Chapter 5   Adding Fixed Predictors to MLMs

5.1   Learning Objectives

In this chapter, we will introduce fixed predictors at both level-1 and level-2.

The learning objectives for this chapter are:

1. Code and interpret fixed effects in multilevel models;
2. Explain the difference between conditional and unconditional effects;
3. Evaluate the utility of predictors in a model by considering the information from regression coefficients and variance reduced.

All materials for this chapter are available for download here.

5.2   Data Demonstration

The data for this chapter were taken from chapter 3 of Heck, R. H., Thomas, S. L., & Tabata, L. N. (2011). *Multilevel and Longitudinal Modeling with IBM SPSS*: Taylor & Francis. Students are clustered within schools in the data.

5.2.1   Load Data and Dependencies

For this data demo, we will use the following packages:

```r
library(dplyr) # for data manipulation
library(ggplot2) # for visualizations
library(lme4) # for multilevel models
library(lmerTest) # for p-values
library(performance) # for intraclass correlation
```

And the same dataset of students' math achievement:

```r
data <- read.csv('heck2011.csv')
```

5.2.2   MLM with Level-1 Predictor

As a reminder, in Chapter 4 we estimated the random-intercept-only model, also called the null model:

```r
null_model <- lmer(math ~ 1 + (1|schcode), data = data)
summary(null_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: math ~ 1 + (1 | schcode)
##    Data: data
##
## REML criterion at convergence: 48877.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6336 -0.5732  0.1921  0.6115  5.2989
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schcode  (Intercept) 10.64    3.262
##  Residual             66.55    8.158
## Number of obs: 6871, groups:  schcode, 419
##
## Fixed effects:
##             Estimate Std. Error       df t value            Pr(>|t|)
## (Intercept)  57.6742     0.1883 416.0655    306.3 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now that we've explored the null model and variance decomposition it gives us access to, let's practice adding a level-1 predictor to our model. Level-1 predictors vary at level-1, which in our example is the student level, meaning that students have different values for a variable. In our data, socioeconomic status (ses) and sex (female) vary across students, at level-1. Let's add a fixed effect for ses as a predictor to our model.

The following equations describe this model:

Level	Equation
Level 1	$math_{ij} = \beta_{0j} + \beta_{1j}ses_{ij} + R_{ij}$
Level 2	$\beta_{0j} = \gamma_{00} + U_{0j}$
	$\beta_{1j} = \gamma_{10}$
Combined	$math_{ij} = \gamma_{00} + \gamma_{10}ses_{ij} + U_{0j} + R_{ij}$

We'll be estimating four parameters:

1. γ_{00}: the fixed effect for the intercept, controlling for `ses`;
2. γ_{10}: the fixed effect for the slope of `ses`;
3. τ_0^2: a random effect for the intercept capturing the variance of schools around the intercept, controlling for `ses`;
4. σ^2: a random effect capturing the variance of students around their school mean math achievement, controlling for `ses`.

Notice that the parameters are now conditional on `ses`. The intercept is no longer interpreted as the intercept across all schools; it's the intercept across all schools *conditional on* `ses` *being equal to 0*, or at the mean `ses` level for the sample given that `ses` is z-scored in these data. Additionally, note that there is no U_j term associated with the coefficient for `ses`; that's because we're only adding a fixed effect for `ses` right now. This implies that the relationship between `ses` and math achievement is the same across all schools (i.e., the slope is fixed, not randomly varying). We'll look at adding random slope effects in the next chapter. For now, let's run our model.

```
ses_l1 <- lmer(math ~ ses + (1|schcode), data = data, REML = TRUE)
summary(ses_l1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: math ~ ses + (1 | schcode)
##    Data: data
##
## REML criterion at convergence: 48215.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7733 -0.5540  0.1303  0.6469  5.6908
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schcode  (Intercept)  3.469   1.863
##  Residual             62.807   7.925
## Number of obs: 6871, groups:  schcode, 419
##
## Fixed effects:
##             Estimate Std. Error       df t value        Pr(>|t|)
## (Intercept)  57.5960     0.1329 375.6989  433.36 <0.0000000000000002 ***
## ses           3.8739     0.1366 3914.6382   28.35 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## ses -0.025
```

Per the intercept, the average math achievement across all schools at mean `ses` is 57.596. A one-standard-deviation increase in `ses` is associated with a 3.87-point increase in math achievement. The variance term describing how schools vary around the intercept is 3.469, whereas the variance term describing how the students vary within schools, about their schools' mean, is 62.807. These variance terms are different from our null model that had no predictors; we can quantify that difference in at least two ways.

One option is to calculate how much level-1 variance was reduced by adding `ses` as a level-1 predictor. If we divide the difference between our null model's level-1 variance and this new model's (l1) level-1 variance by the null model variance, we can see what proportion of variance was reduced.

```
null <- sigma(null_model)^2
l1 <- sigma(ses_l1)^2

(null - l1) / null
```

```
## [1] 0.05624991
```

So we reduced about 5.6% of level-1 variance by adding `ses` as a level-1 predictor. Another way of stating this is that we reduced the unexplained within school variance by 5.6%.

Another option is to calculate the conditional ICC, or the proportion of variance explained by clustering after we account for `ses`. Recall from last chapter that the adjusted ICC accounts only for random effects, while the conditional ICC accounts for both random effects and fixed effects. With the null model, the adjusted and conditional ICC values from `performance` are the same because there are no predictors in the model, but with a fixed level-1 predictor in the model, we should reference the conditional ICC.

```
performance::icc(ses_l1)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.052
##    Unadjusted ICC: 0.046
```

After accounting for the effect of socioeconomic status, 4.6% of the variance in math achievement is accounted for by school membership.

5.2.3 Compare Regular and Multilevel Regression

In the previous chapter, we compared a regular regression to a cluster-robust standard error regression. Now, let's compare those two with a multilevel model.

The regular regression from Chapter 4:

```
model <- lm(math ~ ses, data = data)
summary(model)
```

```
## 
## Call:
## lm(formula = math ~ ses, data = data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.459  -4.678   1.144   5.355  47.560
## 
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 57.59817    0.09819  586.61 <0.0000000000000002 ***
## ses          4.25468    0.12566   33.86 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 8.132 on 6869 degrees of freedom
## Multiple R-squared:  0.143,  Adjusted R-squared:  0.1429
## F-statistic:  1146 on 1 and 6869 DF,  p-value: < 0.00000000000000022
```

The cluster-robust standard error regression from Chapter 4:

```
model_crse <- lmtest::coeftest(model, vcov = sandwich::vcovCL, cluster = ~ schcode)
model_crse
```

```
## 
## t test of coefficients:
## 
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 57.59817    0.13020 442.378 < 0.00000000000000022 ***
## ses          4.25468    0.14981  28.401 < 0.00000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These two models had the same coefficients, with different significance values.

This is our multilevel model:

```
summary(ses_l1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: math ~ ses + (1 | schcode)
##    Data: data
##
## REML criterion at convergence: 48215.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7733 -0.5540  0.1303  0.6469  5.6908
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schcode  (Intercept)  3.469   1.863
##  Residual             62.807   7.925
## Number of obs: 6871, groups:  schcode, 419
##
## Fixed effects:
##              Estimate Std. Error        df t value         Pr(>|t|)
## (Intercept)  57.5960     0.1329  375.6989  433.36 <0.0000000000000002 ***
## ses           3.8739     0.1366 3914.6382   28.35 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## ses -0.025
```

The intercepts are the same between the MLM and regular regressions, but the coefficient for `ses` is not. Why? The coefficient for `ses` represents the mean relationship between SES and math achievement across all schools, weighted by the reliability of the cluster. The weighting reflects cluster-level sample size, and thus varies from the regular regression estimates that treat all observations equally.

5.2.4 MLM with Level-2 Predictor

We added `ses` as a level-1 predictor to explain some of the student-level variance in math achievement. Now, let's add a predictor that varies at level-2, meaning that the value is different across level 2 units, which is the school level. Level-2 predictors are different across schools but the same for all students within a school. There are three possible level-2 predictors:

- `ses_mean` : the mean SES per school (this variable is centered, we'll discuss centering more in Chapter 9)
- `pro4yc` : the percentage of students at a school who intend to study at a 4-year college/university
- `public` : whether the school is private (0) or public (1)

This is where we begin to unlock the potential of MLMs, to ask questions about both individual differences (level-1 variables) and school differences (level-2 variables) at the same time while accounting for clustered data structures. Let's consider the role of school type in our model by adding a fixed effect for `public` as a predictor of our intercept.

The following equations describe this model:

Level	Equation
Level 1	$math_{ij} = \beta_{0j} + \beta_{1j}ses_{ij} + R_{ij}$
Level 2	$\beta_{0j} = \gamma_{00} + \gamma_{01}public_j + U_{0j}$
	$\beta_{1j} = \gamma_{10}$
Combined	$math_{ij} = \gamma_{00} + \gamma_{01}public_j + \gamma_{10}ses_{ij} + U_{0j} + R_{ij}$

A few things to note here: first, `public_j` only has a j subscript because only different schools (j's) have different values of `public`. All students (i's) within a school have the same value. Second, `public` is currently only a predictor for the intercept. In Chapter 6 we'll look at using level-2 variables as predictors of level-1 slopes and the cross-level interactions that result.

We'll be estimating five parameters:

1. γ_{00}: the fixed effect for the intercept, controlling for `ses` and `public`;
2. γ_{01}: the fixed effect for the slope of `public` controlling for `ses`
3. γ_{10}: the fixed effect for the slope of `ses` controlling for `public`;

4. τ_0^2: a random effect for the intercept capturing the variance of schools around the intercept, controlling for `ses` and `public`;

5. σ^2: a random effect capturing the variance of students around their school mean math achievement, controlling for `ses` and `public`.

Notice that the parameters are conditional on both `ses` and on `public` now. Let's run our model.

```r
ses_l1_public_l2 <- lmer(math ~ 1 + ses + public + (1|schcode), data = data, REML = TRUE)
summary(ses_l1_public_l2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
## Formula: math ~ 1 + ses + public + (1 | schcode)
##    Data: data
##
## REML criterion at convergence: 48216
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7718 -0.5541  0.1309  0.6477  5.6916
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schcode  (Intercept)  3.486   1.867
##  Residual             62.807   7.925
## Number of obs: 6871, groups:  schcode, 419
##
## Fixed effects:
##               Estimate Std. Error       df t value         Pr(>|t|)
## (Intercept)   57.63143    0.25535 381.81733 225.693 <0.0000000000000002 ***
## ses            3.87338    0.13673 3928.37427  28.329 <0.0000000000000002 ***
## public        -0.04859    0.29862 385.93649  -0.163            0.871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr) ses
## ses     0.013
## public -0.854 -0.031
```

Let's look at our fixed effects, which describes the conditional mean effect of a variable on the outcome, across all schools. Per the intercept, the average math achievement across all private schools (public = 0) at mean SES (ses = 0) is 57.70. A one-standard-deviation increase in

`ses` across all private schools is associated with a 3.87-point increase in math achievement. Public schools at mean `ses` have a -0.14-point decrease on average in math achievement relative to private schools.

From our random effects, the variance term describing how schools vary around the intercept (at mean SES at private schools) is 3.48, and the variance term describing how students vary around their school means is 62.81.

Let's calculate variance reduced at level 1 and level 2 by adding school type as a predictor.

```
# level-1 variance reduced
sigma2_null <- sigma(null_model)^2
sigma2_public <- sigma(ses_l1_public_l2)^2
(sigma2_null - sigma2_public) / sigma2_null
```

```
## [1] 0.05624525
```

```
# level-2 variance reduced
tau2_null <- VarCorr(null_model)$schcode[1]
tau2_public <- VarCorr(ses_l1_public_l2)$schcode[1]
(tau2_null - tau2_public) / tau2_null
```

```
## [1] 0.6724414
```

We reduced around 5.6% of variance in math achievement at level-1 and 67.2% of variance at level-2 by adding `public` as a level-2 predictor. It makes sense that the variance at level-2 was reduced by so much more, because we added a level-2 predictor that varies at level-2.

So, does it seem like school type is related to math achievement? We have two sources of information to consider so far: the regression coefficient and the variance reduced. While the regression coefficient is relatively small, the intercept variance reduced at level-2 is quite large (67%!), so it seems like school type is a valuable predictor in our model.

5.3 Conclusion

In this chapter, we added level-1 and level-2 fixed effects to our models, considered the difference between conditional and unconditional effects, and used regression coefficients and variance reduced to make a decision about retaining model parameters. In Chapter 6, we'll work with random slopes and explain cross-level interactions.