Problem Set 4

BA 830 - Spring 2021

Due Date: **Tuesday** 4/20 at 6pm ET

[The Effects of Phone Calls on Voter Turnout](#_4aqd83jmex6t)

[How long did this problem set take you in hours? How did you find the level of difficulty?](#_riv14xet91ud)

### The Effects of Phone Calls on Voter Turnout

In this experiment, the researchers wanted to investigate the effects of canvassing calls on voter turnout. You can read more about it [here](https://drive.google.com/open?id=0B_Qj0otlErJqT1NMbFhXejNRZ28) but you don’t have to. This experiment is very similar to the example given in lecture and was conducted by the same researchers. As was the case with the example in lecture, they intentionally included a control group and a placebo group to show the power of a placebo.

This experiment took place in Michigan ahead of the August 2008 primary election. Voters were assigned to one of three conditions:

* Control: no phone call
* Placebo: phone call about recycling
* Treatment: phone call about voting

The outcome variable here is whether or not someone turned out to vote. This is measured for everyone using official government records that are publicly available after election day.

|  |  |
| --- | --- |
| Variables: | |
| voted\_aug2008 | 1 = voted, 0 = did not vote in primary election in August 2008 → outcome variable |
| contacted | Successful contact: 1=answered the phone, 0=did not answer the phone → measures “compliance” |
| treatment\_attempt\_turnout\_call | Indicator for randomized to treatment call encouraging people to vote:   * 1=attempted to call the person with a call encouraging them to vote * 0=(in \_treat.csv, randomized to no call, and in \_placebo.csv, randomized to placebo call)   → randomized treatment variable |
| voted\_nov2002 | 1 = voted in 2002 general election, 0 = did not vote → a covariate |

As mentioned, the researchers employed both a placebo and real treatment. To make things easier for you to analyze, I have separated the data from these interventions into two separate files:

* 'noncompliance\_treat.csv' contains data with control and treatment group observations and no placebo observations. This is the data you would have in an experiment without a placebo group.
* 'noncompliance\_placebo.csv' contains data with placebo and treatment group observations and no control group observations. This is the data you would have in an experiment with a placebo group.

**Method 1 for CACE: ITT and Alpha Comparing Treatment vs Control:**

1. First, let’s start with the data that compares the control and treatment (calls encouraging people to vote). Read the data called 'noncompliance\_treat.csv' using the function fread.   
   Suggestion: To keep this dataset straight with the other one we’ll load in later, I would give it a name like data\_treatvscontrol. (4 points)
2. What is the sample size of this dataset? (6 points)  
   Hint: to figure out how many rows there are in a dataset, you can use the nrow(data\_treatvscontrol) or the dim(data\_treatvscontrol) command.
3. For now, let’s not worry about compliance.
   1. Calculate the intent-to-treat effect (ITT) of using regression like we have in previous assignments. That is, what is the average treatment effect (ATE) of being assigned to the treatment group on voter turnout? You should use a regression function to do this. In this problem set, we'll be using the function 'feols' from the fixest package. This is a more powerful version of the 'lm' function with a very similar syntax. (10 points)  
      feols(outcome ~ treatment, data = data, se = ‘white’).   
      *Note: the se argument allows us to calculate standard errors using different assumptions. We will discuss in class late*r.
   2. Use R to extract the coefficient and std.error on the treatment and create separate variables for each one. The best way to do this is using the ‘tidy’ function in the ‘broom’ package, which places the regression results into a data.frame. (4 points)
4. Now we’ll consider compliance. We’ll first calculate the CACE manually and then use a canned function in R to get (hopefully) the same result.
   1. Before computing anything, what is the equation for CACE that we talked about in lecture from the ITT and the compliance rate? (4 points)
   2. Calculate the compliance rate. In other words, what proportion of people in the treatment condition were successfully contacted? Save this value as a variable called alpha. (8 points)
   3. With alpha in hand, to get the CACE and its standard error, we just need the ITT and its standard error. The regression we ran in 3) already computed the ITT and standard error. Divide your ITT estimate and standard error by your alpha. How do you interpret the effect size and the uncertainty about this result? (6 points)  
      Clarification: how do you interpret the effect size and/or the uncertainty about it.
   4. Instead of calculating the CACE by hand, we can do it using the 'feols' function in R as written below.

# cace\_reg <- feols(outcome ~ 1 | 0 | complied ~ treatment\_assignment, data = this\_data, se = ‘white’)

You need to figure out which variables in your dataset correspond to 'outcome', 'complied', and 'treatment\_assignment'

The point estimate should be the same as when you did the division manually, but the standard error with this approach will be very, very slightly bigger. The reason why is that it is taking into account the uncertainty with which alpha is estimated, too, not just the uncertainty with which the ITT is estimated. (6 points)

**Method 2 for CACE: Comparing compliers in treatment and placebo**

1. Let’s turn to the placebo data. Read in the datafile called. **'Noncompliance\_placebo.csv'**. (4 points)
   * Suggestion: to keep things straight, I would give this dataset a name like **data\_with\_placebo**.
   * *Note: The* ***treatment\_attempt\_turnout\_call*** *variable is still there, but has a slightly different meaning now. When it is set to 1, that still means that someone was in the treatment group attempted with a turnout call. However, when it is set to 0 in this dataset, it now means that they’re in the placebo group attempted with a placebo call.*
2. What is the sample size of this data set? How much smaller or larger is it than the other dataset? ? (6 points)
3. Using regression, examine whether, just within the placebo group, those who answered the phone turn out to vote at the same rates as those who don’t answer the phone. What is your interpretation of this result? Does this indicate that the placebo caused people to vote at a higher rate? Or do you interpret this pattern in another way? (10 points)
4. To estimate the CACE:
   1. Run a regression that calculates the effect of treatment on turnout *only among those who were successfully contacted*. (6 points)
   2. How do we interpret these estimates? (6 points)
   3. This dataset has a much smaller sample size than the first dataset you looked at. Why is the standard error in this dataset not many times bigger? (6 points)
5. Use the feols function to compute the heterogeneous treatment effect of being contacted by whether the person voted in 2002. In words, interpret the interaction term. (10 points)

BONUS: Propose and conduct a test of the assumptions required for a placebo analysis. (5 points)

### How long did this problem set take you in hours? How did you find the level of difficulty?

### 