Assignment 4 - Non-Compliance

Chiebuka Onwuzurike

Questioin 1

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.

```
library(data.table)
library(fixest)
```

```
## Warning: package 'fixest' was built under R version 4.0.5
```

```
library(broom)
library(lfe)
```

```
## Warning: package 'lfe' was built under R version 4.0.4
```

```
## Loading required package: Matrix
```

```
### use the fread function to read the data.
data_treatvscontrol <- fread('noncompliance_treat.csv')
```

Question 2

2. What is the sample size of this dataset?

The sample size of this dataset is 888,238 observations.

```
nrow(data_treatvscontrol)
```

```
## [1] 880238
```

Question 3

3a. 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 ia a more powerful version of the 'lm' function with a very similar syntax.

```
ITT <- feols(voted_aug2008 ~ treatment_attempt_turnout_call, data = data_treatvscontrol, se = 'w
hite')
etable(ITT)
```

```
##                                                  ITT
## Dependent Var.:                        voted_aug2008
##
## (Intercept)                    0.1870*** (0.0004)
## treatment_attempt_turnout_call 0.0132*** (0.0026)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                              880,238
## R2                                        3.02e-5
## Adj. R2                                   2.91e-5
```

3b. Use R to exctract the coefficient and std.error on the treatment into separate variables:

```
# You can create a data frame out of your regression using the package 'broom' and function 'tid
y'
tidy_ITT <- tidy(ITT)

# Extract the coefficient and standard error on treatment:
treat_coef = tidy_ITT$estimate[2]
treat_std = tidy_ITT$std.error[2]
```

Question 4

4a. Before computing anything, what is the equation for CACE that we talked about in lecture from the ITT and the compliance rate?

The CACE is the difference between the average outcome for the control group and average outcome for the treatment group, all divided by the compliance rate.

4b. First, 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.

```
alpha <- mean(data_treatvscontrol[treatment_attempt_turnout_call == 1,contacted])

alpha
```

```
## [1] 0.56659237
```

4c. Divide your ITT estimate and standard error by your alpha. How do you interpret this result?

The ITT divided from the alpha is the complier average causal effect (CACE), the ATE for those who answered the phone. The CACE coeficcient is 0.023335944. The confidence interval is 0.014242206 to 0.032429682 so we can confidently say that CACE is statically significant on the 95% level.

```
treat_coef/alpha
```

```
## [1] 0.023335944
```

```
treat_coef/alpha - treat_std/alpha*1.96
```

```
## [1] 0.014242206
```

```
treat_coef/alpha + treat_std/alpha*1.96
```

```
## [1] 0.032429682
```

4d. Instead of calculating the CACE by hand, we can do it using the ‘feols’ function in R.

```
etable(feols(voted_aug2008 ~ 1 | 0 | contacted ~ treatment_attempt_turnout_call, data = data_tre
atvscontrol, se = 'white'))
```

```
##                   feols(voted_aug2..
## Dependent Var.:      voted_aug2008
##
## (Intercept)      0.1870*** (0.0004)
## contacted        0.0233*** (0.0046)
## _______________  __________________
## S.E. type        Heteroskedas.-rob.
## Observations                880,238
## R2                          0.00018
## Adj. R2                     0.00017
```

Question 5

5. Let's turn to the placebo data. Read in the datafile called. 'Noncompliance_placebo.csv'.

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.

```
data_with_placebo <- fread("Noncompliance_placebo.csv")
```

Question 6

6. What is the sample size of this data set? How much smaller or larger is it than the other dataset?

There is 47,540 observations in the dataset with the placebo which is 832,698 less observations than the dataset without the placebo.

```
nrow(data_with_placebo)
```

```
## [1] 47540
```

```
nrow(data_treatvscontrol)
```

```
## [1] 880238
```

```
nrow(data_treatvscontrol) - nrow(data_with_placebo)
```

```
## [1] 832698
```

Question 7

7. 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?

The people who answered the phone with the placebo had a voting effect of 0.186314 and the people who didn’t answer the phone had a voting effect of 0.186955. The fact that these numbers are close it is a good indication the placebo worked. The placebo did not cause people to vote at a higher rate because it was a similar rate to the control.

```
###  Hint: to run a regression in the subset of the data that is in the placebo group:

etable(feols(voted_aug2008 ~ treatment_attempt_turnout_call, data = data_with_placebo, se = "whi
te"))
```

```
##                                feols(voted_aug2..
## Dependent Var.:                     voted_aug2008
##
## (Intercept)                    0.1863*** (0.0025)
## treatment_attempt_turnout_call 0.0139*** (0.0036)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                               47,540
## R2                                        0.00031
## Adj. R2                                   0.00029
```

```
etable(ITT)
```

```
##                                               ITT
## Dependent Var.:                     voted_aug2008
##
## (Intercept)                    0.1870*** (0.0004)
## treatment_attempt_turnout_call 0.0132*** (0.0026)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                              880,238
## R2                                        3.02e-5
## Adj. R2                                   2.91e-5
```

Question 8

8a. To estimate the CACE, run a regression that calculates the effect of treatment on turnout only among those who were successfully contacted.

```
contacted = subset(data_with_placebo, contacted = 1)
etable(feols(voted_aug2008 ~ treatment_attempt_turnout_call, data = contacted, se = "white"))
```

```
##                                    feols(voted_aug2..
## Dependent Var.:                         voted_aug2008
##
## (Intercept)                    0.1863*** (0.0025)
## treatment_attempt_turnout_call 0.0139*** (0.0036)
## ______________________________ __________________
## S.E. type                      Heteroskedas.-rob.
## Observations                               47,540
## R2                                        0.00031
## Adj. R2                                   0.00029
```

8b. How do we interpret these estimates?

Based on the data calls to encourage voting caused a statistically significant higher voter turn out, when you compare people who picked up the vote call with the placebo call. We are 95% confident there was a 1.3% increase.

8c. 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?

Our smaller data set with the placebo had an extra covariate to reduce the noise and decrease the SE. Although the first dataset was larger, we didn’t know who would have complied. With the smaller dataset we just have to focus if we gave the placebo or the treatment. compliance was 100%.

Question 9

9. Heterogenuous Treatment Effects.

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. Hint, we can add an interaction term by using covariate_variable_name*treatment_variable_name in the formula.

```
feols(voted_aug2008 ~ voted_nov2002*treatment_attempt_turnout_call, data = data_with_placebo, se
="white")
```

```
## OLS estimation, Dep. Var.: voted_aug2008
## Observations: 47,540
## Standard-errors: Heteroskedasticity-robust
##                                              Estimate Std. Error t value
## (Intercept)                                  0.115604   0.003600 32.1110
## voted_nov2002                                0.105855   0.004881 21.6870
## treatment_attempt_turnout_call              -0.008649   0.005024 -1.7216
## voted_nov2002:treatment_attempt_turnout_call  0.032690   0.006905  4.7341
##                                              Pr(>|t|)
## (Intercept)                                 < 2.2e-16 ***
## voted_nov2002                               < 2.2e-16 ***
## treatment_attempt_turnout_call               0.085146 .
## voted_nov2002:treatment_attempt_turnout_call  2.21e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.390512   Adj. R2: 0.021769
```

A person called in 2008 was 3.2% more likely to vote in 2008 if they voted in 2002 compared to someone who didn't vote in 2002.

BONUS: Propose and conduct a test of the assumptions required for a placebo analysis.

We could test to see if the placebo group was randomized properly. To do this we could conduct a t.test on voted_nov2002 to see if the placebo was randomized properly.

```
t.test(data_with_placebo[treatment_attempt_turnout_call==1,voted_nov2002], data_with_placebo[tre
atment_attempt_turnout_call==0,voted_nov2002])
```

```
##
##  Welch Two Sample t-test
##
## data:  data_with_placebo[treatment_attempt_turnout_call == 1, voted_nov2002] and data_with_pl
acebo[treatment_attempt_turnout_call == 0, voted_nov2002]
## t = 1.13115, df = 47537, p-value = 0.258
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.0035738933  0.0133284047
## sample estimates:
##  mean of x  mean of y
## 0.67286261 0.66798535
```

How long did this problem set take you? What is the difficulty level?

Took me about like 4-5 hours. It was pretty difficult.