Stat 506 Assignment 8

Leslie Gains-Germain

1. Fit the data in R using MathAch as the response, with fixed effects for Minority, Sex, and SES and the strongest of the two-way interactions. Include random effects for School and investigate whether or not there is a random Minority, Sex, and or SES effect between schools.

   The strongest two way interaction is between socioeconomic status (SES) and minority. I included a random intercept for school and investigated whether to include a random slope for SES, Sex or Minority. There was weak to moderate evidence of random sex and SES effects (p-values= 0.045 *and* 0.08454 *respectively). There was strong evidence of a random minority effect across schools (p-value*< 0.0001*), so I chose to include a random slope for minority.*

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mixed	7	46387.26	46435.42	-23186.63	46373.26			
slopeMinority	9	46366.28	46428.20	-23174.14	46348.28	24.98	2	0.0000

2. Use the model from no. 1 with random intercept and the strongest one other random term for each school. Extract the random effects and merge it (by school) with the MathAchSchool dataset. Fit each random effect column in turn to the other variables. HIMINTY indicates whether or not a school is classified as having a high proportion of minority students. The other variables are explained in the help page for the dataset. Explain (briefly at this point) what you find.

   There is strong evidence of a linear relationship between the random intercept mong schools and the size of the school (p-value= 0.00024*). There is also strong evidence of a linear relationship between the random intercept and the percentage of students on the academic track (p-value=* 0.000243*). The random intercept for a school is the estimated math achievement score for non-minority, male students with SES=0.*

   There is strong evidence of a linear relationship between the random minority effect among schools and the discrimination climate (p-value= 0.00051*). There is also strong evidence of a linear relationship between the random minority effect among schools and the mean socioeconomic status (p-value=* 0.006901*).*

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.0418	0.3460	-5.90	0.0000
Size	0.0006	0.0002	3.76	0.0002
SectorCatholic	0.6670	0.2795	2.39	0.0182
PRACAD	2.1400	0.5694	3.76	0.0002
DISCLIM	-0.2257	0.1262	-1.79	0.0758
HIMINTY1	0.0709	0.2169	0.33	0.7441
MEANSES	0.6442	0.3040	2.12	0.0357

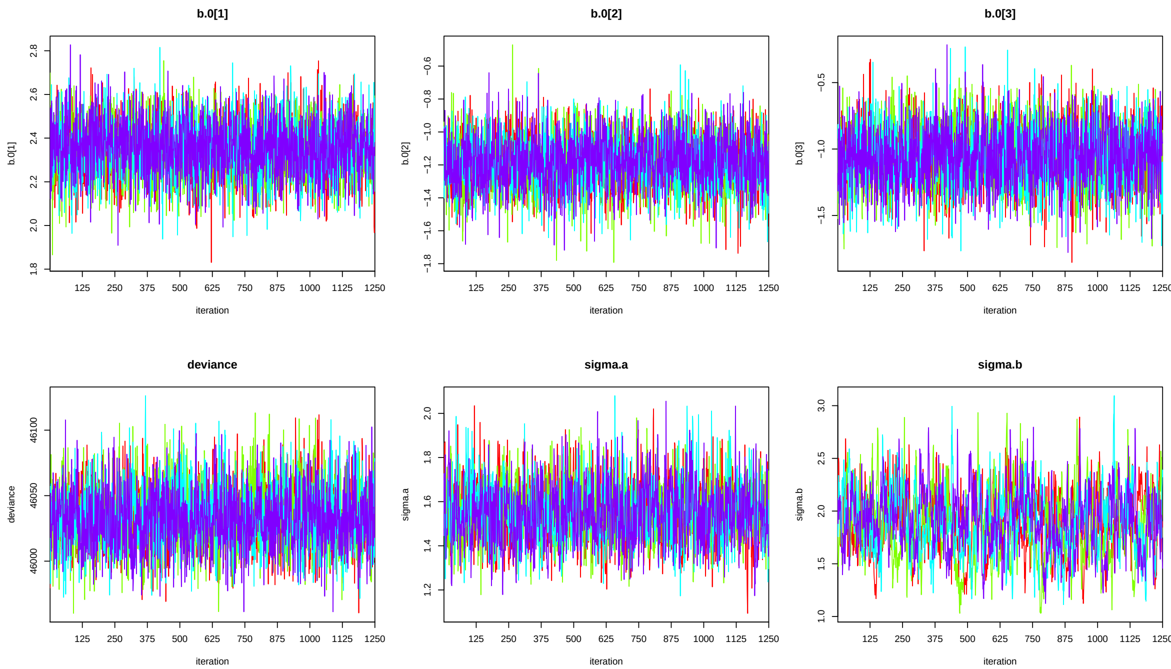
3. Pick a good model from your work above and fit it using jags. Provide enough information to show that it has converged, and compare output to that from the fit above.

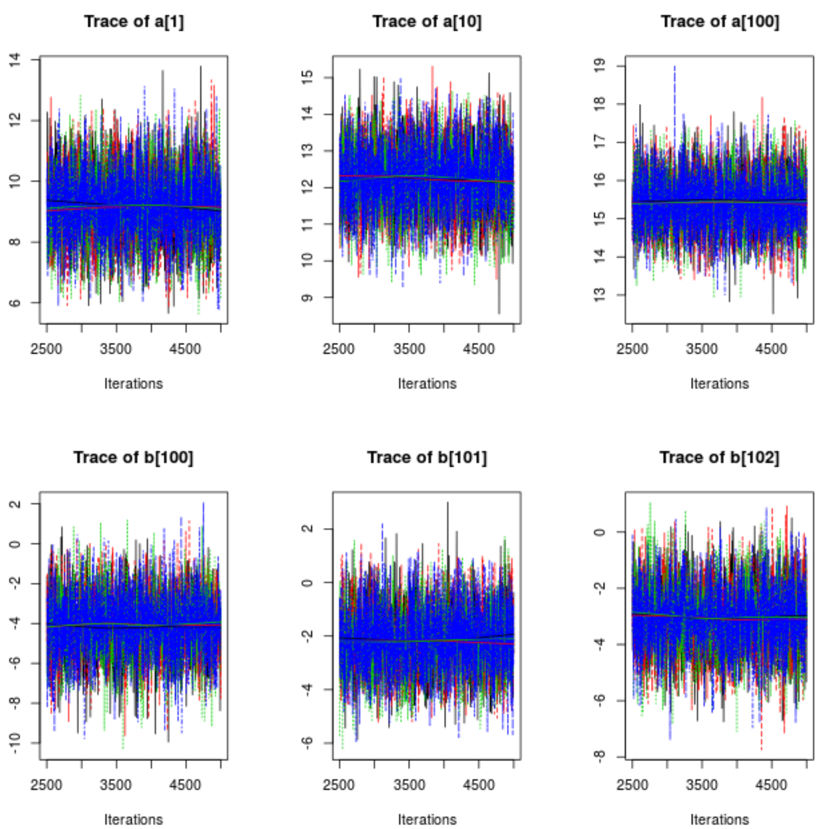
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.7473	0.2431	-3.07	0.0025
Size	0.0001	0.0001	1.31	0.1933
SectorCatholic	0.4725	0.1963	2.41	0.0173
PRACAD	0.6357	0.4000	1.59	0.1140
DISCLIM	-0.3148	0.0887	-3.55	0.0005
HIMINTY1	0.2040	0.1523	1.34	0.1825
MEANSES	0.5848	0.2135	2.74	0.0069

I fit the following model in JAGs. I included fixed effects for sex, minority, socioeconomic status, and a minority by socioeconomic status interaction. I included a random intercept and a random minority effect for school. I then included size and percentage on the academic track (PRACAD) as group level predictors for the random intercept and discrimination climate (DISCLIM) and mean socioeconomic status (MEANSES) as group level predictors for the random slope.

```
setwd("~/Documents/Stat506/Homework/HW8")

##write model file first

cat("
model{
  for(i in 1:n){
  y[i] ~ dnorm(y.hat[i], tau.y)
  y.hat[i] <- inprod(b.0[], X.0[i,])+a[school[i]] + b[school[i]]*x[i]
  }
tau.y <-pow(sigma.y, -2)
sigma.y ~ dunif(0, 100)

for(k in 1:K){
  b.0[k]~dnorm(0,0.0001)
}
for(j in 1:J){
a[j] ~ dnorm(a.hat[j], tau.a)
a.hat[j] <- g.0+g.1*size[j]+g.2*p[j]
b[j] ~ dnorm(b.hat[j], tau.b)
b.hat[j] <- g.b0+g.b1*d[j]+g.b2*ses[j]
}
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif(0, 100)
tau.b <- pow(sigma.b, -2)
sigma.b ~ dunif(0, 100)
g.0 ~ dnorm(0, 0.0001)
g.1 ~ dnorm(0, 0.0001)
g.2 ~ dnorm(0, 0.0001)
g.b0 ~ dnorm(0, 0.0001)
g.b1 ~ dnorm(0, 0.0001)
g.b2 ~ dnorm(0, 0.0001)
}", file="jagsmath.jags")
```

First I show traceplots for the fixed effects and variance parameters, as well as a few of the random slopes and intercepts. All of the traceplots look good. Initially it took me a while to figure out how to get it to converge. The trick that worked, in the end, was to take out the intercept and slope in the model matrix of fixed effects.





Below I show the convergence criteria for all fixed effects and variance parameters, as well as the random effects for schools 1 *and* 10. *The next table shows a summary of all* $\hat{R}$ *values and effective sample sizes. All* $\hat{R}$*'s were below* 1.01, *and all effective sample sizes were larger than* 510.

	mean	sd	2.5%	97.5%	Rhat	n.eff
a[1]	9.20	1.16	6.91	11.45	1.00	5000.00
a[10]	12.25	0.91	10.45	14.06	1.00	4900.00
b[1]	-4.04	1.79	-7.65	-0.70	1.00	2200.00
b[10]	-4.68	1.77	-8.28	-1.37	1.00	1700.00
b.0[1]	2.35	0.13	2.09	2.60	1.00	4400.00
b.0[2]	-1.19	0.16	-1.50	-0.88	1.00	2300.00
b.0[3]	-1.05	0.22	-1.49	-0.62	1.00	5000.00
deviance	46033.37	23.31	45989.14	46082.22	1.00	1800.00
sigma.a	1.55	0.13	1.31	1.83	1.00	4400.00
sigma.b	1.88	0.30	1.32	2.47	1.01	510.00

min	Q1	median	Q3	max	mean	sd	n	missing
1.00	1.00	1.00	1.00	1.01	1.00	0.00	326	0

min	Q1	median	Q3	max	mean	sd	n	missing
510.00	3200.00	5000.00	5000.00	5000.00	4081.66	1250.31	326	0

The overall y-intercept, averaged over all schools, is estimated to be 13.97 *(se=*0.128*), and the overall minority effect is estimated at* -3.10 *(*0.235*) from the JAGs model. The lmer model summary is shown below and is similar. In the JAGs model, the estimates for the SES, SexFem, and SES:Minority effects are* 2.35 *(*0.13*),* -1.19 *(*0.16*), and* -1.05 *(*0.22*) respectively. The estimates and standard errors of these fixed effects are very similar in the lmer model. The estimated standard deviation of the random intercepts is* 1.55 *in the JAGs model and* 1.70 *in the lmer model. The estimated standard deviation of the random slopes is* 1.88 *in the JAGs model and* 1.63 *in the lmer model.*

No correlation was built into the JAGs model a priori, but the correlation of the intercept and slope column in the output is 0.256*, compared to the estimated correlation of* 0.42 *in the lmer model.*

	Estimate	Std. Error	t value
(Intercept)	14.08	0.18	76.53
MinorityYes	-3.21	0.26	-12.28
SexFemale	-1.23	0.16	-7.58
SES	2.40	0.13	19.18
MinorityYes:SES	-1.13	0.22	-5.14

4. Provide a complete summary and explain what weve learned about math achievement scores. The writeup is worth one-third of the points on this assignment, so spend some time polishing it as a report to a legislative committee interested in factors that affect student math achievement scores.

Over all schools, the true mean math achievement score of minority students with a socioeconomic status of 0 *is estimated to be* 3.10 *points lower than the score of non-minority students with a socioeconomic status of* 0*, after controlling for sex, with a* 95% *credible interval from* 2.64 *to* 3.57 *points lower. The school to school variation in the* *minority effect is estimated to be* 1.88^2, *with a* 95% *credible interval from* 1.32^2 *to* 2.47^2.

The mean math achievement score for non minority, male students with a socioeconomic status of 0 *is estimated to be* 13.97 *averaged over all schools, with a* 95% *confidence interval from* 13.72 *to* 14.22. *The school to school variation in the intercept is estimated to be* 1.55^2, *with a* 95% *credible interval from* 1.31 *to* 1.83. *The correlation between the intercept and the minority effect is estimated to be* 0.256 *so that schools with higher starting math achievement scores are estimated to have less negative minority effects. In other words, the minorities aren't as far behind non-minority students at the better schools.*

The true mean math achievement score of female students is estimated to be 1.19 *points lower than the score of male students after controlling for minority and socioeconomic status, with a* 95% *credible interval from* 1.50 *to* 0.88 *points lower.*

For a one point increase on the socioeconomic scale, the true mean math achievement score for non-minority students is estimated to increase by 2.35 *points after controlling for sex, with a* 95% *credible interval from* 2.09 *to* 2.60 *points.*

The socioeconomic effect is estimated to be less extreme for minorities. For a one point increase on the socioeconomic scale, the true mean math achievement score for minority students is estimated to increase by 1.30 *points after controlling for sex, with a* 95% *credible interval from* 0.60 *to* 1.98 *points.*

```
require(nlme)
data(MathAchieve)
math <- MathAchieve
data(MathAchSchool)
school <- MathAchSchool
```

```
require(lme4)
mixed <- lmer(MathAch~Minority+Sex+SES+SES*Minority+(1|School), data=math)

slopeSES <- lmer(MathAch~Minority+Sex+SES+SES*Minority+(1+SES|School), data=math)

slopeSex <- lmer(MathAch~Minority+Sex+SES+SES*Minority+(1+Sex|School), data=math)


###strongest slope effect for minority
slopeMinority <- lmer(MathAch~Minority+Sex+SES+SES*Minority+(1+Minority|School), data=math)

require(xtable)
xtable(anova(mixed, slopeMinority))
```

```
random <- ranef(slopeMinority)$School
random <- cbind.data.frame(rownames(random), random)
names(random) <- c("School", "intercept", "minority")
require(dplyr)
join <- left_join(school, random)
```

```r
require(xtable)
#fit each random effect to the other variables
fit.effect <- lm(intercept~Size+Sector+PRACAD+DISCLIM+HIMINTY+MEANSES, data=join)
#PRACAD percentage of students on academic track
#DISCLIM numeric vector measuring discrimination climate
#MEANSES mean SES score(socioeconomic status)
#HIMINTY indicates whether or not a school is classified as having a high proportion of minority students
#Size of school and PRACAD
xtable(summary(fit.effect))
fit.slope <- lm(minority~Size+Sector+PRACAD+DISCLIM+HIMINTY+MEANSES, data=join)
#DISCLIM, MEANSES, maybe sectorCatholic
xtable(summary(fit.slope))
```

```r
matrix <- left_join(math, school)
```

```r
setwd("~/Documents/Stat506/Homework/HW8")

##write model file first

cat("
model{
  for(i in 1:n){
  y[i] ~ dnorm(y.hat[i], tau.y)
  y.hat[i] <- inprod(b.0[], X.0[i,])+a[school[i]] + b[school[i]]*x[i]
  }
tau.y <-pow(sigma.y, -2)
sigma.y ~ dunif(0, 100)

for(k in 1:K){
  b.0[k]~dnorm(0,0.0001)
}
for(j in 1:J){
a[j] ~ dnorm(a.hat[j], tau.a)
a.hat[j] <- g.0+g.1*size[j]+g.2*p[j]
b[j] ~ dnorm(b.hat[j], tau.b)
b.hat[j] <- g.b0+g.b1*d[j]+g.b2*ses[j]
}
tau.a <- pow(sigma.a, -2)
sigma.a ~ dunif(0, 100)
tau.b <- pow(sigma.b, -2)
sigma.b ~ dunif(0, 100)
g.0 ~ dnorm(0, 0.0001)
g.1 ~ dnorm(0, 0.0001)
g.2 ~ dnorm(0, 0.0001)
g.b0 ~ dnorm(0, 0.0001)
g.b1 ~ dnorm(0, 0.0001)
g.b2 ~ dnorm(0, 0.0001)
}", file="jagsmath.jags")
```

```r
library(R2jags)
set.seed(52)

matrix$MinorityYes <- with(matrix,ifelse(Minority=="No",0,1))
matrix$SexFem <- with(matrix,ifelse(Sex=="Male",0,1))

X.0 <- with(matrix, model.matrix(~SES*MinorityYes+SexFem))[,-c(1,3)]

math.data <-with(matrix,
                 list(n=nrow(matrix), J=length(unique(School)), K=3,
                      school=as.integer(School), y=MathAch,
                      x=MinorityYes,
                      X.0=X.0,
                      size=Size,
                      p=PRACAD, d=DISCLIM, ses=MEANSES))
```

```r
math.param <- c("b.0","a","b", "sigma.a", "sigma.b")

math.model <-  jags(data = math.data, parameters.to.save=math.param,
                   model.file = "jagsmath.jags", n.chains=4, n.iter=5000)
```

```r
require(R2jags)
traceplot(math.model, mfrow=c(1,3), varname=c("b.0"))
#traceplot(withCorM2, mfrow=c(1,2), varname=c("B"))
traceplot(math.model, mfrow=c(1,3), varname=c("deviance","sigma.a", "sigma.b"))
```

```r
require(xtable)
print(xtable(math.model$BUGSoutput$summary[c(1,10,161,170,321:326),c(1:3,7:9)]), table.placement = getOption("xtable.table.place
```

```r
require(mosaic)
xtable(favstats(math.model$BUGSoutput$summary[,8]))
xtable(favstats(math.model$BUGSoutput$summary[,9]))

##get overall estimated y intercept
#mean(math.model£BUGSoutput£summary[1:160,1])
#Now get overall estimated slope
#mean(math.model£BUGSoutput£summary[161:320,2])

#Now let's find estimated overall intercept and standard error
x <- apply(math.model$BUGSoutput$sims.list$a, 1, mean)
#summary(x)
#sd(x)

#now  let's find estimated overall slope and standard error
y <- apply(math.model$BUGSoutput$sims.list$b, 1, mean)
#summary(y)
#sd(y)
#find 95\% credible intervals
require(ffbase)
#quantile(y, probs=c(0.025,0.975))
#quantile(x, probs=c(0.025,0.975))
```

```r
xtable(summary(slopeMinority)$coef)
```