

MLM Project Part 4

Bianca Brusco, Clare Clingain, Kaushik Mohan, & Frankie Wunschel

4/26/2018

Part 4: Kaushik

Reload the data and make person-period file

```
#re-read data
classroom <- read.csv("classroom.csv")
classroom2 <- na.omit(classroom)
#new variables
classroom2 <- classroom2 %>% mutate(math0 = mathkind) %>% mutate(math1 = mathkind+mathgain)
#reshape the data
class_pp <- reshape(classroom2, varying = c("math0", "math1"), v.names = "math", timevar = "year",
times = c(0, 1), direction = "long")
```

Baseline model: unconditional growth model

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

where t represents occasion (in this case, year/grade), i represents students, j represents classrooms and k represents schools. $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, $\zeta_{1k} \sim N(0, \sigma_{\zeta_1}^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_{\epsilon}^2)$ all independent of each other except for ζ_{0k} and ζ_{1k} having a correlation $\rho_{\zeta_0\zeta_1}$.

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

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
## to degrees of freedom [lmerMod]
## Formula: math ~ year + (year | schoolid) + (1 | childid)
## Data: class_pp
##
## REML criterion at convergence: 21391.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6737 -0.4699  0.0038  0.4683  3.4882
##
## Random effects:
##  Groups   Name                Variance Std.Dev. Corr
##  childid  (Intercept)         749.0     27.37
##  schoolid (Intercept)         373.5     19.33
##          year                 112.4     10.60   -0.53
## Residual                    547.8     23.41
## Number of obs: 2162, groups:  childid, 1081; schoolid, 105
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
```

```
## (Intercept) 465.257      2.241 101.270   207.6   <2e-16 ***
## year        58.006      1.491  95.410    38.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## year -0.486
```

Add student, classroom and school level fixed effects

$$MATH_{tijk} = b_0 + \delta_{0ijk} + (b_1 + \zeta_{1k})TIME_{tijk} + b_2SEX_{ijk} + b_3SES_{ijk} + b_4MINORITY_{ijk} + b_5YEARSTEAK_{jk} + b_6MATHKNOW_{ijk}$$

where t represents occasion (in this case, year/grade), i represents students, j represents classrooms and k represents schools. $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, $\zeta_{1k} \sim N(0, \sigma_{\zeta_1}^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ all independent of each other except for ζ_{0k} and ζ_{1k} having a correlation $\rho_{\zeta_0\zeta_1}$.

```
fit2 <- lmer(math ~ year + sex + ses + minority + yearstea + mathknow + mathprep + housepov + (year|schoolid) + (1|childid))
summary(fit2)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
## to degrees of freedom [lmerMod]
## Formula:
## math ~ year + sex + ses + minority + yearstea + mathknow + mathprep +
## housepov + (year | schoolid) + (1 | childid)
## Data: class_pp
##
## REML criterion at convergence: 21275.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6812 -0.4784  0.0040  0.4712  3.4245
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## childid (Intercept) 689.5 26.26
## schoolid (Intercept) 249.2 15.79
## year 114.2 10.69 -0.53
## Residual 547.4 23.40
## Number of obs: 2162, groups: childid, 1081; schoolid, 105
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 483.44962 4.79270 369.60000 100.872 < 2e-16 ***
## year 57.90494 1.49801 95.30000 38.655 < 2e-16 ***
## sex -0.52171 1.95332 1033.80000 -0.267 0.789
## ses 9.59015 1.43816 1071.20000 6.668 4.14e-11 ***
## minority -16.01167 2.82859 705.00000 -5.661 2.19e-08 ***
## yearstea 0.02304 0.11846 876.20000 0.194 0.846
## mathknow -0.22237 1.17471 768.50000 -0.189 0.850
## mathprep -1.08389 1.14195 932.70000 -0.949 0.343
## housepov -18.17699 12.40599 108.10000 -1.465 0.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##      (Intr) year    sex    ses    minrty yearst mthknw mthprp
## year      -0.204
## sex        -0.194 -0.001
## ses         -0.120 -0.001  0.028
## minority   -0.333  0.004 -0.007  0.157
## yearstea    -0.240  0.001  0.018 -0.035  0.021
## mathknow    -0.085 -0.001  0.002 -0.012  0.106  0.038
## mathprep    -0.578 -0.002 -0.014  0.054  0.003 -0.182 -0.001
## housepov    -0.463  0.004 -0.007  0.080 -0.178  0.067  0.060  0.036
```

For year==0:

*what percent of between school differences were explained as you go from the baseline to the second model?

For the baseline model:

$$V_{1BS} = \sigma_{\zeta_0}^2 + 2 * year * \rho_{\zeta_0 \zeta_1} \sigma_{\zeta_0} \sigma_{\zeta_1} + year^2 * \sigma_{\zeta_1}^2$$

$$V_{1BS}(year = 0) = \sigma_{\zeta_0}^2 = 373.5$$

After adding fixed-effects:

$$V_{2BS} = \sigma_{\zeta_0}^2 + 2 * year * \rho_{\zeta_0 \zeta_1} \sigma_{\zeta_0} \sigma_{\zeta_1} + year^2 * \sigma_{\zeta_1}^2$$

$$V_{2BS}(year = 0) = \sigma_{\zeta_0}^2 = 249.2$$

The percent difference in between-school variance for $year = 0$ is be given by:

$$\frac{V_{1BS} - V_{2BS}}{V_{1BS}} = \frac{373.5 - 249.2}{373.5} = 33.28\%$$

Model 2 explains 33.28% of the between-school variance for $year = 0$.

*what percent of between child differences were explained as you go from the baseline to the second model?

For the baseline model:

$$V_{1BC}(year = 0) = \sigma_{\delta_0}^2 = 749.0$$

After adding fixed effects:

$$V_{2BC}(year = 0) = \sigma_{\delta_0}^2 = 689.5$$

The percent difference in between-child variance explained by the second model for $year = 0$ is given by:

$$\frac{V_{1BC} - V_{2BC}}{V_{1BC}} = \frac{749.0 - 689.5}{749.0} = 7.94\%$$

Model 2 explains 7.94% of the between-child variance for $year = 0$.

For year==1:

*what percent of between school differences were explained as you go from the baseline to the second model?

For the baseline model:

$$V_{1BS} = \sigma_{\zeta_0}^2 + 2 * year * \rho_{\zeta_0 \zeta_1} \sigma_{\zeta_0} \sigma_{\zeta_1} + year^2 * \sigma_{\zeta_1}^2$$

$$V_{1BS}(year = 1) = 373.5 + 2(-0.53)(19.33)(10.60) + 112.4 = 268.71$$

After adding fixed-effects:

$$V_{2BS} = \sigma_{\zeta_0}^2 + 2 * year * \rho_{\zeta_0 \zeta_1} \sigma_{\zeta_0} \sigma_{\zeta_1} + year^2 * \sigma_{\zeta_1}^2$$

$$V_{2BS}(year = 1) = 249.2 + 2(-0.53)(15.79)(10.69) + 114.2 = 184.48$$

The percent difference in between-school variance for $year = 0$ is be given by:

$$\frac{V_{1BS} - V_{2BS}}{V_{1BS}} = \frac{268.71 - 184.48}{268.71} = 31.35\%$$

Model 2 explains 31.35% of the between-school variance for $year = 1$.

*what percent of between child differences were explained as you go from the baseline to the second model?

For the baseline model:

$$V_{1BC}(year = 1) = \sigma_{\delta_0}^2 = 749.0$$

After adding fixed effects:

$$V_{2BC}(year = 1) = \sigma_{\delta_0}^2 = 689.5$$

The percent difference in between-child variance explained by the second model for $year = 1$ is given by:

$$\frac{V_{1BC} - V_{2BC}}{V_{1BC}} = \frac{749.0 - 689.5}{749.0} = 7.94\%$$

Model 2 explains 7.94% of the between-child variance for $year = 1$.

Based on significance,

- what factors seem useful in describing (“explaining”) differences between student outcomes?
- Point out the direction of the effect.

SES and *MINORITY* status are the significant fixed-effects terms in the model at $\alpha = 0.05$ implying that these terms (being in Level 1) help to explain the between-student variance conditional on the school.

The coefficient on *SES* is positive meaning that two students in the same school and student-level random effect and all else equal, the one with the higher *SES* has a higher Math score.

The coefficient on *MINORITY* status is negative meaning that two students in the same school and student-level random effect and all else equal, the one who is classified as a Minority student has a lower Math score.

Add random slope for SES

$$MATH_{tijk} = b_0 + \delta_{0ijk} + (b_1 + \zeta_{1k})TIME_{tijk} + b_2SEX_{ijk} + (b_3 + \zeta_{3k})SES_{ijk} + b_4MINORITY_{ijk} + b_5YEARSTEAK_{ijk} + b_6MATH_{tijk}$$

where t represents occasion (in this case, year/grade), i represents students, j represents classrooms and k represents schools. $\delta_{0ijk} \sim N(0, \sigma_{\delta_0}^2)$, $\zeta_{0k} \sim N(0, \sigma_{\zeta_0}^2)$, $\zeta_{1k} \sim N(0, \sigma_{\zeta_1}^2)$, $\zeta_{3k} \sim N(0, \sigma_{\zeta_3}^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_{\epsilon}^2)$ all independent of each other except for ζ_{0k} , ζ_{1k} and ζ_{3k} could be correlated.

```
fit3 <- lmer(math ~ year + sex + ses + minority + yearstea + mathknow + mathprep + housepov + (ses+year
summary(fit3)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
## to degrees of freedom [lmerMod]
## Formula:
## math ~ year + sex + ses + minority + yearstea + mathknow + mathprep +
## housepov + (ses + year | schoolid) + (1 | childid)
## Data: class_pp
##
## REML criterion at convergence: 21273.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6663 -0.4808  0.0048  0.4722  3.4249
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## childid (Intercept) 668.11 25.848
## schoolid (Intercept) 251.26 15.851
## ses 46.49 6.818 -0.03
## year 114.87 10.718 -0.53 0.15
## Residual 547.29 23.394
## Number of obs: 2162, groups: childid, 1081; schoolid, 105
##
## Fixed effects:
## Estimate Std. Error df t value Pr(>|t|)
## (Intercept) 483.20541 4.79250 363.60000 100.825 < 2e-16 ***
## year 57.88435 1.50002 95.30000 38.589 < 2e-16 ***
## sex -0.70446 1.94358 1016.30000 -0.362 0.717
## ses 9.32191 1.63892 69.40000 5.688 2.81e-07 ***
## minority -16.26826 2.83356 658.70000 -5.741 1.43e-08 ***
## yearstea 0.03404 0.11840 876.00000 0.288 0.774
## mathknow -0.28980 1.17539 774.90000 -0.247 0.805
## mathprep -1.04672 1.13615 916.70000 -0.921 0.357
## housepov -17.57739 12.45037 105.20000 -1.412 0.161
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) year sex ses minrty yearst mthknw mthprp
## year -0.205
## sex -0.192 -0.001
## ses -0.098 0.047 0.025
## minority -0.335 0.004 -0.007 0.125
## yearstea -0.240 0.002 0.020 -0.027 0.020
## mathknow -0.080 -0.001 0.001 0.001 0.100 0.037
```

```
## mathprep -0.575 -0.002 -0.015 0.045 0.002 -0.182 -0.003
## housepov -0.463 0.005 -0.007 0.074 -0.180 0.067 0.059 0.037
```

is the estimated s.d. (square root of variance) of the random slope associated with SES large enough so that a value ± 1 s.d. is sufficient to “cancel” (or flip the sign) the fixed effect for this predictor?

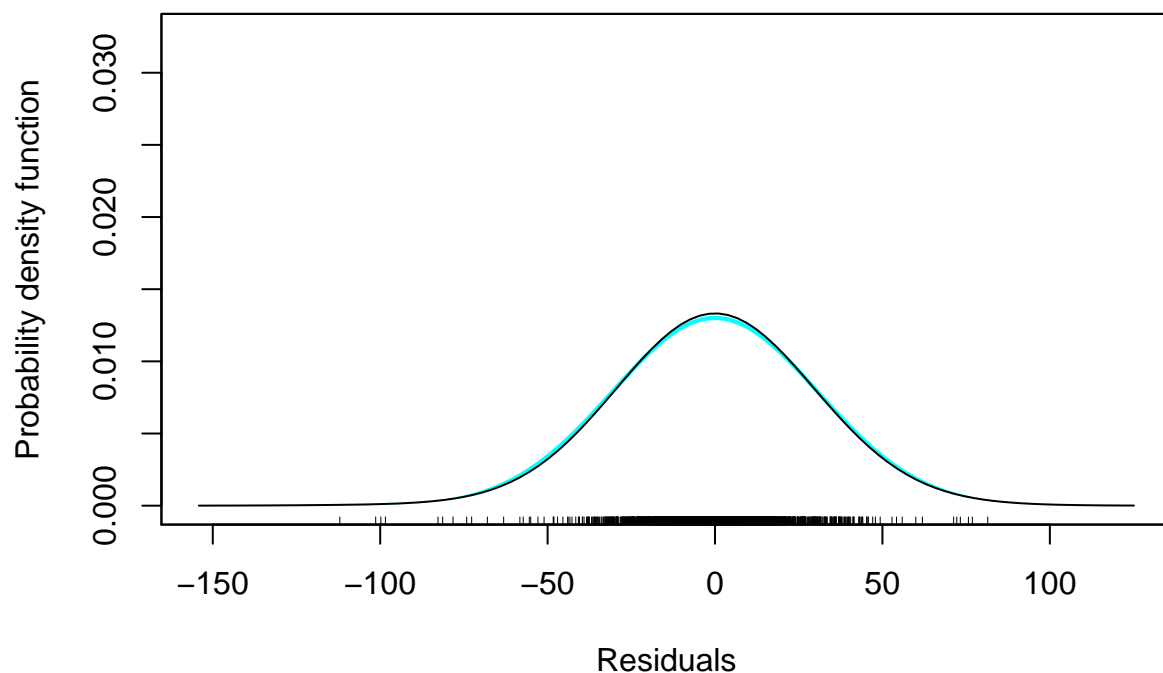
The estimated standard deviation of the random slope associated with SES is 6.818. We note that this is not large enough to “cancel” or flip the sign of the fixed-effect on SES within ± 1 standard deviation.

The majority of the values (middle 68%) for the fixed effect on SES is within the range $[(9.32191 - 6.818), (9.32191 + 6.818)] = [2.50391, 16.13991]$

Residuals and Q-Q Plot

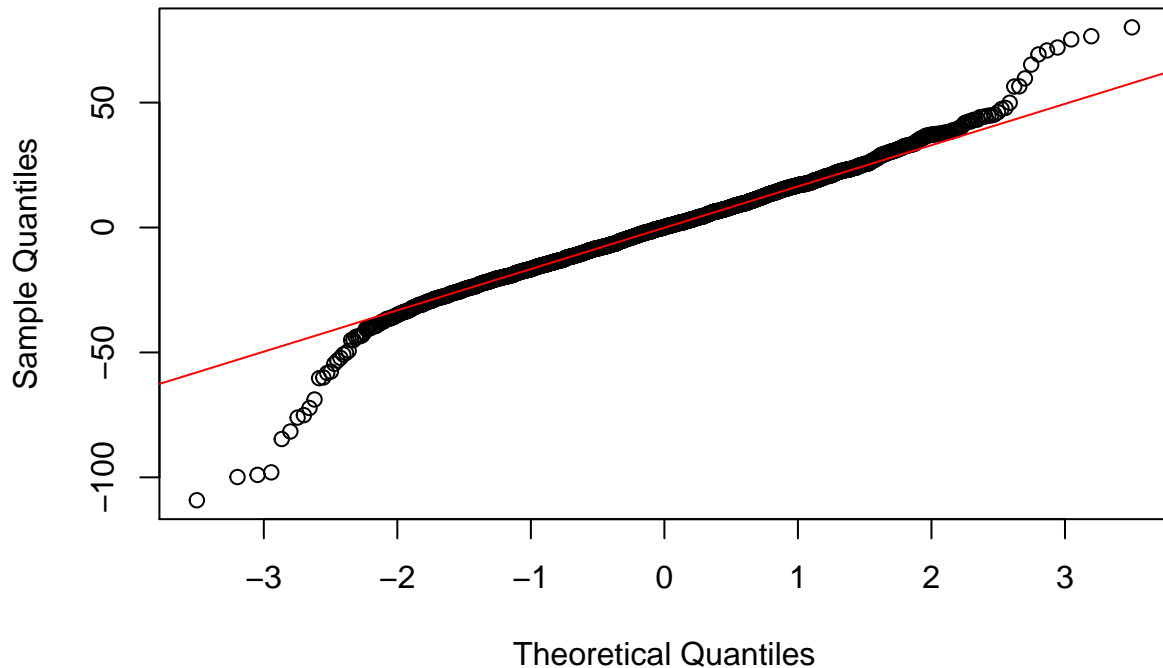
compute residuals in this final model. generate a qq plot and density (STATA: `qnorm`; `kdensity` ..., `normal`)
Is there any reason to question the normality assumption?

```
fit3.residuals <- residuals(fit3)
sm.density(fit3.residuals,model="Normal",xlab="Residuals")
```



```
qqnorm(y=fit3.residuals)
qqline(y=fit3.residuals,col=2)
```

Normal Q-Q Plot



The residuals seem to have slightly heavier tails compared to a standard normal distribution. From the density plot, we also note that given the sample size, the peak is slightly higher than expected for a normal distribution. The distribution of residuals does not fall within the expected range given by the blue region although the deviation is minimal. Within a range of 4-sd (between ± 2 sd), the Residual quantiles are seen to be linear compared to the Theoretical quantiles implying that the residuals are indeed quite normally distributed for the most part.

BLUPs for all 4 random effects & Scatter plots

generate an all pairs scatter plot matrix (4x4) of these * note whether or not you identify any concerns from these scatterplots.

```
ranefs <- ranef(fit3)
Delta0 <- ranefs$childid
idx.school <- match(classroom2$schoolid, sort(unique(classroom2$schoolid)))
Zeta0 <- ranefs$schoolid[idx.school,1]
Zeta1 <- ranefs$schoolid[idx.school,2]
Zeta2 <- ranefs$schoolid[idx.school,3]

ranefs <- data.frame(delta0=Delta0,zeta0=Zeta0,zeta1=Zeta1,zeta2=Zeta2)
colnames(ranefs) <- c("Delta0","Zeta0","Zeta1","Zeta2")
pairs(ranefs,cex=0.5)
```

