UCLA School of Education & Information Studies
Department of Education

EDUC 231D
Advanced Quantitative Methods: Multilevel Analysis
Winter 2025

Introduction to Multilevel Models and Random Effects (Continued)

Lecture 4 Presentation Slides

January 16, 2025

Today’s Topics

- One-way ANOVA models with random effects – unbalanced case
- Unpacking estimation in the unbalanced case
- Empircal Bayes estimates of group means
- Means as outcomes

One-way ANOVA Model: Unbalanced Case

Motivating example

- What was the average math score for Grade 8 students in the United States in 2019? How much did math scores differ across schools?
- Use TIMSS data from a sample of 20 U.S. schools
  - Use the actual number of students tested in each school
  - The data include 661 students
  - Number of students per school ranges from 16 to 51 students

Naïve approach: overall average math score

- Option 1: calculate mean across all students

$$\frac{\sum_{i=1}^{N} Y_{ij}}{N} = 482.64, \text{ where } N = 661$$

- Option 2: calculate mean math score in each school ($\bar{Y}_{.j}$), then calculate average of the $\bar{Y}_{.j}$'s across the 20 schools:

$$\frac{\sum_{j=1}^{J} \bar{Y}_{.j}}{J} = 482.78, \text{ where } J = 20$$

Why are these numbers different?

idschool	j	schnb	schmath_j ($\bar{Y}_{.j}$)
5036	1	41	484.25
5049	2	16	485.51
5050	3	33	535.46
5058	4	42	378.48
5103	5	29	448.60
5110	6	41	619.35
5112	7	37	375.71
5128	8	26	421.43
5143	9	22	561.16
5181	10	51	425.34
5182	11	38	506.35
5198	12	18	520.70
5199	13	36	475.74
5215	14	33	469.31
5239	15	21	413.62
5244	16	29	486.36
5252	17	38	529.84
5269	18	51	528.36
5271	19	22	479.34
5275	20	37	510.69

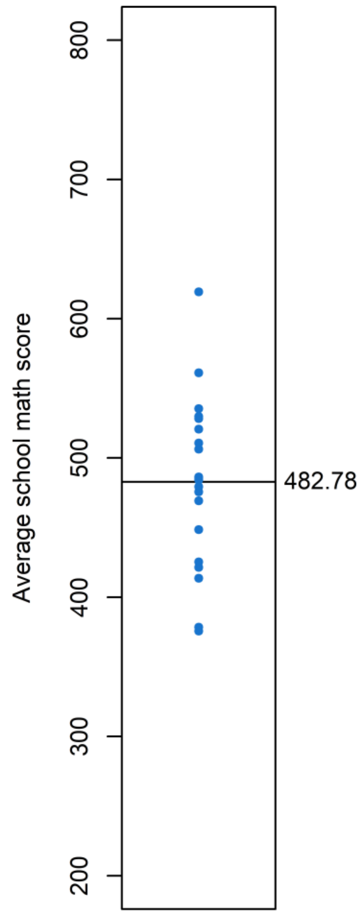
Naïve approach: variation across schools

- Calculate variance of the $\bar{Y}_{.j}$'s across the 20 schools:

$$\frac{\sum_{j=1}^{J}\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2}{J-1} = 3{,}697.67,$$

where $J = 20$ and $\bar{Y}_{..} = 538.98$

- This variance calculation will overestimate the variance if the $\bar{Y}_{.j}$'s are measured with error



Model-based approach: Hierarchical model

- Level-1 (within-school) model: $Y_{ij} = \beta_{0j} + r_{ij}\,, r_{ij} \sim N(0, \sigma^2)$

- Level-2 (between-school) model: $\beta_{0j} = \gamma_{00} + u_{0j}\,, u_{0j} \sim N(0, \tau_{00})$

- Combined model: $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

- Same as a one-way ANOVA model with random effects, where:

$$\mathrm{Var}(Y_{ij}) = \mathrm{Var}(u_{0j}) + \mathrm{Var}(r_{ij}) = \tau_{00} + \sigma^2$$

Estimate the hierarchical model in R

- Combined model: $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

```
m1 <- lmer(bsmmatxx ~ 1 + (1 | idschool), data=td.bx)

print(as_flextable(m1), preview = "pptx")

summary(m1)
```

Estimate the hierarchical model in R

- Combined model:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

$$u_{0j} \sim N(0, \tau_{00})$$

$$r_{ij} \sim N(0, \sigma^2)$$

- $\hat{\gamma}_{00} = 482.757$
- $\hat{\tau}_{00} = (59.647)^2 = 3{,}558$
- $\hat{\sigma}^2 = (73.756)^2 = 5{,}440$

		Estimate	Standard Error	df	statistic	p-value	
		Fixed effects					
	(Intercept)	482.757	13.673	19	35.307	0.0000	***
		Random effects					
idschool	sd__(Intercept)	59.647					
Residual	sd__Observation	73.756					

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

square root of the estimated residual variance: 73.8
data's log-likelihood under the model: -3,807.4
Akaike Information Criterion: 7,620.9
Bayesian Information Criterion: 7,634.3

Unpacking Estimation in the Unbalanced Case

What do we already know from the balanced case?

- Estimate of the grand mean math score is the mean of the school-mean math scores:

$$\hat{\gamma}_{00} = \frac{\sum_{j=1}^{J} \bar{Y}_{.j}}{J}$$

We can use the mean in the balanced case because the error terms are "independent and identically distributed" (iid). In other words, the error associated with $\bar{Y}_{.j}$ is the same for all schools.

- Estimate of Level-1 (within-school) error variance:

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(Y_{ij} - \hat{\beta}_{0j}\right)^2}{N - J}$$

What do we already know from the balanced case?

- Estimated variance of the school means:

$$\hat{\Delta}_j = \hat{\tau}_{00} + \hat{V}_j, \qquad \text{where } \hat{V}_j = \frac{\hat{\sigma}^2}{n_j}$$

Parameter variance

Error variance for school *j*

- In the balanced case, n_j is the same for all schools, so $\hat{V}_j$ and $\hat{\Delta}_j$ are the same for all schools
- Not true for the unbalanced case, which complicates estimation

idschool	j	schn (n_j)	schmath_j
5036	1	41	484.25
5049	2	16	485.51
5050	3	33	535.46
5058	4	42	378.48
5103	5	29	448.60
5110	6	41	619.35
5112	7	37	375.71
5128	8	26	421.43
5143	9	22	561.16
5181	10	51	425.34
5182	11	38	506.35
5198	12	18	520.70
5199	13	36	475.74
5215	14	33	469.31
5239	15	21	413.62
5244	16	29	486.36
5252	17	38	529.84
5269	18	51	528.36
5271	19	22	479.34
5275	20	37	510.69

Challenge of the unbalanced case

- The error variance of the $\hat{\beta}_{0j}$' differs across schools due to the difference in n_j
- School means based on fewer students are measured with less precision (more error)
- School means based on more students are measured with more precision (less error)

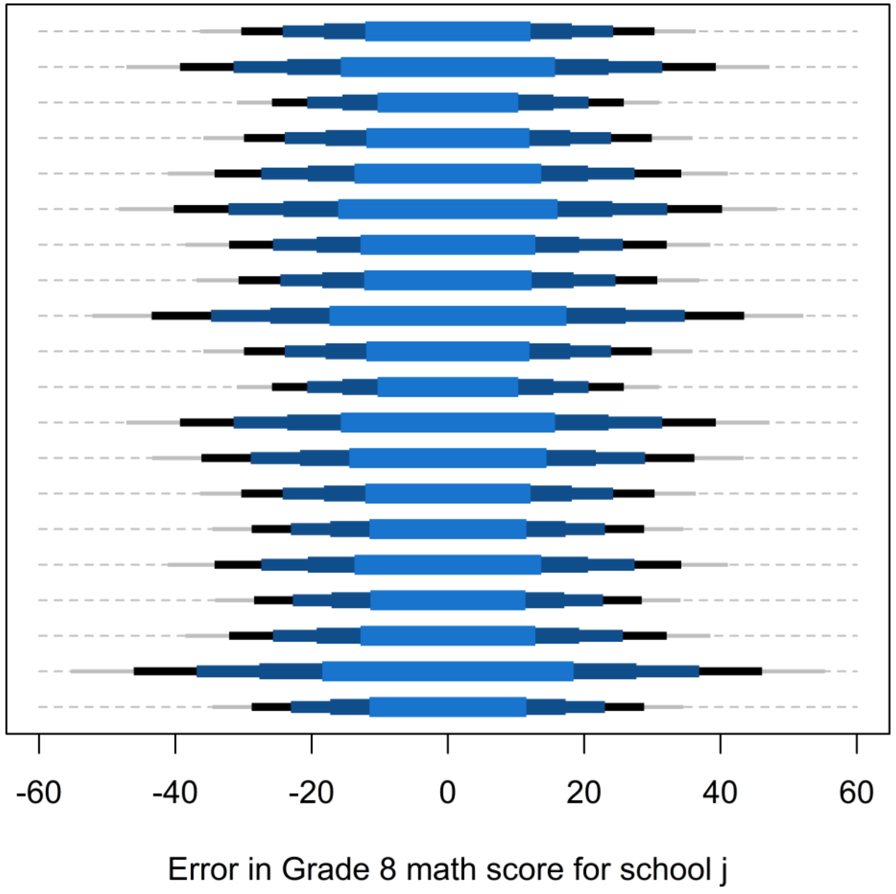
idschool	j	schn	schmath_j	V_j ($\hat{V}_j$)	SE_j
5036	1	41	484.25	132.68	11.52
5049	2	16	485.51	339.99	18.44
5050	3	33	535.46	164.85	12.84
5058	4	42	378.48	129.52	11.38
5103	5	29	448.60	187.58	13.70
5110	6	41	619.35	132.68	11.52
5112	7	37	375.71	147.02	12.13
5128	8	26	421.43	209.23	14.46
5143	9	22	561.16	247.27	15.72
5181	10	51	425.34	106.66	10.33
5182	11	38	506.35	143.16	11.96
5198	12	18	520.70	302.22	17.38
5199	13	36	475.74	151.11	12.29
5215	14	33	469.31	164.85	12.84
5239	15	21	413.62	259.04	16.09
5244	16	29	486.36	187.58	13.70
5252	17	38	529.84	143.16	11.96
5269	18	51	528.36	106.66	10.33
5271	19	22	479.34	247.27	15.72
5275	20	37	510.69	147.02	12.13

Challenge of the unbalanced case

- In the unbalanced case, the error terms are <u>not</u> identically distributed
- Using the balanced approach to estimate $\hat{\tau}_{00}$ and $\hat{\gamma}_{00}$ when the groups are unbalanced will produce biased results

Distribution of within-school error:

$$r_{ij} \sim N(0, \frac{\hat{\sigma}^2}{n_j})$$



A weighted least squares (WLS) approach

- Instead of simply using the mean of $\hat{\beta}_{0j}$ to estimate γ_{00}, using *precision weighting* produces the "minimum variance unbiased estimator of γ_{00}
- The weights are the inverse variance of the of $\hat{\beta}_{0j}$'s:

$$\frac{1}{\hat{\Delta}_j} = \frac{1}{\hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_j}}$$

- Use the model-estimated variance components to calculate the weights: $\hat{\tau}_{00} = 3{,}558$ and $\hat{\sigma}^2 = 5{,}440$

A weighted least squares (WLS) approach

- Precision-weighted grand mean:

$$\hat{\gamma}_{00} = \frac{\sum_j^J \frac{\hat{\beta}_{0j}}{\hat{\Delta}_j}}{\sum_j^J \frac{1}{\hat{\Delta}_j}}$$

$$\hat{\Delta}_j = \hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_j} = 3{,}558 + \frac{5{,}440}{n_j}$$

idschool	j	schn	schmath_j	V_j	SE_j	D_j ($\hat{\Delta}_j$)	wgtpct
5036	1	41	484.25	132.68	11.52	3,690.40	5.07
5049	2	16	485.51	339.99	18.44	3,897.71	4.80
5050	3	33	535.46	164.85	12.84	3,722.56	5.02
5058	4	42	378.48	129.52	11.38	3,687.24	5.07
5103	5	29	448.60	187.58	13.70	3,745.30	4.99
5110	6	41	619.35	132.68	11.52	3,690.40	5.07
5112	7	37	375.71	147.02	12.13	3,704.74	5.05
5128	8	26	421.43	209.23	14.46	3,766.94	4.96
5143	9	22	561.16	247.27	15.72	3,804.98	4.91
5181	10	51	425.34	106.66	10.33	3,664.38	5.10
5182	11	38	506.35	143.16	11.96	3,700.87	5.05
5198	12	18	520.70	302.22	17.38	3,859.93	4.84
5199	13	36	475.74	151.11	12.29	3,708.82	5.04
5215	14	33	469.31	164.85	12.84	3,722.56	5.02
5239	15	21	413.62	259.04	16.09	3,816.76	4.90
5244	16	29	486.36	187.58	13.70	3,745.30	4.99
5252	17	38	529.84	143.16	11.96	3,700.87	5.05
5269	18	51	528.36	106.66	10.33	3,664.38	5.10
5271	19	22	479.34	247.27	15.72	3,804.98	4.91
5275	20	37	510.69	147.02	12.13	3,704.74	5.05

Small group discussion

- In groups of 3-4, take 10 minutes to discuss ...
  - Under what conditions will using the balanced approach instead of weighted least squares approach produce more biased results? Consider the following factors:
    - How the within-group sample size differs across groups
    - The magnitude of between-group variance relative to within-group variance
    - The number of level-2 units
  - In what ways do you think these conditions are common, or could be a factor, in the area(s) of research you are interested in?
- Then share out with the whole class

Empirical Bayes Estimates of Group Means

Group mean estimation

- We want to estimate the true school mean math score (β_{0j})
- Option 1 is based on the level-1 model:

The observed school mean is an estimate of the school's true score

$$\bar{Y}_{.j} = \beta_{0j} + \bar{r}_{.j}\ , \qquad \bar{r}_{.j} \sim N\left(0, \frac{\sigma^2}{n_j}\right)$$

- Option 2 is based on the level-2 model:

$$\beta_{0j} = \gamma_{00} + u_{0j}\ , \qquad u_{0j} \sim N(0, \tau_{00})$$

The grand mean is an estimate of the school's true score

Group mean estimation

- Option 3 is an "optimal" weighted average of the two other options:

Called a shrinkage estimator, an empirical Bayes (EB) estimate, or the best linear unbiased predictor (BLUP)

$$\beta_{0j}^{*} = \lambda_j \bar{Y}_{.j} + (1 - \lambda_j)\hat{\gamma}_{00}$$

- The weight, λ_j, is the reliability of the of the observed means ($\bar{Y}_{.j}$):

$$\lambda_j = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_j}}$$

Group mean estimation

- The more reliable the observed mean, the more weight is given to $\bar{Y}_{.j}$
- The more concentrated group means are around the grand mean (smaller $\hat{\tau}_{00}$), the more weight is given to $\hat{\gamma}_{00}$
- However, the EB estimate (β^*_{0j}) is biased toward $\hat{\gamma}_{00}$ … it "over shrinks" β_{0j} toward $\hat{\gamma}_{00}$

$$\beta^*_{0j} = \lambda_j \bar{Y}_{.j} + (1 - \lambda_j)\hat{\gamma}_{00}$$

$$\lambda_j = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_j}}$$

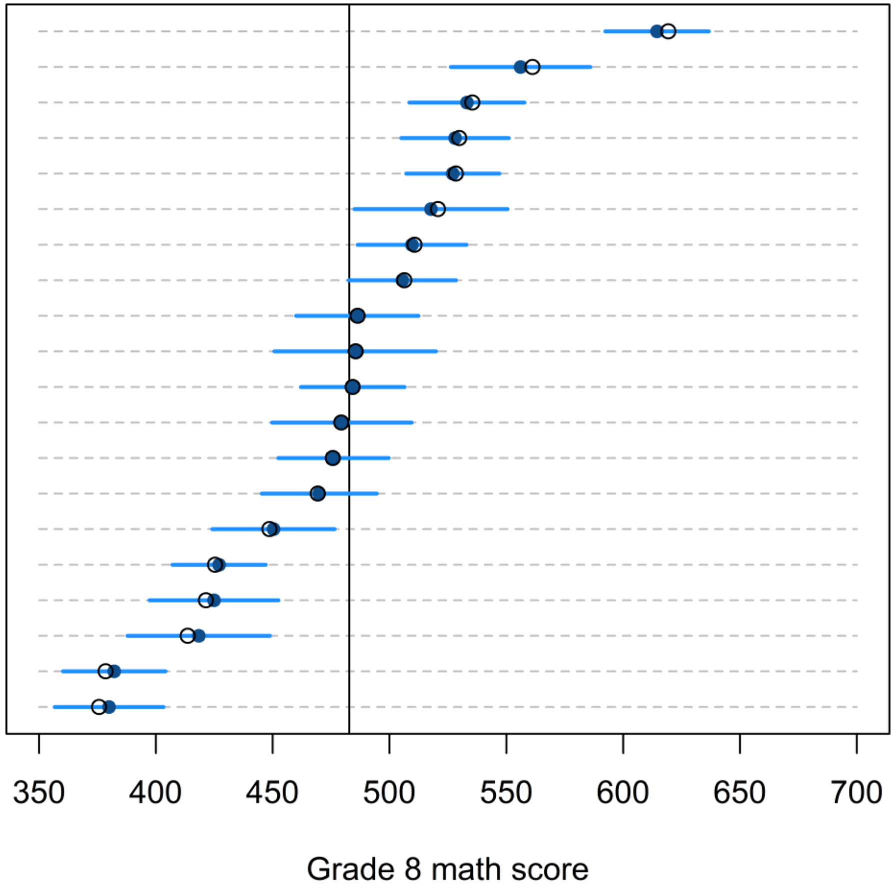
Note: variance formula for β^*_{0j} is not straightforward, but the main takeaway is that it's a weighted average of the variance for $\bar{Y}_{.j}$ and $\hat{\gamma}_{00}$

$$V_j^* = \left(V_j^{-1} + \tau_{00}^{-1}\right)^{-1} + (1 - \lambda_j)^2 \left(\frac{1}{\sum_j^J \frac{1}{\hat{\tau}_{00} + \frac{\hat{\sigma}^2}{n_j}}}\right)^{-1}$$

Group mean estimation

idschool	j	schn	schmath_j	V_j	D_j	EB_j (β^*_{0j})
5036	1	41	484.25	132.68	3,690.40	484.20
5049	2	16	485.51	339.99	3,897.71	485.27
5050	3	33	535.46	164.85	3,722.56	533.12
5058	4	42	378.48	129.52	3,687.24	382.15
5103	5	29	448.60	187.58	3,745.30	450.31
5110	6	41	619.35	132.68	3,690.40	614.44
5112	7	37	375.71	147.02	3,704.74	379.96
5128	8	26	421.43	209.23	3,766.94	424.83
5143	9	22	561.16	247.27	3,804.98	556.06
5181	10	51	425.34	106.66	3,664.38	427.01
5182	11	38	506.35	143.16	3,700.87	505.43
5198	12	18	520.70	302.22	3,859.93	517.73
5199	13	36	475.74	151.11	3,708.82	476.02
5215	14	33	469.31	164.85	3,722.56	469.90
5239	15	21	413.62	259.04	3,816.76	418.31
5244	16	29	486.36	187.58	3,745.30	486.18
5252	17	38	529.84	143.16	3,700.87	528.02
5269	18	51	528.36	106.66	3,664.38	527.03
5271	19	22	479.34	247.27	3,804.98	479.56
5275	20	37	510.69	147.02	3,704.74	509.58

EB estimate of school mean math scores (β^*_{0j}) and 95% confidence interval



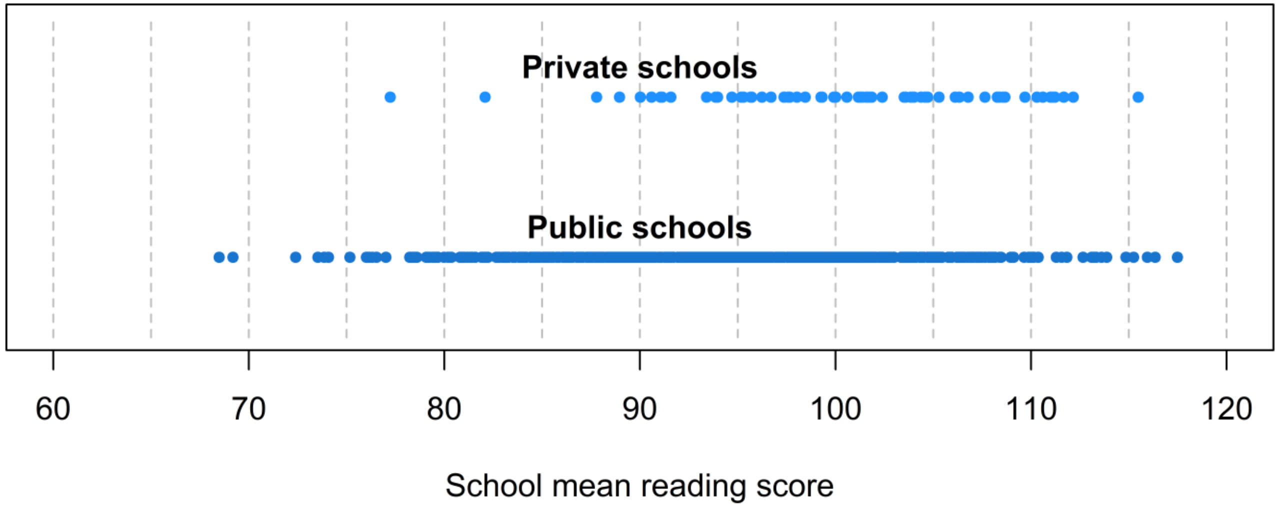
Means as Outcomes

Motivating example

- How much does early grade reading achievement differ between private and public U.S. schools?
- Use sample of data from the ECLS-K:2011
  - The data include 11,091 first grade students, 742 schools
  - Number of students per school ranges from 10 to 25 students
  - 9% of the schools are private schools

Some data descriptives

	Public Schools (N=672)	Private Schools (N=70)	Overall (N=742)
School mean reading score ($\bar{Y}_{.j}$)			
Mean (SD)	94.6 (8.33)	101 (7.48)	95.2 (8.46)
Median [Min, Max]	95.0 [68.5, 117]	102 [77.2, 115]	95.5 [68.5, 117]
School mean SES			
Mean (SD)	-0.0982 (0.517)	0.437 (0.418)	-0.0477 (0.532)
Median [Min, Max]	-0.144 [-1.28, 1.32]	0.516 [-0.798, 1.18]	-0.0576 [-1.28, 1.32]



Naïve approach #1: Student-level analysis

- Estimate an OLS regression based on the student-level data:

$$Y_{ij} = \beta_0 + \beta_1 Sector_j + r_{ij}$$

		Estimate	Standard Error	t value	Pr(>|t|)	
$\hat{\beta}_0$	(Intercept)	94.847	0.175	542.565	0.0000	***
$\hat{\beta}_1$	sector	6.553	0.566	11.584	0.0000	***

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 17.51 on 11089 degrees of freedom
Multiple R-squared: 0.01196, Adjusted R-squared: 0.01187
F-statistic: 134.2 on 11089 and 1 DF, p-value: 0.0000

- What's the predicted value for the average public school (*sector* = 0)?
- What's the predicted values for the average private school (*sector* = 1)?

Naïve approach #1: Student-level analysis

- Standard errors are too small → will lead to invalid inferences

- Degrees of freedom are inflated → each student observation is assumed to be independent

	Estimate	Standard Error	t value	Pr(>|t|)	
(Intercept)	94.847	0.175	542.565	0.0000	***
sector	6.553	0.566	11.584	0.0000	***

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 17.51 on 11089 degrees of freedom
Multiple R-squared: 0.01196, Adjusted R-squared: 0.01187
F-statistic: 134.2 on 11089 and 1 DF, p-value: 0.0000

Naïve approach #2: School-level analysis

- Estimate an OLS regression based on the school-level data:

$$\bar{Y}_{.j} = \beta_0 + \beta_1 Sector_j + e_j$$

		Estimate	Standard Error	t value	Pr(>|t|)	
$\hat{\beta}_0$	(Intercept)	94.586	0.318	297.148	0.0000	***
$\hat{\beta}_1$	sector	6.513	1.036	6.284	0.0000	***

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 8.252 on 740 degrees of freedom
Multiple R-squared: 0.05066, Adjusted R-squared: 0.04938
F-statistic: 39.49 on 740 and 1 DF, p-value: 0.0000

- What's the predicted value for the average public school (*sector* = 0)?
- What's the predicted values for the average private school (*sector* = 1)?

Naïve approach #2: School-level analysis

- Standard error of $\hat{\beta}_1$:

$$\sqrt{\frac{Residual\ variance}{SS\ of\ Predictor}} =$$

$$\sqrt{\frac{\hat{\Delta}}{\sum_{j=1}^{742}\left(Sector_j - \overline{Sector}\right)^2}}$$

		Estimate	Standard Error	t value	Pr(>|t|)	
$\hat{\beta}_0$	(Intercept)	94.586	0.318	297.148	0.0000	***
$\hat{\beta}_1$	sector	6.513	$\boxed{1.036}$	6.284	0.0000	***

*Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05*

Residual standard error: 8.252 on 740 degrees of freedom
Multiple R-squared: 0.05066, Adjusted R-squared: 0.04938
F-statistic: 39.49 on 740 and 1 DF, p-value: 0.0000

Reflects remaining parameter variance + error variance

Multilevel approach

- Level-1 (within-school) model:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

- Level-2 (between-school) model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Sector_j + u_{0j}$$
$$u_{0j} \sim N(0, \tau_{00})$$

Some Definitions

- γ_{00}: the expected school mean achievement for the average public school
- γ_{01}: the expected difference in school mean achievement between private and public schools
- u_{0j}: the random effect for school j; the deviation of the true mean achievement score for school j from an expected value based on school j's sector
- τ_{00}: the variance in true school means conditional on sector; or the parameter variance that remains after taking into account sector

Multilevel approach

- Level-1 (within-school) model:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

- Level-2 (between-school) model:

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

- What's the predicted value for the average public school (*sector* = 0)?
- What's the predicted values for the average private school (*sector* = 1)?

group		Estimate	Standard Error	df	t value	p-value	
		Fixed effects					
$\hat{\gamma}_{00}$	(Intercept)	94.657	0.317	733	298.818	0.0000	***
$\hat{\gamma}_{01}$	sector	6.512	1.030	729	6.323	0.0000	***
		Random effects					
schid	sd__(Intercept)	7.046					
Residual	sd__Observation	16.052					

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

square root of the estimated residual variance: 16.1
data's log-likelihood under the model: -47,021.1
Akaike Information Criterion: 94,050.3
Bayesian Information Criterion: 94,079.5

Multilevel approach

- Standard errors are a little smaller than with the school-level analysis
- Standard error of $\hat{\gamma}_{01}$ uses the school-level residual variance estimate ($\hat{\tau}_{00}$) instead of $\widehat{\Delta}$ → excludes error variance

	group		Estimate	Standard Error	df	t value	p-value	
			Fixed effects					
$\hat{\gamma}_{00}$	(Intercept)		94.657	0.317	733	298.818	0.0000	***
$\hat{\gamma}_{01}$	sector		6.512	$\boxed{1.030}$	729	6.323	0.0000	***
			Random effects					
schid	sd__(Intercept)		7.046					
Residual	sd__Observation		16.052					

$\hat{\tau}_{00}$ = residual parameter variance

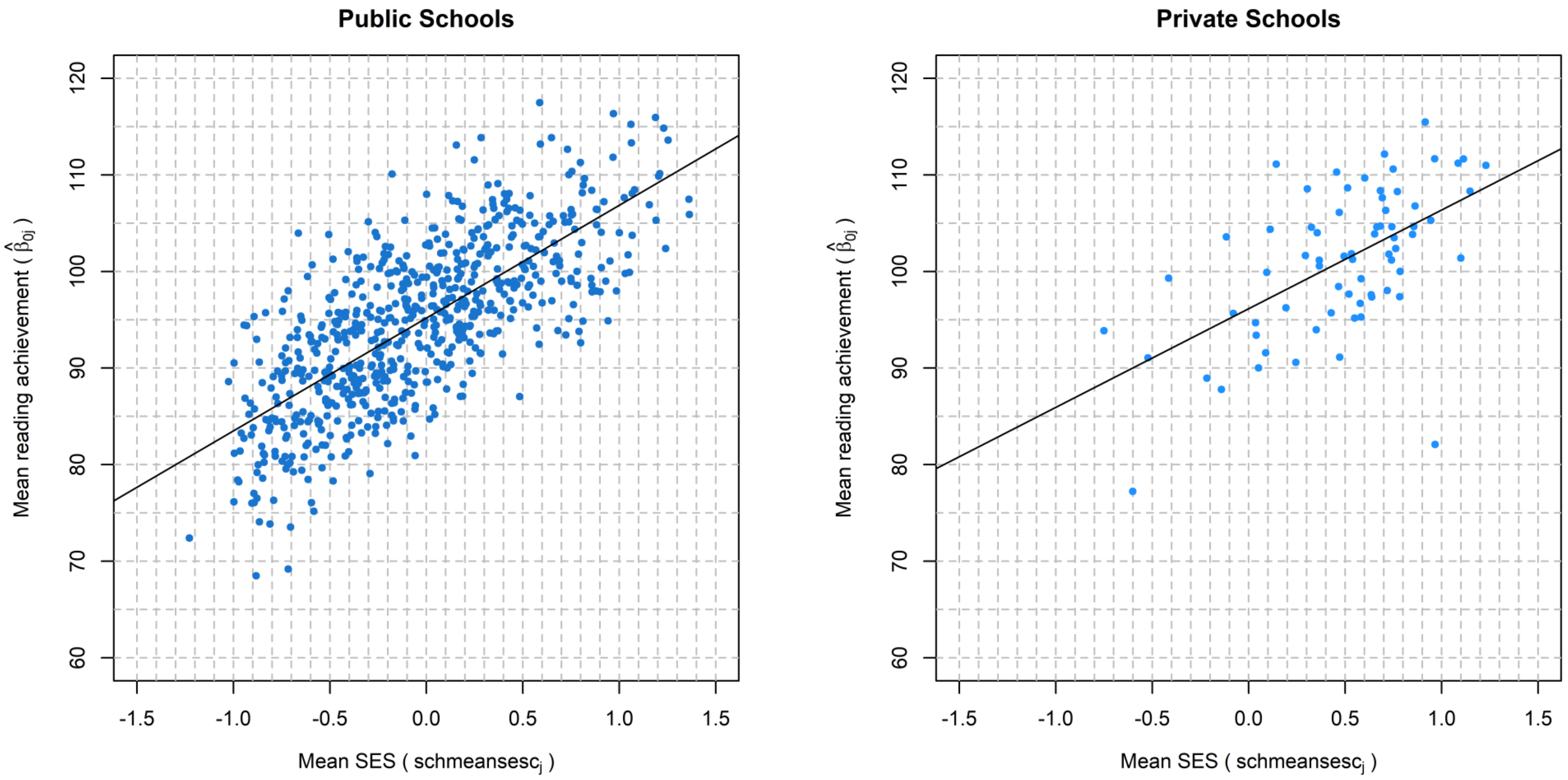
Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05

square root of the estimated residual variance: 16.1
data's log-likelihood under the model: -47,021.1
Akaike Information Criterion: 94,050.3
Bayesian Information Criterion: 94,079.5

Small group exercise, Part 1

- Could the private-public school difference in reading achievement be an artifact of differences in family resources?
- In groups of 3-4, take 10 minutes to discuss the following questions about the graphs on the next slide:
  - Do you think public and private schools serve similar types of families? Why or why not?
  - Do you think that matters when it comes to comparing the reading performance of public and private schools? Why or why not?
  - What do you predict is the mean math reading score for a <u>public</u> school with an average SES student body?
  - What do you predict is the mean math reading score for a <u>private</u> school with an average SES student body?

Small group exercise, Part 1



Small group exercise, Part 2

- In your groups, estimate the following multilevel model using the *lmer* R function and discuss the questions on the next slide
- Level-1 (within-school) model:

$$g1rscore_{ij} = \beta_{0j} + r_{ij}$$

- Level-2 (between-school) model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} sector_j + \gamma_{02} schmeansesc_j + u_{0j}$$

Small group exercise, Part 2

- Define the following parameters in the multilevel model:
  - γ_{00}:
  - γ_{01}:
  - γ_{02}:
  - u_{0j}:
  - τ_{00}:
- What's the expected difference between private and public school reading achievement conditional on the school's mean family SES?
- Given these results, what would you conclude about the early reading performance of private versus public schools?