

Linear Mixed-Effects Models: Alternative Representations and Assumptions

2020-03-05

In this set of notes, you will learn alternative ways of representing the linear mixed-effects model. You will also learn about the underlying assumptions for the linear mixed-effects model, as well as how to evaluate them empirically.

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

# 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.

Statistical Model: Expressed as a Mixed-Effects Model

In the last unit we fitted several linear mixed-effects models to predict variation in students' post-test language scores. The statistical models for these are presented below.

$$\text{Model 1 : Language Score}_{ij} = [\beta_0 + b_{0j}] + \epsilon_{ij}$$

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

$$\text{Model 3 : Language Score}_{ij} = [\beta_0 + b_{0j}] + \beta_1(\text{Verbal IQ}_{ij}) + \beta_2(\text{SES}_{ij}) + \beta_3(\text{Public}_{ij}) + \epsilon_{ij}$$

This representation of the models is referred to as the **composite model** or the **mixed-effects model** since they include both the fixed- and random-effects in the same equation.

Writing the Statistical Model as a Set of Multilevel Equations

Another way we can express the model is by separating the mixed-effects model equation into multiple equations; one for each level of variation. For example, each of the mixed-effects models listed above could be separated into two equations: a student-level equation (Level-1) and a set of school-level equations (Level-2). As an example, take the equation for Model 2:

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

We initially write the student-level, or Level-1 equation. The level-1 equation includes all the fixed-effects and the random error term from the mixed-effects model. When writing the Level-1 model, we add a j subscript to each of the fixed-effects to indicate that the particular effect may be unique to a particular school. The Level-1 equation for Model 2 is:

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

The Level-1 equation describes the variation in students' post-test language scores. It says that this variation is decomposed into that which is explained by differences in students' verbal IQ scores and unexplained random error. The j subscript on the fixed-effects terms indicates that the intercept and effect of verbal IQ scores is the same for all students within a particular school.

After writing the Level-1 equation, we can write out the Level-2, or school-level equation(s). There will be a Level-2 equation for each of the fixed-effects in the Level-1 model. In our example, since we have two fixed-effects in the Level-1 model (β_0 and β_1), there will be two Level-2 equations. In each Level-2 equations, the outcome is one of the fixed-effects from the Level-1 equation. These equations describe how the school-specific intercept and slopes differ across schools. As such, they may include the random-effects and any school-level effects. For example, we can write the Level-2 equations for Model 2 as:

$$\begin{aligned}\beta_{0j} &= \beta_0 + b_{0j} \\ \beta_{1j} &= \beta_1\end{aligned}$$

These equations indicate that the school-specific intercepts are a function of some part common to all schools (β_0 ; a fixed-effect) and some deviation from that (b_{0j} ; a random-effect of intercept). The random-effect is the reason

that schools can have different intercepts. The school-specific slope, on the other hand, does not vary by school; it is the same for each school.

Together these equations are referred to as the set of multilevel equations:

Level-1 :

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_0 + b_{0j}$$

$$\beta_{1j} = \beta_1$$

Multilevel Equation for Model 3

As a second example, consider Model 3:

$$\text{Language Score}_{ij} = [\beta_0 + b_{0j}] + \beta_1(\text{Verbal IQ}_{ij}) + \beta_2(\text{SES}_{ij}) + \beta_3(\text{Public}_{\bullet j}) + \epsilon_{ij}$$

The Level-1 equation is:

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \beta_{2j}(\text{SES}_{ij}) + \epsilon_{ij}$$

This equation indicates that students' post-test language scores are a function of a school-specific intercept, a school-specific effect of verbal IQ scores, a school-specific effect of SES, and random error. There are now three Level-2 equations:

$$\beta_{0j} = \beta_0 + \beta_4(\text{Public}_{\bullet j}) + b_{0j}$$

$$\beta_{1j} = \beta_1$$

$$\beta_{2j} = \beta_2$$

These equations indicate that each school-specific intercept is a function of a fixed, or common, intercept, the type of school, and random error. The effects of verbal IQ scores, pre-test language scores, and SES are constant, or fixed, across schools (there is not a random-effect in any of the last three Level-2 equations).

Going from Multilevel Equations to the Mixed-Effects Model

If we have the multilevel equations, we can substitute the Level-2 equation(s) into the Level-1 equation to get the composite equation or mixed-effects equation. For example, for Model 2, substituting $\beta_0 + b_{0j}$ into β_{0j} and β_1 into β_{1j} gives us:

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

Similarly substituting the Level-2 information into the Level-1 equation for Model 3, we end up with the following mixed-effects representation:

$$\text{Language Score}_{ij} = \left[\beta_0 + \beta_4(\text{Public}_{\bullet j}) + b_{0j} \right] + \beta_1(\text{Verbal IQ}_{ij}) + \beta_2(\text{SES}_{ij}) + \epsilon_{ij}$$

Which, if we re-arrange terms gives us the same mixed-effects model that we started with. This substitution also helps us think about which Level-2 equation we put the the school-level predictors in. For example, what if we would have put the school type (public) predictor into the Level-2 equation associated with verbal IQ?

Level-1 :

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \beta_{2j}(\text{SES}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_0 + b_{0j}$$

$$\beta_{1j} = \beta_1 + \beta_4(\text{Public}_{\bullet j})$$

$$\beta_{2j} = \beta_2$$

$$\beta_{3j} = \beta_3$$

If we substitute back into the Level-1 equation, the composite equation is:

$$\begin{aligned} \text{Language Score}_{ij} &= \left[\beta_0 + b_{0j} \right] + \left[\beta_1 + \beta_4(\text{Public}_{\bullet j}) \right] (\text{Verbal IQ}_{ij}) + \beta_2(\text{SES}_{ij}) + \epsilon_{ij} \\ &= [\beta_0 + b_{0j}] + \beta_1(\text{Verbal IQ}_{ij}) + \beta_4(\text{Public}_{\bullet j})(\text{Verbal IQ}_{ij}) + \beta_2(\text{SES}_{ij}) + \epsilon_{ij} \end{aligned}$$

Adding the school type predictor in the verbal IQ equation produces an interaction term (product terms) in the composite equation! Adding them to the intercept equation produced main-effects. Since our original mixed-effects model included school type as a main-effect, we need to include them in the intercept equation. Note that if the composite equation included both a main-effect and an interaction, we would need to include the predictor in more than one of the Level-2 equations (e.g., in the intercept equation and the verbal IQ equation).

Any predictors included in the Level-2 slope equations also need to be included in the Level-2 intercept equation.

Guidelines for Writing the Multilevel Equations

Here are some guidelines in helping you think about writing multilevel equations.

- Write the Level-1 equation first. This will be an equation that expresses the outcome's relationship to a series of school-specific parameters and a student-specific residual.
- The number of school-specific parameters in the Level-1 equation (aside from the residual) dictate the number of Level-2 equations you will have.
- The school-specific parameters from the Level-1 equation will be the outcomes in the Level-2 equations.
- Random-effects are the residuals in the Level-2 equations, and therefore are in the Level-2 equations; one per equation.
- Variables from the data go to their appropriate level. For example student-level variables will be put in the Level-1 equation, and school-level predictors will be put in one or more of the Level-2 equations.

Multilevel Equations for Fixed-Effects Models

Our conventional fixed-effects regression models (LM) can also be expressed as a multilevel model. For example, consider the fixed-effect model that includes an intercept and effect of verbal IQ score:

$$\text{Language Score}_i = \beta_0 + \beta_1(\text{Verbal IQ}_i) + \epsilon_i$$

The multilevel model would specify that the Level-2 equations would only include fixed-effects (no random-effects). Thus when we substitute them back into the Level-1 model we only have fixed-effects in the model:

Level-1 :

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_0$$

$$\beta_{1j} = \beta_1$$

Why are Multilevel Expressions Helpful?

Expressing the model as a set of multilevel equations can be helpful for readers. First, it explicitly separates the sources of variation and the predictors of these sources of variation into different levels. In our example there are two sources of variation student-level variation (within-school) and school-level variation (between-school). The Level-1 model attempts to describe the within-school variation and, hence, only includes student-level predictors. The Level-2 models attempts to describe the between-school variation and only includes school-level predictors.

Secondly, the multilevel expression of the model helps us think about what the predictors at each level are actually doing. Level-1 predictors explain variation in the outcome. In our example, they are explaining variation in students' post-test language scores. The Level-2 predictors are explaining variation in the school-specific intercepts (or intercepts and slopes)—they explain Level-2 variation.

Thirdly, the multilevel expression of the model helps us see that the random-effects are residuals; they are residuals of the Level-2 models. This helps us think about the more general statistical model. For example, the general statistical model for a model that includes fixed-effects of two predictors and a random-effect of intercept is:

Level-1 :

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij}) + \beta_{2j}(X_{2ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_0 + b_{0j}$$

$$\beta_{1j} = \beta_1$$

$$\beta_{2j} = \beta_2$$

where

$$\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$$

$$b_{0j} \sim \mathcal{N}\left(0, \sigma_0^2\right)$$

In the mixed-effects model we put distributional assumptions on both the Level-1 residuals and the Level-2 residuals.

Evaluating the Assumptions: An Example

Evaluating the assumptions in a mixed-effects model is a bit more complicated than it is in a fixed-effects model, and there are several things to check depending on the model that was fitted (among other things, the number of random-effects and their covariance structure). To simplify things, in this course, we will evaluate the distributional assumptions placed on the Level-1 residuals, and we will also evaluate the normality assumption on the random-effects.

To illustrate assumption checking in practice, we will evaluate the assumptions for fitting Model 3. Recall that the multilevel expression of Model 3 was:

Level-1 :

$$\text{Language Score}_{ij} = \beta_{0j} + \beta_{1j}(\text{Verbal IQ}_{ij}) + \beta_{2j}(\text{SES}_{ij}) + \epsilon_{ij}$$

Level-2 :

$$\beta_{0j} = \beta_0 + \beta_4(\text{Public}_{\bullet j}) + b_{0j}$$

$$\beta_{1j} = \beta_1$$

$$\beta_{2j} = \beta_2$$

The assumptions are based on the Level-1 residuals (ϵ_{ij}) and the Level-2 residuals, or random-effects (b_{0j}). So we need to examine the distributions of those two components. To begin, we will fit the mixed-effects model.

```
# Fit Model 3
lmer.3 = lmer(language_post ~ 1 + verbal_iq + ses + public + (1 | school_id),
              data = joined_data, REML = FALSE)
```

Evaluate Assumptions about the Level-1 Residuals

We will evaluate the Level-1 residuals in the exact same way we evaluated the residuals from a fixed-effects (LM) analysis. The `augment()` function from the **broom** package produces the Level-1 residuals and fitted values.

```
# Augment the model to get the Level-1 residuals and fitted values
out_3 = augment(lmer.3)
head(out_3)
```

	language_post	verbal_iq	ses	public	school_id	.fitted	.resid	.hat	.cooks
1	46	3.166	23	1	1	48.1	-2.073	0.0362	0.0010456
2	45	2.666	10	1	1	44.8	0.217	0.0365	0.0000115
3	33	-2.334	15	1	1	34.4	-1.372	0.0340	0.0004278
4	46	-0.834	23	1	1	39.1	6.923	0.0341	0.0109372
5	20	-3.834	10	1	1	30.2	-10.165	0.0346	0.0239772
6	30	-2.334	10	1	1	33.5	-3.539	0.0341	0.0028579

	.fixed	.mu	.offset	.sqrtXwt	.sqrttrwt	.weights	.wtres
1	46.1	48.1	0	1	1	1	-2.073
2	42.8	44.8	0	1	1	1	0.217
3	32.4	34.4	0	1	1	1	-1.372
4	37.1	39.1	0	1	1	1	6.923
5	28.1	30.2	0	1	1	1	-10.165
6	31.5	33.5	0	1	1	1	-3.539

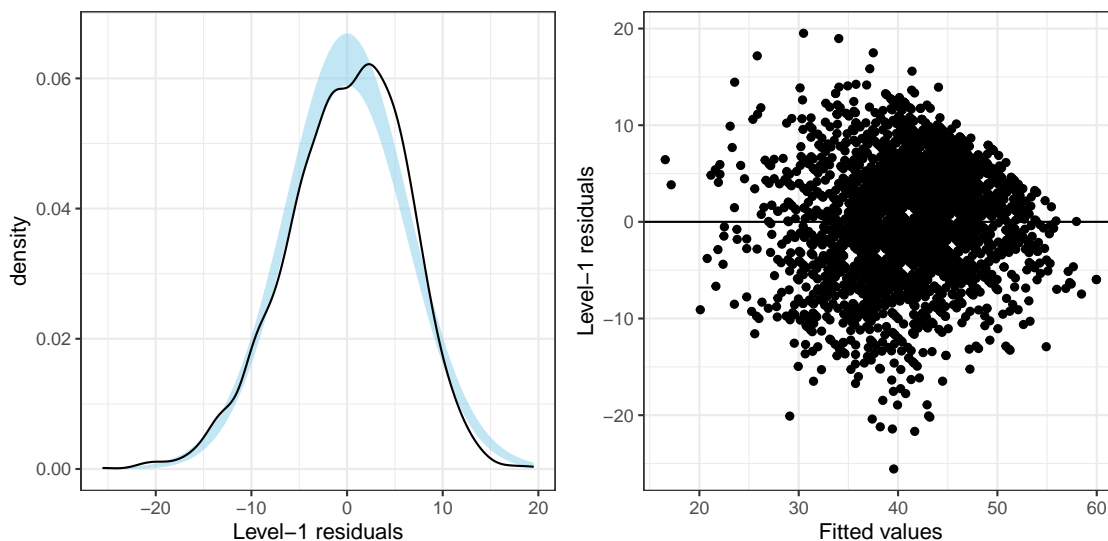
The Level-1 residuals are found in the `.resid` column, and the `.fitted` column contains the \hat{Y} values. As with LM residual analysis, we want to examine the normality of the residuals in a density plot (or some other plot that allows you to evaluate this), and the other assumptions by plotting the residuals against the fitted values in a scatterplot.

```
# Density plot of the level-1 residuals
p1 = ggplot(data = out_3, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals")

# Scatterplot of the Level-1 residuals versus the fitted values
p2 = ggplot(data = out_3, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Level-1 residuals")

# Plot side-by-side
p1 + p2
```

Figure 1
Plots to Evaluate the Level-1 Residuals



Based on the plots, it is unclear that the distributional assumptions for the Level-1 residuals are reasonably satisfied. The density plot suggests that the normality assumption is likely not tenable. The scatterplot shows symmetry around the $Y = 0$ line (average residual is 0), however, the assumption of homoskedasticity is somewhat questionable. The pattern in the residuals shows more variation for fitted values between 30 and 50, and less variation for smaller and larger fitted values.

Assumptions about the Random-Effects

We also need to examine the assumptions for any random-effects included in the model. For this course, we will examine the normality assumption. In our example that means we need to examine the normality assumption about the intercept random-effects. To do this we need to extract the random-effects from the model into a data frame so we can use **ggplot2** functions to evaluate normality.

Unfortunately the name of the column, (Intercept), includes parentheses which are special characters in R (they designate functions). We can either rename this column using the `rename()` function from **dplyr**, or we can enclose the variable name in backticks when we call it in the `aes()` function in `ggplot()`.

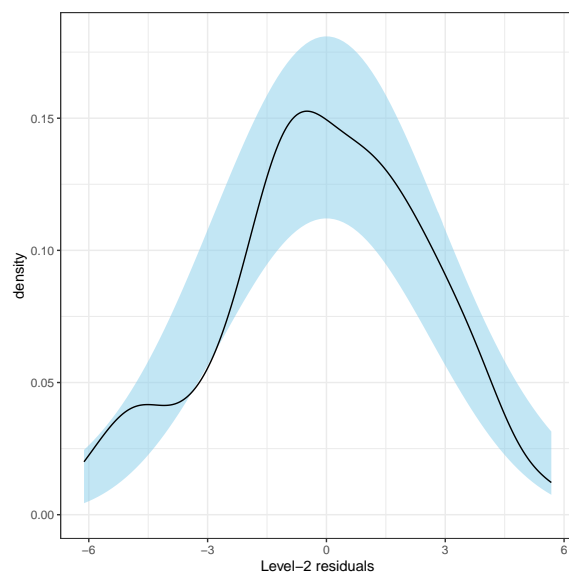
```
# Obtain a data frame of the random-effects
level_2 = ranef(lmer.3)$school_id

#head(level_2)

# Density plot of the RE for intercept
ggplot(data = level_2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")
```

Figure 2

Density Plot of the Estimated Random-Effects for Intercept



This assumption looks reasonably satisfied.

Log-transforming the Outcome to ‘Fix’ the Assumptions

Similar to fixed-effects models, we can apply transformations to the outcome or any of the continuous predictors. Here the log-transform can be applied to the outcome to help alleviate the heteroskedasticity we observed in the

level-1 residuals. (There are no 0 values or negative values in the outcome, so we can directly apply the log-transformation.)

```
# Fit model
lmer.4 = lmer(log(language_post) ~ 1 + verbal_iq + ses + public + (1 | school_id),
              data = joined_data, REML = FALSE)

# Augment model
out_4 = augment(lmer.4)

# Obtain a data frame of the random-effects
level_2 = ranef(lmer.4)$school_id

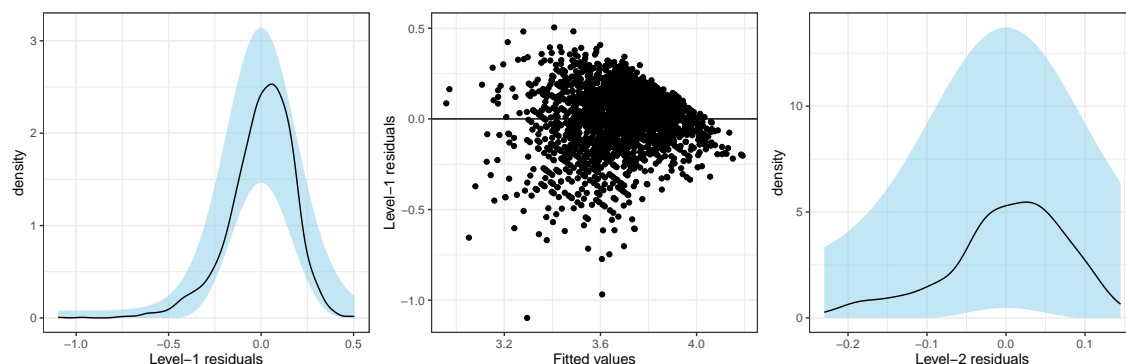
# Density plot of the level-1 residuals
p1 = ggplot(data = out_4, aes(x = .resid)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-1 residuals")

# Scatterplot of the Level-1 residuals versus the fitted values
p2 = ggplot(data = out_4, 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 = level_2, aes(x = `(Intercept)`)) +
  stat_density_confidence(model = "normal") +
  stat_density(geom = "line") +
  theme_bw() +
  xlab("Level-2 residuals")

# Plot side-by-side
p1 + p2 + p3
```

Figure 3
Plots to Evaluate the Level-1 and Level-2 Residuals



The plots suggest that all the distributional assumptions for the Level-1 residuals are tenable. (Although there is a bit of heteroskedasticity, this looks like it might be related to one or two outlying residuals.) The density plot of the level-2 residuals also shows consistency with the assumption of normality.

Interpreting the Output from the Model

We interpret fixed-effect coefficients from the LMER similar to those from the LM. We now just apply the same interpretational changes that we did when previously log-transformed the outcomes.

```
tidy(lmer.4)
```

```
# A tibble: 6 x 5
  term                estimate std.error statistic group
<chr>                <dbl>    <dbl>    <dbl> <chr>
1 (Intercept)         3.56      0.0154    231. fixed
2 verbal_iq           0.0605     0.00202   29.9 fixed
3 ses                 0.00447    0.000419  10.7 fixed
4 public              -0.0363     0.0183   -1.98 fixed
5 sd_(Intercept).school_id 0.0869    NA        NA school_id
6 sd_Observation.Residual 0.179     NA        NA Residual
```

- The intercept is the average log-transformed post-test language score for private schools with verbal IQ scores of 0 and SES of 0. Back-transforming this, we find the average language score is $e^{3.56} = 35.2$.
- Each one-point difference in verbal IQ score is associated with a 0.06-point difference in log-transformed post-test language scores, on average, controlling for differences in SES and sector. Back-transforming this, we find that each one-point difference in verbal IQ score is associated with a 6% increase in post-test language scores, on average, controlling for differences in SES and sector.
- Each one-unit difference in SES is associated with a 0.004-point difference in log-transformed post-test language scores, on average, controlling for differences in verbal IQ score and sector. Back-transforming this, we find that each one-point difference in SES is associated with a 0.4% increase in post-test language scores, on average, controlling for differences in verbal IQ score and sector.
- Public schools have a log-transformed post-test language scores that is 0.04-points lower than private schools, on average, controlling for differences in verbal IQ score and SES. Back-transforming this, we find that public schools have post-test language scores that are 4% lower than private schools, on average, controlling for differences in verbal IQ score and SES.

Note that the variance components change a lot after applying the log-transformation because the metric has changed: Raw scores versus log-transformed scores.

Table 1

Model 3: Outcome of Post-Test Language Score

term	estimate	std.error	statistic	group
sd_(Intercept).school_id	2.93	NA	NA	school_id
sd_Observation.Residual	6.33	NA	NA	Residual

Table 2

Model 4: Outcome of Log-Transformed Post-Test Language Score

term	estimate	std.error	statistic	group
sd_(Intercept).school_id	0.087	NA	NA	school_id
sd_Observation.Residual	0.179	NA	NA	Residual

However, if we look at the reduction in variance components from the baseline models, these values are similar for both the raw and log-transformed models. For the untransformed model the proportion of reduction relative to the unconditional random intercepts model is:

$$\begin{aligned}\text{School-Level} &: \frac{4.41^2 - 2.93^2}{4.41^2} = 0.559 \\ \text{Student-Level} &: \frac{8.04^2 - 6.33^2}{6.33^2} = 0.380\end{aligned}$$

Remember, to obtain the baseline variance components we need to fit the model that only includes the fixed- and random-effects of intercept to predict variation in the outcome. We also need to square the values that `tidy()` produces because we need to work with variance components, not standard deviations.

For the log-transformed model the proportion of reduction relative to the unconditional random intercepts model is:

$$\begin{aligned}\text{School-Level} &: \frac{0.128^2 - 0.087^2}{0.128^2} = 0.538 \\ \text{Student-Level} &: \frac{0.223^2 - 0.179^2}{0.223^2} = 0.356\end{aligned}$$

Although these values are slightly different from those obtained from the model with the raw outcome, they are similar in magnitude, namely:

- Verbal IQ, SES, and sector accounted for roughly 54% of the unexplained variation at the school-level.
- Verbal IQ, SES, and sector accounted for roughly 36% of the unexplained variation at the student-level.