Linear Mixed-Effects Models: Cross-Sectional Analysis

2019-03-06

In this set of notes, you will learn about several of the common linear mixed-effects models fitted in an analysis of cross-sectional data.

Dataset and Research Question

In this set of notes, we will use data from two files, the *netherlands-students.csv* file and the *netherlands-schools.csv* file (see the data codebook here). These data include student- and school-level attributes, respectively, for $n_i = 2287$ 8th-grade students in the Netherlands.

```
# Load libraries
library(AICcmodavg)
library(broom)
library(dplyr)
library(ggplot2)
library(lme4) #for fitting mixed-effects models
library(readr)
library(sm)
library(tidyr)

# Read in student-level data
student_data = read_csv(file = "~/Documents/github/epsy-8252/data/netherlands-students.csv")

# Read in school-level data
school_data = read_csv(file = "~/Documents/github/epsy-8252/data/netherlands-schools.csv")

# Join the two datasets together
joined_data = left_join(student_data, school_data, by = "school_id")
head(joined_data)
```

```
# A tibble: 6 x 13
  student_id school_id language_pre language_post   ses verbal_iq female
       <dbl>     <dbl>        <dbl>         <dbl> <dbl>     <dbl>  <dbl>
1      17001         1           36            46    23      3.17      0
2      17002         1           36            45    10      2.67      0
3      17003         1           33            33    15     -2.33      0
4      17004         1           29            46    23     -0.834     0
5      17005         1           19            20    10     -3.83      0
6      17006         1           22            30    10     -2.33      0
# ... with 6 more variables: minority <dbl>, school_type <chr>,
#   public <dbl>, school_ses <dbl>, school_verbal_iq <dbl>,
#   school_minority <dbl>
```

We will use these data to explore the question of whether verbal IQ scores predict variation in post-test language scores.

Unconditional Random Intercepts Model

As in a conventional fixed-effects regression analysis we begin a mixed-effects analysis by fitting the intercept-only model. This model is referred to as the *unconditional random intercepts model* or the *unconditional means model*. This model includes a fixed-effect of intercept and a random-effect of intercept, and no other predictors. This is the simplest model we can fit while still acounting fo the dependence in the data (e.g., including a random-effect). The statistical model in this example can be expressed as:

$$\text{Language Score}_{ij} = \big[\beta_0 + b_{0j} \big] + \epsilon_{ij}$$

where,

- Language Score$_{ij}$ is the post-test language score for student i in school j;
- β_0 is the fixed-effect of intercept, b_{0j} is the random-effect of intercept for school j; and
- ϵ_{ij} is the error for student i in school j.

As we have been talking about in class, the full specification of a model also includes a mathematical description of the distributional assumptions. Mixed-effects models have distributional assumptions on the errors (ϵ_{ij}) and on each set of random-effects included in the model (b_{0j} in our model). The assumptions on the errors are:

- Independence;
- Conditional normality;
- Conditional means are 0; and
- Homoskedasticity of the conditional variances σ_ϵ^2.

Note that the independence assumption does not assume independence in the original data, but is on the errors which are produced after we account for the dependence in the data by including a random-effect in the model.

The assumptions on each set of random-effects are:

- Independence;
- Normality;
- Mean of 0; and
- There is some variance, $\sigma_{b_0}^2$ (often just denoted σ_0^2)

In mathematical notation the assumptions for the unconditional random intercepts model can be written as:

$$\epsilon_{ij} \overset{i.i.d}{\sim} \mathcal{N}\big(0, \sigma_\epsilon^2\big)$$

$$b_{0j} \overset{i.i.d}{\sim} \mathcal{N}\big(0, \sigma_0^2\big)$$

Fitting and Interpreting the Model

We fit the model and display the output below. We include the argument `REML=FALSE` to force the `lmer()` function to produce maximum likelihood estimates (rather than restricted maximum likelihood estiates). In practice, we will generally want to fit these models using ML estimation.

```
lmer.0 = lmer(language_post ~ 1 + (1 | school_id), data = joined_data, REML = FALSE)

# Coefficient-level output and variance components
summary(lmer.0)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: language_post ~ 1 + (1 | school_id)
   Data: joined_data

     AIC      BIC   logLik deviance df.resid
   16259    16276    -8127    16253     2284

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1166 -0.6573  0.0762  0.7412  2.5058

Random effects:
 Groups   Name        Variance Std.Dev.
 school_id (Intercept) 19.4     4.41
 Residual             64.6     8.04
Number of obs: 2287, groups:  school_id, 131

Fixed effects:
            Estimate Std. Error t value
(Intercept)   40.364      0.426    94.7
```

The `summary()` function displays the fitted coefficients for the fixed-effects, and the variance estimates for the errors ($\hat{\sigma}_\epsilon^2$) and the random-effect of intercept ($\hat{\sigma}_0^2$). Using the fixed-effects estimates, the fitted equation for the fixed-effects model is:

$$\hat{\text{Language Score}}_{ij} = 40.36$$

We interpret coefficients from the fixed-effects model the same way we interpret coefficients produced from the `lm()` output. For example,

- The predicted average post-test language score for all students in all schools is 40.36.

The variance estimates are:

- $\hat{\sigma}_\epsilon^2 = 64.57$
- $\hat{\sigma}_0^2 = 19.43$

We can also use the `tidy()` function to obtain these estimates.

```
tidy(lmer.0)
```

```
# A tibble: 3 x 5
  term                     estimate std.error statistic group
  <chr>                       <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                  40.4     0.426      94.7 fixed
2 sd_(Intercept).school_id      4.41    NA         NA   school_id
3 sd_Observation.Residual       8.04    NA         NA   Residual
```

In the `tidy()` output, the fixed-effect of intercept is the same as that from the `summary()` output,

- $\hat{\beta}_0 = 40.36$

`tidy()` produces estimated standard deviations of the distributions of errors and random-effects rather than variances. To obtain the variances, we square these estimates.

- $\hat{\sigma}_\epsilon^2 = 8.04^2 = 64.6$
- $\hat{\sigma}_0^2 = 4.41^2 = 19.4$

Partitioning Unexplained Variation

To understand how the mixed-effect model allows for a better understanding of the explained variation, let's consider the **intercept-only fixed-effects regression model** (no random-effect term):

$$\text{Language Score}_i = \beta_0 + \epsilon_i$$

In this model, the fixed-effect for intercept represents the global average post-test language score and the error term represents the deviation between Student i's post-test language score and the global average post-test language score. Recall that the error component of the model symbolizes the unexplained variation in language scores. Since the error term, which encompasses all the unexplained variation in the model, is a deviation to the student score, this implies that the *unexplained variation in the fixed-effects model is all at the student-level*. To explain additional variation we would need to include student-level predictors (i.e., predictors that vary between students).

The **random-intercepts regression** model is expressed as (with no parentheses):

$$\text{Language Score}_{ij} = \beta_0 + b_{0j} + \epsilon_{ij},$$

In this model, the fixed-effect for intercept still represents the global average post-test language score, but now Student i's deviation is composed of two separate components: (1) the random-effect represents the deviation from School j's average post-test language score from the global average post-test language score, and (2) the error term represents the deviation between Student i's post-test language score and her school's average post-test language score.

Another way to think about this model is that it has taken the fixed-effects model from earlier and separated the error term from that model into two components school-level deviations (b_{0j}) and student-level deviations (ϵ_{ij}).

$$\text{Language Score}_{ij} = \beta_0 + \overbrace{\left[b_{0j} + \epsilon_{ij}\right]}^{\epsilon_i},$$

In other words, the mixed-effects model partitions the unexplained variation into two parts: (1) school-level variation, and (2) student-level variation. Some statisticians may refer to these as *between-school* variation and *within-school* variation, respectively.

The variance estimates are the quantification of this partitioning. Together the two variance estimates represent variation that is unexplained for by the model (they are errors/deviations after all). Since one mathematical property of variances is that they are additive, we can compute the total unexplained variation by summing the variance estimates:

$$\sigma^2_{\text{Total Unexplained}} = \hat{\sigma}^2_0 + \hat{\sigma}^2_\epsilon$$

$$= 19.43 + 64.57$$

$$= 84$$

We can now use this total value to compute the proportion of unexplained variation at both the school- and student-levels. The proportion of **unexplained variation at the school-level** is:

$$\frac{19.43}{84} = 0.231$$

The proportion of **unexplained variation at the student-level** is:

$$\frac{64.57}{84} = 0.769$$

Interpreting these,

- 23.1% of the unexplained variation is at the school-level (between-school variation).
- 76.9% of the unexplained variation is at the student-level (within-school variation).

Based on this partitioning from the unconditional random intercepts model we have evidence that it may be fortuitous to include both student-level and school-level predictors; there is unaccounted for variation at both levels. To explain the unaccounted for variation at the student-level, include student-level predictors in the model. To explain the unaccounted for variation at the school-level, include school-level predictors in the model. Since more of the unaccounted for variation is at the student-level than the school-level,we may want to focus on the inclusion of student-level predictors rather than school-level predictors.

> This partitioning of variation should be done in every analysis, and ALWAYS is done using the unconditional random intercepts model. The unconditional random intercepts model will serve as our *baseline* model. As we add predictors, we can compare the unexplained variation at each level in the predictor models to the baseline unaccounted for variation in the unconditional means model. This is one way of measuring how effective predictors are at further explaining variation in the model.

Including Predictors

As we begin to include predictors in the model, those predictors might be at the student-level or the school-level. Student-level predictors, as you might expect, would explain variation at the student-level (i.e., within-school variation), and school-level predictors will explain variation at the school-level (i.e., between-school variation). In this analysis, we will examine two student-level predictors, `verbal_iq` (focal predictor) and `ses` (covariate), as well as one school-level predictor, `public` (covariate).

We begin by including the fixed-effect of our focal predictor, verbal_iq, into the random intercepts model. The statistical model for this can be expressed as:

$$\text{Language Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

In this model,

- β_0 is the fixed-effect of intercept;
- b_{0j} is the random-effect of intercept for School j (school deviation);
- β_1 is the fixed-effect of verbal IQ; and
- ϵ_{ij} is the error for Student i in School j (student deviation)

Fitting this model using the lmer() function:

```
# Fit model
lmer.1 = lmer(language_post ~ 1 + verbal_iq + (1 | school_id), data = joined_data, REML = FALSE)

# Output
summary(lmer.1)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: language_post ~ 1 + verbal_iq + (1 | school_id)
   Data: joined_data

     AIC      BIC   logLik deviance df.resid
   15260    15283    -7626    15252     2283

Scaled residuals:
    Min      1Q  Median      3Q     Max
 -4.096  -0.637   0.058   0.707   3.147

Random effects:
 Groups    Name         Variance Std.Dev.
 school_id (Intercept)  9.5      3.08
 Residual               42.2     6.50
Number of obs: 2287, groups:  school_id, 131

Fixed effects:
            Estimate Std. Error t value
(Intercept)  40.6094     0.3069   132.3
verbal_iq     2.4881     0.0701    35.5

Correlation of Fixed Effects:
          (Intr)
verbal_iq 0.018
```

Using the fixed-effects estimates, the fitted equation is:

$$\hat{\text{Language Score}}_{ij} = 40.61 + 2.49(\text{Verbal IQ}_{ij})$$

Interpreting these coefficients,

- The predicted average post-test language score for students with a mean verbal IQ score (=0) is 40.61.
- Each one-point difference in verbal IQ score is associated with a 2.49-point difference in language scores, on average.

The variance estimates are:

- $\hat{\sigma}_\epsilon^2 = 42.23$
- $\hat{\sigma}_0^2 = 9.50$

First note that by including a student-level predictor we REDUCED the unexplained variation at the student-level and at the school-level. Reducing the unexplained student-level variation was intentional (we included a student-level predictor). Reducing the unexplained school-level variation was a mathematical artifact of the estimation process when we included the student-level predictor (Bonus!).

Table 1: Variance Estimates from Fitting the Unconditional Random Intercepts Model (Model 1) and the Conditional Random Intercepts Model with Verbal IQ as a Fixed-Effect (Model 2)

Estimate	Model 1	Model 2
σ_ϵ^2	64.6	42.2
σ_0^2	19.4	9.5

To determine how much we reduced the unexplained variance, we compute the **proportion of reduction relative to the unconditional random intercepts model**.

$$\text{Student-Level}: \frac{64.57 - 42.23}{64.57} = 0.346$$

$$\text{School-Level}: \frac{19.43 - 9.50}{19.43} = 0.511$$

- Verbal IQ accounted for 34.6% of the unexplained variation at the student-level.
- Verbal IQ also accounted for 51.1% of the unexplained variation at the school-level.

Another way that applied researchers write this is to use the language of "explained variation". For example,

- Verbal IQ explains 34.6% of the variation at the student-level.
- Verbal IQ also explained 51.1% of the variation at the school-level.

Including the student-level predictor also CHANGED the amount of unaccounted for variation at the school-level. In this case it happened to reduce this variation, but other times, you will see that the variation stays about the same, or increases! This is a mathematical artifact of the estimation. In a more practical sense, we wouldn't really be too interested in the school-level variation at this point. We are only adding student-level predictors to the model, so that is the variation that we expect to impact.

Evaluating Predictors

As we include predictors in the model, we want to evaluate their overall worth to determine whether they should be retained or omitted from the model. Unlike the fixed-effects models we fitted in EPsy 8251, we cannot use the p-value for the coefficients for evidence of predictor importance as there is no p-value provided in either the `summary()` nor `tidy()` output. Subsequently, we need to look at other evidence.

Typically we compare the model that includes the predictor to the same model without the predictor and evaluate differences between them. There are three pieces of evidence that are commonly evaluated when making this comparison: (1) reduction in unexplained variation; (2) AICc and model evidence; and (3) t-values of the fixed-effect.

In our example, we are comparing the models:

$$\textbf{Model 1}: \text{Language Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \epsilon_{ij}$$
$$\textbf{Model 2}: \text{Language Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

We already examined the reduction in unexplained variation; including Verbal IQ explained roughly 35% of the unexplained variation at the student-level and 50% of the unexplained variation at the school-level. This is evidence to include Verbal IQ in the model.

The table of model evidence (below) also strongly supports inclusion of the Verbal IQ predictor.

```
aictab(
  cand.set = list(lmer.0, lmer.1),
  modnames = c("Model 1", "Model 2")
)
```

```
Model selection based on AICc:

        K  AICc Delta_AICc AICcWt Cum.Wt    LL
Model 2 4 15260          0      1      1 -7626
Model 1 3 16259        999      0      1 -8127
```

Lastly, we can examine the `summary()` output of the conditional model to examine the t-value associated with the fixed-effect of Verbal IQ.

```
Fixed effects:
            Estimate Std. Error t value
(Intercept) 40.60937    0.30686  132.34
verbal_iq    2.48809    0.07005   35.52
```

A rule-of-thumb is that t-values greater than 2 support inclusion of the predictor. Here the t-value associated with Verbal IQ is $t = 35.52$. This is evidence for including Verbal IQ in the model.

In our example, all three pieces of evidence suppported the inclusion of the predictor in the model. This does not always happen. Sometimes the evidence is not congruent in its support. Because of this, it is good to have a plan about which set of evidence you will use to make decisions.

Including the Covariates

Now, we may want to determine whether Verbal IQ is still an important predictor of variation in post-test language scores after we control for differences in SES and whether the school the student attends is a public school. To examine this, we will fit an additional model that includes the fixed-effects of Verbal IQ, SES, and school type. We also continue to include a random-effect of intercept to account for the dependency of post-test scores within schools.

```
# Fit model
lmer.2 = lmer(language_post ~ 1 + verbal_iq + ses + public + (1 | school_id),
              data = joined_data, REML = FALSE)

# Coefficient-level output and variance components
summary(lmer.2)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: language_post ~ 1 + verbal_iq + ses + public + (1 | school_id)
   Data: joined_data

     AIC      BIC   logLik deviance df.resid
   15137    15171    -7562    15125     2281

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.041 -0.649  0.054  0.710  3.084

Random effects:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept)  8.57     2.93
 Residual              40.04     6.33
Number of obs: 2287, groups:  school_id, 131

Fixed effects:
            Estimate Std. Error t value
(Intercept)  36.5197     0.5344   68.34
verbal_iq     2.2489     0.0714   31.51
ses           0.1666     0.0148   11.28
public       -1.4176     0.6233   -2.27

Correlation of Fixed Effects:
         (Intr) vrbl_q ses
verbal_iq  0.215
ses       -0.742 -0.295
public    -0.356  0.045 -0.040
```

Table 2 shows the variance estimates for all three models fitted thus far.

Table 2: Variance Estimates from Fitting the Unconditional Random Intercepts Model (Model 1), the Conditional Random Intercepts Model with Verbal IQ as a Fixed-Effect (Model 2), and the Conditional Random Intercepts Model with Verbal IQ, SES, and School Type as a Fixed-Effects (Model 3)

Estimate	Model 1	Model 2	Model 3
σ^2_ϵ	64.6	42.2	40.04
σ^2_0	19.4	9.5	8.57

Including both student-level and school-level predictors in the model has explained variation at both the student- and school-levels. As expected, the unexplained variance at the student-level was reduced by including SES in the model. Similarly, the unexplained variance at the school-level was also reduced by including school type. To quantify the amount of reduction we compare to the unconditional model.

$$\text{Student-Level}: \frac{64.57 - 40.04}{64.57} = 0.38$$

$$\text{School-Level}: \frac{19.43 - 8.57}{19.43} = 0.559$$

- The model (verbal IQ, SES, and public) explains 38.0% of the variation at the student-level.
- The model (verbal IQ, SES, and public) explains 55.9% of the variation at the school-level.

The model evidence also seems to suggest that this conditional model is more likely given the data than either of the other two candidate model.

```
aictab(
  cand.set = list(lmer.0, lmer.1, lmer.2),
  modnames = c("Model 1", "Model 2", "Model 3")
)
```

```
Model selection based on AICc:

        K  AICc Delta_AICc AICcWt Cum.Wt    LL
Model 3 6 15137          0      1      1 -7562
Model 2 4 15260        123      0      1 -7626
Model 1 3 16259       1122      0      1 -8127
```

Both the model evidence and variance explained values point toward the model that includes fixed-effects of verbal IQ, SES, and school type. What these values don't tell us is whether both covariates (SES and school type) are needed or whether we can drop one or the other. The t-values for each of the fixed-effects can help us think about this. Since all of the t-values in the model are greater than 2, this suggests that each of the predictors seems to be statistically relevant, controlling for the other predictors in the model.

Now that we have adopted a model, we should interpret the fixed-effects. The fitted equation for Model 3 is:

$$\text{Language Score}_{ij} = 36.52 + 2.25(\text{Verbal IQ}_{ij}) + 0.17(\text{SES}_{ij}) - 1.42(\text{Public}_{\bullet j})$$

Notice that the Public variable has a $\bullet j$ subscript. Recall that `public` is a school-level predictor. As such it does not vary for students within a school; it only varies between schools. The lack of an i in the subscript indicates that it does not vary across individuals within a school. Language scores, verbal IQ scores, and SES are all student-level predictors, so they get both an i and a j subscript.

Interpreting the effect associated with verbal IQ, our focal predictor:

- Each one-unit difference in verbal IQ score is associated with a 2.25-point difference in post-test language scores, on average, controlling for differences in SES and school type.

We probably wouldn't bother interpreting the effects of the intercept nor the non-focal covariates, but for pedagogical purposes:

- **Intercept:** The predicted average post-test language score for students with a mean verbal IQ score (=0), a SES value of 0, and enrolled in a private school (reference group) is 36.52. (This is extrapolation as in the data, the lowest SES value is 10.)
- **SES:** Each one-unit difference in SES score is associated with a 0.17-point difference in post-test language scores, on average, controlling for differences in verbal IQ score and school type.
- **Public:** Students enrolled in public school, have a post-test language score that is 1.42-points lower, on average, than students enrolled in a private school, controlling for differences in verbal IQ and SES.

Displaying the Results of the Fitted Models

It is common to display the results of the fitted models in a table or plot. Typically we would use the table to show the results of a subset of fitted models and display the adopted "final" model(s) in a plot. For this example, I would display in a table the results of three different fitted models: (1) the unconditional random intercepts model; (2) the conditional random intercepts model that includes the fixed-effect of verbal IQ (our focal predictor); and (3) the conditional random intercepts model that includes the fixed-effect of verbal IQ and all adopted covariates (SES and school type).

Table of Fitted Models

In displaying the results from fitted mixed-effects models, we typically provide (1) fixed-effects estimates; (2) variance component estimates; and (3) model-level evidence (e.g., LL, AIC). If you are using Markdown, there are several packages that can be used to obtain syntax for these types of tables. Below I use the `stargazer()` function from the **stargazer** package to display the fixed-effects estimates and model-level evidence for each of the three fitted models. If you are using R Markdown, don't forget to set the chunk options for `results='asis'`.

By default the `stargazer()` function shows stars/p-values for the coefficients. In mixed-effects models the coefficient-level p-values are quite controversial, and may be mis-leading. As such, I removed them from the table of regression results using the argument `star.cutoffs = NA`.

```r
library(stargazer)

stargazer(
  lmer.0, lmer.1, lmer.2,
  type = "latex",
  title = "Fixed-Effects Coefficients and Standard Errors for a Taxonomy of Fitted Models to Predict Post-Test Lang
  column.labels = c("Model 1", "Model 2", "Model 3"),
  colnames = FALSE,
  model.numbers = FALSE,
  dep.var.caption = "Outcome: Post-Test Language Scores",
  dep.var.labels.include = FALSE,
  covariate.labels = c("Verbal IQ scores", "SES", "School type"),
  keep.stat = c("ll"),
  notes.align = "l",
  add.lines = list(c("Corrected AIC", round(AICc(lmer.0), 1), round(AICc(lmer.1), 1),
                   round(AICc(lmer.2), 1))),
  star.cutoffs = NA, # Omit stars
  omit.table.layout = "n", #Don't show table notes
  header = FALSE, #Suppress message
  table.placement = "H"
  )
```

Table 3: Fixed-Effects Coefficients and Standard Errors for a Taxonomy of Fitted Models to Predict Post-Test Language Scores for 2,287 Students from 131 Schools. All Models Included a Random-Effect of Intercept and were Fitted using Maximum Likelihood.

| | Outcome: Post-Test Language Scores | | |
	Model 1	Model 2	Model 3
Verbal IQ scores		2.490	2.250
		(0.070)	(0.071)
SES			0.167
			(0.015)
School type			−1.420
			(0.623)
Constant	40.400	40.600	36.500
	(0.426)	(0.307)	(0.534)
Corrected AIC	16259.2	15259.8	15136.9
Log Likelihood	−8,127.000	−7,626.000	−7,562.000

Table 4: Variance Estimates from Fitting the Unconditional Random Intercepts Model (Model 1), the Conditional Random Intercepts Model with Verbal IQ as a Fixed-Effect (Model 2), and the Conditional Random Intercepts Model with Verbal IQ, SES, and School Type as a Fixed-Effects (Model 3)

Estimate	Model 1	Model 2	Model 3
σ^2_{ϵ}	64.6	42.2	40.04
σ^2_0	19.4	9.5	8.57

The variance component estimates should also be provided for each of the models displayed. These can be displayed in the same table as the fixed-effects (generally below the fixed-effects, but prior to the model-level summaries), or in a separate table. Below, I manually enter these in a data frame and use the `kable()` function and functions from the **kableExtra** package to format the table.

```r
data.frame(
  var_comp = c("$\\sigma^2_{\\epsilon}$", "$\\sigma^2_{0}$"),
  mod_1 = c(64.57, 19.43),
  mod_2 = c(42.23, 9.50),
  mod_3 = c(40.04, 8.57)
) %>%
  kable(
    col.names = c("Estimate", "Model 1", "Model 2", "Model 3"),
    caption = "Variance Estimates from Fitting the Unconditional Random Intercepts Model (Model 1), the Conditional
    format = 'latex',
    escape = FALSE,
    booktabs = TRUE,
    align = "c"
    ) %>%
  kable_styling(full_width = FALSE, "HOLD_position") %>%
  column_spec(1:4, width = "1in")
```

Plot of "Final" Model(s)

If you plot the results from the "final" adopted model(s), it is typical to plot only the fixed-effects part of the model. We do this exactly the same way we do for plotting the results from `lm()`. Here I use a sequence of values for verbal IQ scores (focal predictor; plotted on the x-axis), control for the effects of SES (set it to its mean value), and show different lines for each school type.

```r
# Set up plotting data for lmer.2
plot_data = crossing(
  verbal_iq = seq(from = -7.83, to = 6.17, by = 0.01),
  ses = 27.81,
  public = c(0, 1)
)
```

We then use the `predict()` function to obtain the $\hat{Y}$ values for the adopted model(s). We include the `re.form=NA` argument to ignore the random-effects (only compute the $\hat{Y}$ values based on the fixed-effects). Additionally, I coerce the `public` variable into a factor for better plotting labels.

```r
# Predict life satisfaction
plot_data = plot_data %>%
  mutate(
```

```
    yhat = predict(lmer.2, newdata = ., re.form = NA),
    school_type = factor(public, levels = c(0, 1), labels = c("Private", "Public"))
  )


head(plot_data)
```

```
# A tibble: 6 x 5
  verbal_iq   ses public  yhat school_type
      <dbl> <dbl>  <dbl> <dbl> <fct>
1     -7.83  27.8      0  23.5 Private
2     -7.83  27.8      1  22.1 Public
3     -7.82  27.8      0  23.6 Private
4     -7.82  27.8      1  22.1 Public
5     -7.81  27.8      0  23.6 Private
6     -7.81  27.8      1  22.2 Public
```

```
ggplot(data = plot_data, aes(x = verbal_iq, y = yhat, color = school_type, linetype = school_type)) +
  geom_line() +
  theme_bw() +
  xlab("Verbal IQ score") +
  ylab("Predicted post-test language score") +
  ggsci::scale_color_d3(name = "School type") +
  scale_linetype_discrete(name = "School type")
```
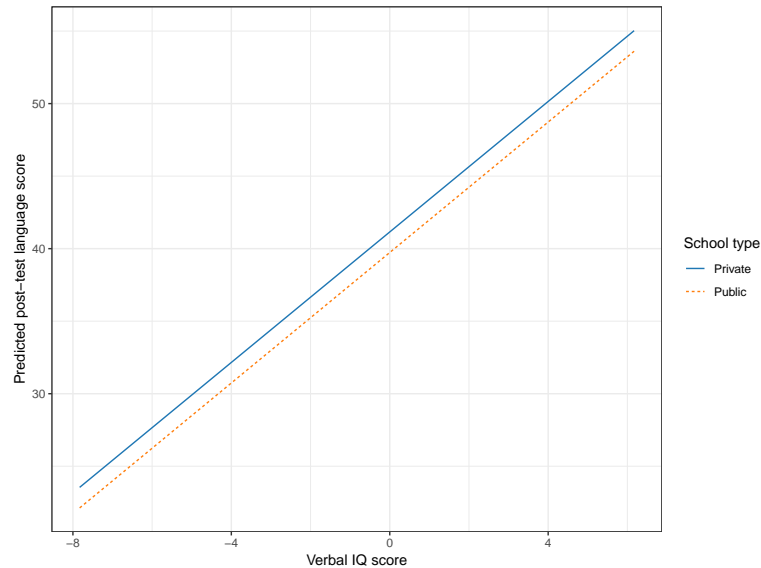

Figure 1: Plot of the model predicted post-test language scores as a function of verbal IQ scores, mean SES, and school type.

Other Resources

In addition to the notes and what we cover in class, there many other resources for learning about using linear mixed-effects models for cross-sectional analysis. Here are some resources that may be helpful in that endeavor:

- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623–641.
- Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347–367.

For **table formatting** using R Markdown, check out:

- Stargazer for lmer Models

If you are interested in the estimation of the coefficients, see:

- Estimation of the Log-Likelihood in lmer

Appendix: p-Values for LMER Models

So long as the random-effects structure is the same across models, we can obtain p-values for the fixed-effects using a likelihood ratio test. This test requires testing nested models, similar to the ΔF-test for fixed-effects models. To illustrate this I will compute a likelihood-based p-value for the effect of Verbal IQ by comparing Model 1 (the unconditional model) to Model 2 (the conditional model including verbal IQ).

$$\textbf{Model 1}: \text{Language Score}_{ij} = \big[\beta_0 + b_{0j}\big] + \epsilon_{ij}$$
$$\textbf{Model 2}: \text{Language Score}_{ij} = \big[\beta_0 + b_{0j}\big] + \beta_1(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

These models have the exact same random-effects structure (random-effect of intercept only) and Model 1 is nested in Model 2. Since the additional fixed-effect(s) in Model 2 only include the effect of verbal IQ, the test will examine whether the inclusion of the verbal IQ effect is warranted. Here the null hypothesis would be:

$$\beta_{\text{Verbal IQ}} = 0$$

To carry out the likelihood ratio test in R, we use the `anova()` function and give it the two models as input. We also include the additional argument `test="LRT"` to signify that we want to carry out a likelihood ratio test.

```
anova(lmer.0, lmer.1, test = "LRT")
```

```
Data: joined_data
Models:
lmer.0: language_post ~ 1 + (1 | school_id)
lmer.1: language_post ~ 1 + verbal_iq + (1 | school_id)
       Df   AIC   BIC logLik deviance Chisq Chi Df          Pr(>Chisq)
lmer.0  3 16259 16276  -8127    16253
lmer.1  4 15260 15283  -7626    15252  1001      1 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of the likelihood ratio test are similat to that from the ΔF-test. In this test, however, we are comparing the difference in deviances (remember, smaller deviance = less error). Model 2 seems to fit the data better as it has a smaller deviance than Model 1. The difference in deviance is 1001. However, Model 2 is also more complex than Model 1 by one parameter.

To determine whether the improved fit is worth the added complexity, we can evaluate the difference in deviance using a χ^2 distribution with degrees of freedom equal to the difference in the number of model parameters. In our example,

$$\chi^2(1) = 1001$$

This distribution, much like the F-distribution is one-sided, so we would compute the cumulative area under the chi-squared distribution that is greater than or equal to 1001. This can be computed using the pchisq() function.

```
1 - pchisq(q = 1001, df = 1)
```

```
[1] 0
```

Note that this is the same p-value presented in the output of the likelihood ratio test results. This would suggest that we reject the null hypothesis; it is likely that there is an effect of verbal IQ on post-test language scores.

To obtain likelihood-based p-values for the effects in Model 2, you would need to carry out many likelihood ratio tests. For example, to compute the likelihood-based p-value for verbal IQ in Model 2 we would need to compare these two models:

$$\textbf{Simple Model}: \text{Language Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{SES}_{ij}) + \beta_1(\text{Public}_j) + \epsilon_{ij}$$
$$\textbf{Complex Model}: \text{Language Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{SES}_{ij}) + \beta_1(\text{Public}_j) + \beta_1(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

Essentially comparing the model with all predictors except verbal IQ to the model that includes those predictors and also includes verbal IQ.

```
# Fit models
lmer.simple = lmer(language_post ~ 1 + ses + public + (1 | school_id),
             data = joined_data, REML = FALSE)

lmer.complex = lmer(language_post ~ 1 + ses + public + verbal_iq + (1 | school_id),
             data = joined_data, REML = FALSE)

# Carry out LRT
anova(lmer.simple, lmer.complex, test = "LRT")
```

```
Data: joined_data
Models:
lmer.simple: language_post ~ 1 + ses + public + (1 | school_id)
lmer.complex: language_post ~ 1 + ses + public + verbal_iq + (1 | school_id)
            Df    AIC    BIC logLik deviance Chisq Chi Df
lmer.simple  5 15959 15988  -7974    15949
lmer.complex 6 15137 15171  -7562    15125   824      1
                  Pr(>Chisq)
lmer.simple
lmer.complex <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The likelihood-based p-value associated with the effect of verbal IQ in Model 2 is $p < .001$. To obtain the likelihood-based p-values for the other predictors in the model, you would need to carry out additional LRTs to compare models that exclude and include each effect.

The likelihood-based p-values could then be reported similarly to how you would report them in a fixed-effects regression analysis. (However, you should indicate that they are likelihood-based p-values.) They can be presented in text, e.g.,

> The effect of verbal IQ on post-test language score was statistically significant even after accounting for differences in SES and school type, $\chi^2(1) = 824$, $p < .001$.

They can also be presented in tables, either by presenting the p-value or via the star system. If you are using `stargazer()`, you may need to manually include the stars and the table note indicating the p-value cutoff associated with each number of stars.