MLM Project-Part2

Junyan Yao

5/10/2017

I examed why multi-level modeling is important in Classroom data example.

```
#use "classroom.dta"
require(foreign)
```

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

```
require(lme4)
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
dat<-read.dta("/Users/YaoJunyan/Documents/NYU/Spring 2017/MLM Nested data/datasets/class
room.dta")

dat2<-dat[(complete.cases(dat)),]  #remove the missing values
math1st<- dat2$mathkind+dat2$mathgain
dat2<- cbind(dat2,math1st)
```

*Fit the first model

```
#mixed math1st housepov yearstea mathprep mathknow ses sex minority || schoolid: || clas
sid:, reml (THIS IS THE MODEL IN STATA)
M1 <- lmer(math1st~housepov+yearstea+mathprep+mathknow+ses+sex+minority+(1|schoolid)+
(1|classid),data=dat2)
print(summary(M1))
```

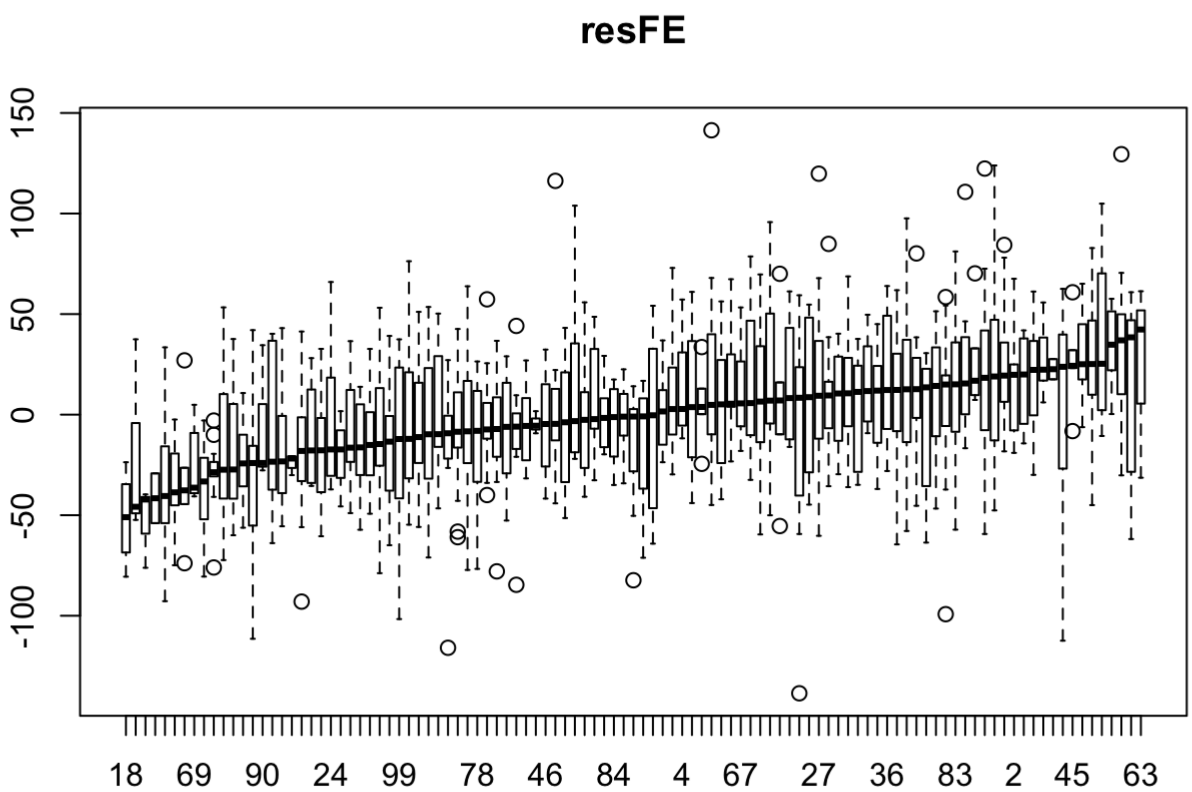
```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +
##     minority + (1 | schoolid) + (1 | classid)
##    Data: dat2
##
## REML criterion at convergence: 10729.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8580 -0.6134 -0.0321  0.5971  3.6598
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  classid  (Intercept)   93.89   9.69
##  schoolid (Intercept)  169.45  13.02
##  Residual             1064.95  32.63
## Number of obs: 1081, groups:  classid, 285; schoolid, 105
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 539.63042    5.31210  101.59
## housepov    -17.64847   13.21757   -1.34
## yearstea      0.01129    0.14141    0.08
## mathprep     -0.27705    1.37583   -0.20
## mathknow      1.35004    1.39168    0.97
## ses          10.05075    1.54484    6.51
## sex          -1.21419    2.09483   -0.58
## minority    -16.18678    3.02605   -5.35
##
## 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' (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)

```
yhat1<- predict(M1, re.form=~0)

resFE<- dat2$math1st - yhat1
#This histogram shows that the res_FE have a trend and it is obviously not independent w
ithin schools
```

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

```
ord <- order(unlist(tapply(resFE, dat2$schoolid, median)))
boxplot(split(resFE, dat2$schoolid)[ord], main="resFE")
```



*Construct the residual that utilizes the BLUPs for the random effects.Do it in these stages:

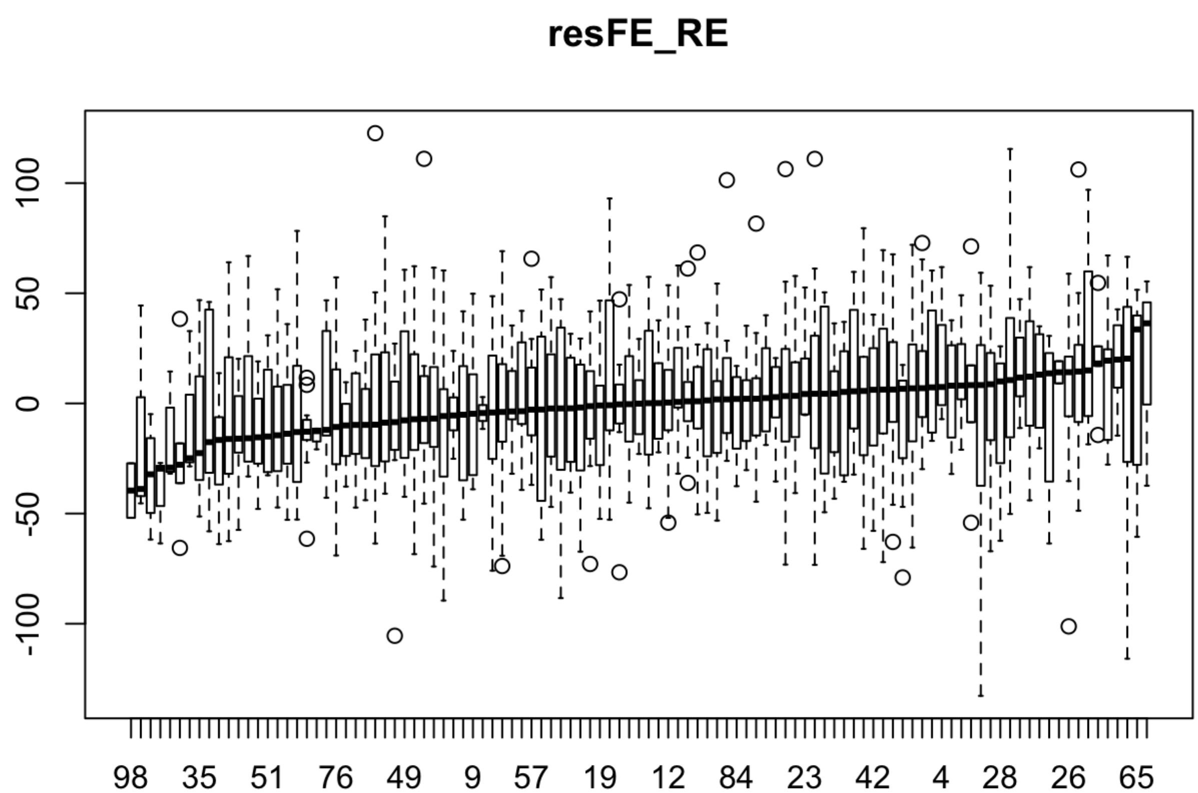
```
# * i) predict and save zeta0
zeta0<- ranef(M1)$schoolid[,1]
#* ii) predit and save eta0
eta0<-ranef(M1)$classid[,1]
#* iii) generate a new residual, called resFE_RE which subtracts yhat, zeta0 and eta0 fr
om the outcome

sch<- unique(dat2$schoolid)
zeta0_new<- cbind(sch,zeta0)

classid<- unique(dat2$classid)
eta0_new<- cbind(classid, eta0)
#merge this to the master dataset
require(plyr)
```

```
## Loading required package: plyr
```

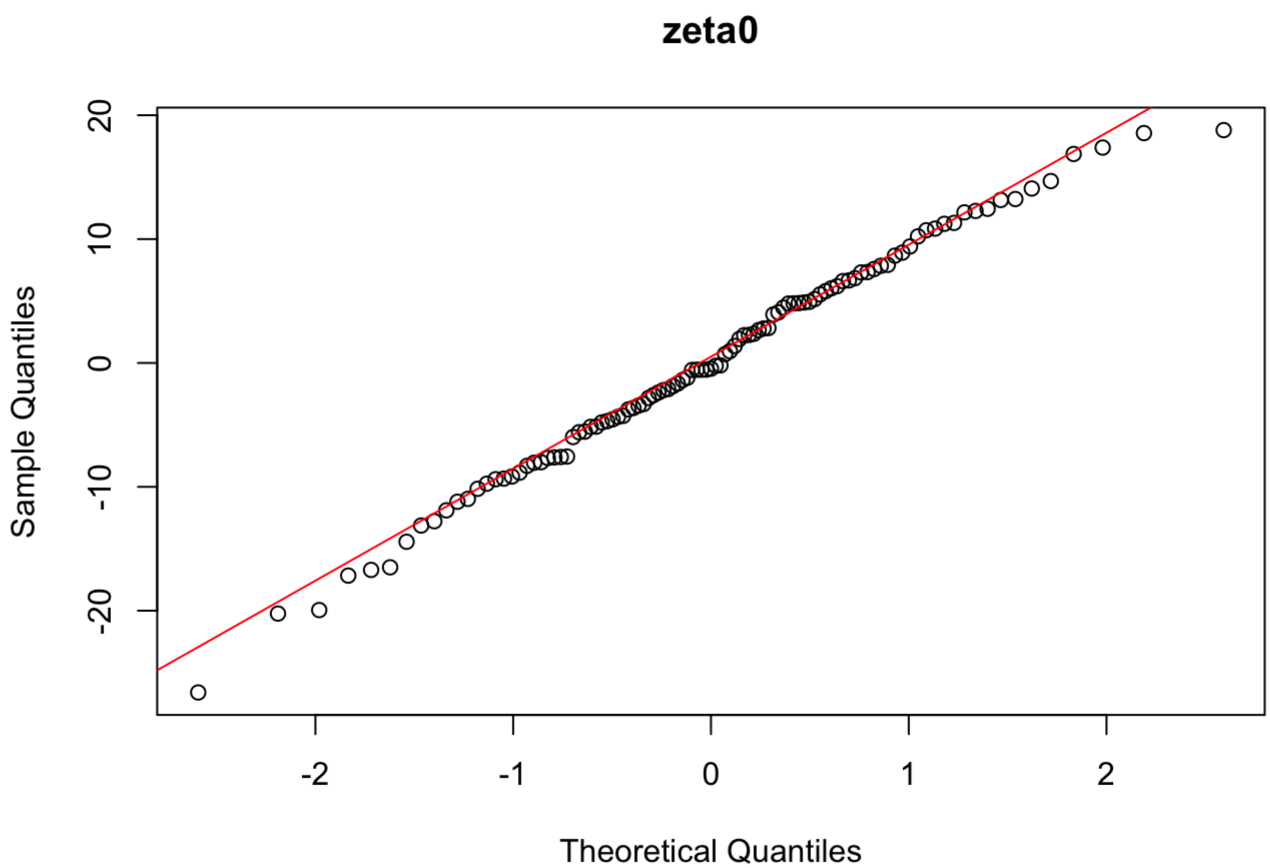
```
zeta0_new<- data.frame(schoolid=zeta0_new[,1], zeta0 = zeta0) #make it to be a data fram
e
newdat<- join(dat2, zeta0_new, by="schoolid", type="left", match="all") #merge zeta0 int
o master dataset

eta0_new<- data.frame(classid=classid, eta0= eta0) #make it to be a data frame
newdat2<- join(newdat, eta0_new, by="classid", type="left", match="all")

resFE_RE <- newdat2$math1st - yhat1 - newdat2$zeta0 - newdat2$eta0
newdat3<- cbind(newdat2,resFE_RE)
```

- show that these new residuals, resFE_RE are MUCH LESS (if not completely un-) correlated within school

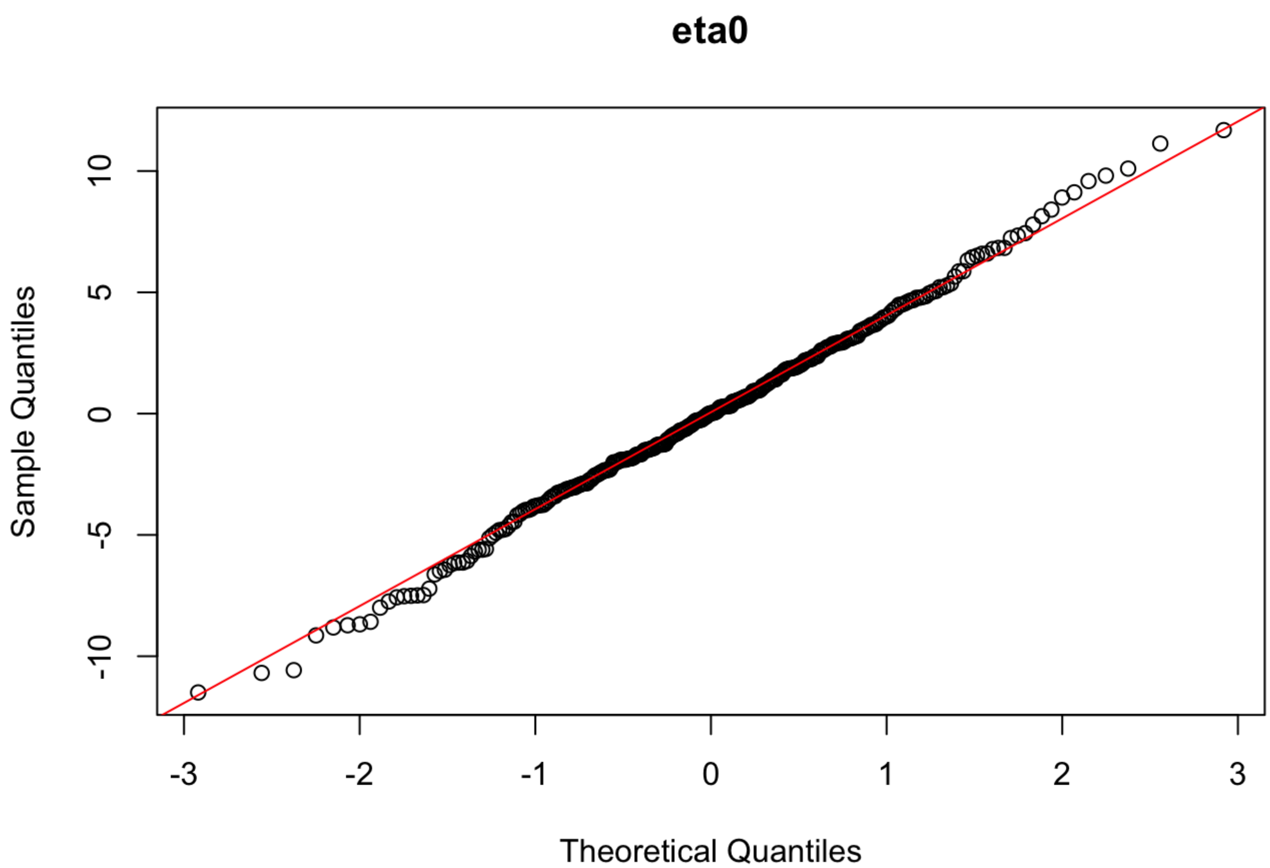
```
ord2 <- order(unlist(tapply(resFE_RE, newdat2$schoolid, median)))
boxplot(split(resFE_RE, newdat2$schoolid)[ord2], main="resFE_RE")
```




```
#examine the two sets of BLUPs (for random effects zeta0 and eta0) for normality

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




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



The new boxplot looks a little flater than the previous boxplot.So there is less correlation within schools now. Both qqplots look overall normal distributed.

Model 2

```
#now reload the data and fit a slightly more complicated model:
#mixed math1st housepov yearstea mathprep mathknow ses sex minority |||schoolid: minorit
y, cov(un) || classid:, reml (THIS IS THE MODEL 2 IN STATA)
M2 <- lmer(math1st~housepov+yearstea+mathprep+mathknow+ses+sex+minority+(minority|school
id)+(1|classid),data=dat2)
print(summary(M2))
```

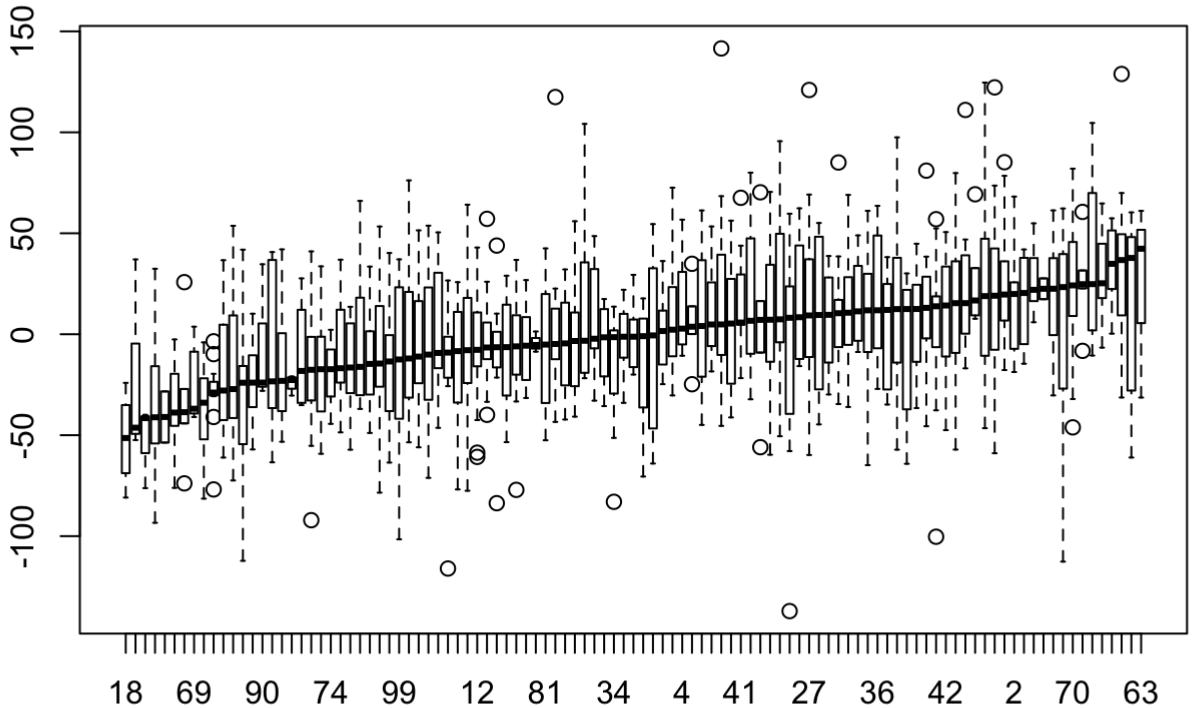
```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex +
##     minority + (minority | schoolid) + (1 | classid)
##    Data: dat2
##
## REML criterion at convergence: 10717.5
##
## Scaled residuals:
##     Min      1Q  Median      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 t value
## (Intercept) 539.493632   5.655135   95.40
## housepov    -16.062540  12.574724   -1.28
## yearstea     -0.004368   0.137649   -0.03
## mathprep     -0.291778   1.335372   -0.22
## mathknow      1.632181   1.359293    1.20
## ses           9.430948   1.543346    6.11
## sex          -0.862802   2.083816   -0.41
## minority    -16.375395   3.896002   -4.20
##
## 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
```

*manually construct the residual that removes only the ‘fixed effects’, call this residual: resFE

```
yhat2<- predict(M2, re.form=~0)
resFE2<- dat2$math1st-yhat2
```

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

```
ord3 <- order(unlist(tapply(resFE2, dat2$schoolid, median)))
boxplot(split(resFE2, dat2$schoolid)[ord3])
```



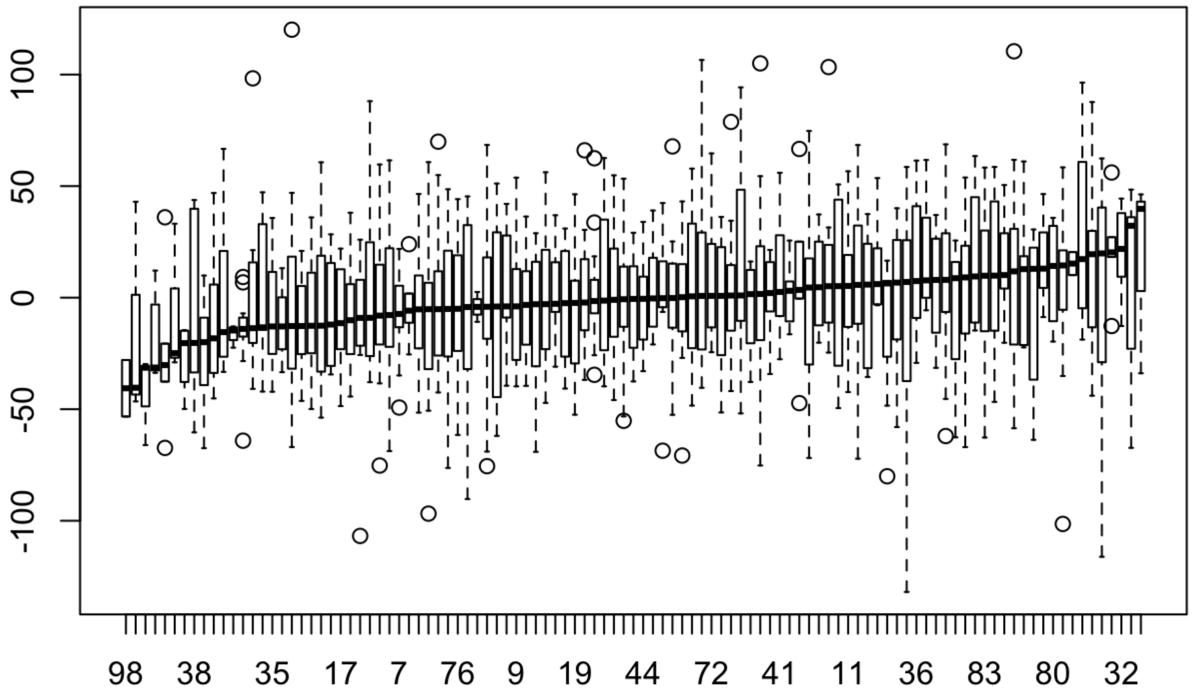

```
#Yes, they are not independent according to the upward trend in this boxplot.
```

*construct the residual that utilizes the BLUPs for the random effects. Do it in these stages:

```
# * i) predict and save zeta0 AND zeta1 (you need to give them in reverse order in STATA
 - ask me why if you want)
zeta0.2<- ranef(M2)$schoolid[,1]
zeta1.2<- ranef(M2)$schoolid[,2]
#* ii) predit and save eta0
eta0.2<-ranef(M2)$classid[,1]
#* iii) generate a new residual, called resFE_RE which subtracts yhat, zeta0, MINORITY*z
eta1 and eta0

zeta0_new.2<- cbind(sch,zeta0.2,zeta1.2)
eta0_new.2<- cbind(classid, eta0.2)

zeta0_new.2<- data.frame(schoolid=zeta0_new.2[,1], zeta0.2 = zeta0.2, zeta1.2=zeta1.2) #
make it to be a data frame

newdat.2<- join(dat2, zeta0_new.2, by="schoolid", type="left", match="all") #merge zeta0
 into master dataset

eta0_new.2<- data.frame(classid=classid, eta0.2= eta0.2) #make it to be a data frame
newdat2.2<- join(newdat.2, eta0_new.2, by="classid", type="left", match="all")


resFE_RE.2 <- newdat2.2$math1st - yhat2 - newdat2.2$zeta0-(newdat2.2$minority)*(newdat2.
2$zeta1.2) - newdat2.2$eta0.2
```

- show that these new residuals, resFE_RE are MUCH LESS (if not completely un-) correlated within school

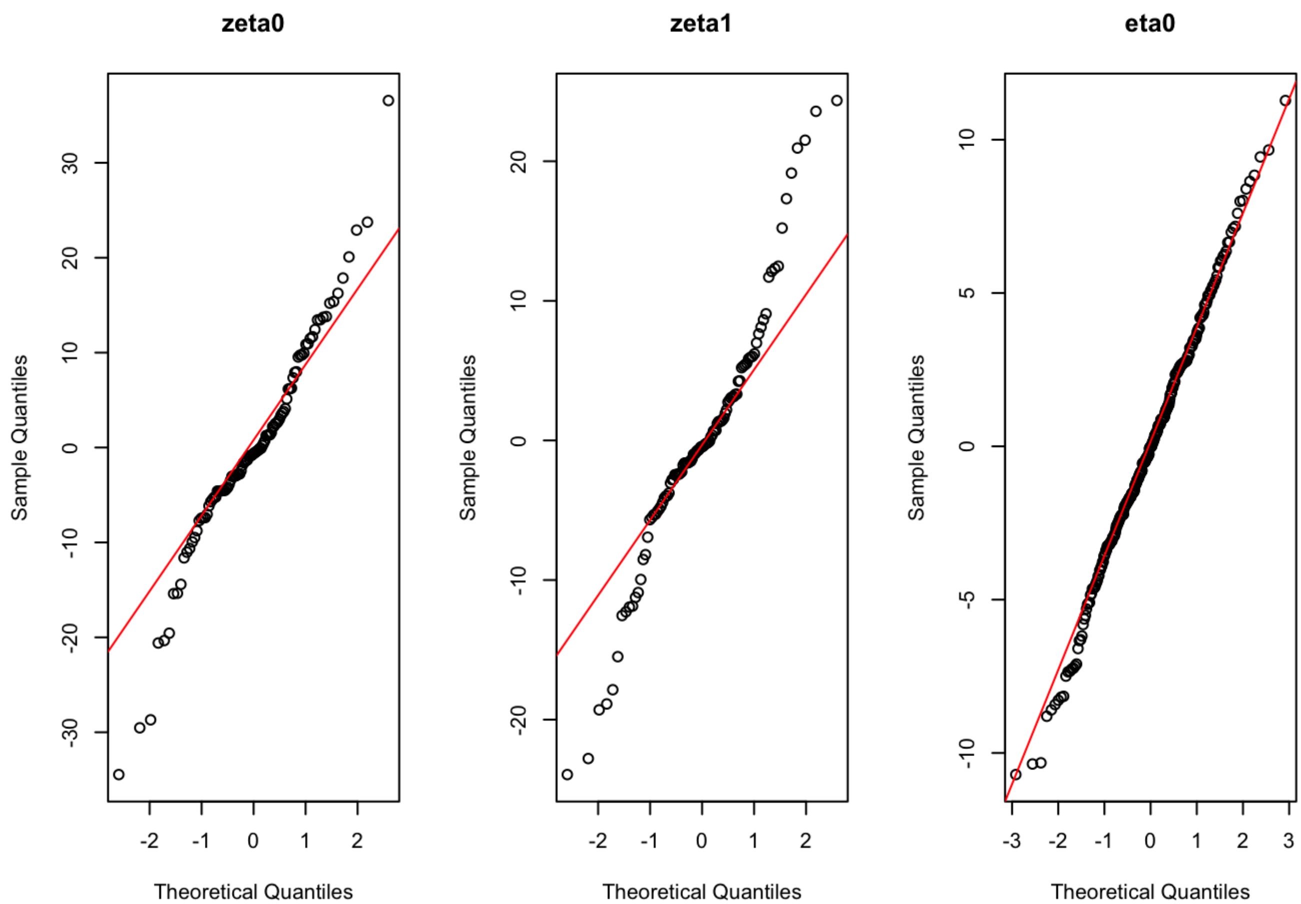
```
ord4 <- order(unlist(tapply(resFE_RE.2, newdat2.2$schoolid, median)))
boxplot(split(resFE_RE.2, newdat2.2$schoolid)[ord4])
```




```
#This boxplot looks only a little flater than the previous boxplot for the same model.
```

*examine the three sets of BLUPs (for random effects zeta0 and eta0) for normality

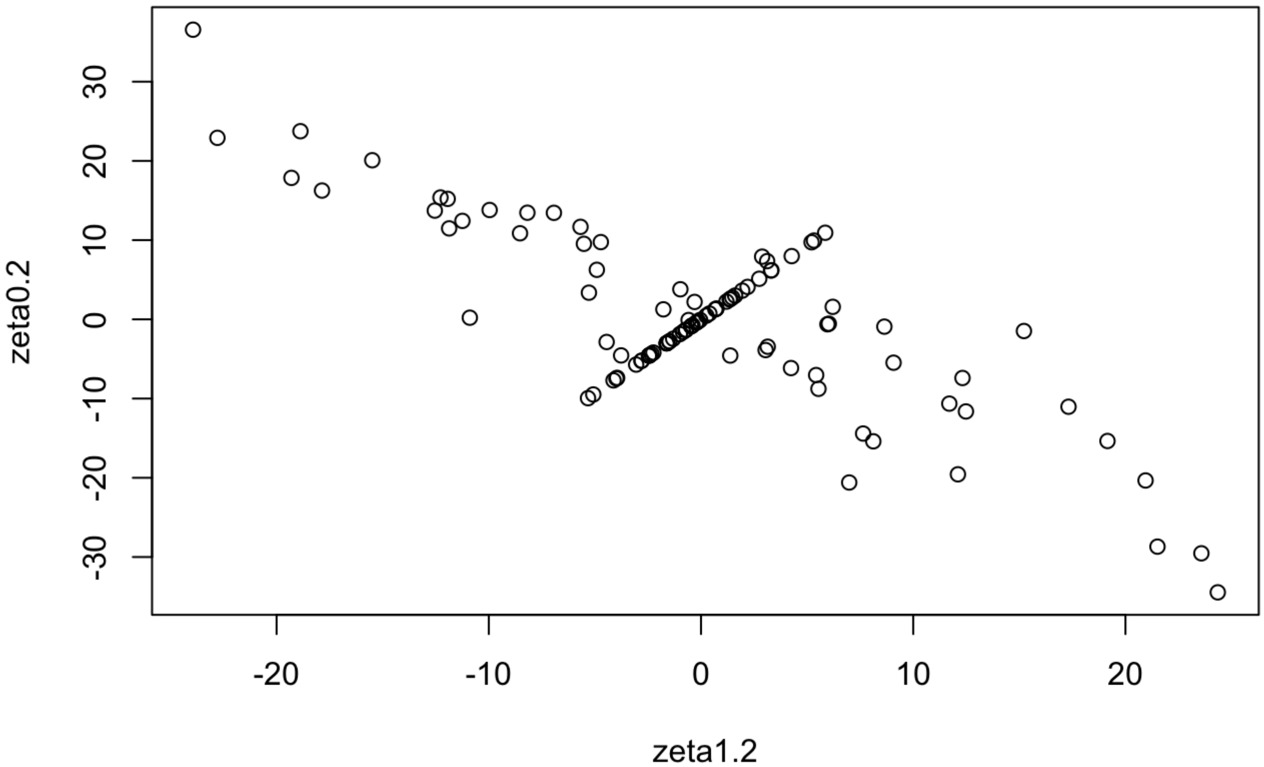
```
par(mfrow=c(1,3))
qqnorm(zeta0.2, main="zeta0")
qqline(zeta0.2, col="red")

qqnorm(zeta1.2, main="zeta1")
qqline(zeta1.2, col="red")

qqnorm(eta0.2, main="eta0")
qqline(eta0.2, col="red")
```




```
#These qqplots look sort of off from normal distribution. It could because of the small
 sample size. Also, zeta0 and zeta1 are correlated in our model. eta0 qqplot looks much
 nicer than other two plots.
```

*plot zeta0 vs. zeta1 to see whether the estimated correlation is consistent with the observed.

```
plot(zeta0.2~zeta1.2)
```



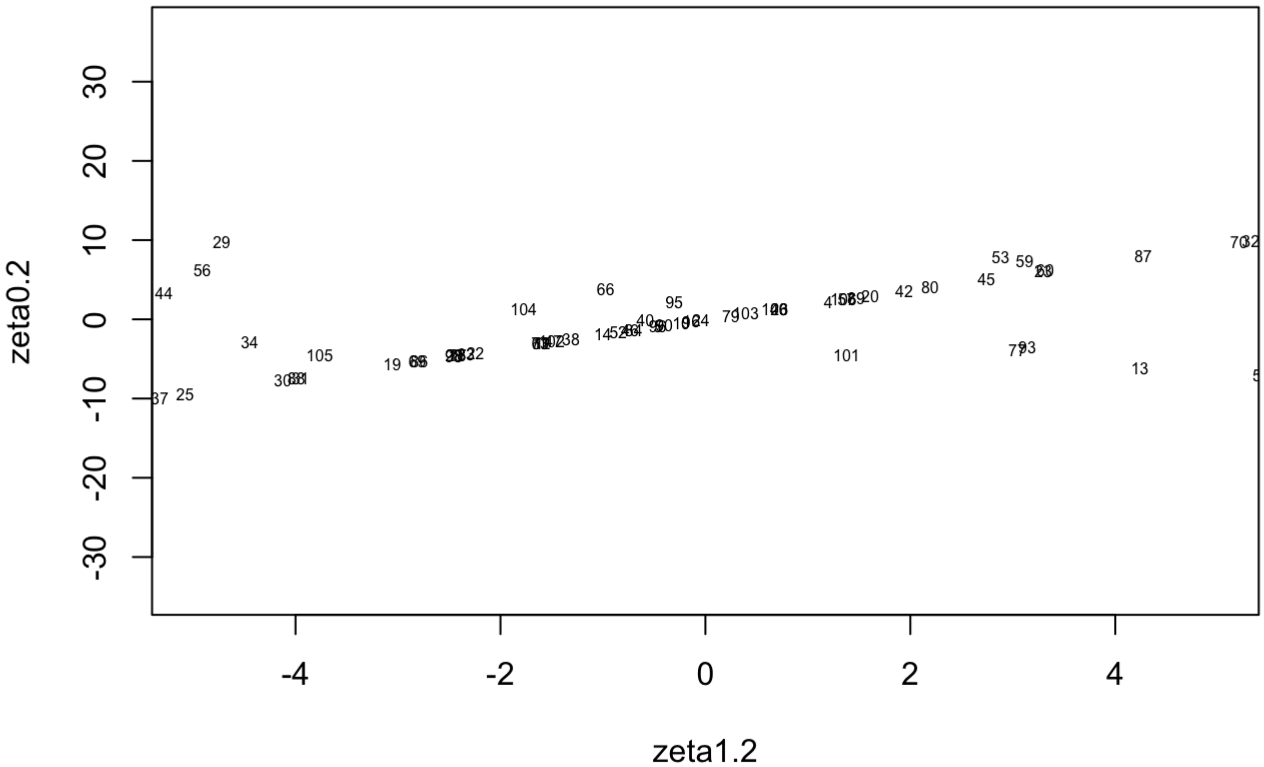

```
#The overall trend of the plot is downward(the slopt is approximate -0.83). However, we
 can see some obvious points between the range of (-5,5) in X-axis looks very upward slo
pe. In the next part, we can label these school IDs for the odd points.
```

*track down those odd points in the scatterplot. What schools are they?

```
#This method can track these odd points using text label;
attach(zeta0_new.2)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     zeta0.2, zeta1.2
```

```
plot(zeta0.2~zeta1.2, xlim=c(-5,5),type='n')
text(zeta0.2~zeta1.2,label = schoolid,cex=.5)
```




```
#We can see these school IDs represent the odd points. By looking closer, I can find out
 that these schools have only one type of minority(either it always equals to 1 or 0), s
uch as school 80, 19, 22, etc.
```

```
```