

Linear Mixed Effects Models

Peter Hoff
Duke STA 610

Introduction

Fixed and random effects

Model fitting

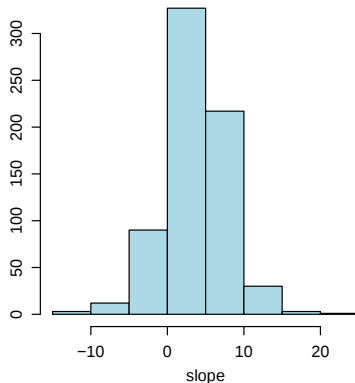
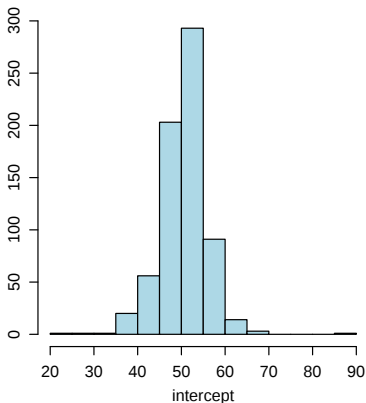
Group-level characteristics

General LME Model

Heterogeneity of $\hat{\beta}_j$'s for the NELS data

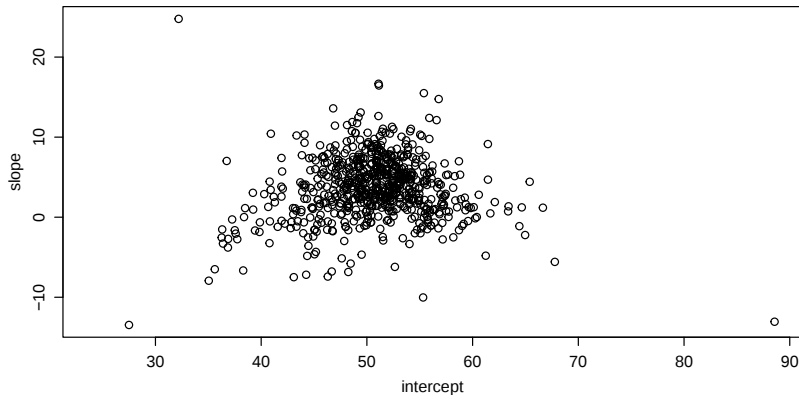
$$\hat{\beta}_j = (\mathbf{X}_j^T \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{y}_j$$

```
hist(BETA.OLS[,1]) hist(BETA.OLS[,2])
```



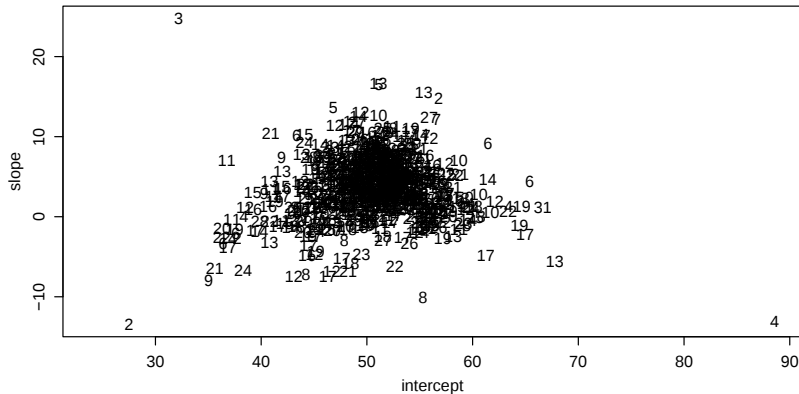
Heterogeneity of $\hat{\beta}_j$'s

```
plot(BETA.OLS)
```



$$\text{Var}[\hat{\beta}_j] = \sigma^2 (\mathbf{X}_j^T \mathbf{X}_j)^{-1}$$

Heterogeneity as a function of sample size



$$\text{Var}[\hat{\beta}_j] = \sigma^2(\mathbf{X}_j^T \mathbf{X}_j)^{-1}$$

Modeling heterogeneity

In the hierarchical normal model:

$$\theta_j = \{\mu_j, \sigma^2\}$$

$$y_{i,j} = \mu_j + \sigma^2, \{\epsilon_{i,j}\} \sim \text{i.i.d. normal}(\mu_j, \sigma^2)$$

$$\mu_1, \dots, \mu_m \sim \text{i.i.d. normal}(\mu, \tau^2)$$

What should we do for a hierarchical regression model?

$$\theta_j = \{\beta_j, \sigma^2\}$$

$$y_{i,j} = \beta_j^T \mathbf{x}_{i,j} + \epsilon_{i,j}, \{\epsilon_{i,j}\} \sim \text{i.i.d. normal}(0, \sigma^2)$$

$$\beta_1, \dots, \beta_m \sim \text{i.i.d. } p(\beta_j)$$

HLM

MVN model for across-group heterogeneity:

$$\beta_1, \dots, \beta_m \sim \text{i.i.d. multivariate normal}(\beta, \Psi)$$

The parameters in this model include

β , an across-group mean regression vector

Ψ , a covariance matrix describing the variability of the β_j 's around β .

Ad-hoc estimates

```
## rough estimate of beta
apply(BETA.OLS, 2, mean, na.rm=TRUE)

## (Intercept)          xj
## 50.618228      3.672483
```

This estimator of β equally weights all schools.
Generally, we want to assign a lower weight to schools with less data.

```
## rough estimate of Sigma_beta
cov(BETA.OLS, use="complete.obs")

##              (Intercept)          xj
## (Intercept)  26.795851  1.001585
## xj           1.001585 15.818939
```

This is a *very rough* estimate of $\Psi = \text{Var}[\beta_j]$:

- It ignores sample size differences;
- It ignores the variability of $\hat{\beta}_j$ around β_j .

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

Sample covariance of $\hat{\beta}_j$'s $\approx$ Ψ $+$ Estimation error

Fixed and random effects

Recall the following:

$$\mu_j \sim N(\mu, \tau^2) \Leftrightarrow \mu_j = \mu + a_j, \quad a_j \sim N(0, \tau^2)$$

Analogously,

$$\beta_j \sim N(\beta, \Psi) \Leftrightarrow \beta_j = \beta + \mathbf{a}_j, \quad \mathbf{a}_j \sim N(\mathbf{0}, \Psi)$$

Therefore, our hierarchical model says that

$$\begin{aligned} \mathbf{y}_j &= \mathbf{X}_j \beta_j + \epsilon_j \\ &= \mathbf{X}_j (\beta + \mathbf{a}_j) + \epsilon_j \\ &= \mathbf{X}_j \beta + \mathbf{X}_j \mathbf{a}_j + \epsilon_j \end{aligned}$$

- β is sometimes called a *fixed effect*, as it is fixed across all groups.
- $\mathbf{a}_j$ is sometimes called a *random effect*
“random” as it varies across groups, or
“random” if the groups were randomly sampled.

A model with fixed and random effects is called a *mixed-effects model*.

Within-group covariance

Recall the HNM:

$$y_{i,j} = \mu + a_j + \epsilon_{i,j}$$

What was the within-group covariance?

$$\begin{aligned}\text{Cov}[y_{i_1,j}, y_{i_2,j}] &= E[(y_{i_1,j} - \mu)(y_{i_2,j} - \mu)] \\ &= E[(a_j + \epsilon_{i_1,j})(a_j + \epsilon_{i_2,j})] \\ &= E[a_j^2] + 0 + 0 + 0 \\ &= \tau^2\end{aligned}$$

Within-group covariance, matrix form

More generally, we might want the *within-group covariance matrix*:

$$\mathbf{y}_j = \begin{pmatrix} y_{1,j} \\ \vdots \\ y_{n,j} \end{pmatrix} \quad \text{Cov}[\mathbf{y}_j] = \begin{pmatrix} \text{Var}[y_{1,j}] & \text{Cov}[y_{1,j}, y_{2,j}] & \cdots & \text{Cov}[y_{1,j}, y_{n,j}] \\ \text{Cov}[y_{1,j}, y_{2,j}] & \text{Var}[y_{2,j}] & \cdots & \text{Cov}[y_{2,j}, y_{n,j}] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[y_{1,j}, y_{n,j}] & \text{Cov}[y_{2,j}, y_{n,j}] & \cdots & \text{Var}[y_{n,j}] \end{pmatrix}$$

Our calculations have shown that for the HNM

$$\text{Cov}[\mathbf{y}_j] = \begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \cdots & \tau^2 \\ \vdots & & & \vdots \\ \tau^2 & \tau^2 & \cdots & \sigma^2 + \tau^2 \end{pmatrix}$$

Within-group covariance, matrix form

In general,

$$\text{Cov}[\mathbf{y}_j] = E[(\mathbf{y}_j - E[\mathbf{y}_j])(\mathbf{y}_j - E[\mathbf{y}_j])^T]$$

For the HLM,

$$\mathbf{y}_j - E[\mathbf{y}_j] = \mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} = \mathbf{X}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j,$$

so

$$\begin{aligned}\text{Cov}[\mathbf{y}_j] &= E[(\mathbf{X}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j)(\mathbf{X}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j)^T] \\ &= E[(\mathbf{X}_j\mathbf{a}_j\mathbf{a}_j^T \mathbf{X}_j^T] + E[\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_j^T] \\ &= \mathbf{X}_j\boldsymbol{\Psi}\mathbf{X}_j^T + \sigma^2\mathbf{I}\end{aligned}$$

$$\text{Cov}[y_{i1,j}, y_{i2,j}] = \mathbf{x}_{i1,j}^T \boldsymbol{\Psi} \mathbf{x}_{i2,j}$$

Dependence and conditional independence

Thus $p(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\Psi}, \sigma^2)$, unconditional on $\mathbf{a}_j$, is

$$\mathbf{y}_j \sim \text{multivariate normal}(\mathbf{X}_j \boldsymbol{\beta}, \mathbf{X}_j \boldsymbol{\Psi} \mathbf{X}_j^T + \sigma^2 \mathbf{I}).$$

On the other hand, conditional on $\mathbf{a}_j$,

$$\mathbf{y}_j \sim \text{multivariate normal}(\mathbf{X}_j \boldsymbol{\beta} + \mathbf{X}_j \mathbf{a}_j, \sigma^2 \mathbf{I}).$$

Dependence and conditional independence

Marginal dependence: If I don't know β_j (or $\mathbf{a}_j$), then knowing $y_{i_1,j}$ gives me a bit of information about β_j , which in turn gives me information about $y_{i_2,j}$, and so the observations are dependent: My information about $y_{i_2,j}$ depends on the value of $y_{i_1,j}$ if I don't know β_j .

Conditional independence: If I know β_j , then knowing $y_{i_1,j}$ doesn't give me any information about $y_{i_2,j}$, and so they are independent. My information about $y_{i_2,j}$ does not depend on the value of $y_{i_1,j}$ if I know β_j .

Note: Within-group covariance can be positive or negative, depending on $\mathbf{X}_j$.

Within-group covariance

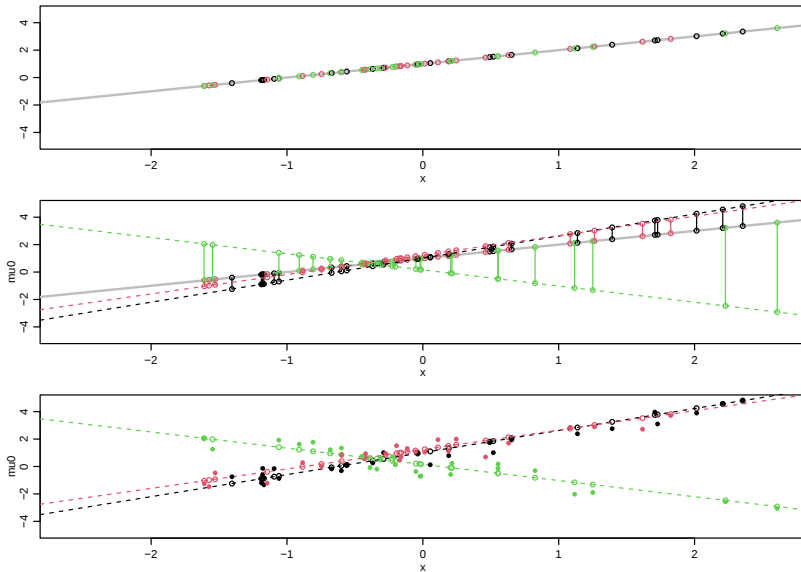
Consider the case that $\mathbf{x}_{i,j} = \{1, x_{i,j}\}$ and $\boldsymbol{\beta}_j = \{\beta_{0,j}, \beta_{1,j}\}$.

- $\mathbf{X}_j$ is $n_j \times 2$
- $\mathbf{X}_j \boldsymbol{\Psi} \mathbf{X}_j^T + \sigma^2 I$ is $n_j \times n_j$, the within-group covariance.

$$\begin{aligned}\text{Cov}[y_{1,j}, y_{2,j}] &= \mathbf{x}_{1,j}^T \boldsymbol{\Psi} \mathbf{x}_{2,j} \\ &= \boldsymbol{\Psi}_{1,1} + \boldsymbol{\Psi}_{1,2}(x_{1,j} + x_{2,j}) + \boldsymbol{\Psi}_{2,2}x_{1,j}x_{2,j} \\ &= \text{Var}[\beta_{0,j}] + \text{Var}[\beta_{1,j}]x_{1,j}x_{2,j} + \text{Cov}[\beta_{0,j}, \beta_{1,j}](x_{1,j} + x_{2,j})\end{aligned}$$

- Intercept variance positively correlates the observations within a group.
- Slope variance can lead to positive or negative correlation, depending on how close $x_{1,j}$ and $x_{2,j}$ are.

Sources of variation and correlation



Fitting a HLM

Assuming data are independent *across* groups, the likelihood at a value (β, Ψ, σ^2) can be computed as follows:

0. Set $ll = 0$.
1. Set $ll = ll + \text{ldmvnorm}(\mathbf{y}_1, \mathbf{X}_1\beta, \mathbf{X}_1\Psi\mathbf{X}_1 + \sigma^2\mathbf{I})$.
2. Set $ll = ll + \text{ldmvnorm}(\mathbf{y}_2, \mathbf{X}_2\beta, \mathbf{X}_2\Psi\mathbf{X}_2 + \sigma^2\mathbf{I})$.
- ⋮
- m . Set $ll = ll + \text{ldmvnorm}(\mathbf{y}_m, \mathbf{X}_m\beta, \mathbf{X}_m\Psi\mathbf{X}_m + \sigma^2\mathbf{I})$.

We can then numerically optimize the likelihood to find the MLEs.

Fitting the HLM with lmer

```
library(lme4)
fit.lme<-lmer( y.nels ~ ses.nels + (ses.nels | g.nels),REML=FALSE)
```

```
summary(fit.lme)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: y.nels ~ ses.nels + (ses.nels | g.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
##   g.nels      (Intercept)  12.223     3.496
##             ses.nels       1.515     1.231    0.11
##   Residual                        67.345     8.206
## Number of obs: 12974, groups: g.nels, 684
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  50.6767    0.1551   326.70
## ses.nels      4.3594    0.1231    35.41
##
## Correlation of Fixed Effects:
##              (Intr)
## ses.nels  0.007
```

Extracting results - fixed effects

```
### fixed effects
beta.hat<-fixef(fit.lme)
beta.hat
```

```
## (Intercept)      ses.nels
##    50.676702      4.359396
```

```
### variance-covariance of fixed effects estimates
VBETA<-vcov(fit.lme)
VBETA
```

```
## 2 x 2 Matrix of class "dpoMatrix"
##              (Intercept)      ses.nels
## (Intercept) 0.0240607576 0.0001310263
## ses.nels    0.0001310263 0.0151611175
```

```
### standard errors
sqrt(diag(VBETA))

## (Intercept)      ses.nels
##    0.1551153    0.1231305
```

```
### t-values
beta.hat/sqrt(diag(VBETA))

## (Intercept)      ses.nels
##    326.70343    35.40469
```

Extracting results - variance components

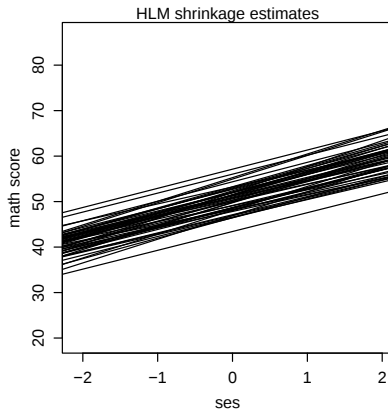
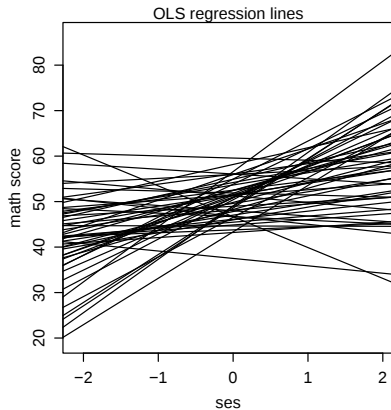
```
### within-group variance  
s2.hat<-sigma(fit.lme)^2
```

```
### across-group variance  
VarCorr(fit.lme)$g.nels  
  
##                (Intercept)  ses.nels  
## (Intercept)   12.2232568  0.4888068  
## ses.nels       0.4888068  1.5148390  
## attr(,"stddev")  
## (Intercept)    ses.nels  
##      3.496177    1.230788  
## attr(,"correlation")  
##                (Intercept)  ses.nels  
## (Intercept)    1.0000000  0.1135954  
## ses.nels       0.1135954  1.0000000
```

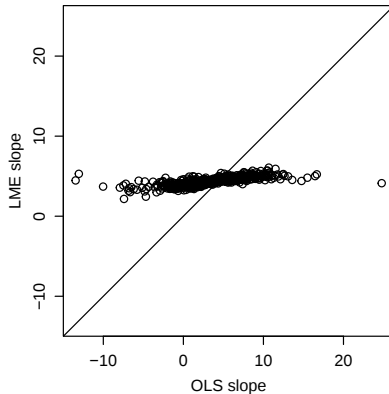
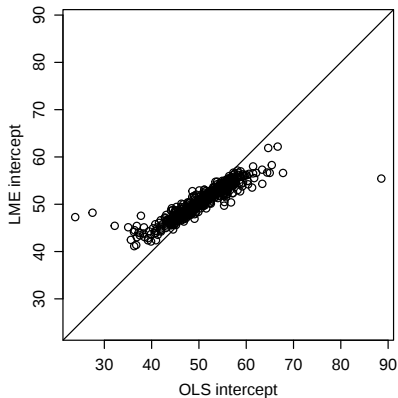
```
### remove the S4 ugliness  
VB<-matrix(VarCorr(fit.lme)$g.nels,2,2)  
  
VB  
  
##                [,1]      [,2]  
## [1,] 12.2232568  0.4888068  
## [2,]  0.4888068  1.5148390
```

Random effects estimates

```
B.LME<-as.matrix(ranef(fit.lme)$g.nels)
BETA.LME<-sweep( B.LME , 2 , beta.hat, "+" )
```



Range of shrinkage estimates



Formula for shrinkage estimates

Intuitively:

$$\tilde{\beta}_j = w_j \hat{\beta}_j + (1 - w_j) \hat{\beta}$$

where w_j depends on Ψ and $\sigma^2(\mathbf{X}_j^T \mathbf{X}_j)^{-1}$:

- w_j is big if $\sigma^2(\mathbf{X}_j^T \mathbf{X}_j)^{-1}$ small compared to Ψ ;
- w_j is small if $\sigma^2(\mathbf{X}_j^T \mathbf{X}_j)^{-1}$ large compared to Ψ .

This is almost right. Averaging has to be done using matrices. The BLUP is:

$$\tilde{\beta}_j = \left(\mathbf{X}_j^T \mathbf{X}_j / \sigma^2 + \Psi^{-1} \right)^{-1} \left(\mathbf{X}_j \mathbf{y}_j / \sigma^2 + \Psi^{-1} \beta \right)$$

In practice, σ^2, Ψ, β are usually replaced with $\hat{\sigma}^2, \hat{\Psi}, \hat{\beta}$.

Quiz: How does $\tilde{\beta}_j$ vary with $\mathbf{X}_j$, σ^2 and Ψ ?

Derivation of shrinkage formula

- $\hat{\beta}_j | \beta_j \sim N(\beta_j, \sigma^2(\mathbf{X}_j^\top \mathbf{X}_j)^{-1})$
- $\beta_j \sim N(\beta, \Psi)$

Then Bayes rule says $\beta_j \sim N(\mathbf{m}, \mathbf{V})$ where

$$\mathbf{V} = (\mathbf{X}_j^\top \mathbf{X}_j / \sigma^2 + \Psi^{-1})^{-1}$$

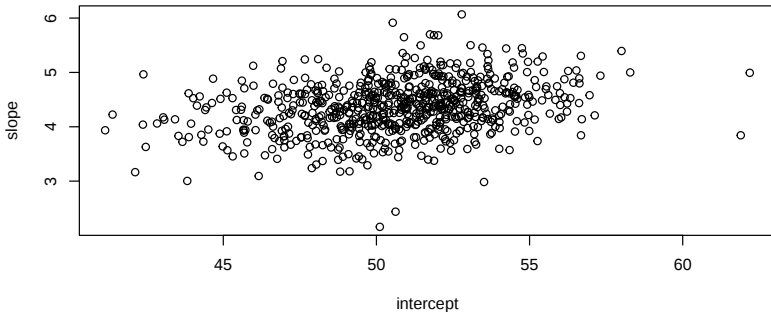
$$\mathbf{m} = V(\mathbf{X}_j^\top \mathbf{y}_j / \sigma^2 + \Psi^{-1} \beta)$$

The BLUP/Bayes estimator is the conditional expectation:

$$\tilde{\beta}_j = \left(\mathbf{X}_j^\top \mathbf{X}_j / \sigma^2 + \Psi^{-1} \right)^{-1} \left(\mathbf{X}_j \mathbf{y}_j / \sigma^2 + \Psi^{-1} \beta \right)$$

Macro-level effects

LME regression estimates:



Questions:

- What kind of schools have big intercepts?
- What kind of schools have big slopes?

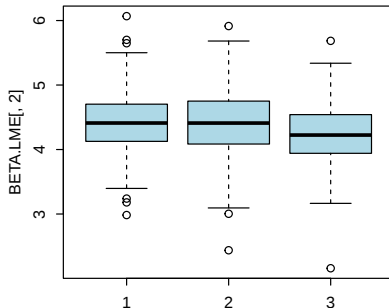
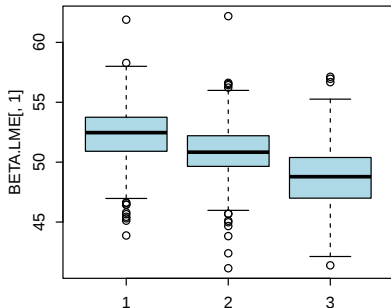
Can we relate *macro-level parameters* to *macro-level effects* ?

Macro-level effects

```
### FLP variable
flp.school<-tapply( flp.nels , g.nels, mean)
table(flp.school)

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

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



Macro-level effects

It seems that $\beta_{0,j}$ and possibly $\beta_{1,j}$ are associated with flp_j .

- Testing: Is there evidence for the association?
- Estimation: What is the association?

These questions can be addressed by expanding the model:

Old model:

$$\begin{aligned}y_{i,j} &= \beta_{0,j} + \beta_{1,j} \times ses_{i,j} + \epsilon_{i,j} \\ &= (\beta_0 + a_{0,j}) + (\beta_1 + a_{1,j}) \times ses_{i,j} + \epsilon_{i,j}\end{aligned}$$

New model:

$$\begin{aligned}y_{i,j} &= \beta_{0,j} + \beta_{1,j} \times ses_{i,j} + \epsilon_{i,j} \\ &= (\beta_{00} + \beta_{01} \times flp_j + a_{0,j}) + (\beta_{10} + \beta_{11} \times flp_j + a_{1,j}) \times ses_{i,j} + \epsilon_{i,j}\end{aligned}$$

Note that under this model,

- The intercept for school j is $\beta_{0,j} = (\beta_{00} + \beta_{01} \times flp_j + a_{0,j})$
- The slope for school j is $\beta_{1,j} = (\beta_{10} + \beta_{11} \times flp_j + a_{1,j})$

(Alternatively, we could treat flp_j as a categorical variable)

Macro-level fixed effects

$$\begin{aligned} y_{i,j} &= \beta_{0,j} + \beta_{1,j} \times \text{ses}_{i,j} + \epsilon_{i,j} \\ &= (\beta_{00} + \beta_{01} \times \text{flp}_j + a_{0,j}) + (\beta_{10} + \beta_{11} \times \text{flp}_j + a_{1,j}) \times \text{ses}_{i,j} + \epsilon_{i,j} \end{aligned}$$

- β_{01} represents the macro effect of flp_j on the intercept/mean in group j
- β_{11} represents the macro effect of flp_j on the slope with $\text{ses}_{i,j}$ in group j

Note: α_0 and α_1 do not vary across groups. If they did, they would be confounded with $a_{0,j}$ and $a_{1,j}$.

Note: As they are fixed across groups, they are in fact *fixed effects*:

Macro-level fixed effects

$$y_{i,j} = (\beta_{00} + \beta_{01} \times flp_j + a_{0,j}) + (\beta_{10} + \beta_{11} \times flp_j + a_{1,j}) \times ses_{i,j} + \epsilon_{i,j}$$

Rearranging, we get

$$y_{i,j} = \beta_{00} + \beta_{01} \times flp_j + \beta_{10} \times ses_{i,j} + \beta_{11} \times flp_j \times ses_{i,j} + \\ a_{0,j} + a_{1,j} \times ses_{i,j} + \\ \epsilon_{i,j}$$

Fixed effects regression: $\beta_{00} + \beta_{01} \times flp_j + \beta_{10} \times ses_{i,j} + \beta_{11} \times flp_j \times ses_{i,j}$

Random effects regression: $a_{0,j} + a_{1,j} \times ses_{i,j}$

Note:

- The predictors for the two regressions are different.
- Macro-effects do not appear in the random effects regression.

Mixed-effects model

$$y_{i,j} = \beta_{00} + \beta_{01} \times flp_j + \beta_{11} \times ses_{i,j} + \beta_{11} \times flp_j \times ses_{i,j} + \\ a_{0,j} + a_{1,j} \times ses_{i,j} + \\ \epsilon_{i,j}$$

We would like to avoid these double subscripts.

We rewrite the model as

$$y_{i,j} = \beta_0 + \beta_1 \times flp_j + \beta_2 \times ses_{i,j} + \beta_3 \times flp_j \times ses_{i,j} + \\ a_{0,j} + a_{1,j} \times ses_{i,j} + \\ \epsilon_{i,j} \\ = \beta^T \mathbf{x}_{i,j} + \mathbf{a}_j^T \mathbf{z}_{i,j} + \epsilon_{i,j}$$

- $\mathbf{x}_{i,j} = (1, flp_j, ses_{i,j}, flp_j \times ses_{i,j})$
- $\mathbf{z}_{i,j} = (1, ses_{i,j})$

Ask yourself: Could flp_j go in $\mathbf{z}_{i,j}$? Why or why not?

Group-level representation

Micro-level representation:

$$y_{i,j} = \boldsymbol{\beta}^T \mathbf{x}_{i,j} + \mathbf{a}_j^T \mathbf{z}_{i,j} + \epsilon_{i,j}$$

Combining observations within a group:

$$\begin{pmatrix} y_{1,j} \\ \vdots \\ y_{n,j} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{1,j} \rightarrow \\ \vdots \\ \mathbf{x}_{n,j} \rightarrow \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \mathbf{z}_{1,j} \rightarrow \\ \vdots \\ \mathbf{z}_{n,j} \rightarrow \end{pmatrix} \begin{pmatrix} \mathbf{a}_{1,j} \\ \vdots \\ \mathbf{a}_{p,j} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,j} \\ \vdots \\ \epsilon_{n,j} \end{pmatrix}$$

Two-level HLM: General form

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{a}_j + \boldsymbol{\epsilon}_j$$

Note: This formulation allows the *fixed effects predictors* to be different from the *random effects predictors*.

Two-level HLM: General form

This is the general form of a *two-level* hierarchical linear model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j$$

where $\mathbf{a}_j$ and $\boldsymbol{\epsilon}_j$ are multivariate normal.

- $\boldsymbol{\beta}$ are the *fixed effects coefficients*;
- $\mathbf{X}_j$ is the *design matrix for the fixed effects*.
- $\mathbf{a}_j$ are the *random effects coefficients for group j* ;
- $\mathbf{Z}_j$ is the *design matrix for the random effects*.

Variance components

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j$$

$$E \begin{bmatrix} \mathbf{a}_j \\ \boldsymbol{\epsilon}_j \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \text{Cov} \begin{bmatrix} \mathbf{a}_j \\ \boldsymbol{\epsilon}_j \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{bmatrix}.$$

Across-group heterogeneity: $\boldsymbol{\Psi}$ is the variance-covariance in $\mathbf{a}_1, \dots, \mathbf{a}_m$.

Within-group heterogeneity: $\boldsymbol{\Sigma}$ is the variance-covariance of $y_{1,j}, \dots, y_{n_j,j}$.

Note: We should write $\boldsymbol{\Sigma}_j$ instead of $\boldsymbol{\Sigma}$, as

$$\text{Cov}[\mathbf{y}_j] = \text{Cov}[\boldsymbol{\epsilon}_j] = \boldsymbol{\Sigma}_j \text{ is an } n_j \times n_j \text{ matrix.}$$

Note: In the examples so far,

$$\boldsymbol{\Sigma}_j = \sigma^2 \mathbf{I}_{n_j}.$$

Question: What other forms for $\boldsymbol{\Sigma}_j$ might be useful?

Example: One-way random effects model, aka the HNM

$$y_{i,j} = \mu + a_j + \epsilon_{i,j}$$

$$\{a_j\} \sim iid N(0, \tau^2)$$

$$\{\epsilon_{i,j}\} \sim iid N(0, \sigma^2)$$

Exercise: Express this model as $\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j$

- Regression parameters:

$$\boldsymbol{\beta} = \mu, \quad \mathbf{a}_j = a_j$$

- Design matrices:

$$\mathbf{X}_j = \mathbf{Z}_j = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{for each } j \in \{1, \dots, m\}$$

- Covariance terms:

$$\boldsymbol{\Psi} = \text{Var}[a_j] = \tau^2, \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$$

Exercise: Check your work by going in reverse.

Example: One-way random effects model, aka the HNM

```
fit.0<-lmer(y.nels~ 1 + (1|g.nels), REML=FALSE)
```

```
summary(fit.0)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: y.nels ~ 1 + (1 | g.nels)
##
##          AIC          BIC    logLik deviance df.resid
##  93919.3   93941.7 -46956.6  93913.3    12971
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8112 -0.6534  0.0093  0.6732  4.6999
##
## Random effects:
##   Groups      Name            Variance Std.Dev.
##   g.nels  (Intercept)    23.63      4.861
##   Residual                    73.71      8.585
## Number of obs: 12974, groups:  g.nels, 684
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  50.9391      0.2026   251.4
```

Group-specific linear regression

$$\begin{aligned}y_{i,j} &= \beta^T \mathbf{x}_{i,j} + \mathbf{a}_j^T \mathbf{x}_{i,j} + \epsilon_{i,j} \\ \{\mathbf{a}_j\} &\sim \text{iid } N(0, \Psi) \\ \{\epsilon_{i,j}\} &\sim \text{iid } N(0, \sigma^2)\end{aligned}$$

Exercise: Express this model as $\mathbf{y}_j = \mathbf{X}_j \beta + \mathbf{Z}_j \mathbf{a}_j + \epsilon_j$

- Design matrices:

$$\mathbf{X}_j = \mathbf{Z}_j = \begin{bmatrix} \mathbf{x}_{1,j} \rightarrow \\ \vdots \\ \mathbf{x}_{n_j,j} \rightarrow \end{bmatrix} \quad \text{for each } j \in \{1, \dots, m\}$$

- Regression parameters:

$$\beta = \beta, \quad \mathbf{a}_j = \mathbf{a}_j$$

- Covariance terms:

$$\Psi = \text{Cov}[\mathbf{a}_j], \quad \Sigma = \sigma^2 \mathbf{I}$$

This is just a special case where $\mathbf{X}_j = \mathbf{Z}_j$.

Group-specific linear regression

```
fit.1<-lmer(y.nels~ ses.nels + (ses.nels|g.nels), REML=FALSE)
```

```
summary(fit.1)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: y.nels ~ ses.nels + (ses.nels | g.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
##   g.nels      (Intercept) 12.223   3.496
##             ses.nels     1.515   1.231   0.11
##   Residual                67.345   8.206
## Number of obs: 12974, groups: g.nels, 684
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  50.6767    0.1551  326.70
## ses.nels      4.3594    0.1231   35.41
##
## Correlation of Fixed Effects:
##              (Intr)
## ses.nels  0.007
```

General LME

$$\begin{aligned}y_{i,j} &= \beta^T \mathbf{x}_{i,j} + \mathbf{a}_j^T \mathbf{z}_{i,j} + \epsilon_{i,j} \\ \{\mathbf{a}_j\} &\sim iid N(0, \Psi) \\ \{\epsilon_j\} &\sim iid N(0, \Sigma)^*\end{aligned}$$

* modulo different sample sizes.

Review of benefits of model extension:

- Group-specific regressors should appear in $\mathbf{X}_j$ but not $\mathbf{Z}_j$;
- If $\{a_{k,1}, \dots, a_{k,m}\}$ shows little variability ($\psi_{k,k}$ small), we may want to remove $x_{i,j,k}$ from the random effects model, and include it as a fixed effect only.
- Within-group covariances other than $\Sigma = \sigma^2 \mathbf{I}$ might be useful:
 - Σ with temporal correlation for longitudinal/panel data;
 - Unrestricted Σ for correlation but unordered outcomes (teeth, eg.)

General LME

```
fit.2<-lmer(y.nels~ flp.nels + ses.nels + flp.nels*ses.nels + (ses.nels | g.nels), REML=FALSE)
summary(fit.2)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: y.nels ~ flp.nels + ses.nels + flp.nels * ses.nels + (ses.nels |
##      g.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
##   g.nels      (Intercept)    9.012     3.002
##              ses.nels       1.571     1.254    0.06
##   Residual                67.260     8.201
## Number of obs: 12974, groups:  g.nels, 684
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    55.3975    0.3860 143.524
## flp.nels        -2.4062    0.1819 -13.230
## ses.nels         4.4909    0.3326  13.500
## flp.nels:ses.nels -0.1931    0.1587  -1.216
##
## Correlation of Fixed Effects:
##              (Intr) flp.nl ss.nls
## flp.nels      -0.930
## ses.nels      -0.158  0.088
## flp.nls:ss.   0.086 -0.007 -0.926
```