Linear Mixed-Effects Models: Longitudinal Analysis

2020-03-11

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(broom)
library(educate)
library(lme4) #for fitting mixed-effects models
library(MuMIn)
library(patchwork)
library(tidyverse)

# 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 females and non-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** (loaded with **tidyverse**) has two functions, `pivot_longer()` (wide → tidy/long) and `pivot_wider()` (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 %>%
  pivot_longer(cols = vocab_08:vocab_11, names_to = "grade", values_to = "vocab_score") %>%
  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 information about using these functions, Google "tidyr pivot" and read through any number of great tutorials or vignettes; for example 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 simultaneously.

```
ggplot(data = vocabulary_long, aes(x = grade, y = vocab_score)) +
  geom_line(aes(group = id), alpha = 0.3) +                    #Add individual profiles
  stat_summary(fun = mean, geom = "line", size = 2, group = 1) + #Add mean profile line
  stat_summary(fun = 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. The Average Growth Profile is Displayed as a Thicker Line
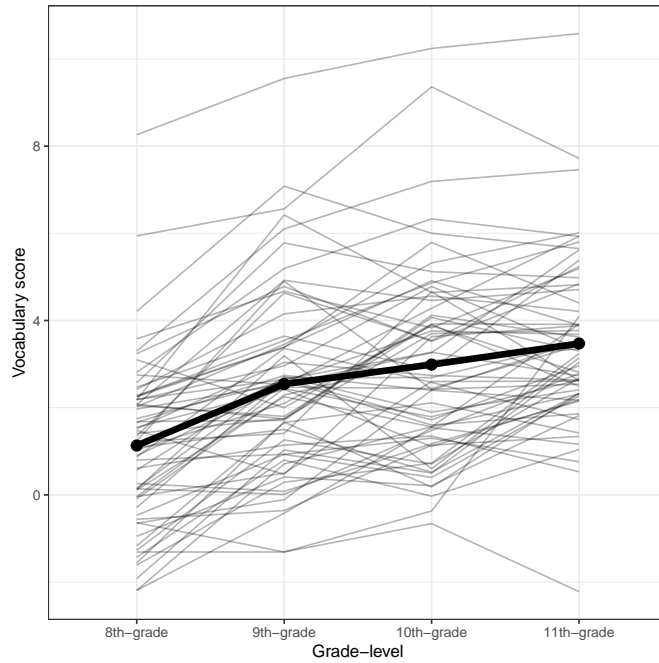
Based on this plot:

- 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:

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

where,

- 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.

```
# Fit unconditional random intercepts model
lmer.0 = lmer(vocab_score ~ 1 + (1|id), data = vocabulary_long, REML = FALSE)

# Coefficient-level output
tidy(lmer.0)
```

```
# A tibble: 3 x 5
  term                    estimate std.error statistic group
  <chr>                      <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                 2.53     0.231      11.0 fixed
2 sd_(Intercept).id           1.72     NA          NA  id
3 sd_Observation.Residual     1.35     NA          NA  Residual
```

```
# Compute variance components
# 1.35^2 #Residual
# 1.72^2 #Intercept
```

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:

$$\text{Vocabulary Score}_{ij} = [\beta_0 + b_{0j}] + \beta_1(\text{Vocab\_09}_{ij}) + \beta_2(\text{Vocab\_10}_{ij}) + \beta_3(\text{Vocab\_11}_{ij}) + \epsilon_{ij}$$

where,

- 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;
- Vocab_09$_{ij}$, Vocab_10$_{ij}$, and Vocab_11$_{ij}$ 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:

```
# Fit unconditional growth model
lmer.1 = lmer(vocab_score ~ 1 + grade + (1|id), data = vocabulary_long, REML = FALSE)

# Coefficient-level output
tidy(lmer.1)
```

```
# A tibble: 6 x 5
  term                    estimate std.error statistic group
  <chr>                      <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                 1.13     0.250      4.52 fixed
2 gradevocab_09               1.41     0.159      8.87 fixed
3 gradevocab_10               1.86     0.159     11.7  fixed
4 gradevocab_11               2.34     0.159     14.7  fixed
5 sd_(Intercept).id           1.79       NA        NA  id
6 sd_Observation.Residual     0.899      NA        NA  Residual
```

```
# Compute variance components
# 0.899^2 #Residual
# 1.791^2 #Intercept
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = 1.13 + 1.41(\text{Vocab\_09}_{ij}) + 1.86(\text{Vocab\_10}_{ij}) + 2.34(\text{Vocab\_11}_{ij})$$

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.

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

- 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.

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:

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

where,

- 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.

We will also look at a model that treats grade as a quantitative predictor, but also centers this predictor on the initial time point. This model is:

$$\text{Vocabulary Score}_{ij} = \left[\beta_0 + b_{0j}\right] + \beta_1(\text{Grade}_{ij} - 8) + \epsilon_{ij}$$

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.

```
# Join the data with the lookup table
vocabulary_long = vocabulary_long %>%
  left_join(lookup_table, by = "grade")
```

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

```
# 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.

```
# Fit unconditional growth model
lmer.1_quant = lmer(vocab_score ~ 1 + grade_quant + (1|id), data = vocabulary_long, REML = FALSE)

# Coefficient-level output
tidy(lmer.1_quant)
```

```
# A tibble: 4 x 5
  term                  estimate std.error statistic group
  <chr>                    <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)              -4.56     0.553     -8.24 fixed
2 grade_quant               0.747    0.0529    14.1  fixed
3 sd_(Intercept).id         1.78    NA         NA    id
4 sd_Observation.Residual   0.947   NA         NA    Residual
```

```
# Compute variance components
# 0.947^2 #Residual
# 1.784^2 #Intercept
```

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 quantitative predictor that was centered on the initial time point.

```
# Fit unconditional growth model with centered grade
lmer.1_quant_cent = lmer(vocab_score ~ 1 + grade_quant_center + (1|id), data = vocabulary_long, REML = FALSE)

# Coefficient-level output
tidy(lmer.1_quant_cent)
```

```
# A tibble: 4 x 5
  term                     estimate std.error statistic group
  <chr>                       <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                  1.41     0.244      5.79 fixed
2 grade_quant_center           0.747    0.0529    14.1  fixed
3 sd_(Intercept).id            1.78    NA         NA    id
4 sd_Observation.Residual      0.947   NA         NA    Residual
```

```
# Compute variance components
# 0.947^2 #Residual
# 1.784^2 #Intercept
```

The fitted equation is:

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

or,

$$\widehat{\text{Vocabulary Score}}_{ij} = 1.41 + 0.75(\text{Grade-level}_{ij} - 8)$$

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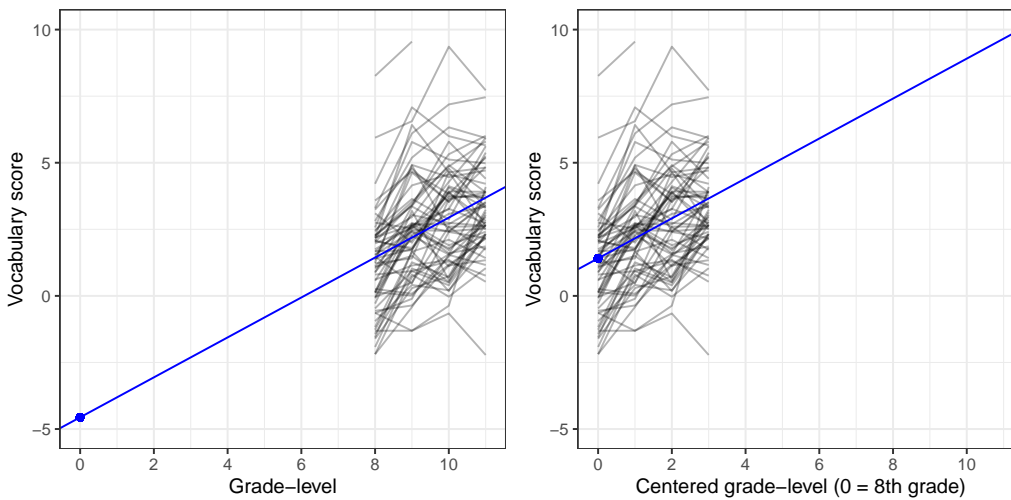
Looking at the variance components and comparing them to the unconditional random intercepts model:

- 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.

Comparing the Unconditional Growth Models

The unconditional growth model was fitted using three different methods.

- The first model treated time (grade) as categorical. This, it turns out, is the basis for a longitudinal method called *Repeated Measures ANOVA*, commonly used in the social sciences.
- The second two models treated time as a continuous variable.

Treating time continuously rather than as a categorical predictor has many advantages:

- In the real-world time is continuous.
- The model is simpler. In the model where time was treated categorically, we had to estimate four regression coefficients and two variance components. In the continuous models, we only had to estimate two regression coefficients and two variance components. If there were additional time points, we would still only need to estimate four parameters for the continuous model, but the number of estimates would increase for the categorical model.
- It allows us too allow for participants being measured at different times.
- It allows us to model nonlinear relationships more easily.

10

In general, **you should always treat time continuously when you have a longitudinal study!** This guidance also implies that you should use linear mixed-effects models rather than RM-ANOVA for the analysis of longitudinal data.

> 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. Since the spaghetti plot suggested that the relationship between grade-level and vocabulary score may be nonlinear, below we consider three potential functional forms for our growth model: (1) a linear relationship (lmer.3); (2) a quadratic relationship; and (3) a log-linear relationship (based on log-transforming grade-level).

Note: Since the centered grade predictor includes values of zero, we need to add one to each value prior to log-transforming this predictor

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

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

Examining the Output for the Log-Linear Fitted Model

```
tidy(lmer.log)
```

```
# A tibble: 4 x 5
  term                        estimate std.error statistic group
  <chr>                          <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                     1.21     0.246      4.91 fixed
2 log(grade_quant_center + 1)     1.67     0.109     15.3  fixed
3 sd_(Intercept).id               1.79    NA        NA     id
4 sd_Observation.Residual         0.907   NA        NA     Residual
```

```
# Compute variance components
# 0.907^2 #Residual
# 1.790^2 #Intercept
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = 1.21 + 1.67\left[\ln(\text{Centered grade-level}_{ij} + 1)\right]$$

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 other unconditional growth models.

Evaluating the Three Functional Forms

We should begin the evaluation process by examining the residuals from the candidate models.

```
# Obtain level-1 resduals and fitted values
mod1_lev1 = augment(lmer.1_quant_cent)
quad_lev1 = augment(lmer.quad)
log_lev1 = augment(lmer.log)

# Obtain a data frame of the random-effects (level-2 residuals)
mod1_lev2 = ranef(lmer.1_quant_cent)$id
quad_lev2 = ranef(lmer.quad)$id
log_lev2 = ranef(lmer.log)$id
```

We will then create the residual plots for the three models

```
##################### LINEAR MODEL #####################

# Density plot of the level-1 residuals
p1 = ggplot(data = mod1_lev1, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals") +
  ggtitle("Linear Model")

# Scatterplot of the Level-1 residuals versus the fitted values
p2 = ggplot(data = mod1_lev1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Density plot of the level-2 residuals (RE of intercept)
p3 = ggplot(data = mod1_lev2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```

```r
##################### NONLINEAR (QUADRATIC) MODEL #####################

# Density plot of the level-1 residuals
p4 = ggplot(data = quad_lev1, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals") +
  ggtitle("Nonlinear Model (Quadratic)")

# Scatterplot of the Level-1 residuals versus the fitted values
p5 = ggplot(data = quad_lev1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Density plot of the level-2 residuals (RE of intercept)
p6 = ggplot(data = quad_lev2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```

```r
##################### NONLINEAR (LOG) MODEL #####################

# Density plot of the level-1 residuals
p7 = ggplot(data = log_lev1, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals") +
  ggtitle("Nonlinear Model (Log)")

# Scatterplot of the Level-1 residuals versus the fitted values
p8 = ggplot(data = log_lev1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Density plot of the level-2 residuals (RE of intercept)
p9 = ggplot(data = log_lev2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```
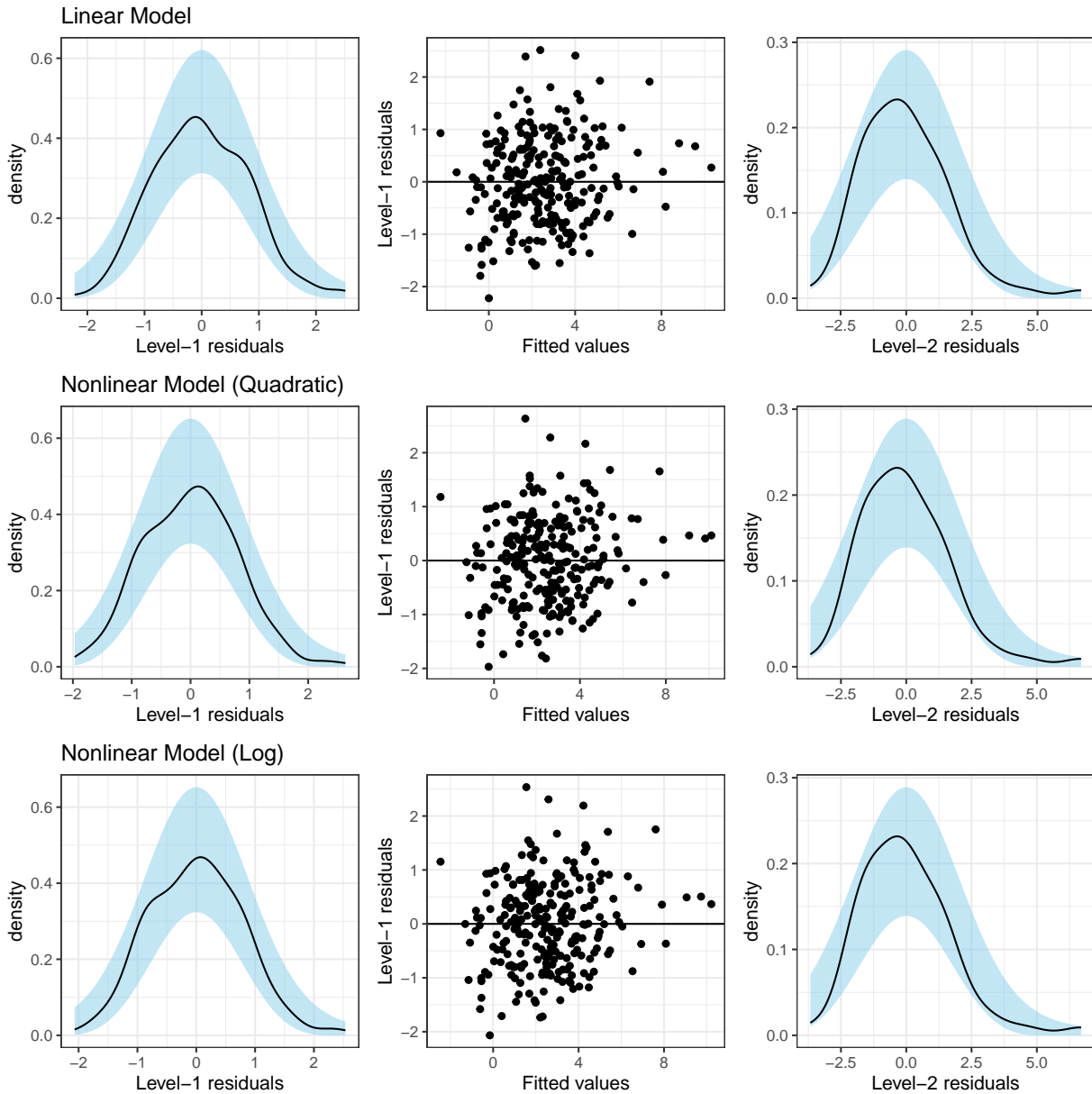
```
# Layout plots
(p1 | p2 | p3) / (p4 | p5 | p6) / (p7   | p8 | p9)
```

Figure 3

*Plots of the Level-1 Residuals and Random-Effects for the Model with a Linear Effect of Grade-Level (Row 1),
Linear and Quadratic Effect of Grade-Level (Row 2), and Log-Transformed Effect of Grade-Level (Row 3)*

The residual plots look similar indicating that all the models seem to meet the assumptions equally well. We will
turn next to information criteria to evaluate the models.

```
# Model-evidence
model.sel(
  object = list(lmer.1_quant_cent, lmer.quad, lmer.log),
  rank = "AICc"
)
```

```
Model selection table
  (Int) grd_qnt_cnt grd_qnt_cnt^2 log(grd_qnt_cnt+1) df logLik AICc delta
3  1.21                                          1.67  4   -428  864  0.00
2  1.18       1.442        -0.232                      5   -428  867  2.67
1  1.41       0.747                                    4   -436  881 16.57
  weight
3  0.791
2  0.208
1  0.000
Models ranked by AICc(x)
Random terms (all models):
'1 | id'
```

Given the data and candidate models, the evidence primarily 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. 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 functional form for our unconditional growth model.

Plot of the Unconditional Growth Model

To better understand the relationship between grade-level and vocabulary score represented in the adopted unconditional growth model, we can plot the predicted mean profile based on the model's fixed-effects.

```
# Create plot
ggplot(data = vocabulary_long, aes(x = grade_quant_center, y = vocab_score)) +
  geom_point(alpha = 0) +
  stat_function(
    fun = function(x) {1.21 + 1.67 * log(x + 1)},
    color = "blue"
    ) +
  theme_bw() +
  scale_x_continuous(
    name = "Grade-level",
    breaks = c(0, 1, 2, 3),
    labels = c("0\n(8th)", "1\n(9th)", "2\n(10th)", "3\n(11th)")
    ) +
  ylab("Vocabulary score")
```

Figure 4

Predicted Change in Vocabulary Score as a Function of Grade-Level
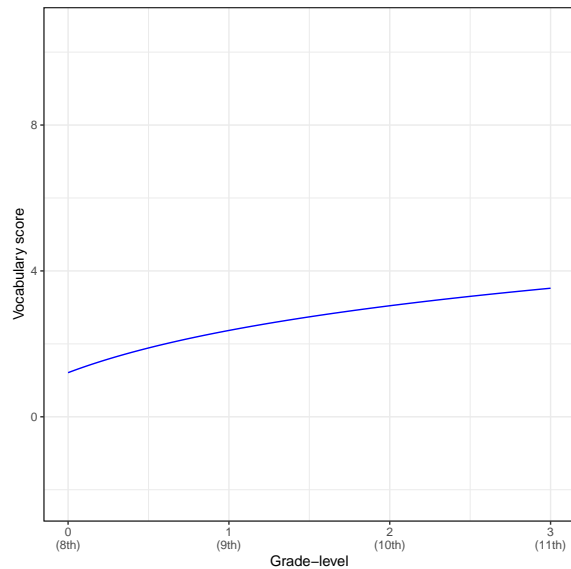
Based on this plot and the coefficient-level output, we can answer the first research question.

> 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 Second Research Question

To answer the second research question about whether the growth pattern is different for females and non-females, we again start by plotting the individual and mean growth profiles, this time conditioned on sex.

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

Figure 5

Plot Showing the Change in Vocabulary Score Over Time Conditioned on Sex. The Average Growth Profile is also Displayed as a Thicker Line.
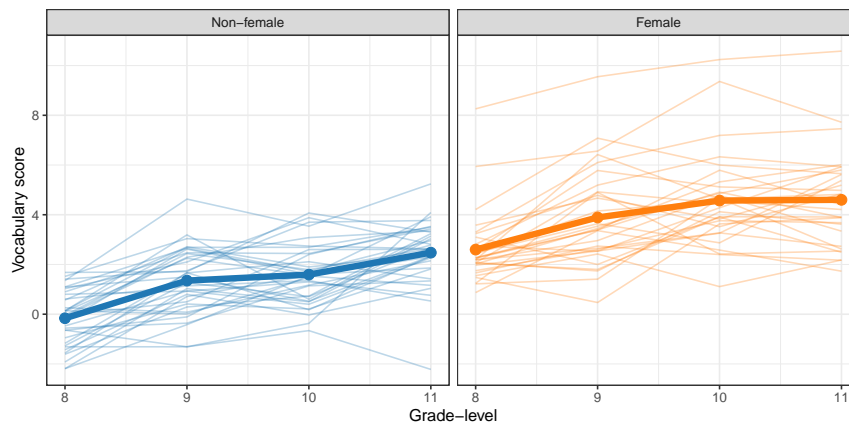
Examining the average profiles for females and non-females in this plot suggests that the females tend to have higher average vocabulary scores than non-females at each grade level. The sample average growth profiles also show slightly different patterns of growth between females and non-females. To further examine these potential differences, we can fit a set of candidate models that include fixed-effects of sex in addition to the grade-level predictor, namely: (1) a model that includes main effects of sex and grade-level, and (2) a model that includes main effects of sex and grade-level and the interaction-effect between sex and grade-level.

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

# Interaction-effect between sex and grade-level
lmer.int = lmer(vocab_score ~ 1 + log(grade_quant_center + 1) + female +
                  log(grade_quant_center + 1):female +(1|id),
                data = vocabulary_long, 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 non-females 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 females and non-females.

To facilitate model selection we will examine a table of model evidence using the following candidate models: (1) the adopted 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
model.sel(
  object = list(lmer.log, lmer.main, lmer.int),
  rank = "AICc"
)
```

```
Model selection table
      (Int) log(grd_qnt_cnt+1)  fml fml:log(grd_qnt_cnt+1) df logLik AICc delta
2 -0.00951                1.67 2.60                          5   -406  823  0.00
3 -0.10850                1.79 2.81               -0.266     6   -406  823  0.61
1  1.21000                1.67                               4   -428  864 41.54
   weight
2  0.575
3  0.425
1  0.000
Models ranked by AICc(x)
Random terms (all models):
'1 | id'
```

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.

```
# Obtain the level-1 residuals and fitted values
main_lev1 = augment(lmer.main)
int_lev1 = augment(lmer.int)

# Obtain a data frame of the random-effects
main_lev2 = ranef(lmer.main)$id
int_lev2 = ranef(lmer.int)$id
```

```
#################### MAIN EFFECTS MODEL ####################

# Density plot of the level-1 residuals
p1 = ggplot(data = main_lev1, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals") +
  ggtitle("Main Effects Model")

# Scatterplot of the Level-1 residuals versus the fitted values
p2 = ggplot(data = main_lev1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Density plot of the level-2 residuals (RE of intercept)
p3 = ggplot(data = main_lev2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```

```
#################### INTERACTION MODEL ####################

# Density plot of the level-1 residuals
p4 = ggplot(data = int_lev1, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals") +
  ggtitle("Interaction Model")

# Scatterplot of the Level-1 residuals versus the fitted values
p5 = ggplot(data = int_lev1, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Density plot of the level-2 residuals (RE of intercept)
p6 = ggplot(data = int_lev2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```
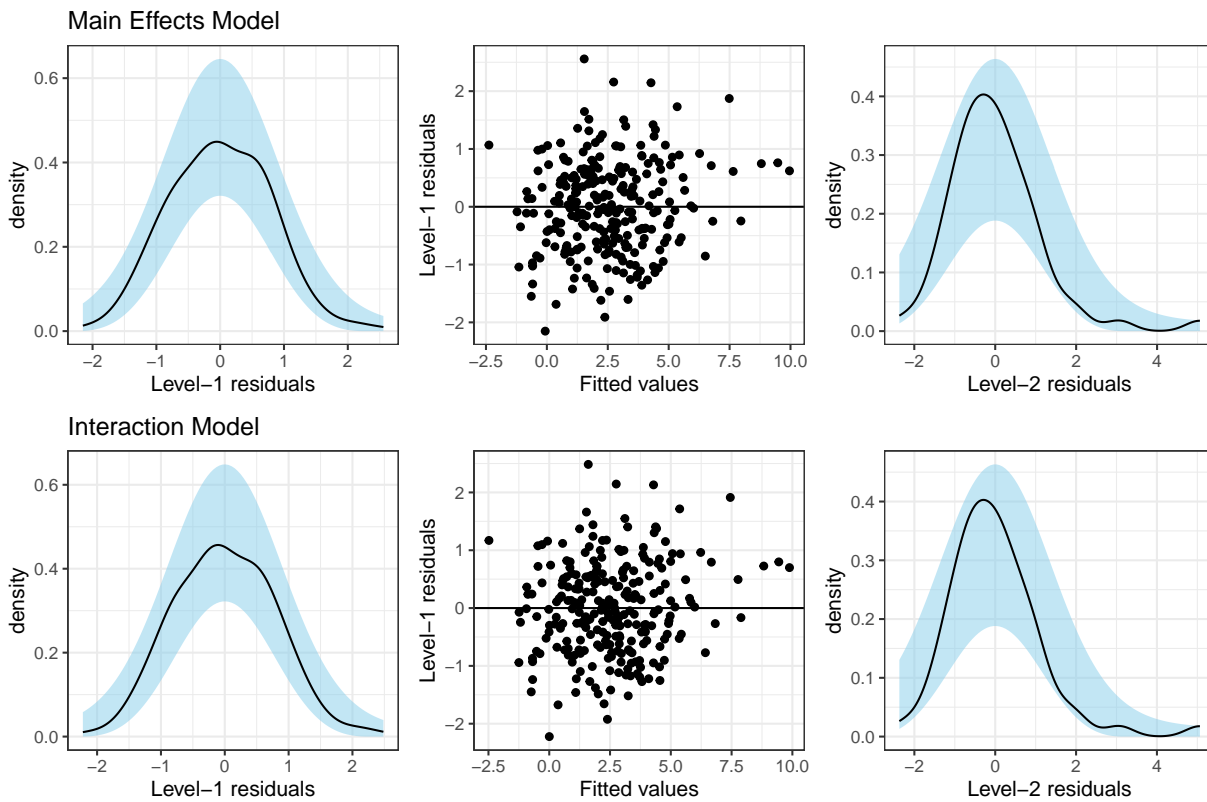
```
# Layoout plots
(p1 | p2 | p3) / (p4 | p5 | p6)
```

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.

Using the results from evaluating the table of model evidence and the residual plots, 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.

If you opted to choose the main effects model because of parsimony, it would be prudent to point out that the evidence also showed some support for the interaction model. This would cue researchers to continue to probe whether this effect is supported in future research.

Main Effects Model

```
tidy(lmer.main)
```

```
# A tibble: 5 x 5
  term                        estimate std.error statistic group
  <chr>                          <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                 -0.00951     0.241   -0.0394 fixed
2 log(grade_quant_center + 1)  1.67        0.109   15.3    fixed
3 female                       2.60        0.329    7.91   fixed
4 sd_(Intercept).id            1.23        NA       NA     id
5 sd_Observation.Residual      0.907       NA       NA     Residual
```

```
# Compute variance components
# 0.907^2 #Residual
# 1.232^2 #Intercept
```

The fitted equation is:

$$\widehat{\text{Vocabulary Score}}_{ij} = -0.01 + 1.67 \left[\ln(\text{Grade-level}_{ij} - 8 + 1) \right] + 2.60(\text{Female}_{\cdot j})$$

Interpreting the coefficients,

- The predicted average vocabulary score for non-female 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 non-females, controlling for differences in grade-level.

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

- 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 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.

```
ggplot(data = vocabulary_long, aes(x = grade_quant_center, y = vocab_score)) +
  geom_point(alpha = 0) +
  stat_function(
    fun = function(x) {-0.01 + 1.67 * log(x + 1)},
    color = "blue",
    linetype = "dashed"
    ) +
  stat_function(
    fun = function(x) {2.59 + 1.67 * log(x + 1)},
    color = "orange"
    ) +
  theme_bw() +
  scale_x_continuous(
    name = "Grade-level",
    breaks = c(0, 1, 2, 3),
    labels = c("0\n(8th)", "1\n(9th)", "2\n(10th)", "3\n(11th)")
    ) +
  ylab("Vocabulary score")
```

Figure 7

Predicted Change in Vocabulary Score as a Function of Grade-Level for Females (Orange, Solid Line) and Non-Females (Blue, Dashed Line
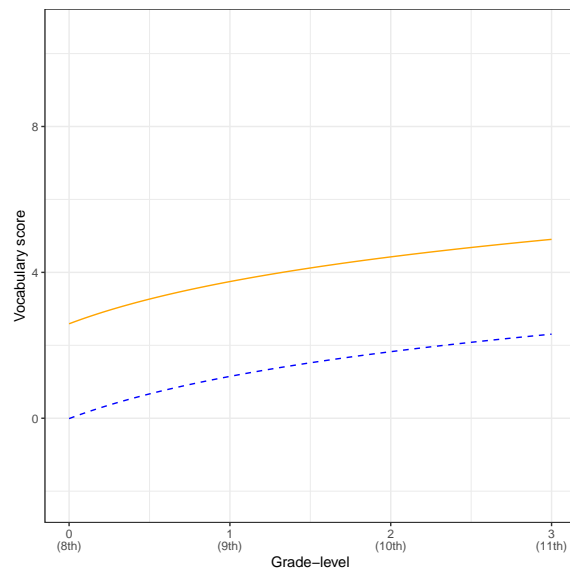
Interaction Model

```
tidy(lmer.int)
```

```
# A tibble: 6 x 5
  term                            estimate std.error statistic group
  <chr>                              <dbl>     <dbl>     <dbl> <chr>
1 (Intercept)                       -0.109     0.254    -0.427 fixed
2 log(grade_quant_center + 1)        1.79      0.149     12.0  fixed
3 female                             2.81      0.371      7.57 fixed
4 log(grade_quant_center + 1):female -0.266    0.217     -1.22 fixed
5 sd_(Intercept).id                  1.23     NA        NA     id
6 sd_Observation.Residual            0.903    NA        NA     Residual
```

```
# Compute variance components
# 0.903^2 #Residual
# 1.233^2 #Intercept
```

The fitted equation is:

$$\hat{\text{Vocabulary Score}}_{ij} = -0.11 + 1.79\left[\ln(\text{Grade-level}_{ij} - 8 + 1)\right] + 2.81(\text{Female}_{.j})$$
$$- 0.27\left[\ln(\text{Grade-level}_{ij} - 8 + 1) \times (\text{Female}_{.j})\right]$$

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 we find that:

- The growth pattern in vocabulary is log-linear over time for both females and non-females.
- While the change in vocabulary score, on average, is positive for females and non-females, 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 non-females.

```
ggplot(data = vocabulary_long, aes(x = grade_quant_center, y = vocab_score)) +
  geom_point(alpha = 0) +
  stat_function(
    fun = function(x) {-0.11 + 1.79 * log(x + 1)},
    color = "blue",
    linetype = "dashed"
    ) +
  stat_function(
    fun = function(x) {2.7 + 1.52 * log(x + 1)},
    color = "orange"
    ) +
  theme_bw() +
  scale_x_continuous(
    name = "Grade-level",
    breaks = c(0, 1, 2, 3),
    labels = c("0\n(8th)", "1\n(9th)", "2\n(10th)", "3\n(11th)")
    ) +
  ylab("Vocabulary score")
```

Figure 8

Predicted Change in Vocabulary Score as a Function of Grade-Level for Females (Orange, Solid Line) and Non-Females (Blue, Dashed Line