Lab 3: Clustered data, longitudinal effects

Robin Mejia

Email: *rjmejia@berkeley.edu*

Office Hours: **Thurs, 9am-10:45am, 109 Haviland**

September 16, 2015

Clustered Data

Consider a scenario where we have randomly selected a set of 10 classrooms from a school. These classrooms have between 28 and 34 students. Within each classroom, we have test scored for each student at the beginning and end of the school year.

- What are the independent units?

- What notation would you use to describe each measurement?

    - How many levels are there?

    - How many subscripts do you need to describe each observation?

Clustered data, notation

- The independent units are the classrooms. This is the level at which we employed a random selection procedure.

- Why would test scores for students in the same classroom be correlated?

Clustered data, notation

- $i \in \{1, \ldots, 10\}$ is the classroom identifier

- $j \in \{1, \ldots n_i\}$ is the child identifier, $n_i \in \{28, \ldots 34\}$

- $k \in \{1, 2\}$ is test score time point

- Y_{ijk} is the test score for student j, in classroom i, at time k.

The data is given to you as described above. What could you do if you wanted to measure the longitudinal effect of the class in each child?