Project A1+A2+B+C

Chongjun Liao, Jeremy Lu, Yu Wang, Xinming Dai

5/7/2022

Contents

A1, Chongjun Liao

0. We will use the classroom.csv data for this project.

a. math1st will be the outcome of interest for this first part

b. Recall that `math1st = mathkind + mathgain`
c. Read in the data (R: store as `dat`)
d. Fit all models using REML
e. It's best if you use `lmerTest::lmer` rather than `lme4::lmer` to call the MLM function. The former provides p-values for fixed effects in the summary.
f. There are 2 common error messages one can get from lmer calls: failed to converge (problem with hessian: negative eigenvalue; max|grad| = ...); and singularity. They may both be problematic in a real problem, but the latter suggests that a variance component is on the boundary of the parameter space.

1. In your discussion/writeup, consider the latter to be a "convergence problem" and ignore the former.

```
dat <- read.csv("~/Documents/GitHub/mlm_final_project/data/classroom.csv")
dat <- dat %>%
  mutate(math1st = mathkind + mathgain)
```

1. Estimate an Unconditional Means Model (UMM) with random intercepts for both schools and classrooms (nested in schools).

```
fit1 <- lmer( math1st ~ (1 | schoolid/classid), dat)
summary(fit1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ (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
```

a. Report the ICC for schools and the ICC for classrooms
   Answer: The ICC for schools is $\frac{\sigma^2_{\zeta_0}}{\sigma^2_{\zeta_0}+\sigma^2_{\eta_0}+\sigma^2_{\varepsilon}} = 0.1855203$ and the ICC for classrooms is $\frac{\sigma^2_{\eta_0}}{\sigma^2_{\zeta_0}+\sigma^2_{\eta_0}+\sigma^2_{\varepsilon}} = 0.0564856$.

b. **WRITE OUT THIS MODEL** using your preferred notation, but use the same choice of notation for the remainder of your project
c. Be mindful and explicit about any assumptions made.

$MATH1ST_{ijk} = b_0 + \zeta_{0k} + \eta_{0jk} + \varepsilon_{ijk}$, with $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, $\eta_{0jk} \sim N(0, \sigma^2_{\eta_0})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2_{\varepsilon})$, independently of one another, j represents classrooms and k represents *schools*.
2. ADD ALL School level predictors

```
fit2 <- lmer( math1st ~ housepov + (1 | schoolid/classid), dat)
anova(fit1,fit2, refit = T)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## fit1: math1st ~ (1 | schoolid/classid)
## fit2: math1st ~ housepov + (1 | schoolid/classid)
##      npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit1    4 11956 11976 -5973.9    11948
## fit2    5 11948 11973 -5968.8    11938 10.125  1   0.001463 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
wald.test(b = fixef(fit2), Sigma = summary(fit2)$vcov, Terms = 2)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 10.3, df = 1, P(> X2) = 0.0013
```

a. Report if adding the predictors as a block is justified
Answer: There is only one school-level predictor which is `housepov`, its p-value is $0.0017029 < 0.05$, and I do a LRT on model with and without the school-level predictor, the p-value is $0.0014627 < 0.05$. So it is reasonable to add school-level predictor. I also do the wald-test, the p-value is also < 0.05.

b. Report change in σ^2_ζ.
The change in σ^2_ζ is 280.6812733-250.9258585 = 29.7554148.

3. ADD ALL Classroom level predictors

```
fit3 <- lmer( math1st ~ yearstea + mathknow + mathprep + housepov + (1 | schoolid/classid),
              dat)
summary(fit3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## math1st ~ yearstea + mathknow + mathprep + housepov + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 10821
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5552 -0.6118 -0.0311  0.5863  3.8315
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:schoolid (Intercept)   94.36   9.714
##  schoolid         (Intercept)  223.31  14.943
##  Residual                      1136.43  33.711
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 532.29852    5.20495 228.85767 102.268  < 2e-16 ***
## yearstea      0.06193    0.14717 223.76570   0.421  0.67432
## mathknow      2.55143    1.44530 231.06560   1.765  0.07883 .
## mathprep     -0.75440    1.42809 203.20755  -0.528  0.59790
## housepov    -41.62117   14.08834 109.83230  -2.954  0.00383 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##          (Intr) yearst mthknw mthprp
## yearstea -0.264
## mathknow -0.052  0.030
## mathprep -0.666 -0.175  0.004
## housepov -0.568  0.077  0.082  0.032
```

```
wald.test(b = fixef(fit3), Sigma = summary(fit3)$vcov, Terms = 2:4)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 3.5, df = 3, P(> X2) = 0.32
```

a. Report if adding the predictors as a block is justified [must use WALD test, not LRT]
Answer: The Wald test generates a p-value = 0.32, which shows that we have no reason to add classroom-level predictors as a block. But it might be reasonable to include `mathknow` since it is significant according to the t-test.

b. Report change in σ^2_η and change in σ^2_ϵ.
Answer: The change in σ^2_η is 94.3625825-82.3601958 = 12.0023867 and change in σ^2_ϵ is 1136.4309806-1146.9548045 = -10.5238239.

c. Give a potential reason as to why σ^2_ϵ is reduced, but not σ^2_η?

One potential reason is that there are only 3~4 sampled student in each classroom. Since the sample size with each classroom is small, the classroom predictors describe aggregrate limited individual characteristics, which would explain student-level variation.

4. ADD (nearly) ALL student level predictors (but not `mathgain` or `mathkind`, as these are outcomes in this context).

```
fit4 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | schoolid/classid), dat)
summary(fit4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8581 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:schoolid (Intercept)   93.89   9.689
```

```
##  schoolid         (Intercept)  169.45  13.017
##  Residual                      1064.96  32.634
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.63041    5.31209  275.39010 101.585  < 2e-16 ***
## ses           10.05076    1.54485 1066.56211   6.506 1.18e-10 ***
## minority     -16.18676    3.02605  704.47787  -5.349 1.20e-07 ***
## sex           -1.21419    2.09483 1022.42110  -0.580    0.562
## yearstea       0.01129    0.14141  226.80861   0.080    0.936
## mathknow       1.35004    1.39168  234.49768   0.970    0.333
## mathprep      -0.27705    1.37583  205.27111  -0.201    0.841
## housepov     -17.64850   13.21755  113.87814  -1.335    0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.190  0.020 -0.011
## yearstea -0.259 -0.028  0.024  0.016
## mathknow -0.083 -0.007  0.115  0.007  0.029
## mathprep -0.631  0.053  0.001 -0.006 -0.172  0.004
## housepov -0.451  0.082 -0.178 -0.007  0.071  0.058  0.038
```

```
wald.test(b = fixef(fit4), Sigma = summary(fit4)$vcov, Terms = 2:4)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 85.1, df = 3, P(> X2) = 0.0
```

a. Report if justified statistically as a block of predictors [must use WALD test, not LRT]
Answer: The wald test gives a p-value less than 0.05, which justifies the significance of adding a block of individual predictors.

b. Report change in variance components for all levels
Answer: The change in σ^2_η is 93.8853485-94.3625825 = -0.477234, increases; the change in σ^2_ζ is 169.4480999-223.3059856 = -53.8578857, decreases; and change in σ^2_ϵ is 1064.9564422-1136.4309806 = -71.4745383, decreases.

c. Give a potential reason as to why the school level variance component drops from prior model
The aggregate effect of student predictors, can be seen as the school-level means and student deviation from the school mean. The school means would account for school-level variance, as a result the school-level variance component drops.

d. **WRITE OUT THIS MODEL** using your chosen notation (include assumptions).

$MATH1ST_{ijk} = b_0 + b_1 SES_{ijk} + b_2 MINORITY_{ijk} + b_3 SEX_{ijk} + b_4 YEARSTEA_{jk} + b_5 MATHKNOW_{jk} + b_6 MATHPREP_{jk} + b_7 HOUSEPOV_k + \zeta_{0k} + \eta_{0jk} + \varepsilon_{ijk}$, with $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, $\eta_{0jk} \sim N(0, \sigma^2_{\eta_0})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, independently of one another, j represents classrooms and k represents *schools*.

5.a. Try to add a random slope for each teacher level predictor (varying at the school level; one by one separately- not all together)

```
fit5.1 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                  housepov + (1 | schoolid/classid) + (0 + yearstea | schoolid),
                dat)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00805459 (tol = 0.002, component 1)
```

```
summary(fit5.1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + yearstea | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8482 -0.6147 -0.0322  0.5979  3.6603
##
## Random effects:
##  Groups           Name        Variance  Std.Dev.
##  classid.schoolid (Intercept) 9.247e+01  9.6159
##  schoolid         (Intercept) 1.684e+02 12.9758
##  schoolid.1       yearstea    1.008e-02  0.1004
##  Residual                     1.065e+03 32.6361
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.59885    5.30780  266.47954 101.662  < 2e-16 ***
## ses           10.04528    1.54492 1066.09816   6.502 1.21e-10 ***
## minority     -16.16715    3.02635  702.61831  -5.342 1.24e-07 ***
## sex           -1.21060    2.09480 1022.21558  -0.578    0.563
## yearstea       0.01128    0.14192  122.87743   0.079    0.937
## mathknow       1.33106    1.39155  234.33195   0.957    0.340
## mathprep      -0.26584    1.37588  204.90504  -0.193    0.847
## housepov     -17.72082   13.21686  113.58577  -1.341    0.183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.191  0.020 -0.010
## yearstea -0.258 -0.027  0.023  0.015
## mathknow -0.082 -0.007  0.115  0.006  0.028
## mathprep -0.632  0.053  0.001 -0.006 -0.172  0.003
```

```
## housepov -0.450  0.082 -0.179 -0.007  0.070  0.057  0.037
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00805459 (tol = 0.002, component 1)
```

```
fit5.2 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                 housepov + (1 | schoolid/classid) + (0 + mathknow | schoolid),
               dat)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(fit5.2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + mathknow | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8580 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups           Name        Variance  Std.Dev.
##  classid.schoolid (Intercept) 9.389e+01  9.689654
##  schoolid         (Intercept) 1.694e+02 13.017245
##  schoolid.1       mathknow    2.323e-07  0.000482
##  Residual                     1.065e+03 32.633630
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.63042    5.31210  275.38873 101.585  < 2e-16 ***
## ses           10.05075    1.54484 1066.56223   6.506 1.18e-10 ***
## minority     -16.18678    3.02605  704.47917  -5.349 1.20e-07 ***
## sex           -1.21419    2.09483 1022.42143  -0.580    0.562
## yearstea       0.01129    0.14141  226.80898   0.080    0.936
## mathknow       1.35004    1.39169  234.49763   0.970    0.333
## mathprep      -0.27705    1.37583  205.27161  -0.201    0.841
## housepov     -17.64848   13.21759  113.87742  -1.335    0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.190  0.020 -0.011
## yearstea -0.259 -0.028  0.024  0.016
## mathknow -0.083 -0.007  0.115  0.007  0.029
## mathprep -0.631  0.053  0.001 -0.006 -0.172  0.004
```

```
## housepov -0.451  0.082 -0.178 -0.007  0.071  0.058  0.038
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
fit5.3 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                 housepov + (1 | schoolid/classid) + (0 + mathprep | schoolid),
               dat)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(fit5.3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + mathprep | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8581 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups           Name        Variance  Std.Dev.
##  classid.schoolid (Intercept) 9.388e+01 9.689e+00
##  schoolid         (Intercept) 1.694e+02 1.302e+01
##  schoolid.1       mathprep    2.171e-07 4.659e-04
##  Residual                     1.065e+03 3.263e+01
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.63039    5.31207  275.39223 101.586  < 2e-16 ***
## ses           10.05076    1.54485 1066.56201   6.506 1.18e-10 ***
## minority     -16.18676    3.02605  704.47629  -5.349 1.20e-07 ***
## sex           -1.21419    2.09483 1022.42070  -0.580    0.562
## yearstea       0.01129    0.14141  226.80838   0.080    0.936
## mathknow       1.35003    1.39167  234.49786   0.970    0.333
## mathprep      -0.27705    1.37582  205.27063  -0.201    0.841
## housepov     -17.64851   13.21749  113.87941  -1.335    0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.190  0.020 -0.011
## yearstea -0.259 -0.028  0.024  0.016
## mathknow -0.083 -0.007  0.115  0.007  0.029
## mathprep -0.631  0.053  0.001 -0.006 -0.172  0.004
```

```
## housepov -0.451  0.082 -0.178 -0.007  0.071  0.058  0.038
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

b. Report the model fit or lack of fit **Answer:** The model with random slope on `mathknow` and the model with random slope on `mathprep` have convergent problem. Besides, all these three random slopes have variation that is close to 0, which indicates that these models are poorly fitted.

c. Retry the above, allowing the slopes to be correlated with the random intercepts (still one by one)

```
fit5.c.1 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                    housepov + (yearstea | schoolid) + (1 | schoolid:classid),
                  dat)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00352934 (tol = 0.002, component 1)
```

```
summary(fit5.c.1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (yearstea | schoolid) + (1 | schoolid:classid)
##    Data: dat
##
## REML criterion at convergence: 10723.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7461 -0.6037 -0.0291  0.6041  3.8451
##
## Random effects:
##  Groups           Name        Variance  Std.Dev. Corr
##  schoolid:classid (Intercept)   37.8479  6.1521
##  schoolid         (Intercept)  366.2230 19.1370
##                   yearstea       0.5527  0.7434  -0.78
##  Residual                     1066.4855 32.6571
## Number of obs: 1081, groups:  schoolid:classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  538.95171    5.48812  222.68165  98.203  < 2e-16 ***
## ses           10.15050    1.53873 1062.66116   6.597 6.62e-11 ***
## minority     -16.44545    2.99653  669.47204  -5.488 5.77e-08 ***
## sex           -1.33563    2.08775 1024.45847  -0.640    0.522
## yearstea       0.02205    0.15767   75.75723   0.140    0.889
## mathknow       1.04618    1.34371  209.64590   0.779    0.437
## mathprep       0.05077    1.34539  190.74479   0.038    0.970
## housepov     -17.14026   13.45947  119.64252  -1.273    0.205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.119
## minority -0.305  0.168
## sex      -0.184  0.022 -0.012
## yearstea -0.370 -0.019  0.032  0.009
## mathknow -0.085 -0.001  0.122  0.008  0.012
## mathprep -0.606  0.049 -0.007 -0.004 -0.139  0.014
## housepov -0.455  0.079 -0.169 -0.004  0.084  0.049  0.050
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00352934 (tol = 0.002, component 1)
```

```
fit5.c.2 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                    housepov + (mathknow| schoolid) + (1 | schoolid:classid),
                  dat)
summary(fit5.c.2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (mathknow | schoolid) + (1 | schoolid:classid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8581 -0.6131 -0.0324  0.5969  3.6603
##
## Random effects:
##  Groups           Name        Variance  Std.Dev. Corr
##  schoolid:classid (Intercept) 9.393e+01  9.6915
##  schoolid         (Intercept) 1.693e+02 13.0118
##                   mathknow    9.182e-04  0.0303  0.97
##  Residual                     1.065e+03 32.6341
## Number of obs: 1081, groups:  schoolid:classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.64041    5.31203  275.38950 101.588  < 2e-16 ***
## ses           10.04788    1.54488 1062.12269   6.504 1.20e-10 ***
## minority     -16.19378    3.02608  703.80365  -5.351 1.18e-07 ***
## sex           -1.21328    2.09485 1021.79810  -0.579    0.563
## yearstea       0.01114    0.14141  226.85275   0.079    0.937
## mathknow       1.35458    1.39201  214.62575   0.973    0.332
## mathprep      -0.27754    1.37599  201.27759  -0.202    0.840
## housepov     -17.64141   13.21242  103.98208  -1.335    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.190  0.020 -0.011
```

```
## yearstea -0.259 -0.028  0.024  0.016
## mathknow -0.082 -0.007  0.115  0.007  0.029
## mathprep -0.631  0.053  0.001 -0.006 -0.173  0.004
## housepov -0.451  0.082 -0.178 -0.007  0.071  0.057  0.038
```

```
fit5.c.3 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                    housepov + (mathprep | schoolid) + (1 | schoolid:classid),
                  dat)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(fit5.c.3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (mathprep | schoolid) + (1 | schoolid:classid)
##    Data: dat
##
## REML criterion at convergence: 10724.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8542 -0.6034 -0.0221  0.5914  3.6475
##
## Random effects:
##  Groups           Name        Variance Std.Dev. Corr
##  schoolid:classid (Intercept)   78.46   8.858
##  schoolid         (Intercept)  552.78  23.511
##                   mathprep      15.89   3.986   -1.00
##  Residual                      1064.26  32.623
## Number of obs: 1081, groups:  schoolid:classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  538.60853    5.60817  159.88504  96.040  < 2e-16 ***
## ses           10.14166    1.53961 1060.93429   6.587 7.04e-11 ***
## minority     -16.46420    2.99525  663.67458  -5.497 5.52e-08 ***
## sex           -1.16760    2.08697 1023.15165  -0.559    0.576
## yearstea      -0.02587    0.13949  223.50105  -0.185    0.853
## mathknow       1.29890    1.37194  229.68059   0.947    0.345
## mathprep       0.04076    1.34846  139.04922   0.030    0.976
## housepov     -14.01322   12.88712  116.05270  -1.087    0.279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.275  0.161
## sex      -0.183  0.024 -0.013
## yearstea -0.260 -0.033  0.025  0.023
## mathknow -0.071 -0.001  0.107  0.002  0.049
```

```
## mathprep -0.692  0.061 -0.035 -0.008 -0.155  0.012
## housepov -0.461  0.095 -0.187  0.003  0.089  0.027  0.107
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

Table 1: variation explained by classroom-level random intercept

	five_b	five_c
yearstea	92.466	37.848
mathknow	93.889	93.925
mathprep	93.882	78.462

Table 2: variation explained by school-level random intercept

	five_b	five_c
yearstea	168.372	366.223
mathknow	169.449	169.306
mathprep	169.446	552.775

Table 3: variation explained by school-level random slope

	five_b	five_c
yearstea	0.01	0.553
mathknow	0.00	0.001
mathprep	0.00	15.886

d. Report anything unusual about the variance components (changes that are in a direction you didn't expect) and any potential explanation for why those changes occurred (hint: what did you add to the model?).

Answer: For `mathknow`, the variation of school-level random slope increase while variation of school-level random intercept decrease. For `yearstea` and `mathprep`, both school-level random slope and school-level random intercept increase variation. Potential reason is that random slope on `mathknow` and random intercept are positively correlated, to explain same amount of school-level variation, the decrease in variation of school-level random intercept would be compensated by the positive covariance. Similarly for `yearstea` and `mathprep` the increase in variance of random slope and random intercept would be compensated by the negative covariance. 6. Question: a. Why is it a bad idea to include a classroom-level variable with random slopes at the classroom level?

Answer: Classroom-level variables does not vary within classroom, if there is no variation on variable, the slope could not be measured, so adding a random slope on classroom variable at classroom level makes no sense.

A2, Jeremy Lu

7. Question:

a. For UMM, write down: V_S, V_C, V_E for the three variance components (simply the estimates)
Answer: We have that V_S = 280.68, V_C = 85.46, and V_E = 1146.8

b. For the most complicated (all fixed effects) random INTERCEPTS ONLY model, what are: V_C, V_S, V_E?
Answer: We have in this model that V_S = 169.45, V_C = 93.89, V_E = 1064.96

c. By what fraction did these each decrease with the new predictors in the model?
Answer: The fraction decrease for V_S, and V_E are 0.396, and 0.071, respectively. But for V_C it actually increased 0.099 fraction-wise.

8. a.

```
fit8.a.1 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | schoolid/classid) + (0 + ses | schoolid),
                dat)
summary(fit8.a.1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + ses | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10724.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6138 -0.6185 -0.0290  0.5798  3.7130
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid.schoolid (Intercept)   88.56   9.411
##  schoolid         (Intercept)  167.98  12.961
##  schoolid.1       ses           72.50   8.515
##  Residual                      1035.12  32.173
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.13754    5.27918  270.54292 102.125  < 2e-16 ***
## ses            9.78982    1.82217   79.01642   5.373 7.62e-07 ***
## minority     -16.52526    3.02189  700.06722  -5.469 6.32e-08 ***
## sex           -1.40185    2.08170 1011.28952  -0.673    0.501
## yearstea       0.03079    0.14052  223.94368   0.219    0.827
## mathknow       1.35576    1.38459  232.20020   0.979    0.329
## mathprep      -0.19801    1.35994  198.59489  -0.146    0.884
## housepov     -16.94561   13.21117  112.82498  -1.283    0.202
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.091
## minority -0.323  0.124
## sex      -0.190  0.017 -0.010
## yearstea -0.260 -0.019  0.024  0.018
## mathknow -0.079  0.006  0.110  0.006  0.028
## mathprep -0.628  0.042  0.001 -0.007 -0.172  0.002
## housepov -0.451  0.076 -0.180 -0.007  0.070  0.056  0.041
```

```
fit8.a.2 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | schoolid/classid) + (0 + sex | schoolid),
                dat)
summary(fit8.a.2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + sex | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10728.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8578 -0.6110 -0.0259  0.5922  3.5557
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid.schoolid (Intercept)   96.08   9.802
##  schoolid         (Intercept)  161.63  12.713
##  schoolid.1       sex           35.84   5.986
##  Residual                      1054.36  32.471
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##              Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.43517    5.30740  272.54993 101.638  < 2e-16 ***
## ses            9.98477    1.54243 1058.27916   6.473 1.46e-10 ***
## minority     -16.16537    3.02861  704.25756  -5.338 1.27e-07 ***
## sex           -1.33535    2.18747  138.09087  -0.610    0.543
## yearstea       0.01448    0.14163  226.44539   0.102    0.919
## mathknow       1.40067    1.39464  234.45909   1.004    0.316
## mathprep      -0.27193    1.38011  205.78530  -0.197    0.844
## housepov     -16.77652   13.22879  112.39634  -1.268    0.207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.120
## minority -0.320  0.161
```

```
## sex      -0.179  0.020 -0.015
## yearstea -0.259 -0.029  0.024  0.013
## mathknow -0.081 -0.007  0.114  0.007  0.028
## mathprep -0.633  0.052  0.001 -0.004 -0.172  0.004
## housepov -0.449  0.081 -0.178 -0.010  0.070  0.055  0.036
```

```
fit8.a.3 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | schoolid/classid) + (0 + minority | schoolid),
                dat)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(fit8.a.3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid/classid) + (0 + minority | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8580 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid.schoolid (Intercept)   93.89   9.69
##  schoolid         (Intercept)  169.45  13.02
##  schoolid.1       minority       0.00   0.00
##  Residual                     1064.96  32.63
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.63041    5.31209  275.39107 101.585  < 2e-16 ***
## ses           10.05075    1.54484 1066.56217   6.506 1.18e-10 ***
## minority     -16.18677    3.02605  704.47765  -5.349 1.20e-07 ***
## sex           -1.21419    2.09483 1022.42106  -0.580    0.562
## yearstea       0.01129    0.14141  226.80889   0.080    0.936
## mathknow       1.35003    1.39168  234.49798   0.970    0.333
## mathprep      -0.27705    1.37583  205.27126  -0.201    0.841
## housepov     -17.64847   13.21752  113.87889  -1.335    0.184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.320  0.162
## sex      -0.190  0.020 -0.011
## yearstea -0.259 -0.028  0.024  0.016
```

```
## mathknow -0.083 -0.007  0.115  0.007  0.029
## mathprep -0.631  0.053  0.001 -0.006 -0.172  0.004
## housepov -0.451  0.082 -0.178 -0.007  0.071  0.058  0.038
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

b. Retry part (a), allowing the slopes to be correlated with the random intercepts.

```
fit8.b.1 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | classid) + (ses | schoolid), dat)
summary(fit8.b.1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | classid) + (ses | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10724.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5646 -0.6166 -0.0264  0.5888  3.7073
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  classid  (Intercept)   86.57   9.305
##  schoolid (Intercept)  171.18  13.083
##           ses           73.36   8.565   0.19
##  Residual             1035.90  32.185
## Number of obs: 1081, groups:  classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  538.72222    5.27648  271.13305 102.099  < 2e-16 ***
## ses            9.72646    1.82985   78.36212   5.315 9.75e-07 ***
## minority     -16.26698    3.03580  668.91588  -5.358 1.16e-07 ***
## sex           -1.40436    2.08074 1011.40322  -0.675    0.500
## yearstea       0.03617    0.14002  220.42240   0.258    0.796
## mathknow       1.26025    1.38201  230.89913   0.912    0.363
## mathprep      -0.21697    1.35642  197.10758  -0.160    0.873
## housepov     -15.89873   13.15396  111.71336  -1.209    0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.062
## minority -0.325  0.117
## sex      -0.188  0.018 -0.011
## yearstea -0.259 -0.021  0.021  0.017
## mathknow -0.077  0.007  0.108  0.005  0.028
## mathprep -0.627  0.045  0.002 -0.008 -0.172  0.001
## housepov -0.449  0.070 -0.182 -0.009  0.073  0.057  0.039
```

```
fit8.b.2 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | schoolid:classid) + (sex | schoolid), dat)
summary(fit8.b.2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | schoolid:classid) + (sex | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10727.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8048 -0.6095 -0.0222  0.5969  3.5525
##
## Random effects:
##  Groups           Name        Variance Std.Dev. Corr
##  schoolid:classid (Intercept)   97.33   9.866
##  schoolid         (Intercept)  206.34  14.364
##                   sex           84.08   9.169   -0.43
##  Residual                      1041.76  32.276
## Number of obs: 1081, groups:  schoolid:classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  5.399e+02  5.363e+00  2.626e+02 100.661  < 2e-16 ***
## ses          9.928e+00  1.540e+00  1.055e+03   6.448 1.72e-10 ***
## minority    -1.642e+01  3.027e+00  7.076e+02  -5.425 7.96e-08 ***
## sex         -1.340e+00  2.301e+00  8.742e+01  -0.582    0.562
## yearstea     6.877e-03  1.418e-01  2.277e+02   0.048    0.961
## mathknow     1.379e+00  1.396e+00  2.364e+02   0.988    0.324
## mathprep    -2.795e-01  1.378e+00  2.061e+02  -0.203    0.839
## housepov    -1.742e+01  1.326e+01  1.136e+02  -1.314    0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.121
## minority -0.319  0.163
## sex      -0.222  0.018 -0.011
## yearstea -0.258 -0.028  0.024  0.014
## mathknow -0.082 -0.006  0.114  0.006  0.027
## mathprep -0.627  0.053  0.004 -0.005 -0.172  0.004
## housepov -0.449  0.083 -0.178 -0.003  0.072  0.060  0.038
```

```
fit8.b.3 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
                housepov + (1 | classid) + (minority | schoolid), dat)
summary(fit8.b.3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
## Formula: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
##     housepov + (1 | classid) + (minority | schoolid)
##    Data: dat
##
## REML criterion at convergence: 10717.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8952 -0.6358 -0.0345  0.6129  3.6444
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  classid  (Intercept)   86.69   9.311
##  schoolid (Intercept)  381.20  19.524
##           minority     343.13  18.524   -0.83
##  Residual             1039.39  32.240
## Number of obs: 1081, groups:  classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  5.395e+02  5.655e+00  1.731e+02  95.399  < 2e-16 ***
## ses          9.431e+00  1.543e+00  1.063e+03   6.111 1.39e-09 ***
## minority    -1.638e+01  3.896e+00  5.824e+01  -4.203 9.17e-05 ***
## sex         -8.628e-01  2.084e+00  1.022e+03  -0.414    0.679
## yearstea    -4.368e-03  1.376e-01  2.172e+02  -0.032    0.975
## mathknow     1.632e+00  1.359e+00  2.248e+02   1.201    0.231
## mathprep    -2.918e-01  1.335e+00  1.981e+02  -0.218    0.827
## housepov    -1.606e+01  1.257e+01  9.999e+01  -1.277    0.204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) ses    minrty sex    yearst mthknw mthprp
## ses      -0.105
## minority -0.494  0.113
## sex      -0.172  0.024 -0.014
## yearstea -0.253 -0.021  0.027  0.014
## mathknow -0.078 -0.005  0.099  0.010  0.024
## mathprep -0.576  0.052 -0.002 -0.005 -0.167 -0.002
## housepov -0.394  0.089 -0.157 -0.013  0.091  0.061  0.037
```

c. Report anything unusual about the variance components (changes that are unexpected)

Answer: Adding the correlation between school-level random slope on any of these student-level predictors, and the school-level random intercept, both the variations captured by the school-level random slope and the variation of random intercept increase substantially, especially for adding correlation between random slope on `minority` and random intercept.

9. a. Take the two predictors that had significant (at .05 level) random slopes, in the forms in which they worked (indep. or correlated) and add both to the model, and test for need of one conditional on already including the other.

```r
# check significance of random slope
anova(fit8.a.1,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
## fit8.a.1: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.a.1   12 10749 10809 -5362.4    10725 4.6972  1    0.03021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit8.b.1,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
## fit8.b.1: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | classid)
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.b.1   13 10750 10815 -5362.2    10724 5.1385  2    0.07659 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit8.a.2,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
## fit8.a.2: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.a.2   12 10753 10813 -5364.4    10729 0.6137  1     0.4334
```

```r
anova(fit8.b.2,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
## fit8.b.2: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.b.2   13 10754 10818 -5363.8    10728 1.8631  2      0.394
```

```r
anova(fit8.a.3,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
```

```
## fit8.a.3: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/
##          npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.a.3   12 10754 10813 -5364.8    10730     0  1          1
```

```
anova(fit8.b.3,fit4,refit=F)
```

```
## Data: dat
## Models:
## fit4: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/clas
## fit8.b.3: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | classid)
##          npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## fit4       11 10752 10806 -5364.8    10730
## fit8.b.3   13 10744 10808 -5358.8    10718 11.967  2    0.00252 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# random slope of ses without correlation and random slope of minority with
# correlation are significant.
fit9 <- lmer( math1st ~ ses + minority + sex + yearstea + mathknow + mathprep +
               housepov + (1 | classid:schoolid) + (0 + ses | schoolid) +
               (minority | schoolid), dat)
anova(fit8.a.1, fit9,refit=F) #P = 0.00204
```

```
## Data: dat
## Models:
## fit8.a.1: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | schoolid/
## fit9: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | classid:schoo
##          npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## fit8.a.1   12 10749 10809 -5362.4    10725
## fit9       14 10740 10810 -5356.2    10712 12.39  2    0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit8.b.3, fit9,refit=F) #P = 0.02365
```

```
## Data: dat
## Models:
## fit8.b.3: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | classid)
## fit9: math1st ~ ses + minority + sex + yearstea + mathknow + mathprep + housepov + (1 | classid:schoo
##          npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## fit8.b.3   13 10744 10808 -5358.8    10718
## fit9       14 10740 10810 -5356.2    10712  5.12  1    0.02365 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b. Is the more complex model (with both random slopes in it) justified?
   Answer: Yes, both random slopes are significant according to the LRT.

c. WRITE OUT THIS MODEL in your preferred notation (include assumptions)
   $MATH1ST_{ijk} = b_0 + (b_1 + \zeta_{1k})SES_{ijk} + (b_2 + \zeta_{2k})MINORITY_{ijk} + b_3 SEX_{ijk} + b_4 YEARSTEA_{jk} + b_5 MATHKNOW_{jk} + b_6 MATHPREP_{jk} + b_7 HOUSEPOV_k + \zeta_{0k} + \eta_{0jk} + \varepsilon_{ijk}$, with $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, $\zeta_{1k} \sim N(0, \sigma^2_{\zeta_1})$, $\zeta_{3k} \sim N(0, \sigma^2_{\zeta_3})$ $\eta_{0jk} \sim N(0, \sigma^2_{\eta_0})$ and $\varepsilon_{ijk} \sim N(0, \sigma^2_{\varepsilon})$, $corr(\zeta_{0k}, \zeta_{1k}) = 0$, and $corr(\zeta_{0k}, \zeta_{2k}) = 0$, the other random components are independent of each others.

10. Now consider the model with a random slope *only* in minority. We will make predictions at levels of minority in the range 0 to 1 for illustrative purposes.

    a. What are: V_C, V_S(minority=0), V_E? i. We need to list 'minority=0' here, or we don't know how to use the slope variance. **Answer:** V_C = 86.69, V_S = 381.20, V_E = 1039.39

```
data.frame(VarCorr(fit8.b.3))
```

```
##        grp        var1     var2       vcov      sdcor
## 1  classid (Intercept)     <NA>   86.69412  9.3109677
## 2 schoolid (Intercept)     <NA>  381.20088 19.5243664
## 3 schoolid    minority     <NA>  343.12842 18.5237258
## 4 schoolid (Intercept) minority -299.26986 -0.8274803
## 5 Residual        <NA>     <NA> 1039.38897 32.2395560
```

    b. What are: V_S(minority=0.25), V_S(minority=+0.50), V_S(minority=+0.75)?

```
# minority = 0.25
paste0("V_S(minority=0.25) =", data.frame(VarCorr(fit8.b.3))[2,4]+0.25^2 *
         data.frame(VarCorr(fit8.b.3))[3,4] + 0.25*2*
         data.frame(VarCorr(fit8.b.3))[4,5]*data.frame(VarCorr(fit8.b.3))[3,5]*
         data.frame(VarCorr(fit8.b.3))[2,5])
```

```
## [1] "V_S(minority=0.25) =253.011478742454"
```

```
# minority = 0.5
paste0("V_S(minority=0.5) =", data.frame(VarCorr(fit8.b.3))[2,4]+0.5^2 *
         data.frame(VarCorr(fit8.b.3))[3,4] + 0.5*2*
         data.frame(VarCorr(fit8.b.3))[4,5]*data.frame(VarCorr(fit8.b.3))[3,5]*
         data.frame(VarCorr(fit8.b.3))[2,5])
```

```
## [1] "V_S(minority=0.5) =167.713127497944"
```

```
# minority = 0.75
paste0("V_S(minority=0.75) =", data.frame(VarCorr(fit8.b.3))[2,4]+0.75^2 *
         data.frame(VarCorr(fit8.b.3))[3,4] + 0.75*2*
         data.frame(VarCorr(fit8.b.3))[4,5]*data.frame(VarCorr(fit8.b.3))[3,5]*
         data.frame(VarCorr(fit8.b.3))[2,5])
```

```
## [1] "V_S(minority=0.75) =125.305828492755"
```

Answer:
V_S(minority=0.25) $= \sigma^2_{\zeta_{0k}} + 2 \times 0.25 \times \rho_{\zeta_{0k},\zeta_{2k}} + 0.25^2\sigma^2_{\zeta_{2k}} = 253.0114787$,
V_S(minority=0.50) $= \sigma^2_{\zeta_{0k}} + 2 \times 0.5 \times \rho_{\zeta_{0k},\zeta_{2k}} + 0.5^2\sigma^2_{\zeta_{2k}} = 167.7131275$,
V_S(minority=0.75) $= \sigma^2_{\zeta_{0k}} + 2 \times 0.75 \times \rho_{\zeta_{0k},\zeta_{2k}} + 0.75^2\sigma^2_{\zeta_{2k}} = 125.3058285$
c. Is the variance between schools monotonically *increasing* in the value of minority?
Answer: No, it seems to be decreasing from minority 0 to 0.75 given the variance calculated.

Project B, Xinming Dai

Question 0: read data and process missingness

```
dat <- read.csv("classroom.csv")

# construct new outcome math1st
dat <-
  dat %>%
  mutate(math1st = mathkind + mathgain)

# remove missing data
dat <-
  dat %>%
  filter(complete.cases(dat))
```

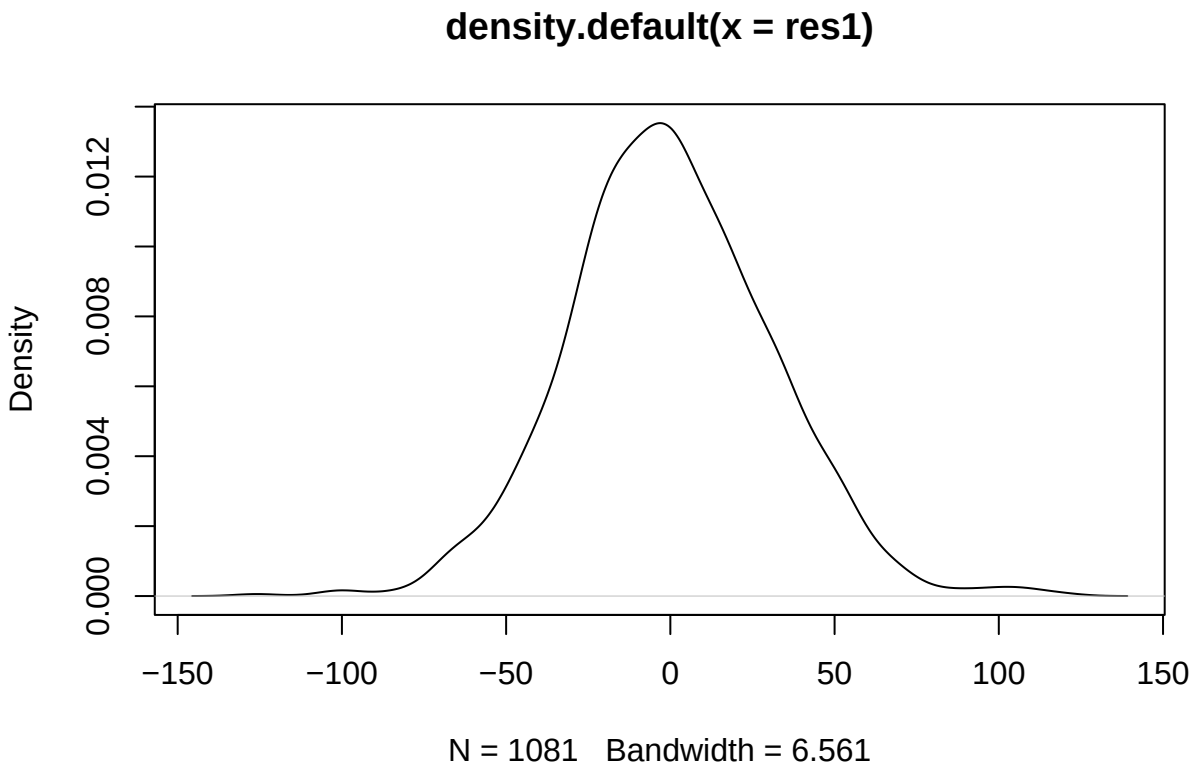
Question 1

```
# fit a model
fit1 <- lmerTest::lmer(math1st ~ housepov + yearstea + mathprep + mathknow +
                         ses + sex + minority + (1|schoolid/classid), data = dat)
summary(fit1)
```

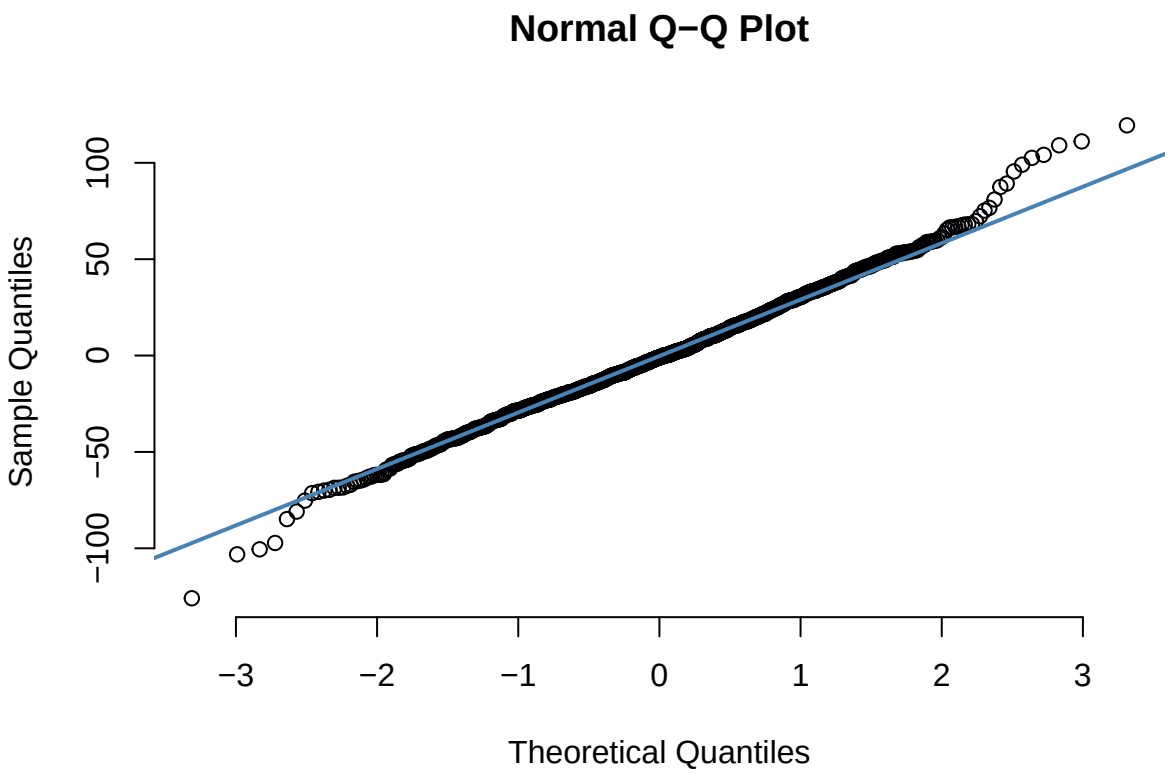
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +
##     minority + (1 | schoolid/classid)
##    Data: dat
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8581 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  classid:schoolid (Intercept)   93.89   9.689
##  schoolid         (Intercept)  169.45  13.017
##  Residual                     1064.96  32.634
## Number of obs: 1081, groups:  classid:schoolid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  539.63041    5.31209  275.39010 101.585  < 2e-16 ***
## housepov     -17.64850   13.21755  113.87814  -1.335    0.184
## yearstea       0.01129    0.14141  226.80861   0.080    0.936
## mathprep      -0.27705    1.37583  205.27111  -0.201    0.841
## mathknow       1.35004    1.39168  234.49768   0.970    0.333
## ses           10.05076    1.54485 1066.56211   6.506 1.18e-10 ***
```

```
## sex           -1.21419    2.09483 1022.42110  -0.580    0.562
## minority     -16.18676    3.02605  704.47787  -5.349 1.20e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) houspv yearst mthprp mthknw ses    sex
## housepov -0.451
## yearstea -0.259  0.071
## mathprep -0.631  0.038 -0.172
## mathknow -0.083  0.058  0.029  0.004
## ses      -0.121  0.082 -0.028  0.053 -0.007
## sex      -0.190 -0.007  0.016 -0.006  0.007  0.020
## minority -0.320 -0.178  0.024  0.001  0.115  0.162 -0.011
```

```r
# plot residuals to test normality assumption
res1 <- residuals(fit1)
# density plot
plot(density(res1))
```



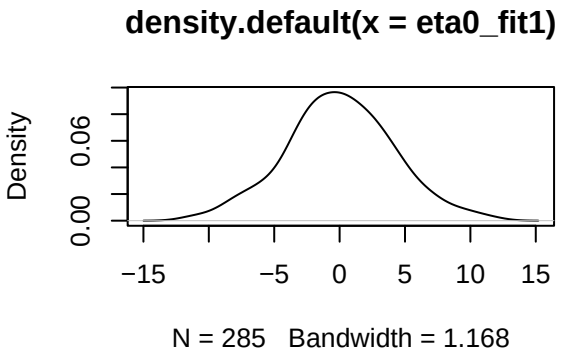

```r
# QQ plot
qqnorm(res1, pch = 1, frame = FALSE)
qqline(res1, col = "steelblue", lwd = 2)
```

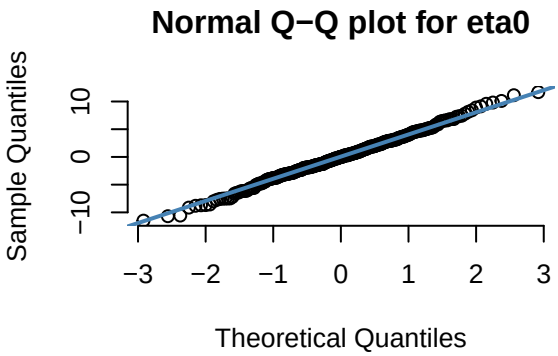


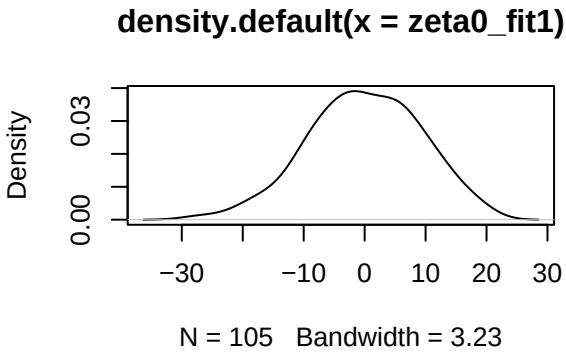
QQ plot shows that points are around the line, and thus we believe the normality assumption holds.

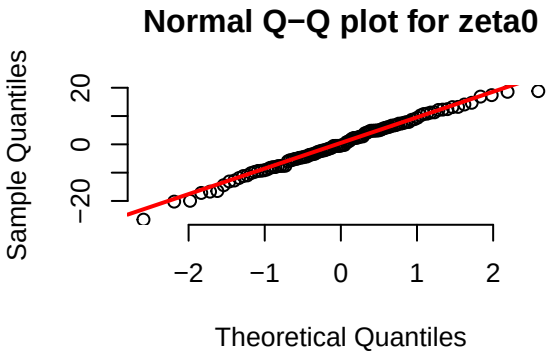
Question 2

```
# Generate the two sets of BLUPs (for random effects zeta0 and eta0)
blups_fit1 <- ranef(fit1)

par(mfrow=c(2,2))
# examine normality for eta0 (class-level)
eta0_fit1 <- blups_fit1$`classid:schoolid`$`(Intercept)`
# density plot
plot(density(eta0_fit1))
# QQ plot
qqnorm(eta0_fit1, pch = 1, frame = FALSE, main = "Normal Q-Q plot for eta0")
qqline(eta0_fit1, col = "steelblue", lwd = 2)

# examine normality for zeta0 (school-level)
zeta0_fit1 <- blups_fit1$schoolid$`(Intercept)`
# density plot
plot(density(zeta0_fit1))
# QQ plot
qqnorm(zeta0_fit1, pch = 1, frame = FALSE, main = "Normal Q-Q plot for zeta0")
qqline(zeta0_fit1, col = "red", lwd = 2)
```










```
par(mfrow=c(1,1))
```

QQ plot shows that both sets of BLUPs of zeta0 and eta0 are around the line, and thus we believe the normality assumption holds.

Question 3

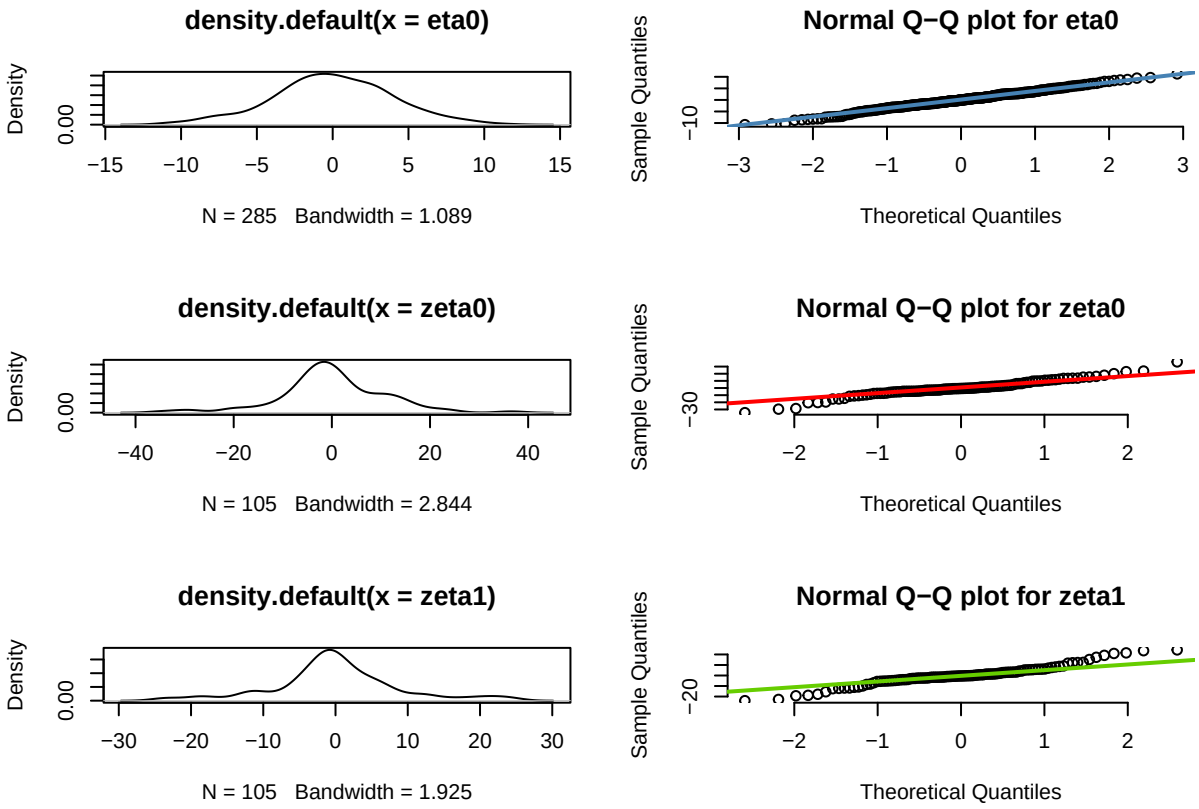
```
# a
# add a random slope for minority, correlated with the random intercept, at the school level
fit2 <- lmerTest::lmer(math1st ~ housepov + yearstea + mathprep +
                         mathknow + ses + sex + minority + (minority|schoolid) +
                         (1 | classid), data = dat)
print(summary(fit2))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +
##     minority + (minority | schoolid) + (1 | classid)
##    Data: dat
##
## REML criterion at convergence: 10717.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8952 -0.6358 -0.0345  0.6129  3.6444
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  classid  (Intercept)   86.69   9.311
##  schoolid (Intercept)  381.20  19.524
##           minority     343.13  18.524   -0.83
##  Residual             1039.39  32.240
## Number of obs: 1081, groups:  classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  5.395e+02  5.655e+00  1.731e+02  95.399  < 2e-16 ***
## housepov    -1.606e+01  1.257e+01  9.999e+01  -1.277    0.204
## yearstea    -4.368e-03  1.376e-01  2.172e+02  -0.032    0.975
## mathprep    -2.918e-01  1.335e+00  1.981e+02  -0.218    0.827
## mathknow     1.632e+00  1.359e+00  2.248e+02   1.201    0.231
## ses          9.431e+00  1.543e+00  1.063e+03   6.111 1.39e-09 ***
## sex         -8.628e-01  2.084e+00  1.022e+03  -0.414    0.679
## minority    -1.638e+01  3.896e+00  5.824e+01  -4.203 9.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) houspv yearst mthprp mthknw ses    sex
## housepov -0.394
## yearstea -0.253  0.091
## mathprep -0.576  0.037 -0.167
## mathknow -0.078  0.061  0.024 -0.002
## ses      -0.105  0.089 -0.021  0.052 -0.005
## sex      -0.172 -0.013  0.014 -0.005  0.010  0.024
## minority -0.494 -0.157  0.027 -0.002  0.099  0.113 -0.014
```

```r
# b
# residual
blups_fit2 <- ranef(fit2)
# BULPs
zeta0 <- blups_fit2$schoolid$`(Intercept)`
zeta1 <- blups_fit2$schoolid$minority
eta0 <- blups_fit2$classid$`(Intercept)`

# c
# check normality
par(mfrow=c(3,2))
# density plot
plot(density(eta0))
# QQ plot
# examine normality for eta0 (class-level)
qqnorm(eta0, pch = 1, frame = FALSE, main = "Normal Q-Q plot for eta0")
qqline(eta0, col = "steelblue", lwd = 2)

# examine normality for zeta0 (school-level)
# density plot
plot(density(zeta0))
# QQ plot
qqnorm(zeta0, pch = 1, frame = FALSE, main = "Normal Q-Q plot for zeta0")
```

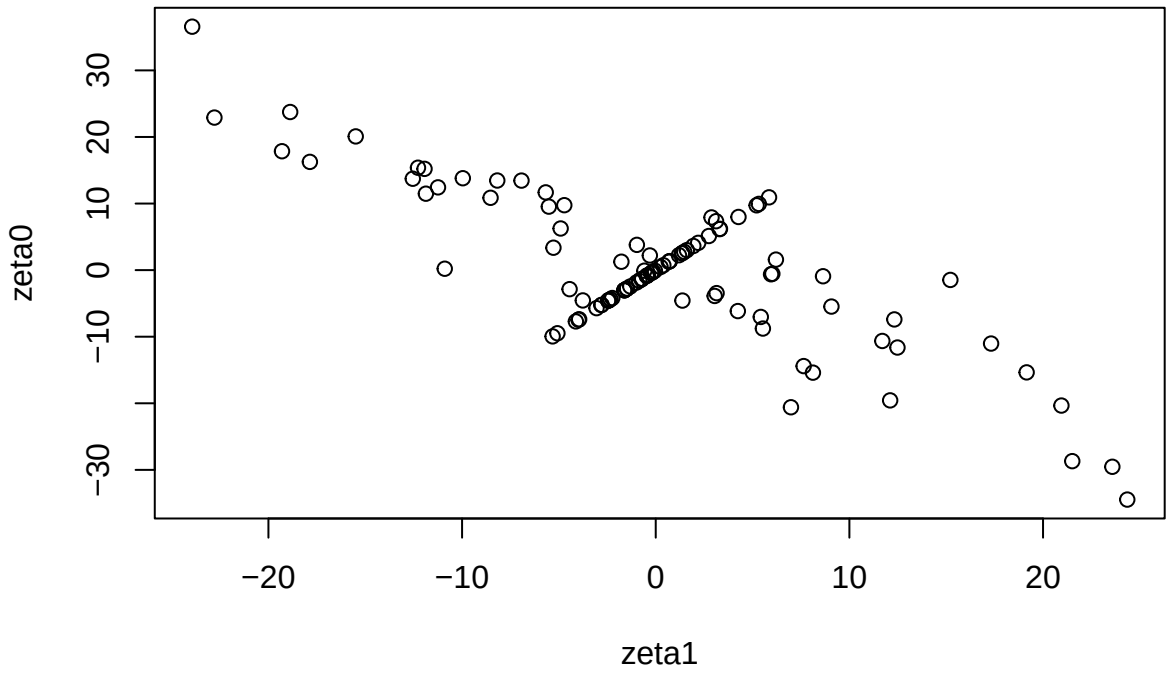
```
qqline(zeta0, col = "red", lwd = 2)

# examine normality for zeta1 (random slop)
# density plot
plot(density(zeta1))
# QQ plot
qqnorm(zeta1, pch = 1, frame = FALSE, main = "Normal Q-Q plot for zeta1")
qqline(zeta1, col = "chartreuse3", lwd = 2)
```




```
par(mfrow=c(1,1))
```

QQ plot shows that BLUPs of eta0 are around the line, and thus we believe the normality assumption holds. However, BLUPs of zeta0, and zeta1 deviate from the line too much, and therefore we don't think the normality assumption holds.

```
# d
plot(zeta1, zeta0)
```



Overall, zeta0 and zeta1 are negative correlated. However, some odd points are positive correlated.

```
# e
# points from first quadrant
fq <- (3-abs(blups_fit2$schoolid$minority)>0)&(zeta1>=0&zeta0>=0)
# points from third quadrant
sq <- (3-abs(blups_fit2$schoolid$minority)>0)&(zeta1<=0&zeta0<=0)

# these schools are
unique(dat$schoolid)[fq|sq]
```

```
##  [1]   1   4   5   9  10  12  14  16  17  20  22  24  26  28  33  38  40  42  43
## [20]  45  46  47  52  53  57  61  69  73  78  79  80  84  86  89  90  96  98 100
## [39] 102 103 106
```

```
odd_point <-
  dat %>%
  filter(schoolid %in% unique(dat$schoolid)[fq|sq])
```

Almost all students in these school are minority.

Question 4

a

$V_S = 169.45$, $V_C = 93.89$, and $V_E = 1064.96$.

```
# fit a model
fit3 <- lmerTest::lmer(math1st ~ housepov + yearstea + mathprep + mathknow + ses
                       + sex + minority + (ses|schoolid) + (1 | classid),
                       data = dat)
summary(fit3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +
##     minority + (ses | schoolid) + (1 | classid)
##    Data: dat
##
## REML criterion at convergence: 10724.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5646 -0.6166 -0.0264  0.5888  3.7073
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  classid  (Intercept)   86.57   9.305
##  schoolid (Intercept)  171.18  13.083
##           ses           73.36   8.565   0.19
##  Residual             1035.90  32.185
## Number of obs: 1081, groups:  classid, 285; schoolid, 105
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  538.72222    5.27648  271.13305 102.099  < 2e-16 ***
## housepov     -15.89873   13.15396  111.71336  -1.209    0.229
## yearstea       0.03617    0.14002  220.42240   0.258    0.796
## mathprep      -0.21697    1.35642  197.10758  -0.160    0.873
## mathknow       1.26025    1.38201  230.89913   0.912    0.363
## ses            9.72646    1.82985   78.36212   5.315 9.75e-07 ***
## sex           -1.40436    2.08074 1011.40322  -0.675    0.500
## minority     -16.26698    3.03580  668.91588  -5.358 1.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) houspv yearst mthprp mthknw ses    sex
## housepov -0.449
## yearstea -0.259  0.073
## mathprep -0.627  0.039 -0.172
## mathknow -0.077  0.057  0.028  0.001
## ses      -0.062  0.070 -0.021  0.045  0.007
## sex      -0.188 -0.009  0.017 -0.008  0.005  0.018
## minority -0.325 -0.182  0.021  0.002  0.108  0.117 -0.011
```

c

$V_C = 86.57$, $V_{S(ses=0)} = 171.18$, and $V_E = 1035.90$.

d

$V_{S(ses=-0.50)} = 171.18 + 2 * (-0.5) * 13.083 * 8.565 * 0.19 + (-0.5)^2 * 73.36 = 168.23$

$V_{S(ses=0.50)} = 171.18 + 2 * (0.5) * 13.083 * 8.565 * 0.19 + (0.5)^2 * 73.36 = 210.81$

e

There is heteroscedasticity at school level (3) because $V_{S(ses=0.50)}$ and $V_{S(ses=-0.50)}$ are not approximate and V_S are depend on ses.

Project C – Longitudinal Data, Yu Wang

1. Make a person-period file with math score (Kindergarten and First grade). That is, math0 <- mathkind; math1 <- mathkind+mathgain (you have to make this work in the dataframe). Using reshape in R, you have to be careful to specify the name of the math variable (math0 and math1) as varying.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v stringr 1.4.0
## v tidyr   1.2.0     v forcats 0.5.1
## v readr   2.1.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()
```

```
library(lme4)
library(lmerTest)
library(lattice)
require(foreign)
```

```
## Loading required package: foreign
```

```
dat<- read.csv('classroom.csv')
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")
```

2.We ignore classrooms in this analysis, but keep it in the notation

2a. Fit a model with math as outcome, and fixed effect for time trend (year), and random intercepts for schools.

ANS:

```
# a
fit1 <- lmer(math ~ year + (1 | schoolid), data = class_pp)
summary(fit1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (1 | schoolid)
##    Data: class_pp
##
## REML criterion at convergence: 23951.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.2833 -0.6084  0.0037  0.6329  3.7761
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  348.7   18.67
##  Residual             1268.4   35.62
## Number of obs: 2380, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  464.932      2.116  132.154  219.73   <2e-16 ***
## year          57.566      1.460 2270.855   39.43   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.345
```

2b. Write down the model (include assumptions).

ANS : $MATH_{tijk} = b_0 + \zeta_{0k} + b_1 TIME_{tijk} + \varepsilon_{tijk}$ and assume $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, and $\varepsilon_{tijk} \sim N(0, \sigma^2_{\varepsilon})$, independently.

2c. Add random intercepts for child

```
fit2 <- lmer(math ~ year + (1 |schoolid/childid), data = class_pp)
summary(fit2)
```

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

2d. Write down the model (include assumptions).

ANS : $MATH_{tijk} = b_0 + \delta_{0ijk} + \zeta_{0k} + b_1 TIME_{tijk} + \varepsilon_{tijk}$ and assume $\delta_{0ijk} \sim N(0, \sigma^2_{\delta_0})$, $\zeta_{0k} \sim N(0, \sigma^2_{\zeta_0})$, and $\varepsilon_{tijk} \sim N(0, \sigma^2_{\varepsilon})$, independently.

3. Report original and new variance estimates for $\sigma^2_{\zeta_0}$ (between schools) and σ^2_{ε}(within schools)

ANS: The old between school variance is 348.7 where the old within school variance is 1268.4 The new between school variance is 307.5 where the new within school variance is 599.1

a Compute a pseudo R2 relating the between school variation and ignoring between students in the same school. In other words, what fraction of the between-school variance in the first model is 'explained' by the addition of a student random effect?

ANS: R^2 = (348.7-307.5)/348.7 = 0.11

b. Does the total variation stay about the same (adding between children within schools variance as well, to the second model results)?

ANS: The total variance in the fit1 is 348.7 + 1268.4 = 1617.1 The total variance in the fit2 is 702 + 307.5 + 599.1 = 1608.6 Hence, the total variation seems stay about the same although their values are not exactly same.

4. Add a random slope (zeta1) for time trend (year) within schools (uncorrelated with random intercept (zeta0))

ANS:

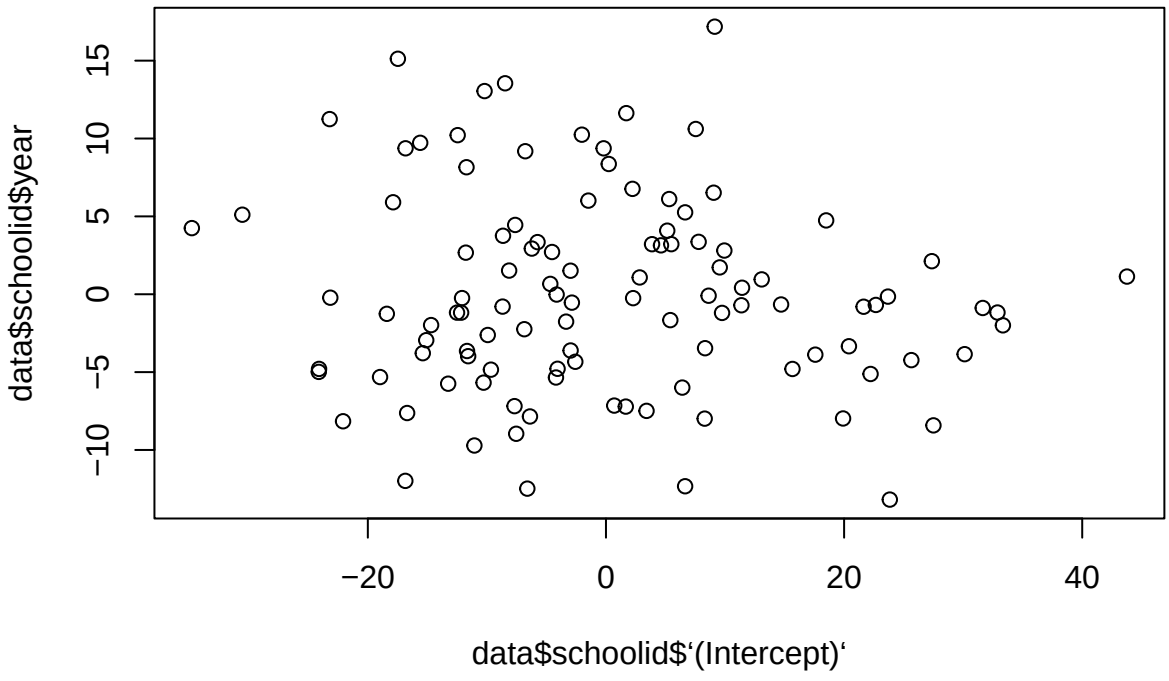
```
fit3 <- lmer(math ~ year + (year || schoolid) + (1 | schoolid:childid),
             data = class_pp)
summary(fit3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (year || schoolid) + (1 | schoolid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23529.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7665 -0.4721  0.0139  0.4686  3.6080
##
## Random effects:
##  Groups           Name        Variance Std.Dev.
##  schoolid.childid (Intercept) 725.12   26.928
##  schoolid         year         88.67    9.417
##  schoolid.1       (Intercept) 324.79   18.022
##  Residual                     552.21   23.499
## Number of obs: 2380, groups:  schoolid:childid, 1190; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.087      2.081 109.955  223.45   <2e-16 ***
## year          57.499      1.370  99.917   41.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.178
```

a Generate the BLUPs for the random effects and examine whether the independence between zeta0 and zeta1 is REFLECTED in a scatterplot of these two sets of effects.

ANS:

```r
# obtain the random effects
data <- ranef(fit3)

# plot the correlation
plot(data$schoolid$`(Intercept)`, data$schoolid$year)
```



Based on the above scatter plot, I would say zeta0 and zeta1 are independent because there does exist a patter which can explain the relationship between year(x) and intercept(y)

b. Compute V_S(year=0) and V_S(year=1). Since there are only two years, this is a form of heteroscedasticity in the random effects.

i. In which year is there more between school variation, net of all else (year=0 or year=1)?
ANS: V_S(year=0) = $\sigma^2_{\zeta_0}$ + 0 * $\sigma^2_{\zeta_0}$ = 324.79 V_S(year=1) = $\sigma^2_{\zeta_0}$ + 1 * $\sigma^2_{\zeta_0}$ = 324.79 + 88.67 = 413.46
Based on the hand calculation, I would say (year=1) has more between school variation

5.If you ran the model separately BY YEAR, and removed the year trend from the model, would you get the same estimates for the variance between schools? TRY IT

ANS

```
dat_temp = class_pp %>% filter(year==0)
fit_year0 = lmer(math ~ (1 | schoolid),data = dat_temp)
summary(fit_year0)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ (1 | schoolid)
##    Data: dat_temp
##
## REML criterion at convergence: 12085.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8223 -0.5749  0.0005  0.6454  3.6237
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  364.3   19.09
##  Residual             1344.5   36.67
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   465.23       2.19 103.20   212.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dat_temp2 = class_pp %>% filter(year==1)
fit_year1 = lmer(math ~ (1| schoolid), data = dat_temp2)
summary(fit_year1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ (1 | schoolid)
##    Data: dat_temp2
##
## REML criterion at convergence: 11950.8
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -5.291 -0.612 -0.005  0.613  3.793
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  306.8   17.52
##  Residual             1205.0   34.71
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  522.698      2.027 103.069   257.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the above summary table, we would say year1(364.3) has more between school variation than year0(306.8).

6. Rerun the last nested longitudinal model, allowing correlation between intercept and slope.

```
fit4 <- lmer(math ~ year + (year | schoolid) + (1| schoolid:childid),
             data = class_pp)
summary(fit4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
## Formula: math ~ year + (year | schoolid) + (1 | schoolid:childid)
##    Data: class_pp
##
## REML criterion at convergence: 23520.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7030 -0.4686  0.0066  0.4669  3.5142
##
## Random effects:
##  Groups           Name        Variance Std.Dev. Corr
##  schoolid:childid (Intercept) 728.0    26.98
##  schoolid         (Intercept) 370.6    19.25
##                   year        109.1    10.44    -0.45
##  Residual                     547.0    23.39
## Number of obs: 2380, groups:  schoolid:childid, 1190; schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  465.099      2.188 102.918  212.60   <2e-16 ***
## year          57.668      1.440  94.574   40.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.439
```

a. Is the correlation signif.?

```
anova(fit3, fit4, refit = F)
```

```
## Data: class_pp
## Models:
## fit3: math ~ year + (year || schoolid) + (1 | schoolid:childid)
## fit4: math ~ year + (year | schoolid) + (1 | schoolid:childid)
##      npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## fit3    6 23541 23576 -11764    23529
## fit4    7 23534 23575 -11760    23520 8.8241  1   0.002973 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANS: Since the p value is much less than 0.05, we can conclude that the correlation is significant

b.Compute V_S(year=0) and V_S(year=1) for this new model (your formula should include covariance terms).

ANS: covariance = -0.45 * 19.25 * 10.44 = -90. V_S(Year=0) = 370.6 + 2 * 0 * (-90) + 0^2 * 109.1 = 370.6. V_S(Year=1) = 370.6 + 2 * 1 * (-90) +1^2 * 109.1 = 299.7.

Is this result (and thus model) more consistent with the separate grade analysis? You are implicitly testing model fit here.

ANS: The calculated value for V_S(Year=0) is 370.6 where the actual between school variations 364.3. Also, the calculated value for V_S(Year=1) is 299.7 where the actual between school variance is about 306.6. As we can see, they are pretty similar, suggesting that the model is consistent with the separate grade analysis