ECON 379: Program Evaluation for International Development

Empirical Exercise #2

Due 3/2 by 11:30 AM

In this exercise, we'll use Stata's `rnormal()` command to generate draws from a normally-distributed random variable. This approach – simulating data according to a known data-generating process – is an incredibly useful tool in empirical microeconomics (both for checking your econometric intuitions and your anlayis code).

We'll use "locals" (also know as "local macros") to easily change the number of observations and other parameters of our data set. This will allow us to explore the way the properties of randomly-assigned treatment groups in larger and smaller samples.

This exercise introduces a range of practical coding tools: `rnormal()`, locals, and the `return list` and `display` commands. By varying the sample size, we'll build a better understanding of the role that the Law of large Numbers plays in randomized evaluations.

Create a do file containing the code below, and then use/modify it to answer the questions on the next page:

```
// PRELIMINARIES

** start with a clean workspace
clear all
set more off // setting more off prevents your code from stopping halfway through
set seed 12345 // setting the seed makes pseudo-random draws replicable
set scheme s1mono // the scheme is only relevant when making graphs/figures
version 16.1 // make sure you use a specific version of Stata (for replicability)

** change working directory as appropriate to where you want to save
cd "C:\Users\pj\Dropbox\econ379-2021\exercises\E2-selection-bias"

** save your do file to a local directory no (do this by hand, not in code)


// GENERATE A DATA SET

** define a local that we'll use to indicate the number of observations
local myobs = 10

** use the localto create an empty data set with N=myobs rows
set obs `myobs'
```

```
** define some variables
gen y = rnormal()
gen z = 5*rnormal() + 10

** assign half the variables (observations 1 through N/2) to treatment
count
return list // this shows you all the local macros saved by your last command
local cutoff = (r(N)/1)/2
gen treatment = 1 in 1/‘cutoff’
replace treatment = 0 if treatment==.
```

Use the code above to help you answer the following questions:

1. How many observations are in the data set?
2. The variable `act_any` is a dummy for assignment to any treatment (positive subsidy) What is the mean of the variable `act_any` (to three decimal places)?
3. How many people received a positive subsidy?
4. The variable `c_act` is a dummy for using ACT treatment during a malaria episode. What is the standard deviation of the variable `c_act`?
5. How many respondents report using ACT treatment for malaria?
6. Regress `c_act` on `act_any`. What is the R-squared?
7. What is the coefficient associated with the `act_any` variable?
8. What is the associated standard error?
9. What do you get when you divide the coefficient by the standard error?
10. What is the t-statistic associated with the `act_any` variable?