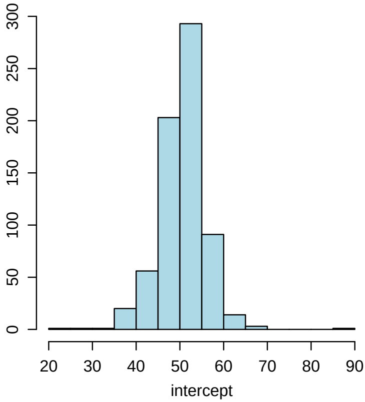
# Linear Mixed Effects Models

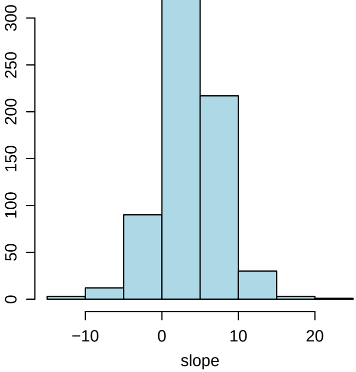
Peter Hoff
Duke STA 610

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

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

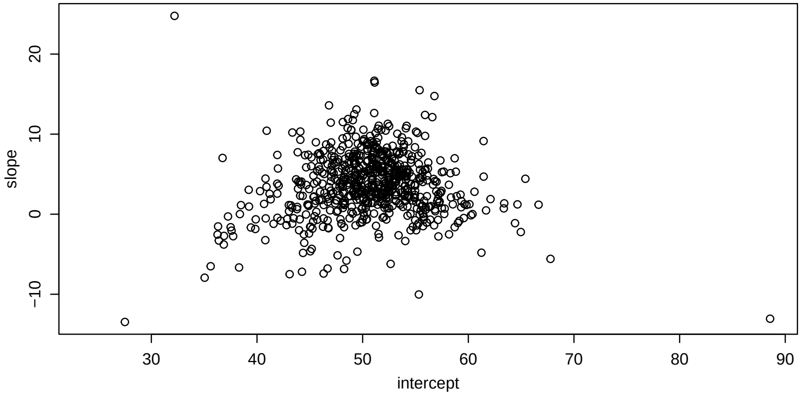
```
hist(BETA.OLS[,1]) hist(BETA.OLS[,2])
```



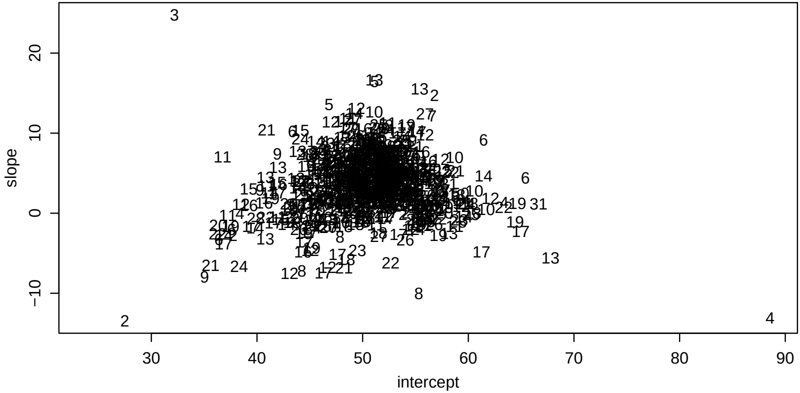
# Heterogeneity of $\hat{\boldsymbol{\beta}}_j$'s

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




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

# Heterogeneity as a function of sample size




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

# Modeling heterogeneity

In the hierarchical normal model:

$$y_{i,j} = \theta_j + \epsilon_{i,j},$$
$$\{\epsilon_{i,j}\} \sim \text{i.i.d normal}(\mu_j, \sigma^2),$$
$$\theta_1, \ldots, \theta_m \sim \text{i.i.d. normal } (\mu, \tau^2).$$

What should we do for a hierarchical regression model?

$$y_{i,j} = \boldsymbol{\beta}_j^T \mathbf{x}_{i,j} + \epsilon_{i,j},$$
$$\{\epsilon_{i,j}\} \sim \text{i.i.d. normal}(0, \sigma^2),$$
$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \sim \text{i.i.d. } P.$$

What should $P$ be?

# HLM

**MVN model for across-group heterogeneity:**

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \sim \text{i.i.d. multivariate normal}(\boldsymbol{\beta}, \Psi)$$

The parameters in this model include

$\boldsymbol{\beta}$, an across-group mean regression vector

$\Psi$, a covariance matrix describing the variability of the $\boldsymbol{\beta}_j$'s around $\boldsymbol{\beta}$.

# Ad-hoc estimates

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

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

This estimator of $\boldsymbol{\beta}$ is unbiased, but not efficient.
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 $\Psi = \text{Var}[\boldsymbol{\beta}_j]$:

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

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

$$\text{Sample covariance of } \hat{\boldsymbol{\beta}}_j\text{'s} \approx \qquad \Psi \qquad + \qquad \text{Estimation error}$$

# Fixed and random effects

Recall the following:

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

Analogously,

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

Therefore, our hierarchical model says that

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

- $\boldsymbol{\beta}$ is sometimes called a *fixed effect*, as it is fixed across all groups.
- $\boldsymbol{a}_j$ 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?

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

# Within-group covariance, matrix form

We will need the *within-group covariance matrix* to compute the likelihood:

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

Our calculations have shown that for the HNM

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

# Within-group covariance, matrix form

In general,

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

For the HLM,

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

so

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

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

## Dependence and conditional independence

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

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

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 $\boldsymbol{\beta}_j$ (or $\mathbf{a}_j$), then knowing $y_{i_1,j}$ gives me a bit of information about $\boldsymbol{\beta}_j$, 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 $\boldsymbol{\beta}_j$.

**Conditional independence:** If I know $\boldsymbol{\beta}_j$, 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 $\boldsymbol{\beta}_j$.

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

# Within-group covariance

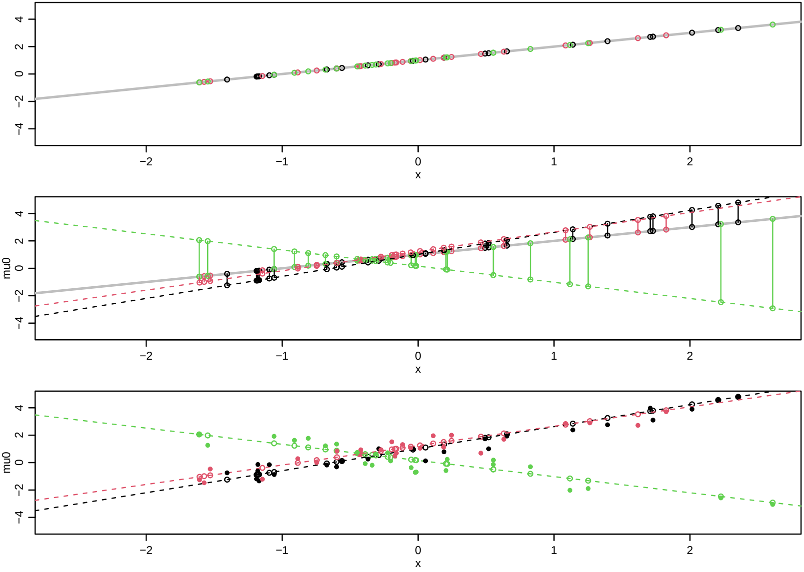
Consider the case that $\mathbf{x}_{i,j} = \{1, x_{i,j}\}$ and $\boldsymbol{\beta}_j = \{\beta_{0,j}, \beta_{1,j}\}$.

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

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

- Intercept variance positivly 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 $(\boldsymbol{\beta}, \Psi, \sigma^2)$ can be computed as follows:

0. Set `ll`$=0$.
1. Set `ll`$=$ `ll` $+$ `ldmvnorm(` $\mathbf{y}_1$ , $\mathbf{X}_1\boldsymbol{\beta}$ , $\mathbf{X}_1\Psi\mathbf{X}_1+\sigma^2\mathbf{I}$`)`.
2. Set `ll`$=$ `ll` $+$ `ldmvnorm(` $\mathbf{y}_2$ , $\mathbf{X}_2\boldsymbol{\beta}$ , $\mathbf{X}_2\Psi\mathbf{X}_2+\sigma^2\mathbf{I}$`)`.

$\vdots$

m. Set `ll`$=$ `ll` $+$ `ldmvnorm(` $\mathbf{y}_m$ , $\mathbf{X}_m\boldsymbol{\beta}$ , $\mathbf{X}_m\Psi\mathbf{X}_m+\sigma^2\mathbf{I}$`)`.

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

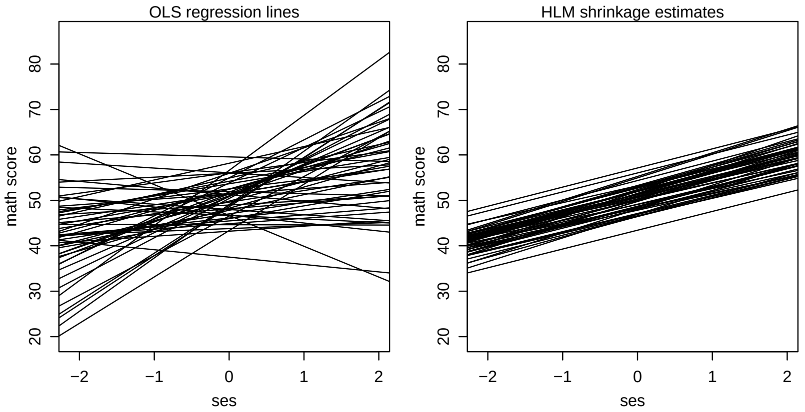
```r
### within-group variance
s2.hat<-sigma(fit.lme)^2
```

```r
### 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
```

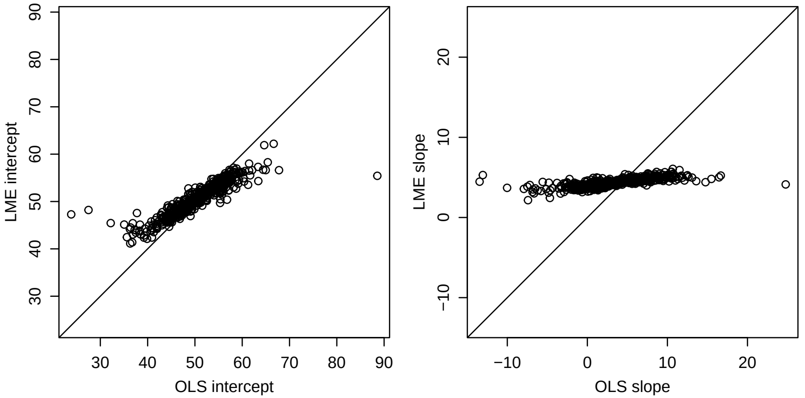
```r
### remove the S4 ugliness
VB<-matrix(VarCorr(fit.lme)$g.nels,2,2)

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:** Let $\hat{\boldsymbol{\beta}}_j = (\mathbf{X}_j^\top \mathbf{X}_j)^{-1}\mathbf{X}_j^\top \mathbf{y}_j$.

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

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

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

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

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

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

**Quiz:** How does $\tilde{\boldsymbol{\beta}}_j$ vary with $\mathbf{X}_j$, $\sigma^2$ and $\Psi$?

# Derivation of shrinkage formula

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

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

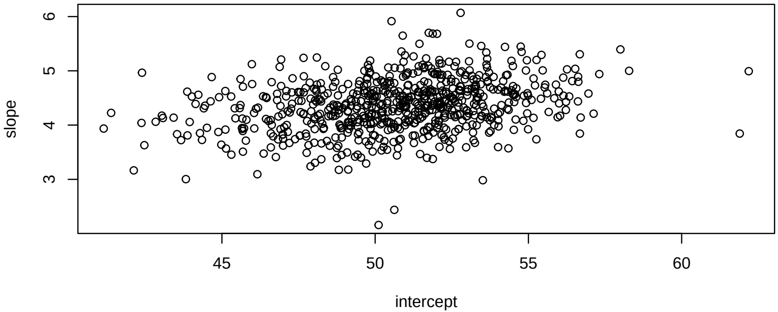
$$\mathbf{V} = (\mathbf{X}_j^\top \mathbf{X}_j / \sigma^2 + \Psi^{-1})^{-1}$$
$$\mathbf{m} = V(\mathbf{X}_j^\top \mathbf{y}_j / \sigma^2 + \Psi^{-1} \boldsymbol{\beta})$$

The BLUP/Bayes estimator is the conditional expectation:

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

# Macro-level effects

**LME regression estimates:**



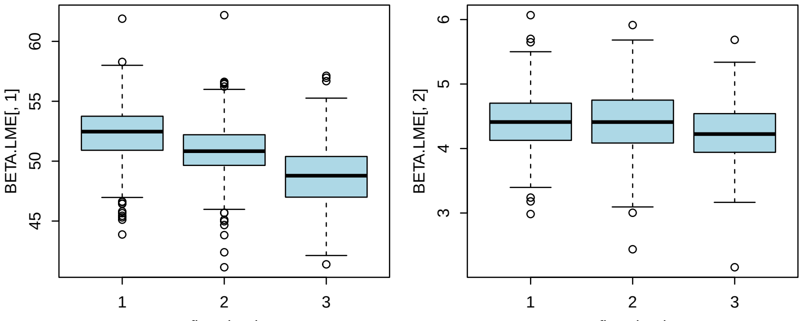

**Questions:**

- 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 $\text{flp}_j$.

- 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

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

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

**Note:** $\beta_{01}$ and $\beta_{11}$ do not vary across groups. If they did, they would be confounded with $a_{0,j}$ and $a_{1,j}$.

**Note:** As $\beta_{01}$ and $\beta_{11}$ are fixed across groups, they are called *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

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

**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

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

We would like to avoid these double subscripts.

We rewrite the model as

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

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

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:**

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

**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*.

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

# Variance components

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

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

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

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

**Note:** We should write $\Sigma_j$ instead of $\Sigma$, as

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

**Note:** In the examples so far,

$$\Sigma_j = \sigma^2 \mathrm{I}_{n_j}.$$

**Question:** What other forms for $\Sigma_j$ 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:

$$\beta = \mu \; , \; a_j = a_j$$

- 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:

$$\Psi = \mathrm{Var}[a_j] = \tau^2 \; , \; \Sigma = \sigma^2 \mathbf{I}$$

**Exercise:** Check your work by going in reverse.

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

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

```r
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

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

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

- Design matrices:

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

- Regression parameters:

$$\boldsymbol{\beta} = \boldsymbol{\beta} \; , \; \mathbf{a}_j = \mathbf{a}_j$$

- Covariance terms:

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

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

# Group-specific linear regression

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

```r
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

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

* 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 $\{a_{k,1}, \ldots, a_{k,m}\}$ 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:
  - $\Sigma$ with temporal correlation for longitudinal/panel data;
  - Unrestricted $\Sigma$ 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
```