MLM Project 5

Jemar Bather & Zarni Htet [with Martha Moreno, Sarah Rosenbach, and 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
```

$$MATH1ST_{ijk} = b_0 + b_1 HOUSEPOV_k + b_2 YEARSTEA_{jk} + b_3 MATHPREP_{jk} + b_4 MATHKNOW_{jk} + b_5 SES_{ijk} + b_6 SEX_{ijk} + b_7 MINORITY_{ijk} + \eta_{jk} + \zeta_k + \varepsilon_{ijk}$$

With $\zeta_k \sim N(0,\sigma_\zeta^2)$, $\eta_{jk} \sim N(0,\sigma_\eta^2)$, and $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$, independent of one another.

```
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")
```

resFE 1st BoxPlot
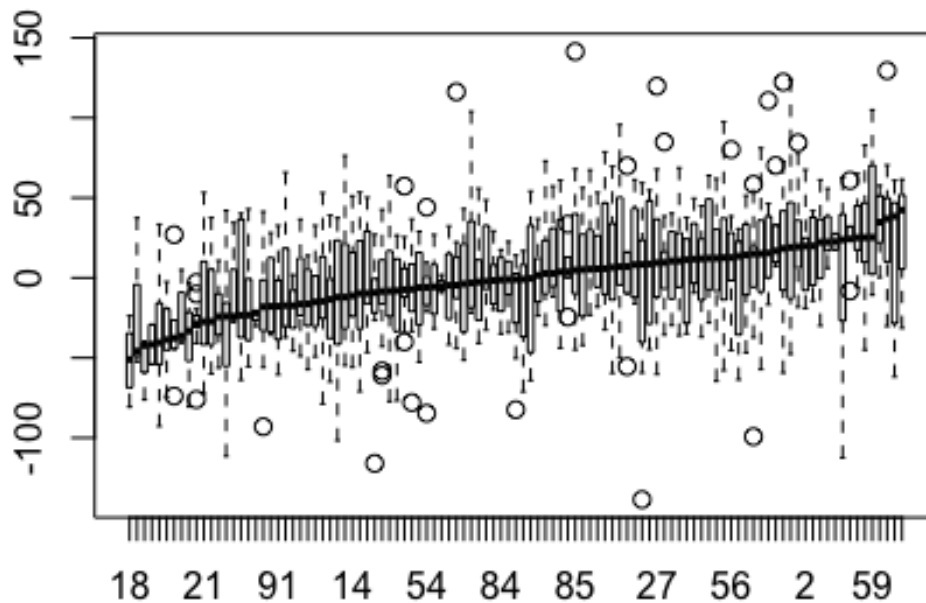
construct the residual that utilizes the BLUPs for the random effects. Do it in these stages: i) predict and save zeta0 * ii) predit and save eta0 * iii) generate a new residual, called resFE_RE which subtracts yhat, zeta0 and eta0 from the outcome *note: there is an easier way to get the residuals in 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$zeta0 <- zeta0[idx.sch] # Putting Random Intercept of School back into the data s
et
cc.data$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 removin
g both fixed and random effects and the random effects are removed manually. Therefore, r
esiduals 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")
```

resFE_RE_BoxPlot 2
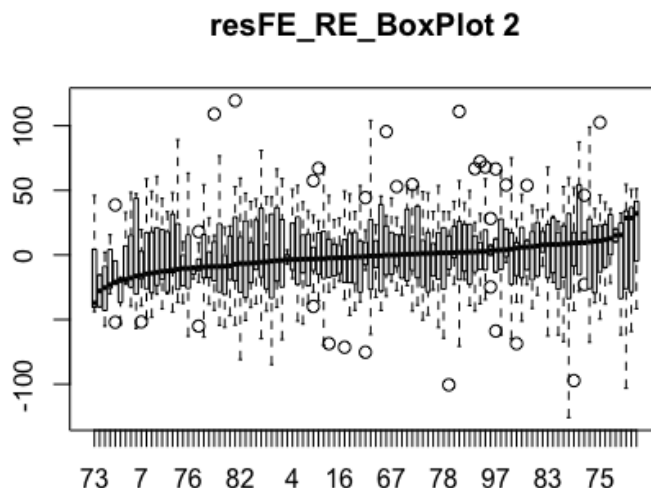
examine the two sets of BLUPs (for random effects zeta0 and eta0) for normality first 'tag' a single value from each grouping (e.g., school or classroom) so that you only have *as many BLUPs as the grouping factor (should be less of an issue in R)

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")
```
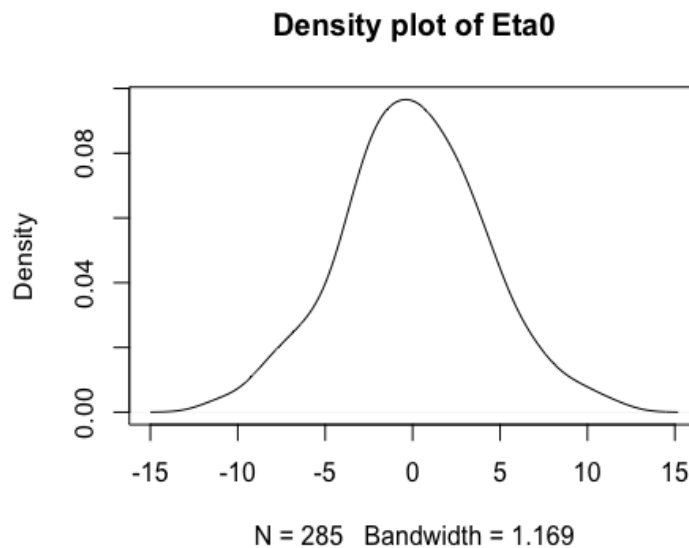
Density Plot of Zeta0

N = 105   Bandwidth = 3.23
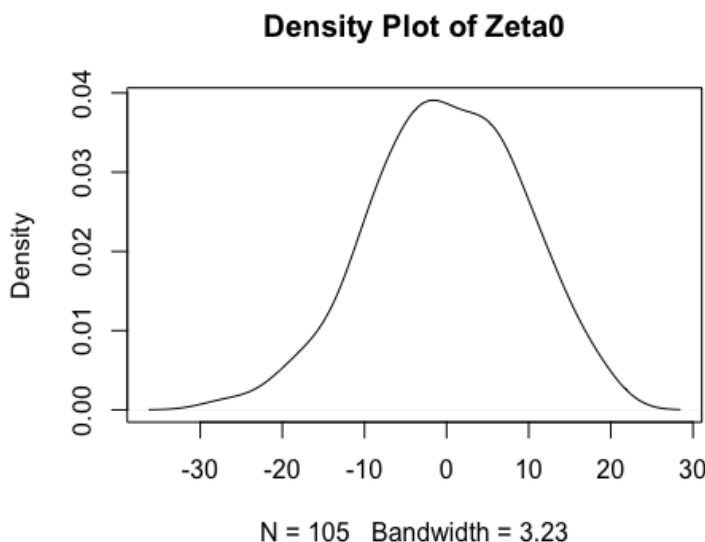
```r
qqnorm(zeta0, main = "zeta0")
qqline(zeta0, col ="red")
```
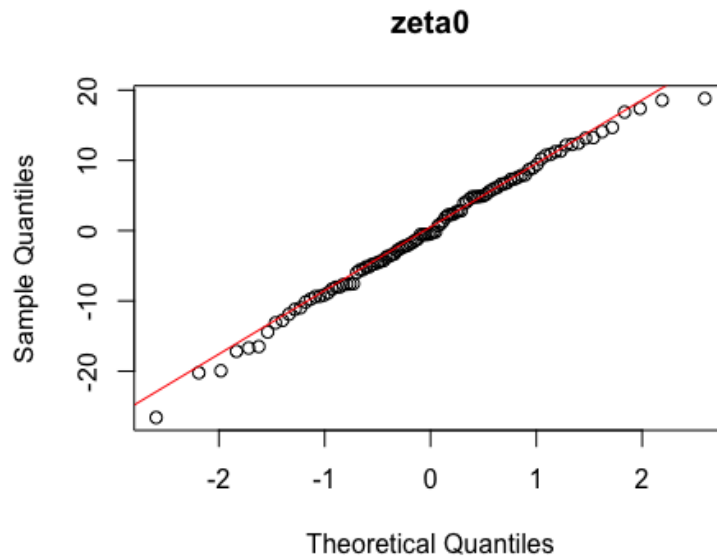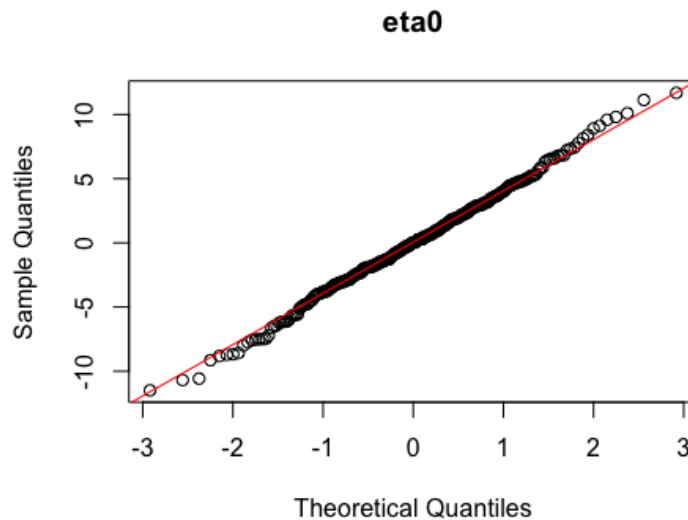
zeta0

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.

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

Density plot of Eta0

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

eta0

*now reload the data and fit a slightly more complicated model: use "classroom.dta" gen math1st = mathkind + mathgain

```r
dat <- read.csv("../classroom.csv")
dat$math1st <- dat$mathkind + dat$mathgain
dat <- dat[(complete.cases(dat)),]
attach(dat)

## The following objects are masked from cc.data (pos = 3):
##
##     childid, classid, housepov, math1st, mathgain, mathkind,
##     mathknow, mathprep, minority, schoolid, ses, sex, yearstea

## The following objects are masked from cc.data (pos = 7):
##
##     childid, classid, housepov, mathgain, mathkind, mathknow,
##     mathprep, minority, schoolid, ses, sex, yearstea
```

$$MATH1ST_{ijk}$$
$$= b_0 + b_1 HOUSEPOV_k + b_2 YEARSTEA_{jk} + b_3 MATHPREP_{jk} + b_4 MATHKNOW_{jk} + b_5 SES_{ijk} + b_6 SEX_{ijk} + b_7 MINORITY_{ijk}$$
$$+ \eta_{jk} + \zeta_k + \zeta_{1k} MINORITY_{ijk} + \varepsilon_{ijk}$$

With $\zeta_k \sim N(0,\sigma_{\zeta_0}^2)$, $\zeta_{1k} \sim N(0,\sigma_{\zeta_1}^2)$, $\eta_{jk} \sim N(0,\sigma_\eta^2)$, and $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$, BUT NOW $corr(\zeta_{0k}, \zeta_{1k}) = \rho_{\zeta_0 \zeta_1}$, which may not be zero, and all other pairs of random terms are independent of one another.

```r
#The model is slightly more complicated with an uncorrelated random slope school level
variability of minority.
M2 <- lmer(math1st ~ housepov + yearstea + mathprep + mathknow + ses + sex+ minority+ (mi
nority|schoolid) + (1|classid))
print(summary(M2))

## 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 + (minority | schoolid) + (1 | classid)
```

```
## 
## 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  1.000e+02  -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  5.820e+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

idx.sch <- match(dat$schoolid, sort(unique(dat$schoolid)))
idx.cls <- match(dat$classid, sort(unique(dat$classid)))
```

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

```
pred.yhat.2 <- predict(M2, re.form = ~0)
resFE.2 <- dat$math1st - pred.yhat.2
```

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 appears to be dependent on school like the results of the less complicated model above.

```
ord3<-order(unlist(tapply(resFE.2,schoolid,median)))
boxplot(split(resFE.2,schoolid)[ord3], main = "resFE.2 Boxplot_3")
```
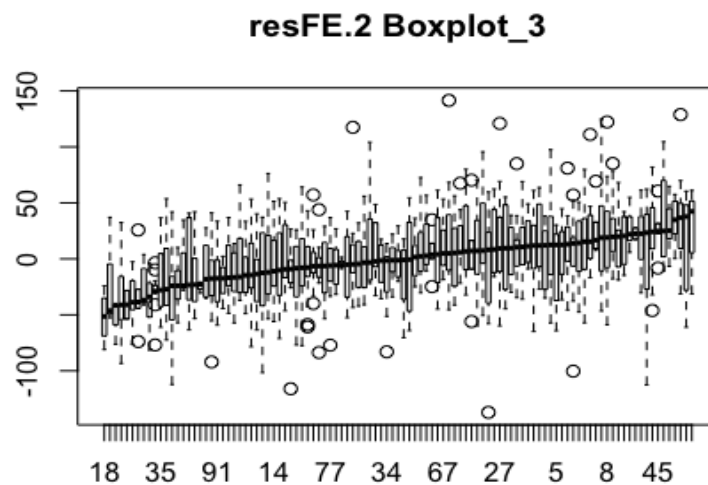
resFE.2 Boxplot_3

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) * ii) predit and save eta0 * iii) generate a new residual, called resFE_RE which subtracts yhat, zeta0, MINORITY*zeta1 and eta0

```
ranerr <- ranef(M2)
zeta0 <- ranerr$schoolid[,1]
zeta1 <- ranerr$schoolid[,2] #It is the random slope of the minority intercept
eta0 <- ranerr$classid[,1] #This is the random intercept for the classroom level.
dat$zeta0 <- zeta0[idx.sch]
dat$zeta1 <- zeta1[idx.sch]
dat$eta0 <- eta0[idx.cls]
resFE_RE.2 <- dat$math1st - (pred.yhat.2 + dat$zeta0 + (dat$minority * dat$zeta1) + dat$e
ta0)#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. Th
us, 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. Additonally, the model also incorporates a correlated random effect on minority at the school level so, we are expecting a much flatter line than the less complicated model before as our total variance explained in this model is higher.

```
ord4<-order(unlist(tapply(resFE_RE.2,schoolid,median)))
boxplot(split(resFE_RE.2,schoolid)[ord4], main = "resFE_RE.2 Boxlot 4")
```

resFE_RE.2 Boxlot 4

73 74 17 99 41 16 84 81 29 87 8
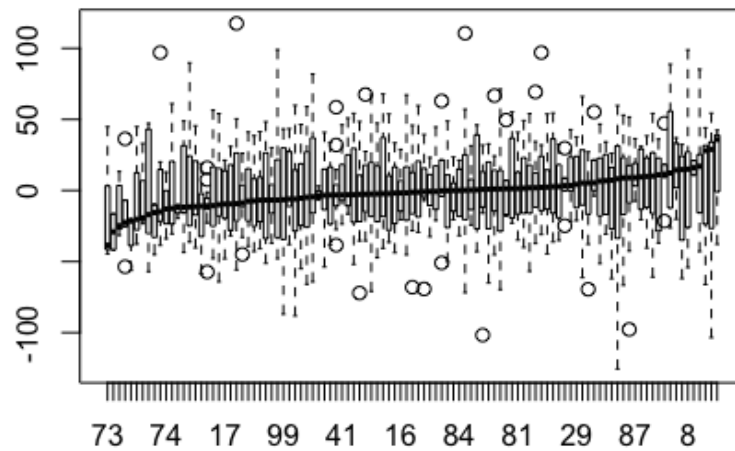
*examine the three sets of BLUPs (for random effects zeta0 and eta0) for normality

If we examine the density and qqplot of the Zeta0 BLUP below, it's not that close to symmetric, and heavy tails on both ends. Even though the sample size is 105, we find it hard to tolerate this lack of normality.
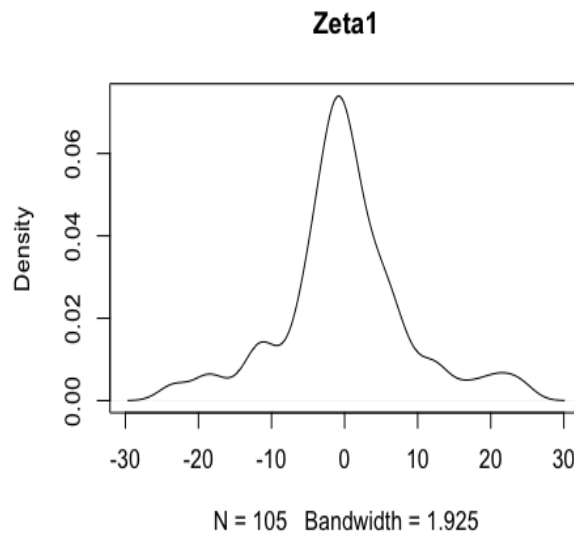
```
plot(density(zeta0), main = "Zeta0")
```

Zeta0

N = 105   Bandwidth = 2.844
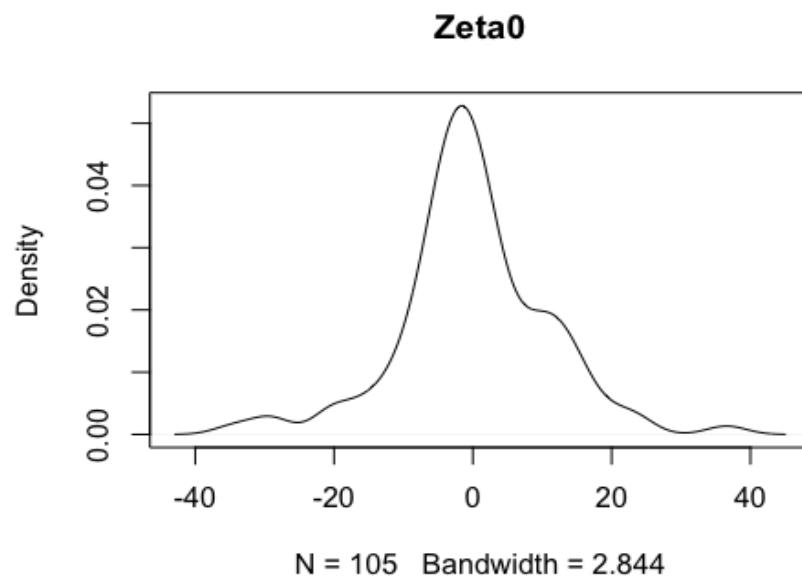
```
qqnorm(zeta0, main = "Zeta0") #This is already tagged by schoolid
qqline(zeta0, col = "red") #This is already tagged by schoolid
```
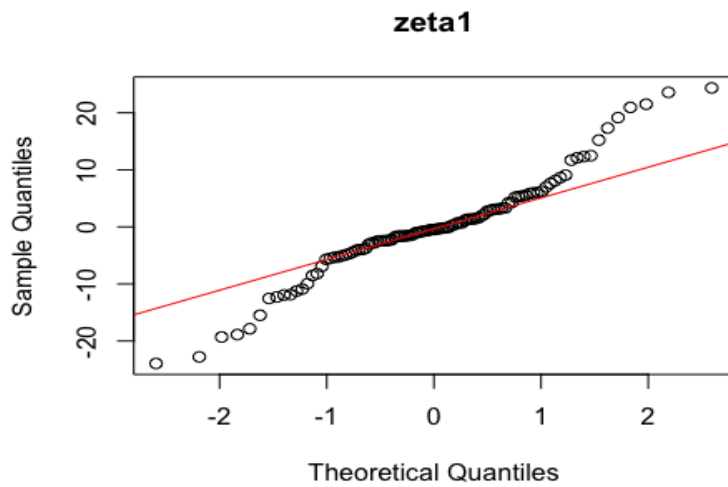
Zeta0
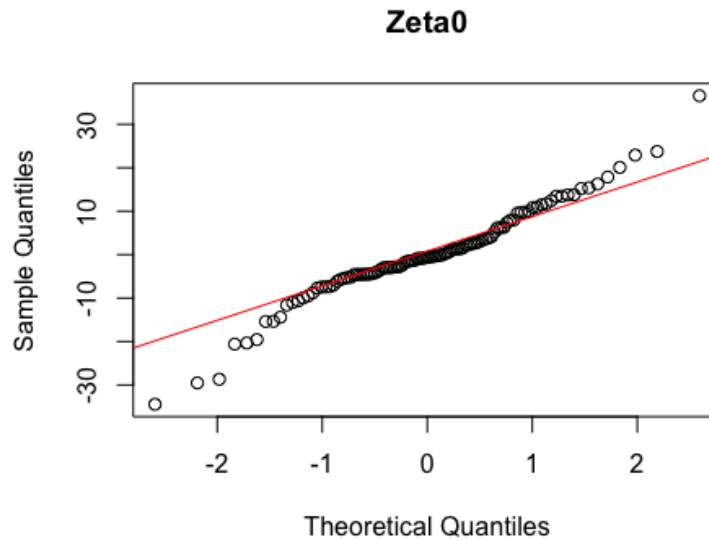

```
plot(density(zeta1), main = "Zeta1")
```
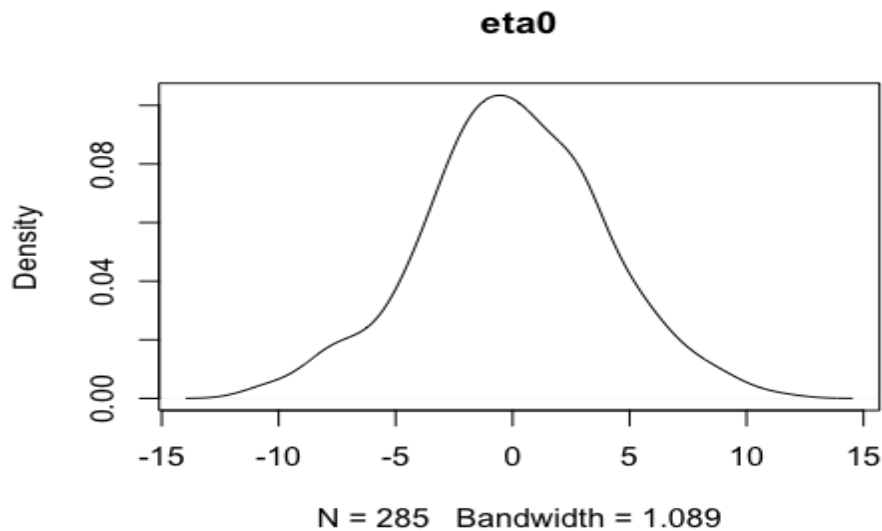
Zeta1

If we examine the density and qqplot of the Zeta1 BLUP below, it's not that close to symmetric, and heavy tails on both ends. Even though the sample size is 105, we find it hard to tolerate this lack of normality.

```
qqnorm(zeta1, main = "zeta1")
qqline(zeta1, col = "red")
```
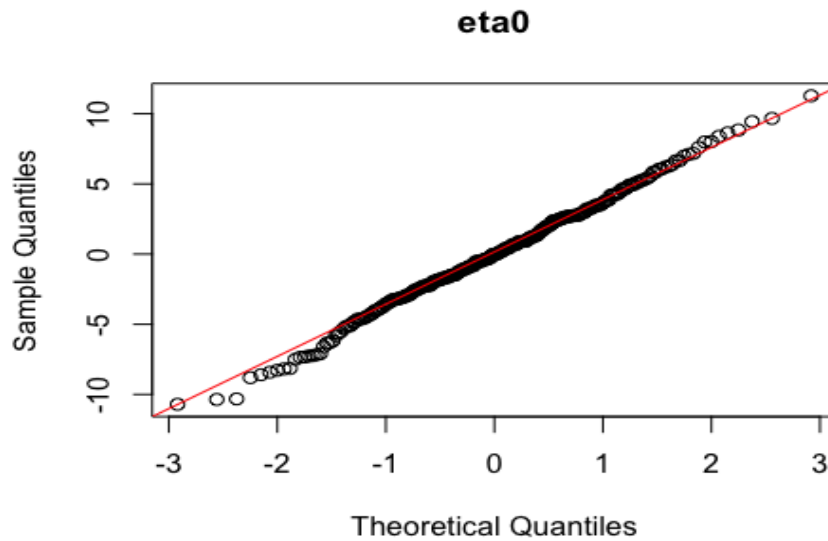
zeta1

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 = "eta0")
```

eta0

N = 285   Bandwidth = 1.089

```
qqnorm(eta0, main = "eta0") #this is already tagged by classid
qqline(eta0, col = "red") #this is already tagged by classid
```

eta0

plot zeta0 vs. zeta1 to see whether the estimated correlation is consistent with the observed. Use tag to subset as before

```
plot(zeta0~zeta1, main = "Zeta 0 and Zeta 1 scatter plot")
```

Zeta 0 and Zeta 1 scatter plot