Following on from our simulation of omitted variable bias last week, we will now simulate experimental conditions (you can adapt your script from last week or start again):

1. Generate data on a population of 1,000 people. Specifically, create a variable (a vector) that randomly assigns these people to be male or female (50:50). Hint: In R, try rbinom and in Stata, try rbinomial.

2. Now we are going to simulate the potential outcomes - a measure of attitudes - under control (y0) for our population. Create another variable of random normally-distributed values with mean of 5 and standard deviation of 1.

3. One problem with observational data is that potential outcomes are often correlated with other variables such as gender. Adjust your value of y0 to add 1 (one) for all units who are male.

4. Now simulate potential outcomes under treatment (y1) for all units. Define a constant treatment effect of c = 2 and create another vector y1 = y0 + c.

5. Next, let’s assume a specific **Treatment Assignment Mechanism – random assignment.** Create a treatment variable D that gives each unit a 50\% chance of binary treatment (like a coin flip).

6. Are gender and treatment correlated? Calculate the correlation between x and D.

7. In practice, we only observe one value: yobs. Create a new variable yobs which equals y1 if D = 1 but which equals y0 if D = 0.

8. Compare the balance of observed covariates (x, gender) in the treatment and control groups using a difference-in-means test.

9. Compare the balance of potential outcomes (y\_0, y\_1) in the treatment and control groups using a difference-in-means test. (Remember we can’t do this in reality!)

10. Based on the observable data, analyze the results of the experiment using a difference-in-means test of outcomes (y) by treatment status (D).

11. As an alternative, run the basic regression of observable outcomes (yobs) on treatment (D). Interpret the result. Is this an accurate estimate of the treatment effect that we created at the start?

12. Now add the control variable $x$ to the regression. How do the results change?

13. Repeat the above experiment 100 times with the same x, y\_0 and y\_1 every time, but re-randomize treatment assignment. Collect the 100 resulting p-values. How many of them are less than 0.05?

14. Now let’s try to understand how things change when treatment is *clustered*.

a. In one of your datasets, add an additional column which transforms (‘bins’) the $y\_0$ variable into a categorical variable with 20 categories, where category ‘1’ represents the lowest values of $y\_0$ and category ’20’ contains the highest values of $y\_0$. This reflects the fact that people in the same cluster have more similar characteristics.

b. Add a new binary treatment variable, D\_clustered, that randomly assigns each cluster to treatment or control.

c. Now run the regression of observable outcomes (yobs) on clustered treatment (D\_clustered).

d. Now run the regression of observable outcomes (yobs) on clustered treatment (D\_clustered), and cluster standard errors at the cluster level.

How would you analyze the following field experiment? List any qualitative tests, quantitative checks, and the statistical or regression analyses that you would apply to the data. Assume the authors find a positive 0.2 standard deviation average treatment effect – how would you interpret this effect? How much do we learn from this experiment?

1. Classroom-level treatment, indi-level results, blocked on school performance,

How would you design and implement a field experiment for the following research question?