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

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

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