

MLM Final Project Part 2ab

May 12 2020

Team Members and division of work:

Question 1

Refit the model in Part 1 that has all fixed effects as well as random intercepts (in schools and classrooms). Recall that `math1st = mathkind + mathgain` is the outcome. The model is `math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex + minority + (1|schoolid/classid)`, REML = T)

```
lm1 <- lmerTest::lmer(math1st ~ housepov + yearstea + mathprep + mathknow +  
                      ses + sex + minority + (1|schoolid/classid), REML = T, data = classroom)  
summary(lm1)
```

```
## 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: classroom  
##  
## 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  275.39009  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
```

- a. Construct the residual that removes only the 'fixed effects' then subtract it from the outcome; call this residual `resFE`
 - i. R hint 1: `predict` has an option to generate the prediction based on the fixed effects only.
 - ii. R hint 2: If you decide to add a column to your data frame with `resFE`, note that `predict` only generates predictions for cases uses in the model *after listwise deletion*.

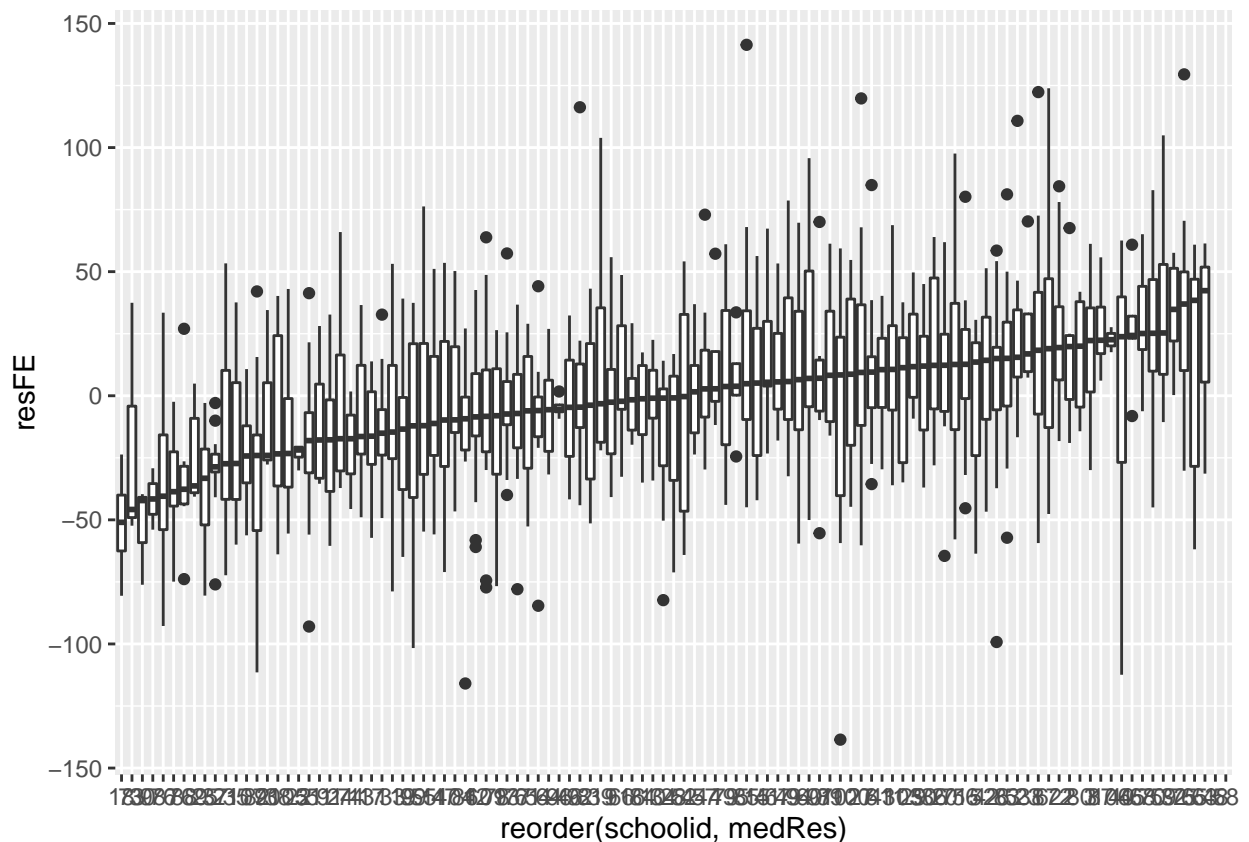
```
# Calculate predictions using fixed effects only:
predsFE <- predict(lm1, re.form = ~0)

# Calculate residual and add to dataframe:
resFE <- classroom[complete.cases(classroom), "math1st"] - predsFE
classroom[complete.cases(classroom), "resFE"] = resFE
```

Question 2

Show that the residual is not independent within schools in some manner.

```
# Insert code to show that the residual, resFE, is not independent within schools
classroom %>% group_by(schoolid) %>% mutate(medRes = median(resFE, na.rm = T)) %>% ggplot(., aes(x = reo
## Warning: Removed 109 rows containing non-finite values (stat_boxplot).
```



- The boxplots of residuals show evidence of a relationship of scores within schools. After excluding random effects due to schools, the variation between each school is no longer accounted for, and the

plot shows that some schools have residuals below the overall average, and some are above, indicative of heterogeneity.

Question 3

a. Construct the residual that utilizes the BLUPs for the random effects using the R command `residuals`.

i. Call the new residual `resFE_RE`

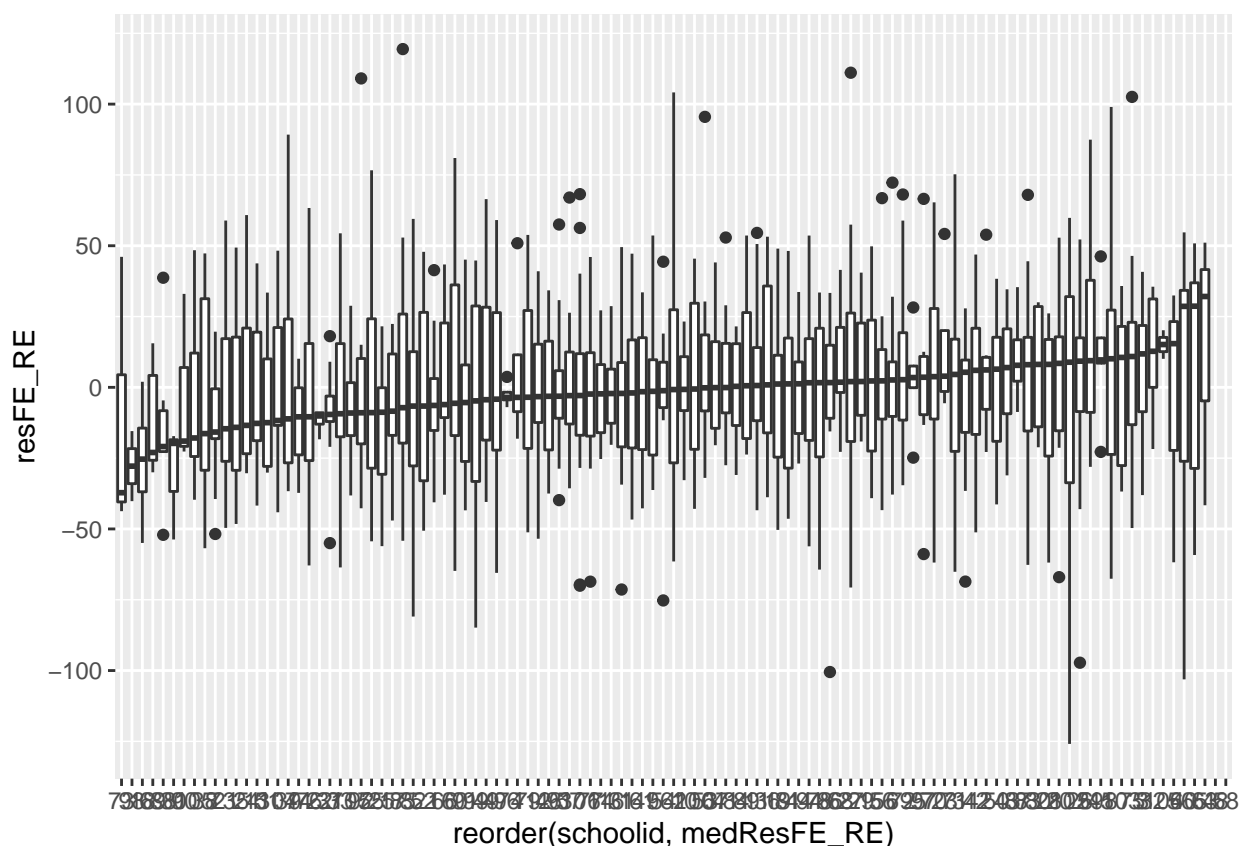
```
# Insert code to construct the residual
resFE_RE <- residuals(lm1)
classroom[complete.cases(classroom), "resFE_RE"] <- resFE_RE
```

Question 4

a. Show that these new residuals, `resFE_RE` are MUCH LESS (if not completely un-) correlated within schools, using the same method as before (boxplot?) (you should comment)

```
classroom %>% group_by(schoolid) %>% mutate(medResFE_RE = median(resFE_RE, na.rm = T)) %>% ggplot(., aes(
```

```
## Warning: Removed 109 rows containing non-finite values (stat_boxplot).
```



Response:

The relationship within schools appears to be much less than before. The mean residuals for each school

Question 5

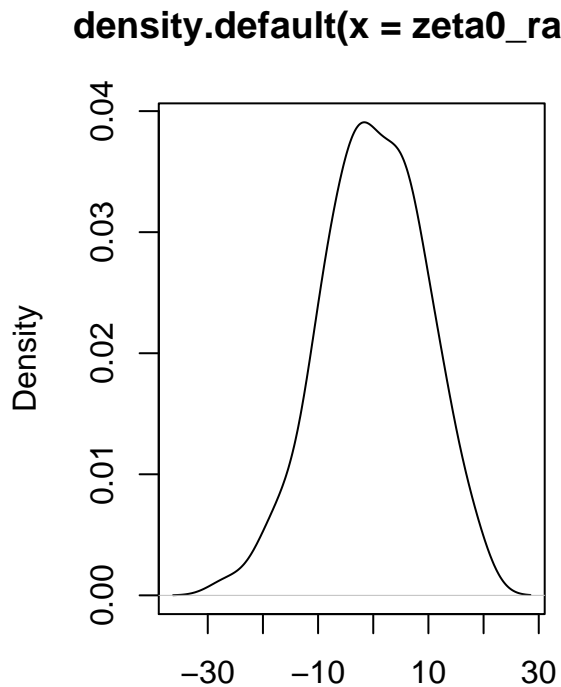
a. Generate the two sets of BLUPs (for random effects zeta0 and eta0)

```
# Insert code to generate the two sets of BLUPs (zeta0 and eta0)
ranefs <- ranef(lm1)
zeta0_ranef <- ranefs$schoolid[,1]
eta0_ranef <- ranefs$classid[,1]
```

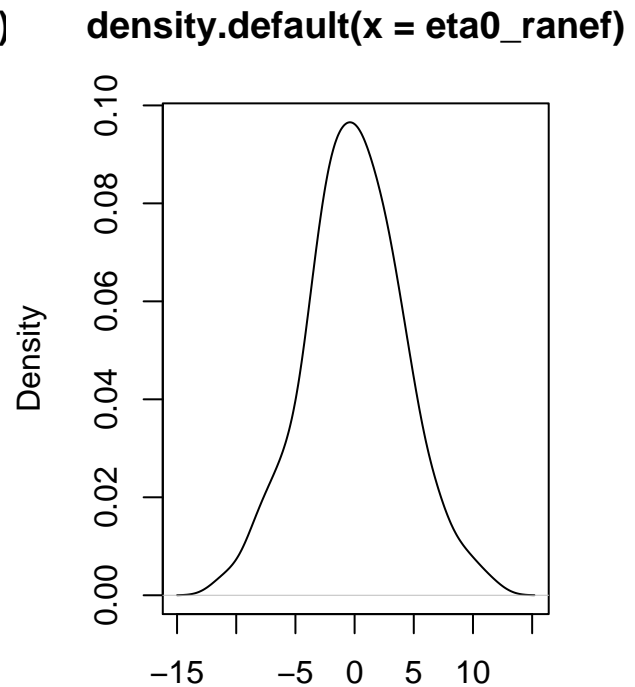
b. Examine these for normality (include evidence), and comment.

```
# Insert code to examine BLUPs for normality
# par(mfrow=c(1,2)) produces palette for one row of plots with two columns

par(mfrow = c(1,2))
plot(density(zeta0_ranef))
plot(density(eta0_ranef))
```



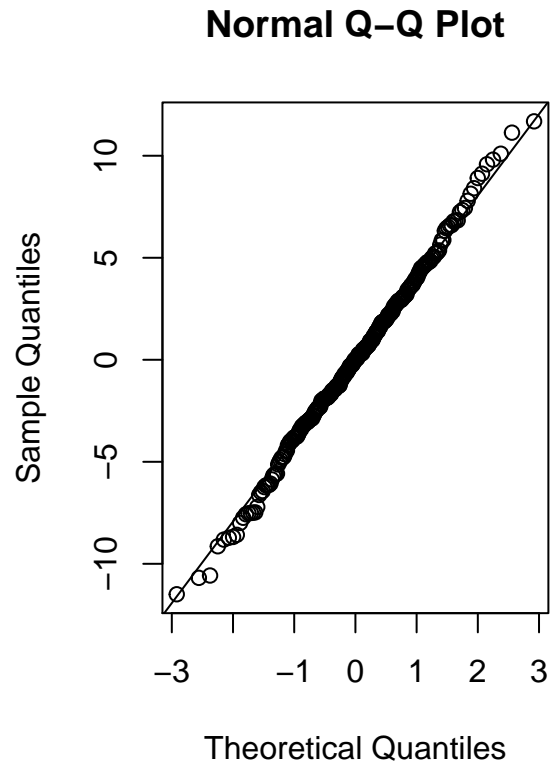
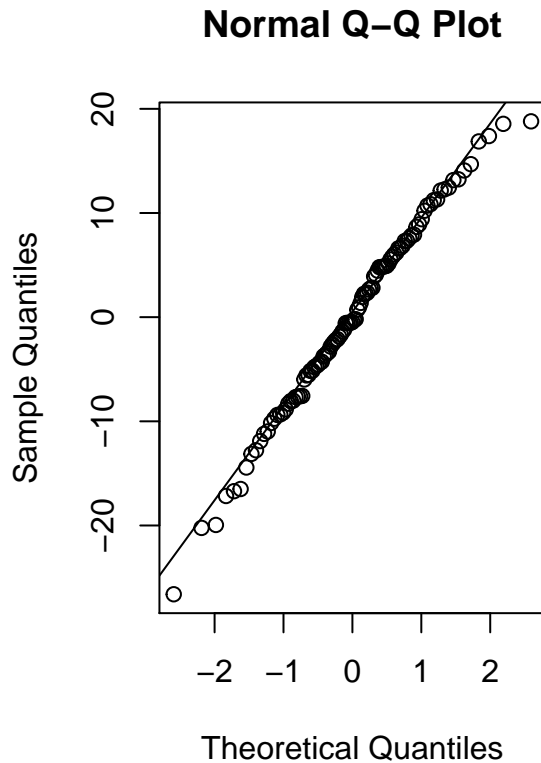
N = 105 Bandwidth = 3.23



N = 285 Bandwidth = 1.168

```
qqnorm(zeta0_ranef)
qqline(zeta0_ranef)

qqnorm(eta0_ranef)
qqline(eta0_ranef)
```



Response:

The density and QQ-plots of the random effects for schools and classrooms (ζ_0 and η_0) show evidence of normality.

Question 6

a. Fit a slightly more complicated model with the same fixed effects, but now add a random slope for minority, correlated with the random intercept, at the school level (keep the classroom level random intercept).

```
# Insert code to fit the slightly more complicated model and print the summary
lm2 <- lmerTest::lmer(math1st ~ housepov + yearstea + mathprep + mathknow +
  ses + sex + minority + (minority | schoolid) + (1 | classid), REML = T, data = classr)
summary(lm2)
```

```
## 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: classroom
##
## 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)  86.7    9.311
## schoolid (Intercept) 381.2   19.524
##      minority      343.2   18.525  -0.83
## Residual      1039.4   32.240
## Number of obs: 1081, groups: classid, 285; schoolid, 105
##
## Fixed effects:
##      Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  539.49369    5.65513  173.09178  95.399 < 2e-16 ***
## housepov     -16.06251   12.57477   99.99134  -1.277  0.204
## yearstea     -0.00437    0.13765  217.17884  -0.032  0.975
## mathprep     -0.29178    1.33537  198.06922  -0.218  0.827
## mathknow      1.63216    1.35929  224.78144   1.201  0.231
## ses           9.43095    1.54335 1063.13485   6.111 1.39e-09 ***
## sex          -0.86278    2.08382 1021.81437  -0.414  0.679
## minority     -16.37547    3.89604   58.24604  -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
```

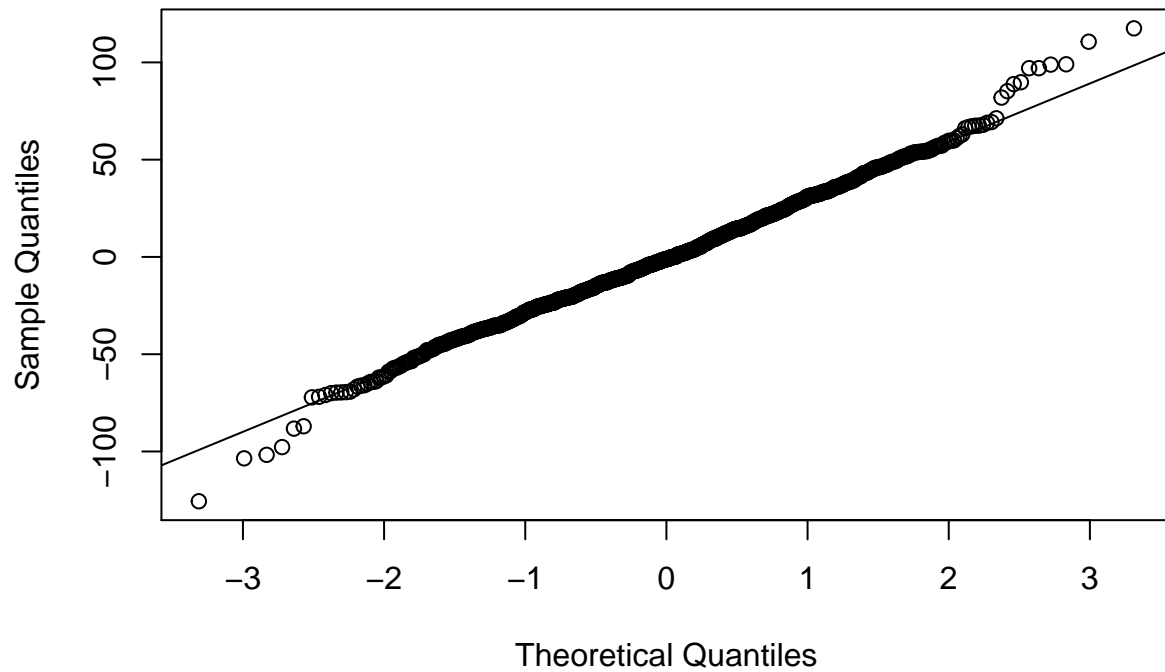
b. Construct the residual (individual, level 1) and the BLUPs for the remaining random effects. Call the new residual `resFE_RE` as before.

```
# Insert code to construct residual and BLUPs
resFE_RE <- residuals(lm2)
```

c. Examine all error estimates (individual level residuals, BLUPs (school and classroom level) for normality (and comment)).

```
qqnorm(resFE_RE)
qqline(resFE_RE)
```

Normal Q-Q Plot

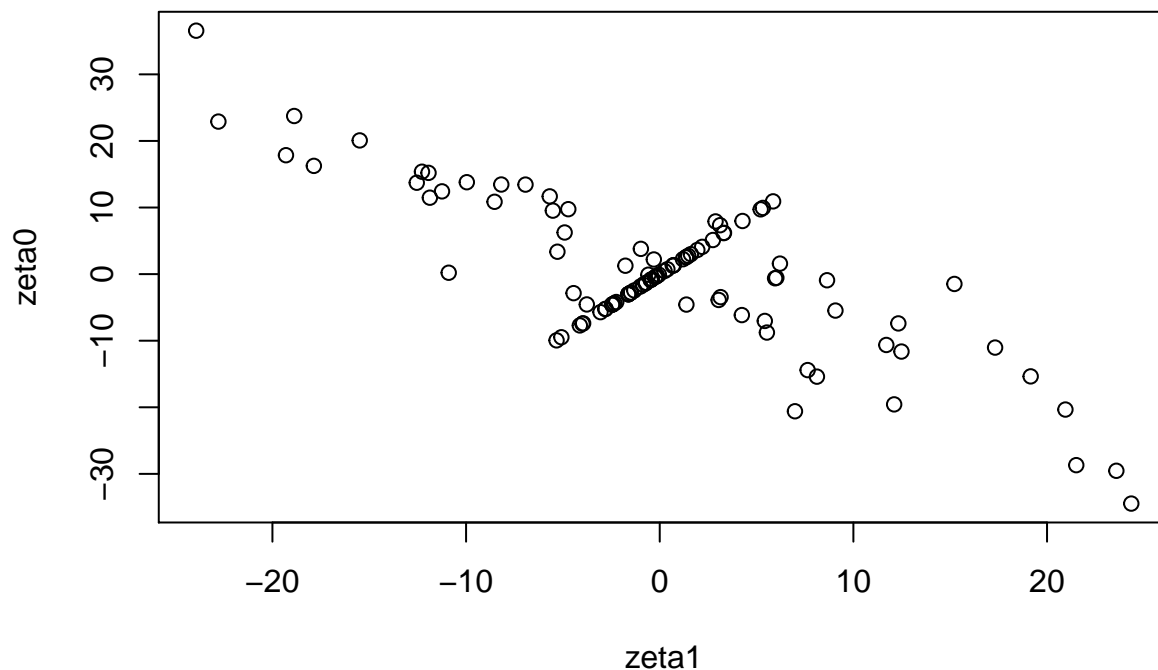


Response:

d. Plot zeta0 vs. zeta1 to see whether the estimated correlation is consistent with the observed. Briefly comment.

```
zeta0 <- ranef(lm2)$schoolid[,1]
zeta1 <- ranef(lm2)$schoolid[,2]

plot(x = zeta1, y = zeta0)
```



Response:

e. Track down those odd points in the scatterplot. What schools are they? Do they have anything in common? (You should comment)

```
# Insert code if you want to examine odd points

# Identify which schools are odd on the scatterplot:
test_df <- data.frame(zeta0 = zeta0, zeta1 = zeta1, z0z1 = zeta0*zeta1)
which(test_df$z0z1 > 0)

## [1] 1 4 5 9 10 12 14 16 17 19 20 22 23 24 25 26 28
## [18] 30 31 32 33 34 37 38 40 42 43 45 46 47 48 51 52 56
## [35] 57 58 59 66 67 68 71 76 77 78 82 84 85 86 87 88 94
## [52] 96 98 100 101 103 104

# Add "oddschools" indicator to dataset:
classroom$oddschools <- classroom$schoolid %in% which(test_df$z0z1 > 0)

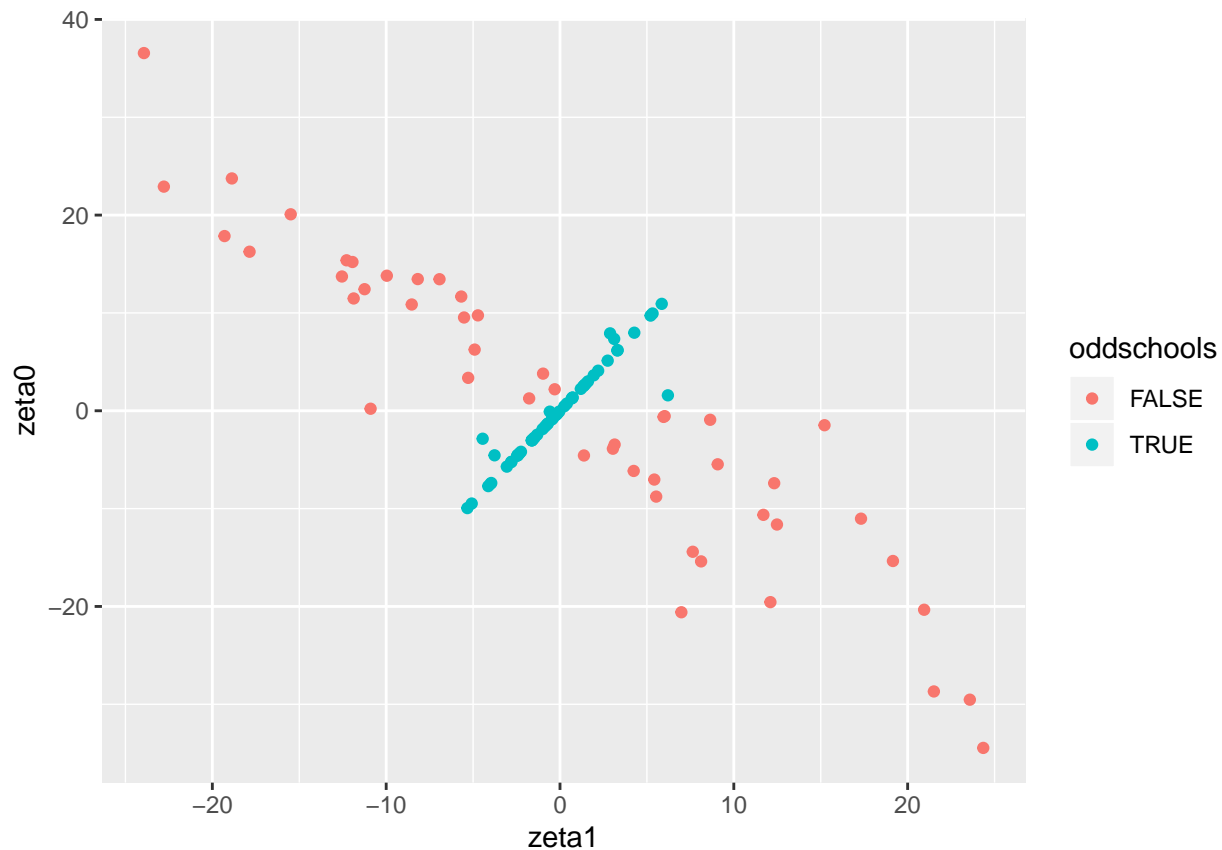
# Calculate percentage of minority students in each odd school:
classroom %>% group_by(oddschools) %>% summarize(minority_avg = mean(minority))

## # A tibble: 2 x 2
## oddschools minority_avg
## <lgl> <dbl>
## 1 FALSE 0.569
## 2 TRUE 0.772

# Show odd schools in plot of zeta1 v. zeta0:
distinct(test_df %>%
  mutate(schoolid = row_number()) %>%
  left_join(classroom[, c("schoolid", "oddschools")])) %>%
  ggplot(., aes(x = zeta1, y = zeta0, color = oddschools)) + geom_point()
```



```
## Joining, by = "schoolid"
```



Response:

The “odd” schools in the scatterplot are those schools that have mostly minority populations. This makes it difficult to estimate a random slope for these schools because there is little variation in minority (i.e. slope estimates close to 0).

Question 7

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*.

```
# Insert code to create the variables math0 and math1 and to reshape data
personperiod <- classroom %>% mutate(math0 = mathkind, math1 = mathkind + mathgain)

class_pp <- reshape(personperiod, varying = c("math0", "math1"), v.names = "math", timevar = "year",
times = c(0, 1), direction = "long")
```

Question 8

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

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

```
lm3 <- lmerTest::lmer(math ~ year + (1 | schoolid), data = class_pp)
summary(lm3)
```

```
## 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
```

b. Write down the model

Equation:

$$MATH_{tijk} = b_0 + \zeta_{0k} + b_1 TIME_{tijk} + \epsilon_{tijk}$$

with $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, $\epsilon_{tijk} \sim N(0, \sigma_{\epsilon}^2)$, independent of each other

c. Add random intercepts for child

```
# Insert code to fit new model and print summary output
lm4 <- lmerTest::lmer(math ~ year + (1 | schoolid/classid), data = class_pp)
summary(lm4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: math ~ year + (1 | schoolid/classid)
## Data: class_pp
##
## REML criterion at convergence: 23908.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -5.2194 -0.5853 0.0012 0.6163 3.8138
##
## Random effects:
## Groups Name Variance Std.Dev.
## classid:schoolid (Intercept) 145.9 12.08
## schoolid (Intercept) 294.9 17.17
## Residual 1179.1 34.34
## Number of obs: 2380, groups: classid:schoolid, 312; schoolid, 107
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 464.757 2.113 131.898 219.95 <2e-16 ***
## year 57.566 1.408 2061.878 40.89 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr)
## year -0.333
```

d. Write down the model

Equation:

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

with $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\epsilon_{tijk} \sim N(0, \sigma_{\epsilon}^2)$, independent of each other

Question 9

Report original and new variance estimates of $\sigma_{\zeta_0}^2$ (between schools) and σ_{ϵ}^2 (within schools):

$\sigma_{\zeta_0}^2$:

- Original 348.7
- New: 294.9

σ_{ϵ}^2 :

- Original: 1268.4
- New: 1179.1

a. Compute a pseudo R^2 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?

Insert code to compute psuedo R^2 or do this inline

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

Response:

Question 10

Add a random slope (ζ_1) for the trend (year) within schools (uncorrelated with random intercept (ζ_0))

```
# Insert code to fit model and print out summary
```

a. Generate the BLUPs for the random effects and examine whether the independence between zeta_0 and zeta_1 is reflected in a scatterplot of these two sets of effects. (you should comment)

```
# Insert code to generate BLUPs
```

Response:

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

```
# Insert code to compute terms or do this inline
```

i. In which year is there more between school variation, net of all else, (you should comment)?

Response:

Question 11

If you ran the model BY YEAR, and removed the year trend from the model, would you get the same estimates for the variances between schools? ***(you should comment)***

```
# Insert code to fit the two models by year and print out the summary
```

Response:

Question 12

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

a. Is the correlation significant? (you should comment)

```
# Insert code to fit model, print the summary output, and compare models
```

Response:

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

```
# Insert code to compute terms or do this inline
```

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

Response: