The People Lab

**Thank you for applying to work at The People Lab!**

This task is designed to get a sense of your level of proficiency with data manipulation and analysis. Please clean and analyze the attached data using Stata or R.

## Task instructions

Please save your code (do file or R script) entitled “[Your Last Name]_DataTask_code” and one-page executive summary of your findings entitled “[Your Last Name]_DataTask_summary”.

In the accompanying excel documents you will find data from a field experiment that aimed to test what type of language is most effective at encouraging college students to apply for CalFresh, California’s Supplemental Nutrition Assistance Program. Specifically, this study sought to answer the following research questions:

1. Does sending an informational message to students who are likely eligible for CalFresh increase their likelihood of applying relative to no message?
2. Does sending a social norms message to students who are likely eligible for CalFresh increase their likelihood of applying relative to no message?
3. Does sending a social norms message to students who are likely eligible for CalFresh increase their likelihood of applying relative to an informational message?

**Study Description:**

The study sample is composed of 62,773 college students in two counties in California who were identified as being likely eligible for CalFresh. In stratified randomization, all college students were randomly assigned to one of the following three conditions with equal probability:

1. *Control:* Students in the control group did not receive any communication as part of this study.
2. *Information:* Students in the information group received a simple informational email with instructions for enrolling in CalFresh.
3. *Social norms:* Students in the social norms group received an email with instructions for enrolling in CalFresh, as well as information on how many other students in California have already enrolled.

The randomization was stratified by county.

**About the Data:**

The experiment was conducted in all universities in two different California counties (variable *county*). After we collected the data related to the experiment (Experiment data.csv), we asked both counties to send demographic data of their students to use as covariates; they were sent in different files (County 1 data.csv and County 2 data.csv).
Please be sure to:

- Recode all variables such that they are consistent across both counties.
- Merge the Experiment data with the County data, keeping in mind that the student ID is unique only within each county. That means the unique identifier for each participant is *county* and *student ID.*

Some important variables:

- Control_Condition: 1 = assignment to the control condition
- Info_Condition: 1 = assignment to the information condition
- Norms_Condition: 1 = assignment to the social norms condition
- Applied: 1 = Submitted an application for CalFresh; Other Numbers = Did not submit an application for CalFresh

**Analysis:**
Please analyze the accompanying data to answer the research questions above. Show results (1) using a simple comparison of conditions (without covariates), and (2) controlling for relevant covariates.

**Executive summary:**
After completing your analysis, write a one-page (no more!) summary of your findings to share with your manager and research collaborators.

**Questions:**
If you have any questions, please do not hesitate to email jessica_lasky-fink@hks.harvard.edu.