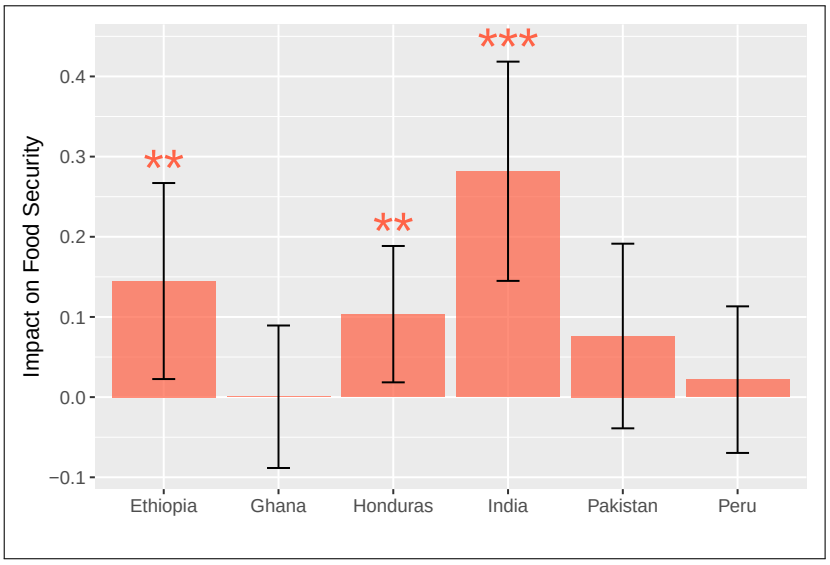
ECON 204: Graduation Programs Lab (R Version)

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 the R script editor and run the code; 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 script as you answer the questions below.

1. **Generating the X variable.**

   (a) Generate a data frame, `results`, containing six rows that includes the observed values of the country IDs and the country names.

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

   (a) Add empty variables called `coefficient` and `se` to the `results` data frame.

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 a t-test. We can also do this using a regression: 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.

   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.

```
ols <- lm(assets ~ treatment, data = mydata, subset = (cid == 1))
estimates <- summary(ols)
results$beta[1] <- estimates$coefficients[2,1]
results$se[1] <- estimates$coefficients[2,2]
```

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) 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 `ggplot`. 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.