STA 610L: Module 2.8

Random effects ANCOVA (cont'd)

Dr. Olanrewaju Michael Akande

STA 610L

NELS DATA RECAP

We fit a random intercepts ANCOVA to assess the association between math score and SES.

That is, we allowed school-specific intercepts while including SES as a covariate x_{ij}:

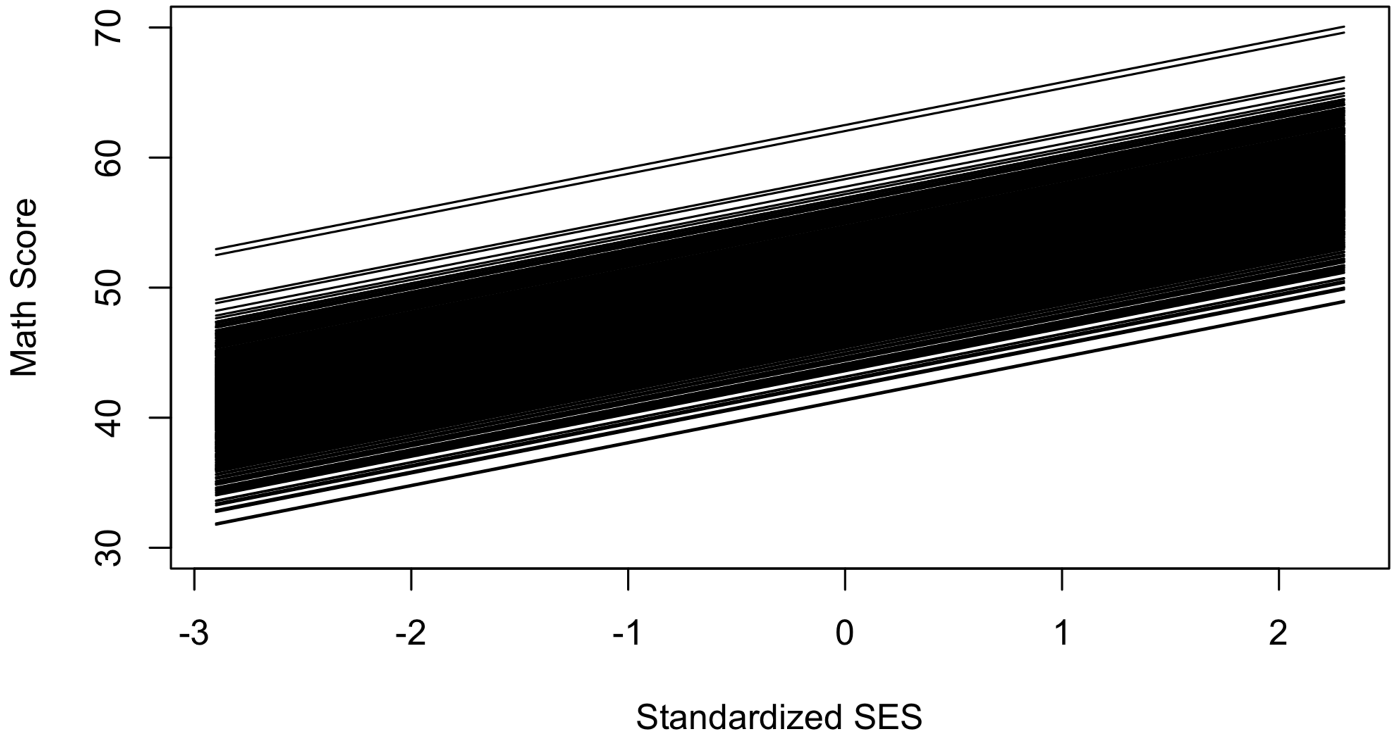
$$
\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij}; \quad i = 1, \ldots, n_j \\
\varepsilon_{ij} &\overset{iid}{\sim} N(0, \sigma^2); \quad \beta_{0j} \sim N(\beta_0, \tau^2); \quad j = 1, \ldots, J,
\end{aligned}
$$

This model allows separate intercepts for each school but assumes a common slope.

STA 610L

Random intercepts model

Currently, we have the fitted lines:



STA 610L

Extending the model

As an initial step towards extending our random intercepts model, we can examine the variation in slopes across 685 separate regression models fit separately in each school:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}, \ \ \varepsilon_{ij} \sim N(0, \sigma_j^2).$$

Here, we will fit an "unpooled" model with completely different regressions for each school.

In each case, if we write $\beta_j = (\beta_{0j}, \beta_{1j})'$, then the OLS estimaste of β_j is

$$\widehat{\beta}_j = (X_j' X_j)^{-1} X_j' y_j,$$

where

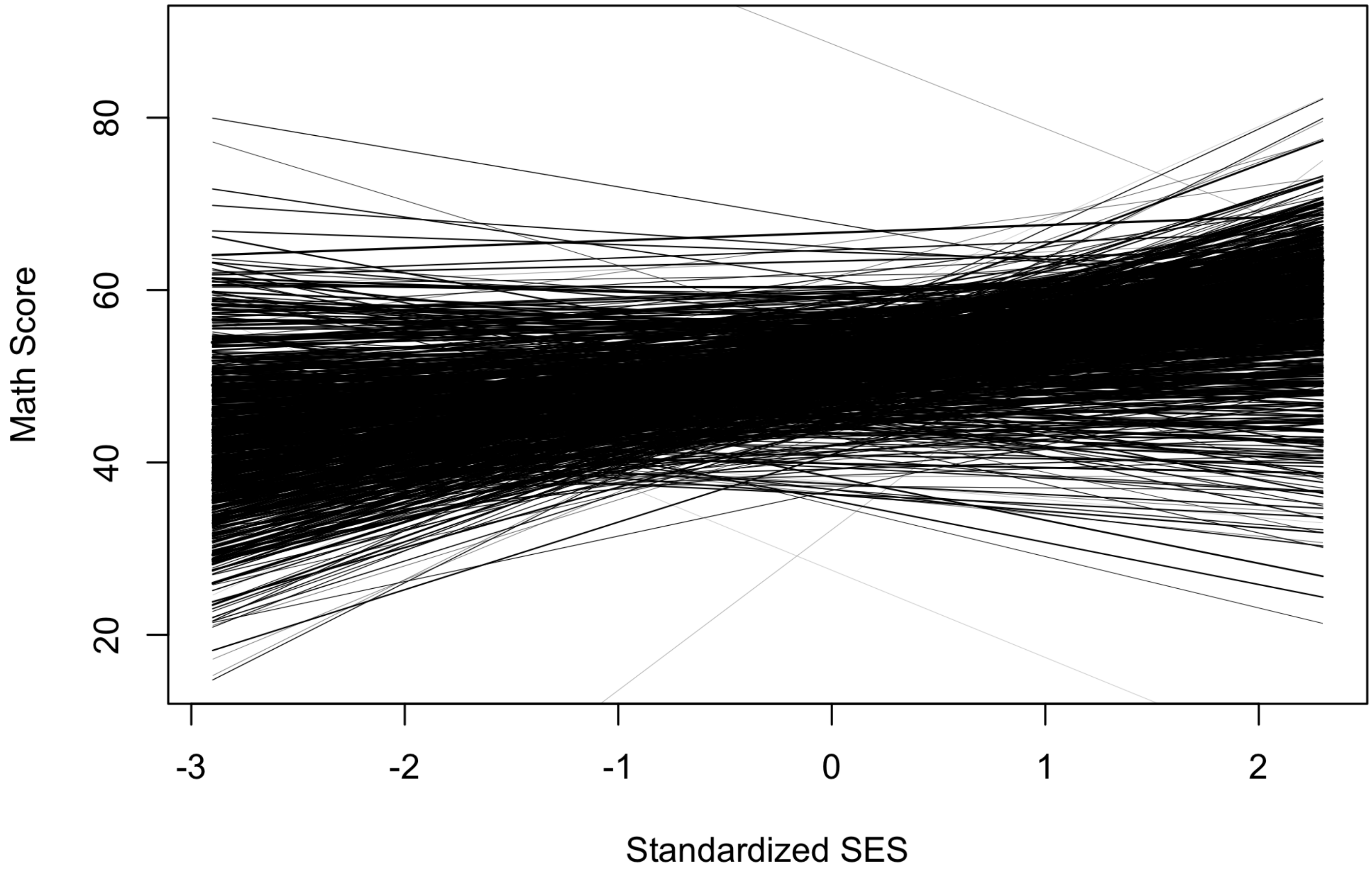
- y_j is a vector containing the math scores for all students in school j,
- X_j contains a column of 1's for the intercept and a column containing the SES of each student in school j.

Different regressions for each school

```
xplot <- seq(-2.9,2.3,by=.1)
yplot <- rep(60,length(xplot))
plot(xplot,yplot,type="n",ylim=c(15,90),xlab="Standardized SES",ylab="Math Score")
for(school in
    id.schools[order( avmscore.schools )[seq(1,length(avmscore.schools), by=1)]])
{
  y <- nels$mscore[nels$school==school]
  x <- nels$sesstd[nels$school==school]
  m <- lm(y~x)
  yplot <- coef(m)[1]+coef(m)[2]*xplot
  lines(xplot,yplot,lwd=length(y)/30)
}
```

STA 610L

Different regressions for each school



STA 610L

Different regressions for each school

The plot looks pretty different from what we had before!

By the way, line width is proportional to the number of students tested in each school.

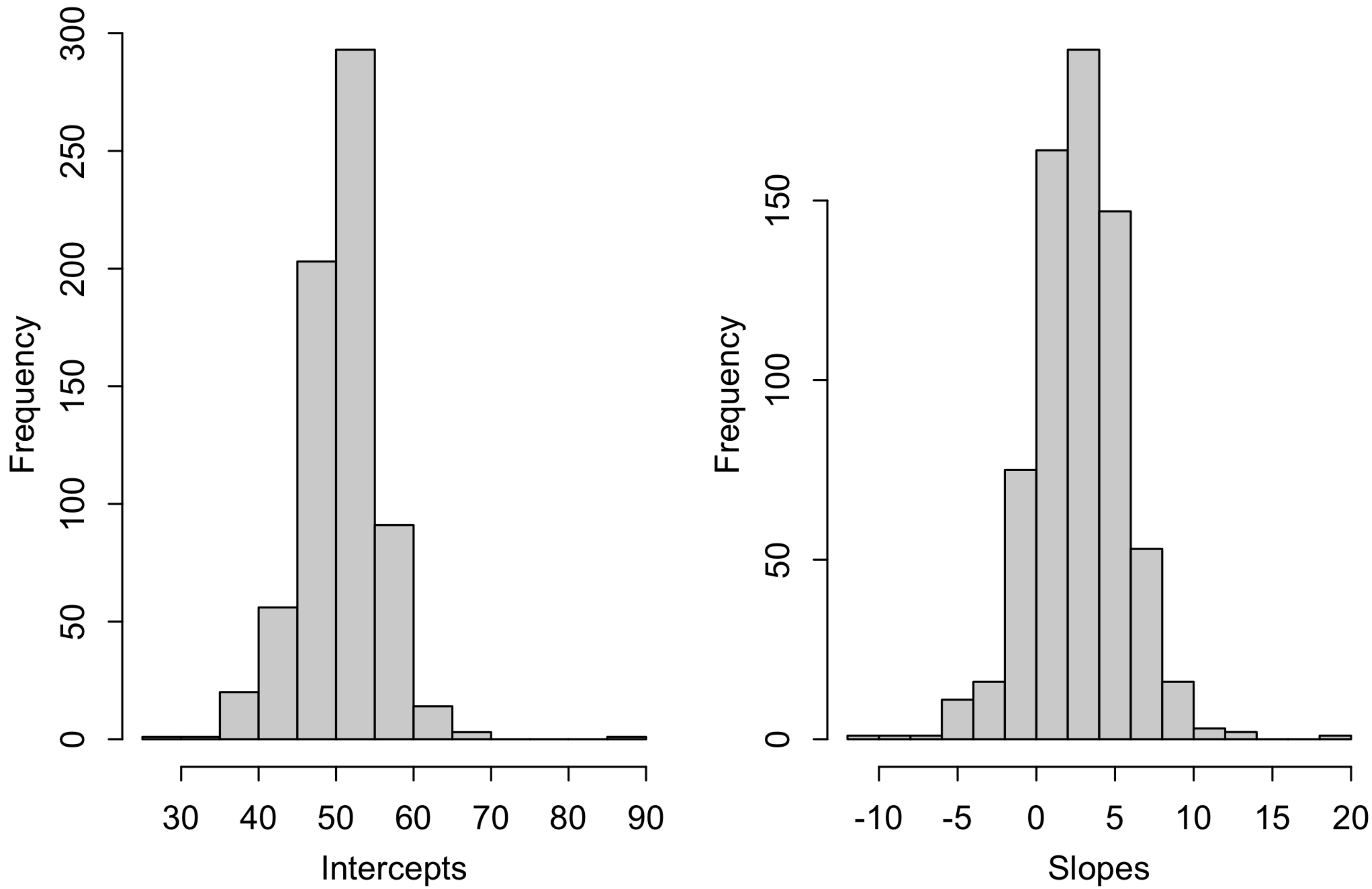
Among the schools, roughly 85% have positive slope estimates.

Also, the steepest slopes (positive and negative) tend to occur in the schools with smaller sample sizes.

So, there may be "significant" differences in the slopes across schools.

How do we get good estimates of the school-specific slopes?

Histograms of school-specific intercepts and slopes



STA 610L

School-specific slopes

As always, the goal is to slowly work our way up from pooled and unpooled models.

Building on our knowledge of random intercept models (including ANOVA and ANCOVA), our usual three approaches will lead to the following estimates in this case.

- $\widehat{\beta}_j = \widehat{\beta}_j^{UNPOOL} = (X_j' X_j)^{-1} X_j' y_j$, relying only on the data from school j
- $\widehat{\beta}_j = \widehat{\beta}^{POOL} = (X'X)^{-1} X'y$, using all the data and pooling across schools
- $\widehat{\beta}_j = w_j \widehat{\beta}_j^{UNPOOL} + (1 - w_j) \widehat{\beta}^{POOL}$, doing something in between those two extremes. This is of course desirable here.

STA 610L

School-specific slopes

Basically, we can play our usual game. We can fit a single model with school-specific slopes and intercepts.

Now, these factors could be fixed or random effects.

Let's actually take a step back from the random effects model, and start with a model with ONLY fixed effects.

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

If we wish to evaluate whether there is heterogeneity across schools, an easy approach is to fit the model as a linear regression using indicator variables.

STA 610L

Fixed effects model

That is,

$$y_{ij} = \beta_0 + \alpha_j I(\text{school} = j) + \beta_1 x_{ij} + \gamma_j x_{ij} I(\text{school} = j) + \varepsilon_{ij},$$

where we assume $\alpha_1 = \gamma_1 = 0$ (reference cell coding).

In this case, an F test can be used to evaluate the hypothesis

$$H_0 : \gamma_j = 0, \quad j = 1, \ldots, J - 1,$$

which corresponds to a constant effect of SES, β_1, across groups.

FIXED EFFECTS MODEL

In R, we have

```
m6 <- lm(mscore~school+sesstd, data=nels) #pooled slope
m7 <- lm(mscore~school+sesstd+school:sesstd, data=nels) #school-specific slopes
anova(m6,m7)
```

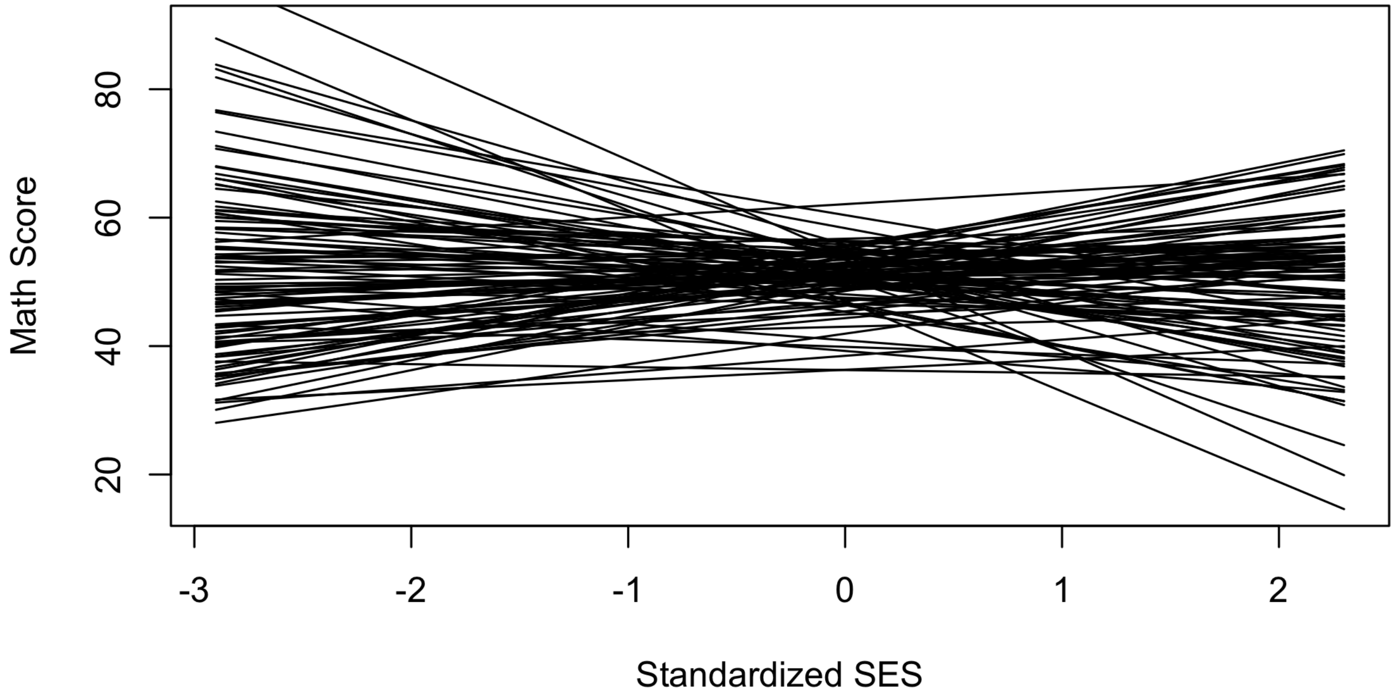
```
## Analysis of Variance Table
##
## Model 1: mscore ~ school + sesstd
## Model 2: mscore ~ school + sesstd + school:sesstd
##   Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
## 1  12289 832772
## 2  11607 776507 682     56264 1.2332 4.865e-05
```

Here we have evidence in favor of school-specific slopes in the fixed effects model.

However, our estimates of school-specific slopes in small schools may have high standard errors.

STA 610L

Estimated lines



Remember that unlike the previous plot of estimated lines, we have now estimated a common variance.

Hierarchical regression models

Our usual hierarchical normal model involves two levels:

- within-group model $p(y_{1j}, \ldots, y_{n_j j} \mid \theta_j)$ describing heterogeneity in group j; and
- among-groups model $p(\theta_1, \ldots, \theta_J)$.

Specifically, we let

- $\theta_j = (\mu_j, \sigma^2)$ (or $\theta_j = (\mu_j, \sigma_j^2)$ when desired)
- $y_{1j}, \ldots y_{n_j j} \mid \theta_j \overset{iid}{\sim} N\left(\mu_j, \sigma^2\right)$
- $\mu_1, \ldots, \mu_j \overset{iid}{\sim} N\left(\mu, \tau^2\right)$.

STA 610L

Hierarchical regression models

In the standard regression setting, we can then write

- $\theta_j = (\beta_j, \sigma^2)$
- $y_{ij} = \beta_j' \boldsymbol{x}_{ij} + \varepsilon_{ij}, \;\; \varepsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right)$
- $\beta_1, \ldots, \beta_J \overset{iid}{\sim} p(\beta_j | \beta)$

How should we model $p(\beta_j|\beta)$, the heterogeneity across groups in the vector of regression coefficients?

STA 610L

Hierarchical regression models

It is often the case that intercepts and slopes are correlated, so that we should try to account for that when including varying intercepts and slopes.

- In a study of income over time, people who start off making more money may have larger raises over time.
- In a study of exercise, people who exercise a lot at the start of the study may have lower changes over time than those who exercise less

A natural choice for the β_j model is the multivariate normal distribution, which allows for correlation among the group-specific regression coefficients.

Hierarchical regression models

We can specify our model in the context of maximum likelihood estimation as

- $y_j \mid \beta_j \sim MVN(X_j\beta_j, \sigma^2 I)$
- $\beta_j \sim MVN(\beta, \Sigma_\beta)$

  where

$$\beta_j \sim MVN(\beta, \Sigma_\beta) \iff \beta_j = \beta + b_j, \;\; b_j \sim MVN(0, \Sigma_\beta).$$

The parameters are

- β, an "overall" mean vector of regression coefficients,
- b_j's, the vectors of group specific coefficients, and
- Σ_β, a covariance matrix describing the variability of the β_j's around β.

STA 610L

Hierarchical regression models

We can combine terms and write the model as

$$y_j = X_j\beta_j + \varepsilon_j = X_j(\beta + b_j) + \varepsilon_j = X_j\beta + X_jb_j + \varepsilon_j$$

Here

- β is usually called a fixed effect (fixed across all groups);
- b_j is usually called a random effect (varies across groups and can be considered random if groups were randomly sampled); and
- a model with both fixed and random effects is often called a mixed-effects model.

STA 610L

Ad hoc ESTIMATES

We can get a rough estimate of β by averaging the estimates from our 685 school-specific regression models.

```
BETA.OLS<-NULL
DF<-SSE<-0
y.nels=nels$mscore
ses.nels=nels$sesstd
for(j in sort(unique(nels$school)))
{
  yj<-y.nels[nels$school==j]
  xj<-ses.nels[nels$school==j]
  fitj<-lm(yj~xj)
  BETA.OLS<-rbind(BETA.OLS,fitj$coef)
  if(length(yj)>=2)  {SSE<-SSE+sum(fitj$res^2) ; DF<-DF+length(yj)-2 }
}
s2.ols<-SSE/DF
apply(BETA.OLS,2,mean,na.rm=TRUE)
```

```
## (Intercept)          xj
##   50.618228    2.760704
```

This estimator is not perfect -- it equally weights all the schools, regardless of size. We would prefer to assign a lower weight to schools with less data.

Ad hoc estimates

We can get a *very rough* estimate of Σ_β:

```
cov(BETA.OLS,use="complete.obs") #dropped n=1 schools
```

```
##             (Intercept)        xj
## (Intercept)  26.7958507 0.7529181
## xj            0.7529181 8.9391754
```

This estimate not only ignores sample size differences, it also ignores the variability of $\widehat{\beta}_j$ around β_j:

$$\text{Var}[\widehat{\beta}_j\text{'s around } \widehat{\beta}] \approx \text{Var}[\beta_j\text{'s around } \beta] + \text{Var}[\widehat{\beta}_j\text{'s around } \beta_j\text{'s}] :$$

basically, the sample covariance of the $\widehat{\beta}_j$'s is approximately

$$\Sigma_\beta + \text{estimation error}$$

STA 610L

Covariance within Groups

$$Cov(y_j) = E[(y_j - E(y_j))(y_j - E(y_j))']$$

In our model

$$y_j = X_j\beta_j + \varepsilon_j = X_j(\beta + b_j) + \varepsilon_j = X_j\beta + X_jb_j + \varepsilon_j,$$

$$y_j - E[y_j] = y_j - X_j\beta = X_jb_j + \varepsilon_j, \;\; b_j \sim N(0, \Sigma_\beta), \;\; \varepsilon_j \sim N(0, \sigma^2 I)$$

and because we specify $b_j \perp \varepsilon_j$,

$$Cov(y_j) = E[(X_jb_j + \varepsilon_j)(X_jb_j + \varepsilon_j)']$$

$$= E[X_jb_jb_j'X_j'] + E[\varepsilon_j\varepsilon_j'] = X_j\Sigma_\beta X_j' + \sigma^2 I.$$

STA 610L

Marginal and conditional distributions of Y

So conditional on b_j,

$$y_j \sim MVN(X_j\beta + X_j b_j, \sigma^2 I)$$

and unconditional on b_j we have

$$p(y_j \mid \beta, \Sigma_\beta, \sigma^2) = MVN(X_j\beta, X_j\Sigma_\beta X_j' + \sigma^2 I).$$

STA 610L

Dependence and conditional independence

Marginal dependence: If we don't know β_j (or b_j), then knowing the response y_{ij} gives me some information about β_j, which gives us some information about $y_{i'j}$, so the observations within a group are dependent.

Conditional independence: If I do know β_j, then knowing y_{ij} does not give me any extra information about $y_{i'j}$, and they are independent. My information about $y_{ij} \perp y_{i'j}$ if I know β_j.

STA 610L

Fitting the model

```
m8 <- lmer(mscore~sesstd+(sesstd|school),data=nels,REML=FALSE)
summary(m8)
```

```
## 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
```

Do we need the random slope in addition to the random intercept?

Let's test whether the slope should be random or fixed.

```
m9 <- lmer(mscore~sesstd+(1|school),data=nels,REML=FALSE)
anova(m9,m8)
```

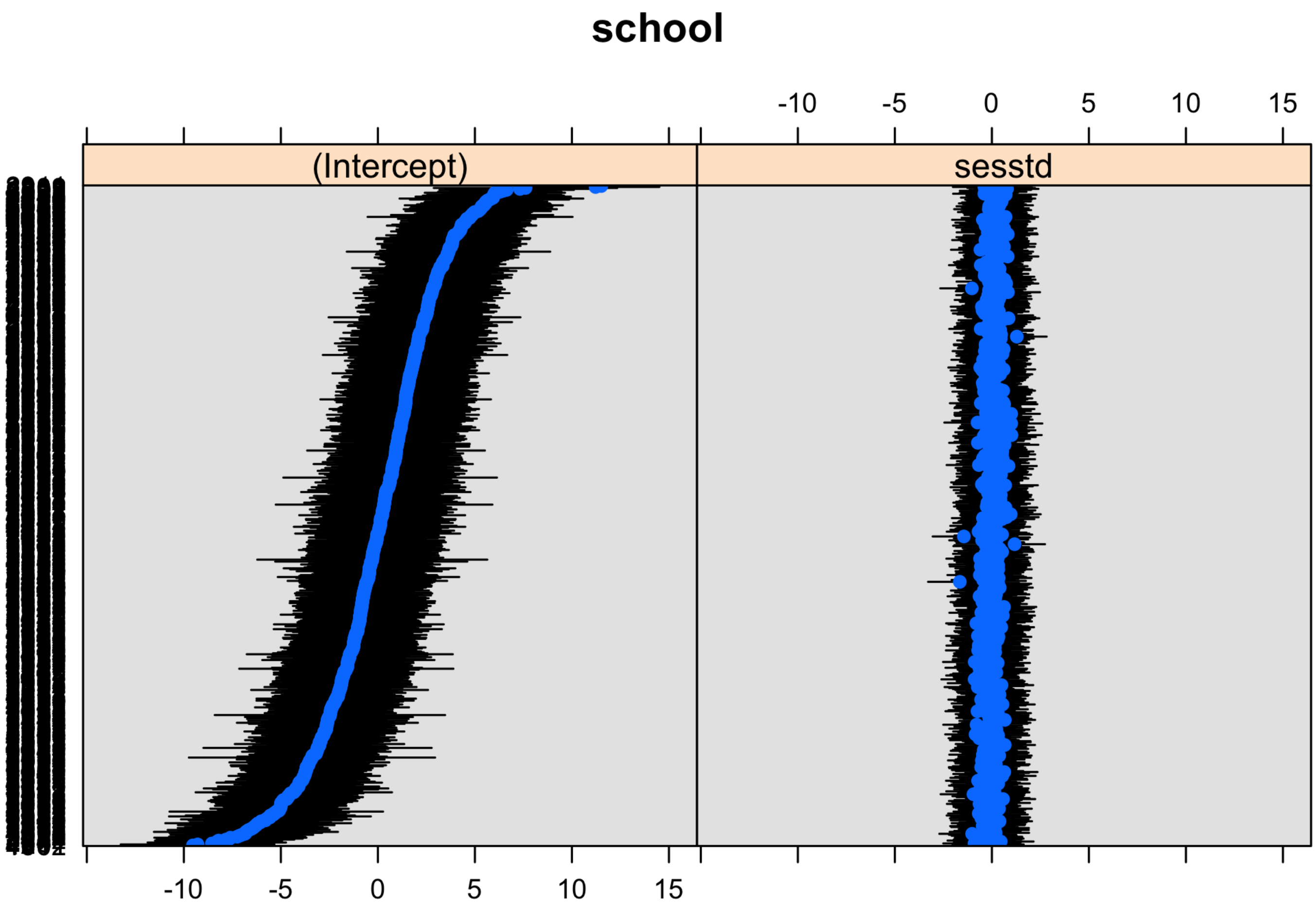
```
## Data: nels
## Models:
## m9: mscore ~ sesstd + (1 | school)
## m8: mscore ~ sesstd + (sesstd | school)
##    npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## m9    4 92562 92592 -46277    92554
## m8    6 92553 92598 -46271    92541 12.582  2   0.001853
```

Yes, looks like the random slope explains additional variance.

STA 610L

Group effects

```
dotplot(ranef(m8, condVar=TRUE))$school
```

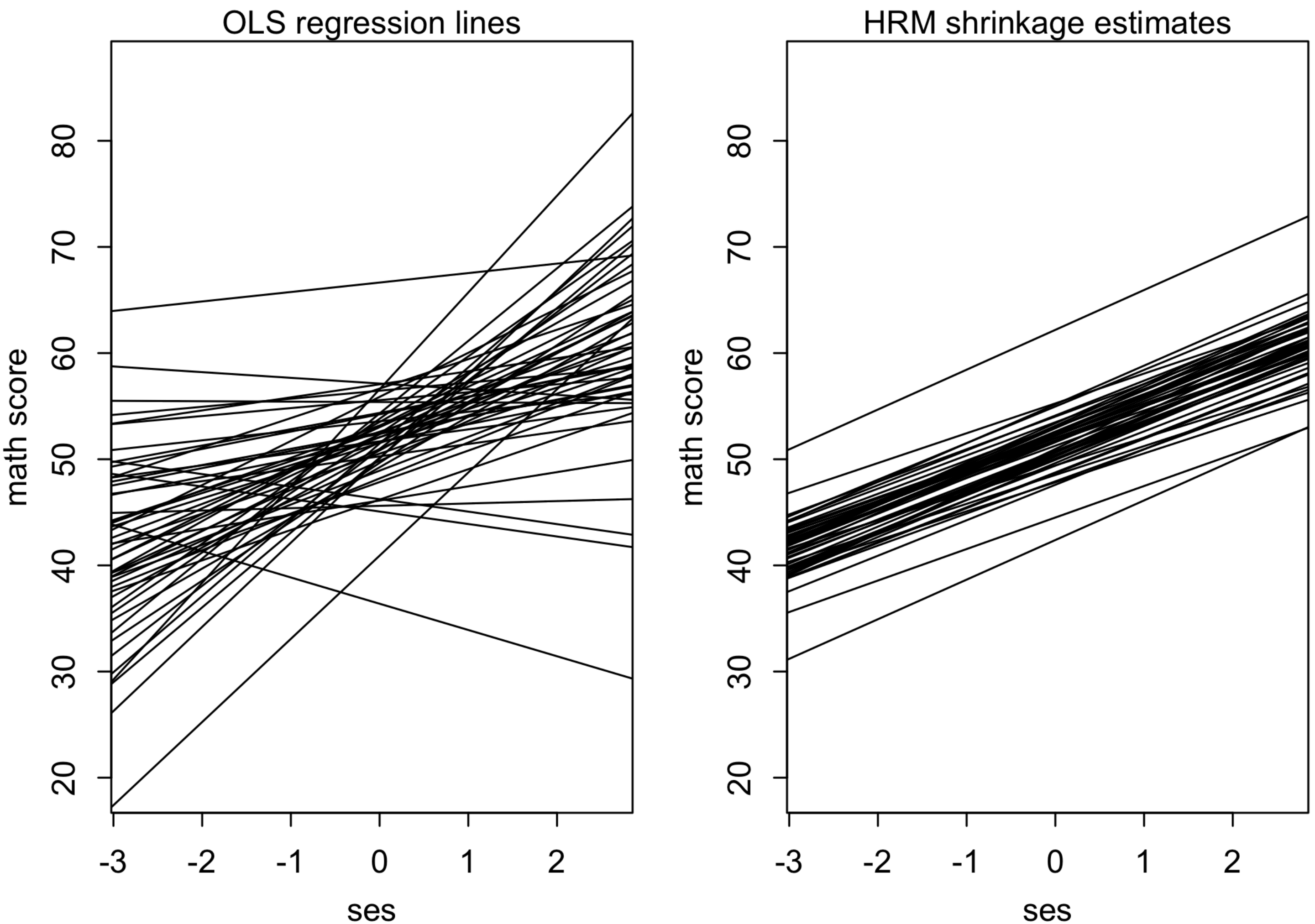


Comparing estimates

```r
options(warn=-1)
B.LME<-as.matrix(ranef(m8)$school)
BETA.LME<-sweep( B.LME , 2 , fixef(m8), "+" )

mpar()
par(mfrow=c(1,2))

ssample<-sample(sort(unique(nels$school)),50) #less clutter take half schools
plot(range(nels$sesstd),range(nels$mscore),type="n",xlab="ses",ylab="math score")
apply( BETA.OLS[ ssample,] ,1,abline)
mtext("OLS regression lines",3)

plot(range(nels$sesstd),range(nels$mscore),type="n",xlab="ses", ylab="math score")
apply( BETA.LME[ ssample,] ,1,abline)
mtext("HRM shrinkage estimates",3)
#options(warn=0)
```

STA 610L

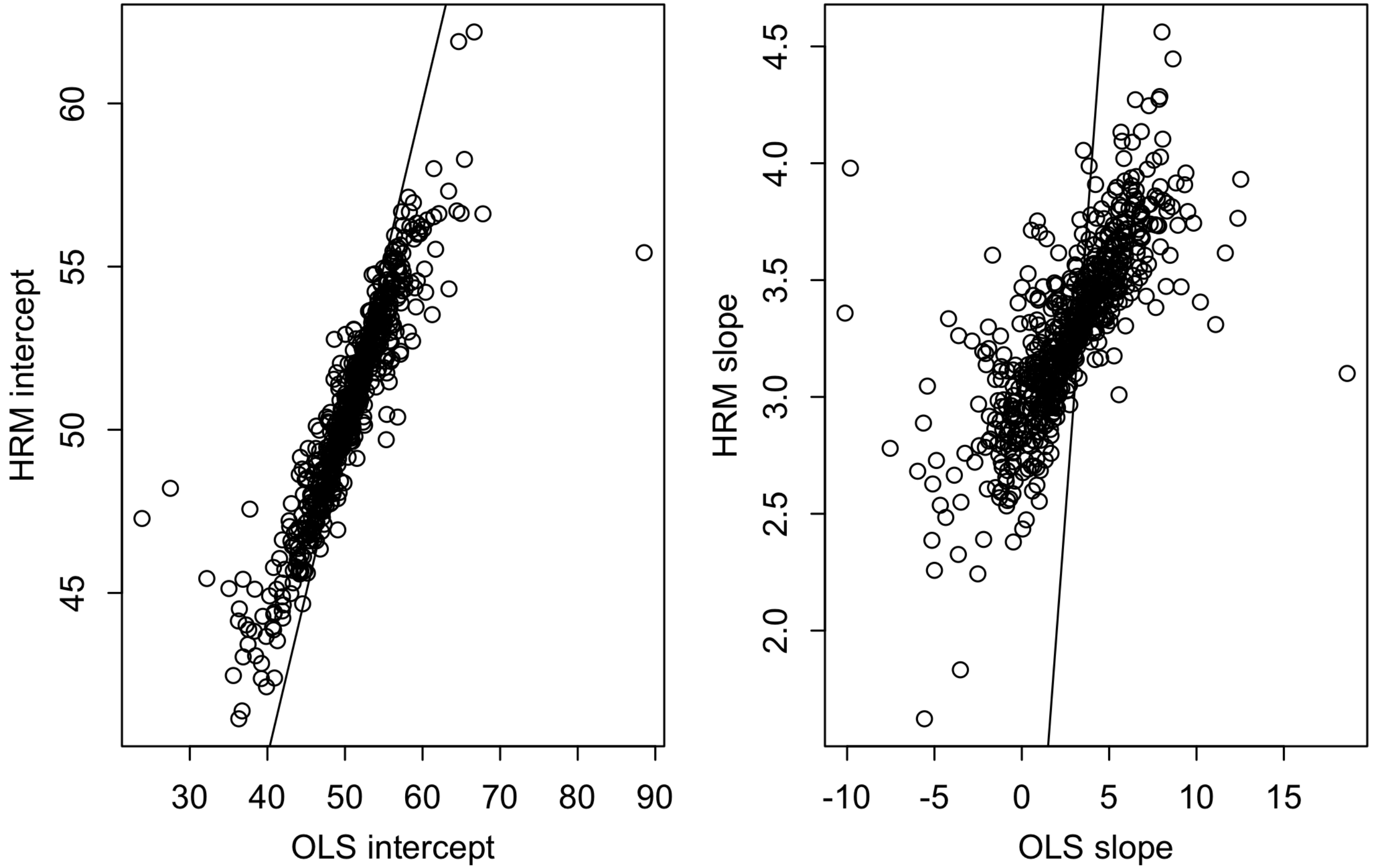
Comparing estimates



Shrinkage Estimates

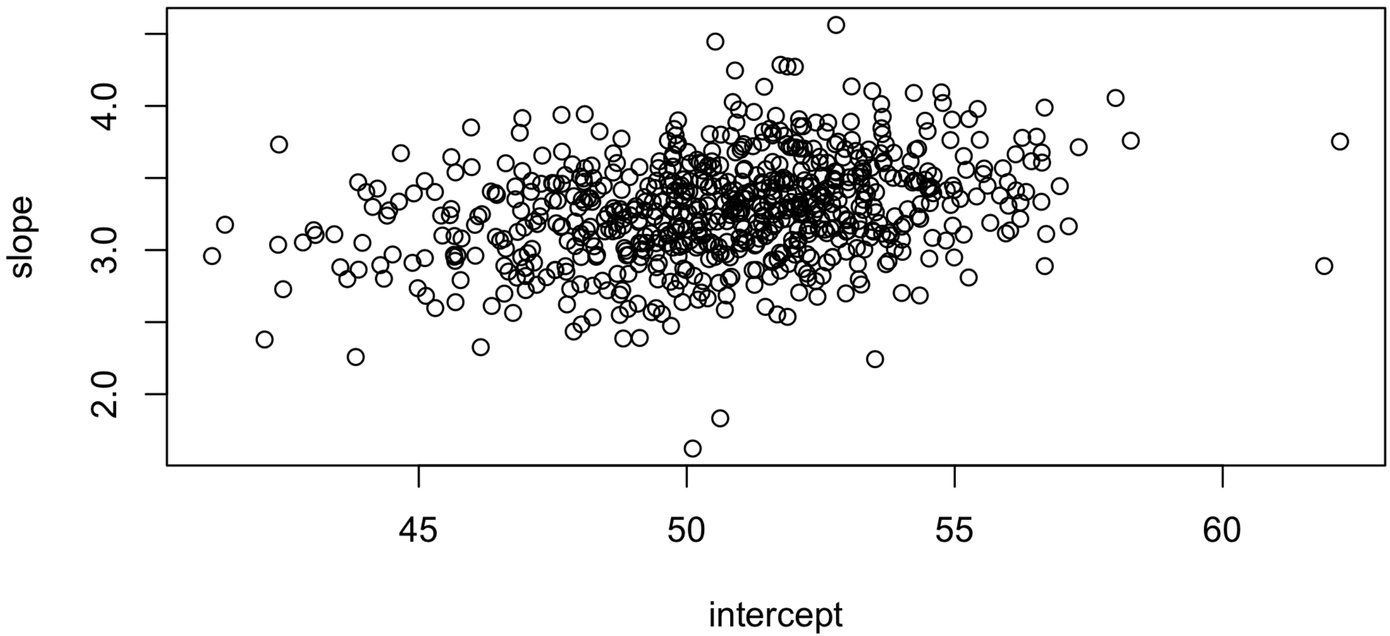
```
mpar()
par(mfrow=c(1,2))
plot(BETA.OLS[,1],BETA.LME[,1],xlab="OLS intercept",ylab="HRM intercept")
abline(0,1)
plot(BETA.OLS[,2],BETA.LME[,2],xlab="OLS slope",ylab="HRM slope")
abline(0,1)
```

Shrinkage Estimates



STA 610L

What kind of schools have big intercepts and big slopes?



Free lunch variable

The US government has programs to provide free or reduced-price lunches to students based on their family economic status.

The percentage of children in a school who are eligible to receive free or reduced-price lunches is an indicator of the school-level socioeconomic status.

In our data, the variable is defined as follows.

- flp=1 if 0-5% of children are eligible to receive free or reduced-price lunch
- flp=2 if 5-30% of children are eligible for benefits
- flp=3 if >30% of children are eligible for benefits

So higher levels of the flp variable are associated with lower school-level socio-economic status

Let's examine whether as a school-level indicator, flp can explain additional variability in school-level intercepts and slopes.

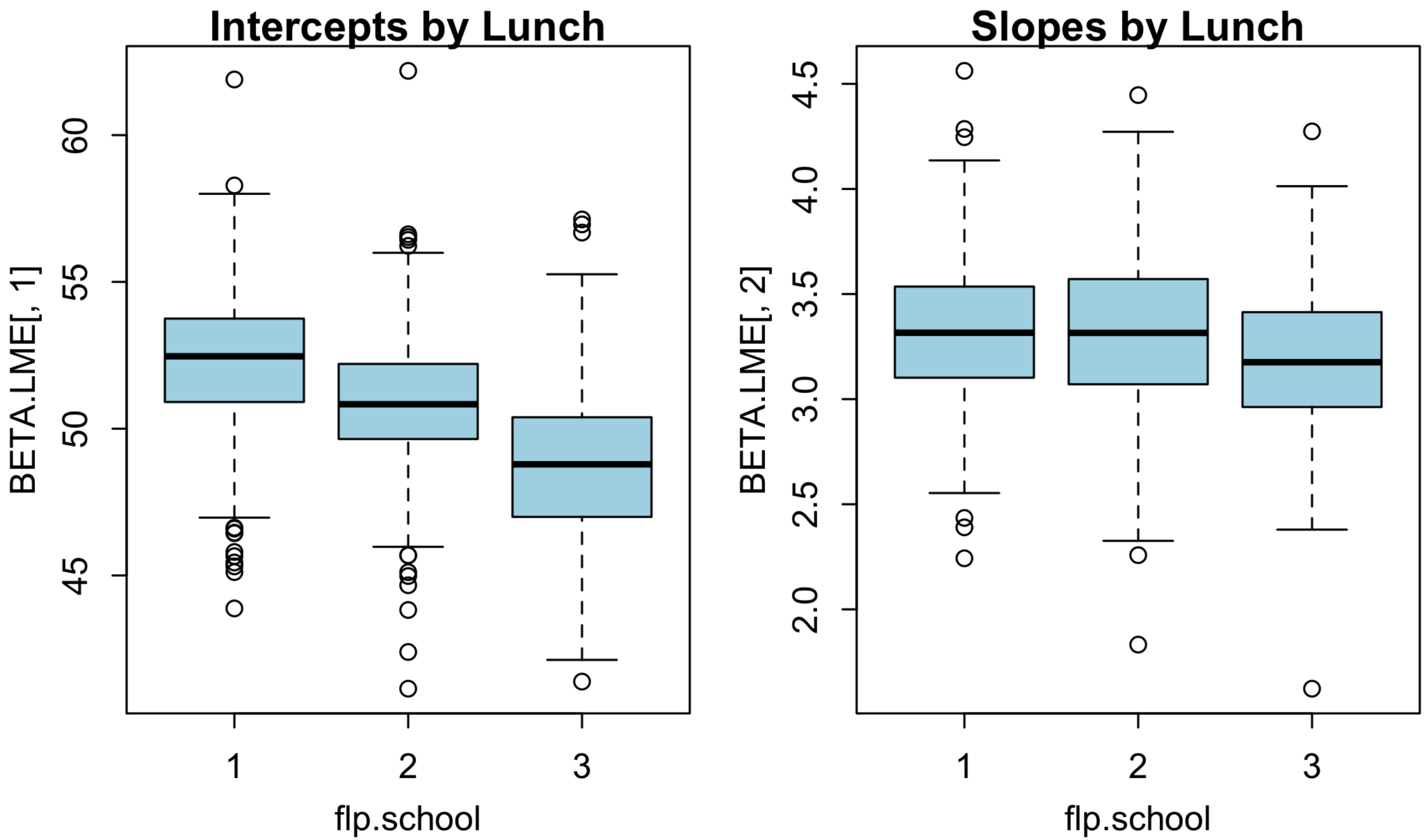
Free lunch variable

```r
flp.school<-tapply( nels$flp , nels$school, mean)
table(flp.school)

### RE and FLP association
mpar()
par(mfrow=c(1,2))
boxplot(BETA.LME[,1]~flp.school,col="lightblue", main="Intercepts by Lunch")
boxplot(BETA.LME[,2]~flp.school,col="lightblue", main="Slopes by Lunch")
```

STA 610L

Free lunch variable

```
## flp.school
##   1   2   3
## 226 257 201
```



Results

Based on the box plots, it seems that the β_{0j} and maybe the β_{1j} are associated with school-level SES, measured by the percentage of kids eligible for free and reduced-price lunch.

We may be interested in the following:

- Testing: is there evidence of a relationship?
- Estimation: what kind of relationship is there?

Let's expand our model so that we can investigate.

STA 610L

Model extension

Our current model can be written

$$y_{ij} = \beta_{0j} + \beta_{1j}\text{ses}_{ij} + \varepsilon_{ij}$$

where

$$\beta_{0j} = \beta_0 + b_{0j} \quad \text{and} \quad \beta_{1j} = \beta_1 + b_{1j}$$

To investigate whether the school-level SES variable explains additional variance, we treat it as an discrete variable (the better way is probably to treat as ordinal categorical) and expand the models for β_{hj} so that

$$\beta_{0j} = \beta_0 + \psi_0\text{flp}_j + b_{0j} \quad \text{and} \quad \beta_{1j} = \beta_1 + \psi_1\text{flp}_j + b_{1j}.$$

Putting things all together, we get

$$y_{ij} = \beta_0 + \psi_0\text{flp}_j + \beta_1\text{ses}_{ij} + \psi_1\text{flp}_j\text{ses}_{ij} + b_{0j} + b_{1j}\text{ses}_{ij} + \varepsilon_{ij}$$

Model extension

Note it does not matter if we use ψ or β notationally, so it may be simpler to write

$$y_{ij} = \beta_0 + \beta_1 \mathrm{flp}_j + \beta_2 \mathrm{ses}_{ij} + \beta_3 \mathrm{flp}_j \mathrm{ses}_{ij} + b_{0j} + b_{1j} \mathrm{ses}_{ij} + \varepsilon_{ij}$$

or more succinctly,

$$y_j = X_j \beta + Z_j b_j + \varepsilon_j,$$

where X_j is a matrix containing a column of 1's, a column for flp, a column for SES, and a column for the flp and SES interaction, and Z_j contains colums for the random intercept and random SES effect.

We'll return to this latter notation in the general context of the linear mixed effects model.

STA 610L

Fitting the model

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: mscore ~ sesstd + flp + sesstd:flp + (sesstd | school)
##    Data: nels
##
##      AIC      BIC   logLik deviance df.resid
##  92396.3  92456.0 -46190.1  92380.3    12966
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9773 -0.6417  0.0201  0.6659  4.5202
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  school   (Intercept)  9.0123  3.0020
##           sesstd       0.8881  0.9424   0.06
##  Residual             67.2595  8.2012
## Number of obs: 12974, groups:  school, 684
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  55.3975     0.3860 143.525
## sesstd        3.3759     0.2501  13.500
## flp          -2.4062     0.1819 -13.230
## sesstd:flp   -0.1451     0.1193  -1.216
##
## Correlation of Fixed Effects:
##            (Intr) sesstd flp
## sesstd     -0.158
## flp        -0.930  0.088
## sesstd:flp  0.086 -0.926 -0.007
```

Certainly flp is doing something, though maybe we don't need that interaction term. We'll come back to this issue shortly.

What's next?

Move on to the readings for the next module!