The Random Intercept Model

PSYC 575

August 6, 2020 (updated: 26 November 2022)

Week Learning Objectives

- Explain the components of a **random intercept model**
- Interpret **intraclass correlations**
- Use the **design effect** to decide whether MLM is needed
- Explain why ignoring clustering (e.g., regression) leads to inflated chances of Type I errors
- Describe how MLM **pools information** to obtain more stable inferences of groups

Data 1982 High School and Beyond Survey[1]

- 7,185 students (10-12th graders) from 160 schools (90 public and 70 Catholic)

- Level 1: Student
  - `id`: group identifier
  - `minority`: (1 = minority, 0 = not)
  - `female`: 1 = female, 0 = male
  - `ses`
  - `mathach`: Mathematics achievement

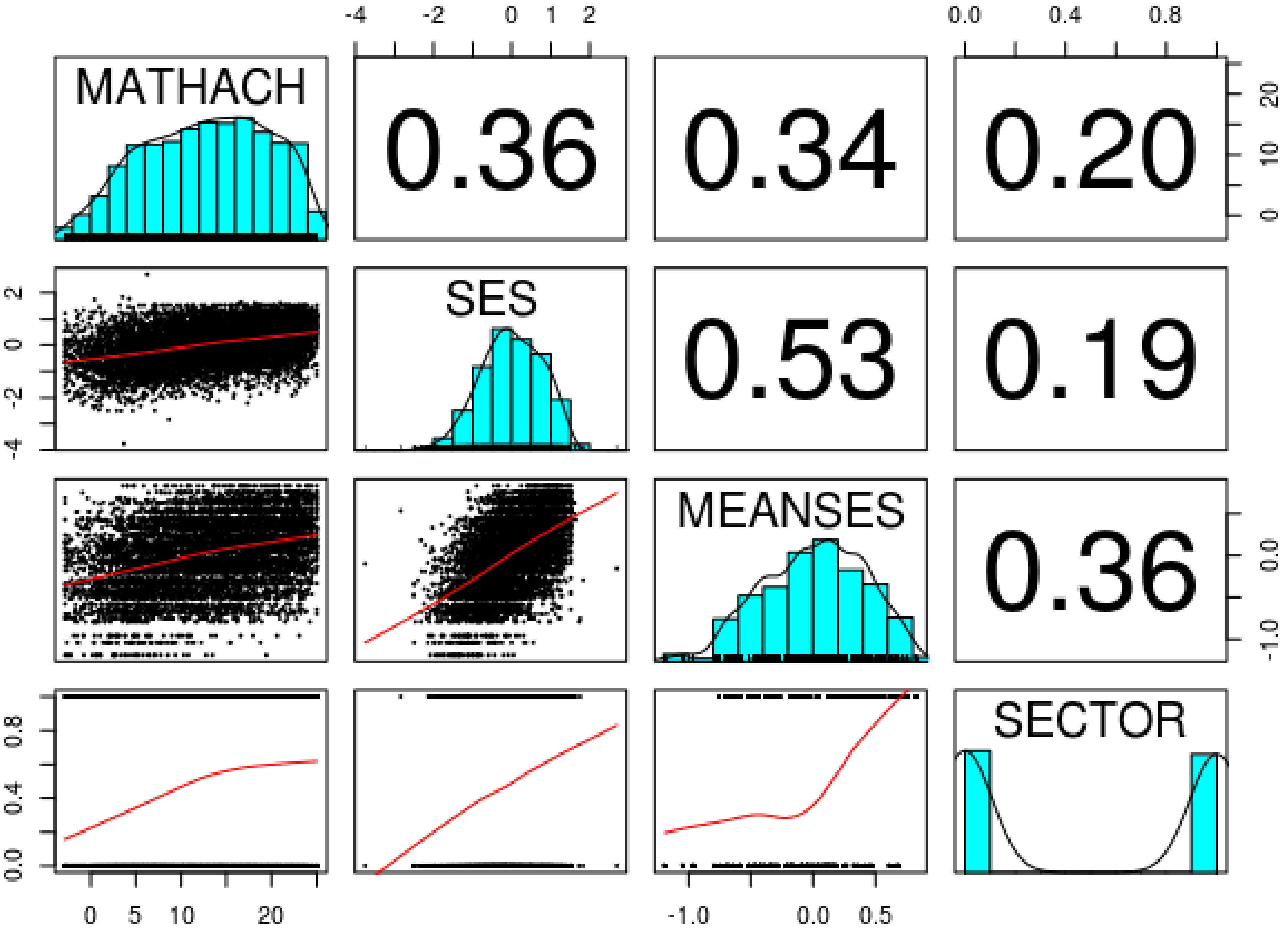
- Level 2: School
  - `size`: school size
  - `sector` (1 = Catholic, 0 = Public)
  - `pracad`: proportion in academic track
  - `disclim`: disciplinary climate
  - `himnty`: 1 = > 40% minority, 0 = < 40% minority
  - `meanses`: mean of Lv-1 SES

[1]: Check https://nces.ed.gov/surveys/hsb/ for more information

	ID	MINORITY	FEMALE	SES	MATHACH	SIZE	SECTOR	PRACAD	DISCLIM	HIMINTY	MEANSES
1	1224	0	1	-1.528	5.876	842	0	0.35	1.597	0	-0.428
2	1224	0	1	-0.588	19.708	842	0	0.35	1.597	0	-0.428
3	1224	0	0	-0.528	20.349	842	0	0.35	1.597	0	-0.428
4	1224	0	0	-0.668	8.781	842	0	0.35	1.597	0	-0.428
5	1224	0	0	-0.158	17.898	842	0	0.35	1.597	0	-0.428
6	1224	0	0	0.022	4.583	842	0	0.35	1.597	0	-0.428
7	1224	0	1	-0.618	-2.832	842	0	0.35	1.597	0	-0.428
8	1224	0	0	-0.998	0.523	842	0	0.35	1.597	0	-0.428
9	1224	0	1	-0.888	1.527	842	0	0.35	1.597	0	-0.428
10	1224	0	0	-0.458	21.521	842	0	0.35	1.597	0	-0.428

Student-level variables **School-level variables**

	ID	MINORITY	FEMALE	SES	MATHACH	SIZE	SECTOR	PRACAD	DISCLIM	HIMINTY	MEANSES
996	2458	1	1	0.852	22.743	545	1	0.89	-1.484	1	0.234
997	2458	1	1	0.262	17.205	545	1	0.89	-1.484	1	0.234
998	2458	1	1	0.052	12.071	545	1	0.89	-1.484	1	0.234
999	2458	1	1	-0.468	19.161	545	1	0.89	-1.484	1	0.234
1000	2458	1	1	-0.268	12.332	545	1	0.89	-1.484	1	0.234
1001	2458	0	1	1.512	22.681	545	1	0.89	-1.484	1	0.234
1002	2458	1	1	0.182	4.928	545	1	0.89	-1.484	1	0.234
1003	2458	1	1	0.242	9.142	545	1	0.89	-1.484	1	0.234
1004	2458	0	1	1.072	24.488	545	1	0.89	-1.484	1	0.234
1005	2458	1	1	1.172	13.666	545	1	0.89	-1.484	1	0.234


MATHACH
SES
MEANSES
SECTOR
0.36
0.34
0.20
0.53
0.19
0.36

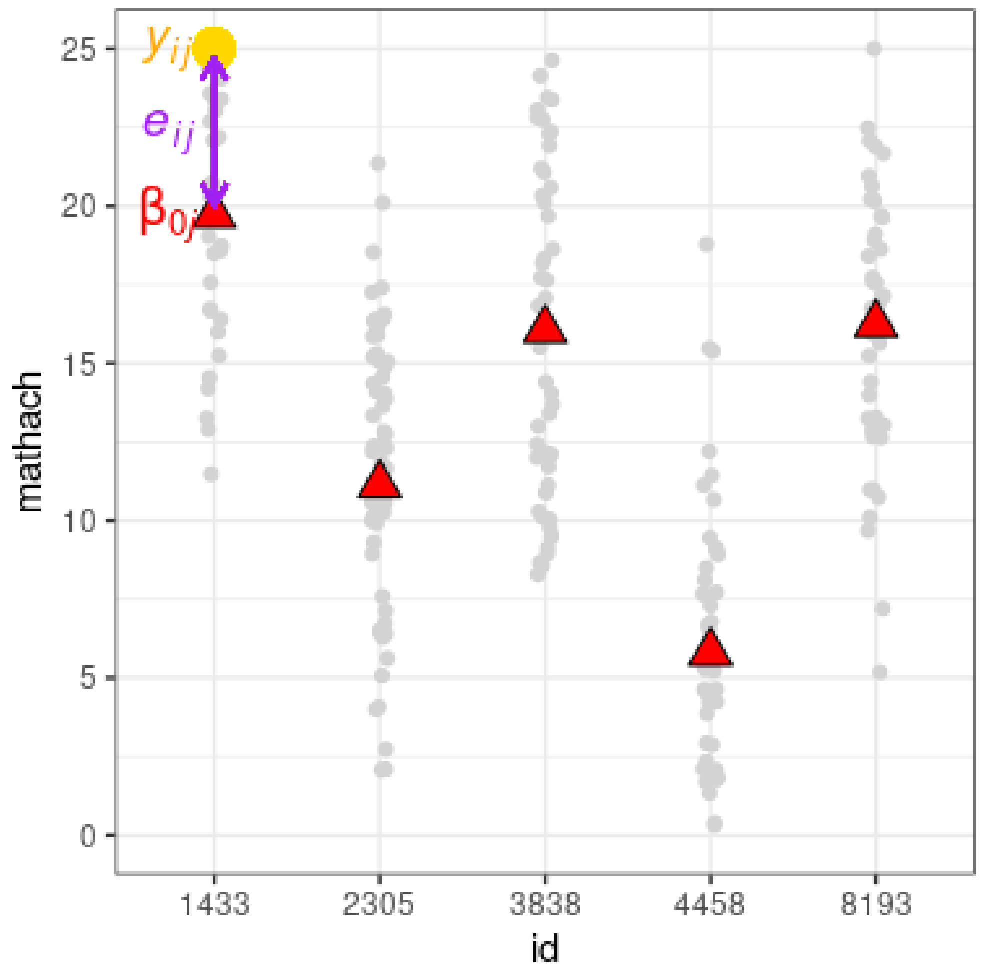
Research Questions

- Does math achievement vary across schools? How much is the variation?
- Do schools with higher mean SES have students with higher math achievement?

Random Intercept Model

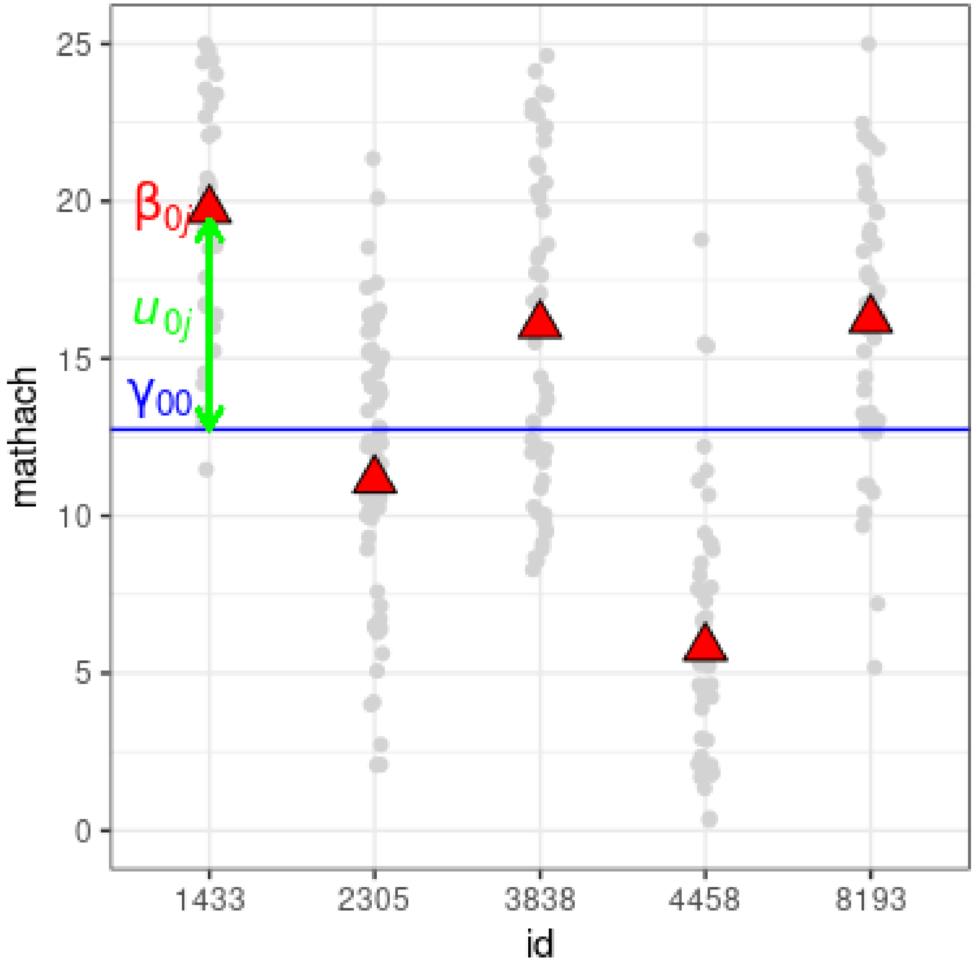
(Unconditional) Random Intercept Model

- Student level (Lv 1)
  - $\text{mathach}_{ij} = \beta_{0j} + e_{ij}$



(Unconditional) Random Intercept Model

- School level (Lv 2)
  - $\beta_{0j} = \gamma_{00} + u_{0j}$



(Unconditional) Random Intercept Model

- Student level (Lv 1)
  - $\text{mathach}_{ij} = \beta_{0j} + e_{ij}$
- School level (Lv 2)
  - $\beta_{0j} = \gamma_{00} + u_{0j}$

Combined:

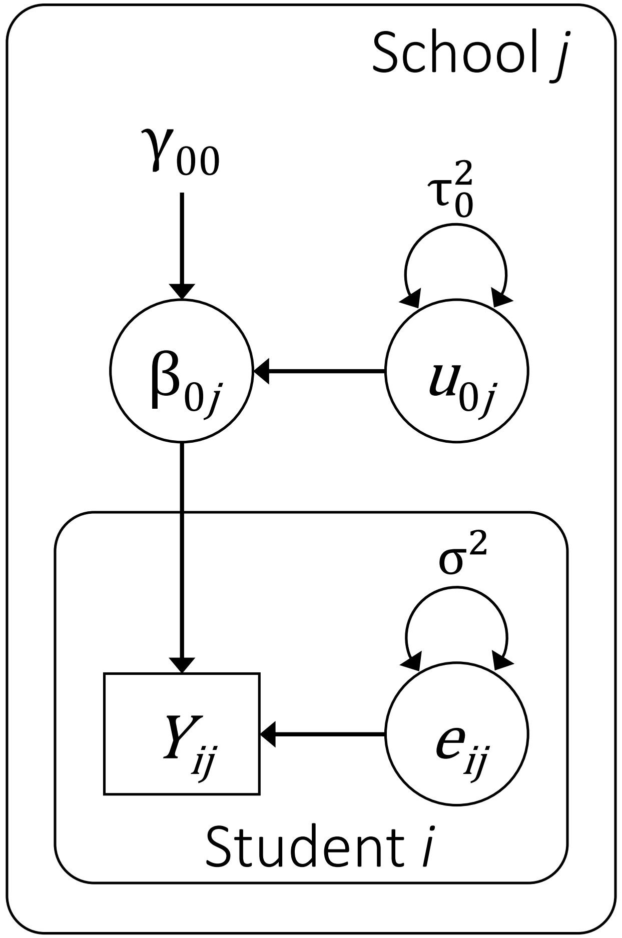
$$\text{mathach}_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

Score of student *i* in school *j*

= Grand mean (γ_{00}) + school deviation (u_{0j})
+ student deviation (e_{ij})

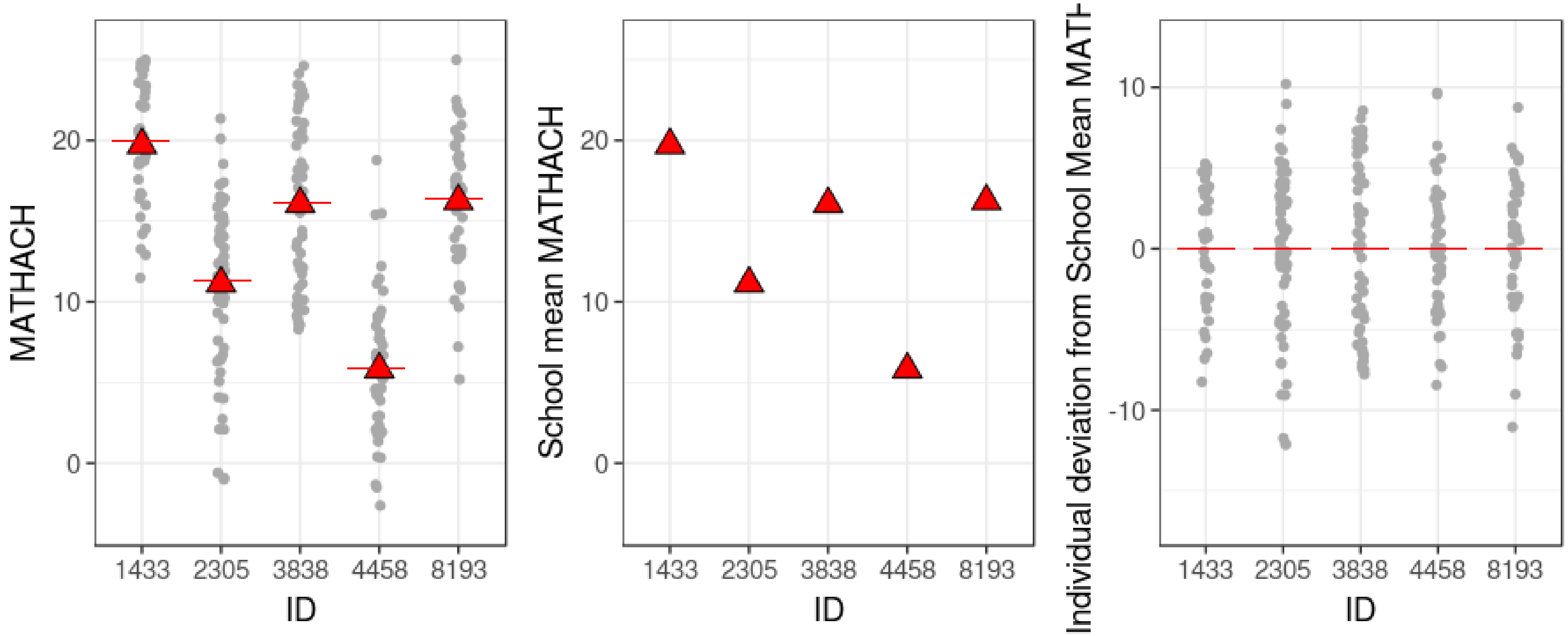
Model Diagram

- Student level (Lv 1)
  - $mathach_{ij} = \beta_{0j} + e_{ij}, \quad e_{ij} \sim N(0, \sigma)$
- School level (Lv 2)
  - $\beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim N(0, \tau_0)$
- Combined:
  - $mathach_{ij} = \gamma_{00} + u_{0j} + e_{ij}$



Decomposing School- and Student-Level Information

- `mathach` = School info + Student info (Relative to School)



Terminology

- Fixed effects (γ): constant for everyone
- Random effects (e_{ij}, u_{0j}): varies for different observations/clusters
  - Describe by some probability distributions (e.g., normal)
  - Variance components: variance of random effects

Fixed Effects (R Output)

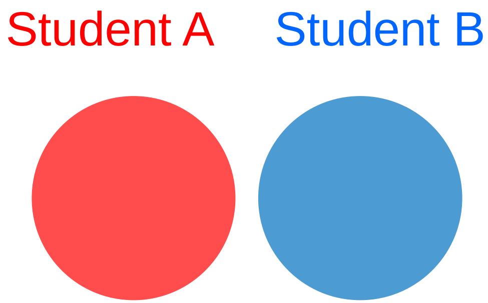
```
>#  Fixed effects:
>#              Estimate Std. Error t value
>#  (Intercept)  12.6370     0.2444   51.71
```

The estimated grand mean of MATHACH for all students is γ_{00} **= 12.64**, *SE* = **0.24**

Intraclass Correlation

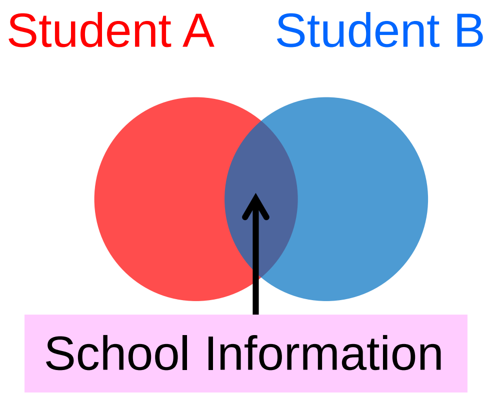
Intraclass Correlations (ICC; ρ)

- Independent



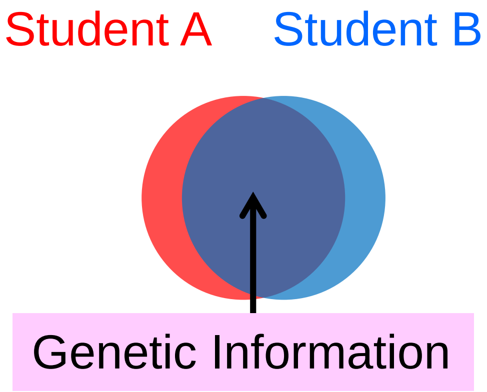
- ICC = 0

- Weakly Correlated



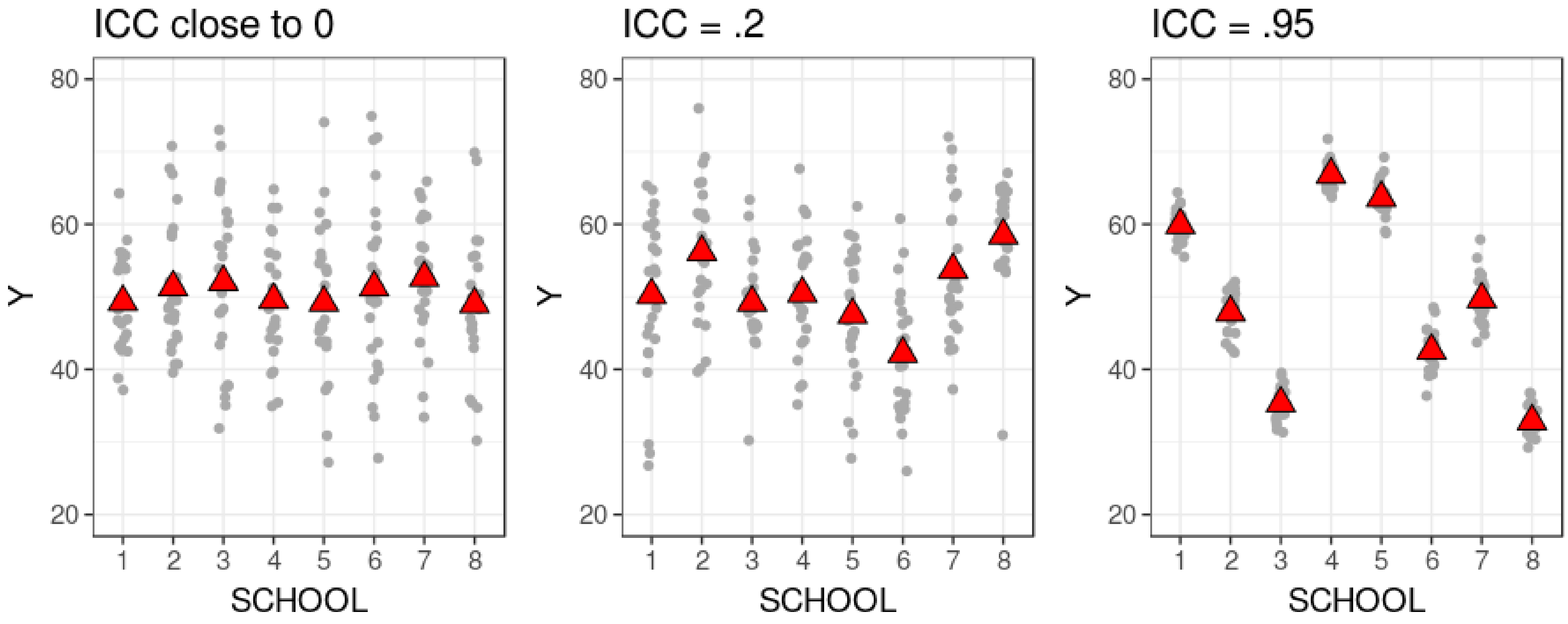
- ICC = .2

- Strongly Correlated



- ICC = .8

- ICC =
  1. Proportion of variance due to the higher (school-) level
  2. Average correlation between observations (students) in the same cluster (school)

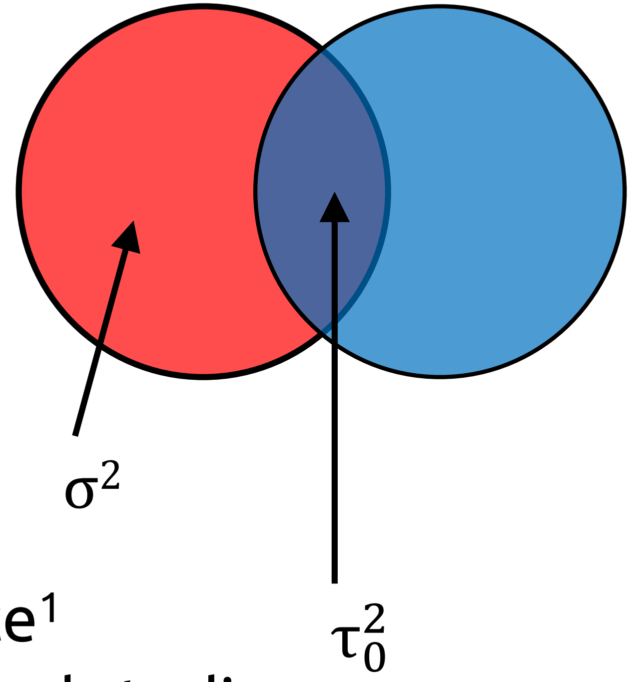


Variance Components

- Var(u_{0j}) = τ_0^2 = between-school variance
- Var(e_{ij}) = σ^2 = within-school variance
- ICC:

$$\rho = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$$

  - Typical ICC = .1 to .25 for educational performance[1]
  - Higher ICCs for repeated measures and longitudinal studies



[1]: Hedges and Hedberg (2007), https://doi.org/10.3102/0162373707299706

R Output

```
># Random effects:
>#  Groups    Name        Variance Std.Dev.
>#  id        (Intercept)  8.614    2.935
>#  Residual              39.148    6.257
># Number of obs: 7185, groups:  id, 160
```

Variance of school means = 8.61
Variance of individual scores within a school = 39.15
ICC = 8.61 / (8.61 + 39.15) = **0.18**

Question: Does math achievement varies across schools? How much is the variation?

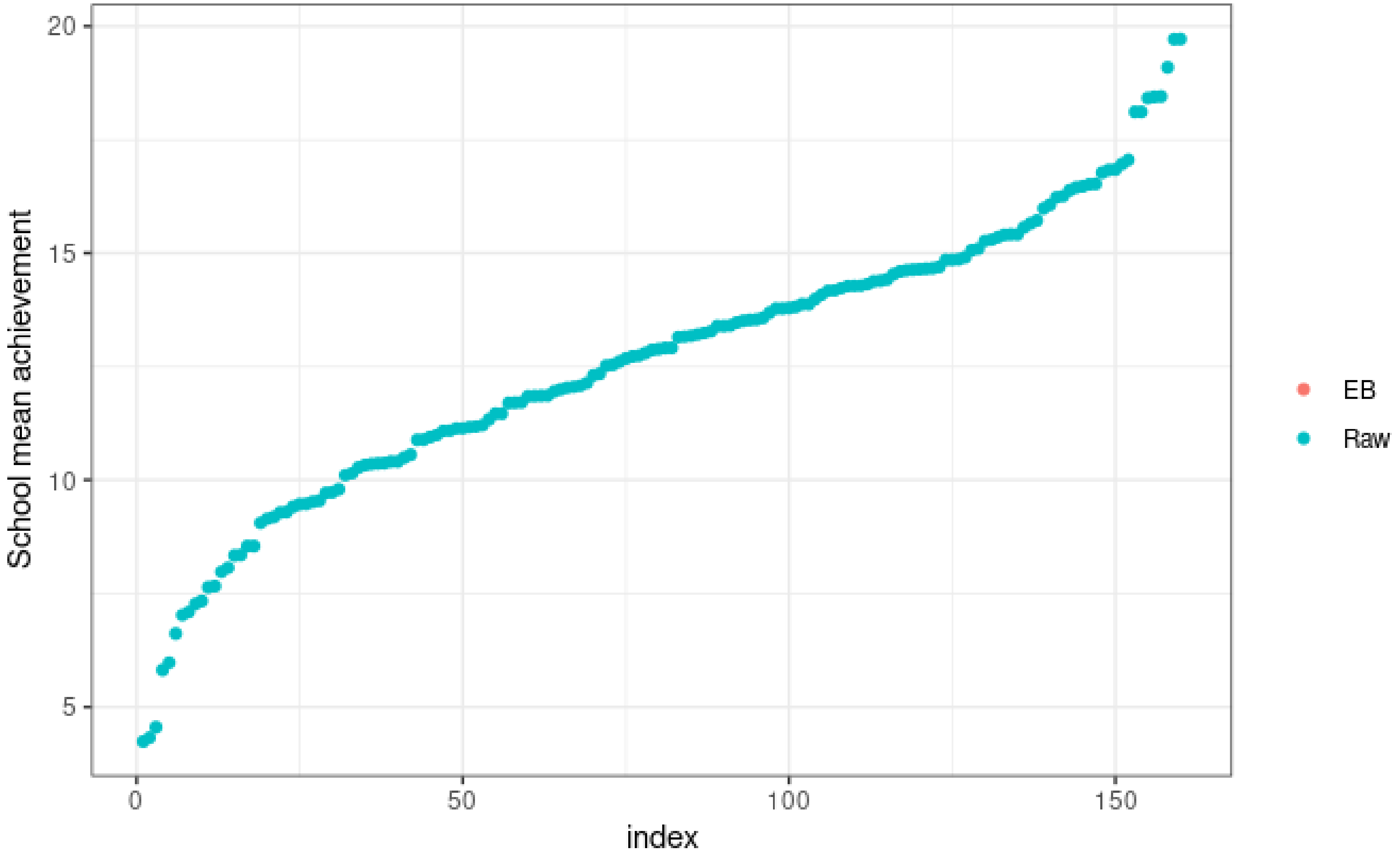
- Yes, there is evidence that students’ math achievement varies across schools.
- Variability at the school level accounts for 18% of the total variability of math achievement

Empirical Bayes Estimates

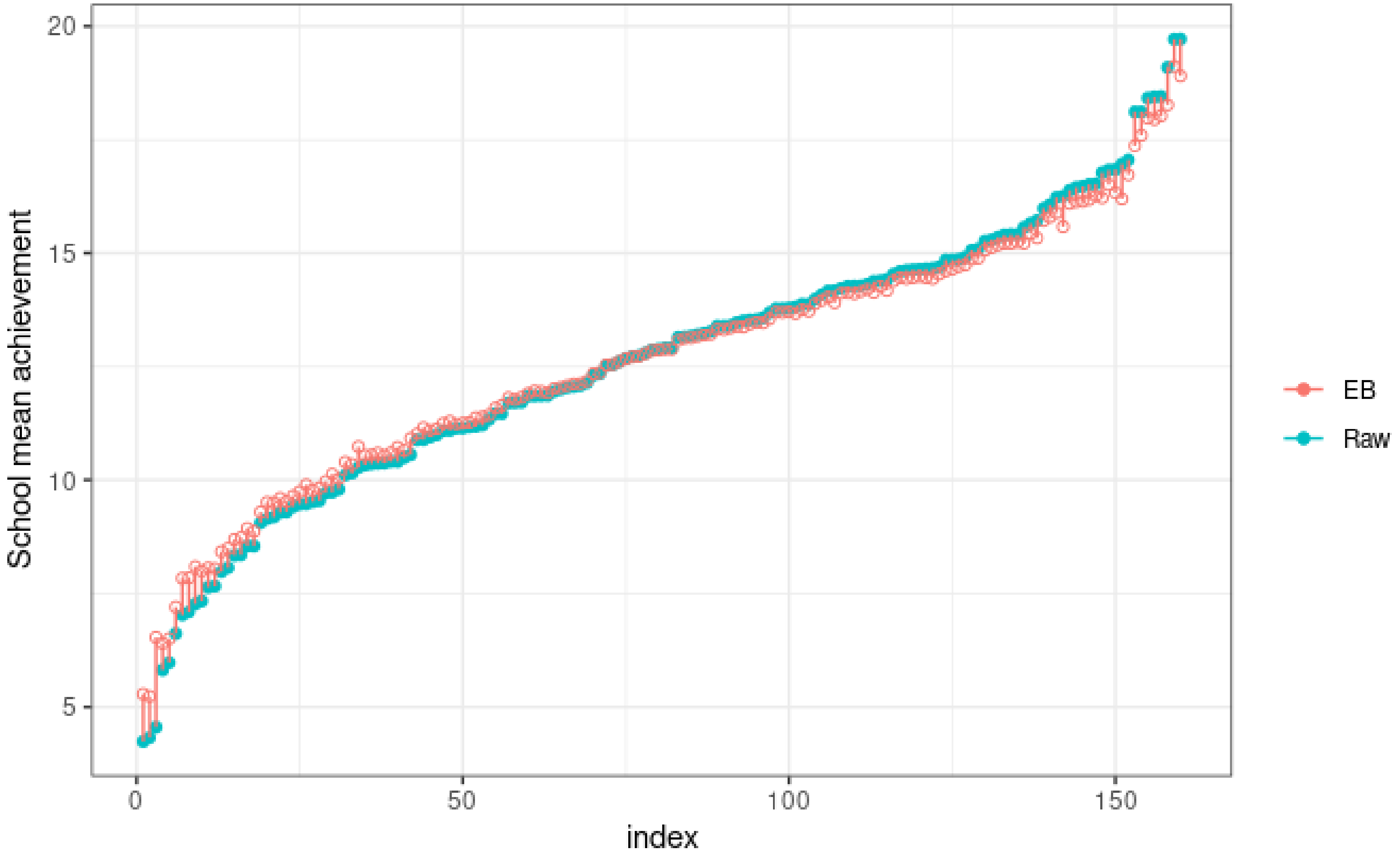
MLM Borrows Information

- β_{0j} = (population) mean math achievement of school *j*
- Most straightforward way to estimate β_{0j}:
  - Take the average of everyone in the sample in school *j*
- It may be unstable in small samples
- Instead, MLM borrows information from other schools

Also called *Shrinkage estimates*, *Best unbiased linear predictor* (BLUP), *Posterior modes*



Also called *Shrinkage estimates*, *Best unbiased linear predictor* (BLUP), *Posterior modes*



Empirical Bayes "Estimates"

$$\hat{\beta}_{0j}^{\text{EB}} = \lambda_j \hat{\beta}_{0j}^{\text{OLS}} + (1 - \lambda_j)\gamma_{00},$$

where

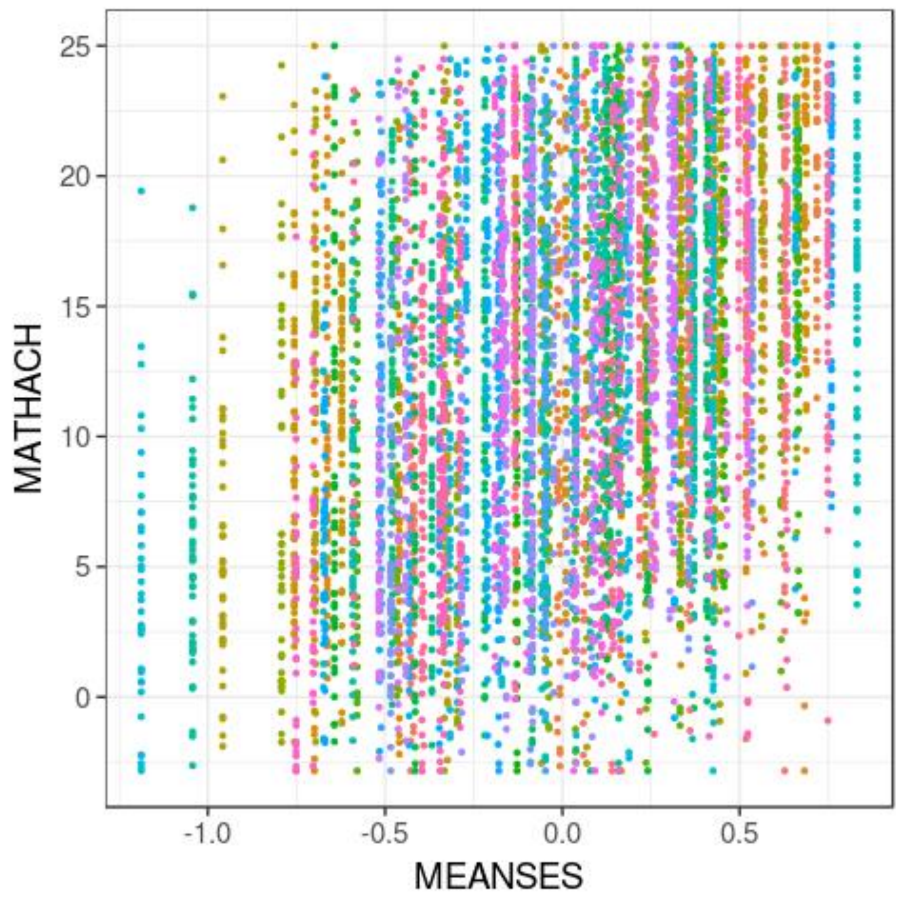
- $\lambda_j = \tau_0^2/(\tau_0^2 + \sigma^2/n_j)$ = reliability of group means
- In practice, the variance components need to be estimated

- Think: what happens when ICC = 0 (i.e., τ_0^2 = 0)? Or ICC = 1 (i.e., σ^2 = 0)?
- Read more on Snijders & Bosker, 4.8

Do schools with higher mean SES have students with higher math achievement?

Adding Predictors

- Why do some schools have higher mean math achievement than others?



Why Not Simple Regression?

- `mathach` and `meanses` are at different levels
- Two (problematic) approaches:
  - <u>Disaggregation</u> (both variables as lv 1)
  - Aggregation (both variables as lv 2)

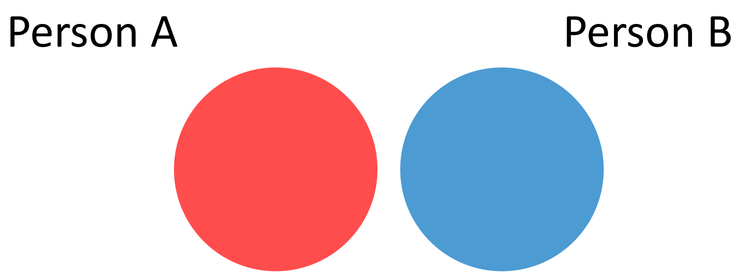
Problem of Disaggregation

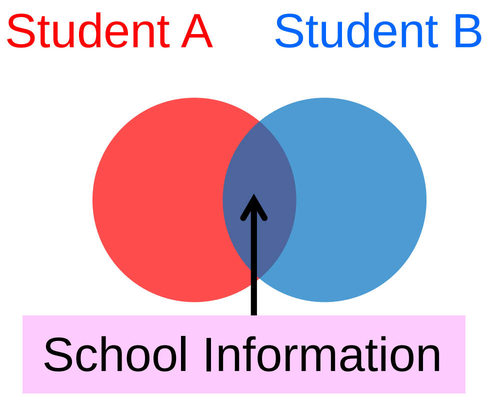
“Miraculous multiplication of the number of units”
(Snijders & Bosker, p. 16)

- Only 160 schools, but regression uses $N = 7{,}185$

Dependent Observations

- Regression assumes *independent* observations



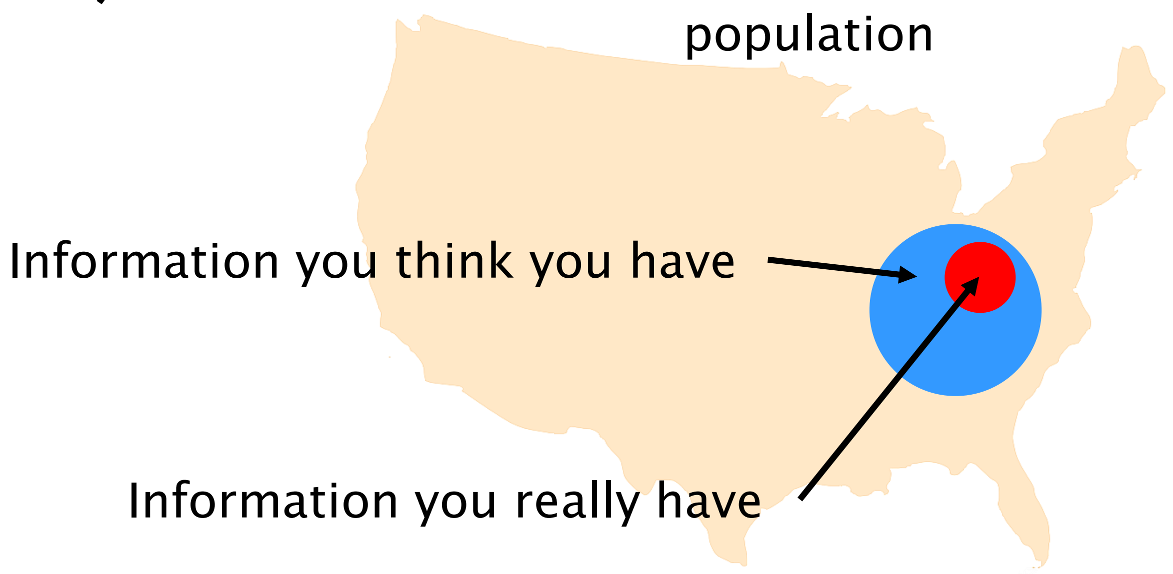


Design Effect

Design Effect (*Deff*)

- Dependent observations ➔ reduces information
  - Depends on overlap (ICC)

- $Deff = 1 + (\text{average cluster size} - 1) \times \text{ICC}$
- $N_{eff} = N / Deff$



Underestimated Standard Error

- OLS on 7,185 students

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.71276    0.07622  166.80   <2e-16 ***
meanses      5.71680    0.18429   31.02   <2e-16 ***
```

- MLM

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6494     0.1493   84.74
meanses       5.8635     0.3615   16.22
```

$$t = \frac{\text{Est}}{SE}$$

(Optional Material)
Approximate Standard Errors

- N = 7,185 students; J = 160 schools
- $s^2_{\text{means es}}$ = .170 = variance of MEANSES

```
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  2.639   1.624
 Residual             39.157   6.258
Number of obs: 7185, groups:  id, 160
```

Approximate Standard Errors

- $SE_{\text{OLS}} \approx \sqrt{\frac{1}{s^2_{\text{MEANSES}}}\left(\frac{\tau_0^2+\sigma^2}{N}\right)} = \sqrt{\frac{1}{.170}\left(\frac{2.639+39.157}{7185}\right)} = .185$

τ_0^2 (lv-2) is divided by an incorrect sample size (lv-1)

- $SE_{\text{MLM}} \approx \sqrt{\frac{1}{s^2_{\text{MEANSES}}}\left(\frac{\tau_0^2}{J} + \frac{\sigma^2}{N}\right)}$

$$= \sqrt{\frac{1}{.170}\left(\frac{2.639}{160} + \frac{39.157}{7185}\right)} = .359$$

Type I Error Inflation[1]

Cluster size	**ICC**	***Deff***	**Type I Error**	**Cluster size**	**ICC**	***Deff***	**Type I Error**
10	0	1.00	.05	10	.20	2.80	.28
25	0	1.00	.05	25	.20	5.80	.46
100	0	1.00	.05	100	.20	20.80	.70
10	.05	1.45	.11	10	.40	5.50	.46
25	.05	2.20	.19	25	.40	13.00	.63
100	.05	5.95	.43	100	.40	50.50	.81

For the HSB data, *Deff* = ??

- Lai & Kwok (2015):[2] MLM needed when *Deff* > 1.1

[1]: Table adapted from Barcikowski (1983) [2]: https://doi.org/10.1080/00220973.2014.907229

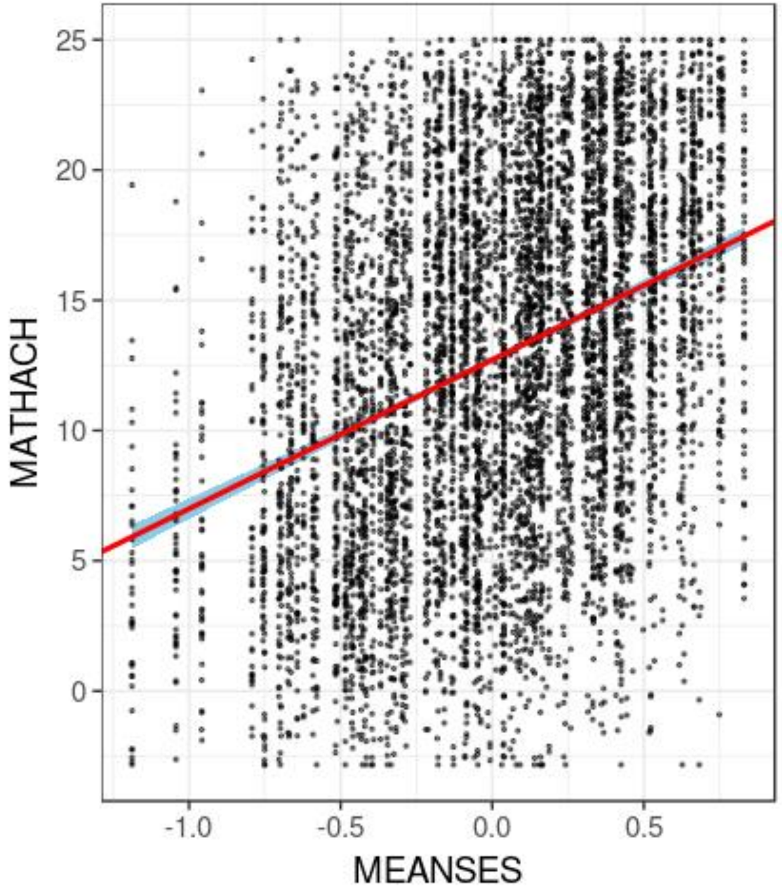
Exercise

- *Deff* = 1 + (average cluster size – 1) × ICC
- Average cluster size = 7,185 / 160 ≈ 44.91
- ICC = 0.18

- Bonus Challenge: What is the design effect for a longitudinal study of 5 waves with 30 individuals, and the ICC for the outcome is 0.5?

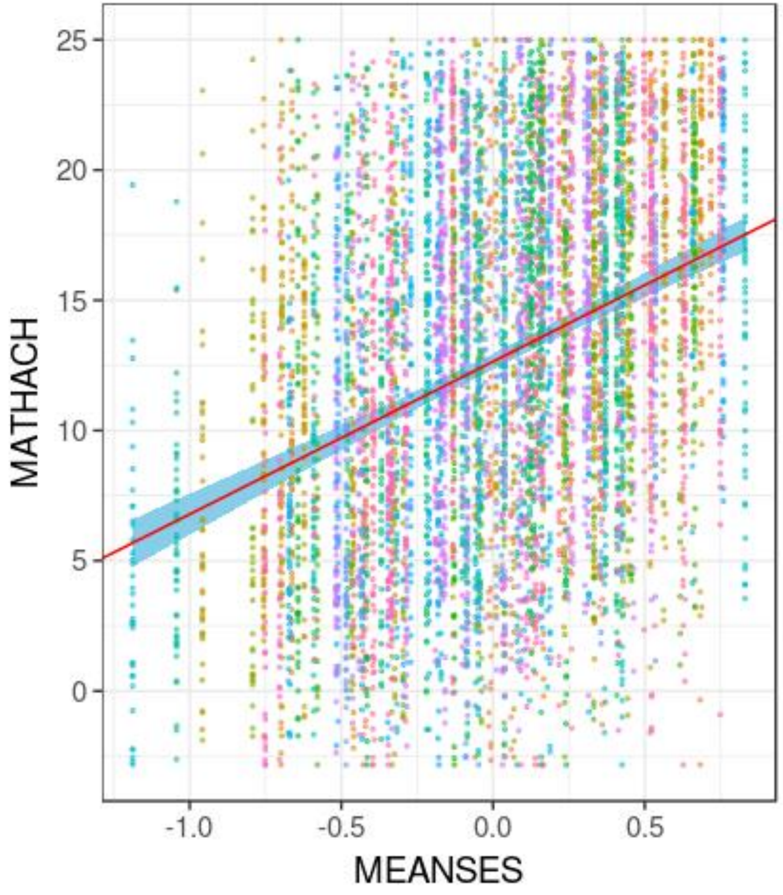
Overconfidence (Disaggregation)

OLS



95 % CI of slope = [5.36, 6.08]

MLM



95 % CI of slope = **[5.16, 6.57]**

Problem of Aggregation

- Student-level information is ignored
- OLS on <u>160 schools</u>

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.6219     0.1533   82.35   <2e-16 ***
MEANSES        5.9093     0.3714   15.91   <2e-16 ***
```

- MLM

```
Fixed effects:
             Estimate Std. Error t value
(Intercept)   12.6494     0.1493   84.74
MEANSES        5.8635     0.3615   16.22
```

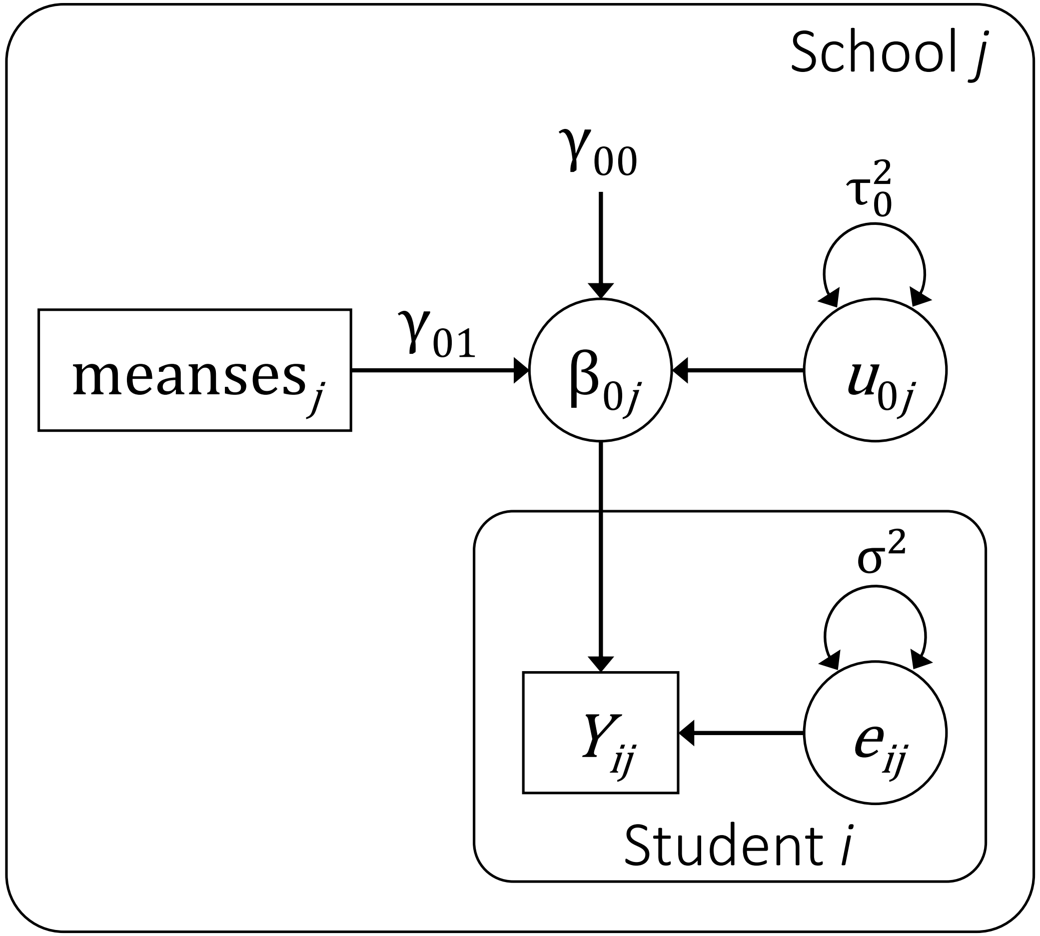
SE is slightly overestimated

Model Equations

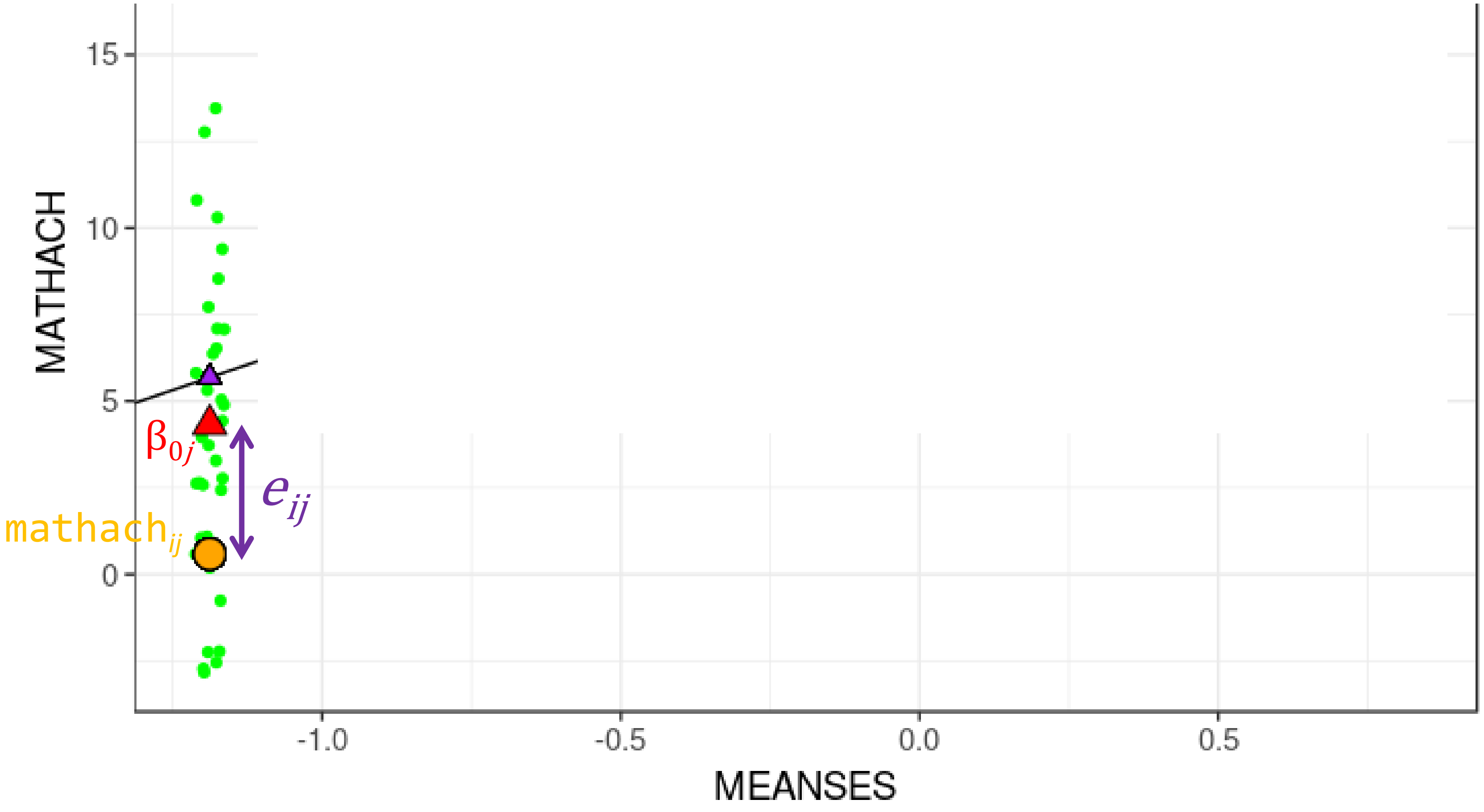
- Lv 1: $\text{mathach}_{ij} = \beta_{0j} + e_{ij}$
- Lv 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\ \text{meanses}_j + u_{0j}$
- Combined: $\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\ \text{meanses}_j + u_{0j} + e_{ij}$

Model Equations

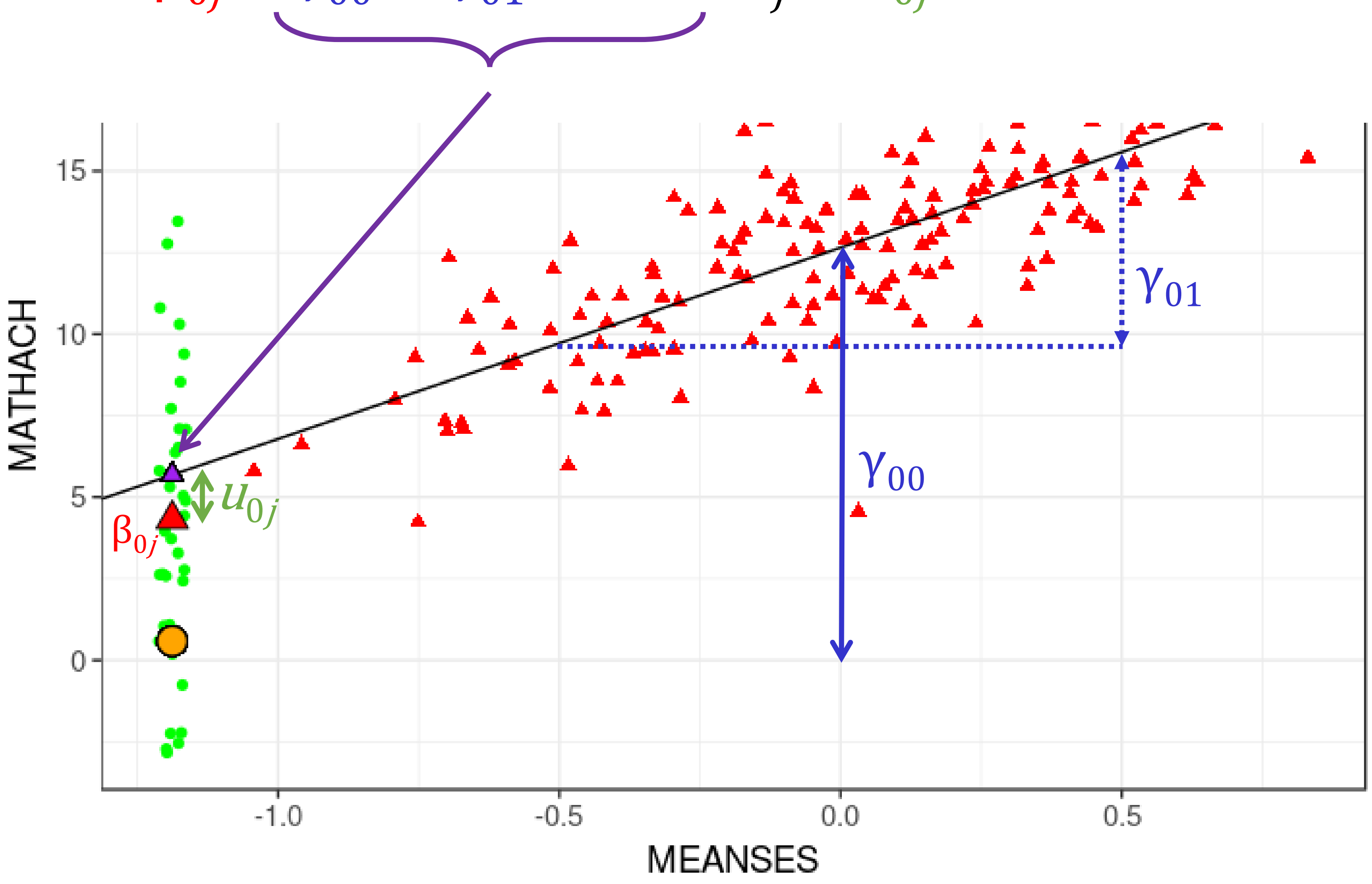
- Lv 1: $\text{mathach}_{ij} = \beta_{0j} + e_{ij}$
  $e_{ij} \sim N(0, \sigma)$
- Lv 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{meanses}_j + u_{0j}$
  $u_{0j} \sim N(0, \tau_0)$
- Combined:
  $\text{mathach}_{ij} = \gamma_{00} + \gamma_{01}\,\text{meanses}_j + u_{0j} + e_{ij}$



Lv 1: $\text{mathach}_{ij} = \beta_{0j} + e_{ij}$



Lv 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}\ \text{meanses}_j + u_{0j}$



Run the Model in R

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6494     0.1493   84.74
meanses       5.8635     0.3615   16.22
```

The estimated school mean of mathach when meanses = 0 is γ_{00} = **12.65 (*SE* = 0.15)**

The model predicts that students from two schools with 1 unit difference in meanses will have an average difference of γ_{01} = **5.86 (*SE* = 0.36)** units in mathach

Run the Model in R

```
Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  2.639   1.624
 Residual             39.157   6.258
Number of obs: 7185, groups:  id, 160
```

Variance of deviations of school means from the regression line = Var(u_{0j}) = 2.64

Variance of individual scores within a school = Var(e_{ij}) = 39.16

Statistical Inference

- It's important to understand that the coefficients you obtained in software are merely *estimates*, which involves *uncertainty*
- Confidence intervals
  - Wald intervals
  - Likelihood-based intervals
- Hypothesis testing (to be discussed later)

Confidence Interval (Wald)

- 95% CI for γ_{01} = 5.86 ± 2 × 0.36 = [5.16, 6.57]
  - Can be obtained in most software

At 95% confidence level, one unit difference in school-level MEANSES is associated with an average difference in MATHACH of **5.16** to **6.57** units

Confidence Interval (Likelihood-Based)

```
> confint(m_lv2, parm = "beta_")
Computing profile confidence intervals ...
                2.5 %    97.5 %
(Intercept) 12.356615 12.941707
meanses      5.155769  6.572415
```

- Easily obtained in the R package `lme4`
- Usually more accurate than Wald intervals, especially with smaller sample sizes
- With a large sample size, the difference is minimal