Linear Mixed-Effects Models II

Andrew Zieffler

Recall our LME model that included fixed-effects for intercept and slope, and random-effects for intercept and slope.

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \big[b_{0i} + b_{1i}(\text{Grade}_{ij})\big] + \epsilon_{ij}$$

We also considered the LME model that included fixed-effects for intercept, slope, and sex, as well as, random-effects for intercept and slope.

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta 2(\text{Female}_{ij}) + \big[b_{0i} + b_{1i}(\text{Grade}_{ij})\big] + \epsilon_{ij}$$

We fitted these models in the previous set of notes to introduce you to the LME model and to introduce the `lmer()` syntax. In this set of notes, we will explore the model-fitting process. For example, how do we determine if the structural form is linear? Or quadratic? How do we know whether we should include random-effects for both intercept and slope? Which covariate or set of covariates should we retain in the model?

Expressing the Mixed-Effects Model as a Multi-level Model

To better describe the fitting process, we will first introduce an alternative method for expressing the LME model.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$
$$\beta_1^* = \beta_1 + b_{1i}$$

This expression is referred to as the **multi-level model**.

- The level-1 model describes the *within-student variation*. Here it describes the student-specific change profile for student *i*. It only includes time-varying predictors. Here we denote the average-effect with an asterisk.
- The level-2 models describe the *between-student variation*. There is one level-2 model for each effect included in the level-1 model. The random-effects are part of the level-2 models as they describe between-student variation.

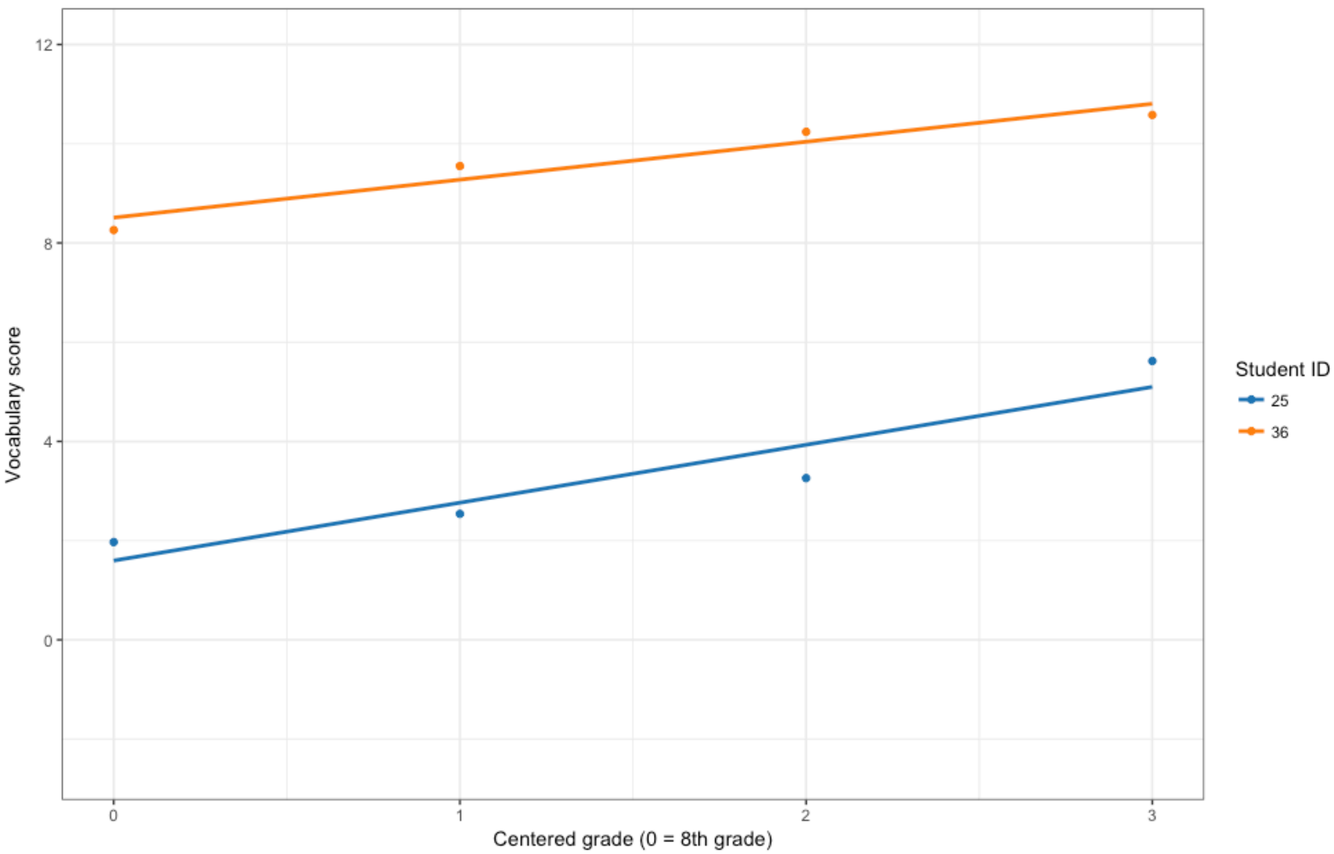
The level-1 model describes within-student variation by indicating the student-specific change profile.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta^*_0 + \beta^*_1(\text{Grade}_{ij}) + \epsilon_{ij}$$

The beta-star values represent the student-specific intercept and linear effect of grade on vocabulary.

Note that the epsilon term measures the unaccounted for within-student variation; the student-specific residuals.



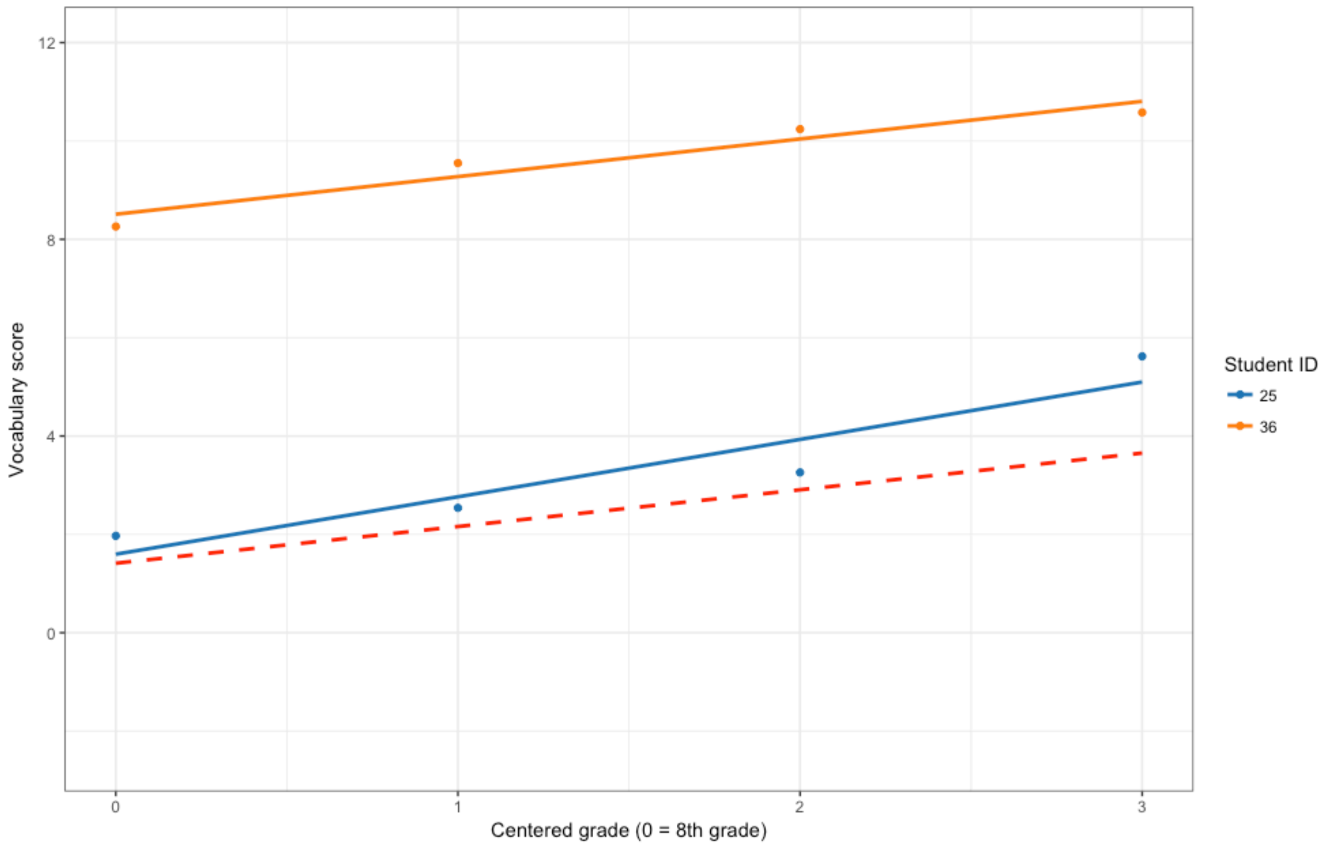
The level-2 models describe the between-student differences by indicating how the student-specific parameters vary from an average effect.

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$
$$\beta_1^* = \beta_1 + b_{1i}$$

The student-specific intercept can be written as the average intercept plus some deviation (the random-effect)

The student-specific slope can be written as the average slope plus some deviation (the random-effect)



The red line shows the average line; the fixed-effects.

If we include covariates, they need to be put into the appropriate model. In general, if the covariate VARIES BY TIMEPOINT, it goes in the level-1 model. If the covariate VARIES BY INDIVIDUAL, but does not vary by timepoint within individuals, then it goes in the level-2 models.

The sex covariate is an example of a level-2 predictor. It varies across individuals, but not across timepoints within an individual. Which specific level-2 model(s) you put it in depends on whether you believe sex explains differences in the intercept or slope.

Level-2 Models

$$\beta_0^* = \beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}$$
$$\beta_1^* = \beta_1 + b_{1i}$$

Here we believe that sex explains differences in the subject-specific intercepts, but not in slopes.

Level-2 Models

$$\beta_0^* = \beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}$$
$$\beta_1^* = \beta_1 + \beta_3(\text{Female}_{ij}) + b_{1i}$$

Here we believe that sex explains differences in both the subject-specific intercepts andslopes.

Let's further examine the differences between these two sets of level-2 models.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}$$

$$\beta_1^* = \beta_1 + b_{1i}$$

Substituting we get...

$$\text{Vocabulary}_{ij} = \left[\beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}\right] + \left[\beta_1 + b_{1i}\right](\text{Grade}_{ij}) + \epsilon_{ij}$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta_2(\text{Female}_{ij}) + \left[b_{0i} + b_{1i}(\text{Grade}_{ij})\right] + \epsilon_{ij}$$

When we include a covariate in the level-2 intercept model, we end up with a MAIN-EFFECT in the mixed-effects model.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}$$
$$\beta_1^* = \beta_1 + \beta_3(\text{Female}_{ij}) + b_{1i}$$

Substituting we get…

$$\text{Vocabulary}_{ij} = \left[\beta_0 + \beta_2(\text{Female}_{ij}) + b_{0i}\right] + \left[\beta_1 + \beta_3(\text{Female}_{ij} + b_{1i}\right](\text{Grade}_{ij}) + \epsilon_{ij}$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta_2(\text{Female}_{ij}) + \beta_3(\text{Grade}_{ij})(\text{Female}_{ij}) + \left[b_{0i} + b_{1i}(\text{Grade}_{ij})\right] + \epsilon_{ij}$$

When we include a covariate in the level-2 intercept AND slope models, we end up with a MAIN-EFFECT AND AN INTERACTION-EFFECT in the mixed-effects model.

What if we only included the sex covariate in the level-2 slope model?

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$

$$\beta_1^* = \beta_1 + \beta_3(\text{Female}_{ij}) + b_{1i}$$

Substituting we get…

$$\text{Vocabulary}_{ij} = \left[\beta_0 + b_{0i}\right] + \left[\beta_1 + \beta_3(\text{Female}_{ij} + b_{1i}\right](\text{Grade}_{ij}) + \epsilon_{ij}$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta_3(\text{Grade}_{ij})(\text{Female}_{ij}) + \left[b_{0i} + b_{1i}(\text{Grade}_{ij})\right] + \epsilon_{ij}$$

When we include a covariate in the level-2 slope model, but not in the level-2 intercept model, we end up with an INTERACTION-EFFECT with no constituent main-effect in the mixed-effects model.

No bueno for interpretation!

Unconditional Means Model

It is typical, in practice, to initially fit an unconditional intercept-only model. This model includes a fixed-effect of intercept and a random-effect of intercept, with no other terms.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + b_{0i} + \epsilon_{ij}$$

The is the simplest longitudinal model we can fit.

In longitudinal analysis, we will always include a random-effect for intercept. This accounts for the correlation that arises because of non-independence.

Since this model expresses the student-specific models via their mean vocabulary score, we refer to it as the unconditional means model.

This simple model is quite useful as it partitions the residual variation into that which is within-students and that which is between-students.

$$\text{Vocabulary}_{ij} = \beta_0 + \underbrace{\boxed{b_{0i}}}_{\text{Between-student variation}} + \underbrace{\boxed{\epsilon_{ij}}}_{\text{Within-student variation}}$$

This partitioning is shown in the `Random Effects` part of the `lmer()` output.

```
> lmer.0 = lmer(score ~ 1 + (1|id), data = vocab_long, REML = FALSE)
> summary(lmer.0)

Random effects:
 Groups   Name        Std.Dev.
 id       (Intercept) 1.718
 Residual             1.351
```

Between-student variation

Within-student variation

To better understand this partitioning we typically compute VARIANCES.

Computing the variances we get…

$$\mathrm{Var}(\epsilon_{ij}) = \hat{\sigma}_\epsilon^2 = 1.351^2 = 1.825$$

$$\mathrm{Var}(b_{0i}) = \hat{\sigma}_0^2 = 1.718^2 = 2.952$$

The sum of these variances give the total unaccounted for variation. Conventionally we report these values and also indicate the proportion of that variance which is within-student and between-student.

$$\frac{1.825}{1.825 + 2.952} = 0.38$$

← **38% of the unaccounted variation is within-student or level-1 variation**

$$\frac{2.952}{1.825 + 2.952} = 0.62$$

← **62% of the unaccounted variation is between-student or level-2 variation**

There is unaccounted for variation at both level-1 and level-2. To explain unaccounted for variation at level-1 we include level-1 predictors (e.g., grade, grade^2, etc.). To explain unaccounted for variation at level-2 we include level-2 predictors (e.g., female, special education status, etc.).

Unconditional Linear Growth Model

Let's add grade as a predictor in the level-1 model.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$

$$\beta_1^* = \beta_1$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + b_{0i} + \epsilon_{ij}$$

While we include a level-2 equation for each effect in the level-1 model, we will keep the random-effect only in the intercept. The rationale for only including the random-effect in the intercept is that we are interested in selecting the fixed-effects structure. It is hard to evaluate differences in the fixed-effects if you are manipulating both the fixed- and random-effects simultaneously.

Comparing the two mixed-effects models, we can see the only difference is in the additional fixed-effect of grade in the second model.

Unconditional means model: $$\text{Vocabulary}_{ij} = \beta_0 + b_{0i} + \epsilon_{ij}$$

Unconditional linear growth model: $$\text{Vocabulary}_{ij} = \beta_0 + \boxed{\beta_1(\text{Grade}_{ij})} + b_{0i} + \epsilon_{ij}$$

```
> lmer.1 = lmer(score ~ 1 + grade + (1|id), data = vocab_long, REML = FALSE)
> summary(lmer.1)

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 3.1838   1.7843  
 Residual             0.8965   0.9468  
Number of obs: 256, groups:  id, 64
```

In models that include predictors (aside from the intercept), comparing the within- and between-student variance to each other is less helpful. It is, however, useful to compare the variances from these models to the variances from the intercept-only model.

Intercept-only model	**Linear grade model**
$\sigma^2_\epsilon = 1.825$	$\sigma^2_\epsilon = 0.896$

By adding the grade predictor to the level-1 model, we have reduced the unaccounted for variation at level-1 from 1.825 to 0.896; a 51% reduction!

This is evidence that we should adopt the linear effect of grade into our level-1 model.

What about the variation at level-2

Intercept-only model

$$\sigma_0^2 = 2.952$$

Linear grade model

$$\sigma_0^2 = 3.183$$

The unexplained variation at level-2 actually increased! This is just a mathematical artifact of the estimation.

In a more practical sense, we wouldn't really be too interested in the level-2 variation at this point. We are only adding predictors to the level-1 model, so that is the level we expect to have an impact on the variation at.

Let's also examine the fixed-effect estimates from `lmer()`.

```
> summary(lmer.1)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.41330    0.24403   5.791
grade        0.74666    0.05293  14.107
```

While the evidence from comparing the variance components suggests that by including a linear-effect of grade-level we reduced the unexplained within-student variation, one question we might have is: whether or not this reduction is more than we would expect by chance.

In fixed-effects only regression, we could answer this by examining the *p*-value associated with the particular regression coefficient. The `lmer()` function, however, does not produce *p*-values for individual coefficients.

Likelihood Ratio Test for Comparing Models

There are valid reasons that `lmer()` does not provide p-values. This was a design-choice by tits programmers. Statisticians strongly disagree about the effectiveness of the tests for individual coefficients in mixed-effects models.

An alternative method of inference (that has more universal agreement) is based on the deviance statistic. These tests allow you to both examine individual coefficients, and also sets of coefficients. They also produce better Type I error rates.

Recall that the log-likelihood gives the joint likelihood of the data given the model. If you fit multiple models, the idea is that higher likelihoods correspond to better fitting (more likely) models.

```
> logLik(lmer.0)
'log Lik.' -504.6293 (df=3)

> logLik(lmer.1)
'log Lik.' -436.352 (df=4)
```

Recall that a saturated model will give perfect fit and would be the "BEST" model one could fit in terms of residual fit. The deviance measures the difference in log-likelihood between the model under consideration and this THEORETICALLY BEST POSSIBLE model.

$$\text{Deviance} = -2\left[\text{Log-Likelihood}_{\text{Model}} - \text{Log-Likelihood}_{\text{Saturated Model}}\right]$$

We multiply by –2 to allow for testing under particular statistical distributions.

The likelihood that the data fit the model in the saturated model is 1. Thus, the log-likelihood is 0.

$$\text{Deviance} = -2\left[\text{Log-Likelihood}_{\text{Model}} - 0\right]$$

$$\text{Deviance} = -2 \times \text{Log-Likelihood}_{\text{Model}}$$

Smaller deviance values indicate the fitted-model is close to this BEST model.

$$\text{Deviance} = -2 \times \text{Log-Likelihood}_{\text{Model}}$$

```
> -2 * logLik(lmer.0)[1]
[1] 1009.259

> -2 * logLik(lmer.1)[1]
[1] 872.704
```

Here, the model that includes the linear-effect of grade has a lower deviance than the intercept-only model. This suggests that the more complex model has better fit than the less complex model.

HOWEVER…more complex models always have better fit than simpler models.

NOTE: You cannot directly interpret the sign nor magnitude of a deviance statistic.

We can compare deviances by computing the difference between them.

$$\begin{aligned}\Delta\text{Deviance} &= \text{Deviance}_{\text{Model A}} - \text{Deviance}_{\text{Model B}} \\ &= -2(\text{Log-Likelihood}_{\text{Model A}}) - \big[-2(\text{Log-Likelihood}_{\text{Model B}})\big] \\ &= -2\big[\text{Log-Likelihood}_{\text{Model A}} - \text{Log-Likelihood}_{\text{Model B}}\big]\end{aligned}$$

Remember that the difference of two logarithms is equal to the logarithm of a ratio.

$$\log(A) - \log(B) = \log(\frac{A}{B})$$

Using this, we can re-write the difference in deviance as a ratio.

$$\Delta\text{Deviance} = -2\ln\left(\frac{\text{Likelihood}_{\text{Model A}}}{\text{Likelihood}_{\text{Model B}}}\right)$$

The statistical test, referred to as the **Likelihood Ratio Test** (LRT) is a test of the ratio of the two likelihoods relative to the difference in complexity as measured through the degrees of freedom. This test evaluates the null hypothesis that the difference in deviances is 0.

$$H_0 : \Delta\text{Deviance} = 0$$

From the `logLik()` function we saw that the deviance for the unconditional intercept-only model was 1009.259 and the deviance for the unconditional linear-growth model was 872.704. The difference in deviances is 136.55.

The complexity of the models are given by their degrees of freedom. From the `logLik()` function we saw that the unconditional intercept-only model was estimating 3 parameters, and the unconditional linear-growth model was estimating 4 parameters. The difference here is 1 additional parameter for the unconditional linear-growth model.

To fit it we use the `anova()` function, and give it both models as arguments.

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

Data: vocab_long
Models:
lmer.0: score ~ 1 + (1 | id)
lmer.1: score ~ 1 + grade + (1 | id)
       Df    AIC     BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer.0  3 1015.3 1025.89 -504.63   1009.3
lmer.1  4  880.7  894.88 -436.35    872.7 136.55      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$$\chi^2(1) = 136.55,\ p < .001$$

Based on the results of the LRT, we reject the null hypothesis that the two models fit equally well. It is likely that the unconditional linear growth model fits better than the unconditional intercept-only model.

There are several caveats to using the LRT in practice.

- The models must be estimated using **identical data**. In other words, if a case is omitted in the estimation of one model, it must also be omitted in the estimation of another model.
- The more simple model **must be nested** within the more complex model.
- The models must be **estimated using full maximum likelihood (ML)**. You should not use this test if you have estimated the parameters using reduced maximum likelihood estimation (REML).

Here we add a quadratic fixed-effect of grade, but keep the RE to that of RE for intercept.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \beta_2^*(\text{Grade}_{ij}^2) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$
$$\beta_1^* = \beta_1$$
$$\beta_2^* = \beta_2$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta_2(\text{Grade}_{ij}^2) + b_{0i} + \epsilon_{ij}$$

```
> lmer.2 = lmer(score ~ 1 + grade + I(grade^2) + (1 | id),
     data = vocab_long, REML = FALSE)

> logLik(lmer.2)
'log Lik.' -428.3619 (df=5)

> -2 * logLik(lmer.2)[1]
[1] 856.7239
```

Now we can test the unconditional quadratic growth model relative to the unconditional linear-growth model using the LRT

```
> anova(lmer.1, lmer.2)

Data: vocab_long
Models:
lmer.1: score ~ 1 + grade + (1 | id)
lmer.2: score ~ 1 + grade + I(grade^2) + (1 | id)
       Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer.1  4 880.70 894.88 -436.35   872.70
lmer.2  5 866.72 884.45 -428.36   856.72 15.98      1  6.401e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$$\chi^2(1) = 15.98,\ p < .001$$

Based on the results of the LRT, we reject the null hypothesis that the two models fit equally well. There is evidence that the unconditional quadratic growth model fits better than the unconditional linear growth model.

One problem with the LRT is that it is an assymptotic test. That is it is only approximately distributed as a chi-square distribution under certain conditions. Because of this, the *p*-values are only approximate.

Long (2012) points out that because of this approximation, which only seems valid with large sample sizes, it is just as appropriate to make a "descriptive inference" for the fixed-effects by evaluating the size of the *t*-statistic associated with the fixed-effect. If the t-statistic is greater than 2 (or less than –2), the fixed-effect is likely to be statistically important.

```
> summary(lmer.1)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.41330    0.24403   5.791
grade        0.74666    0.05293  14.107
```

Here the *t*-statistic associated with the linear-effect of grade is 5.79, which indicates that we should adopt this effect into the model, over and above the intercept.

```
> summary(lmer.2)

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.18158    0.24954   4.735
grade        1.44181    0.17770   8.114
I(grade^2)  -0.23172    0.05676  -4.082
```

Here the *t*-statistic associated with the quadratic-effect of grade is –4.08, which indicates that we should adopt this effect into the model, over and above the linear effect.

Let's fit a model that includes fixed-effects of intercept and linear grade level, and random-effects for intercept and linear grade level.

Level-1 Model

$$\text{Vocabulary}_{ij} = \beta_0^* + \beta_1^*(\text{Grade}_{ij}) + \epsilon_{ij}$$

Level-2 Models

$$\beta_0^* = \beta_0 + b_{0i}$$
$$\beta_1^* = \beta_1 + b_{1i}$$

Mixed-Effects Model

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + b_{0i} + b_{1i}(\text{Grade}_{ij}) + \epsilon_{ij}$$

How can we tell if we should adopt the RE for grade level?

```
> lmer.3 = lmer(score ~ 1 + grade + (1 + grade | id), data = vocab_long, REML = FALSE)

> summary(lmer.3)

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 id       (Intercept) 3.138e+00 1.771334
          grade       7.514e-05 0.008668 1.00
 Residual             8.963e-01 0.946748
Number of obs: 256, groups:  id, 64
```

Unfortunately, there is not an analog of the *t*-statistic evaluation for random effects.

We can however use the LRT…at least to compare nested models.

Computing the log-likelihood and deviance for this model…

```
> logLik(lmer.2)
'log Lik.' -436.3394 (df=6)

> -2 * logLik(lmer.2)[1]
[1] 872.6787
```

Note the with this model, even though it appears as if we added a single additional term (the RE of slope), we actually added TWO additional terms to estimate 6 total parameters.

```
> anova(lmer.1, lmer.3)

Data: vocab_long
Models:
lmer.1: score ~ 1 + grade + (1 | id)
lmer.3: score ~ 1 + grade + (1 + grade | id)
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lmer.1  4 880.70 894.88 -436.35   872.70
lmer.3  6 884.68 905.95 -436.34   872.68 0.0253      2     0.9874
```

$$\chi^2(2) = 0.025,\ p = 0.987$$

Based on the results of the LRT, we fail to reject the null hypothesis that the two models fit equally well. There is no evidence that the model that includes RE for both the intercept and slope terms fits any better than the model that only includes RE for the intercept.

SIDE NOTE ON PARAMETERS

In the unconditional means model, we were estimating three parameters:

- A fixed-effect of intercept
- A variance estimate for the random-effects of intercept
- A variance estimate for the level-1 residuals

$$\hat{\beta}_0, \quad \hat{\sigma}_0^2, \quad \hat{\sigma}_\epsilon^2$$

In the unconditional linear growth model, we were estimating four parameters:

- A fixed-effect of intercept
- A fixed-effect of linear growth
- A variance estimate for the random-effects of intercept
- A variance estimate for the level-1 residuals

$$\hat{\beta}_0, \quad \hat{\beta}_1, \quad \hat{\sigma}_0^2, \quad \hat{\sigma}_\epsilon^2$$

In the unconditional quadratic growth model, we were estimating five parameters:

- A fixed-effect of intercept
- A fixed-effect of linear growth
- A fixed-effect of quadratic growth
- A variance estimate for the random-effects of intercept
- A variance estimate for the level-1 residuals

$$\hat{\beta}_0, \quad \hat{\beta}_1, \quad \hat{\beta}_2, \quad \hat{\sigma}_0^2, \quad \hat{\sigma}_\epsilon^2$$

In the linear growth model with fixed- and random-effects of intercept and slope, we were estimating six parameters:

- A fixed-effect of intercept
- A fixed-effect of linear growth
- A variance estimate for the random-effects of intercept
- A variance estimate for the random-effects of linear growth
- A covariance between the REs for intercept and REs for slopes
- A variance estimate for the level-1 residuals

$$\hat{\beta}_0, \quad \hat{\beta}_1, \quad \hat{\sigma}_0^2, \quad \hat{\sigma}_1^2, \quad \hat{\sigma}_{0,1}, \quad \hat{\sigma}_\epsilon^2$$

What if we wanted to compare this model to the unconditional quadratic growth model?

```
> anova(lmer.2, lmer.3)

Data: vocab_long
Models:
lmer.2: score ~ 1 + grade + I(grade^2) + (1 | id)
lmer.3: score ~ 1 + grade + (1 + grade | id)
       Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer.2  5 866.72 884.45 -428.36   856.72
lmer.3  6 884.68 905.95 -436.34   872.68     0      1          1
```

The results here are gibberish. There is a difference in deviance, but the chi-squared value is 0. The problem is that these two models are not nested.

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + \beta_2(\text{Grade}^2_{ij}) + b_{0i} + \epsilon_{ij}$$

$$\text{Vocabulary}_{ij} = \beta_0 + \beta_1(\text{Grade}_{ij}) + b_{0i} + b_{1i}(\text{Grade}_{ij}) + \epsilon_{ij}$$

You can not set parameters equal to zero in one model and end up with the parameters in the other model!

References and Source Material

- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis*. New York: Wiley.
- Kohli, N. (2016). *Linear mixed-effects models (course notes)*. Minneapolis, MN: Author.
- Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. Thousand Oaks, CA: Sage. (Minneapolis reading data)

- Graphics used from Open Clip Art Library (https://openclipart.org/)