

## Problem Set 2

Note: The dataset is difficult to parse, there isn't a convenient interpretation on hand for exactly what each value means. I've chosen to only use the value of `wjtest01` for this assignment because I do not know which `wjtest` represents which actual test, and the results from a cursory glance show that looking at one test for the questions should be enough.

(1) What is the research question? What causal effect are the authors attempting to estimate?

This research is attempting to answer the question of what causal effect there may be of prekindergarten programs on student performance. Or, in the paper's words: "What can we expect from prekindergarten programs?"

(2) What feature of the data allow them to estimate this causal effect? How do they go about doing it?

They have two groups of students, who are roughly the same age but due to the yearly school entry cutoff are a school-year apart. They are all given the same test at the beginning of pre-k and at the beginning of kindergarten.

<methodology-rant>

I, however, am going to express a small amount of skepticism at this methodology.

They shouldn't have had the year-ahead group to take the test before pre-k, as well as before kindergarten. Otherwise, how do they account for trial-and-error? A student who has a decent memory, and saw what things they got wrong the first time, may vastly improve as compared to a student who didn't get the opportunity to do that in the first place.

Essentially, how do they isolate the effects of pre-k from the effects of good feedback on completed work?

</methodology-rant>

(3) Produce a plot that shows graphically how the authors estimate this causal effect.

I'll just be looking at the 3-month margin for now.

The below stargazer table includes some variables I would include:

Table 1:

|                         | <i>Dependent variable:</i>  |
|-------------------------|-----------------------------|
|                         | wjtest01                    |
| age                     | 0.007***<br>(0.001)         |
| I(age >= 0)             | 3.618***<br>(0.323)         |
| white                   | 0.369<br>(0.275)            |
| black                   | 0.187<br>(0.278)            |
| hispanic                | -1.685***<br>(0.299)        |
| freelunch               | -2.093***<br>(0.176)        |
| female                  | 0.806***<br>(0.161)         |
| age:I(age >= 0)         | -0.0003<br>(0.002)          |
| Constant                | 6.717***<br>(0.336)         |
| Observations            | 2,756                       |
| R <sup>2</sup>          | 0.397                       |
| Adjusted R <sup>2</sup> | 0.395                       |
| Residual Std. Error     | 4.230 (df = 2747)           |
| F Statistic             | 226.235*** (df = 8; 2747)   |
| <i>Note:</i>            | *p<0.1; **p<0.05; ***p<0.01 |

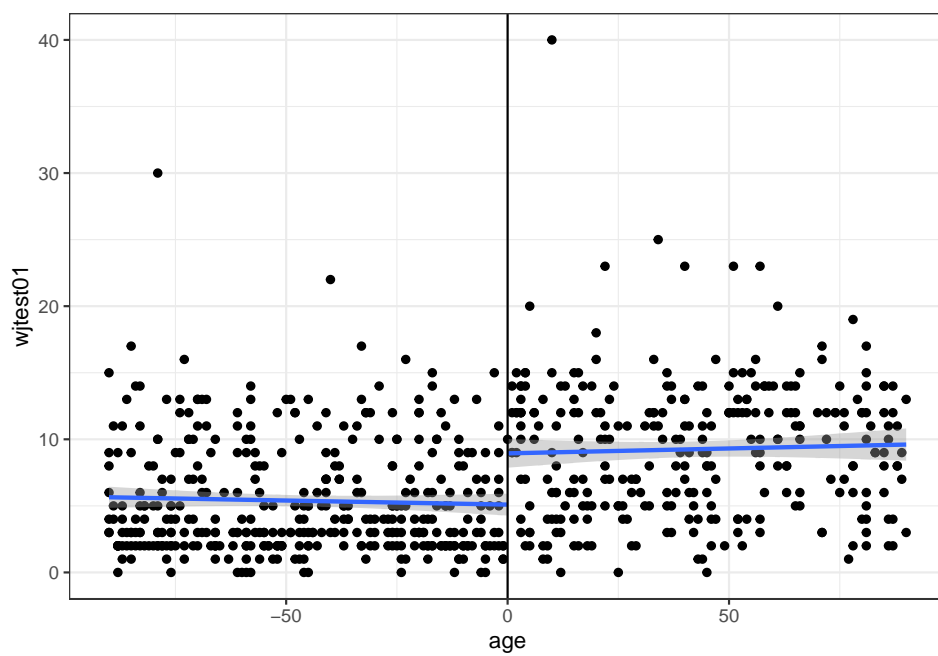


Figure 1: ...