STA 610L: Module 2.14

Random effects ANCOVA (holistic analysis)

Dr. Olanrewaju Michael Akande

STA 610L

Getting more serious about NELS

Until now, we have used the NELS data to illustrate different aspects of model fitting for the multilevel model.

Now let's step back and think about these data more holistically, as if we're seeing them for the first time.

NELS VARIABLES

We will consider the following variables of interest in NELS:

- Math score (individual-level outcome)
- SES (individual-level socio-economic status)
- FLP (school level % of kids eligible for free or reduced-price lunch)
  - 1: 0-5% eligible
  - 2: 5-30% eligible
  - 3: >30% eligible
- Enrollment (school level # of kids in 10th grade, rounded and measured in hundreds, so 0=<100, 1=around 100, ..., 5=around 500)
- Public (school level, takes value 1 if public school and 0 if private school)
- Urbanicity (school level factor with levels rural, suburban, and urban)

Model selection

As we think about models, we'll keep in mind a couple of methods for comparison.

- Likelihood ratio test for nested models
  - For tests involving fixed effects only, we can use a χ^2_d for testing whether d fixed effects all equal 0 (ML, not ok for REML)
  - For tests involving random effects only, we can use a 50-50 mixture of χ^2_{p-1} and χ^2_p, where p is the number of random effect variances in the larger model
  - Non-nested models or testing both fixed and random effects, not so simple.

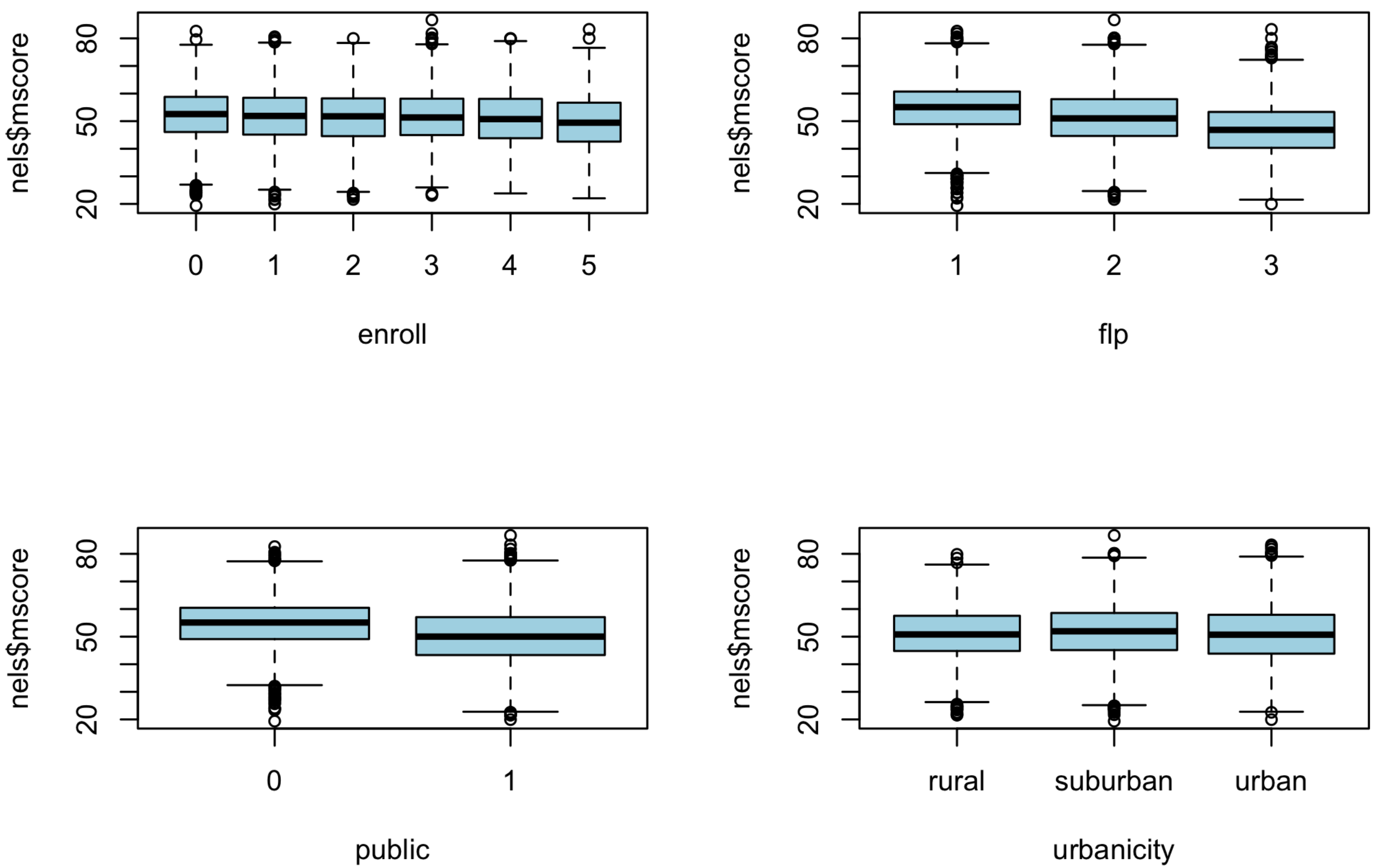
STA 610L

Model selection

- BIC
  - smaller-is-better coding
  - already adjusted for model complexity
  - approximation to posterior model probability
  - model selection consistent
  - nested models not required

STA 610L

Descriptive statistics



STA 610L

What's wrong with ANOVA?

Suppose I don't really care about school effects one way or the other. Why not just use ANOVA (or other fixed effects model) here?

Under a fixed effects model,

$$\text{Cov}(y_j) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

STA 610L

What's wrong with ANOVA?

Under a random intercepts model,

$$\text{Cov}(y_j) = \begin{pmatrix} \sigma^2+\tau^2 & \tau^2 & \dots & \tau^2 \\ \tau^2 & \sigma^2+\tau^2 & \dots & \tau^2 \\ \vdots & & & \vdots \\ \tau^2 & \tau^2 & \dots & \sigma^2+\tau^2 \end{pmatrix},$$

and

$Corr(y_{ij}, y_{i'j}) = \frac{\tau^2}{\tau^2+\sigma^2}$

We generally don't believe independence within the same school environment holds.

This type of covariance structure is often called *exchangeable* or *compound symmetric*.

STA 610L

Other considerations

Why not treat school as a fixed effect? That should handle the school heterogeneity.

```
m10 <- lm(mscore~school+as.factor(enroll)+as.factor(flp)+as.factor(public)+
         urbanicity, data=nels)
#summary(m10)
coef(m10)[(length(coef(m10))-30):length(coef(m10))]
```

STA 610L

OTHER CONSIDERATIONS

```
##         school4513         school4521         school4522         school4531
##           1.771737           3.085000           4.330590          -5.556333
##         school4532         school4541         school4542         school4551
##           1.619069           1.912625           4.158000           1.240200
##         school4552         school4553         school4561         school4562
##           2.027769           7.574857           8.552385           1.357000
##         school4571         school4572         school4582         school4591
##          -3.348000           4.821000           9.443250           6.169727
##         school4592         school4601         school4602         school4611
##          12.405182         -13.559667           3.622333           5.820846
##         school4612 as.factor(enroll)1 as.factor(enroll)2 as.factor(enroll)3
##          -7.980692                 NA                 NA                 NA
## as.factor(enroll)4 as.factor(enroll)5    as.factor(flp)2    as.factor(flp)3
##                 NA                 NA                 NA                 NA
## as.factor(public)1 urbanicitysuburban    urbanicityurban
##                 NA                 NA                 NA
```

What happened to the estimates for enrollment, eligibility for free lunch, public/private status, and urbanicity?

STA 610L

Other considerations

The school-specific fixed effects explain approximately *all* heterogeneity in means across schools, leaving basically no room for the other factors (which we care more about in terms of learning about patterns in the data) to explain any heterogeneity.

So this approach does not allow us to evaluate school-level predictors, and it is also very expensive in terms of spending degrees of freedom (estimating a lot of parameters).

This is a relatively common phenomenon when dealing with categorical group-level predictors.

Heterogeneity across schools

Let's take a more detailed look at the heterogeneity across schools and how much of that can be explained by measured school-level factors including urbanicity, public/private status, free lunch percentage, and school size.

In a model with only a random intercept, let's calculate the intraclass correlation -- the correlation between two kids in the same school.

$$y_{ij} = \beta_{0,j} + \varepsilon_{ij}, \;\; \beta_{0,j} \overset{iid}{\sim} N(\beta_0, \tau^2) \perp \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

```
fit0 <- lmer(mscore~(1|school),data=nels, REML=FALSE)
sigma2hat <- sigma(fit0)*sigma(fit0) #pick off estimate of sigma2
tau2hat <- as.numeric(VarCorr(fit0)$school) #pick off est of tau2
c(sigma2hat,tau2hat,tau2hat/(tau2hat+sigma2hat)) #show vars and correlation
```

```
## [1] 73.7084447 23.6341046  0.2427932
```

STA 610L

Heterogeneity across schools

How much of the heterogeneity across schools is explained by enrollment?

```
fit1 <- lmer(mscore~as.factor(enroll)+(1|school),data=nels, REML=FALSE)
sigma2hat <- sigma(fit1)*sigma(fit1) #pick off estimate of sigma2
tau2hat <- as.numeric(VarCorr(fit1)$school) #pick off est of tau2
c(sigma2hat,tau2hat,tau2hat/(tau2hat+sigma2hat)) #show vars and correlation
```

```
## [1] 73.7202995 23.0948621  0.2385459
```

Not much!

STA 610L

Heterogeneity across schools

How much of the remaining heterogeneity across schools is explained by the percentage of kids eligible for free or reduced price lunch?

```
fit2=lmer(mscore~as.factor(enroll)+as.factor(flp)+(1|school),
          data=nels, REML=FALSE)
sigma2hat <- sigma(fit2)*sigma(fit2) #pick off estimate of sigma2
tau2hat <- as.numeric(VarCorr(fit2)$school) #pick off est of tau2
c(sigma2hat,tau2hat,tau2hat/(tau2hat+sigma2hat)) #show vars and correlation
```

```
## [1] 73.7655328 13.5097521  0.1547947
```

Wow, "school-level SES" explained a lot of that heterogeneity.

STA 610L

Heterogeneity across schools

What if we add public/private status?

```
fit3=lmer(mscore~as.factor(enroll)+as.factor(flp)+as.factor(public)+
            (1|school),data=nels, REML=FALSE)
sigma2hat=sigma(fit3)*sigma(fit3) #pick off estimate of sigma2
tau2hat=as.numeric(VarCorr(fit3)$school) #pick off est of tau2
c(sigma2hat,tau2hat,tau2hat/(tau2hat+sigma2hat)) #show vars and correlation
```

```
## [1] 73.7749179 13.2366759  0.1521254
```

Heterogeneity across schools

Now we add urbanicity.

```
fit4=lmer(mscore~as.factor(enroll)+as.factor(flp)+as.factor(public)+
            urbanicity+(1|school),data=nels, REML=FALSE)
sigma2hat=sigma(fit4)*sigma(fit4) #pick off estimate of sigma2
tau2hat=as.numeric(VarCorr(fit4)$school) #pick off est of tau2
c(sigma2hat,tau2hat,tau2hat/(tau2hat+sigma2hat)) #show vars and correlation
```

```
## [1] 73.779034 12.908770  0.148911
```

STA 610L

Summary

As we add more group-level predictors,

- $\hat{\tau}^2$ decreases
- $\widehat{\sigma}^2$ stays about the same
- the within-group correlation is nonincreasing (and with the addition of some variables decreases substantially)

STA 610L

NELS Data

Let's return to our data from a data analysis perspective (rather than just illustrating aspects of the multi-level model), considering the hypotheses regarding the role of school-specific and individual-specific factors in math test scores.

We'll start with a simple model and build from there, using the BIC as our primary selection criterion.

$$y_{ij} = \beta_{0,j} + \beta_{1,j}\text{ses}_{ij} + \varepsilon_{ij}, \;\; \beta_{0,j} = \beta_0 + b_{0,j} \quad \beta_{1,j} = \beta_1 + b_{1,j}$$

$$\begin{pmatrix} b_{0,j} \\ b_{1,j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{12} & \tau_{22} \end{pmatrix} \right) \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

This model allows random intercepts and slopes across schools.

STA 610L

NELS Data

We saw previously that the random slope did explain additional heterogeneity in a model without school-level predictors.

We'll come back to that question again once we add a few school level predictors to the model.

Let's first compare our starting model to models that add enrollment to the mix, so that

$$\beta_{0,j} = \beta_0 + \alpha_{0,k} I(\text{enroll}_j = k) + b_{0,j}$$

$$\beta_{1,j} = \beta_1 + \alpha_{1,k} I(\text{enroll}_j = k) + b_{1,j}$$

$$k = 1, \ldots, 5$$

We'll use ML estimation because we may wish to consider likelihood ratio tests of the mean parameters.

NELS Data

First, check out the base model.

```
mod1=lmer(mscore~sesstd+(sesstd|school),data=nels, REML=FALSE)
summary(mod1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mscore ~ sesstd + (sesstd | school)
##    Data: nels
##
##       AIC      BIC   logLik deviance df.resid
##   92553.1  92597.9 -46270.5  92541.1    12968
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8910 -0.6382  0.0179  0.6669  4.4613
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  school   (Intercept) 12.2231  3.4961
##           sesstd       0.8562  0.9253   0.11
##  Residual             67.3451  8.2064
## Number of obs: 12974, groups:  school, 684
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 50.67670    0.15511   326.7
## sesstd       3.27708    0.09256    35.4
##
## Correlation of Fixed Effects:
##        (Intr)
## sesstd 0.007
```

STA 610L

NELS Data

```
mod2a=lmer(mscore~as.factor(enroll)+sesstd+(sesstd|school),
           data=nels, REML=FALSE)
mod2b=lmer(mscore~as.factor(enroll)+sesstd+as.factor(enroll):sesstd+
             (sesstd|school),data=nels, REML=FALSE)
anova(mod2b,mod2a)
```

```
## Data: nels
## Models:
## mod2a: mscore ~ as.factor(enroll) + sesstd + (sesstd | school)
## mod2b: mscore ~ as.factor(enroll) + sesstd + as.factor(enroll):sesstd +
## mod2b:     (sesstd | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod2a   11 92557 92639 -46267    92535
## mod2b   16 92559 92678 -46263    92527 7.9798  5     0.1574
```

```
anova(mod2a,mod1)
```

```
## Data: nels
## Models:
## mod1: mscore ~ sesstd + (sesstd | school)
## mod2a: mscore ~ as.factor(enroll) + sesstd + (sesstd | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod1     6 92553 92598 -46271    92541
## mod2a   11 92557 92639 -46267    92535 6.1315  5     0.2936
```

Here we don't see much evidence that enrollment is useful, so we don't need to use it.

NELS Data

Next we can consider eligibility for free and reduced lunch, so that

Here we'll consider a variety of models, including the one above, a model without the interaction with flp, a model that has the flp main effect but drops the SES random effect, and a model that drops all the school random effects given that flp is in the model $(\tau = 0)$.

```
mod3a=lmer(mscore~as.factor(flp)+sesstd+(sesstd|school),
           data=nels, REML=FALSE)
mod3b=lmer(mscore~as.factor(flp)+sesstd+as.factor(flp)*sesstd+
             (sesstd|school),data=nels, REML=FALSE)
mod3c=lmer(mscore~as.factor(flp)+sesstd+(1|school),
           data=nels, REML=FALSE)
mod3d=lm(mscore~as.factor(flp)+sesstd,data=nels)
anova(mod3b,mod3a)
anova(mod3a,mod1)
anova(mod3c,mod3a) #just look at BIC here
BIC(mod3d) #check if random intercept needed by comparing to BIC from 3c
```

STA 610L

Model selection

```
mod3a=lmer(mscore~as.factor(flp)+sesstd+(sesstd|school),
           data=nels, REML=FALSE)
mod3b=lmer(mscore~as.factor(flp)+sesstd+as.factor(flp)*sesstd+(sesstd|school),
           data=nels, REML=FALSE)
anova(mod3b,mod3a)
```

```
## Data: nels
## Models:
## mod3a: mscore ~ as.factor(flp) + sesstd + (sesstd | school)
## mod3b: mscore ~ as.factor(flp) + sesstd + as.factor(flp) * sesstd +
## mod3b:     (sesstd | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod3a    8 92395 92454 -46189    92379
## mod3b   10 92384 92459 -46182    92364 14.384  2  0.0007525
```

STA 610L

Model selection

```r
anova(mod3a,mod1)
```

```
## Data: nels
## Models:
## mod1: mscore ~ sesstd + (sesstd | school)
## mod3a: mscore ~ as.factor(flp) + sesstd + (sesstd | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod1     6 92553 92598 -46271    92541
## mod3a    8 92395 92454 -46189    92379 162.51  2  < 2.2e-16
```

STA 610L

Model selection

```
mod3c=lmer(mscore~as.factor(flp)+sesstd+(1|school),
           data=nels, REML=FALSE)
anova(mod3c,mod3a) #just look at BIC here
```

```
## Data: nels
## Models:
## mod3c: mscore ~ as.factor(flp) + sesstd + (1 | school)
## mod3a: mscore ~ as.factor(flp) + sesstd + (sesstd | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod3c    6 92404 92449 -46196    92392
## mod3a    8 92395 92454 -46189    92379 13.371  2   0.001249
```

```
mod3d=lm(mscore~as.factor(flp)+sesstd,data=nels)
BIC(mod3d) #check if random intercept needed by comparing to BIC from 3c
```

```
## [1] 93151.9
```

STA 610L

MODEL SELECTION

Note that BIC now likes the model without a random slope -- we evaluated that because we thought that after introducing a school-level SES variable to the model, the importance of the individual-level SES variable may change.

It also prefers a model without an interaction between individual-level SES and school-level SES (measured by flp).

Model selection

Now our model for the coefficients is

```
summary(mod3c)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mscore ~ as.factor(flp) + sesstd + (1 | school)
##    Data: nels
## 
##      AIC      BIC   logLik deviance df.resid 
##  92403.9  92448.7 -46196.0  92391.9    12968 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.9560 -0.6434  0.0178  0.6710  4.4906 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept)  9.004   3.001   
##  Residual             67.959   8.244   
## Number of obs: 12974, groups:  school, 684
## 
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)     52.84307    0.24462 216.020
## as.factor(flp)2 -1.87992    0.33042  -5.689
## as.factor(flp)3 -4.79607    0.36150 -13.267
## sesstd           3.10819    0.08578  36.233
## 
## Correlation of Fixed Effects:
##             (Intr) as.()2 as.()3
## as.fctr(f)2 -0.739              
## as.fctr(f)3 -0.697  0.514       
## sesstd      -0.202  0.143  0.237
```

STA 610L

Model selection

The more students we have eligible for the free and reduced price lunch program, the lower the math scores.

In addition, the coefficient on individual-level SES did not change much in magnitude -- so SES operates both on the school level and the individual level.

Let's now add the public school indicator.

```
mod4a=lmer(mscore~as.factor(flp)+as.factor(public)+sesstd+
            (1|school),data=nels, REML=FALSE)
mod4b=lmer(mscore~as.factor(flp)+as.factor(public) +
             sesstd+as.factor(public)*sesstd+(1|school),
           data=nels, REML=FALSE)
anova(mod4b,mod4a)
anova(mod4b,mod3c)
#summary(mod4b)
```

Model selection

```
mod4a=lmer(mscore~as.factor(flp)+as.factor(public)+sesstd+(1|school),
           data=nels, REML=FALSE)
mod4b=lmer(mscore~as.factor(flp)+as.factor(public) + sesstd+
             as.factor(public)*sesstd+(1|school),data=nels, REML=FALSE)
anova(mod4b,mod4a)
```

```
## Data: nels
## Models:
## mod4a: mscore ~ as.factor(flp) + as.factor(public) + sesstd + (1 | school)
## mod4b: mscore ~ as.factor(flp) + as.factor(public) + sesstd + as.factor(public) *
## mod4b:     sesstd + (1 | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod4a    7 92406 92458 -46196    92392
## mod4b    8 92396 92456 -46190    92380 11.948  1   0.000547
```

STA 610L

Model selection

```
## Data: nels
## Models:
## mod3c: mscore ~ as.factor(flp) + sesstd + (1 | school)
## mod4b: mscore ~ as.factor(flp) + as.factor(public) + sesstd + as.factor(public) *
## mod4b:     sesstd + (1 | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod3c    6 92404 92449 -46196    92392
## mod4b    8 92396 92456 -46190    92380 12.081  2   0.002381
```

The BIC suggests leaving public/private out of the model.

STA 610L

Model selection

Now let's consider urban/suburban/rural status.

```
mod5a=lmer(mscore~as.factor(flp)+urbanicity+sesstd+(1|school),
           data=nels, REML=FALSE)
mod5b=lmer(mscore~as.factor(flp)+urbanicity+sesstd+
             urbanicity*sesstd+(1|school),
           data=nels, REML=FALSE)
anova(mod5b,mod5a)
anova(mod5a,mod3c)
summary(mod5b)
```

STA 610L

Model selection

```
mod5a=lmer(mscore~as.factor(flp)+urbanicity+sesstd+(1|school),
           data=nels, REML=FALSE)
mod5b=lmer(mscore~as.factor(flp)+urbanicity+sesstd+
             urbanicity*sesstd+(1|school),
           data=nels, REML=FALSE)
anova(mod5b,mod5a)
```

```
## Data: nels
## Models:
## mod5a: mscore ~ as.factor(flp) + urbanicity + sesstd + (1 | school)
## mod5b: mscore ~ as.factor(flp) + urbanicity + sesstd + urbanicity *
## mod5b:     sesstd + (1 | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod5a    8 92400 92460 -46192    92384
## mod5b   10 92390 92465 -46185    92370 13.373  2   0.001248
```

STA 610L

Model selection

```
anova(mod5a,mod3c)
```

```
## Data: nels
## Models:
## mod3c: mscore ~ as.factor(flp) + sesstd + (1 | school)
## mod5a: mscore ~ as.factor(flp) + urbanicity + sesstd + (1 | school)
##       npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod3c    6 92404 92449 -46196    92392
## mod5a    8 92400 92460 -46192    92384 8.0578  2    0.01779
```

BIC suggests leaving urbanicity out of the model.

STA 610L

Summary of Selection using BIC

- Enrollment, urbanicity, and public/private status did not add much to our model using the BIC as our selection criterion
- The lower the SES status of the whole school (measured by percent eligible for free and reduced-price lunch), the lower the math scores on average
- Having higher individual-level SES was associated with higher math scores regardless of the school environment
- A random intercept for school explained significant variability across schools and controlled for lack of independence within schools

$$y_{ij} = \beta_{0,j} + \beta_1 \text{ses}_{ij} + \varepsilon_{ij}$$

$$\beta_{0,j} = \beta_0 + \psi_{0,l} I(\text{flp}_j = l) + b_{0,j}$$

$$b_{0,j} \sim N\left(0, \tau^2\right) \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

STA 610L

Final model again

```
summary(mod3c)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mscore ~ as.factor(flp) + sesstd + (1 | school)
##    Data: nels
## 
##      AIC      BIC   logLik deviance df.resid 
##  92403.9  92448.7 -46196.0  92391.9    12968 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.9560 -0.6434  0.0178  0.6710  4.4906 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept)  9.004   3.001   
##  Residual             67.959   8.244   
## Number of obs: 12974, groups:  school, 684
## 
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)     52.84307    0.24462 216.020
## as.factor(flp)2 -1.87992    0.33042  -5.689
## as.factor(flp)3 -4.79607    0.36150 -13.267
## sesstd           3.10819    0.08578  36.233
## 
## Correlation of Fixed Effects:
##             (Intr) as.()2 as.()3
## as.fctr(f)2 -0.739              
## as.fctr(f)3 -0.697  0.514       
## sesstd      -0.202  0.143  0.237
```

STA 610L

What's next?

Move on to the readings for the next module!