ECON 523: Program Evaluation for International Development
In-Class Activity 9

In this exercise, we'll be learning how to randomly assign treatment status in a way that is transparent and reproducible. After assigning treatments, we'll check whether we've succeeded in creating a treatment group and a control group that are comparable in terms of their observable characteristics.

Start by creating a new do file that runs the following Stata code:

```
clear
set obs 4
gen id = _n
gen rand_num = rnormal()
sort rand_num
egen treatment = seq(), from(0) to(1)
sort id
```

What happens when you run the code? Use Stata's data editor to view the (very small) data set you created. Which ID numbers are assigned to treatment? Run the code several times. Are the same ID numbers assigned to treatment each time?

The code above contains the three key parts of every randomization do file:

(1) A command that generates a pseudo-random number

(2) A command that sorts the data based on that random number

(3) A command that assigns treatment based on that random sort order

The idea behind random assignment is that we can generate a variable using Stata's pseudo-random number generator and then sort the data set based on that variable; when we do this, the observations in the data set are listed in a random order. If we want to randomly assign observations to treatment and comparison groups, we can assign every other observation to treatment – after we've sorted them based on our random x variable.

The command

```
egen treatment = seq(), from(0) to(1)
```

generates a repeating sequence from 0 to 1: the first row (observation) in the data set will get a 0, the second row will get a 1, the third row will get a 0, and so on. Familiarize yourself with this command. How might you assign observations in your data set to four different treatment groups?

In the example above, we failed to set the seed, so each time we run our code, we get a completely new random treatment assignment. Insert the command

```
set seed 1234
```

between `clear` and `set obs 4`. This will guarantee that Stata uses the same sequence of pseudo-random numbers every time you run the file (you can also set the version for additional confidence in your code's reproducibility). Run the file a few times to confirm that this is the case.