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

Unit 2, Lecture 4:
Within vs. Between variation
or
Group-level versions of individual level predictors

Instructor: Prof. Luke W. Miratrix

lmiratrix@g.harvard.edu - Larsen 603

A note on upcoming Project 1

The next assignment is a “project”

Projects are done in pairs

Goal is to mimic (but scaffold) a real-life data analysis

- ★ You will be given data that is not completely cleaned.
- ★ We will guide you through some simple cleaning exercises
- ★ You will be asked to fit models to answer specific research questions of interest
- ★ We will assess your work by reading the presentation of your results.

Some good things people like

Lecture Goals

Actually acknowledge and discuss the assumptions behind multilevel modeling

Convey concept of between vs. within variation

Show how both can be estimated using multilevel models

Continue to develop interpretation of models, in particular coefficients of level-2 predictors.

Modeling Assumptions

The MLM Assumptions

1. **Level 1 Exogeneity**
2. **Level 2 Exogeneity**
3. **Level 1 Homoskedasticity**
4. **Level 2 Homoskedasticity**
5. **Uncorrelation of residuals**
6. **Uncorrelation of random effects**
7. **Uncorrelation of residuals with random effects.**
8. **Normal distribution of random effects**

Equivalent Assumptions (in Math)

Level 1

1. $r_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$

2. $Cov(X_{qij}, r_{ij}) = 0$

The X and W are level 1 and level 2 covariates

These are not quite the prior page. But the overall set is equivalent. Let's continue to unpack.

Level 2

3. $u_j = (u_{0j}, \ldots, u_{Qj})' \overset{i.i.d.}{\sim} N(0, \Sigma)$

4. $Cov(W_{sj}, u_{qj}) = 0$

Cross Level

5. $Cov(r_{ij}, u_{qj}) = 0$

6. $Cov(X_{qij}, u_{q'j}) = 0 \qquad Cov(W_{sj}, r_{ij}) = 0$

E.g., Normality (#7 prior slide) gets tucked away into the other assumptions (#1, #3 above). The "no correlation" is same as "covariance = 0".

A HS&B Model to make this a bit more concrete

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \beta_2 fem_{ij} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} SEC_j + \gamma_{02} MnSES_j + u_{0j}$$

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

Random effect equations have been omitted

For our assumption slides:

X_1 = SES of student, X_2 = indicator of female

$W_{01} = W_{11}$ = Sector, W_{02} = Mean SES

How many parameters will be estimated?

What is reduced form of this model?

How do you fit it in lmer()?

Assumptions of Level 1 of the Classic Two-Level Model

#1 All the r_{ij} are independent, mean 0, identically distributed, and normally distributed.

$$r_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

These are the classic OLS assumptions

#2 The level-1 predictors are independent of the r_{ij}

$$Cov(X_{qij}, r_{ij}) = 0$$

Conditional on student SES, within-school errors are normal and independent. Equal variances across schools.
Any missing student-level predictors (regulated to error term) are independent of student's social class.

Assumptions of Level 2 of the Classic Two-Level Model

#3 The random errors (the u) are multivariate normal, mean 0, and i.i.d.:

$$u_j = (u_{0j}, \ldots, u_{Qj})' \overset{i.i.d.}{\sim} N(0, \Sigma)$$

#4 Level-2 predictors are independent of the *u*:

$$Cov(W_{sj}, u_{qj}) = 0$$

The residual school effects are indeed bivariate normal, etc.
Any missing school predictors of school-specific intercept or slope are independent of Mean SES and Sector.

Assumptions of Level 1/Level 2 relations of the Classic Two-Level Model

#5 The errors at level 1 and level 2 are independent

$$Cov(r_{ij}, u_{qj}) = 0$$

#6 The predictors at a level are not correlated with the random effects at another level:

$$Cov(X_{qij}, u_{q'j}) = 0 \quad Cov(W_{sj}, r_{ij}) = 0$$

#5: Whatever student-level predictors **excluded** from level-1 are independent of the school-level predictors
Similarly, any missing school-level predictors are uncorrelated with student SES.

More tangibly, what are we assuming?

- ★ The expected value of y is in fact linear:

$$\mathbb{E}\left[y_i | x_i\right] = \alpha_{j[i]} + \beta_{j[i]} x_i$$

This is a consequence of the residual assumptions

- ★ That the schools are a random sample from some super-population. (This is why their offsets are random variables.)
- ★ The school-level random terms are i.i.d.
- ★ The individual residuals are i.i.d and independent of the covariates.
- ★ The individual residuals are independent of school random effects and identically distributed (homoskedastic) across schools.
- ★ The distributions are normal.

So, how could things go wrong?

Understanding assumptions by breaking them

Level 1 Homoskedasticity: If high SES students have more variable achievement than low SES students.

Level 2 Homoskedasticity: If Catholic schools are more variable (in their random offsets) than public schools.

More things to go wrong?

Still breaking things, or trying to.

Level 1 Exogeneity: If the association b/w SES and Math achievement is not linear

Level 2 Exogeneity: This one is fine, as long as we have categorical level-2 predictors. (With categorical predictors, we are always "linear.")

For continuous, we would need a linear relationship again.

Even more things wrong!?!

Break, break, break.

Uncorrelation of student residuals: Our school samples are generated by taking a couple of classes at random and measuring SES and math achievement in those classes only. (This would be unobserved clustering.)

Uncorrelation of random effects: Our schools are sampled by first sampling districts, and then sampling schools within those districts. (More unobserved clustering.)

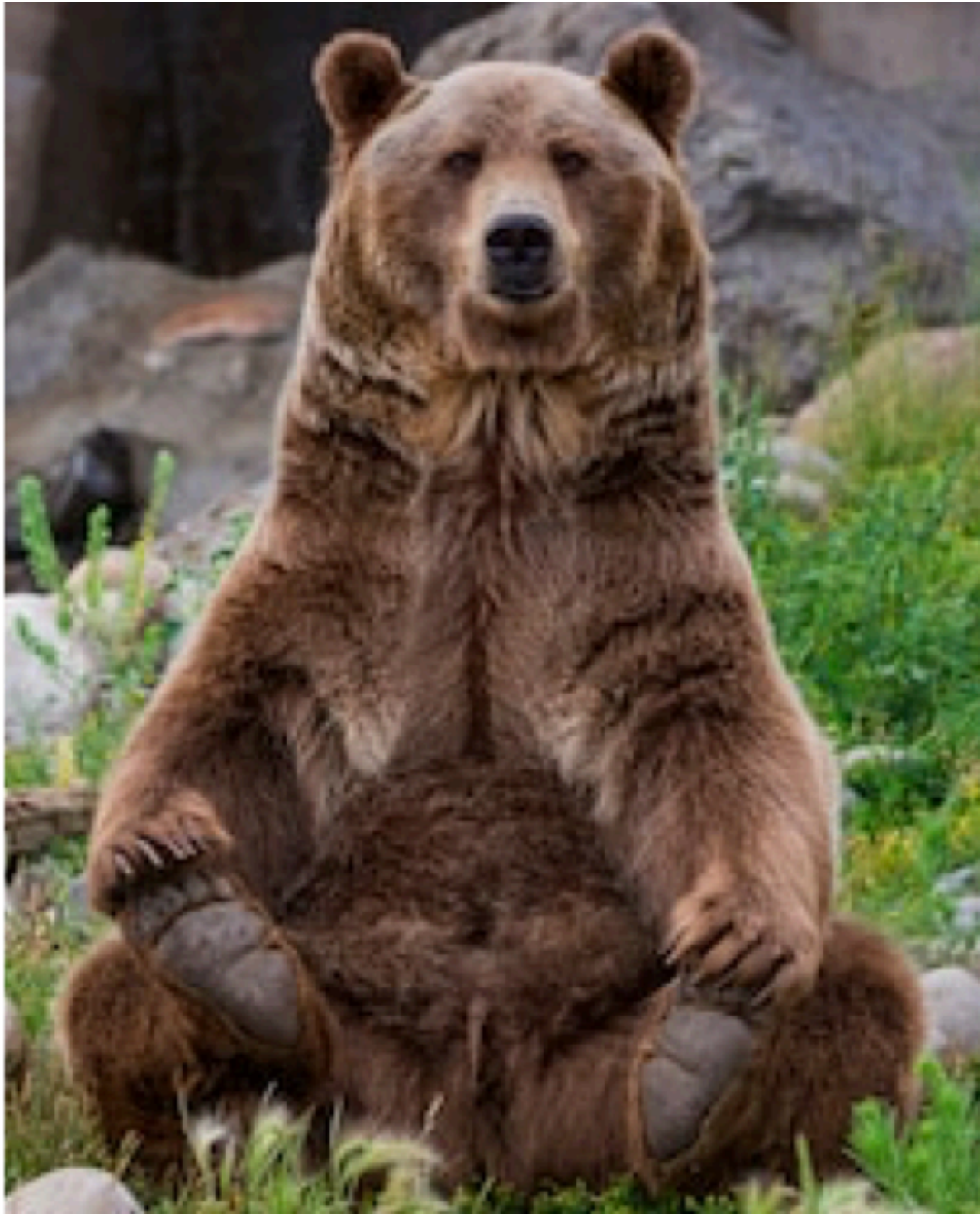
Impact of assumptions

Assumptions 2, 4 & 6: Relationship between variables in the "structural" portion and error terms. Violation can give *bias.* (**This is a big deal**.)

Assumptions 1, 3 & 5: These are on random portions only. Their tenability affects

- ★ consistency of the standard error estimates
- ★ accuracy of variance estimates
- ★ accuracy of hypothesis tests and confidence intervals. **(Ok, this can be a big deal too.)**

Centering your covariates

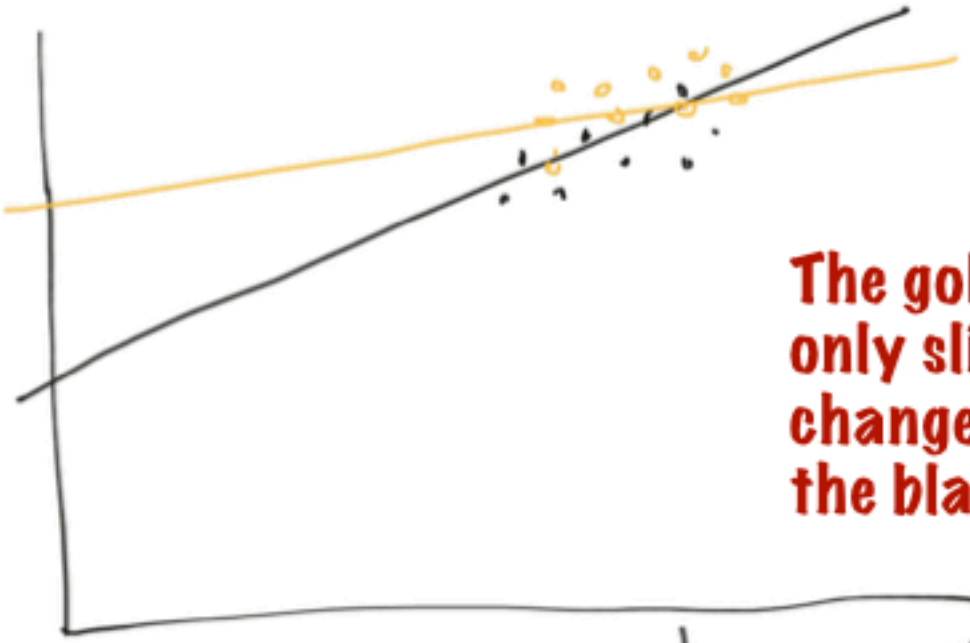


Interpreting intercepts is often ignored

- ★ What does the school-level intercept even mean?
- ★ Often the intercept is a nuisance, and ignorable.
- ★ Random slope models tie the intercept and slope together; if the intercept is meaningless, then this is an odd activity.
- ★ We have enough nonsense; let's make sure we are tying ***meaningful*** things together.

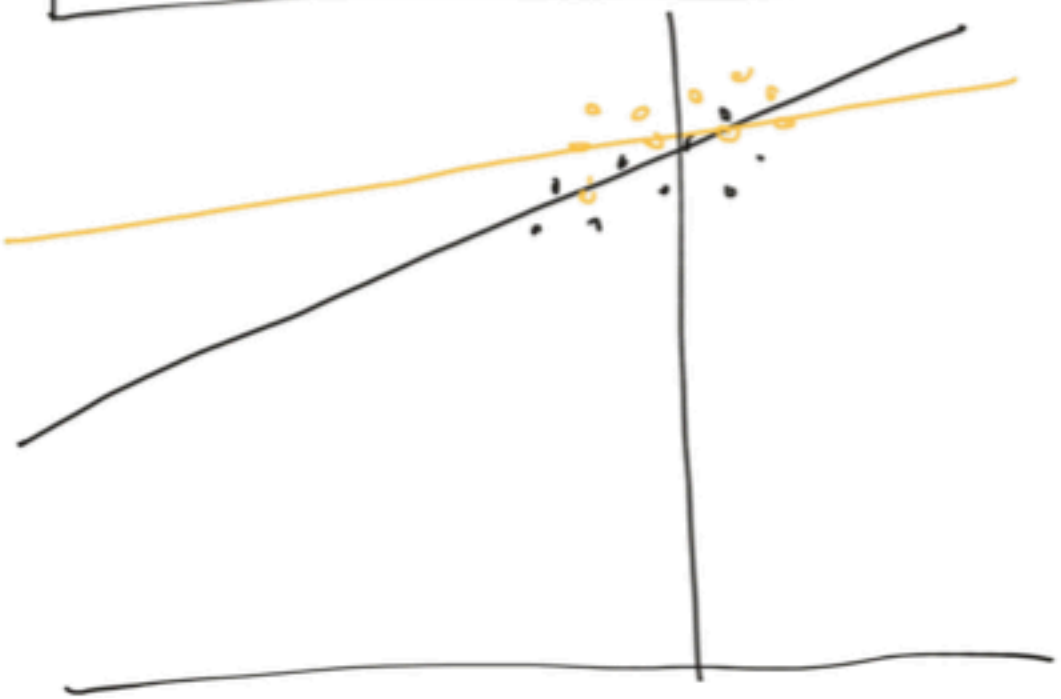
Why intercepts can be unstable

Intercept far away. We get slight change in slope going with big change in intercept. They are quite correlated.



The gold dots are only slightly changed versions of the black dots.

With centering, changing the slope doesn't really change the intercept. The intercept is more the overall mean. Useful.



Group Mean Centering

$$y_i = \alpha_{j[i]} + \beta_{j[i]} \left(x_i - \bar{x}_{j[i]} \right) + \epsilon_i$$

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

- ★ The intercept and the slope have been decoupled.
- ★ The intercept is a relevant, overall school summary.
- ★ The slope is a statement about deviations of students *within* the school.

How to Group Mean Center (using dplyr package)

```
> head( df )
  SID         x         y
1   1 29.91908 20.47426
2   2 29.93382 26.24375
3   3 30.06064 22.24806
4   4 30.77270 20.15656
```

```
> df2 = df %>%
        group_by( id ) %>%
        mutate( x = x-mean(x) )
```

```
> head( df2 )
  SID           x         y
1   1 -0.2674487 20.47426
2   1 -0.9359377 19.25521
3   1  0.5486244 22.49033
4   1 -0.6924070 22.87696
```

Different Versions of X

Always consider the meaning of X!

★ Covariate left alone

- Intercept often of questionable interpretation.
- Intercept often effectively **determined by the slope.**
- Shrinkage estimates of random coefficients are shrinking to what?

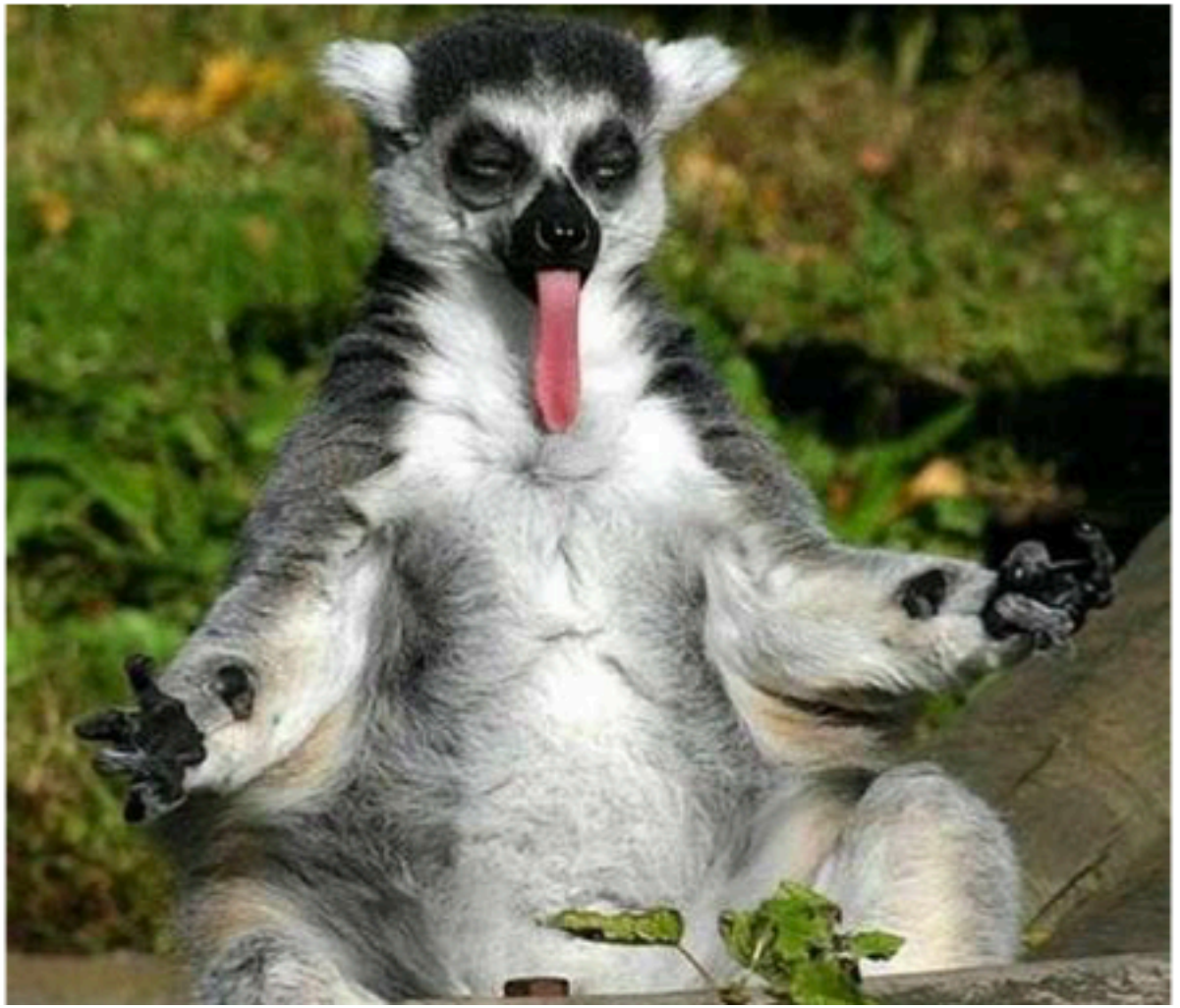
★ Grand-mean centered $\alpha_j = \mu_j - \beta_j(\bar{x}_j - \bar{x})$

- Standard form for ANCOVA.
- Intercepts are now "adjusted means"
- They are the mean prediction for a "canonical individual" in that group.

★ Group-mean centered

- Individual-level covariate now measures departure from the group individual is in.
- Intercepts are now means for the whole group $\alpha_j = \mu_j$

Within vs Between Variation (which leads to more centering)



Dataset: Birthweight and smoking

Multiple births from same mothers

Outcome: birthweight

Primary covariate: smoking during pregnancy

Other covariates: education, married, age, history of pre-natal care, etc.

See RH&S Chapter 3

Remarks:

Some mothers always (or never) smoke, while others change their behavior between pregnancies

Our Core Issue

We want to regress weight onto smoking, controlling for other effects (heading towards a causal argument)

But it is “unrealistic to assume that birthweights of children born to the same mother are uncorrelated given the observed covariates.”

What do we do?

(RH&S 3.3)

Potential comparisons

Between Mother:

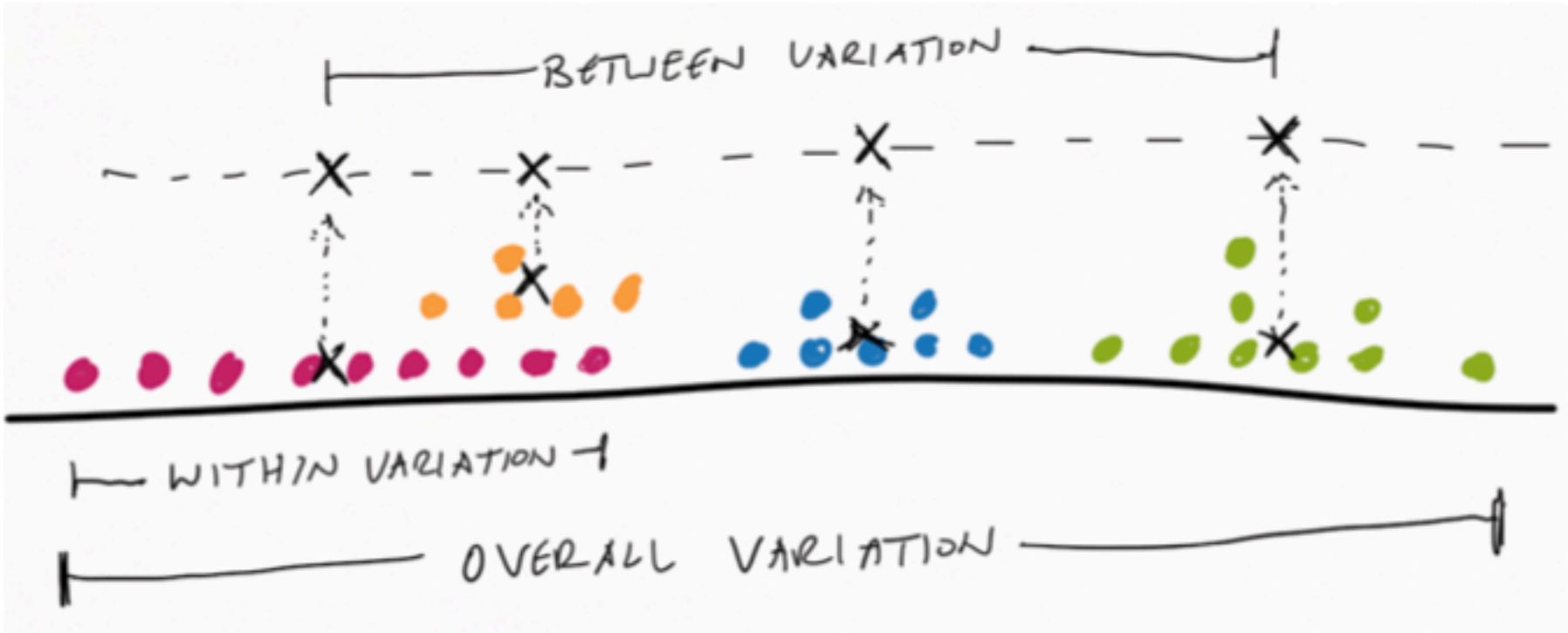
Take two mothers with the same covariates, but one who smokes and one who does not. What is the expected difference in birthweight of their babies?

Within Mother:

Take a mother. What is the expected difference in birthweight between her babies from when she was smoking vs. from when she was not smoking?

Turn and talk to your neighbors: Come up with a concrete reason/ illustration/thought experiment to show why each of these could give biased (confounded) estimates.

$$Var_O = Var_W + Var_B$$



https://anthonyartino.wordpress.com/2016/04/23/between-vs-within-person-variation-a-major-problem-in-behavioral-science/

Some summary statistics

Overall standard deviation

$$S_{xO} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{..})^2}$$

Between

$$S_{xB} = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}(\bar{x}_{.j} - \bar{x}_{..})^2}$$

Within

$$S_{xO} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{.j})^2}$$

Anyway...

Estimating between (only)

Collapse data to the group level

- ★ For each mother, calculate average birthweight of her babies, average smoking status, average everything else
- ★ Regress average birthweight onto these averages

Estimating within (only)

Take out between variation and then regress

- ★ Add a fixed-effect (completely unpooled) intercept for each mother.
- ★ Then all between mother variation is taken up by these unrestricted fixed effects
- ★ Any other variation explained by covariates must be within-mother variation
- ★ Question: What about the mothers with no variation in smoking?

Alternative: ***recentering***

- ★ Subtract mother-level means from outcomes and all covariates (i.e., subtract the between-mother model from the overall model). See RH&S pg 145

Estimating Both with MLM!

We allow for different within and between effects. So, for birth *i* of mother *j:*

$$y_{ij} = \beta_1 + \beta^W (s_{ij} - \bar{s}_{\cdot j}) + \beta^B \bar{s}_{\cdot j} + \xi_j + \epsilon_{ij}$$

with s_{ij} smoking status, and $s_{\cdot j}$ average smoking status for mother *j*.

Remarks:

- ★ Our recentered variable is *not correlated* with our random effect by construction.
- ★ We may want to recenter all our level-1 covariates so none of them are correlated, so as to improve validity of inference for our target β^W.

RH&S pg145
"Table 3.2"

Rand Int. (w/ means)

	Random effects **Rand Int. (Simple)** $\widehat{\beta}^{ML}$		Between effects **OLS (aggregated)** $\widehat{\beta}^{B}$		Within effects **FE for mothers** $\widehat{\beta}^{W}$		Random effects +clust. mean $\widehat{\beta}^{ML}$	
	Est	(SE)	Est	(SE)	Est	(SE)	Est	(SE)
Fixed part								
β_1 [_cons]	3,117	(41)	3,241	(46)	2,768	(86)	3,238	(46)
β_2 [smoke]	−218	(18)	−286	(23)	−105	(29)	−105	(29)
β_3 [male]	121	(10)	105	(19)	126	(11)	126	(11)
β_4 [mage]	8	(1)	4	(2)	23	(3)	23	(3)
β_5 [hsgrad]	57	(25)	59	(26)			56	(25)
β_6 [somecoll]	81	(27)	85	(28)			83	(28)
β_7 [collgrad]	91	(28)	100	(29)			98	(29)
β_8 [married]	50	(26)	42	(26)			42	(26)
β_9 [black]	−211	(28)	−218	(29)			−219	(28)
[illegible]	[illegible]	[illegible]	241	(101)	155	(60)	155	(60)
β_{15} [m_smok]							−183	(37)
β_{16} [m_male]							−20	(22)
[illegible]							45	(92)
β_{21} [m_pretri3]							96	(117)
Random part								
$\sqrt{\psi}$	339				440 [a]		338	
$\sqrt{\theta}$	371				369 [a]		369	

Some output truncated for space

These are not parameter estimates, but sd of raw residuals and Yes

Back to our friend, HS&B

Math achievement and SES, again

First, our model:

$$\text{mathach} \sim \text{ses} + \text{meanses} + (1|\text{id})$$

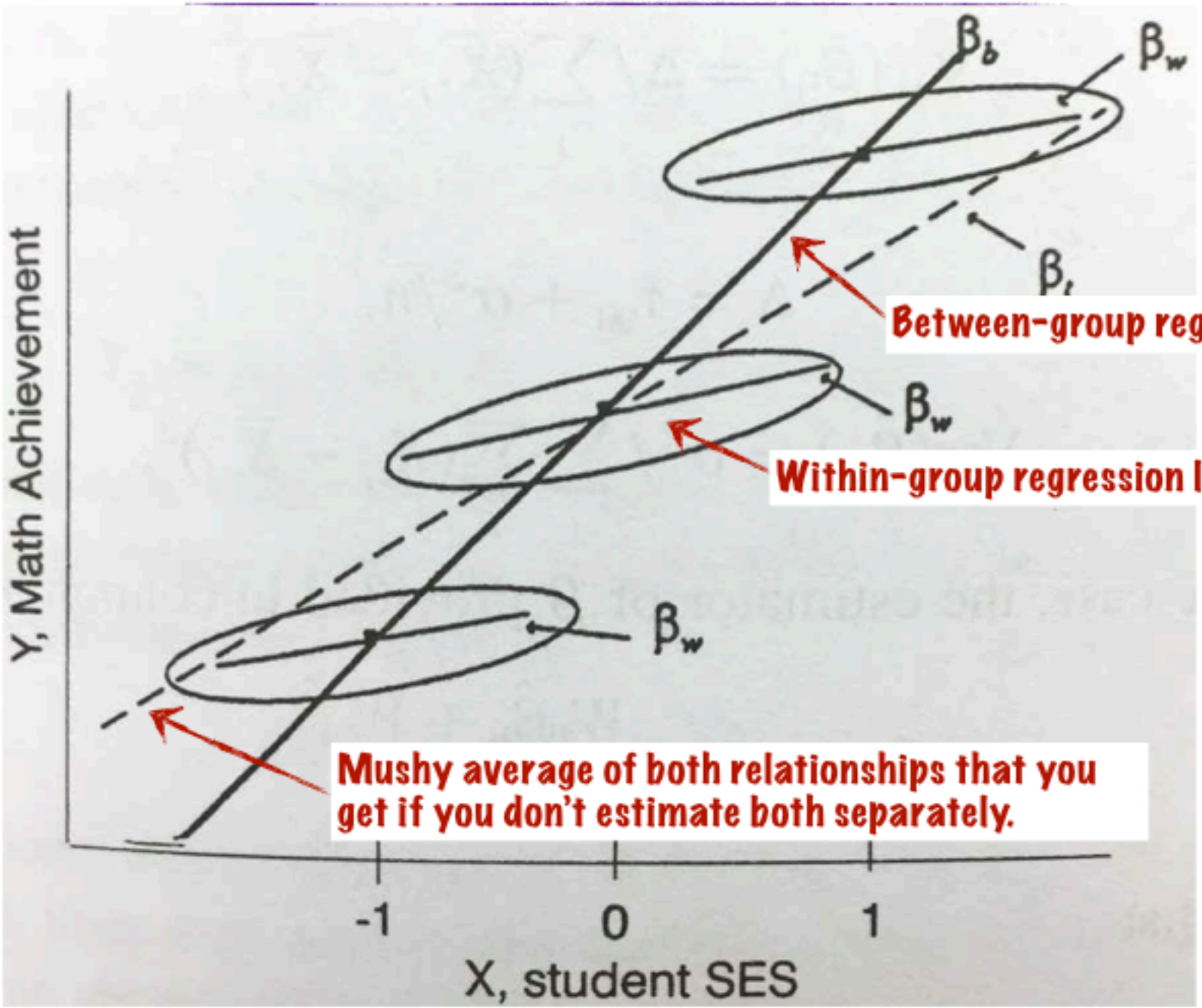
ses is an *individual-level* covariate

meanses is a *school-level* covariate

Questions:

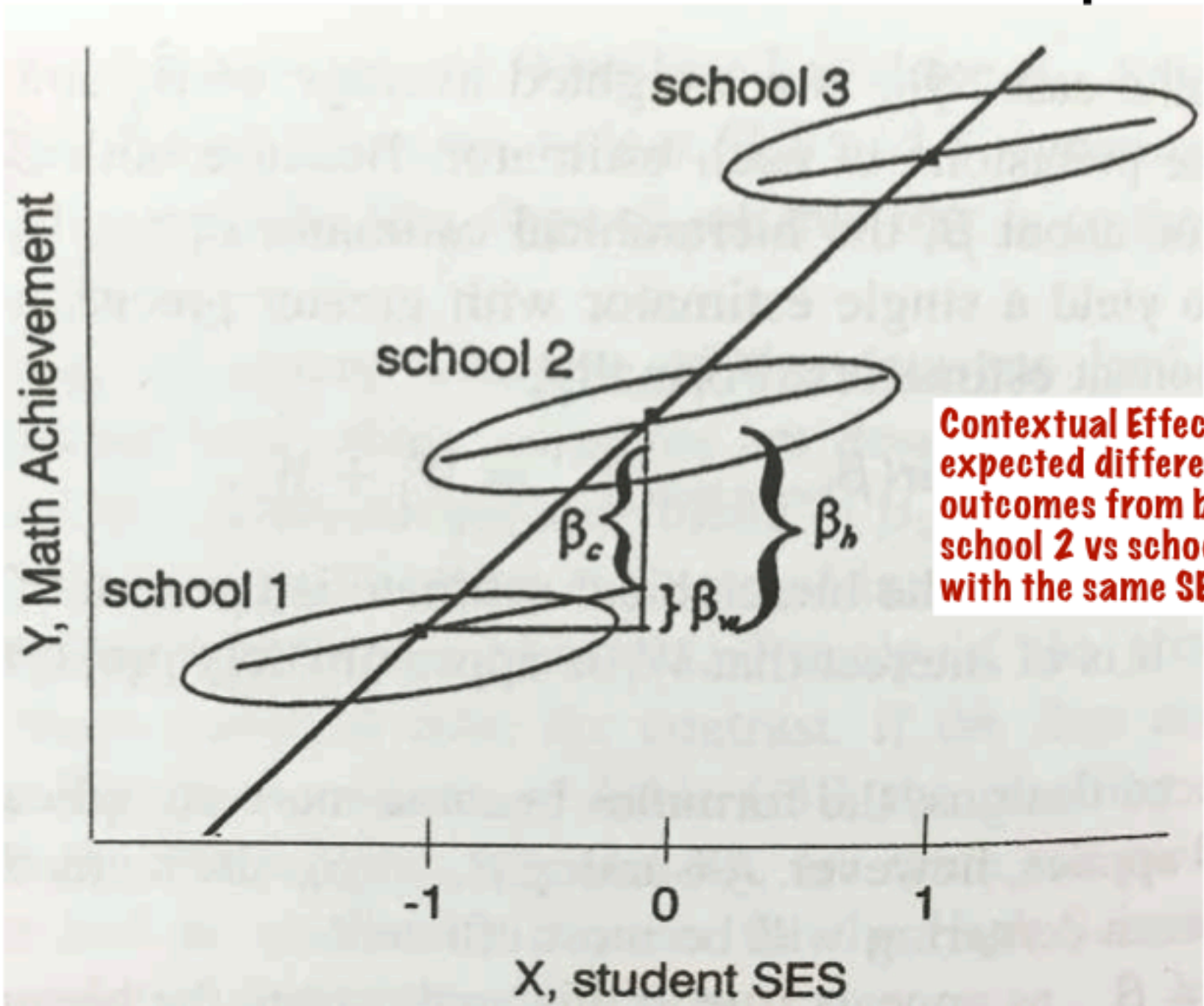
- ★What model are we estimating?
- ★What coefficients will we get?

Within vs Between Relationships



R&B
pg 137

The **Contextual Effect** β_c



Contextual Effect: The expected difference in outcomes from being in school 2 vs school 1 (but with the same SES).

R&B
pg 140

Fitting our model with between and within effects

```
> dat = dat %>% group_by( id ) %>%
      mutate( meanses = mean( ses ) )

> M2 = lmer( mathach ~ ses + meanses + (1|id), data=dat )
> display( M2 )

(Intercept) 12.66       0.15
ses          2.19       0.11
meanses      3.68       0.38

Error terms:
 Groups   Name          Std.Dev.
 id       (Intercept)   1.64
 Residual               6.08
---
number of obs: 7185, groups: id, 160
```

What would you predict average math achievement to be for a school with an average SES 0.5 above average?
What about for a student with 0 SES in such a school?

Alternative fitting (note group recentering)

```
> dat = dat %>% group_by( id ) %>%
        mutate( ses.cent = ses - mean( ses ) )
> M3 = lmer( mathach ~ ses.cent + meanses + (1|id),
                       data=dat )

> display( M3 )
            coef.est coef.se
(Intercept) 12.68     0.15
ses.cent     2.19     0.11
meanses      5.87     0.36

Error terms:
 Groups   Name          Std.Dev.
 id       (Intercept)   1.64
 Residual               6.08
---
number of obs: 7185, groups: id, 160
AIC = 46578.6, DIC = 46559
deviance = 46563.8
```

Consider our prior questions from the earlier model. Easier to answer?

Contextual vs. Overall Effects

```
> stargazer( M2, M3, type = "text" )
```

	Dependent variable:	
	mathach	
	(1) grand mean cent	(2) Group centered
ses	2.191***	
	(0.109)	
ses.cent		2.191***
		(0.109)
meanses	3.675***	5.866***
	(0.378)	(0.362)
Constant	12.680***	12.680***
	(0.149)	(0.149)
Observations	7,185	7,185
Log Likelihood	-23,284.000	-23,284.000
Akaike Inf. Crit.	46,579.000	46,579.000

Slope is same for both models

Note: 5.866 - 2.191 = 3.675

Aside: You can't include Level-2 variables in fixed effect models

Consider doing this the "fixed effects" way:

```
a = lm( mathach ~ ses + meanses + id,
                                  data=dat )
```

This will **fail** due to **collinearity.**

When covariates are collinear, we cannot get unique estimates for their coefficients in the linear model.

G&H Pg 269

Aside: Violations of level 2 endogeneity of level 1 covariates

The assumption: $Cov(X_{qij}, u_{q'j}) = 0$

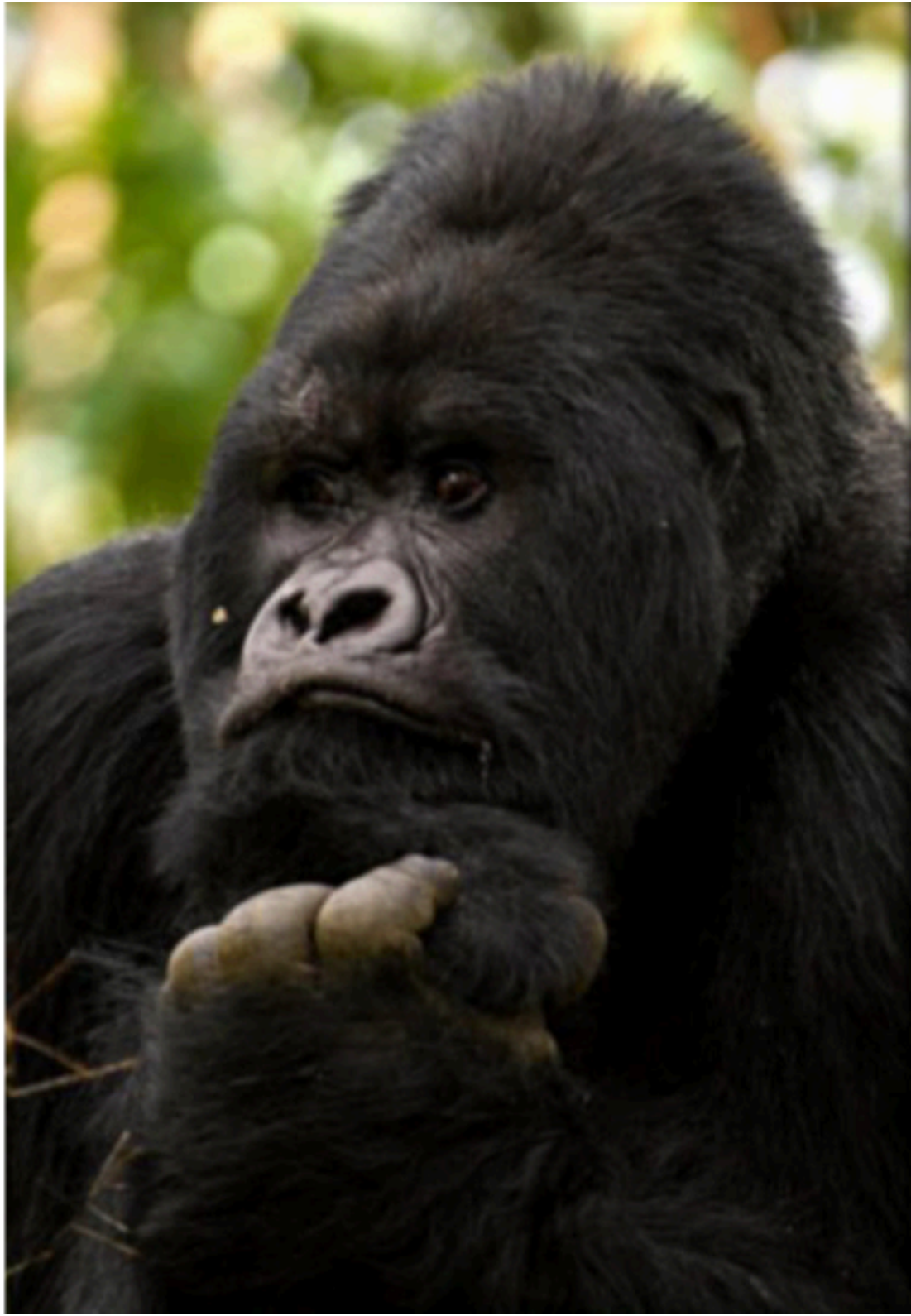
Fixed Effect Regression:

- ★ All level 2 covariates (unobserved and observed) are implicitly included.
- ★ Everything is within-cluster variation

Group mean centering & inclusion of cluster mean covariates

- ★ Deviations from cluster means are automatically uncorrelated with cluster means and random offsets
- ★ So this makes our core assumption true by construction!

Recap &
reflect
time



Recap

Check-In
http://cs179.org/lec24

There are a lot of modeling assumptions, but the ones about independence are the ones to worry about.

Within vs. between is a way of identifying whether a covariate is associated with outcome due to grouping of individuals by that covariate, or a more direct link

This is a slippery concept, and it takes careful thinking to get it right. Examples are helpful.

Appendix:
How much data?

Reading: G&H 12.9

When J is small, worry.

When J is small, people often elect to do a “no pooling” or “fixed effect” analysis.

MLM will *tend* to overestimate group variance terms (although for small J there is a tendency to collapse estimates to 0).

And if you have 1-2 groups? No dice, unless you go full Bayes

What is small? 10? 30? Depends on who you ask.

When n_j is small, worry about that group

When you have few observations in a group then:

- ★ The predictions and individual regression for that group will be unreliable.
- ★ Worse, the SEs will tend to be too small.

However:

- ★ The overall picture of trends across groups and individual covariate regression trends across groups can still be ok.
- ★ Even a large number of size-1 groups is fine.
- ★ You do need some multi-unit groups to measure within-group variance well.

Appendix
General form of two-level models

(Two-level models allow for lots of relationships)

General Notation Note: The general form of two-level model

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + r_{ij}$$

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj}$$

X are our level-1 predictors. W are our level-2 predictors. We have Q+1 things varying by group.

Fixed effects are generated by dropping u_{qj} terms.

Example: yet another HS&B Model

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \beta_2 fem_{ij} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} SEC_j + \gamma_{02} MnSES_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} SEC_j + u_{1j}$$

We have Q = 1 random slopes so Q+1=2 level 2 equations.

$S_1 = 2$, $S_2 = 1$

X_1 = SES of student, X_2 = indicator of female

$W_{01} = W_{11}$ = Sector, W_{02} = Mean SES

How many parameters will be estimated?

What is reduced form of this model?

How do you fit it in lmer()?

Model fitting results

```
> M3 = lmer( mathach ~ 1 + female + ses * sector + meanses +
                                        (1 + ses|id), data=dat )

> display( M3 )
            coef.est coef.se
(Intercept) 12.79     0.21
female      -1.18     0.16
ses          2.73     0.14
sector       1.29     0.29
meanses      3.04     0.37
ses:sector  -1.31     0.21

Error terms:
 Groups   Name        Std.Dev. Corr
 id       (Intercept) 1.45
          ses         0.18     0.65
 Residual             6.05
---
number of obs: 7185, groups: id, 160
AIC = 46482.9, DIC = 46445
deviance = 46454.0
```