2

THE UNCONDITIONAL MEANS MODEL

LEARNING OBJECTIVES

After reading and studying this chapter, students should be able to:

- Describe multilevel models using notation.
- Fit an unconditional multilevel model using `lmer`.
- Calculate the unconditional intraclass correlation coefficient (ICC).
- Explain the relevance of the ICC.

2.1 UNDERSTANDING MLM NOTATION

Before fitting multilevel models using software, it is important to be able to express the regression model using some form of notation. Understanding notation is important as authors will present models using these symbols and readers will undoubtedly come across these equations in manuscripts. Using notation also helps communicate what exactly is being modeled.

There are at least three systems of notation used when discussing multilevel models: (1) level, (2) composite, and (3) matrix notation (Scott, Shrout, & Weinberg, 2013). I stick to level notation, which was popularized by Raudenbush and Bryk (2002). Using level notation, formulas from different levels can be combined resulting in a type of composite notation. The advantage of level notation is that readers can explicitly see the models at different levels and is easily combined. Level notation is used in software such as HLM (Raudenbush, Bryk, & Congdon, 2013), and composite/combined notation is most similar in structure to code used in SAS, Stata, or R (Scott et al., 2013). Note however that even if authors use the same system of notation (e.g., level notation), the symbols and subscripts used may differ.

In a simple, single-level regression formula, readers should be familiar with a regression equation expressed as: $Y = \beta_0 + \beta_1 X_1 + r$ where Y is the outcome predicted by the intercept (β_0), plus the regression coefficient β_1 times the values of the predictor X_1, plus some random error, r. Capital, italicized Roman letters (e.g., Y, X) represent the observed variables in the dataset. With MLM, additional subscripts are added to more clearly represent that the outcome is from unit i in cluster j. For example, to represent the outcome for student i in school j, the dependent variable is written as Y_{ij}.

Level notation is a system of equations used to represent the model at different levels. In a two-level model, level-1 units belong to a higher level-2 group. There is a single level-1 formula where the coefficients are represented by βs ("betas") with subscripts representing 0 for the intercept, 1 for the first coefficient, 2 for the second, and so on.

There are different strategies one can take when constructing multilevel models. For practical and pedagogical purposes, I begin with the simplest model and will work up to the more complex models. Often, when running multilevel analyses, the first step is to fit an **unconditional means** or **null model** which is a model with no predictors (which might sound strange). The model is referred to as unconditional as the results are not *conditioned* on (or do not depend on) any predictors. Fitting the null model first is not universally done, but a majority of studies do so (Luo et al., 2021) and many journal reviewers/editors (especially in education and psychology) will look for the results from this model.

The focus of the null model is on the *variance* of the outcome and there are two reasons for using a null model: (1) to partition the variance of the outcome at the different levels; and (2) to establish baseline levels of variance that can be used in succeeding models for computing (pseudo) R^2 measures. A null model can be written as:

$$\begin{aligned} \text{Level 1:} \quad Y_{ij} &= \beta_{0j} + r_{ij} \\ \text{Level 2:} \quad \beta_{0j} &= \gamma_{00} + u_{0j} \end{aligned}$$

At level one, β_{0j} ("beta zero *j*" sometimes referred to as "beta naught *j*") represents the level-1 intercept for group *j*. The term r_{ij} is the residual for unit *i* in group *j* and represents the individual difference from the mean of group *j*. At level two, γ_{00} ("gamma zero zero") is the intercept which represents the overall grand mean of the outcome and u_{0j} represents group *j*s deviation from the overall mean. The u_{0j} at level two indicates that the intercept (γ_{00}) can randomly vary by group. The lowercase, italicized Roman letters (e.g., r_{ij}, u_{0j}) represent the residuals at each level of the model.

In a multilevel model, even in a model with no predictors, there are at least two sets of residuals, one at level one (i.e., the unit level), and at least one at level two (i.e., the cluster/group level). If the level-2 equation for β_{0j} is substituted in the level-1 equation, the resulting combined formula is: $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$. Readers familiar with ANOVA formulas will recognize this as a one-way random effects ANOVA.

When using OLS regression, the residuals are often not of interest aside from their use when conducting regression diagnostics. However, with a multilevel model, the variability of these residuals is of interest. Once variance at the different levels is known, we can determine how much of the variability can be attributed to the cluster/group level and the individual level.

The level-2 residuals are assumed to be normally distributed with a mean of 0 and a variance of τ_{00} (referred to as "tau zero zero"); written as $u_{0j} \sim \mathcal{N}(0, \tau_{00})$. In addition, the level-1 residuals are also assumed to be normally distributed with a mean of 0 with a variance of σ^2 ("sigma squared"); $r_{ij} \sim \mathcal{N}(0, \sigma^2)$.[1] Both u_{0j} and r_{ij} are referred to as *random effects* while γ_{00} is referred to as a *fixed effect*. In a two-level model, fixed effects refer to the regression coefficients and are represented by the Greek letters of β and γ. The random effects refer to the residuals and are represented by italicized, lowercase Roman letters. **Residual variance** is the variability that is not explained by the fixed effects (or the predictors in the model). The residual variance in a multilevel model can be partitioned between the variance resulting from group and individual differences.

[1]When using MLM, τ_{00} and σ^2 are estimated, not the residuals themselves.

A NOTE ON NOTATION

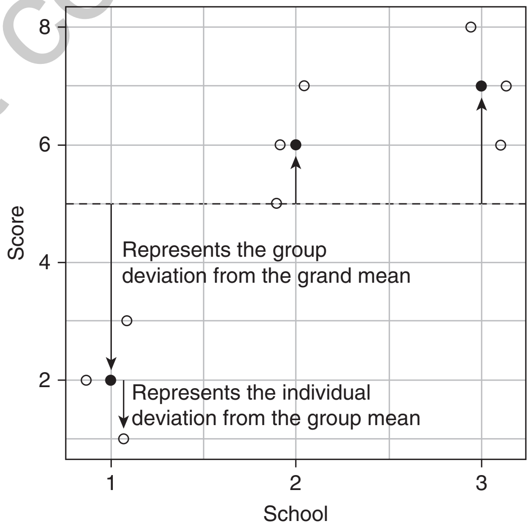
For this book, I adopt the convention of Raudenbush and Bryk (2002) and use τ_{00} to represent the variance of the level-2 intercept. Other authors may use different notation to represent the same thing such as σ^2_{school}, σ^2_{u0}, or τ^2.

At times, the level-1 residual term may also be shown using ε_{ij}, ϵ_{ij}, or e_{ij}. The differing types of notation used can be a source of confusion, so it is important to recognize what these symbols are referring to.

The inclusion of u_{0j} is what indicates that this is a multilevel model. If the level-2 formula were written as $\beta_{0j} = \gamma_{00}$ instead (i.e., no u_{0j}), the combined model would be: $Y_{ij} = \gamma_{00} + r_{ij}$ except that in this case, r_{ij} would represent the residual of the individual from the overall mean (regardless of group membership) represented by γ_{00}. The inclusion of u_{0j} in the formula shows that there is more than one source of random variability in the model.

To illustrate this graphically, using an educational example, we might have a dataset consisting of quiz scores from three students from three different schools. We show the outcomes for nine students from the three schools in Figure 2.1 using hollow circles (the points are jittered slightly so the points don't overlap). For simplicity, scores for the three students in School 1 are 1, 2, and 3; scores of students in School 2 are 5, 6, and 7; and scores of students in

FIGURE 2.1 ● Quiz Scores of Nine Students From Three Schools. Filled Circles Represent the Average per Group



School 3 are 6, 7, and 8. If the mean of all the scores are taken (without regard to group membership), this is simply referred to as the overall **grand mean** (see dashed line in Figure 2.1) or $\overline{X} = 5.0$. Each school will have its own average quiz scores per group (shown as filled circles at 2, 6, and 7) or **group means** which differ from the overall mean. These differences of the group means from the overall grand mean are deviations which represent variability in scores due to between-group differences. At the same time, there is variability in the outcomes within schools with some students scoring higher and some scoring lower than the school average.

2.2 FITTING AN UNCONDITIONAL/ NULL MODEL

For our first example, we can use the `Exam` dataset in the `mlmRev` package. The package must first be installed using `install.packages ("mlmRev")`.[2] If the package is already installed, load it by using `library(mlmRev)`. Afterward, the dataset in the package can be loaded into memory by using `data(Exam)`.

The `Exam` dataset consists of 4,059 students from 65 schools from London (Goldstein et al., 1993). To learn more about the dataset, entering `?Exam` in the R console is a quick way to get information about the dataset. As a reminder, R is case sensitive, so entering `Exam` is not the same as `exam`.

The R code shown also has comments/notes following the # symbol. Comments are not executed and are only included to provide additional notes for the reader and do not need to be entered.

```
library(mlmRev) #load in mlmRev package, must be installed first
data(Exam) #load the dataset
#an alternative way is to use data(Exam, package = 'mlmRev')
head(Exam) #inspect the first few observations in the dataset
  school  normexam  schgend  schavg      vr     intake  standLRT  sex  type  student
1      1    0.2613    mixed  0.1662  mid 50%  bottom 25%   0.6191    F   Mxd      143
2      1    0.1341    mixed  0.1662  mid 50%     mid 50%   0.2058    F   Mxd      145
3      1   -1.7239    mixed  0.1662  mid 50%     top 25%  -1.3646    M   Mxd      142
4      1    0.9676    mixed  0.1662  mid 50%     mid 50%   0.2058    F   Mxd      141
5      1    0.5443    mixed  0.1662  mid 50%     mid 50%   0.3711    F   Mxd      138
6      1    1.7349    mixed  0.1662  mid 50%  bottom 25%   2.1894    M   Mxd      155
```

[2]To see all datasets within a package, the following syntax can be entered in the R console: `data(package = "mlmRev")`.

```
str(Exam) #to view the internal structure of the dataset

'data.frame':  4059 obs. of  10 variables:
 $ school     : Factor w/ 65 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ normexam   : num  0.261 0.134 -1.724 0.968 0.544 ...
 $ schgend    : Factor w/ 3 levels "mixed","boys",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ schavg     : num  0.166 0.166 0.166 0.166 0.166 ...
 $ vr         : Factor w/ 3 levels "bottom 25%","mid 50%",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ intake     : Factor w/ 3 levels "bottom 25%","mid 50%",..: 1 2 3 2 2 1 3 2 2 3 ...
 $ standLRT   : num  0.619 0.206 -1.365 0.206 0.371 ...
 $ sex        : Factor w/ 2 levels "F","M": 1 1 2 1 1 2 2 2 1 2 ...
 $ type       : Factor w/ 2 levels "Mxd","Sngl": 1 1 1 1 1 1 1 1 1 1 ...
 $ student    : Factor w/ 650 levels "1","2","3","4",..: 143 145 142 141 138 155 158 115 117 113 ...
```

Think of functions as prespecified commands that you can enter in R to accomplish certain tasks. Functions can be applied to objects that are enclosed within parenthesis. Objects in R are generally data containers—the containers can consist of one value, a set of values, a dataset, or a collection of data.

The `head` function can be used to show the first few (i.e., 6) observations in the object named `Exam` (which happens to be a dataset). The `str` (structure) function gives a glimpse into the dataset showing what kinds of variables are contained (e.g., whether a variable is numeric or a factor). Our outcome of interest is the `normexam` variable which is a standardized exam score (i.e., it is a z-score). In a traditional multilevel model, the outcome is a variable at the lowest level of the hierarchy (i.e., at level one). Basic descriptives can be checked by using the `mean` and the `var` function to get the mean and variance of the `normexam` variable, respectively (which confirms that the outcome is standardized with a $M = 0$ and SD [and variance] $= 1$). We can also get the minimum and maximum value by using the `range` function. A variable can be directly accessed from a dataset by specifying the variable in the data frame separated by a `$` operator (e.g., `dataframe$variable`). In R, multiple datasets can be loaded into memory.

```
mean(Exam$normexam)
[1]  -0.0001138
var(Exam$normexam)
[1]  0.9979
range(Exam$normexam)
[1]  -3.666 3.666
```

The unconditional/null model can be fit to answer the question of how much variability can group-level effects account for (i.e., how much does the cluster account for in the variability of the outcome)? The model is also known as an **unconditional random intercept model**. In addition to variance that can be due to differences in student characteristics, the schools that the students attend may play a role in their performance. In other words, there may be some variability that is associated with school-level factors.

WHY BE CONCERNED ABOUT CLUSTER-LEVEL VARIABILITY?

The question of how much variability group-level factors account for has often been used as a way to assess the contribution of group- vs. individual-level variables. For example, in one of the earliest studies of educational effectiveness, the Equality of Educational Opportunity (EEO) or the Coleman et al. (1966) Report, which involved over 650,000 students and 4,000 schools, indicated that approximately 20% of the variability of student outcomes could be attributed to group-level factors. At the group level, the educational background and aspirations of fellow students were strong predictors of academic achievement. The report indicated that the most important variable related to a child's academic achievement was the students' family educational background, holding all variables constant. The strongest implication that the report made was that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (p. 325).

To fit the unconditional/null model and operationalize the composite formula we had indicated earlier, we will use the `lmer` function in the `lmerTest` package. The package must be loaded once per R session (if the user quits or restarts R, the package has to be loaded again for that session) by using `library(lmerTest)`. Using formula notation, we specify the outcome predicted by only the intercept (`normexam ~ 1`). The variable on the left-hand side of the tilde (~) is the outcome or dependent variable of interest. The tilde is shorthand for saying "is defined by" or "is predicted by". On the right-hand side of the tilde, we can include multiple predictors, covariates, or independent variables of interest joined by a + sign. But, being an unconditional model, we only include a `1` to indicate that there are no predictors aside from the intercept (which the `1` represents).

In addition, we specify the random effects or that the intercepts can randomly vary by the grouping/cluster variable by indicating `+ (1|school)`. Specifying the random effects term using `lmer` consists of two expressions within parenthesis separated by a vertical bar (|) which is read as "by." The left side of the | is a `1` which indicates that intercepts are allowed to randomly vary by group/cluster, which is indicated on the right side of the |. The cluster/grouping variable can be a factor, character, or a numeric variable. In this case, we can say that intercepts are allowed to randomly vary by school.

In R, additional options or arguments can be specified in a function by including the options separated by a comma. In this case, I specify which dataset to use by including `data = Exam` and save the output into an object named `nullmodel`. I use the `<-` operator which assigns the output to our `nullmodel` object (the = sign also does the same thing but `<-` is conventionally used). The choice of the name of the object is up to the researcher (the name though cannot start with a number or contain spaces). Putting this all together, we can fit the model by running the following syntax in the R console:

```
library(lmerTest) #needs to be installed first
nullmodel <- lmer(normexam ~ 1 + (1|school), data = Exam)
```

If all goes well, no errors will be reported and you have now fit your first multilevel model using R! If however some errors are reported, check your syntax to make sure there are no typos and that you have installed the `lmerTest` package. The model can also be specified by using: `nullmodel <- lmer(normexam ~ (1|school), data = Exam)`. By default, the intercept will be estimated in a regression model anyway. Afterward, results can be displayed by using the `summary` function on the created `nullmodel` object by running:

```
summary(nullmodel)

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: normexam ~ 1 + (1 | school)
  Data: Exam

REML criterion at convergence: 11015

Scaled residuals:
   Min      1Q  Median     3Q    Max
-3.947  -0.649   0.012  0.699  3.657


Random effects:
Groups    Name         Variance  Std.Dev.
school    (Intercept)  0.172     0.414
Residual               0.848     0.921
Number of obs: 4059, groups: school, 65

Fixed effects:
             Estimate  Std. Error       df  t value  Pr(>|t|)
(Intercept)   -0.0133      0.0541  62.5289    -0.25      0.81
```

Of interest at this point are the variance components under the `Random effects` section of the output. The `school` and `Residual` (under the `Groups` heading) variance from the random effects output refer to τ_{00} and σ^2, respectively. Note that the `Std.Dev.` column refers to the standard deviation which is just the square root of the variance. The variance is now partitioned to between (τ_{00}) and within groups (σ^2). Based on the output, $\tau_{00} = 0.172$ and $\sigma^2 = 0.848$. Readers should also check that the number of observations (4,059) and groups (65) are what is expected. Adding both the between- and within-group variance results in the total variance of 1.02.

The unconditional variance components represent baseline measures of variance which will be used in succeeding models to compute how much variance has been reduced with the addition of predictors to the model. When we add predictors to the model, we would expect variance to be reduced and the predictors are likely to account for some of the unexplained variance.

Although not really of interest at this point, the `Fixed effects` section shows the intercept or $\gamma_{00} = -0.01$. This merely shows the estimated average value of the respondents on `normexam` based on the model (i.e., the overall grand mean). The estimated value should be close to zero since the outcome was standardized to begin with (i.e., $M = 0$, $SD = 1$).

2.2.1 Computing the Intraclass Correlation Coefficient

Of primary interest in examining the random effects of an unconditional model is to obtain the **intraclass correlation coefficient (ICC)** or ρ ("rho"). This is also referred to in some texts as the variance partition coefficient or VPC. The ICC indicates the proportion of variance in the outcome that can be attributed to the grouping structure in the population.[3] The ICC has also been defined as a measure of the degree of similarity of units within the same cluster, and the ICC can be interpreted as the correlation in the outcome between two individuals randomly selected from the same group/cluster. The unconditional ICC estimated from the variance components for a model with no predictors is:

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{0.172}{0.172 + 0.848} = .169.$$

[3]In the measurement literature, ICC has been referred to as a measure of reliability. ICC has also been referred to as the Item Characteristic Curve. Note that both of these are different from what we are discussing here!

The ICC indicates that approximately 16.9% of the variance in the outcome can be attributed to school-level factors. Knowing how to compute the ICC manually (as we had just done) is strongly recommended as this is a basic formula which will often be encountered and makes intuitive sense. For academic achievement–related outcomes, ICCs for cross-sectional datasets often range from .10 to .25 in an American educational context (Hedges & Hedberg, 2007).

The `icc` function in the `performance` package can also be used to double-check computations. After installing the `performance` package (i.e., `install.packages('performance')`), the ICC can be computed by using `performance::icc(nullmodel)`. The `::` is used to access the function directly without having to load in the `performance` package using the `library` function. Remember that you only have to install packages once on the computer that you are using.

In older studies, ICCs have been used to justify the need to use MLM or not with some indicating that if the ICC is low (i.e., ICC $<.05$), MLM is not needed. This can make sense in some contexts but cannot be universally applied as even with an ICC as low as .01, incorrect inferential statistics are possible (Huang, 2018b). Even if a low ICC is found, accounting for the clustering using MLM (or some other alternative procedure, see Chapter 12) may be more prudent as it is not the ICC alone that has an impact on the inferential statistical tests used (Huang, 2016). In a later chapter (see Section 4.3), after discussing some other concepts, we discuss statistical significance testing for the ICC.

2.2.2 Understanding the ICC Further

The ICC based on the formula shown can range from 0 to 1 where 0 indicates that all variability is due to individual-level factors while a 1 indicates that all variability is due to group-level factors. For example, imagine in Figure 2.2, the heights of individuals within four clusters are measured. However, within each group, there is no variation in height and are all the same within cluster (think of them as clones; this is just a thought experiment!). As a result, all variation is between clusters, which results in an ICC of 1.

However, if another scenario is considered where the variability of heights within a group is the same in each cluster (and there is no variability between groups as shown in Figure 2.3), the ICC = 0. None of the variability is between groups, and all the variability is within groups.

FIGURE 2.2 ● No Within-Group Variation

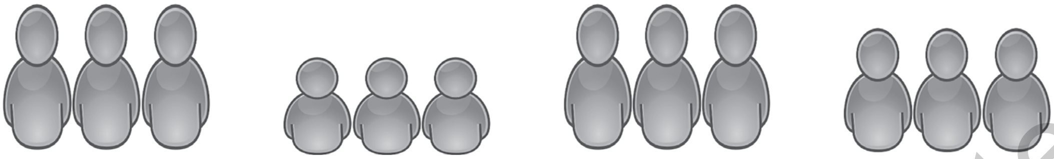
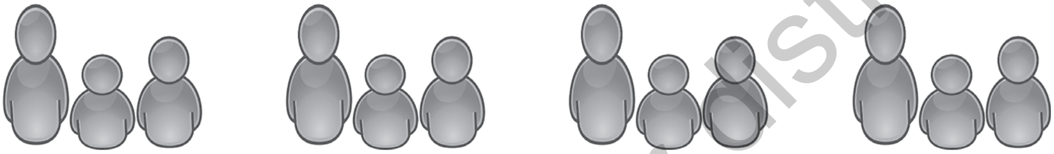


FIGURE 2.3 ● No Between-Group Variation



This simple example illustrates what it means when all variability in the outcome is a result of group differences or individual differences. Note that another way of computing ICCs uses mean square values from ANOVA-based computations which can result in negative ICC values (in those cases, an ICC of 0 is reported). However, the ICC will always be positive using the formula shown using the between-group and within-group variance estimates.

2.3 SUMMARY

In this chapter, basic level notation was introduced and illustrated how we can use formulas to describe and communicate our models of interest. The most basic of multilevel models, the unconditional means or null model (i.e., a model with no predictors) was described and fit using the `lmer` function. The null model provides us with partitioned variance estimates at the group and at the individual level. The partitioned variance components allow us to compute the unconditional ICC which indicates how much variability is attributable to the group and individual levels. The ICC, between-group variability, and within-group variability are basic concepts often encountered when discussing multilevel models and will be constantly revisited in succeeding chapters in this book, so having a firm grasp of their meaning is important.

Test Yourself

Use the `Hsb82` dataset in the `mlmRev` package which can be loaded using `data(Hsb82, package = 'mlmRev')`. Using an unconditional means model with `mAch` as the outcome and `school` as the cluster variable, determine:

1. τ_{00}.
2. σ^2.
3. The total variance of `mAch` using both the answers from #1 and #2.
4. The intraclass correlation coefficient (ICC).