MLM Project 5 Part 1

Jemar Bather, Zarni Htet, Martha Moreno, Sarah Rosenbach, James Wu

#library(readr)  
dat <- read.csv("../classroom.csv")  
cc.data <- dat[(complete.cases(dat)),]

\*we have two models to consider, but we first focus on the simpler one:

attach(cc.data)  
cc.data$math1st <- mathkind + mathgain

require(lme4)

## Loading required package: lme4

## Loading required package: Matrix

require(lmerTest)

## Loading required package: lmerTest

##   
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':  
##   
## lmer

## The following object is masked from 'package:stats':  
##   
## step

#options(digits = 5)  
fit1 <- lmer(math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex + minority + (1|schoolid/classid), data=cc.data)  
summary(fit1)

## Linear mixed model fit by REML t-tests use Satterthwaite approximations  
## to degrees of freedom [lmerMod]  
## Formula:   
## math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +   
## minority + (1 | schoolid/classid)  
## Data: cc.data  
##   
## 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   
## Residual 1064.95 32.63   
## 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 275.40000 101.585 < 2e-16 \*\*\*  
## housepov -17.64847 13.21757 113.90000 -1.335 0.184   
## yearstea 0.01129 0.14141 226.80000 0.080 0.936   
## mathprep -0.27705 1.37583 205.30000 -0.201 0.841   
## mathknow 1.35004 1.39168 234.50000 0.970 0.333   
## ses 10.05075 1.54484 1066.50000 6.506 1.18e-10 \*\*\*  
## sex -1.21419 2.09483 1022.40000 -0.580 0.562   
## minority -16.18678 3.02605 704.50000 -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

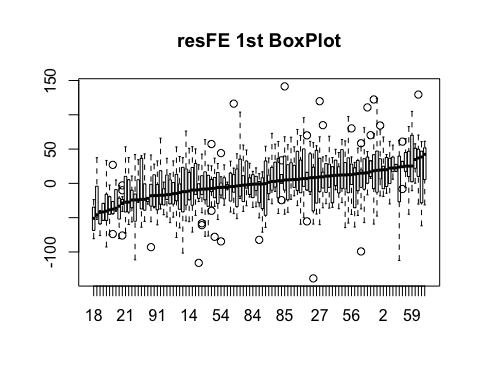
*manually construct the residual that removes only the 'fixed effects'* hint: predict yhat, xb will generate the prediction for the outcome based on the fixed effects only \*then subtract it from the outcome; call this residual: resFE

pred.yhat <- predict(fit1, re.form = ~0) #This ignore random effects  
resFE <- cc.data$math1st - pred.yhat

\*show that this residual is not independent within schools in some manner.

Below, we have sorted each of the fixed effect residuals for the schools by their median from lowest to highest. If we are assuming independence of residuals within schools, then, we would expect to see a flat-line showing that residuals across schools is random. However, since we are seeing a positive trending line, we can conclude that residuals are not random and it may be dependent on school.

ord<-order(unlist(tapply(resFE,schoolid,median)))  
boxplot(split(resFE,schoolid)[ord], main = "resFE 1st BoxPlot")



*construct the residual that utilizes the BLUPs for the random effects. Do it in these stages:*  i) predict and save zeta0 \* ii) predit and save eta0 \* iii) generate a new residual, called resFE\_RE which subtracts yhat, zeta0 and eta0 from the outcome \*note: there is an easier way to get the residuals in this case, predict ..., residuals, but we need to do it manually.

idx.sch <- match(cc.data$schoolid, sort(unique(cc.data$schoolid)))  
idx.cls <- match(cc.data$classid, sort(unique(cc.data$classid)))

attach(cc.data)

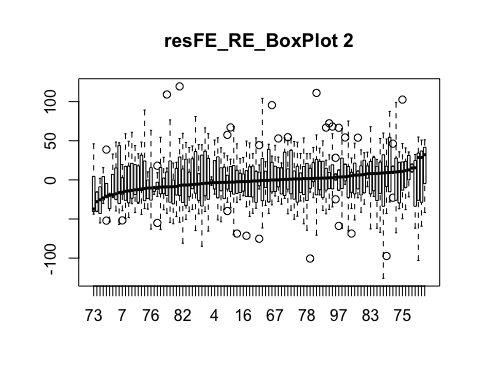
## The following objects are masked from cc.data (pos = 6):  
##   
## childid, classid, housepov, mathgain, mathkind, mathknow,  
## mathprep, minority, schoolid, ses, sex, yearstea

ranefs <- ranef(fit1) #Pulling out the random effects  
zeta0 <- ranefs$schoolid[, 1] #Random Intercept for School  
eta0 <- ranefs$classid[, 1] #Random Intercept for Class  
cc.data$zeta0 <- zeta0[idx.sch] # Putting Random Intercept of School back into the data set  
cc.data$eta0 <- eta0[idx.cls] #Putting Random Intercept of Class back into the data set  
resFE\_RE <- math1st - pred.yhat - cc.data$zeta0 - cc.data$eta0 #This accounts for removing both fixed and random effects and the random effects are removed manually. Therefore, residuals should be less dependent than the previous one. Thus, we expect to see a flatter line.

* show that these new residuals, resFE\_RE are MUCH LESS (if not completely un-) correlated within school
* using the same method as before (boxplot?)

We expected a flatter residual line or something flat on 0 because the random effects as well as fixed were removed from the residuals. If we don't have any additional model misspecification, we would have a flatter line or something close to it which is we see below.

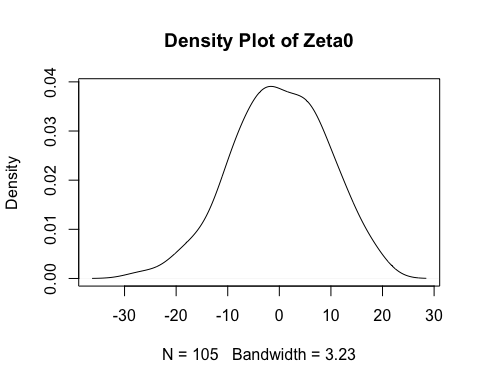
ord2<-order(unlist(tapply(resFE\_RE,schoolid,median)))  
boxplot(split(resFE\_RE,schoolid)[ord2], main = "resFE\_RE\_BoxPlot 2")



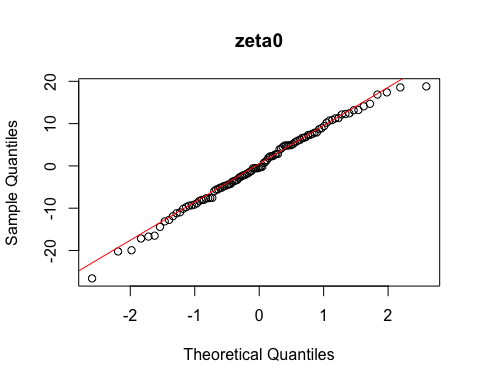
*examine the two sets of BLUPs (for random effects zeta0 and eta0) for normality* first 'tag' a single value from each grouping (e.g., school or classroom) so that you only have \*as many BLUPs as the grouping factor (should be less of an issue in R)

If we examine the density and qqplot of the Zeta0 below, it's close to symmetric, but with seeminlgy heavy tail on left. Since the sample size is 105, we can tolerate this lack of perfect and normality.

plot(density(zeta0), main = "Density Plot of Zeta0")

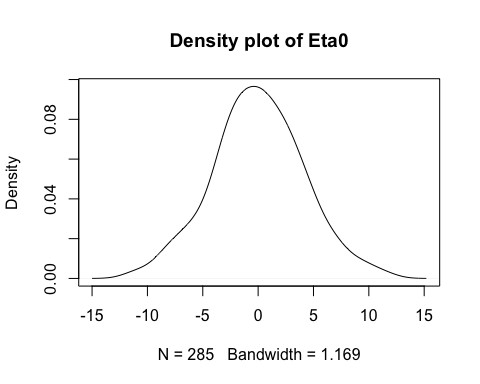


qqnorm(zeta0, main = "zeta0")  
qqline(zeta0, col ="red")



If we examine the density and qqplot of the eta0 below, it's close to symmetric, and the tails look less heavy than the ones before. Since the sample size is 285, we think this is reasonable to accept it as normal.

plot(density(eta0), main = "Density plot of Eta0")



qqnorm(eta0, main = "eta0")  
qqline(eta0, col = "red")

