**Assignment 3: Randomization inference with regression**

**Due date: Thursday May 21, 2020 by 12:00PM**

Randomization inference calculates exact p-values under Fisher’s sharp null of no treatment effect. The method asserts a sharp null, calculates a test statistic for the true effect, and then for every combination of treatment assignments possible, calculates placebo test statistics. The distribution of these test statistics is then used to calculate a two tailed p-value based on the position of the true effect in the left or right tail. In this exercise, you will extend Thorton’s 2008 article Demand for Learning HIV status to include a regression model, as opposed to the simple difference in means estimator that I did in the thornton\_ri.do and thornton\_ri.R programs.

1. As before, create a set of directories in your github repository corresponding to this assignment. Feel free to create subdirectories within your broader directory. Send me the link to that github respoitory.
2. You will be using the thornton\_hiv.dta file which is at the following link:

**https://github.com/scunning1975/mixtape/raw/master/thornton\_hiv.dta**

Repeat the randomization inference in the thornton\_ri.do or thornton\_ri.R only instead of calculating the simple difference in outcomes, run the following regression:

Your test statistic will be the coefficient on *ANY.* You will need to save the coefficient each iteration, store it similar to how I did, append all the iterations and calculate the p-value using simulation since all possible combinations are impossible to calculate.

1. Do the simulation three times: with 100 simulations, 1000 simulations and 10,000 simulations. Report p-values under each.
2. Create a histogram for the 10,000 simulations showing the distribution of all placebos as well as the true effect.
3. What is the approximate probability that the true effect came from the placebo distribution?