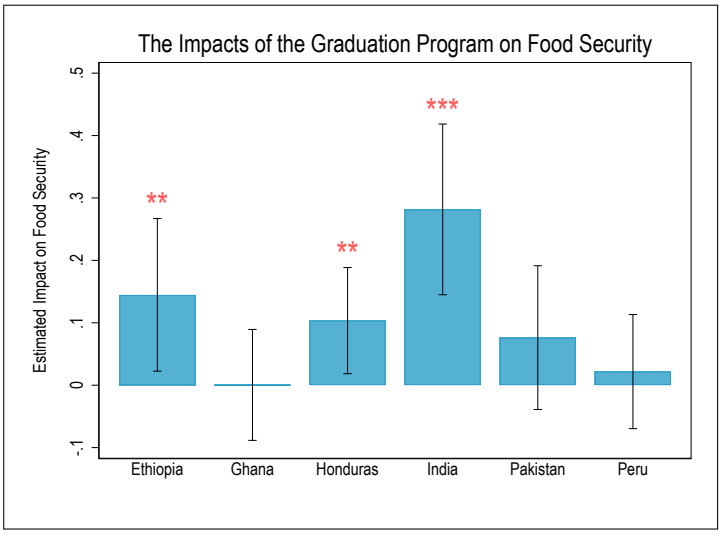
ECON 204: Graduation Programs Lab

This exercise makes use of the data set `ECON-204-graduation-data.dta`, which contains data from the paper "A multifaceted program causes lasting progress for the very poor: Evidence from six countries" by Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, William Pariente, Jeremy Shapiro, Bram Thuysbaert, and Chris Udry, published in *Science* in 2015.

The authors examine the impacts of a "graduation" program first designed by the Bangladeshi NGO BRAC. The program offers extremely poor households an asset transfer, temporary consumption support, skills training, home visits, and access to savings technologies. The program was evaluated through a randomized trial in six countries.

The figure below summarizes the impact of the graduation program on food security:



You will generate a figure similar to this one that summarizes the impact of the graduation program on household assets. To do this, you will estimate the treatment effect of the program separately for each country. Open the lab template in Stata's do file editor and then run the do file; this will load the required data set directly from github. Familiarize yourself with the data set. It contains the following variables:

- `treatment` indicates whether the household was in the treatment (1) or comparison (0) group
- `country` is the country
- `cid` is a numeric ID number for the country
- `assets` is an index of household assets, measured after the graduation program ended
- `food_security` is an index of food security, also measured after the program

Now extend the template do file as you answer the questions below.

1. **Generating the X variable.**

   (a) Use the command `generate graphid = _n` to generate a variable equal to the row number for each observation in the data set.

   (b) Use the `replace` command to replace all values of `graphid` greater than 6 with missing values (`graphid == .`) Summarize or tabulate the `graphid` to make sure that it goes from 1 to 6 (for the 6 countries in the data set).

2. **Generating the `coefficient` and `se` variables.**

   (a) Use the `generate` command to generate new, empty variables called `coefficient` and `se`. You will use these to store the results of your regressions (below).

3. **Estimating the impact of the graduation program in Ethiopia.**

   In an earlier lab, we saw that we could estimate the impact of a program that has been randomized by comparing the means in the treatment and control groups using Stata's `ttest` command. We can also do this using the `regress` command: as we've discussed, when we regress an outcome variable y on a dummy variable t, the regression coefficient on t tells us the difference in means between the $t = 1$ group and the $t = 0$ group (convince yourself of this by running the `ttest` and `regress` commands).

   Use the code below to (1) regress `assets` on `treatment` using only the data from Ethiopia (which is Country 1 in the data set, so observations from Ethiopia all have `cid` equal to 1) and (2) store your regression results in the variables `coefficient` and `se` that we defined above.

```
reg assets treatment if cid == 1
replace coefficient = _b[treatment] if graphid == 1
replace se = _se[treatment] if graphid == 1
```

   Notice that the results of our regression (specifically, the coefficient and standard error from) were stored as `_b[treatment]` and `_se[treatment]`.

4. **Estimating the impact of the program in the other countries.**

   (a) Extend and modify the code above so that you also store the estimated impact of the graduation program in Ghana (`cid==2`), Honduras (`cid==3`), India (`cid==4`), Pakistan (`cid==5`), and Peru (`cid==6`).

   (b) Use the command `list graphid coefficient se` to confirm that you have successfully recorded an estimate of the impact of the graduation program for each country in the data set.

5. **Generating the upper and lower bounds of the confidence interval.**

   (a) Generate a new variable, `lower`, that is the **lower** bound of the 95 percent confidence interval (i.e. the estimated coefficient **minus** 1.96 times the estimated standard error.)

(b) Generate a new variable, `upper`, that is the **upper** bound of the 95 percent confidence interval (i.e. the estimated coefficient **plus** 1.96 times the estimated standard error.)

6. **Presenting Treatment Effects Graphically.**

   Make a bar graph that summarizes the impact of graduation program on household assets in each country using the `twoway` command. When you finish your graph, export or save it as a `.png` or `.pdf` file. Your graph should have correctly labeled and scaled axes, a title, bars representing the estimated treatment effects, and lines indicating the confidence intervals. You should use color effectively. If you can, figure out how to display the stars indicating the level of statistical significance.

   When you submit your assignment on gradescope. You will be asked to upload both the pdf or png file containing an image of your graph and the code that you used to generate it.