1 | school
       (Intercept) Residual
StdDev: 0.3167654 1.940760

Fixed effects: geread ~ gevocab + senroll
                Value   Std.Error     DF   t-value  p-value
(Intercept) 2.0748819  0.11400758  10159  18.19951   0.0000
gevocab     0.5128708  0.00837340  10159  61.25000   0.0000
senroll    -0.0001026  0.00020511    158  -0.50012   0.6177
  Correlation:
          (Intr)  gevocb
gevocab   -0.327
senroll   -0.901  -0.002
```

```
Standardized Within-Group Residuals:
       Min          Q1         Med          Q3         Max
-3.0834462  -0.5728938  -0.2103480  0.3212091   4.4335881

Number of Observations: 10320
Number of Groups: 160
```

Note that in this specific function call, `senroll`, is included only in the fixed part of the model and not in the random part. This variable thus has only a fixed (average) effect and is the same across all schools. We will see shortly how to incorporate a random coefficient in this model.

From these results we can see that enrollment did not have a statistically significant relationship with reading achievement. In addition, notice some minor changes in the estimates of the other model parameters and a fairly large change in the correlation between the fixed effect of `gevocab` slope and the fixed effect of the intercept. The slope for `senroll` and intercept were strongly negatively correlated and the slopes of the fixed effects exhibited virtually no correlation. As noted earlier, these correlations are typically not very helpful for explaining the dependent variable and are rarely discussed in any detail in reports of analysis results. The R^2 values for Levels 1 and 2 appear below.

$$R_1^2 = 1 - \frac{\sigma_{M1}^2 + \tau_{M1}^2}{\sigma_{M0}^2 + \tau_{M0}^2}$$

$$= 1 - \frac{1.940760 + 0.3167654}{2.24611 + 0.6257119}$$

$$= 1 - \frac{2.2575254}{2.8718219} = 1 - 0.7860952 = 0.2139048$$

$$R_2^2 = 1 - \frac{\sigma_{M1}^2 / B + \tau_{M1}^2}{\sigma_{M0}^2 / B + \tau_{M0}^2}$$

$$= 1 - \frac{1.940760/64.5 + 0.3167654}{2.24611/64.5 + 0.6257119}$$

$$= 1 - \frac{0.34685}{0.66053} = 1 - 0.52378 = 0.474884$$

3.2.2 Random Coefficient Models Using `nlme`

In Chapter 2, we described the random coefficients model in which the impact of the independent variable on the dependent is allowed to vary across the Level 2 effects. In the context of the current research problem, this would mean that we allow the impact of `gevocab` on `gread` to vary from one school to another. Incorporating such random coefficient effects

into a multilevel model using `lme` occurs in the `random` part of the model syntax. When defining random effects, as mentioned above, `1` stands for the intercept, so that if all we desire is a random intercepts model as in the previous example, the syntax `~1|school` is sufficient. If, however, we want to allow a Level 1 slope to vary randomly, we will change this part of the syntax (recall that `gevocab` is already included in the fixed part of the model). Let us return to the Model 3.1 scenario, but this time allow both the slope and intercept for `gevocab` to vary randomly from one school to another. The syntax for this model would now become

```
Model3.3 <- lme(fixed = geread~gevocab, random =
                ~gevocab|school, data = Achieve)
```

This model differs from Model 3.1 only in that the `1` in the random line is replaced by the variable name whose effect we want to be random. Notice that we no longer explicitly state a random intercept in the specification. After a random slope is defined, the random intercept becomes implicit so we no longer need to specify it (i.e., it is included by default). If we do not want the random intercept while modeling the random coefficient, we would include a `–1` immediately prior to `gevocab`. The random slope and intercept syntax will generate the following model summary:

```
summary(Model3.3)
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC       BIC    logLik
  43004.85   43048.3 -21496.43

Random effects:
Formula: ~gevocab | school
Structure: General positive-definite, Log-Cholesky
           parametrization
              StdDev     Corr
(Intercept)   0.5316640  (Intr)
gevocab       0.1389372  -0.858
Residual      1.9146629

Fixed effects: geread ~ gevocab
                 Value   Std.Error     DF   t-value  p-value
(Intercept)  2.0057073  0.06108846  10159  32.83283        0
gevocab      0.5203554  0.01441502  10159  36.09815        0
 Correlation:
        (Intr)
gevocab -0.866

Standardized Within-Group Residuals:
       Min           Q1           Med          Q3         Max
-3.7101835  -0.5674382   -0.2074307  0.3176354   4.6774104
```

```
Number of Observations: 10320
Number of Groups: 160
```

An examination of the results shows that `gevocab` is statistically significantly related to `geread` across schools. The estimated coefficient 0.5203554 corresponds to γ_{10} from Chapter 2, and is interpreted as the average impact of the predictor on the outcome across schools. In addition, the value 0.1389372 represents the estimate of τ_1^2 from Chapter 2, and reflects the variation in coefficients across schools. A relatively larger value of this estimate indicates that the coefficient varies from one school to another; i.e., the relationship of the independent and dependent variables differs across schools. As before, we also have the estimates of τ_0^2 (0.5316640) and σ^2 (1.9146629). Taken together these results show that the largest source of random variation in `geread` is variation among students within schools, with lesser variation from differences in the conditional mean (intercept) and coefficient for `gevocab` across schools.

A model with two random slopes can be defined in much the same way as defining a single slope. As an example, suppose a researcher is interested in determining whether the age of a student also impacts reading performance, and wants to allow this effect to vary from one school to another. Such incorporation of two random slopes can be modeled as:

```
Model3.4 <- lme(fixed = geread~gevocab + age,
                random = ~gevocab + age|school, data = Achieve)

summary(Model3.4)
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC        BIC      logLik
  43015.77   43088.18   -21497.88

Random effects:
Formula: ~gevocab + age | school
Structure: General positive-definite, Log-Cholesky
           parametrization
              StdDev       Corr
(Intercept)   0.492561805  (Intr) gevocb
gevocab       0.137974552  -0.073
age           0.006388612  -0.649 -0.601
Residual      1.914030323

Fixed effects: geread ~ gevocab + age
                 Value  Std.Error     DF   t-value   p-value
(Intercept)  2.9614102  0.4151894  10158   7.13267    0.0000
gevocab      0.5191491  0.0143562  10158  36.16205    0.0000
age         -0.0088390  0.0038396  10158  -2.30208    0.0214
Correlation:
        (Intr) gevocb
gevocab -0.095
age     -0.989 -0.032
```

```
Standardized Within-Group Residuals:
       Min         Q1          Med         Q3        Max
-3.6805437  -0.5686992  -0.2091111  0.3180592  4.6850568

Number of Observations: 10320
Number of Groups: 160
```

Here we see that age is significantly related to `geread` ($p = 0.0214$), with a negative coefficient indicating that older students had lower scores. In addition, the random variance of coefficients for this variable across schools (0.006388612) is much smaller than that of `gevocab` (0.137974552), leading us to conclude that the relationship of vocabulary on reading varies more across schools than does the impact of age.

3.2.3 Interactions and Cross-Level Interactions Using `nlme`

Interactions among the predictor variables, particularly cross-level interactions, can be very important in the application of multilevel models. Cross-level interactions occur when the impact of a Level 1 variable on an outcome (e.g., vocabulary score) differs based on the value of the Level 2 predictor (e.g., school enrollment). Interactions, whether within the same level or across levels, are simply the products of two predictors. Thus, incorporation of interactions and cross-level interactions in multilevel modeling is accomplished in much the same manner as we saw for the `lm()` function in Chapter 1. Following are examples for fitting an interaction model for two Level 1 variables (Model 3.5) and a cross-level interaction involving Level 1 and Level 2 variables (Model 3.6).

```
Model3.5 <- lme(fixed = geread~gevocab + age + gevocab*age,
                random = ~1|school, data = Achieve)

Model3.6 <- lme(fixed = geread~gevocab + senroll +
                gevocab*senroll, random = ~1|school, data =
                Achieve)
```

Model 3.5 defines a multilevel model in which two Level 1 (student level) predictors interact with each other. Model 3.5 defines a multilevel model with a cross-level interaction in which a Level 1 (student level) and Level 2 (school level) predictor interact. Note that no difference exists in the treatment of variables at different levels when computing interactions.

```
summary(Model3.5)
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC       BIC    logLik
  43155.49  43198.94 -21571.75
```

```
Random effects:
 Formula: ~1 | school
        (Intercept) Residual
StdDev:   0.3142524 1.939708

Fixed effects: geread ~ gevocab + age + gevocab * age
                 Value  Std.Error     DF   t-value   p-value
(Intercept)  5.187208  0.8667857  10157  5.984418    0.0000
gevocab     -0.028078  0.1881452  10157 -0.149233    0.8814
age         -0.029368  0.0080348  10157 -3.655077    0.0003
gevocab:age  0.005027  0.0017496  10157  2.873204    0.0041
Correlation:
            (Intr) gevocb    age
gevocab     -0.879
age         -0.998  0.879
gevocab:age 0.877  -0.999 -0.879

Standardized Within-Group Residuals:
       Min          Q1          Med          Q3        Max
-3.0635106  -0.5706179  -0.2108349  0.3190991  4.4467448

Number of Observations: 10320
Number of Groups: 160
```

We can see from the output of Model 3.5 that both age ($t = -3.65$, $p < 0.01$) and the interaction (`gevocab:age`) between age and vocabulary ($t = 2.87$, $p < 0.01$) are significant predictors of reading. Focusing on the interaction, the sign on the coefficient is positive. This indicates an enhancing effect: as age increases, the relationship of reading and vocabulary becomes stronger.

```
summary(Model3.6)
Linear mixed-effects model fit by REML
 Data: Achieve
        AIC       BIC     logLik
   43175.57  43219.02  -21581.79

Random effects:
 Formula: ~1 | school
        (Intercept)   Residual
StdDev:    0.316492   1.940268

Fixed effects: geread ~ gevocab + senroll + gevocab * senroll
                     Value  Std.Error    DF    t-value  p-value
(Intercept)      1.7477004 0.17274011 10158 10.117513   0.0000
gevocab          0.5851202 0.02986497 10158 19.592189   0.0000
senroll          0.0005121 0.00031863   158  1.607242   0.1100
gevocab:senroll -0.0001356 0.00005379 10158 -2.519975   0.0118
```

```
  Correlation:
                (Intr)   gevocb   senrll
gevocab         -0.782
senroll         -0.958   0.735
gevocab:senroll  0.752  -0.960   -0.766

Standardized Within-Group Residuals:
       Min           Q1          Med         Q3       Max
-3.1228018   -0.5697103   -0.2090374  0.3187827 4.4358936

Number of Observations: 10320
Number of Groups: 160
```

The output from Model 3.6 has a similar interpretation. When school enrollment is used instead of age as a predictor, the main effect of vocabulary ($t = 19.59$, $p < 0.001$) and the interaction between vocabulary and school enrollment ($t = -2.51$, $p < 0.05$) are significant predictors of reading achievement. Focusing on the interaction, since the sign on the coefficient is negative we would conclude that there is a buffering or inhibitory effect. In other words, as school size increases, the relationship between vocabulary and reading achievement becomes weaker.

3.2.4 Centering Predictors

Based on discussions in Chapter 2, it may be advantageous to center predictors, especially when interactions are incorporated. Centering predictors can provide slightly easier interpretation of interaction terms and also help alleviate multicollinearity arising from inclusion of both main effects and interactions in the same model. Recall that centering of a variable entails the subtraction of a mean value from each score in the variable. Centering of predictors can be accomplished through R by the creation of new variables. For example, returning to Model 3.5, grand mean centered `gevocab` and `age` variables can be created with the following syntax:

```
Cgevocab <- Achieve$gevocab - mean(Achieve$gevocab)
Cage <- Achieve$age - mean(Achieve$age)
```

After mean centered versions of the predictors are created, they can be incorporated into the model in the same manner used earlier.

```
Model3.5.C <- lme(fixed = geread~Cgevocab + Cage +
                  Cgevocab*Cage,
                  random = ~1|school, data = Achieve)

summary(Model3.5.C)
Linear mixed-effects model fit by REML
 Data: Achieve
       AIC      BIC    logLik
  43155.49 43198.94 -21571.75
```

```
Random effects:
Formula: ~1 | school
       (Intercept)  Residual
StdDev:  0.3142524  1.939708

Fixed effects: geread ~ Cgevocab + Cage + Cgevocab * Cage
                 Value    Std.Error     DF    t-value p-value
(Intercept)   4.332326  0.03206185  10157  135.12403  0.0000
Cgevocab      0.512480  0.00837950  10157   61.15878  0.0000
Cage         -0.006777  0.00391727  10157   -1.72999  0.0837
Cgevocab:Cage 0.005027  0.00174965  10157    2.87320  0.0041
  Correlation:
              (Intr) Cgevcb  Cage
Cgevocab      0.008
Cage          0.007  0.053
Cgevocab:Cage 0.043  0.021  0.205

Standardized Within-Group Residuals:
       Min         Q1        Med        Q3       Max
-3.0635106 -0.5706179 -0.2108349 0.3190991 4.4467448

Number of Observations: 10320
Number of Groups: 160
```

First, notice the identical model fit (compare AIC, BIC, and log likelihood) of the centered and uncentered models. This is a good way to ensure that centering worked. Looking now to the fixed effects of the model, we see some changes in their interpretation. These differences are likely due to multicollinearity issues in the original uncentered model. The interaction is still significant ($t = 2.87$, $p < 0.05$) but we now see a significant effect of vocabulary ($t = 61.15$, $p < 0.01$). Age is no longer a significant predictor ($t = –1.73$ $p > 0.05$). Focusing on the interaction, recall that when predictors are centered, an interaction can be interpreted as the effect of one variable while holding the second variable constant. Since the sign on the interaction is positive, vocabulary has a positive impact on reading ability if we hold age constant.

3.3 The `lme4` Package

3.3.1 Random Intercept Models Using `lme4`

The previous discussion focused on using the `lme` function from the `nlme` library to fit multilevel models in R. As noted previously in this chapter, a second function for fitting such models, called `lme4`, is available in the `lmer` library. We will see that in some ways the syntax and output from these two functions are virtually identical. However, they exhibit some fundamental

differences that we must consider as we apply them. We will focus on some of these differences and their implications for practice. In particular, the `lme4` package offers a slightly more streamlined syntax for fitting multilevel models. It also provides a more flexible framework for definition of complex models. In `lme4`, we would fit Model 3.1 using the following syntax:

```
Model3.7 <- lmer(geread~gevocab +(1|school), data = Achieve)
```

The model is defined in much the same way as we defined the `lme` function, where the outcome variable is the sum or linear combination of all of the random and fixed effects. The only difference in treatment of fixed and random effects is that the random effects require information on the nesting structure (students within schools in this case) for the parameter within which they vary. The primary difference in model syntax between `lme` and `lmer` is that the random effect is denoted by its appearance within parentheses rather than through explicit assignment using the `random` statement. This syntax will yield the following output:

```
Model3.7
Linear mixed model fit by REML
Formula: geread ~ gevocab + (1 | school)
   Data: Achieve
   AIC   BIC   logLik   deviance  REMLdev
 43145 43174   -21569      43124    43137
Random effects:
   Groups  Name          Variance   Std.Dev.
   school  (Intercept)   0.099779    0.31588
   Residual              3.766470    1.94074
Number of obs: 10320,    groups: school, 160

Fixed effects:
            Estimate   Std. Error    t value
(Intercept) 2.023343     0.049305      41.04
gevocab     0.512901     0.008373      61.26

Correlation of Fixed Effects:
        (Intr)
gevocab -0.758
```

From this output we can see one obvious benefit of the `lme4` package is that all important information is presented without requiring the use of a `summary` statement. The function call alone is enough to provide model fit statistics, parameter estimates, parameter significance tests, parameter estimate correlations, residuals, and sample summaries. We can also see that the `lme4` package includes deviance and `REML` estimated deviance values in the model fit statistics in addition to the AIC, BIC, and log likelihood reported in the `nlme` package. What the `lme4` package does not include are *p* values for model coefficients.

In comparing the outputs of `lme` and `lmer`, we notice that while both *t* values and accompanying *p* values are reported in the `nlme` package, only the *t* values for fixed effects are reported in `lme4`. The reason for this discrepancy in the reported results, and specifically for the lack of *p* values is somewhat complex and is not within the scope of this book. However, we should note that the standard approach for finding *p* values based on using the reference *t* distribution, which would seem to be the intuitively correct step, does in fact not yield correct values in many cases. Therefore, some alternative approach for obtaining them is necessary.

Douglas Bates, the developer of `lme4`, recommends the use of Markov chain Monte Carlo (MCMC) methods to obtain *p* values for mixed model effects. We review MCMC in greater detail in Chapter 9 so that readers may gain an understanding of how this method works. We can say at this point that the computer-intensive MCMC approach relies on generating a posterior distribution for each model parameter, then using the distributions to obtain *p* values and confidence intervals for each parameter estimate. To obtain MCMC *p* values and confidence intervals for `lme` objects, we must install the `coda` and `languageR` packages and then use the following command sequence to obtain the desired statistics for Model 3.7.

```
library(coda)
library(languageR)
Model3.7.pvals<-pvals.fnc(Model3.7, nsim = 10000, withMCMC =
  TRUE)
```

These commands first load the two libraries we need. We then create an object that contains the *p* values and confidence intervals for the various terms in Model 3.7 in the object `Model3.7.pvals`. The actual function that we use is `pvals.fnc`, which is part of the `languageR` library. In turn, this function calls the `mcmcsamp` function from the `coda` library. Three elements are included in this function call, including the name of the `lmer` object that contains the model fit results (`Model3.7`), the number of simulated data sets we want to sample by using MCMC (`nsim`), and whether we want results of each of these 10000 MCMC draws to be saved (`withMCMC = TRUE`). Setting this last condition to `TRUE` is not necessary, as we are interested only in summary statistics. We can obtain the relevant information for the fixed and random portions of the model by typing the following commands.

```
Model3.7.pvals$fixed
```

```
             Estimate  MCMCmean  HPD95lower  HPD95upper   pMCMC  Pr(>|t|)
(Intercept)   2.0233    2.0218      1.9243       2.118  0.0001         0
gevocab       0.5129    0.5134      0.4966       0.530  0.0001         0
```

```
Model3.7.pvals$random

  Groups       Name Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1 school (Intercept)   0.3159     0.3065   0.3074      0.2532      0.3637
2 Residual             1.9407     1.9413   1.9413      1.9134      1.9665
```

From these results, we can determine that the vocabulary score was statistically significantly related to the reading score, and that the random effects `school` and `Residual`, were both different from 0 as well, since neither of their confidence intervals included 0.

Returning to model definition using `lmer()`, multiple predictors at any level and interactions between predictors at any level are again entered in the model in the same manner as using the `lm()` or `lme()` functions. The following is the syntax for fitting Model 3.8 using `lmer`.

```
Model3.8 <- lmer(geread~gevocab + senroll +(1|school), data =
                 Achieve)
Model3.8
Linear mixed model fit by REML
Formula: geread ~ gevocab + senroll + (1 | school)
   Data: Achieve
   AIC   BIC logLik   deviance REMLdev
 43162 43198 -21576      43124   43152
Random effects:
  Groups   Name        Variance  Std.Dev.
  school   (Intercept)  0.10034   0.31676
  Residual              3.76655   1.94076
Number of obs: 10320, groups: school, 160

Fixed effects:
              Estimate    Std. Error   t value
(Intercept)  2.0748764     0.1139915     18.20
gevocab      0.5128742     0.0083733     61.25
senroll     -0.0001026     0.0002051     -0.50

Correlation of Fixed Effects:
        (Intr) gevocb
gevocab -0.327
senroll -0.901 -0.002

Model3.8.pvals<-pvals.fnc(Model3.8, nsim = 10000, withMCMC =
                          TRUE)
Model3.8.pvals$fixed

            Estimate  MCMCmean  HPD95lower  HPD95upper   pMCMC  Pr(>|t|)
(Intercept)   2.0749    2.0752      1.8493      2.2950  0.0001    0.0000
gevocab       0.5129    0.5133      0.4970      0.5295  0.0001    0.0000
senroll      -0.0001   -0.0001     -0.0005      0.0003  0.5960    0.6169
```

```
Model3.8.pvals$random

  Groups        Name  Std.Dev. MCMCmedian MCMCmean HPD95lower HPD95upper
1 school (Intercept)    0.3168     0.3076   0.3085     0.2501     0.3633
2 Residual              1.9408     1.9415   1.9415     1.9140     1.9673
```

3.3.2 Random Coefficient Models Using `lme4`

The definition of random effects for slopes in `lme4` is very similar to that in `nlme`. The only real difference is that again, as in the random intercepts model, the random effects are defined in parentheses as a linear combination of effects. Returning to Model 3.3, we may express the same multilevel model using `lmer` as:

```
Model3.9 <- lmer(geread~gevocab + (gevocab|school), data =
  Achieve)
Model3.9
Linear mixed model fit by REML
Formula: geread ~ gevocab + (gevocab | school)
   Data: Achieve
   AIC     BIC  logLik  deviance   REMLdev
   43005 43048  -21496     42981     42993
Random effects:
  Groups   Name         Variance    Std.Dev.    Corr
  school   (Intercept)  0.282692    0.53169
           gevocab      0.019305    0.13894    -0.859
Residual                3.665937    1.91466
Number of obs: 10320, groups: school, 160

Fixed effects:
            Estimate    Std. Error    t value
(Intercept)  2.00570       0.06109      32.83
gevocab      0.52036       0.01442      36.09

Correlation of Fixed Effects:
        (Intr)
gevocab -0.867
```

We must note here that the MCMC approach for obtaining hypothesis test results for models estimated using `lmer` is not currently available for random coefficient models.

Although, for the most part, the syntax of `lme4` is fairly similar to that of `lme` for relatively simple models, incorporating multiple random slopes into multilevel models using `lme4` is somewhat different. The random effects discussed for the `nlme` package assume correlated or nested levels. Random effects in `lme4` may be either correlated or uncorrelated. In this respect, `lme4` provides greater modeling flexibility. This difference in model specification

is communicated through a different model syntax. As an example, refer to Models 3.10 and 3.11, each of which has the same fixed and random effects. However, the random slopes in Model 3.10 are treated as correlated with one another; in Model 3.11, they are specified as uncorrelated. This lack of correlation in Model 3.11 is expressed by having separate random effect terms (`gevocab|school`) and (`age|school`). In contrast, Model 3.10 includes both random effects in a single term (`gevocab + age|school`).

```
Model3.10 <- lmer(geread~gevocab + age+(gevocab + age|school),
                  Achieve)

Model3.11 <- lmer(geread~gevocab + age+ (gevocab|school) +
                  age|school), Achieve)
```

```
Model3.10
Linear mixed model fit by REML
Formula: geread ~ gevocab + age + (gevocab + age | school)
   Data: Achieve
   AIC   BIC  logLik   deviance   REMLdev
 43015 43088  -21498      42974     42995
Random effects:
 Groups    Name         Variance     Std.Dev.    Corr
 school    (Intercept)  1.8361e-02   0.135503
           gevocab      1.9026e-02   0.137936     0.465
           age          2.4641e-05   0.004964    -0.197    -0.960
 Residual               3.6641e+00   1.914182
Number of obs: 10320, groups: school, 160

Fixed effects:
             Estimate    Std. Error     t value
(Intercept)   2.965272      0.413052        7.18
gevocab       0.519278      0.014351       36.18
age          -0.008881      0.003822       -2.32

Correlation of Fixed Effects:
          (Intr)     gevocb
gevocab   -0.081
age       -0.989     -0.047

Model3.11
Linear mixed model fit by REML
Formula: geread ~ gevocab + age + (gevocab | school) + (age |
         school)
   Data: Achieve
   AIC   BIC  logLik   deviance   REMLdev
 43017 43089  -21498      42975     42997
Random effects:
 Groups    Name        Variance     Std.Dev.     Corr
 school    (Intercept) 2.1436e-01   0.46299441
           gevocab     1.9194e-02   0.13854364   -0.976
```

```
 school   (Intercept) 2.2262e-02  0.14920466
          age         8.8027e-07  0.00093822    1.000
 Residual             3.6649e+00  1.91439622
Number of obs: 10320, groups: school, 160

Fixed effects:
             Estimate  Std. Error   t value
(Intercept)  2.973619    0.414551     7.17
gevocab      0.519191    0.014397    36.06
age         -0.008956    0.003798    -2.36

Correlation of Fixed Effects:
         (Intr)   gevocb
gevocab  -0.159
age      -0.989    0.033
```

Notice the difference in how random effects are expressed in `lmer` between Models 3.10 and 3.11. Output in Model 3.10 provides identical estimates to those of the `nlme` Model 3.4. With random effects, R reports estimates for the variability of the random intercept, variability for each random slope, and the correlations between the random intercept and random slopes. Output in Model 3.11, however, reports two different sets of uncorrelated random effects.

The first set reports variability for the random intercept and variability for the random slope for vocabulary and correlation between the random intercept and random slope for vocabulary. The second set of random effects reports variability of a second random intercept, variability in the random slope for age, and the correlation between the random intercept and the random slope for age. The random slope for vocabulary and the random slope for age are not allowed to correlate. Finally, we can obtain *p* values and confidence intervals for each model term using the `pvals.fnc` function based on the MCMC approach reviewed earlier in this chapter.

3.4 Additional Options

R provides several additional options for applying multilevel models through both the `nlme` and `lme4` packages.

3.4.1 Parameter Estimation Method

Both `nlme` and `lme4` by default use restricted maximum likelihood (REML) estimation. However, each package also allows use of maximum likelihood (ML) estimation instead. Model 3.12 demonstrates syntax for fitting a multilevel model using ML in the `nlme` package. To change the estimation

method in `nlme`, the call is `method = "ML"`. Model 3.13 depicts fitting of the same multilevel model using the `lme4` package. The call to designate the use of the ML to be used is `REML = FALSE`.

```
Model3.12 <- lme(fixed = geread~gevocab, random = ~1|school,
                 data = Achieve, method = "ML")

Model3.13 <- lmer(geread~gevocab + (1|school), data = Achieve,
                  REML = FALSE)
```

3.4.2 Estimation Controls

Sometimes a correctly specified model will not reach a solution (converge) in the default settings for model convergence. This problem often can be fixed by changing the default estimation controls using the `control` option. Convergence issues can be fixed frequently by changing the model iteration limit (`maxIter`) or by changing the model optimizer (`opt`). To specify which controls will be changed, R must be given a list of controls and their new values. For example, `control = list(maxIter = 100, opt = "optim")` will change the maximum number of iterations to 100 and the optimizer to *optim*. These control options are placed in the R code in the same manner as choice of estimation method (separated from the rest of the syntax by a comma). They are the same for both the `nlme` and `lme4` packages. See Models 3.14 and 3.15 below. A comprehensive list of estimation controls can be found on the R `help ?lme and ?lme4` pages.

```
Model3.14 <- lme(fixed = geread~gevocab, random = ~1|school,
                 data = Achieve, method = "ML", control =
                 list(maxIter = 100, opt = "optim"))

Model3.15 <- lmer(geread~gevocab + (1|school), data = Achieve,
                  REML = FALSE, control = list(maxIter = 100,
                  opt = "optim"))
```

3.4.3 Chi Square Test for Comparing Model Fit

We previously explained how the fits of various models can be compared using the AIC and BIC information indices. However, these statistics are descriptive in nature so that no hypotheses about relative model fit can be tested formally. Thus, if the AIC for one model is 1000.5 and 999 for another models, we cannot know whether the apparently small difference in fit within the sample is truly representative of a difference in fit in the general population. Therefore, when we work with nested models and one model is a more constrained (i.e., simpler) version of another, we may wish to test whether overall fit of the two models differs. Such hypothesis testing is possible using the chi-square difference test based on the deviance statistic. When the fits of nested models are compared, the difference in chi-square

values for each model deviance can be used to compare model fit. After each of the models in question has been fit, the difference in chi-square values can be obtained using the `anova()` function call.

For models run using the `nlme` package, the `anova()` command will provide accurate comparisons only if maximum likelihood estimation is used. For models run using `lme4`, the `anova()` command will work for both maximum likelihood and restricted maximum likelihood. When maximum likelihood is used, both fixed and random effects are compared simultaneously. When restricted maximum likelihood is used, only random effects are compared. The following is an example of comparing fit with the chi-square difference statistic for Models 3.1 and 3.2 that were discussed in detail above.

```
Model3.1 <- lme(fixed = geread~gevocab, random = ~1|school,
                data = Achieve, method = "ML")

Model3.2 <- lme(fixed = geread~gevocab + senroll, random =
                ~1|school, data = Achieve, method = "ML")

anova(Model3.1, Model3.2)

anova(Model3.1 Model3.2)

Model3.1 1
4 43132.43 43161.40 -21562.22
Model3.2 2 5 43134.18 43170.39 -21562.09 1 vs 2 0.2550617
  0.6135
```

3.4.4 Confidence Intervals for Parameter Estimates

Readers who are familiar with multilevel modeling may have noticed that neither `nlme` nor `lme4` output provides statistical significance tests for the variance of random effects. As outlined in Chapter 2, statistical significance of random effects provides very useful information about the variability of the clusters under study. Using the example from this chapter, the significance of the random intercept indicates variations in reading ability among schools in the sample; i.e., different schools exhibit significantly different mean reading scores. Similarly, a significant random slope for vocabulary would indicate significant variation in the impact of vocabulary on reading ability across the schools. This is often very useful information by providing insights into the factors that contribute to score differences. However, the current packages do not provide an option for testing the significance of random effects.

It is still possible, however, to obtain information about significance of random effects by creating confidence intervals. With the `nlme` package, the function call `intervals()` can be used to generate 95% confidence intervals for the fixed effects and the variances of the random effects. The confidence intervals obtained for the variances of the random effects can

be used to determine the significance of the random effects. For example, returning to Model 3.3 covered earlier in this chapter, we determined that vocabulary was a significant predictor of reading ability. However, we could not determine from the output of Model 3.3 whether the variability in the random intercept or random slope was significantly different from 0. If not different, the result would indicate that the mean reading achievement and/or the relationship of vocabulary score to reading achievement did not differ across schools. To determine the significance of the random effects we can use the `intervals()` function call.

```
intervals(Model3.3)

  Approximate 95% confidence intervals

Fixed effects:
                 lower          est.         upper
(Intercept)  1.8859621     2.0057064     2.1254506
gevocab      0.4920982     0.5203554     0.5486126
attr(,"label")
[1] "Fixed effects:"

Random Effects:
  Level: school
                                lower          est.         upper
sd((Intercept))            0.4250700     0.5316531     0.6649611
sd(gevocab)                0.1153701     0.1389443     0.1673356
cor((Intercept),gevocab)  -0.9178709    -0.8585096    -0.7615768

Within-group standard error:
   lower     est.    upper
1.888327 1.914663 1.941365
```

For the intercept, the 95% confidence interval lies between 0.425 and 0.665. Thus, we are 95% confident that the actual variance component for the intercept was between these two values. Likewise, the 95% confidence interval for the random slope variance was between 0.115 and 0.167. From these values, we can see that 0 did not lie in the interval for either random effect, intercept, or slope. Thus, we can conclude that both the random intercept and random slope were significantly different from 0.

Summary

This chapter put to work the concepts learned in Chapter 2 to work using R. We learned the basics of fitting two-level models when a dependent variable is continuous using the `lme` and `lmer` packages. Within this multilevel

framework, we learned how to fit the null, random intercept, and random slopes models. We also covered independent variables at both levels of data and learned how to compare the fits of models with one another. This last point will prove particularly useful as we engage in the process of selecting the most parsimonious (simplest) model that also explains the dependent variable adequately. Of greatest import in this chapter, however, is the ability to fit multilevel models using both `lme` and `lme4` in R and correctly interpreting the resultant output. If you have mastered those skills, you are ready to move to Chapter 4, where we extend the model to include a third level in the hierarchy. As we will see, the actual fitting of three-level models is very similar to fitting two-level models studied in the chapter.