MLM Nested Handout 5 (Spring 2022)

Nested Longitudinal Data

1. Quick introduction to longitudinal data
   a. Longitudinal (panel) data models involve repeated measurement of the same outcome at different time points.
   b. Using our nested data modeling framework, we can view those repeated measures as nested *within subject*.
   c. The simplest growth curve model could be written:

   $$Y_{ti} = b_0 + \delta_{0i} + \varepsilon_{ti},$$

   with $\delta_{0i} \sim N(0, \sigma^2_{\delta_0})$ and $\varepsilon_{ti} \sim N(0, \sigma^2_\varepsilon)$. We assume δ_{0i}, ε_{ti} are independent, throughout.
      i. There is an overall mean for the outcome (here, b_0), and it is constant over time. This is the *unconditional means model* in this setting.
      ii. You should be familiar with this as a random intercept model, with a random intercept for each subject.
      iii. No 'growth' is predicted by this model, either at the population (mean) level or individual level.
   d. This model is clearly misspecified: we expect *time* to play some role.
      i. Time can be lots of different things: time since a study began, age, years of schooling, grade level, years on the job, etc. We often represent it with a 'specially named' variable, TIME, to reflect this. So the time at occasion t for subject i is $TIME_{ti}$.
         1. Under this formulation, $TIME_{ti}$ and occasion 't' are not necessarily the same (e.g., infant's age in days).
      ii. Adding time to the basic model for growth, we have:

      $$Y_{ti} = b_0 + b_1 TIME_{ti} + \delta_{0i} + \varepsilon_{ti},$$

      with $\delta_{0i} \sim N(0, \sigma^2_{\delta_0})$ and $\varepsilon_{ti} \sim N(0, \sigma^2_\varepsilon)$.
         1. This is not the unconditional growth model (UGM) described by Singer & Willett; the usual UGM allows individual-specific growth as well.
   e. Of course, other covariates can also influence the mean:

   $$Y_{ti} = b_0 + b_1 TIME_{ti} + b_2 X_{2i} + b_3 X_{3ti} + \delta_{0i} + \varepsilon_{ti}$$

      i. Here, subscripts on the X terms reflect whether or not they are time-dependent. Those with subscripts 'ti' allow for change over time. Those with subscript 'i' effectively repeat (in the dataset) for all occasions t.
      ii. X_2 above is constant for each subject (e.g., race); X_3 above can vary over time (e.g., education).
   f. Complexity can be added to any of the three components:
      i. More covariates, interactions or 'higher order functions of TIME' (quadratics, cubics) may be added, potentially improving our predictions for the mean, or fixed effects.
      ii. The systematic variation, δ_{0i} above, can be made more complex (through additional terms and the relationships between them).
      iii. The residual error structure can also be made more complex.
   g. We now look at the level or hierarchical form of notation for this type of model. In the top level, we usually limit the covariates to those that either change over time (and thus require reference

to both coefficients '$_{ti}$') or vary at the next 'level,' which in this case is the subject-level. So our model is greatly simplified, and can be written:

$$\text{Level 1: } Y_{ti} = \pi_{0i} + \pi_{1i}TIME_{ti} + \pi_{2i}X_{2ti} + \varepsilon_{ti}$$
$$\text{Level 2: } \pi_{0i} = b_0 + b_3X_{3i} + \delta_{0i};$$
$$\pi_{1i} = b_1;$$
$$\pi_{2i} = b_2;$$
$$\delta_{0i} \sim N(0, \sigma^2_{\delta_0}), \ \varepsilon_{ti} \sim N(0, \sigma^2_\varepsilon), \text{ indep.}$$

1. The use of π here will eventually make things simpler when we add levels.

h. It may seem odd to see X_3 as part of the subject-specific constant, but that's all that a subject-constant predictor can do (a single shift).

i. NOTE: we reiterate that we assume normal, independent distributions for the random components: $\delta_{0i} \sim N(0, \sigma^2_{\delta_0})$ and $\varepsilon_{ti} \sim N(0, \sigma^2_\varepsilon)$.

j. A random slope model is easily added at level 2 as follows:

$$\text{Level 2: } \pi_{0i} = b_0 + b_3X_{3i} + \delta_{0i}$$
$$\pi_{1i} = b_1 + \delta_{1i}$$
$$\pi_{2i} = b_2$$
$$\delta_{0i} \sim N(0, \sigma^2_{\delta_0}), \ \delta_{1i} \sim N(0, \sigma^2_{\delta_1}), \ \varepsilon_{ti} \sim N(0, \sigma^2_\varepsilon), \text{ indep.}$$

2. A common design to study *change over time* might also involve a nested sampling structure, such as students within classrooms within schools, or individuals within households within buildings within census tracts.

   a. At the individual level, the subject is *tracked over time*.

   b. We have just conceptualized time as another level of grouping, so this type of model is not substantially different from any nested model – except that time has an "elevated" status as a covariate.

      i. Elevated simply means that it is the first covariate we are likely to consider. *Why? Because it is usually a bad idea to leave out a time trend.*

      ii. *However, to start with a simpler model, we will again leave out the time trend!*

   c. Let us look at an unconditional means model and continue to use level notation. We will introduce an additional level of grouping. Let's assume a *single* school, and at first, a sample of classrooms indexed by j within it. We proceed mostly by adding a 'j' to the prior longitudinal model specification (but also note the use of π and β):

$$\text{Level 1: } Y_{tij} = \pi_{0ij} + \varepsilon_{tij}$$
$$\text{Level 2: } \pi_{0ij} = \beta_{0j} + \delta_{0ij}$$
$$\text{Level 3: } \beta_{0j} = b_0 + \eta_{0j}$$
$$\eta_{0j} \sim N(0, \sigma^2_{\eta_0}), \ \delta_{0ij} \sim N(0, \sigma^2_{\delta_0}), \ \varepsilon_{tij} \sim N(0, \sigma^2_\varepsilon), \text{ indep.}$$

We intentionally leave out the random slope in TIME to keep the model "simple." You can call $\eta_{0j}, \ \delta_{0ij}, \varepsilon_{tij}$ the level 3, 2, and 1 errors, respectively.

   d. You might feel more comfortable with the model in composite notation:

$$Y_{tij} = b_0 + \delta_{0ij} + \eta_{0j} + \varepsilon_{tij}$$

      i. Let's add a time trend. Where does it go? Since time can vary at the individual level, it would go on level 1:

ii. Let $TIME_{tij}$ be the time of measurement associated with occasion t on person i in classroom j. TIME could be grade, age, etc.

iii. Then this model at least contains a fixed effect for TIME (displayed in composite form):

$$Y_{tij} = b_0 + b_1 TIME_{tij} + \delta_{0ij} + \eta_{0j} + \varepsilon_{tij}$$

with $\delta_{0ij} \sim N(0, \sigma_{\delta_0}^2)$, $\eta_{0j} \sim N(0, \sigma_{\eta_0}^2)$, and $\varepsilon_{tij} \sim N(0, \sigma_\varepsilon^2)$ independently.

e. What if we have random slopes (in TIME), say across classrooms:

$$Y_{tij} = b_0 + b_1 TIME_{tij} + \delta_{0ij} + \eta_{0j} + \eta_{1j} TIME_{tij} + \varepsilon_{tij}$$

with $\delta_{0ij} \sim N(0, \sigma_{\delta_0}^2)$, $\eta_{0j} \sim N(0, \sigma_{\eta_0}^2)$, $\eta_{1j} \sim N(0, \sigma_{\eta_1}^2)$, and $\varepsilon_{tij} \sim N(0, \sigma_\varepsilon^2)$ independently.[1]

3. Adding a new level – schools – should be relatively straightforward: just add a final subscript, k.

$$Y_{tijk} = b_0 + b_1 TIME_{tijk} + \delta_{0ijk} + \eta_{0jk} + \eta_{1jk} TIME_{tijk} + \zeta_{0k} + \varepsilon_{tijk}$$

with $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\eta_{0jk} \sim N(0, \sigma_{\eta_0}^2)$, $\eta_{1jk} \sim N(0, \sigma_{\eta_1}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$ and $\varepsilon_{tijk} \sim N(0, \sigma_\varepsilon^2)$ independently.

b. Another way to understand this model is that we have really just "inserted" a person-specific level shift in the form of δ_{0ijk} into our previous model specifications and prepended a subscript 't' to our indices.

4. We now examine several examples of models such as those sketched above.

a. We have modified the classroom data to reflect its longitudinal nature. Namely, we have a kindergarten and first grade test score, so we can think of this as a within-subject repeated measure

i. It takes slightly complicated code to generate a data frame with a separate entry for each school year from one that contains both, primarily because the names of variables must conform to a specific convention. We then take what is known as a 'wide' format to a 'long' (or person-period) format as follows:

```
dat <- read.csv("../Datasets/classroom.csv")
# *make a person-period file
dat$math0 <- dat$mathkind
dat$math1 <- dat$mathkind + dat$mathgain
class_pp <- reshape(dat, varying = c("math0", "math1"), v.names = "math", timevar = "year",
    times = c(0, 1), direction = "long")
```

b. Our baseline model has TIME as the only fixed effect predictor. We include random (intercept, or level) effects for schools and students but not classrooms.

i. The random effect for students captures some aspects of the average performance, or level, of each student, net of all else.

1. Netting out a student-specific level is similar to controlling for a pre-test, or more accurately, unmeasured, pre-existing abilities or factors.

ii. Composite notation:

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \zeta_{0k} + b_1 TIME_{tijk} + \varepsilon_{tijk},$$

and we assume $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, and $\varepsilon_{tijk} \sim N(0, \sigma_\varepsilon^2)$.

iii. R code and model fit:

```
fit.M00 <- lmer(math ~ year + (1 | schoolid/childid), data = class_pp)
print(summary(fit.M00))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (1 | schoolid/childid)
```

---

[1] Independence between slope and intercept within classroom is a strong assumption that can be changed to allow correlation.

```
##    Data: class_pp
##
## REML criterion at convergence: 23554.7
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -4.7492 -0.4811  0.0085  0.4881  3.4957
##
## Random effects:
##  Groups             Name         Variance Std.Dev.
##  childid:schoolid (Intercept) 702.0     26.50
##  schoolid         (Intercept) 307.5     17.54
##  Residual                     599.1     24.48
## Number of obs: 2380, groups:  childid:schoolid, 1190; schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  465.118     2.042  117.023  227.74   <2e-16 ***
## year          57.566     1.003 1189.000   57.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.246
```

4. (cont.)

    b. (cont.)

        1. Comments:

           a. Math scores are going up, on average, net of school and pupil effects

           b. The variance components are: $\sigma^2_{\delta_0} = 702$ (between student, within-school variation); $\sigma^2_{\zeta_0} = 308$ (between-school variation for the average student in each school); $\sigma^2_\varepsilon = 599$ (within-school, within-student variation; this is variation over time).

           c. *Notice the slightly different way we frame these effects in this context.*

           d. *Notice that uncertainty within student is as large as between.*

5. Our next model adds a random intercept for classrooms, to see if it is 'needed.'

    i. Composite notation:

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \eta_{0jk} + \zeta_{0k} + b_1 TIME_{tijk} + \varepsilon_{tijk}$$

    and assume $\delta_{0ijk} \sim N(0, \sigma^2_{\delta_0})$, $\eta_{0jk} \sim N(0, \sigma^2_{\eta_{00}})$, $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, and $\varepsilon_{tijk} \sim N(0, \sigma^2_\varepsilon)$, independently.

    ii. R code and model fit (with a preliminary comment).

        1. This is the first time we have three levels of nesting, so it is worth examining the nested structure closely.

```
fit.MO <- lmer(math ~ year + (1 | schoolid) + (1 | schoolid:classid) + (1 | classid:childid),
    data = class_pp)  #shortcut to get model to fit in reasonable amount of time.
# alternative: fit.MO <-
# lmer(math~year+(1|schoolid/classid/childid),data=class_pp) #shortcut to get
# model to fit in reasonable amount of time.
print(summary(fit.MO))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (1 | schoolid) + (1 | schoolid:classid) + (1 |
```

```
##      classid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23552.8
##
## Scaled residuals:
##      Min      1Q  Median      3Q     Max
## -4.7533 -0.4887  0.0124  0.4941  3.4835
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:childid  (Intercept) 675.98   26.000
##  schoolid:classid (Intercept)  39.01    6.246
##  schoolid         (Intercept) 294.31   17.156
##  Residual                     599.05   24.476
## Number of obs: 2380, groups:
## classid:childid, 1190; schoolid:classid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  465.062      2.045   117.683  227.46   <2e-16 ***
## year          57.566      1.003  1189.005   57.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.245
```

```
# lrtest M00 M0
anova(fit.M00, fit.M0, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M00: math ~ year + (1 | schoolid/childid)
## fit.M0: math ~ year + (1 | schoolid) + (1 | schoolid:classid) + (1 | classid:childid)
##         npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.M00    5 23565 23594 -11777    23555
## fit.M0     6 23565 23599 -11776    23553 1.8954  1     0.1686
```

5. (cont.)
    iv. The LR test is non-significant.
        1. However, we will leave in classroom effects for discussion purposes.
    v. Ask yourself: is there enough information in the data to estimate all of the parameters in this model? $\sigma^2_{\delta_0}$ in particular is the most challenging. Would you be able to estimate a random slope in time at the person-level? Maybe; think about degrees of freedom.
6. Our next model adds a random slope effect (in TIME) for schools. This is closer to what Singer & Willett call an unconditional growth model: we have differences (at the school level) in how those trajectories change over time... *or do we?*
    i. Composite notation:

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \eta_{0jk} + \zeta_{0k} + (b_1 + \zeta_{1k})TIME_{tijk} + \varepsilon_{tijk}$$

    and assume $\delta_{0ijk} \sim N(0, \sigma^2_{\delta_0})$, $\eta_{0jk} \sim N(0, \sigma^2_{\eta_0})$, $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, $\zeta_{1k} \sim N(0, \sigma^2_{\zeta_1})$ and $\varepsilon_{tijk} \sim N(0, \sigma^2_\varepsilon)$.
    ii. R code and model fit:

```
fit.M1 <- lmer(math ~ year + (year || schoolid) + (1 | schoolid:classid) + (1 | classid:childid),
    data = class_pp)  #shortcut to get model to fit in reasonable amount of time.
print(summary(fit.M1))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (year || schoolid) + (1 | schoolid:classid) + (1 |
##      classid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23527.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7699 -0.4749  0.0113  0.4644  3.5957
##
## Random effects:
##  Groups          Name        Variance Std.Dev.
##  classid.childid (Intercept) 699.50   26.448
##  schoolid.classid (Intercept)  38.33    6.191
##  schoolid        year         88.57    9.411
##  schoolid.1      (Intercept) 312.12   17.667
##  Residual                    552.23   23.499
## Number of obs: 2380, groups:
## classid:childid, 1190; schoolid:classid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.028      2.084 110.506  223.12   <2e-16 ***
## year          57.501      1.370  99.876   41.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.177
```

```
# lrtest M0 M1
anova(fit.M0, fit.M1, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M0: math ~ year + (1 | schoolid) + (1 | schoolid:classid) + (1 | classid:childid)
## fit.M1: math ~ year + (year || schoolid) + (1 | schoolid:classid) + (1 | classid:childid)
##        npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## fit.M0    6 23565 23599 -11776    23553
## fit.M1    7 23541 23582 -11764    23527 25.52  1  4.377e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6. (cont.)
    iv. We have evidence from the LRT that random slopes add explanatory power (test of H0: $\sigma^2_{\zeta_1} = 0$):
    v. The 'new' variance components are: $\sigma^2_{\zeta_0} = 312$ and $\sigma^2_{\zeta_1} = 89$.
        1. The corresponding school-level effects enter the model in these two terms: $\zeta_{0k} + \zeta_{1k}TIME_{tijk}$.
        2. This translates into two different between-school variances (call this a function, $V_{B\text{-}S}(t)$), one

at time 0 and one at time 1:

    a. $V_{B-S}(0) = \sigma_{\zeta_0}^2 + 0 \cdot \sigma_{\zeta_1}^2 = 312$

    b. $V_{B-S}(1) = \sigma_{\zeta_0}^2 + 1^2 \cdot \sigma_{\zeta_1}^2 = 312 + 89 = 401$

We will eventually check whether the variance between schools is statistically larger in first grade. *It's important to be aware of how heteroscedasticity is implicitly included in a model this way, through the random slope terms.*

7. Our next model *moves* the random slope effect (in TIME) to classrooms – a much smaller unit on which to base estimation. Again, we examine whether we have differences (at the classroom level) in how those trajectories change over time.

    i. Composite notation:

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \eta_{0jk} + \zeta_{0k} + (b_1 + \eta_{1jk})TIME_{tijk} + \varepsilon_{tijk}$$

    ii. For both notations, we assume $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\eta_{0jk} \sim N(0, \sigma_{\eta_0}^2)$, $\eta_{1jk} \sim N(0, \sigma_{\eta_1}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$ and $\varepsilon_{tijk} \sim N(0, \sigma_{\varepsilon}^2)$.

    iii. R code and model fit:

```
fit.M2 <- lmer(math ~ year + (1 | schoolid) + (year || schoolid:classid) + (1 | classid:childid),
    data = class_pp)  #shortcut to get model to fit in reasonable amount of time.
print(summary(fit.M2))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (1 | schoolid) + (year || schoolid:classid) + (1 |
##     classid:childid)
##     Data: class_pp
##
## REML criterion at convergence: 23530.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7720 -0.4608  0.0021  0.4687  3.6375
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  classid.childid    (Intercept) 710.07   26.647
##  schoolid.classid   year        117.47   10.838
##  schoolid.classid.1 (Intercept)  21.57    4.644
##  schoolid           (Intercept) 310.69   17.626
##  Residual                       534.75   23.125
## Number of obs: 2380, groups:
## classid:childid, 1190; schoolid:classid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.112      2.062 111.926  225.58   <2e-16 ***
## year          57.311      1.162 313.622   49.33   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.192
```

```
# lrtest M0 M2
anova(fit.M0, fit.M2, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M0: math ~ year + (1 | schoolid) + (1 | schoolid:classid) + (1 | classid:childid)
## fit.M2: math ~ year + (1 | schoolid) + (year || schoolid:classid) + (1 | classid:childid)
##        npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.M0    6 23565 23599 -11776    23553
## fit.M2    7 23545 23585 -11765    23531 21.956  1  2.789e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7. (cont.)

    iv. LRT suggests that random slopes for classrooms are needed (when they have been omitted at the school level):

    v. The classroom-based variance components are: $\sigma^2_{\eta_0} = 22$ and $\sigma^2_{\eta_1} = 117$.

        1. The corresponding classroom-level effects enter the model in these two terms: $\eta_{0jk} + \eta_{1jk}TIME_{tijk}$.

        2. This translates into two different between-classroom variances ($V_{B\text{-}C}(t)$):

            a. $V_{B-C}(0) = \sigma^2_{\eta_0} + 0 \cdot \sigma^2_{\eta_1} = 22$

            b. $V_{B-C}(1) = \sigma^2_{\eta_0} + 1 \cdot \sigma^2_{\eta_1} = 22 + 117 = 139$

            c. This dramatic difference between the two grade levels really should (and will) be tested formally.

            d. A graph can be made of V(t) or V(X), any X.

8. Our next model contains a random slope effect (in TIME) for both classrooms and schools. Again, we examine whether we have differences (at both levels) in how those trajectories change over time.

    i. Composite notation:

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \eta_{0jk} + \zeta_{0k} + (b_1 + \eta_{1jk} + \zeta_{1k})TIME_{tijk} + \varepsilon_{tijk}$$

    assume $\delta_{0ijk} \sim N(0, \sigma^2_{\delta_0})$, $\eta_{0jk} \sim N(0, \sigma^2_{\eta_0})$, $\eta_{1jk} \sim N(0, \sigma^2_{\eta_1})$, $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, $\zeta_{1k} \sim N(0, \sigma^2_{\zeta_1})$ and $\varepsilon_{tijk} \sim N(0, \sigma^2_{\varepsilon})$.

    ii. R code and model fit:

```
fit.M3 <- lmer(math ~ year + (year || schoolid/classid) + (1 | classid:childid),
    data = class_pp)  #shortcut to get model to fit in reasonable amount of time.
print(summary(fit.M3))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (year || schoolid/classid) + (1 | classid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23521.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7783 -0.4702  0.0115  0.4617  3.6485
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  classid.childid    (Intercept) 711.99   26.683
##  classid.schoolid   year         60.92    7.805
##  classid.schoolid.1 (Intercept)  25.32    5.032
##  schoolid           year         69.36    8.328
```

```
## schoolid.1        (Intercept) 318.69   17.852
## Residual                      529.96   23.021
## Number of obs: 2380, groups:
## classid:childid, 1190; classid:schoolid, 312; schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  465.071      2.085  109.673  223.08   <2e-16 ***
## year          57.349      1.371   98.610   41.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.169
```

The variance components generate a fair amount of heteroscedasticity. The significance of the random slope terms is confirmed through these two LRTs:

```
anova(fit.M1, fit.M3, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M1: math ~ year + (year || schoolid) + (1 | schoolid:classid) + (1 | classid:childid)
## fit.M3: math ~ year + (year || schoolid/classid) + (1 | classid:childid)
##        npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.M1    7 23541 23582 -11764    23527
## fit.M3    8 23538 23584 -11761    23522 5.5304  1    0.01869 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit.M2, fit.M3, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M2: math ~ year + (1 | schoolid) + (year || schoolid:classid) + (1 | classid:childid)
## fit.M3: math ~ year + (year || schoolid/classid) + (1 | classid:childid)
##        npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.M2    7 23545 23585 -11765    23531
## fit.M3    8 23538 23584 -11761    23522 9.0945  1   0.002564 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

NOTE: You should be able to state what the assumptions are (the base model) and what the null hypothesis is (whatever is added has coefficients equal to zero) in both LRTs, above.

Importantly, we do not know, from the above tests, whether each of the random *intercepts* is still needed.

In particular, it seems unlikely that the random intercept for classrooms is needed, once the same is included for schools. We can test this formally by fitting yet another model:

```
fit.MM3 <- lmer(math ~ year + (year || schoolid) + (0 + year | schoolid:classid) +
    (1 | classid:childid), data = class_pp)  #shortcut to get model to fit in reasonable amount of time
print(summary(fit.MM3))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (year || schoolid) + (0 + year | schoolid:classid) +
```

```
##      (1 | classid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23522.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7799 -0.4658  0.0112  0.4606  3.6624
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid.childid  (Intercept) 727.57   26.974
##  schoolid.classid year         64.17    8.011
##  schoolid         year         68.43    8.272
##  schoolid.1       (Intercept) 327.61   18.100
##  Residual                     529.30   23.006
## Number of obs: 2380, groups:
## classid:childid, 1190; schoolid:classid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.102      2.084 109.253  223.22   <2e-16 ***
## year          57.338      1.371  98.538   41.81   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.168
```

```
# lrtest MM3 M3
anova(fit.M3, fit.MM3, refit = F)
```

```
## Data: class_pp
## Models:
## fit.MM3: math ~ year + (year || schoolid) + (0 + year | schoolid:classid) + (1 | classid:childid)
## fit.M3: math ~ year + (year || schoolid/classid) + (1 | classid:childid)
##         npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.MM3    7 23536 23577 -11761    23522
## fit.M3     8 23538 23584 -11761    23522 0.7306  1     0.3927
```

It seems we do not need the level effects for classrooms (so school level intercept suffices), with school effects in the model, but the random slopes are still useful at the classroom level. We have not tested the last point formally, but we can expect the likelihood to essentially be unchanged with the random classroom intercept dropped; under this scenario, the prior LRT that examined the 'need' for random classroom slopes should prevail.

9. One last thing to explore is whether our models for heteroscedasticity are consistent with prior findings.
   i. These models assign greater variation at the classroom or school level to first grade math scores.
   ii. We check this by reloading the original data (wide format) and running some simple models (child-level effects cannot enter into models such as these as there is no replication at that level):

```
dat <- read.csv("../Datasets/classroom.csv")
# gen math1st = mathgain + mathkind
dat$math1st <- dat$mathkind + dat$mathgain
# xtmixed mathkind || schoolid: || classid:, var
```

```
fit.M_K <- lmer(mathkind ~ 1 + (1 | schoolid/classid), data = dat)
print(summary(fit.M_K))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: mathkind ~ 1 + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 12085.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8198 -0.5749  0.0012  0.6303  3.5893
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:schoolid (Intercept)   32.0    5.657
##  schoolid         (Intercept)  352.8   18.784
##  Residual                     1323.4   36.379
## Number of obs: 1190, groups:  classid:schoolid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.220      2.191 103.579   212.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# xtmixed math1st || schoolid: || classid:, var
fit.M_1 <- lmer(math1st ~ 1 + (1 | schoolid/classid), data = dat)
print(summary(fit.M_1))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ 1 + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 11944.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.1872 -0.6174 -0.0204  0.5821  3.8339
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:schoolid (Intercept)   85.46   9.244
##  schoolid         (Intercept)  280.68  16.754
##  Residual                     1146.80  33.864
## Number of obs: 1190, groups:  classid:schoolid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  522.540      2.037 104.407   256.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9. (cont.)
     i. The variation at the classroom level follows the basic trend of increasing in first grade, but the effect is muted.
    ii. The variation at the school level drops from kindergarten to first grade, suggesting a misspecification in our nested longitudinal models.
   iii. A reasonable addition is a correlation between the random intercept and slope at the school level in the nested longitudinal data (to leave this out is to potentially misspecify the model):

```
fit.M4 <- lmer(math ~ year + (year | schoolid) + (year || schoolid:classid) + (1 |
    classid:childid), data = class_pp)  #shortcut to get model to fit in reasonable amount of time.
print(summary(fit.M4))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (year | schoolid) + (year || schoolid:classid) +
##     (1 | classid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23512.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7195 -0.4674  0.0034  0.4598  3.5668
##
## Random effects:
##  Groups            Name        Variance Std.Dev. Corr
##  classid.childid   (Intercept) 714.68   26.734
##  schoolid.classid  year         60.44    7.774
##  schoolid.classid.1 (Intercept) 25.14    5.014
##  schoolid          (Intercept) 363.23   19.059
##                    year         89.36    9.453  -0.50
##  Residual                      525.06   22.914
## Number of obs: 2380, groups:
## classid:childid, 1190; schoolid:classid, 312; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.081      2.188 103.208  212.59   <2e-16 ***
## year          57.532      1.439  93.791   39.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.432
```

```
# lrtest M3 M4
anova(fit.M3, fit.M4, refit = F)
```

```
## Data: class_pp
## Models:
## fit.M3: math ~ year + (year || schoolid/classid) + (1 | classid:childid)
## fit.M4: math ~ year + (year | schoolid) + (year || schoolid:classid) + (1 | classid:childid)
##        npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit.M3    8 23538 23584 -11761    23522
```

```
## fit.M4    9 23531 23583 -11756    23513 8.8936  1   0.002862 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The LRT suggests we 'need' the correlation.

10. What does this new model imply, in terms of variance between schools at different grades?
    a. Using the usual formula for variance of a linear combination of (correlated, random) variables, if we have a combination of variables, such as $X + aY + c$, in which 'a' and 'c' are constant, but X & Y vary from observation to observation, or group to group, we have, $Var(X) + 2aCov(X,Y) + a^2 Var(Y)$ (anything constant that does not multiply something random drops out).
    b. In our example, the prediction for a school is

$$MATH_{tijk} = b_0 + \delta_{0ijk} + \eta_{0jk} + \zeta_{0k} + (b_1 + \eta_{1jk} + \zeta_{1k})TIME_{tijk}$$

but we are focusing on the school-to-school differences, so that everything other than $\zeta_{0k} + \zeta_{1k}TIME_{tijk}$ behaves like a constant; the expression for variance is

$$V_{B-S}(TIME_{tijk}) = \sigma_{\zeta_0}^2 + 2TIME_{tijk}\rho_{\zeta_0\zeta_1} + TIME_{tijk}^2\sigma_{\zeta_1}^2$$

yielding:
$$V_{B-S}(0) = \sigma_{\zeta_0}^2 + 2 \cdot 0 \cdot \rho_{\zeta_0\zeta_1} + 0^2 \cdot \sigma_{\zeta_1}^2 = 363$$

$$V_{B-S}(1) = \sigma_{\zeta_0}^2 + 2 \cdot 1 \cdot \rho_{\zeta_0\zeta_1} + 1^2 \cdot \sigma_{\zeta_1}^2 = 363 + 2 \cdot (-90) + 89 = 272$$

So the more complex model picks up the 'drop' in between-school variance in first grade.