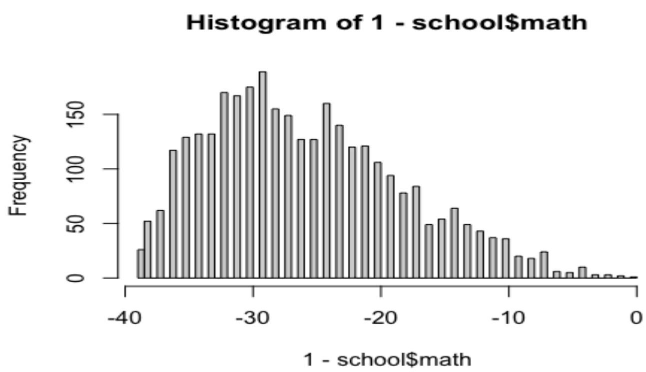
STA442 HW2 Q1

Insu Yoon, 1003180933

Introduction

We analyze the data set that finding what factors influence on math scores."School" is an ordered factor identifying the 49 schools and 4 different classes with number of students and we have social class, gender, grade. First, I assume that the math score's of student depending on which school they attend. So, I would set the prior and posterior about the random effect concerned about the school, class, and students Then, I would compare the sd of each random effects so that I can see which random effects make the biggest difference among school, class, and students.

Method



$Y_{ijk}|U_i \sim Pois(O_i\lambda_i)$

$log(\lambda_i) = X_i\beta$

U_i

Y_{ijk}: $The\ response\ variable, the\ number\ of\ questions\ the\ student$
$gets\ right\ of\ the\ kth\ student\ who\ is\ from\ jth\ classroom\ of\ the\ ith\ school$

X_i: $Covariates(Socialclass, gender, grade)$

$score\ = school\ effect\ +\ class\ effect\ +\ student\ effect\ +\ Covariates$

```
##                          mean     0.5quant   0.025quant  0.975quant
## (Intercept)         2.499168512  2.499144562  2.376670234  2.62168232
## genderm            -0.002430596 -0.002434355 -0.060072100  0.05517721
## socialClassII      -0.174216065 -0.174185223 -0.308812924 -0.03991251
## socialClassIIIn     0.014329630  0.014359076 -0.129760745  0.15812343
## socialClassIIIm     0.137910070  0.137925443  0.017377547  0.25824815
## socialClassIV       0.096046279  0.096074692 -0.046461835  0.23826695
```

```
## socialClassV          0.219076498  0.219099494  0.070978498  0.36691164
## socialClasslongUnemp  0.164384877  0.164422209  0.008452235  0.31996405
## socialClasscurrUnemp  0.170898781  0.170954917 -0.042711820  0.38399436
## socialClassabsent     0.167045670  0.167067202  0.036300077  0.29755319
## grade1               -0.001743735 -0.001743924 -0.023283912  0.01977752
## grade2               -0.420378384 -0.420369899 -0.446519081 -0.39431020
## SD for school         0.009412887  0.008042652  0.003395152  0.02331429
## SD for classUnique    0.180066250  0.179479392  0.135542616  0.22875613
## SD for studentUnique  0.456912500  0.456798097  0.433946450  0.48054001
```

In conclusion, we can find that the individual-level variation is the largest effect with $2\tau \approx 0.9$. Class level effects have similar maginitue as differences between social classes. Therefore, we can conclude that to raise the math score we need to give extra attention to students who perform poor in math class.

Appendix

```
library("Pmisc")
dir.create(file.path("..", "data"), showWarnings = FALSE)
school = read.fwf("../HW2/JSP.DAT", widths = c(2, 1, 1, 1, 2, 4, 2, 2, 1), co
l.names = c("school", "class", "gender", "socialClass", "ravensTest", "studen
t", "english", "math", "year"))
school$socialClass = factor(school$socialClass, labels = c("I", "II", "IIIn",
"IIIm", "IV", "V", "longUnemp", "currUnemp", "absent"))
school$gender = factor(school$gender, labels = c("f", "m"))
school$classUnique = paste(school$school, school$class)
school$studentUnique = paste(school$school, school$class,school$student)
school$grade = factor(school$year)
schoolLme = glmmTMB::glmmTMB(math ~ gender + socialClass + grade + (1 | schoo
l) + (1 | classUnique) + (1 | studentUnique), data = school)
schoolFormula<-(math~gender+socialClass+grade+school+classUnique+studentUniqu
e)
summary(schoolLme)
```

```
##  Family: gaussian  ( identity )
## Formula:
## math ~ gender + socialClass + grade + (1 | school) + (1 | classUnique) +
##     (1 | studentUnique)
```

```
## Data: school
##
##      AIC      BIC   logLik deviance df.resid
##  20218.9  20316.2 -10093.5  20186.9     3220
##
## Random effects:
##
## Conditional model:
##  Groups        Name        Variance  Std.Dev.
##  school        (Intercept) 3.615e-06 0.001901
##  classUnique   (Intercept) 4.520e+00 2.126011
##  studentUnique (Intercept) 3.173e+01 5.632823
##  Residual                  1.436e+01 3.789021
## Number of obs: 3236, groups:  school, 49; classUnique, 94; studentUnique,
1192
##
## Dispersion estimate for gaussian family (sigma^2): 14.4
##
## Conditional model:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           27.83238    1.14474  24.313  < 2e-16 ***
## genderm               -0.17138    0.36730  -0.467 0.640790
## socialClassII          0.07457    1.21075   0.062 0.950891
## socialClassIIIn       -1.73338    1.27328  -1.361 0.173404
## socialClassIIIm       -3.19815    1.14590  -2.791 0.005256 **
## socialClassIV         -2.76230    1.26486  -2.184 0.028971 *
## socialClassV          -4.77784    1.30756  -3.654 0.000258 ***
## socialClasslongUnemp -3.91251    1.35566  -2.886 0.003901 **
## socialClasscurrUnemp -4.68072    1.81660  -2.577 0.009977 **
## socialClassabsent     -3.55569    1.19894  -2.966 0.003020 **
## grade1                 0.02234    0.16037   0.139 0.889206
## grade2                 4.99848    0.17050  29.316  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
hist(1 - school$math, breaks = 100)

school$nwrong = 40- school$math
fResP = inla(nwrong~gender+socialClass+grade+f(school,model = "iid")+f(classU
nique, model = "iid")+f(studentUnique, model="iid"), data= school, family = '
poisson', control.fixed= list(
   mean=0, mean.intercept=0,
   prec = 0.2^(-2), prec.intercept = 10^(-2)))

rbind(fResP$summary.fixed[, c('mean','0.5quant','0.025quant','0.975quant')],
    Pmisc::priorPostSd(fResP)$summary[, c('mean','0.5quant','0.025quant','0.9
75quant')])
```