Beginning linear mixed effects models

Data: flipped classrooms?

- A *flipped classroom* involves students watching lectures at home, and doing activities during class time
- There is debate about the pros and cons of this teaching method
- Here we will look at simulated data from an experiment with flipped classrooms

Data: flipped classrooms?

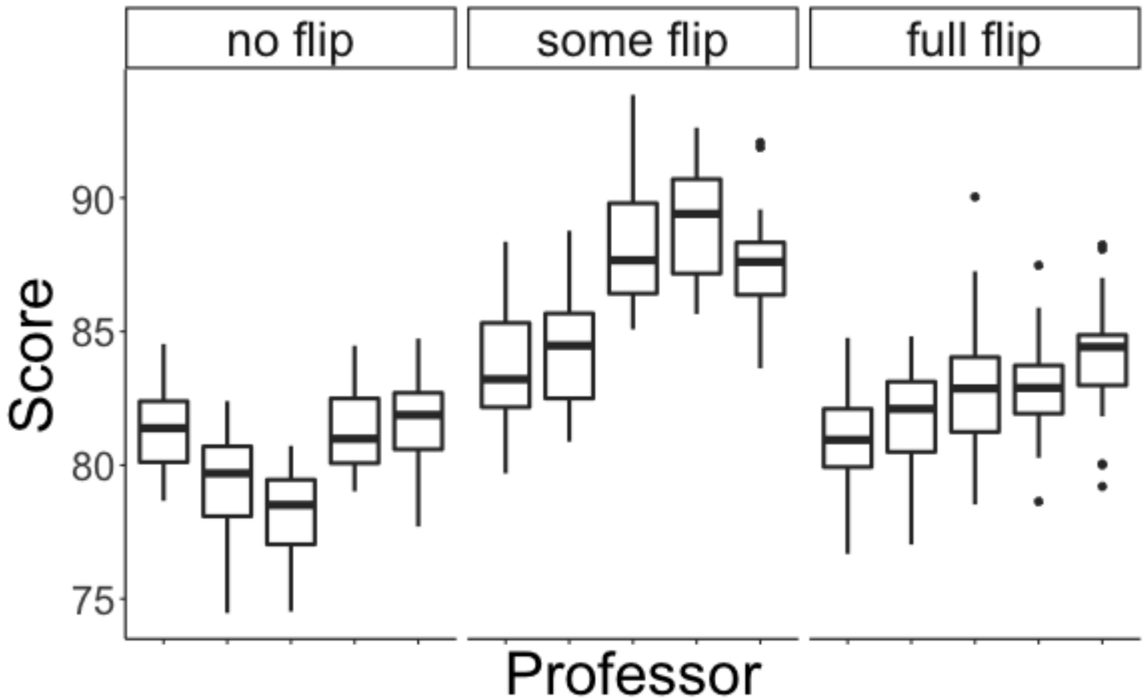
- 15 classes of introductory statistics
- 25 students in each class (so 375 students total)
- Each class taught by a different professor
- Each professor randomly assigned a teaching style: No flip, Some flip, and Fully flipped
- At the end of the semester, we give all the students in all the classes the same exam, and compare their results

Data: flipped classrooms?

Data set has 375 rows (one per student), and the following variables:

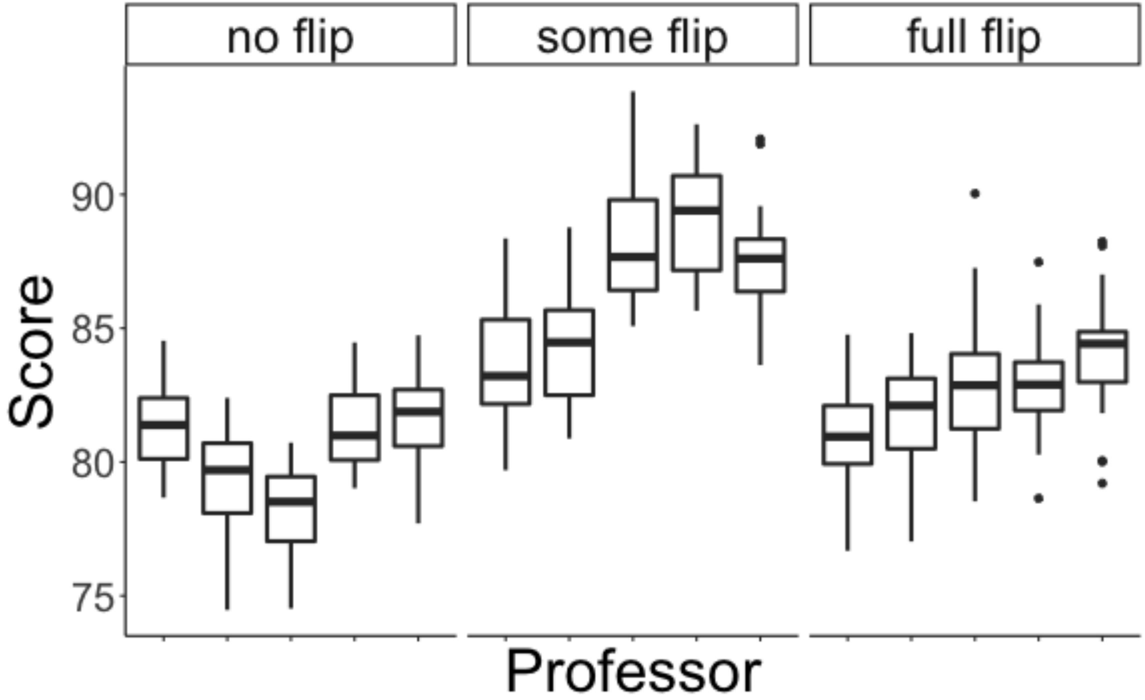
- `professor`: which professor taught the class (1 -- 15)
- `style`: which teaching style the professor used (no flip, some flip, fully flipped)
- `score`: the student's score on the final exam

Considering results



What do you notice about the scores?

Considering results



- There may be some differences between styles
- There may be some differences between professors

Considering results

> Suppose we notice that, on average, students in the "Some Flipped" classes have higher scores than students in the "Fully Flipped" classes. What might explain this difference?

Considering results

> Suppose we notice that, on average, students in the “Some Flipped” classes have higher scores than students in the “Fully Flipped” classes. What might explain this difference?

- The "Some Flipped" method may lead to higher test results.
- The professors assigned to teach "Some Flipped" may teach in such a way that their scores are higher than those in the "Fully Flipped" group (more experience, etc.).
- The students in the "Some Flipped" classes may have been stronger than those in the "Fully Flipped" group.

Different effects

(e.g., average difference in score between "some flipped" and "fully flipped" classes)

- *Effect of interest* (*treatment effect*): The "Some Flipped" method may lead to higher test results; *the treatment imposed by the researchers has an effect on the outcome.*

Different effects

- *Effect of interest* (*treatment effect*): The "Some Flipped" method may lead to higher test results; *the treatment imposed by the researchers has an effect on the outcome.*

- *Group effect*: The professors assigned to teach "Some Flipped" may have had an impact on the test scores; *the group the students are in has an effect on the outcome.*

Different effects

- *Effect of interest* (*treatment effect*): The "Some Flipped" method may lead to higher test results; *the treatment imposed by the researchers has an effect on the outcome.*
- *Group effect*: The professors assigned to teach "Some Flipped" may have had an impact on the test scores; *the group the students are in has an effect on the outcome.*
- *Individual effect*: The students in the "Some Flipped" classes may have been stronger than those in the "Fully Flipped" group; *the individuals' characteristics or abilities have an effect on the outcome.*

Writing down a model

Score is a continuous response, so we can go back to linear models:

$$Score_i = \beta_0 + \underbrace{\beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i}_{\text{Treatment effect}} + \varepsilon_i$$

individual effect

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> Which effects does this model capture?

Writing down a model

Score is a continuous response, so we can go back to linear models:

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> Which effects does this model capture?

- Treatment effect (β_0 is the average score in the no flip group, and β_1 and β_2 tell us how the score changes in the other groups)
- Individual effect (ε_i is the difference from the mean for student i)

Assumptions

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$$

> What does this model assume about group effects (differences between professors)?

Assumptions

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$

> What does this model assume about group effects (differences between professors)?

That there are no systematic differences between professors (i.e., no group effects)

Assumptions

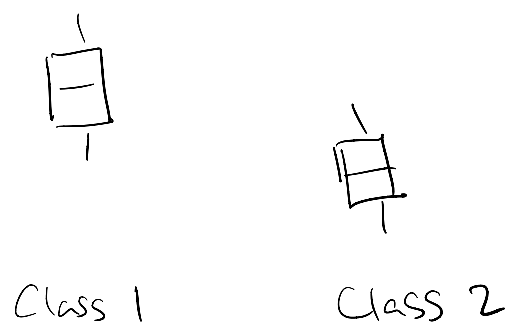
$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

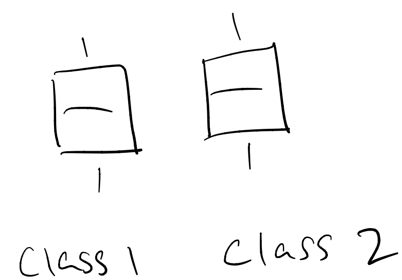
$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

we assume noise is independent
(i.e. students are independent)
We have not included Professor in the model

> What does this model assume about correlation within a class?

If there are differences between classes:





Assumptions

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$$

> What does this model assume about correlation within a class?

That there is no correlation between student scores within the same class

> Is this a good assumption?

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> How can I incorporate systematic differences between classes?

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> How can I incorporate systematic differences between classes?

Add a variable for the different professors:

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> How many parameters did we add to the model to capture class differences?

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$

> How many parameters did we add to the model to capture class differences?

14 (β_3,..., β_{16})

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> Do we want to do inference on β_3,..., β_{16} ?

Writing down a model

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$

> Do we want to do inference on β_3,..., β_{16} ?

No -- we only care about inference for the treatment effect parameters (β_1 and β_2)

> Can we do something *different* to capture group effects?

Our first mixed effects model

Linear model:

student i

$$Score_i = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + \\ \beta_3 \text{Class2}_i + \cdots + \beta_{16} \text{Class15}_i + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$

Linear mixed effects model: Let $Score_{ij}$ be the score of student j in class i

$$Score_{ij} = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + u_i + \varepsilon_{ij}$$

only observed at group level

$\underbrace{\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)}_{\text{variation between individual students}}$ $\quad$ $\underbrace{u_i \overset{iid}{\sim} N(0, \sigma_u^2)}_{\text{variation between classes}}$

(i.e. groups of students

Anatomy of the mixed effects model

Linear mixed effects model: Let $Score_{ij}$ be the score of student j in class i

$$Score_{ij} = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + u_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \qquad u_i \overset{iid}{\sim} N(0, \sigma_u^2)$$

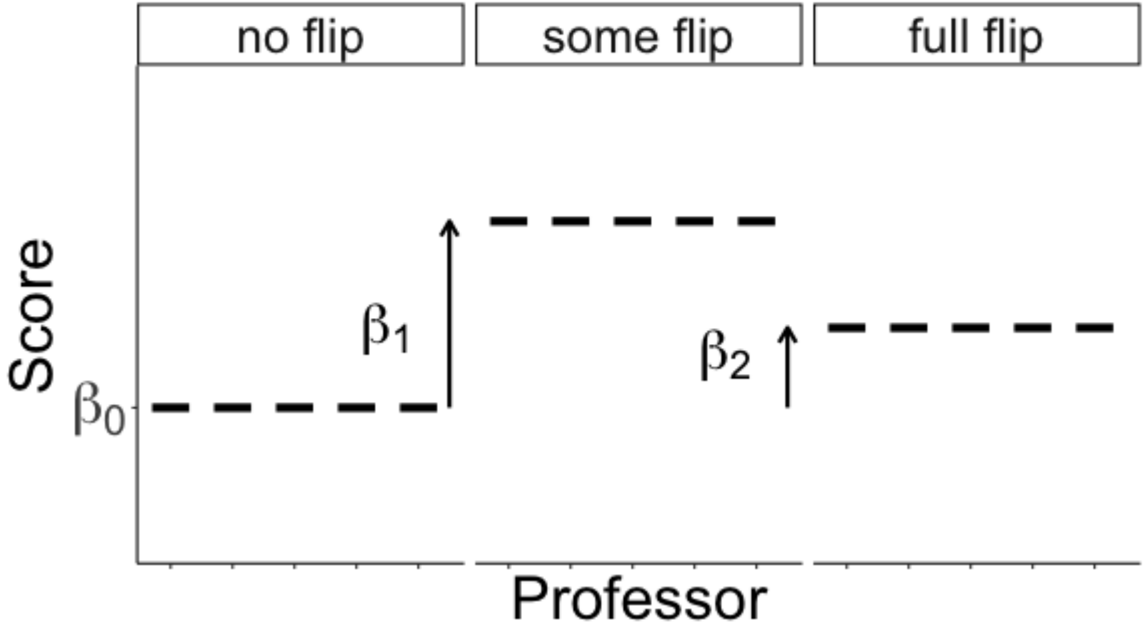
- $\beta_0, \beta_1, \beta_2$: **fixed effect** terms (representing treatment effect)
- u_i : **random effect** terms (representing group effects)
- ε_{ij} : **noise** terms (representing individual effects)

Anatomy of the mixed effects model

$$Score_{ij} = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + u_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon) \qquad u_i \overset{iid}{\sim} N(0, \sigma^2_u)$$

Part 1: Fixed effects (treatment effects)

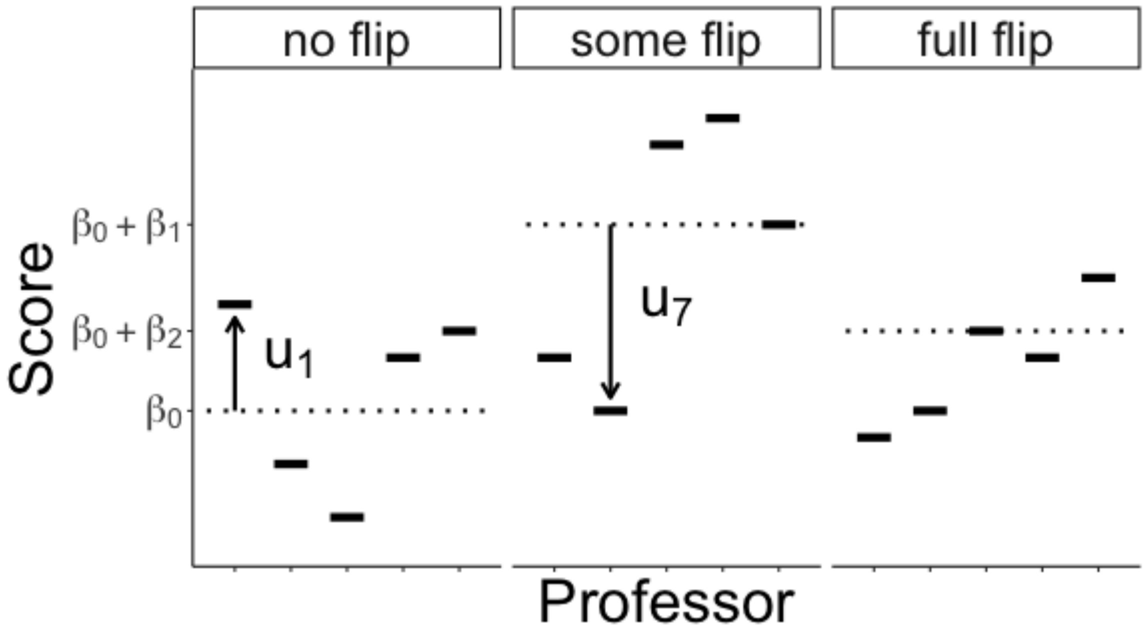


Anatomy of the mixed effects model

$$Score_{ij} = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + u_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon) \qquad u_i \overset{iid}{\sim} N(0, \sigma^2_u)$$

Part 2: Random effects (group effects)



Anatomy of the mixed effects model

$$Score_{ij} = \beta_0 + \beta_1 \text{SomeFlipped}_i + \beta_2 \text{FullyFlipped}_i + u_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon) \qquad u_i \overset{iid}{\sim} N(0, \sigma^2_u)$$

Part 3: Noise (individual effects)