ECON 523: Program Evaluation for International Development
Empirical Exercise 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.

We're going to use the same data set on potential microfinance clients in urban India that we worked with in Empirical Exercise 6. The data set comes from the paper "The Miracle of Microfinance? Evidence from a Randomized Evaluation" by Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. The authors worked with an Indian MFI (microfinance institution) called Spandana that was expanding into the city of Hyderabad. Spandana identified 104 neighborhoods where it would be willing to open branches. They couldn't open branches in all the neighborhoods simultaneously, so they worked with the researchers to assign half of them to a treatment group where branches would be opened immediately. Spandana held off on opening branches in the control neighborhoods until after the study.

The data set contains information on 6,853 households. Suppose you want to work with a local NGO to offer business training and mentoring to microentrepreneurs, and you want to stratify treatment assignments by treatment status in the Spandana RCT to see whether impacts depend on the availability of microcredit.

Create a do file that reads the Spandana data from Empirical Exercise 6 into Stata. Then extend your do file so that it randomizes treatment assignments, as described below.

1. You want to stratify treatment assignments in your evaluation by four variables:

   ○ Treatment status in the original Banerjee et al. (2015) Spandana RCT
   ○ A new variable that you create indicating whether a household has taken a formal loan from either an MFI or a bank (by the time of Endline 1 in the original Spandana study)
   ○ An indicator for operating a household business (by Endline 1)
   ○ An indicator for ever having been late with a loan repayment (by Endline 1)

   Construct stratification cells based on these four variables.

2. Now randomly assign the households in the sample to treatment and control, stratifying by the four variables described above.

3. Once you have randomly assigned treatment status, we will typically want to check whether our treatment and comparison groups look similar in terms of observable characteristics. Make a **balance check table** that reports, for each of a set of covariates,

   ○ The mean and standard deviation of the covariate in the entire sample

- ◦ The mean and standard deviation of the covariate in the treatment group
  - ◦ The mean and standard deviation of the covariate in the control group
  - ◦ The p-value from a t-test of the hypothesis that the mean does not differ between the treatment and comparison groups

  To do this, you can adapt the Stata program that you wrote for Empirical Exercise 8. Report tests for balance for each of your stratification variables plus the variables capturing whether a household operates a business (as of Endline 1), the number of household businesses, business assets, business revenues, business expenses, business profits. Save a copy of your finished balance check table as a pdf so that you can upload it to gradescope.

4. If you tested 1,000 baseline covariates for balance, how many of those variables would you expect to be imbalanced enough that you could reject the hypothesis that the mean in the treatment group was equal to be mean in the control group at the 95 percent confidence level?

Extensions

1. How would you modify your code to stratify on baseline business revenues?

2. How would you modify your code to assign households to either a control group or one of three different treatment arms?

3. How would you modify your code to assign treatment at the neighborhood level, and then check for balance on household-level covariates?