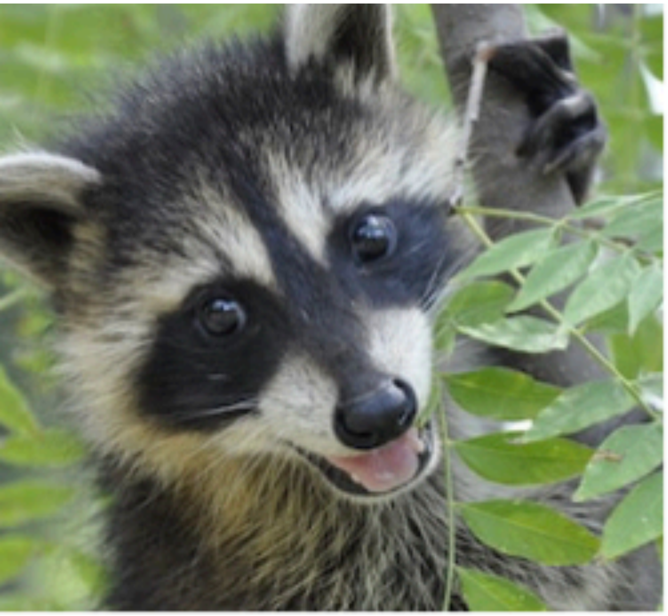


Hi! I'm back! Please get ready for class by:

1) Downloading the R scripts for lecture into a folder with the hsb data

2) Open and run the script up to the "Stop Here" note.

3) If bored, modify the code to show 6 randomly selected schools instead of 3

4) If still bored, remove "method="lm"" from the ggplot code.

Code tip:
To step through code, try hitting "Command-Enter" on the first line of code. It should run it and auto-advance to the next line. You can walk through code line by line that way, seeing how it works.

S-043/Stat-151
Analysis for Clustered and Longitudinal Data
(Multilevel & Longitudinal Models)

Unit 2, Lecture 3
Random Slope Models

Instructor: Prof. Luke W. Miratrix

lmiratrix@g.harvard.edu

Larsen 603

Todays Goals

★ Practice model building with an in class exercise

★ Learn about random slope models!
- How to write them down
- How to fit them in R
- How to interpret their coefficients
- Some notational conventions we need to talk about them

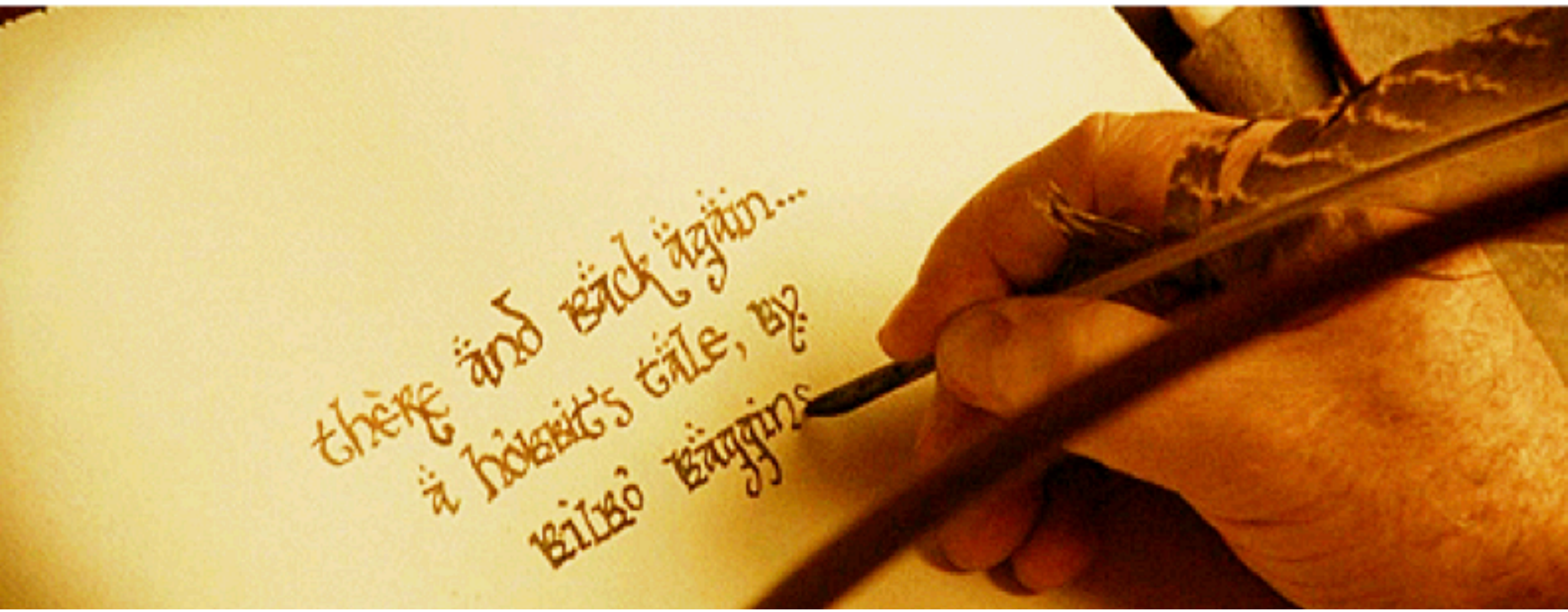
★ Further discussion of Empirical Bayes estimates
- How we can get an estimated regression line for each school
- How they are “over-shrunk” which is better than “over-dispersed.”

Announcement: Canvas going private

We think the enrollment stuff is resolved so are switching canvas to enrolled folks only.

Sing out if it turns out that this impacts you and we will fix.

There and back again: Translating models to R & identifying R output



In-class exercise

1. Write down a mathematical model (in two-level notation) for investigating
   A. whether there is a gender gap in math achievement.
2. Make your model an R command and fit it (use the “female” variable.
3. Identify the estimated values for all parameters in the original model.
4. Extend your model to answer:
   B. Is the gap is different for Catholic and public schools?
5. (Extension) Does the gap vary by school?
   (We will need the rest of the lecture do do this.)

{Collective in class R coding.}

Random slopes:
letting slopes vary

Why should intercepts have all the fun?

They shouldn’t. that wouldn’t be fair

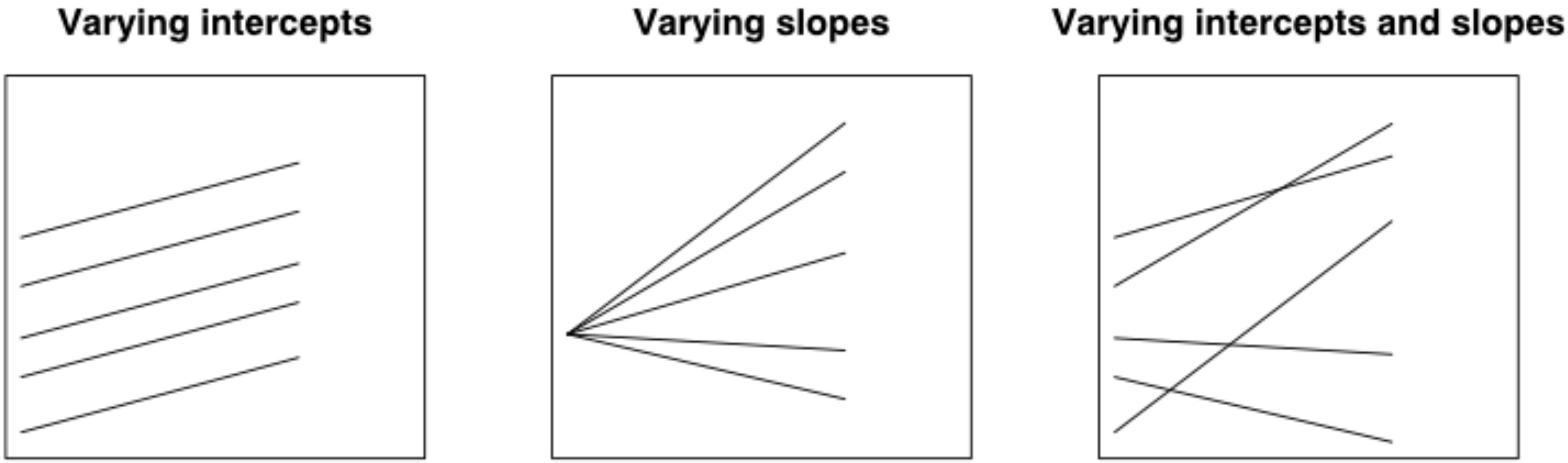


Figure 11.1 *Linear regression models with (a) varying intercepts ($y = \alpha_j + \beta x$), (b) varying slopes ($y = \alpha + \beta_j x$), and (c) both ($y = \alpha_j + \beta_j x$). The varying intercepts correspond to group indicators as regression predictors, and the varying slopes represent interactions between x and the group indicators.*

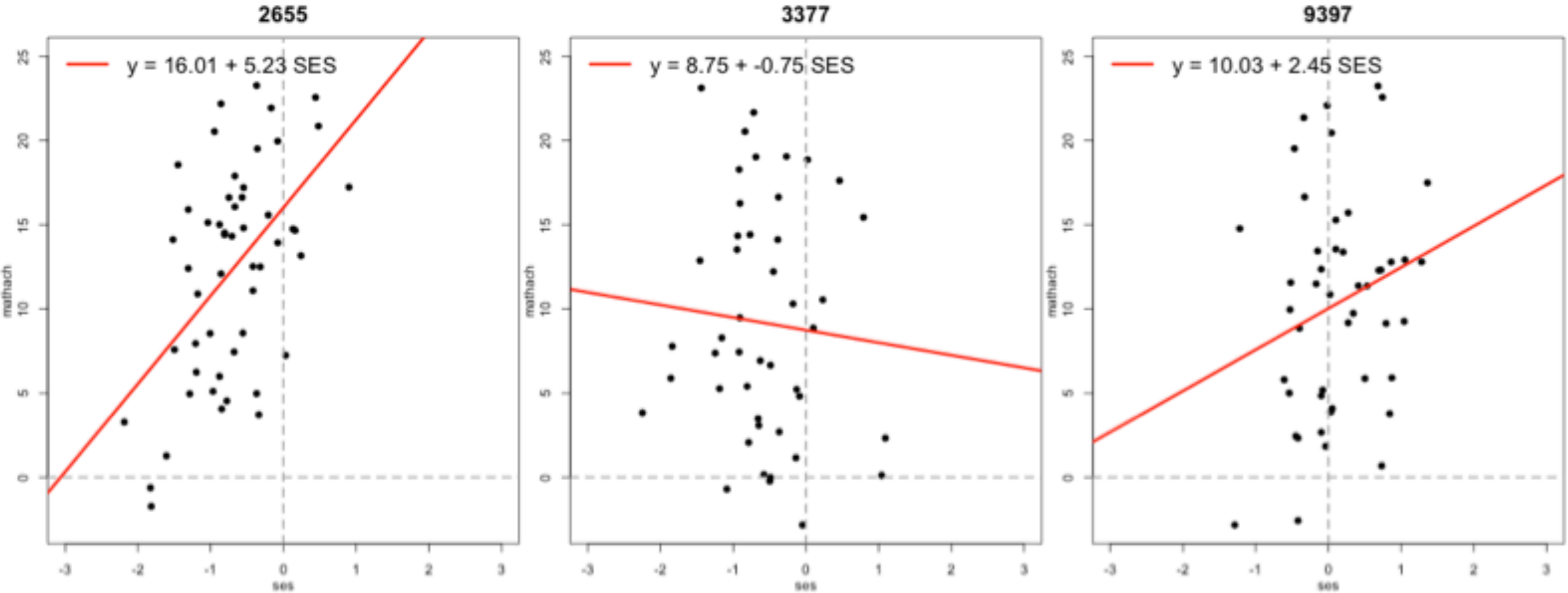
We did that

This one is just weird

(Ok, this is just a special case)

Today we are doing this

Three Schools from HS&B



Slopes seem to vary by school.

Intercepts do also.

So does the distribution of SES

- » Some schools wealthier than others.

Clearly lots of noise.

RQ: Does the **relationship** of SES and achievement vary by school? How much, and does anything predict this?

1. Here we are asking about how a relationship between variables inside a school varies across schools.

2. Also it implicitly asks whether *inside a school* do we tend to see a relationship between SES and math achievement?

The Random Slope Model

Each school has its *own* intercept and slope

Level 1

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

Two things vary by school

We then have one equation for the intercept and one for the slope

Average intercept (across schools)

$$\alpha_j = \mu_\alpha + u_{0j}$$

Level 2

$$\beta_j = \mu_\beta + u_{1j}$$

Average slope (across schools)

Now we have two level 2 random effects…

Now we see why those annoying subscripts on mu (μ)

The Random Effect Distribution

We say the intercept and slope follow a "multivariate normal distribution"

The slopes and intercepts are "tied together"

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ & \sigma_\beta^2 \end{pmatrix}\right]$$

our random effects

zero-centered

How much intercepts vary

How much slopes vary

How intercepts and slopes co-vary. This is covariance, not correlation

The covariance term lets the intercept and slope co-vary: maybe schools with high intercepts have low slopes, for example.

Some commentary

1) Both the slope and intercept has its own variance, but they also covary.

What "Covary" means: Having an unusual intercept can be predictive of having an unusual slope.

2) Sometimes we write the entire level 2 model like this:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \quad \Sigma\right]$$

Correlation vs. Covariance

The random intercept model is in terms of **Covariance**

Law:

$$cor(A, B) = \frac{cov(A, B)}{sd(A)sd(B)}$$

Covariance is correlation on the scale of the original variables. **Correlation** *is "unitless."*

Same thing in math:

$$\rho = \frac{cov(\alpha_j, \beta_j)}{\sigma_\alpha \sigma_\beta}$$

Equivalent Model: Collapsed version

We can write our model as a single equation with all the parts in it

$$y_i = \mu_\alpha + u_{0j[i]} + \left(\mu_\beta + u_{1j[i]}\right) x_i + \varepsilon_i$$

$$= \mu_\alpha + \mu_\beta x_i + \left(u_{0j[i]} + u_{1j[i]} x_i + \varepsilon_i\right)$$

$$\varepsilon_i \sim N(0, \sigma_y^2)$$

The random elements are the same:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right]$$

We just plugged in our level 2 equations for our α_j and β_j

This notation corresponds with lmer()

Case Study:
Random Slopes in HS&B

The Model

This is the same model we just saw, but the parameters are now different.

Level 1

$$y_{ij} = \beta_{0j} + \beta_{1j} ses_{ij} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_y^2)$$

This is one of the canonical ways of writing our model. The gamma (γ) allow for ready extension to more complex models with covariates.

Level 2

$$\beta_{0j} = \gamma_0 + u_{0j}$$

$$\beta_{1j} = \gamma_1 + u_{1j}$$

Also note using only β for level 1.

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ & \sigma_1^2 \end{pmatrix}\right]$$

Fitting the Random-Slope Model

```
> mlm= lmer( mathach ~ 1 + ses + (1+ses|id), data=dat )

> display( mlm )
lmer(formula = mathach ~ ses + (ses | id), data = dat)
            coef.est coef.se
(Intercept) 12.67     0.19
ses          2.39     0.12

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 2.20
          ses         0.64     -0.11
 Residual             6.07
---
number of obs: 7185, groups: id, 160
AIC = 46652.4, DIC = 46632.5
deviance = 46636.5
```

Correlation not Covariance!
Covariance would be
-0.11 X 2.2 X 0.64

Interpretations

- $\gamma_{00} = 12.7$, so in the average school district, students with mean SES (SES = 0) score 12.7, on average
- $\gamma_{10} = 2.4$, so in the average school district, a one-unit difference in SES predicts a 2.4 point difference in achievement
- $\sigma_0 = 2.2$
- $\sigma_1 = 0.6$
- $\sigma_\varepsilon = 6.1$
- $\rho = -0.1$, so there is (or may be) a very weak negative association between intercept and slope; schools with large intercepts (may) have smaller slopes

These are standard deviations, ***not*** variances; looks like there's school level variance in intercept and slope, and of course residual variability

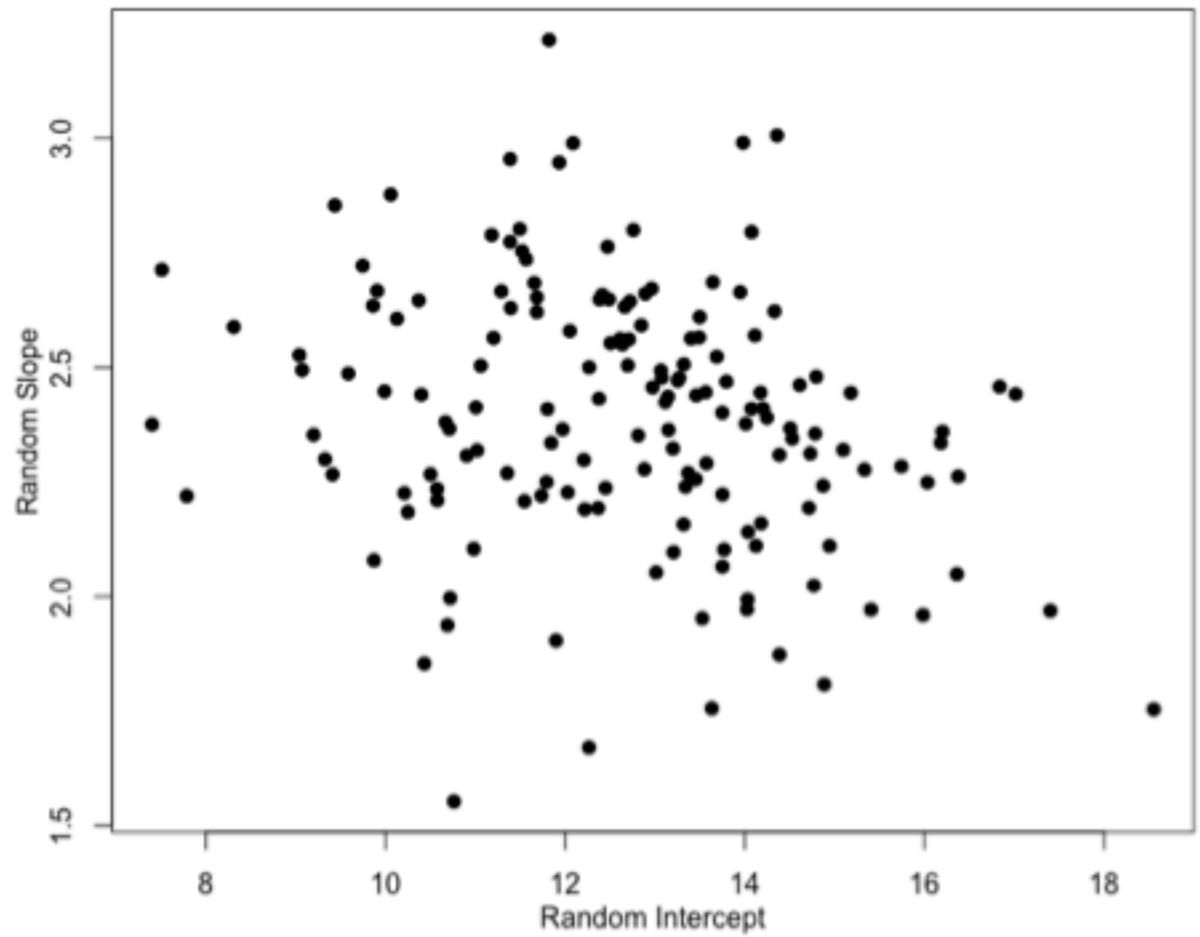
Get school-level estimates

```
> re = coef( mlm )$id
> names( res ) = c( "alpha", "beta" )
> res$id = rownames( re )
> head( re )
         alpha     beta   id
1224 11.060027 2.504085 1224
1288 13.073613 2.477931 1288
1296  9.197303 2.352988 1296
1308 14.384640 2.309299 1308
1317 12.449141 2.237438 1317
1358 11.520356 2.753442 1358
```

We generally want to name our columns

```
> cor( re$alpha, re$beta )
[1] -0.2343
```

Note this is a bit larger than the model estimate; this often happens due to "over shrinkage"



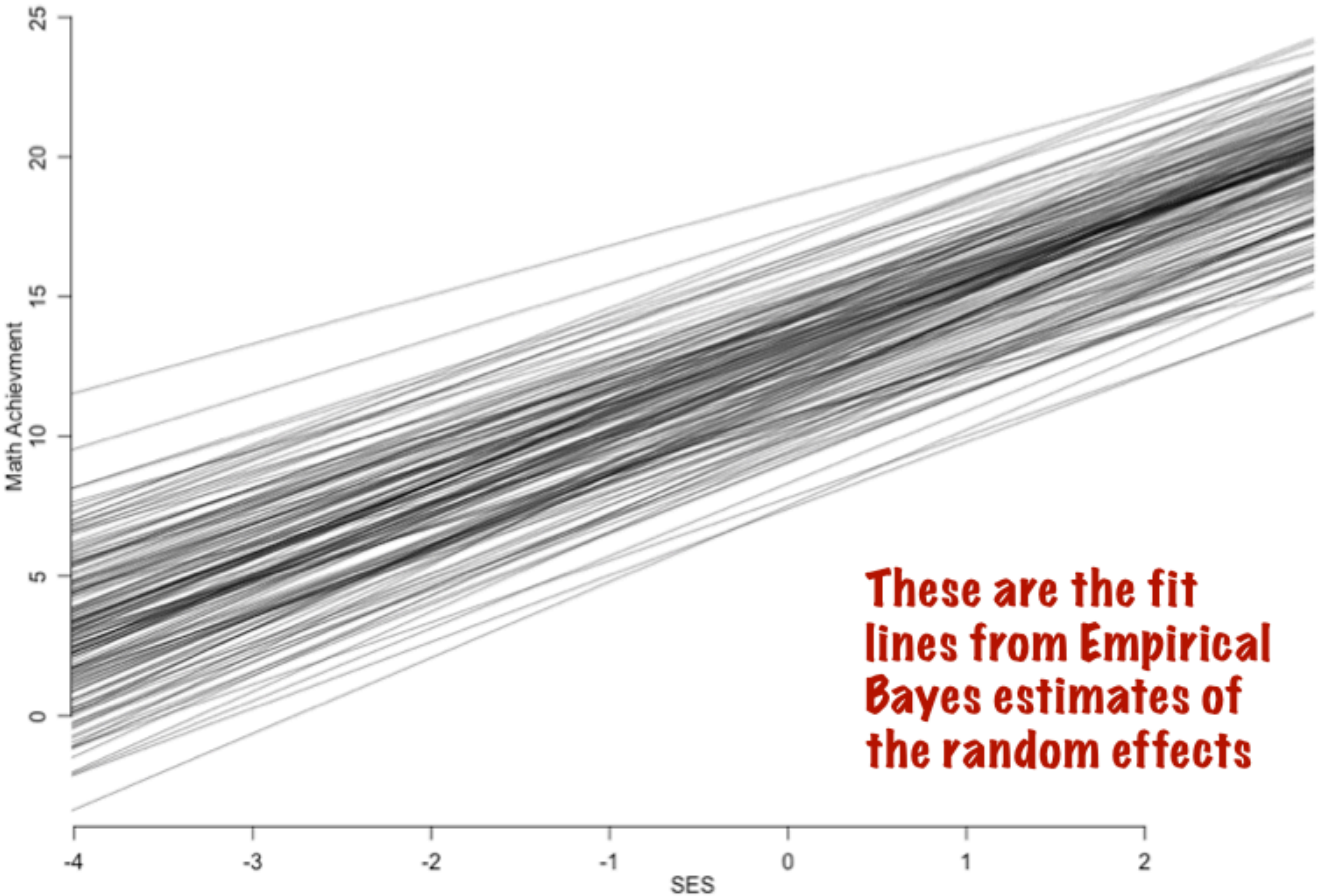
We now have an estimated slope and intercept for each school.

These are Empirical Bayes estimates of the slope and intercept.

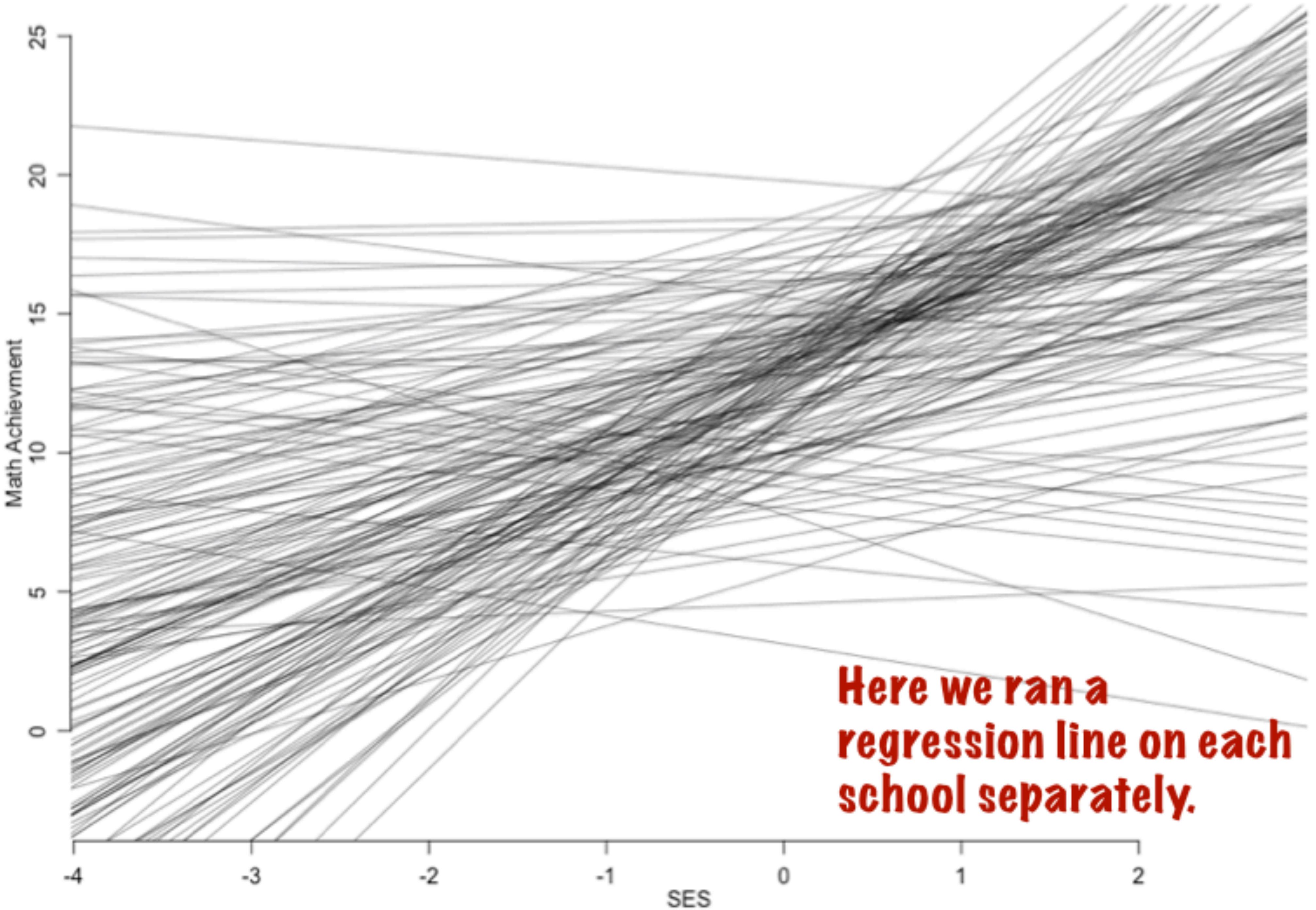
These are the random effects WITH the fixed effects included. That is, the actual within-school intercepts and slopes.

Looking at individual schools' regression lines

Individual school-level regression lines



Individual school-level regression lines without pooling



Smoothing and Over-Smoothing

No pooling

- ★ The individual variation of the schools is making the intercepts and slopes vary **too much**

Empirical Bayes

- ★ We end up heavily borrowing trends from the population. Individual schools (especially those with few students) will look more like the general trend than they actually are. They are **over smoothed.**

How do we get our Goldilocks on?

The Idea

- ★ We start with our fit model which tells us what a population of schools looks like.
- ★ Let's use our model to generate random schools and look at them to see!

Simulating Data: pretend to be mother nature

- ★ First take our estimated parameters as the "truth"
- ★ Now draw random intercepts and slopes from this distribution 160 times to make 160 fake schools.
- ★ Plot the resulting lines.

Generating some random schools

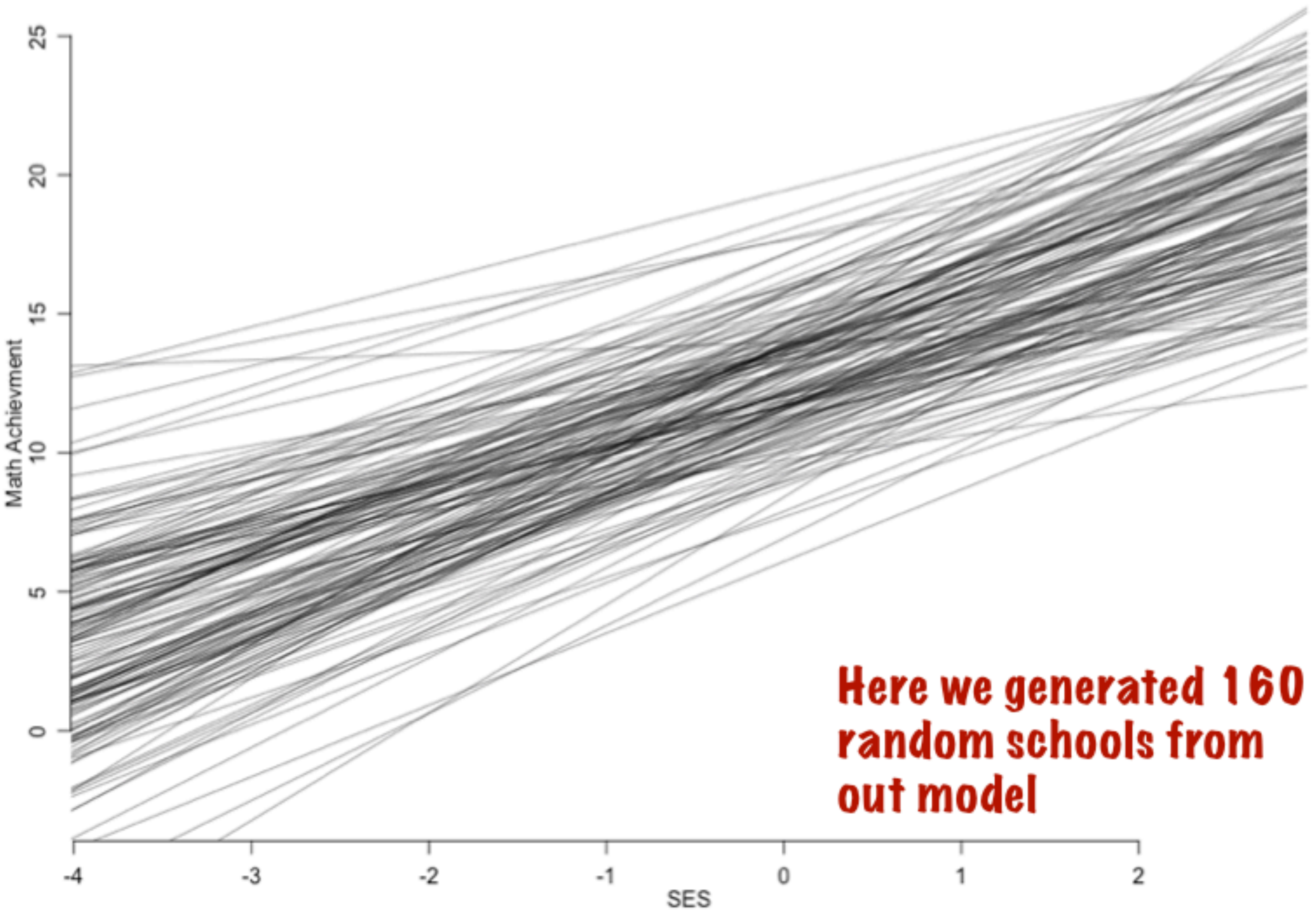
```
> fixef( M1 )
(Intercept)         ses
     12.665       2.394

> VarCorr( M1 )$id
            (Intercept)     ses
(Intercept)      4.8286 -0.1543
ses             -0.1543  0.4129

> res.fake = mvrnorm( 160,
                      mu=fixef( M1 ),
                      Sigma=VarCorr( M1 )$id  )
```

Generate 160 random draws from a multivariate normal distribution with mean mu and variance Sigma

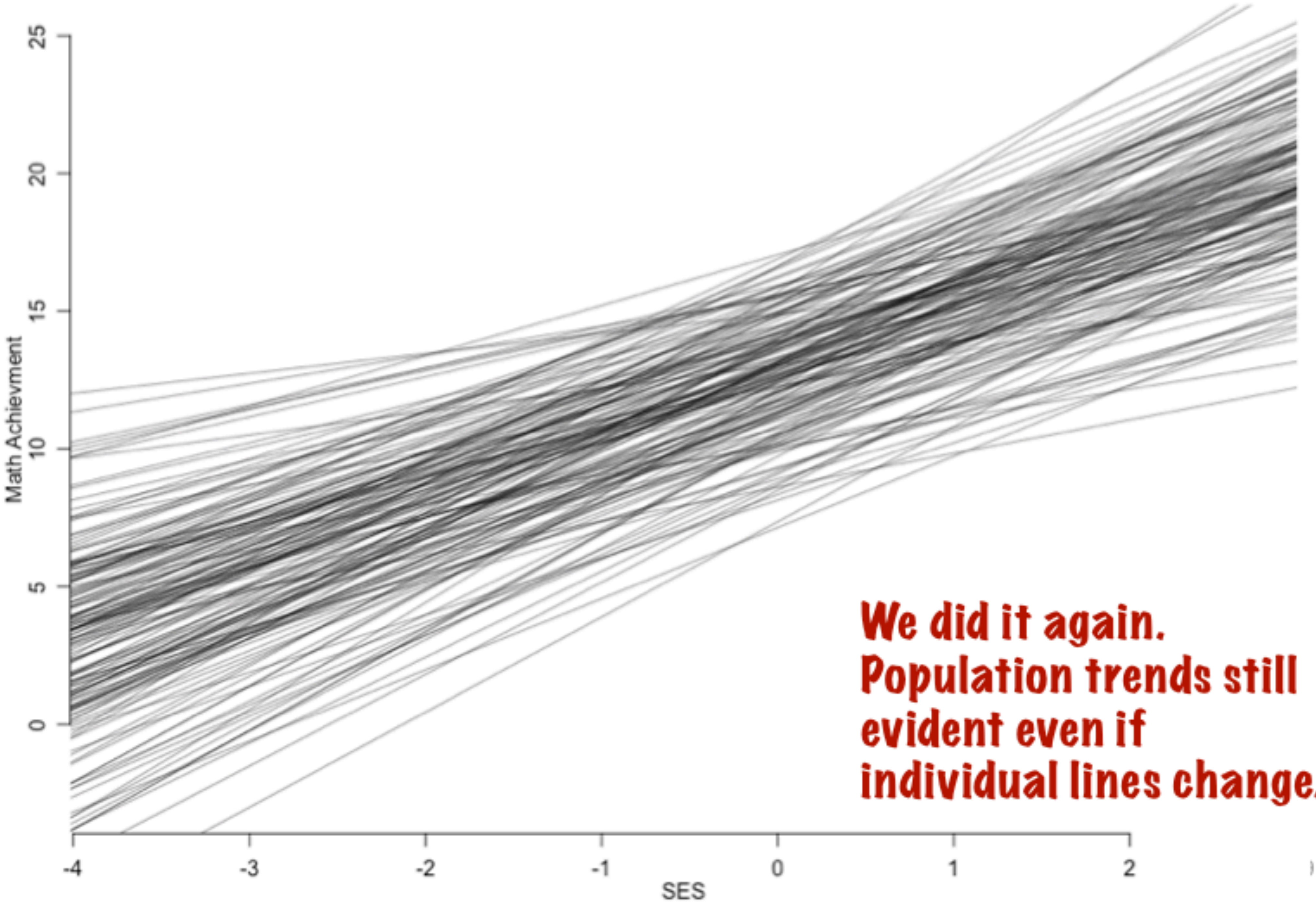
```
> res.fake = data.frame( res.fake )
> names( res.fake ) = c( "alpha", "beta" )

> head( res.fake )
      alpha     beta
1 17.779661 2.332997
2 12.327428 2.697634
```

We get a matrix from mvrnorm. We convert to a data frame and give our columns some nice names.

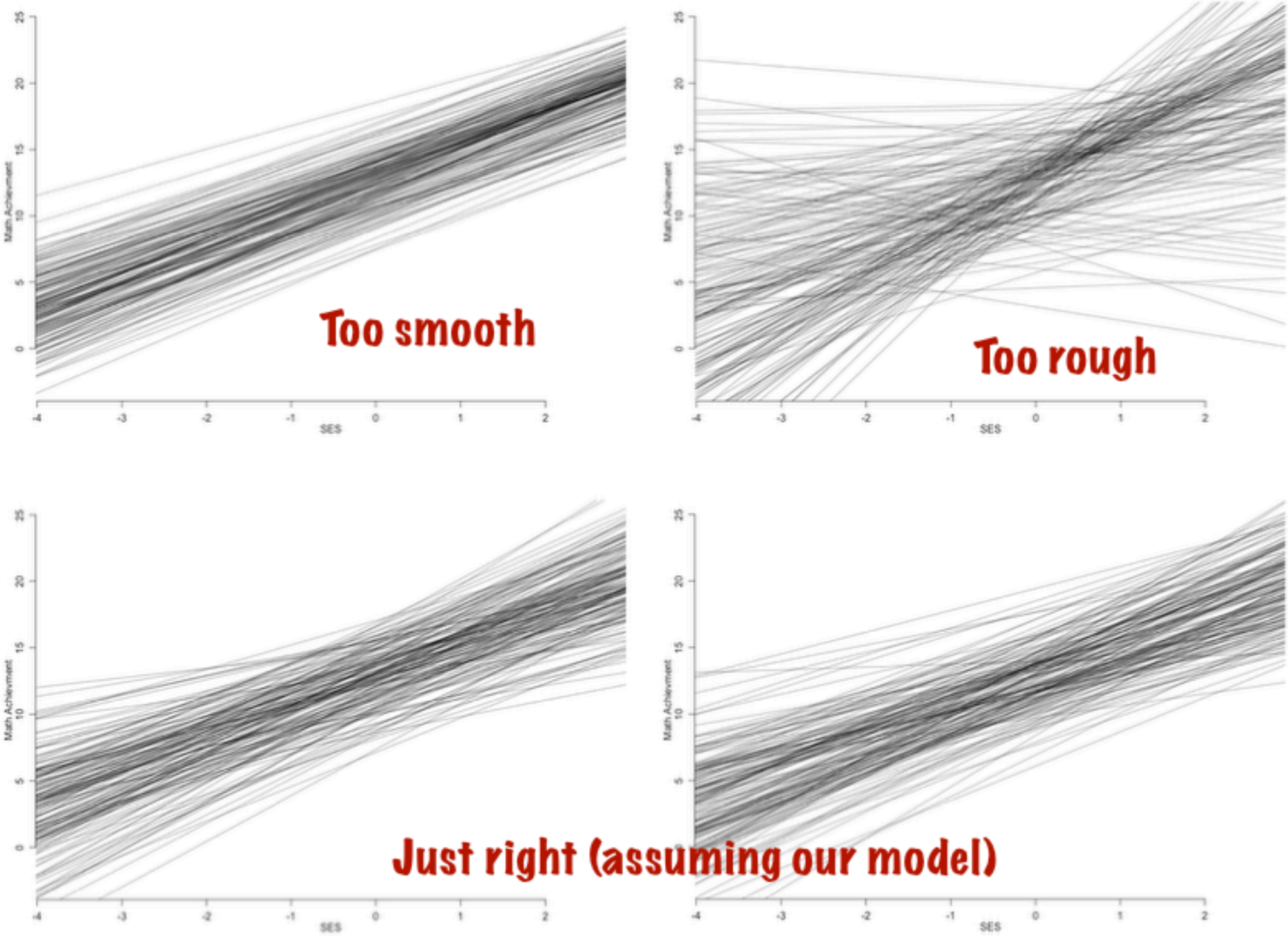
A fake population of 160 schools



Another fake population of 160 schools



We did it again. Population trends still evident even if individual lines change.


Too smooth
Too rough
Just right (assuming our model)
Math Achievement
SES
-4
-3
-2
-1
0
1
2
0
5
10
15
20
25

Recap on the differences

Lines from Empirical Bayes estimates

After partially pooling slopes and intercepts, it looks like the differences in intercepts are much larger than the difference in slopes. But this is smoother than the lines would really be if we knew the truth, because they're all biased in towards the mean values! (Over shrinkage.)

Lines from individual OLS on each school

These lines are too variable, because they contain sampling error; some of these lines are fit to only a handful of students within a school, all of whom have very similar values of SES: that's REALLY bad for the slopes (Over dispersion)

Lines from a randomly generated set of 160 schools

We see more variability than our empirical bayes estimates, but less than the unfettered chaos of each school getting its own line.

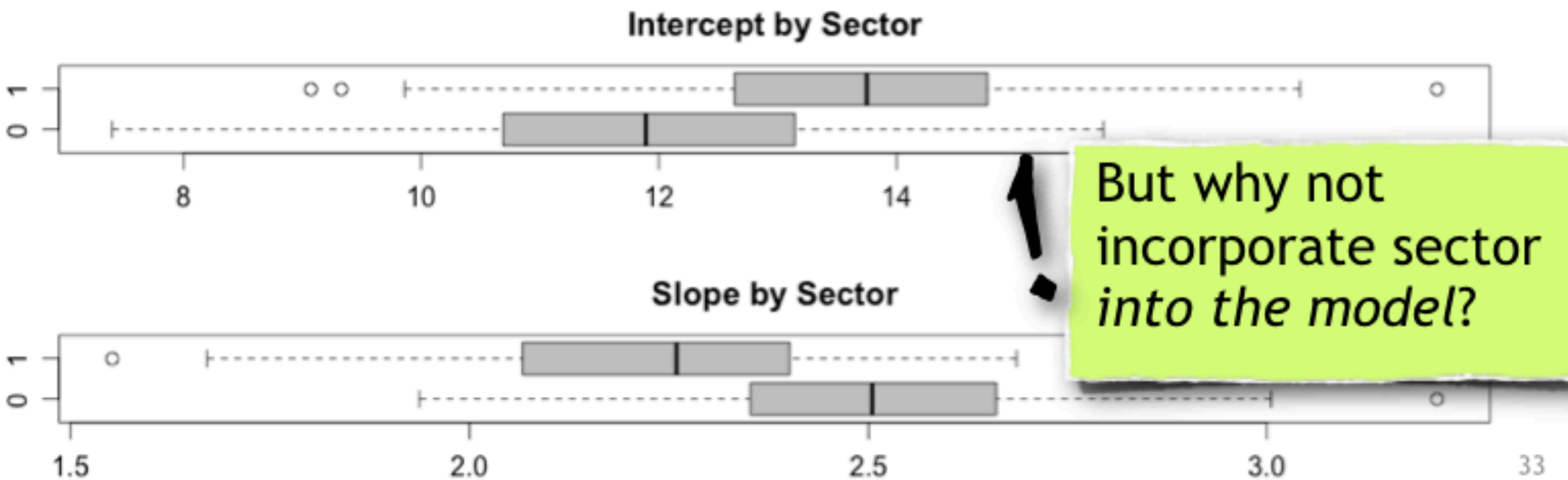
This is the best picture of what the world looks like overall, even though individual lines no longer have any connection to our individual schools in the data

Adding in covariates to predict slope (Level 2 covariates for random slopes)

How do schools vary by sector?

RQ: Does the mean achievement or *relationship of achievement and SES* vary by sector (public vs. catholic)?

Aggregation way: make box-plots on our estimated alphas and betas:



But why not incorporate sector *into the model*?

A model including group-level predictors in the random slope...

Level 1 Model

$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

Now we are seeing real regression equations for each level.

sj is our sector (level 2) variable

Level 2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01} s_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} s_j + u_{1j}$$

Group-level errors

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

The "tau notation" is best.

$$cov(u_{0j}, \epsilon_i) = cov(u_{1j}, \epsilon_i) = 0$$

...collapsed into a single model

Intercept terms

Slope terms

$$
\begin{aligned}
y_i &= \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_i \\
&= (\gamma_{00} + \gamma_{01} s_j + u_{0j}) + (\gamma_{10} + \gamma_{11} s_j + u_{1j}) x_i + \epsilon_i \\
&= \gamma_{00} + \gamma_{01} s_j + u_{0j} + \gamma_{10} x_i + \gamma_{11} s_j x_i + u_{1j} x_i + \epsilon_i \\
&= \gamma_{00} + \gamma_{01} s_j + \gamma_{10} x_i + \gamma_{11} s_j x_i + u_{0j} + u_{1j} x_i + \epsilon_i
\end{aligned}
$$

Covariates in the Intercept component turn into main effects

Covariates in random slope components turn into interactions

The random error term. These are NOT independent, NOT homoskedastic either

OMG! So many parameters!!!

γ_{00} Grand mean intercept (mean of public schools)

γ_{01} How Catholic schools are different from public on average

γ_{10} Average ses slope for public (baseline) schools

γ_{11} Extra bump to slope for Catholic schools

σ_y^2 Residual student variation *within schools*

τ_{00} Variation in the random effect for intercept.

Read as "tau zero zero". The zeros are "row, column" of covariance matrix

τ_{11} Variation in the random effect for slope

τ_{01} Covariance (not correlation) of random intercepts and slopes

Why do the parameters keep *changing??*

α_j, β_j vs. β_{0j}, β_{1j}

As we move to more than 1 covariate at level 1, all beta is best.

$\sigma_0^2, \sigma_1^2, \rho$ vs. $\sigma_\alpha^2, \sigma_\beta^2, \rho$ vs. $\tau_{00}, \tau_{11}, \tau_{01}$

If we need more than 2 random effects, tau is the only way

γ_0 vs. γ_{00} vs. μ_α vs. μ

γ_1 vs. γ_{10} vs. μ_β

The gammas indicates that we have possible level 2 covariates.

Single subscript "0" rather than double ("00") when no covariates involved.

I think these are generally the best

$\beta_{0j}, \beta_{1j}, \tau_{00}, \tau_{11}, \tau_{01}, \gamma_{00}, \gamma_{10}, \gamma_{01}, \cdots$

Our slope model with level-2 covariates

```
> dat = merge( dat, sdat, by="id", all.x=TRUE )
> M2 = lmer(mathach ~ 1 + ses*sector + (1+ses|id), data=dat)


> display( M2 )
              coef.est coef.se
(Intercept) 11.75     0.23
ses          2.96     0.14
sector       2.13     0.35
ses:sector  -1.31     0.22

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 1.95
          ses         0.28     1.00
 Residual             6.07
---
number of obs: 7185, groups: id, 160
AIC = 46585.1, DIC = 46557.2
deviance = 46563.2
```

This is not good, and may indicate a modeling problem

Happy Notation Fun Time!
Let's write out the mathematical model with these numbers on the board.

```
            coef.est coef.se
(Intercept) 11.75     0.23
ses          2.96     0.14
sector       2.13     0.35
ses:sector  -1.31     0.22

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 1.95
          ses         0.28     1.00
 Residual             6.07
---
number of obs: 7185, groups: id, 160
AIC = 46585.1, DIC = 46557.2
deviance = 46563.2
```

γ_{00}

γ_{10}

γ_{01}

γ_{11}

$\sqrt{\tau_{00}}$

$\sqrt{\tau_{11}}$

ρ

σ_y

$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} s_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} s_j + u_{1j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

$$\tau_{01} = \rho\sqrt{\tau_{00}\tau_{11}}$$

R gives correlation (ρ or rho), not covariance. Correlation is better.

Interpretations

- ★ We estimate that, in an average public school, students of SES = 0 (mean SES) will have average math achievement of 11.75
- ★ We estimate that, in an average Catholic school, students of SES = 0 will have math achievement 2.96 higher than students of SES = 0 in an average public school (so we estimate that the intercept in an average Catholic school is about 11.75 + 2.13 = 13.88).
- ★ We estimate that in an average public school, a one-unit difference in SES predicts a 2.96 unit difference in math achievement.
- ★ We estimate that in an average Catholic school, a one-unit difference in SES predicts a difference in math achievement that is 1.31 less than in an average public school (so we estimate that a one-unit difference in SES predicts a 2.96 + -1.31 = 1.65 unit difference in math achievement in an average Catholic school).

Revisiting the gender gap:

Does the gap vary by school?

Previously:
Is there a gender gap?
Is the gender gap different for Catholic vs. Public schools?

Our original model in our new random slope formation

$$mathach_i = \beta_{0j[i]} + \beta_{1j[i]} female_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma_y^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + u_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} sector_j$$

$$u_j \sim N(0, \sigma_0^2)$$

Does it vary by sector?

Is there a gender gap?

{More collective in class R coding if time}

What just happened?

Recap

Check-In
http://cs179.org/lec23

Random slope models allow us to have each school have its own trend between outcome and individual covariate

We can add school level covariates that predict this trend.

★ For example, we can ask whether Catholic schools have different ses-achievement relationships than public schools.

We can fit all this in R

There are lots of ways of writing down these models, because the field is not unified in notation (and different notation is better for different things).

Appendix
Other notation for our same random slope model

Collapsing Just Level-2

$$y_i = \beta_{0j[i]} + \beta_{1j[i]} x_i + \epsilon_i$$

Level 1 Model

$$\epsilon_i \sim N(0, \sigma_y^2)$$

Level 2 Model

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} \gamma_{00} + \gamma_{01} s_j \\ \gamma_{10} + \gamma_{11} s_j \end{pmatrix}, \quad \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right]$$

In the other models we've been presenting, we've treated random effects as offsets from means; in this one we're treating them as the coefficients themselves. The model is the same, but the representation is different.

General case:
We can **really** make things look like random variables

Level 1 Model

$$y_i \sim N(X_i'\beta_j, \sigma_i^2)$$

Level 2 Model

$$\beta_j \sim N(S_j'\gamma, \Sigma_\beta)$$

$$X_j \in R^p$$

$$\beta_j \in R^p$$

$$S_j \in R^q$$

$$\gamma \in R^{q \times p}$$

$$\Sigma_\beta \in R^{p \times p}$$

I wouldn't worry much about this formulation

Don't forget, R represents models too!

R lmer() gives you a formula:

```
mathach ~ 1 + ses * sector +
    (1 + ses | id)
```

It is the econometric view. It has:

- ★ A fixed-effects (OLS) model (with interactions)
- ★ Grouping information to determine the correlation structure of errors