Linear Mixed-Effects Models: Longitudinal Analysis

2019-03-06

In this set of notes, you will learn how to use the linear mixed-effects model to analyze longitudinal data.

Dataset and Research Question

In this set of notes, we will use data from the file *vocabulary.csv* (see the data codebook here). These data include repeated measurements of scaled vocabulary scores for $n = 64$ students.

```
# 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 data
vocabulary = read_csv(file = "~/Documents/github/epsy-8252/data/vocabulary.csv")
head(vocabulary)
```

```
# A tibble: 6 x 6
     id vocab_08 vocab_09 vocab_10 vocab_11 female
  <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  <dbl>
1     1     1.75     2.6      3.76     3.68       1
2     2     0.9      2.47     2.44     3.43       0
3     3     0.8      0.93     0.4      2.27       0
4     4     2.42     4.15     4.56     4.21       1
5     5    -1.31    -1.31    -0.66    -2.22       0
6     6    -1.56     1.67     0.18     2.33       0
```

We will use these data to explore the change in vocabulary over time (longitudinal variation in the vocabulary scores). We will focus on two primary research questions: (1) What is the growth pattern in the average vocabulary score over time? and (2) Is this growth pattern different for males and females?

Data Structure: Tidy/Long Data vs. Wide Data

Before doing any analysis of the data, it is worth understanding the structure of the data. There are two common structures for repeated measures data: *tidy/long structured data* and *wide structured data*.

- In **tidy/long** structured data, there is a single column per variable. For example, the outcome variable (vocabulary scores) would be organized into a single column. Similarly, the predictor that designates time (grade-level in our example) would also be organized into a single column.
- In **wide** structured data, the outcome variable (or predictor variables) is typically spread out over multiple columns. Often there are not columns that include data on the time predictor; instead this information is typically embedded in the column name.

The vocabulary data is currently structured as wide data; the vocabulary scores are organized into four separate columns and the information about grade-level (the time predictor) is embedded in the variable names (e.g., `vocab_08` indicates 8th-grade). The same data are presented below in the tidy/long structure.

Table 1: Vocabulary Data Presented in the Tidy/Long Format

id	female	grade	vocab_score
1	1	vocab_08	1.75
1	1	vocab_09	2.60
1	1	vocab_10	3.76
1	1	vocab_11	3.68
2	0	vocab_08	0.90
2	0	vocab_09	2.47
2	0	vocab_10	2.44
2	0	vocab_11	3.43
3	0	vocab_08	0.80
3	0	vocab_09	0.93
3	0	vocab_10	0.40
3	0	vocab_11	2.27

Notice that in the tidy/long structured data that the vocabulary scores (outcome) are now organized into a single column. Grade-level (the time predictor) is also now explicitly included in the data and is also organized as a single column in the data. Note that in the long structure, each row now represents a particular student at a particular grade-level, and that each student's data now encompasses several rows.

There are advantages to each of the structures. For example the wide structure has the advantage of being a better structure for data entry. Since each row corresponds to a different student, there are fewer rows and therefore less redundancy in the data entry process. Compare this to the tidy/long data where each student's data encompasses four rows. If you were doing data entry in this structure you would need to record the student's sex four times rather than once in the wide structure.

The tidy/long structure is the structure that is needed for modeling. Thus, if one of the analytic goals is to fit a linear mixed-effects model to explain variation or examine predictor effects, the tidy/long data structure is key. Note that the wide structured data is also used in some analyses (e.g., computing correlations).

Switching between the Two Data Structures

The library **tidyr** has two functions, `gather()` (wide → tidy/long) and `spread()` (tidy/long → wide), that convert data between these two structures. Below, I show the code for going from the wide structured data (`vocabulary`) to the tidy/long structure.

```
# Convert from wide to long structured data
vocabulary_long = vocabulary %>%
  gather(key = "grade", value = "vocab_score", vocab_08:vocab_11) %>%
  arrange(id, grade)

# View data
head(vocabulary_long, 12)
```

```
# A tibble: 12 x 4
      id female grade    vocab_score
   <dbl>  <dbl> <chr>          <dbl>
 1     1      1 vocab_08        1.75
 2     1      1 vocab_09        2.6
 3     1      1 vocab_10        3.76
 4     1      1 vocab_11        3.68
 5     2      0 vocab_08        0.9
 6     2      0 vocab_09        2.47
 7     2      0 vocab_10        2.44
 8     2      0 vocab_11        3.43
 9     3      0 vocab_08        0.8
10     3      0 vocab_09        0.93
11     3      0 vocab_10        0.4
12     3      0 vocab_11        2.27
```

For more infomation about using these functions, google "tidyr" and read through any number of great tutorials or vignettes; for example here or here. You can also read Hadley Wickham's original paper on tidy data.

Exploration: Plot of the Mean and Individual Profiles

There are two plots that are particularly useful in exploring longitudinal data. The first is a plot of the mean value of the outcome at each time point (mean profile plot). This shows the average growth profile and is useful for determining the functional form of the fixed-effects part of the model; is the mean change over time linear? Quadratic? Log-linear? Another plot that is often examined is a spaghetti plot. A spaghetti plot shows the individual growth patterns or profiles and is useful for determining whether there is variation from the average profile. This helps us to consider the set of random-effects to include in the model. Below we examine both the mean profile and individual profiles simulateneously.

```
ggplot(data = vocabulary_long, aes(x = grade, y = vocab_score)) +
  geom_line(aes(group = id), alpha = 0.3) +                      #Add individual profiles
  stat_summary(fun.y = mean, geom = "line", size = 2, group = 1) + #Add mean profile line
  stat_summary(fun.y = mean, geom = "point", size = 3) +         #Add mean profile points
  theme_bw() +
  scale_x_discrete(
    name = "Grade-level",
    labels = c("8th-grade", "9th-grade", "10th-grade", "11th-grade")
    ) +
  ylab("Vocabulary score")
```

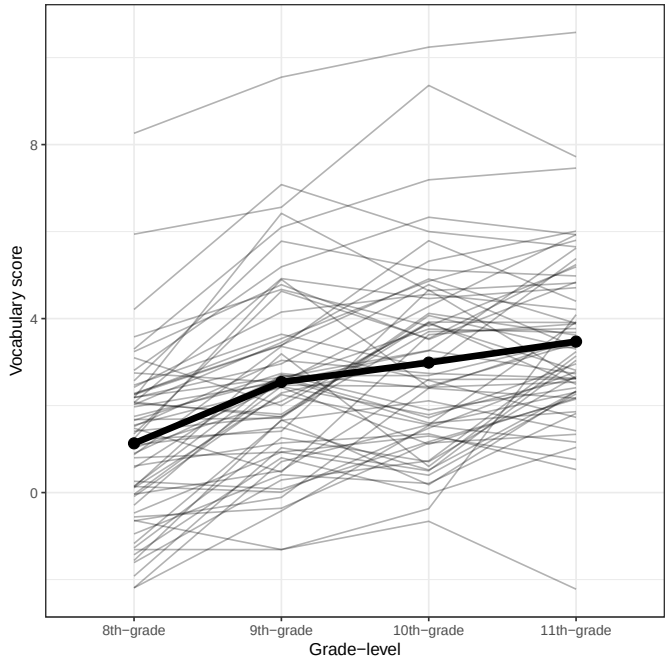

Figure 1: Plot showing the change in vocabulary score over time for 64 students. The average growth profile is also displayed.

Based on this plot:

- The average profile displays change over time that is positive (growth) and linear (or perhaps log-linear).
- The individual profiles show variation from the average profile; they have different vocabulary scores in 8th-grade and the profiles themselves vary in terms of their change (some show more change; others show decline)

Modeling: Unconditional Random Intercepts Model

As in a cross-sectional analysis we begin a longitudinal analysis by fitting the unconditional random intercepts model. The statistical model in this example can be expressed as:

$$\text{Vocabulary Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \epsilon_{ij}$$

where,

- $\text{Vocabulary Score}_{ij}$ is the vocabulary score at time point i for student j;
- β_0 is the fixed-effect of intercept;
- b_{0j} is the random-effect of intercept for student j; and
- ϵ_{ij} is the error at time point i for student j.

```
lmer.0 = lmer(vocab_score ~ 1 + (1|id), data = vocabulary_long, REML = FALSE)
summary(lmer.0)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + (1 | id)
   Data: vocabulary_long

     AIC      BIC   logLik deviance df.resid
    1015     1026     -505     1009      253

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2549 -0.6783  0.0471  0.6431  2.4923

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 2.95     1.72
 Residual             1.83     1.35
Number of obs: 256, groups:  id, 64

Fixed effects:
            Estimate Std. Error t value
(Intercept)    2.533      0.231      11
```

Fitting the unconditional means model gives us our baseline comparison model. The variance components suggest that there is unexplained within-student variation ($\hat{\sigma}^2_\epsilon = 1.83$) and unexplained between-student variation ($\hat{\sigma}^2_{\text{ID}} = 2.95$). Most of the unexplained variation seems to be between-student variation (61.8%).

Modeling: Unconditional Growth Model

We can now add the fixed-effect of time (the time predictor) to the model. In this data set, the time predictor is `grade`, which is a categorical predictor. We could create dummy variables, or simply add `grade` into the model and let R choose the reference group alphabetically (`vocab_08` in this example). The statistical model in this example can be expressed as:

$$\text{Vocabulary Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{9th-grade}_{ij}) + \beta_2(\text{10th-grade}_{ij}) + \beta_3(\text{11th-grade}_{ij}) + \epsilon_{ij}$$

where,

- $\text{Vocabulary Score}_{ij}$ is the vocabulary score at time point i for student j;
- β_0 is the fixed-effect of intercept;
- b_{0j} is the random-effect of intercept for student j;
- 9th-grade_{ij}, 10th-grade_{ij}, and 11th-grade_{ij} are dummy coded variable indicating grade-level,
- β_1 is the effect of 9th-grade (i.e., mean vocabulary score difference between 8th- and 9th-grade),
- β_2 is the effect of 10th-grade (i.e., mean vocabulary score difference between 8th- and 10th-grade),
- β_3 is the effect of 11th-grade (i.e., mean vocabulary score difference between 8th- and 11th-grade), and
- ϵ_{ij} is the error at time point i for student j.

Fitting the model:

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + grade + (1 | id)
   Data: vocabulary_long

     AIC      BIC   logLik deviance df.resid
     865      886     -426      853      250

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.2414 -0.6546  0.0426  0.6122  2.8757

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 3.206    1.791
 Residual             0.808    0.899
Number of obs: 256, groups:  id, 64

Fixed effects:
              Estimate Std. Error t value
(Intercept)      1.132      0.250    4.52
gradevocab_09    1.410      0.159    8.87
gradevocab_10    1.857      0.159   11.68
gradevocab_11    2.340      0.159   14.72

Correlation of Fixed Effects:
            (Intr) grd_09 grd_10
gradevcb_09 -0.317
gradevcb_10 -0.317  0.500
gradevcb_11 -0.317  0.500  0.500
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = 1.13 + 1.41(\text{9th-grade}_{ij}) + 1.86(\text{10th-grade}_{ij}) + 2.34(\text{11th-grade}_{ij})$$

Interpreting the coefficients,

- The predicted average vocabulary score for 8th-grade students (intercept) is 1.13.
- On average, 9th-grade students have a vocabulary score that is 1.41-points higher than 8th-grade students.
- On average, 10th-grade students have a vocabulary score that is 1.86-points higher than 8th-grade students.
- On average, 11th-grade students have a vocabulary score that is 2.34-points higher than 8th-grade students.

Based on the *t*-statistics:

- The average 8th-grade vocabulary score is statistically different than 0 ($t = 4.52$).
- There is a statistical differences between the average 8th-grade vocabulary score and the average 9th-grade vocabulary score ($t = 8.87$).
- There is a statistical differences between the average 8th-grade vocabulary score and the average 10th-grade vocabulary score ($t = 11.68$).
- There is a statistical differences between the average 8th-grade vocabulary score and the average 11th-grade vocabulary score ($t = 14.72$).

Looking at the variance components:

- The model has explained 55.8% of the within-student variation. This is because grade is a within-student predictor (it has values that vary within each student).
- The model has *increased* the variation between-students (-8.7%). This is a mathematical artifact of the estimation process.

Likelihood Ratio Test: p-Values for Mixed-Effects Models

So long as the assumptions of the linear mixed-effects model have been met (see Assumptions notes), we can obtain a p-value for the effect of grade. The way we do this is by taking advanatage of the fact that the unconditional random intercepts model is a nested model of the unconditional growth model. If we have nested models, they can be compared using a *Likelihood Ratio Test*. To carry out this test, we use the `anova()` function and input the two mixed-effects models we want to compare. (Note: Both models need to be fitted with ML.)

```
anova(lmer.0, lmer.1)
```

```
Data: vocabulary_long
Models:
lmer.0: vocab_score ~ 1 + (1 | id)
lmer.1: vocab_score ~ 1 + grade + (1 | id)
       Df  AIC  BIC logLik deviance Chisq Chi Df          Pr(>Chisq)
lmer.0  3 1015 1026   -505     1009
lmer.1  6  865  886   -426      853   156      3 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis being tested is that the reduced model (`lmer.0`) and the full model (`lmer.1`) fit the data equally well. The way that we measure fit is via the deviance. The deviance of the reduced model is 1009.3 and that for the full model is 852.8. If the two models fit equally well, we would expect the difference in deviances to be zero. The actual difference in deviances is 156.46. (This is often referred to as ΔG^2, for goodness-of-fit, or as χ^2.) This indicates that the fuller model fits the sample data better than the reduced model; the full model has a smaller deviance.

As with any difference, we wonder whether this is within what would be expected because of sampling (chance) variation. To test this, we evaluate ΔG^2 in a χ^2-distribution with *df* equal to the difference in K between the two models (K is the *df* for each model). This difference should be the difference in the complexity between the two models; the difference in the estimated number of parameters. Our reduced model has three parameters being estimated ($\hat{\beta}_0$, $\hat{\sigma}^2_\epsilon$, and $\hat{\sigma}^2_0$), and our full model has six parameters being estimated ($\hat{\beta}_0$, $\hat{\beta}_{\text{9th-grade}}$, $\hat{\beta}_{\text{10th-grade}}$, $\hat{\beta}_{\text{11th-grade}}$, $\hat{\sigma}^2_\epsilon$, and $\hat{\sigma}^2_0$). The difference in complexity between these models is $6 - 3 = 3$.

```
1 - pchisq(156.46, df = 3)
```

```
[1] 0
```

Note that all of these results are given in the `anova()` output. This is typically reported as something like:

> A likelihood ratio test indicated that the model that included the fixed-effects of grade-level fitted the data significantly better that the unconditional random intercepts model, $\chi^2(3) = 156.46$, $p < .001$. This suggests that there is an effect of time on average vocabulary scores.

Mathematics Behind the Likelihood Ratio Test

Why is this called a *likelihood ratio test*? Remember that the deviance is equal to $-2\ln(\mathcal{L})$. Thus the difference in deviances can be written as:

$$\Delta G^2 = -2\ln\big[\mathcal{L}(\text{Reduced Model})\big] - \Big[-2\ln\big[\mathcal{L}(\text{Full Model})\big]\Big]$$

Pulling out the -2 we get

$$\Delta G^2 = -2\Big[\ln\big[\mathcal{L}(\text{Reduced Model})\big] - \ln\big[\mathcal{L}(\text{Full Model})\big]\Big]$$

The difference between two logarithms, e.g., $\log(A) - \log(B)$ is the logarithm of the quotient ($\log(\frac{A}{B})$). Thus, we can re-write this as,

$$\Delta G^2 = -2\ln\left[\frac{\mathcal{L}(\text{Reduced Model})}{\mathcal{L}(\text{Full Model}}\right]$$

Now it should be a little more apparent why this test is called a likelihood ratio test. Note that if both models fit the data equally well, their likelihood values would be equivalent and thus this equation would reduce to:

$$\begin{aligned}\Delta G^2 &= -2\ln\Big[1\Big] \\ &= -2(0) \\ &= 0\end{aligned}$$

Thus if the difference in the goodness-of-fit between the two models turns out to be zero (or are within chance variation of zero), both models fit the data equally and thus we should adopt the reduced model (Occam's Razor).

Quantitative Time Predictor: A More Flexible Model for Repeated Measures Data

One advantage to using the linear mixed-effects model to analyze repeated measures data over traditional methods (e.g., RM-ANOVA or MANOVA) is that the regression model allows for both categorical and quantitative variables. For example, rather than code our grade-levels categorically (as `vocab_08`, `vocab_09`, `vocab_10` and `vocab_11`), which was a necessity in days of yore, we could have simply coded them as 8, 9, 10, and 11. Then we could have fitted the LME model using this quantitative predictor. The statistical model when time is quantitative would be:

$$\text{Vocabulary Score}_{ij} = \big[\beta_0 + b_{0j}\big] + \beta_1(\text{Grade}_{ij}) + \epsilon_{ij}$$

where,

- $\text{Vocabulary Score}_{ij}$ is the vocabulary score at time point i for student j;
- β_0 is the fixed-effect of intercept;
- b_{0j} is the random-effect of intercept for student j;
- Grade_{ij} is a quantitative variable indicating grade-level,
- β_1 is the effect of a one-unit change in grade, and
- ϵ_{ij} is the error at time point i for student j.

This is still referred to as the *unconditional growth model* since the only predictor is a fixed-effect of time.

Lookup Table: Mapping Categories to Quantities

To convert grade to a quantitative variable, we create a **lookup table** which maps the levels of the categorical time predictor to the values we want to use in our new quantitative predictor. Below I show this mapping for two quantitative predictors, grade_quant which is a straight mapping to the relevant grade-level and grade_quant_center which centers the grade_quant predictor by subtracting 8 from each value.

```
# Create lookup table
lookup_table = data.frame(
  grade = c("vocab_08", "vocab_09", "vocab_10", "vocab_11"),
  grade_quant = c(8, 9, 10, 11),
  grade_quant_center = c(0, 1, 2, 3)
)

# View lookup table
lookup_table
```

```
     grade grade_quant grade_quant_center
1 vocab_08           8                  0
2 vocab_09           9                  1
3 vocab_10          10                  2
4 vocab_11          11                  3
```

Then, we join the tidy/long data with the lookup table.

```
vocabulary_long_2 = left_join(vocabulary_long, lookup_table, by = "grade")
head(vocabulary_long_2)
```

```
# A tibble: 6 x 6
     id female grade    vocab_score grade_quant grade_quant_center
  <dbl>  <dbl> <chr>          <dbl>       <dbl>              <dbl>
1     1      1 vocab_08        1.75           8                  0
2     1      1 vocab_09        2.6            9                  1
3     1      1 vocab_10        3.76          10                  2
4     1      1 vocab_11        3.68          11                  3
5     2      0 vocab_08        0.9            8                  0
6     2      0 vocab_09        2.47           9                  1
```

Fitting the Unconditional Growth Model with a Quantitative Time Predictor

Below we fit the linear mixed-effects model using the `grade_quant` predictor.

```
lmer.2 = lmer(vocab_score ~ 1 + grade_quant + (1|id), data = vocabulary_long_2, REML = FALSE)
summary(lmer.2)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + grade_quant + (1 | id)
   Data: vocabulary_long_2

     AIC      BIC   logLik deviance df.resid
     881      895     -436      873      252

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.347 -0.657 -0.028  0.644  2.655

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 3.184    1.784
 Residual             0.896    0.947
Number of obs: 256, groups:  id, 64

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -4.5600     0.5532   -8.24
grade_quant   0.7467     0.0529   14.11

Correlation of Fixed Effects:
            (Intr)
grade_quant -0.909
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = -4.56 + 0.75(\text{Grade-level}_{ij})$$

The model using the quantitative predictor of grade-level is simpler than the model using the categorical version of grade-level since it has two fewer fixed-effects to estimate (fewer model degrees-of-freedom).

Interpreting the coefficients,

- The predicted average vocabulary score for 0th-grade students (intercept) is -4.55 (extrapolation).
- Each one-unit difference in grade-level is associated with a 0.75-point difference in vocabulary score, on average.

Looking at the variance components and comparing them to the unconditional random intercepts model:

- The unconditional growth model has explained 50.8% of the within-student variation.
- The unconditional growth model has *increased* the variation between-students (-7.8%). This is a mathematical artifact of the estimation process.

This is similar to the variance components obtained from the model using the categorical predictors of grade level.

Centering the Time Predictor: Better Interpretations of the Intercept

Now, let's fit the model using the centered quantitative predictor.

```
lmer.3 = lmer(vocab_score ~ 1 + grade_quant_center + (1|id), data = vocabulary_long_2, REML = FALSE)
summary(lmer.3)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + grade_quant_center + (1 | id)
   Data: vocabulary_long_2

     AIC      BIC   logLik deviance df.resid
     881      895     -436      873      252

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.347 -0.657 -0.028  0.644  2.655

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 3.184    1.784
 Residual             0.896    0.947
Number of obs: 256, groups:  id, 64

Fixed effects:
                   Estimate Std. Error t value
(Intercept)          1.4133     0.2440    5.79
grade_quant_center   0.7467     0.0529   14.11

Correlation of Fixed Effects:
            (Intr)
grd_qnt_cnt -0.325
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = 1.41 + 0.75(\text{Centered grade-level}_{ij})$$

Interpreting the coefficients,

- The predicted average vocabulary score for 8th-grade students is 1.41. Centering removes the problem of extrapolation in the interpretation because we have now made 0 a legitimate value in the predictor.
- Each one-unit difference in grade-level is associated with a 0.75-point difference in vocabulary score, on average. This is identical to the previous model since we have not changed what a one-unit difference in the predictor represents.

We can see why the intercepts are different but the slopes are the same by comparing the plots of the individual growth profiles and the fitted fixed-effects models for the two predictors.

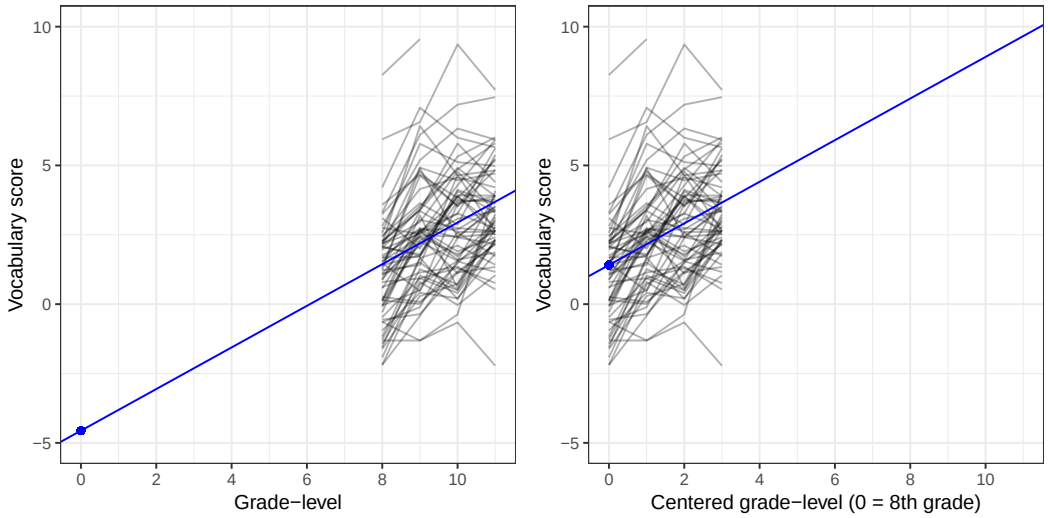

Figure 2: Plot showing the change in vocabulary score over time for 64 students. The average growth profile is also displayed. This is shown for the non-centered (left) and 8th-grade centered (right) grade-level. A large blue point is shown at the intercept value in both plots.

Looking at the variance components and comparing them to the unconditional random intercepts model:

- The model has explained 50.8% of the within-student variation.
- The model has *increased* the variation between-students (−7.8%). This is a mathematical artifact of the estimation process.

These values are identical to the variance components obtained from the previous model.

> Because of the interpretive value of the intercept when we center the grade-level predictor, we will fit all future models using the 8th-grade centered grade-level.

Examining the Functional Form of the Growth Model

As in any regression analysis. we need to specify the functional form of the growth model. Below we consider three potential functional forms between grade-level and vocabulary score: (1) a linear relationship (`lmer.3`); (2) a quadratic relatinship; and (3) a log-linear relationship (based on log-transforming grade-level).

```
# Quadratic model
lmer.4 = lmer(vocab_score ~ 1 + grade_quant_center + I(grade_quant_center^2) + (1|id),
              data = vocabulary_long_2, REML = FALSE)

# Log-linear model
lmer.5 = lmer(vocab_score ~ 1 + log(grade_quant_center + 1) + (1|id),
              data = vocabulary_long_2, REML = FALSE)
```

```r
# Model-evidence
aictab(
  cand.set = list(lmer.3, lmer.4, lmer.5),
  modnames = c("Linear", "Quadratic", "Log-linear")
)
```

```
Model selection based on AICc:

           K AICc Delta_AICc AICcWt Cum.Wt   LL
Log-linear 4  864       0.00   0.79   0.79 -428
Quadratic  5  867       2.67   0.21   1.00 -428
Linear     4  881      16.57   0.00   1.00 -436
```

Given the data and candidate models, the evidence supports the log-linear model. There is some slight evidence for the quadratic model and almost no evidence for the linear model. This is consistent with the nonlinearity we observed in the mean profile earlier. We should also evaluate the residuals for both models.

```r
out_4 = augment(lmer.4)
out_5 = augment(lmer.5)


# Log-linear model
ggplot(data = out_5, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals") +
  ggtitle("Log-linear")

sm.density(out_5$.resid, model = "normal", main = "Main-Effect", xlab = "Level-1 residuals")
sm.density(ranef(lmer.5)$id[ , 1], model = "normal", xlab = "Random effects of the intercept")


# Quadratic model
ggplot(data = out_4, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals") +
  ggtitle("Quadratic")

sm.density(out_4$.resid, model = "normal", main = "Interaction Effect", xlab = "Level-1 residuals")
sm.density(ranef(lmer.4)$id[ , 1], model = "normal", xlab = "Random effects of the intercept")
```

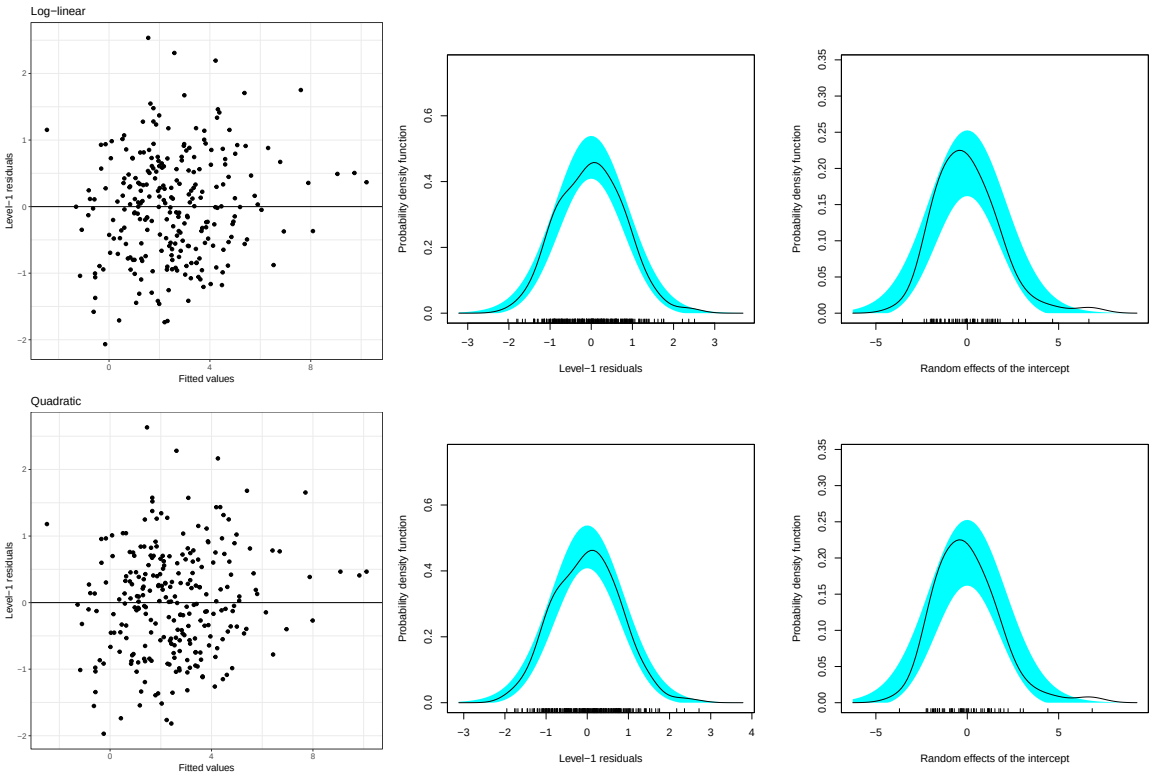

Figure 3: Plots of the Level-1 residuals and random-effects for the log-linear model (Row 1) and the quadratic model (Row 2).

The residual plots look similar indicating that neither model meets the assumptions better than the other. Given this, the higher evidence for the log-linear model, and the simplicity of the log-linear model relative to the quadratic model, we will adopt the log-linear growth profile.

Examining the Output for the Log-Linear Fitted Model

```
summary(lmer.5)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + log(grade_quant_center + 1) + (1 | id)
   Data: vocabulary_long_2

     AIC      BIC   logLik deviance df.resid
     864      878     -428      856      252

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.279 -0.648 -0.001  0.620  2.796

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 3.202    1.790
 Residual             0.822    0.907
Number of obs: 256, groups:  id, 64

Fixed effects:
                           Estimate Std. Error t value
(Intercept)                   1.210      0.246    4.91
log(grade_quant_center + 1)   1.666      0.109   15.30

Correlation of Fixed Effects:
            (Intr)
lg(grd__+1) -0.351
```

The fitted equation is:

$$\hat{\text{Vocabulary Score}}_{ij} = 1.21 + 1.66\Big[\ln(\text{Centered grade-level}_{ij} + 1)\Big]$$

Interpreting the coefficients,

- The predicted average vocabulary score for 8th-grade students is 1.21.
- Each one-percent difference in grade-level is associated with a 0.0166-point difference in vocabulary score, on average.

Looking at the variance components:

- The model has explained 54.9% of the within-student variation.
- The model has *increased* the variation between-students (-8.4%). This is a mathematical artifact of the estimation process.

These values are quite similar to the variance components obtained from the previous model.

Plot of the Fixed-Effects Part of the Model

The t-value associated with the fixed-effect of grade-level ($t = 15.31$) indicates that the log-linear relationship between grade-level and vocabulary score is statistically important. To better understand this relationship we can plot the fixed-effects part of the unconditional growth model.

```
# Set up data
plot_data = crossing(
  grade_quant_center = seq(from = 0, to = 3, by = 0.01)
) %>%
  mutate(
    yhat = predict(lmer.5, newdata = ., re.form = NA)
  )

head(plot_data)
```

```
# A tibble: 6 x 2
  grade_quant_center  yhat
               <dbl> <dbl>
1               0     1.21
2               0.01  1.23
3               0.02  1.24
4               0.03  1.26
5               0.04  1.27
6               0.05  1.29
```

```
# Create plot
ggplot(data = plot_data, aes(x = grade_quant_center, y = yhat)) +
  geom_line() +
  theme_bw() +
  scale_x_continuous(
    name = "Grade-level",
    breaks = c(0, 1, 2, 3),
    labels = c(8, 9, 10, 11)
    ) +
  ylab("Vocabulary score")
```

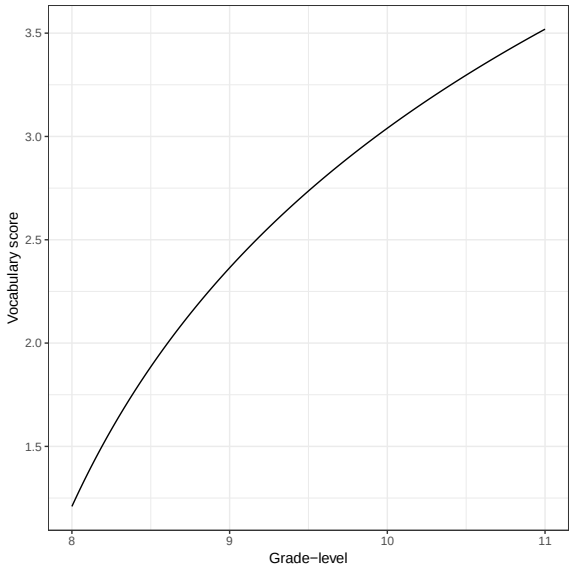

Figure 4: Predicted change in vocabulary score as a function of grade-level.

Based on this and the coefficient-level output, we can answer the first research question.

> The growth pattern in vocabulary is log-linear over time. While the change in vocabulary score, on average, is positive, the growth rate somewhat diminishes over time.

Examining the Male and Female Growth Profiles

To answer the second research question about whether the growth pattern is different for males and females, we again plot the individual and mean growth profiles for females and males.

```
# Turn female into factor for better plotting
vocabulary_long_2 %>%
  mutate(
    Sex = factor(female, levels = c(0, 1), labels = c("Male", "Female"))
  ) %>%
ggplot(aes(x = grade_quant, y = vocab_score, color = Sex)) +
  geom_line(aes(group = id), alpha = 0.3) +
  stat_summary(fun.y = mean, geom = "line", size = 2, group = 1) +
  stat_summary(fun.y = mean, geom = "point", size = 3) +
  theme_bw() +
  xlab("Grade-level") +
  ylab("Vocabulary score") +
  facet_wrap(~Sex) +
  ggsci::scale_color_d3()
```

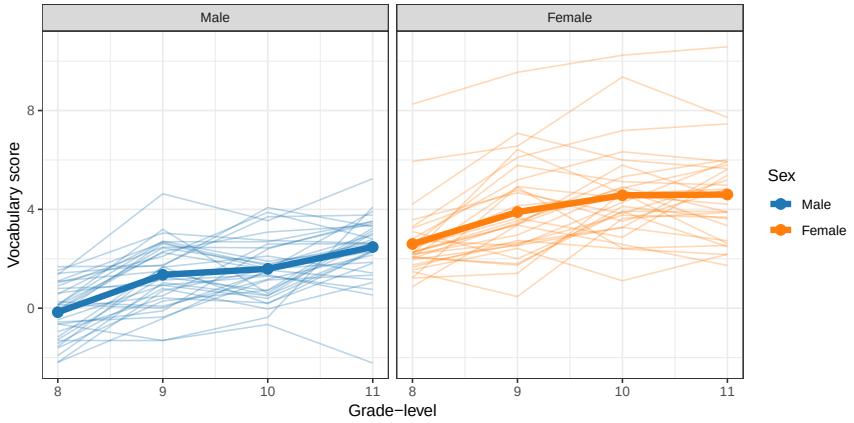

Figure 5: Plot showing the change in vocabulary score over time conditioned on sex. The average growth profile is also displayed for each sex.

Examining the average profiles for males and females in this plot suggests that the females tend to have higher average vocabulary scores than males at each grade level. The sample average growth profiles also show slightly different patterns of growth between males and females. To examine whether this is due to chance, we can fit a set of models that include fixed-effects of sex in addition to the grade-level predictor.

Here we fit a model that includes a main-effect of sex and another that include both the main-effect of sex and the interaction-effect between sex and grade-level.

```
# Main-effect of sex
lmer.6 = lmer(vocab_score ~ 1 + log(grade_quant_center + 1) + female +
                (1|id), data = vocabulary_long_2, REML = FALSE)

# Interaction-effect between sex and grade-level
lmer.7 = lmer(vocab_score ~ 1 + log(grade_quant_center + 1) + female + log(grade_quant_center + 1):female +
                (1|id), data = vocabulary_long_2, REML = FALSE)
```

These two models allow us to test different hypotheses about the patterns of growth between males and females:

- If there is a main-effect of sex, it will allow us to conclude that the growth pattern is the same, but that the average females vocabulary score is systematically different than that for males at each time point (e.g., always lower or higher by the same amount).
- If there is an interaction-effect between sex and grade-level, it will allow us to conclude that the pattern of change over time is different between males and females.

To facilitate model selection we will examine a table of model evidence using the following candidate models: (1) the unconditional growth model, (2) the model that includes main-effects of grade-level and sex, and (3) the model that includes an interaction-effect between grade-level and sex.

```
# Model-evidence
aictab(
  cand.set = list(lmer.5, lmer.6, lmer.7),
  modnames = c("Unconditional growth", "Main-effect of sex", "Interaction-effect")
)
```

```
Model selection based on AICc:

                     K AICc Delta_AICc AICcWt Cum.Wt   LL
Main-effect of sex   5  823       0.00   0.58   0.58 -406
Interaction-effect   6  823       0.61   0.42   1.00 -406
Unconditional growth 4  864      41.54   0.00   1.00 -428
```

Here the evidence slightly favors the model that includes the main-effect of sex. There is also a fair bit of evidence to support the interaction model. Again, we should probably examine the residuals from both of these models and adopt the model that better meets the assumptions.

```
out_6 = augment(lmer.6)
out_7 = augment(lmer.7)


# Main-effect model
ggplot(data = out_6, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals") +
  ggtitle("Main-Effect")

sm.density(out_6$.resid, model = "normal", main = "Main-Effect", xlab = "Level-1 residuals")
sm.density(ranef(lmer.6)$id[ , 1], model = "normal", xlab = "Random effects of the intercept")


# Interaction model
ggplot(data = out_7, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals") +
  ggtitle("Interaction Effect")

sm.density(out_7$.resid, model = "normal", main = "Interaction Effect", xlab = "Level-1 residuals")
sm.density(ranef(lmer.7)$id[ , 1], model = "normal", xlab = "Random effects of the intercept")
```

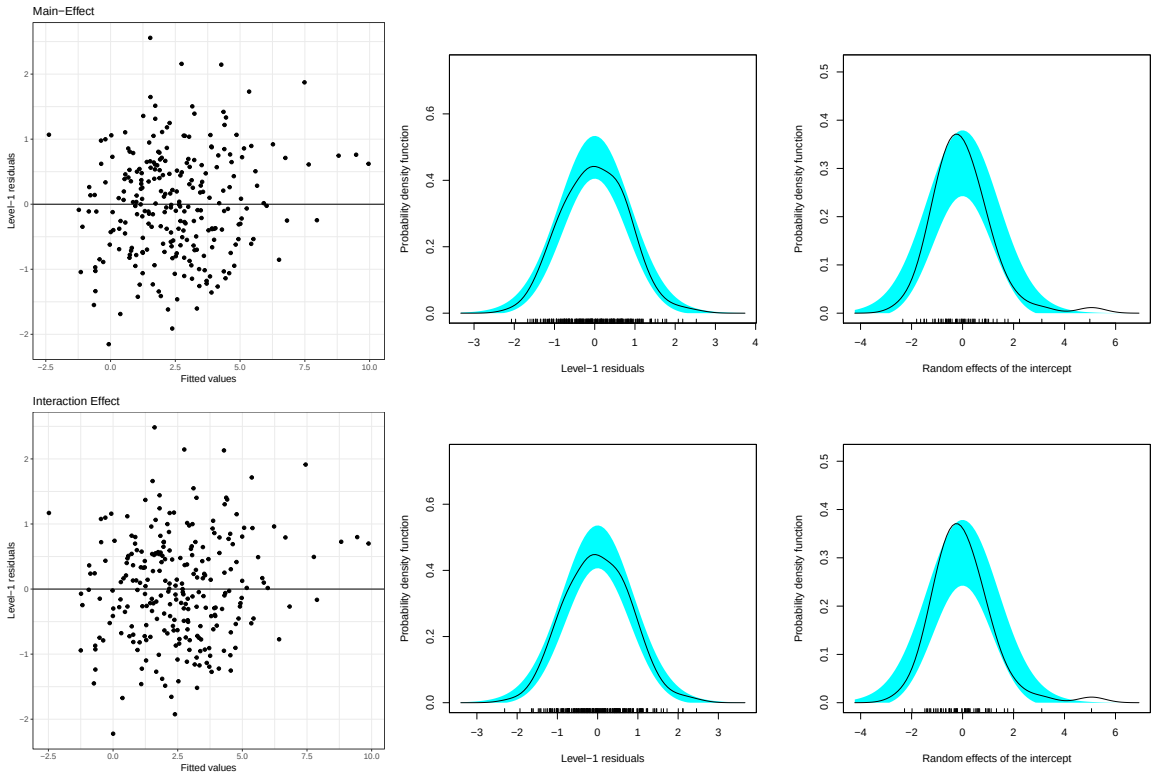

Figure 6: Plots of the Level-1 residuals and random-effects for the main-effects model (Row 1) and the interaction model (Row 2).

The residuals for both models look fairly reasonable. Neither model shows improved fit based on these plots.

Testing Hypotheses about Sex Differences in the Growth Profiles

Since the unconditional growth model is nested in the growth model with the main-effect of `female` and that, in turn, is nested inside the interaction model, we could have also used a series of likelihood ratio tests to evaluate these models.

```
# LRT
anova(lmer.5, lmer.6, lmer.7)
```

```
Data: vocabulary_long_2
Models:
lmer.5: vocab_score ~ 1 + log(grade_quant_center + 1) + (1 | id)
lmer.6: vocab_score ~ 1 + log(grade_quant_center + 1) + female + (1 |
lmer.6:     id)
lmer.7: vocab_score ~ 1 + log(grade_quant_center + 1) + female + log(grade_quant_center +
lmer.7:     1):female + (1 | id)
       Df AIC BIC logLik deviance Chisq Chi Df    Pr(>Chisq)
lmer.5  4 864 878   -428      856
lmer.6  5 823 840   -406      813 43.62      1 0.00000000004 ***
lmer.7  6 823 844   -406      811  1.49      1          0.22
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first χ^2-test (on the `lmer.6` line) compares the growth model that includes the main-effect of sex to the unconditional growth model (the model on the previous line). This result is statistically significant, $\chi^2(1) = 43.62$, $p < 0.001$. This suggests that the model that includes the main-effect of sex has significantly less error than the unconditional growth model.

The second χ^2-test (on the `lmer.7` line) compares the growth model that includes the interaction-effect between sex and grade-level to the growth model that includes the main-effect of sex. This result is not statistically significant, $\chi^2(1) = 1.49$, $p = 0.222$. This suggests that the model that includes the interaction-effect between sex and grade-level does not have significantly less error than the model that includes the main-effect of sex.

This series of tests suggests that of these models, we should adopt the model that includes the main-effect of sex.

As a reminder, the results from evaluating models using a *p*-value approach and those employing a model evidence approach are not always consistent with each other. This is because they represent two very different philosophical approaches to model selection. As such, the decision about how you will make decisions about model adoption (*p*-value vs. model evidence) should be decided prior to carrying out any data analysis.

Using the results from evaluating the table of model evidence, there is evidence to support both the models that include the main-effects of grade-level and sex, as well as the model that includes the interaction-effect between grade-level and sex. We will interpret the coefficients and variance components from each of these models, and also plot their fixed-effects.

Main-Effect of Sex

```
summary(lmer.6)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: vocab_score ~ 1 + log(grade_quant_center + 1) + female + (1 |
    id)
   Data: vocabulary_long_2

     AIC      BIC   logLik deviance df.resid
     822      840     -406      812      251

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.3728 -0.6708 -0.0194  0.6392  2.8203

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 1.518    1.232
 Residual             0.822    0.907
Number of obs: 256, groups:  id, 64

Fixed effects:
                            Estimate Std. Error t value
(Intercept)                 -0.00951    0.24121   -0.04
log(grade_quant_center + 1)  1.66615    0.10886   15.30
female                       2.60057    0.32888    7.91

Correlation of Fixed Effects:
            (Intr) l(__+1
lg(grd__+1) -0.359
female      -0.639  0.000
```

The fitted equation is:

$$\hat{\text{Vocabulary Score}}_{ij} = -0.01 + 1.67\Big[\ln(\text{Centered grade-level}_{ij} + 1)\Big] + 2.60(\text{Female.}_j)$$

Interpreting the coefficients,

- The predicted average vocabulary score for male 8th-grade students is -0.01.
- Each one-percent difference in grade-level is associated with a 0.0167-point difference in vocabulary score, on average, controlling for differences in sex.
- Females have an average vocabulary score that is 2.60-points higher than males, controlling for differences in grade-level.

Looking at the variance components:

- The model has explained 54.9% of the within-student variation.
- The model has explained 16.8% of the between-student variation.

We expect the model to explain variation between-students as the `female` predictor we included was a between-students predictor. Plotting this model (see below) we find that the growth pattern in vocabulary is log-linear over time for both females and males. While the change in vocabulary score, on average, is positive for both sexes, the growth rate somewhat diminishes over time. Moreover, while females tend to have a higher vocabulary score at each grade level, the change patterns seem to have the same rate of growth for both sexes.

```
# Set up data
plot_data = crossing(
  grade_quant_center = seq(from = 0, to = 3, by = 0.01),
  female = c(0, 1)
) %>%
  mutate(
    yhat = predict(lmer.6, newdata = ., re.form = NA),
    Sex = factor(female, levels = c(0, 1), labels = c("Male", "Female"))
  )

head(plot_data)
```

```
# A tibble: 6 x 4
  grade_quant_center female     yhat Sex
               <dbl>  <dbl>    <dbl> <fct>
1               0         0 -0.00951 Male
2               0         1  2.59    Female
3               0.01      0  0.00707 Male
4               0.01      1  2.61    Female
5               0.02      0  0.0235  Male
6               0.02      1  2.62    Female
```

```
# Create plot
ggplot(data = plot_data, aes(x = grade_quant_center, y = yhat, color = Sex, linetype = Sex)) +
  geom_line() +
  theme_bw() +
  scale_x_continuous(name = "Grade-level", breaks = c(0, 1, 2, 3), labels = c(8, 9, 10, 11)) +
  ylab("Vocabulary score") +
  ggsci::scale_color_d3()
```

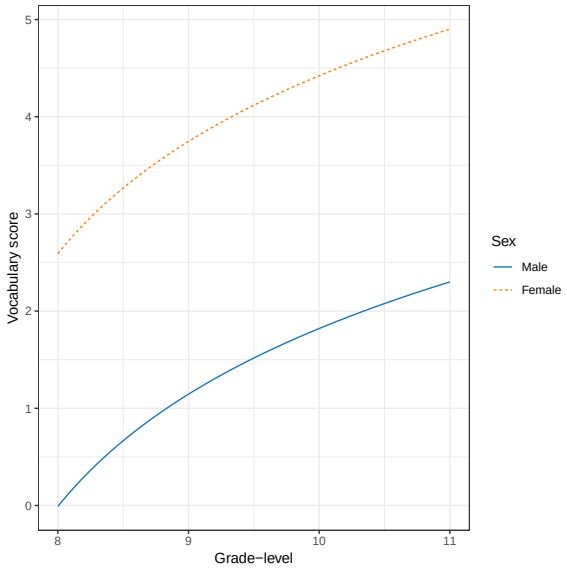

Figure 7: Predicted change in vocabulary score as a function of grade-level and sex.

Interaction-Effect between Sex and Grade-Level

```
summary(lmer.7)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula:
vocab_score ~ 1 + log(grade_quant_center + 1) + female + log(grade_quant_center +
    1):female + (1 | id)
   Data: vocabulary_long_2

     AIC      BIC   logLik deviance df.resid
     823      844     -406      811      250

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4609 -0.6316 -0.0285  0.6069  2.7505

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 1.520    1.233
 Residual             0.816    0.903
Number of obs: 256, groups:  id, 64

Fixed effects:
                                Estimate Std. Error t value
(Intercept)                       -0.109      0.254   -0.43
log(grade_quant_center + 1)        1.791      0.149   12.04
female                             2.812      0.371    7.57
log(grade_quant_center + 1):female -0.266      0.217   -1.22

Correlation of Fixed Effects:
```

```
            (Intr) lg(__+1) female
lg(grd__+1) -0.465
female      -0.685  0.318
lg(gr__+1):  0.318 -0.685   -0.465
```

The fitted equation is:

$$\hat{\text{Vocabulary Score}}_{ij} = -0.11 + 1.79\Big[\ln(\text{Centered grade-level}_{ij}+1)\Big] + 2.81(\text{Female.}_j) - 0.27\Big[\ln(\text{Centered grade-level}_{ij}+1)\times(\text{Fem}$$

Interpreting the coefficients ,

- The predicted average vocabulary score for male 8th-grade students is -0.11.
- For males, each one-percent difference in grade-level is associated with a 0.0179-point difference in vocabulary score, on average.
- Eighth-grade females have an average vocabulary score that is 2.81-points higher than 8th-grade males.
- For females, each one-percent difference in grade-level is associated with a 0.0152-point difference in vocabulary score, on average. This is less than the effect for males by 0.0027.

More generally, we might say: - The effect of grade-level differs by sex. - The effect of sex differs by grade-level

Looking at the variance components:

- The model has explained 55.3% of the within-student variation.
- The model has explained 16.7% of the between-student variation.

Plotting this model (see below; syntax not displayed) we find that the growth pattern in vocabulary is log-linear over time for both females and males. While the change in vocabulary score, on average, is positive for both sexes, the growth rate somewhat diminishes over time. Moreover, while females tend to have a higher vocabulary score at each grade level, the growth rate for females is slightly smaller than that for males.

```
# Set up data
plot_data_2 = crossing(
  grade_quant_center = seq(from = 0, to = 3, by = 0.01),
  female = c(0, 1)
) %>%
  mutate(
    yhat = predict(lmer.7, newdata = ., re.form = NA),
    Sex = factor(female, levels = c(0, 1), labels = c("Male", "Female"))
  )

head(plot_data_2)
```

```
# A tibble: 6 x 4
  grade_quant_center female    yhat Sex
               <dbl>  <dbl>   <dbl> <fct>
1               0         0 -0.109  Male
2               0         1  2.70   Female
3               0.01      0 -0.0907 Male
4               0.01      1  2.72   Female
5               0.02      0 -0.0731 Male
6               0.02      1  2.73   Female
```

```r
# Create plot
ggplot(data = plot_data_2, aes(x = grade_quant_center, y = yhat, color = Sex, linetype = Sex)) +
  geom_line() +
  theme_bw() +
  scale_x_continuous(name = "Grade-level", breaks = c(0, 1, 2, 3), labels = c(8, 9, 10, 11)) +
  ylab("Vocabulary score") +
  ggsci::scale_color_d3()
```

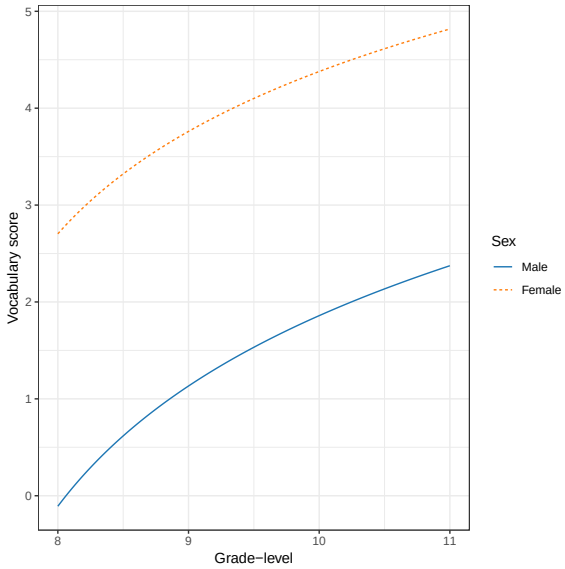

Figure 8: Predicted change in vocabulary score as a function of grade-level.

The similarity between the predicted values from the two models, and the fact that the residuals from the two models look quite similar might suggest that we should ultimately adopt the model that includes the main-effects of grade-level and sex over the interaction model.

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 longitudinal analysis. Here are some resources that may be helpful in that endeavor:

- Long, J. D. (2012). Longitudinal data analysis for the behavioral sciences using R. Thousand Oaks, CA: Sage.
- Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). Lasagna plots: A saucy alternative to spaghetti plots. *Epidemiology, 21*(5), 621–625.