



# INTRODUCTION

## 1.1 WHY BOTHER WITH MULTILEVEL MODELING?

---

Multilevel modeling (MLM) has grown in popularity over the past 30 years. MLM is known by various names across different disciplines. MLM is also known as hierarchical linear modeling (HLM), mixed/random effects modeling, variance components modeling, and random coefficient modeling. For models with noncontinuous (e.g., binary) outcomes, models have also been referred to as generalized linear mixed models (GLMMs) or hierarchical generalized linear models (HGLMs).

MLM is an extension of multiple regression—so a solid understanding of multiple regression is needed in order to make sense of MLM concepts. MLM can be used to study clustered data which are common in fields such as education (e.g., students within schools), biostatistics (e.g., patients within hospitals), business management (e.g., employees within companies), and political science (e.g., voters within districts). Multilevel data structures are quite common and can be described as having lower-level units (level-1 units; e.g., students) nested within higher-level groups or clusters (level-2 units; e.g., schools). Clustered data structures are also present in longitudinal studies where outcomes over time are nested within individuals (repeated measures). In this case, measurements are clustered within individuals (e.g., collecting a standardized test measure from the same students at the start, middle, and end of the school year). Clustered data are commonly referred to as nested, hierarchical, or dependent (as opposed to independent) data.

When analyzing clustered data, there are statistical and theoretical reasons for using MLM. MLM can be used for the analysis of clustered data in order to arrive at the correct inferences/conclusions based on the data and/or when there is an interest in the effect of group-level variables on some outcome. Not properly accounting for the nested data structure, through the use of MLM or some alternative procedure, can result in researchers making incorrect study conclusions based on highly misleading results.

When using ordinary least squares (OLS) regression, observations are assumed to be independent of each other (e.g., knowing information about one individual does not provide you information about another individual). However, with clustered data, this assumption of observation independence is violated. Units (e.g., students) within groups or clusters (e.g., schools) tend to be more similar with each other compared to units in other groups. For example, students in the same school may share the same teachers, resources, or curricula used compared to students in other schools. If a researcher is interested in how certain individual-level predictors (e.g., motivation, interest in the topic) are related to some outcome (e.g., reading scores) but fails to account for factors at the group level (e.g., presence of motivated peers, more engaging teachers), then conclusions drawn from such analyses may be questionable. Not accounting for clustering can seriously affect model results and MLM addresses this important assumption violation of observation independence.

When analyzing a data structure represented by individuals within groups, a researcher may be interested in group-level effects, individual-level effects, or both. This raises the question of what is the unit of analysis. A frequently cited study that brought up the unit of analysis problem comes from Robinson (1950). Robinson's study has been called "profoundly influential" (Subramanian, Jones, Kaddour, & Krieger, 2009, p. 354) and Grotenhuis, Eisinga, and Subramanian (2011, p. 1123) referred to it as an "an all-time classic and one of the most influential methodological papers."

In the study, Robinson analyzed US Census data from the 1930s<sup>1</sup> and, using individual-level data, computed the correlation coefficient between an individual's foreign-born status (referred to as nativity) and illiteracy where  $r = .12$  (p. 354). In the study, illiteracy was defined by the US Census as the inability to read and write for those aged 10 years and older. However, when data were

---

<sup>1</sup>Remember, the data are almost 100 years old and come from a different era in the nation's history. On their website (<https://www.ru.nl/sociology/mt/rob/downloads/>), Grotenhuis et al. have also provided the data, with some corrections, to allow for the reanalysis of Robinson's original paper.

aggregated by state, the correlation between the percent of foreign-born individuals with the percent illiterate in the state was negative ( $r = -.53$ ) showing the opposite relationship. The negative correlation indicated that as the percent of foreign-born individuals increased, illiteracy rates dropped, which was the opposite conclusion resulting from the analysis of individual-level data. Robinson concluded that the group-level correlation (known as the **ecological correlation**) did not have to be the same as the individual-level correlation.

Generalizing group-level effects to the individual has come to be known as an **ecological fallacy** and generalizing from the individual-level effect to the group is known as the **atomistic fallacy**. The result of these differing relationships across levels is that regression coefficients (if clustering is not accounted for) can be biased, which is another way of saying that the coefficients can be too high or too low, on average. When using MLM, we can investigate separate cluster-level and unit-level effects appropriately.

Another commonly cited statistical reason for using MLM has to do with the precision of the regression coefficients estimated. Each regression coefficient has a corresponding standard error which represents the amount of variability in the estimated coefficient. The less variability associated with the regression coefficient makes the estimate more precise (i.e., there is less error). If standard errors are estimated too low (i.e., they are smaller than they should be), we get a false sense of precision possibly resulting in a **Type I** error where statistical significance is found when they should not be.

The standard  $t$ -tests used to check for statistical significance are computed by taking the estimate and dividing it by the standard error. Generally, when the  $t$  statistic is above 2 (or more specifically 1.96 for large samples), results are referred to as statistically significant ( $p < .05$ ). Standard errors are influenced by the sample size ( $n$ ) and as  $n$  increases, standard errors decrease, making it easier to find statistical significance (which is a reason you may often hear that with a large enough sample size, everything will be statistically significant, even if effects are trivial). The proper estimation of standard errors is not a mere technical issue but can result in supporting or overturning study conclusions. For example, in a review of 33 studies that used nested datasets but did not account for the hierarchical nature of the data, 19 of the studies no longer had statistically significant results after adjustments for clustering were made (Baldwin, Murray, & Shadish, 2005).

With clustered datasets, higher- or group-level variables are repeated for all individuals in the same group. For example, the first four and last four observations of the often-used High School & Beyond (HSB) dataset of 7,185 students in 160 schools are shown (see Table 1.1). Notice that students in the same school (e.g., students in `school #1224` or `school #9586`) have data from school-level variables repeating, once per student. In the example, the `meanses` (mean socioeconomic status [SES] at the school) and `sector` (Public or Catholic school) are school-level variables and repeat for those students in the same school.

If OLS regression is used to run a model predicting math achievement (`mAch`) using the group-level predictors, the software will think there are 7,185 values for the predictors (never mind that they are just repeating) when these actually only come from 160 schools. As a result, standard errors are much lower than they should be (because of the large  $n$ ), and this is one reason we often hear that ignoring clustering raises the risk of finding statistically significant results for group-level variables when they should not have been.

Aside from the statistical reasons, researchers may be interested in a particular theory how variables at the higher-level may affect individual-level outcomes. Much research has discussed how individual outcomes can be shaped by the different environments an individual is part of and also how the micro- (individual) and macro- (group level) characteristics may interact with each other (Bronfenbrenner, 1992). For example:

**TABLE 1.1** First and Last Four Observations From the High School and Beyond Dataset

school	minority	sex	ses	mAch	meanses	sector	cses
1224	No	Female	-1.53	5.88	-0.43	Public	-1.09
1224	No	Female	-0.59	19.71	-0.43	Public	-0.15
1224	No	Male	-0.53	20.35	-0.43	Public	-0.09
1224	No	Male	-0.67	8.78	-0.43	Public	-0.23
			...	...	...		...
9586	No	Female	-0.04	14.79	0.62	Catholic	-0.66
9586	No	Female	1.33	19.64	0.62	Catholic	0.71
9586	No	Female	-0.01	16.24	0.62	Catholic	-0.63
9586	No	Female	0.79	22.73	0.62	Catholic	0.17

- Does the type of school (a group-level variable) attended (e.g., public or private) and group-level SES (i.e., the school average of the individual-level SES) have a relationship with student mathematics achievement (Raudenbush & Bryk, 2002; Singer, 1998)?
- Does the quality of leadership climate and group consensus affect individual morale and depression (Bliese & Britt, 2001)?
- Is neighborhood-level violent crime related to individual mental health outcomes (Goldman-Mellor, Margerison-Zilko, Allen, & Cerda, 2016)?

Although several reasons for using MLM have been discussed, on many occasions, all of these reasons may come together as justification to use MLM to account for the nesting effect. For example, a standard experimental design has individuals randomly assigned to a treatment or a control condition. However, a group or **cluster randomized controlled trial** (CRT) is a group-based version of an experiment where the unit of randomization is the group or cluster. All units (e.g., students or patients) within the same cluster (e.g., school or hospital) will then receive the treatment or intervention. After administering the treatment, outcomes of all the participants are compared to see if there are differences as a result of the intervention.

An illustration of a CRT involves researchers testing if marketing vegetables in school cafeterias can lead to healthier eating habits of children (Hanks, Just, & Brumberg, 2016). An intervention involved placing branded vegetable characters in a school cafeteria, and in the experiment, schools were randomly assigned to the intervention or control condition. The outcome was the number of children taking vegetables at the school salad bar over some time period. In such an instance, the intervention was administered at the cluster level, and all students in the same school experienced only either the intervention or control (i.e., business-as-usual) condition. In this case, intervention status was a group-level variable and individuals in the same school were all in the same condition. Not properly accounting for the clustering effect has been shown to result in questionable claims of intervention efficacy, often resulting in claims that an intervention worked when statistical evidence was actually lacking (Musca et al., 2011).

## 1.2 WHY ANOTHER MLM BOOK?

---

There is no shortage of excellent books on MLM. However, books can differ in their accessibility and technical complexity. At times, substantive researchers

(already familiar with multiple regression) may want to know how to quickly get up and run using MLM for the analysis of clustered data.

Multilevel models can be confusing given the terms and the notation used. However, readers can be guided through the concepts and steps necessary to fit their models. As such, this book is devoid of mathematical proofs and derivations, light on equations, and is written with the applied researcher in mind. Practical examples are included to help illustrate concepts and are used throughout this book. Some equations are provided but only requiring basic algebra to understand. The book's goal is not to turn readers into master multilevel modelers, but to help readers understand the different concepts related to MLM, which in turn will be necessary if (or when) more advanced topics are pursued.

In addition, the focus of the book is on using R (R Core Team, 2020) specifically to fit multilevel models. There are several excellent software packages, some specialized and some general purpose, that can fit multilevel models, and for a comparison of statistical software/packages capable of conducting MLM analyses, readers can consult McCoach et al. (2018). However, using R for MLM is nice for several reasons.

First, R is available as a free download from <http://cran.r-project.org/>. The add-ons (referred to as packages) needed to fit multilevel models are free as well. Commercial MLM software can cost hundreds of dollars as a one-time or recurring cost.

Second, R is available for a variety of computing platforms (PC, Mac, Linux). I use R on a Chromebook using Linux, and R can be run from a web browser at <http://rstudio.cloud>.

Finally, and importantly as well, all data management, graphical visualization, modeling, and diagnostics can be performed using the same software. This means that there is no need to use multiple programs to handle each specific task, and there is no need to switch back and forth between programs.

R may have a bit of a learning curve. However, R syntax for performing the analyses are provided, allowing readers the ability to replicate results shown by simply copying and pasting (or typing in) the syntax.

## 1.3 USING R FOR MLM

---

Throughout the book, R code (and output) will be shown using `monospace` fonts which can be entered directly in the R console to perform the different

tasks. A function in R is a command or a set of instructions to perform certain tasks. R functions begin with text (never numbers), will not include spaces, and are followed by parentheses (these are important). Examples of basic functions include `mean()` (to compute the mean), `sd()` (to compute the standard deviation), or `lm()` (to fit a linear model).

We will rely on several packages throughout this book that extend base R's (i.e., R without additional packages) features. A **package** is a set of functions (i.e., bundled code that can be run to perform some calculation or output results) and/or datasets that can extend the functionality of R. Requested packages are downloaded automatically by R from a repository when connected to the internet (if not using R Studio, users may be prompted which server to download packages from; just pick the one geographically closest to you).

Packages in R must be installed once (on each computer that you use) by entering `install.packages("packagename")` in the R console and pressing Enter (to immediately execute the function) where `packagename` is the name of the package to be installed. For example: `install.packages("mlmRev")` installs the `mlmRev` package where `mlmRev` (Bates, Maechler, & Bolker, 2014) is a package that contains several clustered datasets that will be used throughout the book. Don't forget to include the quotes. Single or double quotes can be used but must be consistent, so indicating `'mlmRev'` is the same as indicating `"mlmRev"` but `"mlmRev'` or `'mlmRev"` will not work. Of importance is that R syntax is case sensitive (like Stata but unlike SAS) so take note of how functions/objects/variables are named. In this example, entering `mlmrev` or `MLMREV` is not the same as `mlmRev`.

Then, once within each R session, the package needs to be loaded using the `library` function using `library(packagename)`: e.g., `library(mlmRev)`. A **session** is an instance where R is run on your computer and the session ends when you quit or exit R. Unlike the `install.packages` function, the quotes are not required. Once loaded, the additional functions and/or datasets in the package will be available to the user. If a package is installed *but* not loaded using the `library` function, the added functions will not work and an error will be returned. Although the terms `package` and `library` are often used interchangeably, they are not the same: the `library` is where the files related to the package are stored on your computer and the `library` function is used to load the package in an R session.

There are several excellent packages that can be used to fit multilevel models using R. Examples of such packages include:

- `nlme` (Pinheiro, Bates, & R Core Team, 2022),
- `WeMix` (Bailey, Kelley, Nguyen, Huo, & Kjeldsen, 2020), or
- `GLMMadaptive` (Rizopoulos, 2020).

For this book, we will focus on the popularly used `lme4` package (Bates, Mächler, Bolker, & Walker, 2015). The main functions used in `lme4` are `lmer` for linear mixed effects models and `glmer` for generalized linear mixed effects models.

Of note is that the authors of the `lmer` (pronounced “elmer”) function have purposely not included  $p$  values in the output. Many researchers though may want to report the  $p$  values for their regression coefficients, and there are several alternative approaches (discussed in Chapter 3) for obtaining these  $p$  values (Luke, 2017). A simple, straightforward, and acceptable way though is to install and use the `lmerTest` package (Kuznetsova, Brockhoff, & Christensen, 2017) which we will use throughout this book. Once installed, users only need to load the package using `library(lmerTest)` (without having to load the `lme4` package) which provides an alternate `lmer` function. The output of the updated `lmer` function will automatically show the  $p$  values. If you are viewing the output of your `lmer` model and do not see  $p$  values, that is likely because the `lmerTest` library was not installed and/or loaded.

In addition, several other packages will be used:

1. The companion R package to this book, `MLMusingR`, contains some datasets (referred to as data frames in R) and functions that will be used throughout. This can be installed using `install.packages("MLMusingR")`.
2. Additional nested datasets can be obtained using the `mlmRev` package (Bates et al., 2014). `mlmRev` contains over a dozen clustered datasets (some of which will be used in certain chapters) that have been commonly used in the MLM literature as part of tutorials.
3. For basic data management (e.g., selecting variables from a dataset, sorting observations), I use the `dplyr` package (Wickham, François,



Henry, & Müller, 2020). Data management becomes more important when we focus on how to reshape data (Chapter 5), conduct model diagnostics (Chapter 7), or investigate missing data (Chapter 10).

In some cases where a single function is used in a package, I may refer to the function using the `packagename::function()` syntax which can be used to access a function directly without having to load the package using the `library` function (though the package still needs to be installed first). For example, I may indicate: `performance::icc(model1)` which computes the intraclass correlation coefficient (ICC; to be discussed in the next chapter) of a variable using the `icc` function found in the `performance` package (Lüdtke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021). Other packages and functions will be discussed when used in the book.

To install all R packages that are used in this book, the following syntax can be run (this only has to be run once per computer):

```
install.packages(c('car', 'clubSandwich', 'dplyr', 'estimatr', 'geepack',
'influence.ME', 'jomo', 'jtools', 'lme4', 'lmerTest', 'lmtest', 'mice',
'mitml', 'mlmRev', 'MLMusingR', 'naniar', 'performance', 'PowerUpR',
'psych', 'sandwich', 'sjmisc', 'tidyr'))
```

Do not copy, post, or distribute