Problem Set - Non-Compliance

Prateek Naharia

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_small.csv' using the function fread. What is the sample size of this data?

Suggestion: To keep this dataset straight with the other one we'll load in later, I would give it a name like data_treatvscontrol. Remember that the variables are defined on the assignment Google Drive.

```
library(data.table)
library(fixest)
library(broom)
### use the fread function to read the data.
this_data <- fread("noncompliance_treat_small.csv")
sample_size_treat_small = nrow(this_data)
print(paste0("The Sample Size : ",sample_size_treat_small))
```

```
[1] "The Sample Size : 293412"
```

```
#this_data$names
#this_data
```

2. 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. Note the code below, which provides a template for the regression.

```
# Your regression
this_reg <- feols(voted_aug2008 ~ treatment_attempt_turnout_call,
                  data = this_data, se = "hetero")
# Use the function 'etable' or 'modelsummary' to output it nicely.
etable(this_reg)
```

```
                                       this_reg
Dependent Var.:                     voted_aug2008

Constant                          0.1870*** (0.0007)
treatment_attempt_turnout_call     0.0120** (0.0045)
------------------------------ ------------------
S.E. type                         Heteroskedas.-rob.
Observations                                293,412
R2                                          2.48e-5
Adj. R2                                     2.14e-5
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Intent to Treat (ITT) = 0.0120**

4a. 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.

```
ds_turnoutcall_contacted <- nrow(this_data[this_data$
                                           treatment_attempt_turnout_call == 1 &
                                           this_data$contacted == 1, ])

ds_turnoutcall <- nrow(this_data[this_data$treatment_attempt_turnout_call == 1, ])

alpha <- ds_turnoutcall_contacted / ds_turnoutcall
print(paste0('The Compliance Rate: ',alpha))
```

```
[1] "The Compliance Rate: 0.574092741935484"
```

From the above result, we can say that 57.4 % of people in the treatment were successfully contacted

4b. Divide your ITT estimate by alpha and the standard error by alpha. How do you interpret this result? (4 points)

```
# ITT estimate divided by alpha
ITT_by_alpha <- coef(this_reg)[2] / alpha
print(paste0('ITT/Alpha : ',ITT_by_alpha))
```

```
[1] "ITT/Alpha : 0.020856877076219"
```

```
itt_estimate = coef(this_reg, 'treatment_attempt_turnout_call')
standard_error = sqrt(vcov(this_reg)['treatment_attempt_turnout_call',
                                     'treatment_attempt_turnout_call'])
print(paste0('The Standard Error: ',standard_error))
```

```
[1] "The Standard Error: 0.00454045301599279"
```

```
a <- standard_error/alpha
print(paste0('StandardError/Alpha : ',a))
```

```
[1] "StandardError/Alpha : 0.00790891903751509"
```

Interpretation :- CACE - Complied means that - > who got the call and received it.

ITT estimate gives us the average impact of being assigned to the treatment group on voter turnout. The resulting values are the complier average causal effect (CACE) estimate - 0.02 or 2%. Which states that the treatment group has approx 2% highter voter turnout.

StandardError/Alpha can be interpreted as the standard error estimate for those who actually complied.

4c. Instead of calculating the CACE by hand, we can do it using the 'feols' function in R. Notice the standard errors are slightly different.

```
# this_reg <- feols(voted_aug2008 ~ 1 | actually_treated ~ treatment_assignment,
# data = this_data, se = 'hetero')
this_reg_cace <- feols(voted_aug2008 ~ 1 | contacted ~ treatment_attempt_turnout_call,
                       data = this_data, se = 'hetero' )
etable(this_reg_cace)
```

```
                    this_reg_cace
Dependent Var.:      voted_aug2008

Constant        0.1870*** (0.0007)
contacted        0.0209** (0.0079)
--------------- ------------------
S.E. type       Heteroskedas.-rob.
Observations               293,412
R2                         0.00018
Adj. R2                    0.00017
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#this_reg_cace
```

5. Let's turn to the placebo data. Read the datafile called 'noncompliance_placebo.csv'. What is the sample size of this data set? How much smaller or larger is it than the other data?

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_placebo = fread("noncompliance_placebo.csv")
sample_size_placebo = nrow(data_placebo)
#print(paste0("The Sample Size for noncompliance_placebo : ",sample_size))
difference_sample_size = sample_size_treat_small - sample_size_placebo
print(paste0("The Difference: ", difference_sample_size,
             ", & The Sample Size for noncompliance_placebo :",sample_size_placebo ))
```

```
[1] "The Difference: 245872, & The Sample Size for noncompliance_placebo :47540"
```

Sample Size for noncompliance_placebo is 47540 Sample Size for noncompliance_treat_small.csv is 293412 Difference: 245872

6. 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? No more than two sentences.

```
### Hint: to run a regression in the subset of the data that is in the placebo group:
sub_group = data_placebo[data_placebo$treatment_attempt_turnout_call==0]

this_reg_placebo <- feols(voted_aug2008 ~ contacted,
                          data = sub_group, se = 'hetero')
etable(this_reg_placebo)
```

```
                 this_reg_placebo
Dependent Var.:      voted_aug2008

Constant        0.1618*** (0.0036)
contacted       0.0438*** (0.0050)
--------------- ------------------
S.E. type       Heteroskedas.-rob.
Observations                23,761
R2                         0.00312
```

```
Adj. R2                     0.00307
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The regression results indicate that within the placebo group, (A positive correlation is observed between contacted & voter turnout) those who answered the phone were more likely to turn out to vote (coeff = 0.0438***, p < 0.001). However, the effect size is small and the R-squared value is low (0.0031), suggesting that this pattern may not be practically significant. Therefore, It is not appropriate to interpret this as evidence that the placebo caused people to vote at a higher rate.

7a. To estimate the CACE, run a regression that calculates the effect of treatment on turnout only among those who were successfully contacted.

```
sub_group_contacted = data_placebo[data_placebo$contacted==1,]

reg_cace <- feols(voted_aug2008 ~ treatment_attempt_turnout_call,
                          data = sub_group_contacted, se = 'hetero')
etable(reg_cace)
```

```
                                    reg_cace
Dependent Var.:                  voted_aug2008

Constant                      0.2056*** (0.0035)
treatment_attempt_turnout_call 0.0306*** (0.0051)
------------------------------ ------------------
S.E. type                      Heteroskedas.-rob.
Observations                                26,793
R2                                         0.00136
Adj. R2                                    0.00132
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer - treatment_attempt_turnout_call 0.0306*** (0.0051) 3%

7b. How do we interpret these estimates?

Regression Interpretation:- The coefficient for the treatment_attempt_turnout_call variable is 0.0306 with a standard error of 0.0051. This means that among those who were successfully

contacted, being assigned to the treatment group (receiving a call encouraging them to vote) increases the probability of voting in the August 2008 primary election by 3.06 percentage points on average, compared to the control group (not receiving a call or receiving a placebo call). This effect is statistically significant at the ***/0.001 level. However, the R-squared value of the model is quite low (0.00136), indicating that the treatment variable does not explain much of the variation in the outcome variable.

7c. 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?

The standard error (SE) is a measure of the precision of the estimate of the treatment effect. It is calculated as the standard deviation of the sampling distribution of the estimate. While a smaller sample size would generally lead to a larger standard error,

In this case, it's possible that the smaller sample size is offset by a lower level of variability in the outcome variable or a stronger treatment effect, which would result in a smaller standard error. However, it's important to note that a smaller standard error doesn't necessarily mean that the estimate is more reliable or that the treatment effect is larger. It simply means that the estimate is more precise. Additionally, the use of a heteroskedasticity-robust standard error calculation method may also contribute to a smaller standard error.

8: Run a regression to test whether the effects of the turnout call vary by whether the person voted in 2002. Interpret the results.

```
# Reminder, we can add an interaction effect as follows:
#subset_voted <- data_placebo[data_placebo$contacted ==1]
reg_personvoted2002 <- feols(voted_aug2008 ~ treatment_attempt_turnout_call*voted_nov2002,
                             data = data_placebo, se = 'hetero')
etable(reg_personvoted2002)
```

```
                                          reg_personvo..2002
Dependent Var.:                                voted_aug2008

Constant                                  0.1156*** (0.0036)
treatment_attempt_turnout_call             -0.0086. (0.0050)
voted_nov2002                             0.1059*** (0.0049)
treatment_attempt_turnout_call x voted_nov2002 0.0327*** (0.0069)
----------------------------------------  ------------------
S.E. type                                 Heteroskedas.-rob.
Observations                                          47,540
```

```
R2                                                          0.02183
Adj. R2                                                     0.02177
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above results we can interpret that:- The treatment effect of the turnout call on voting behavior is moderated, by whether the individual voted in the previous election (voted_nov2002).

For individuals who did not vote in the previous election, the treatment effect appears to be -0.0086, stating decrease in the probability of voting & is not statistically significant.

For individuals who did vote in the previous election, the treatment effect appears to be 0.0327*** or 3.27%. Which states the increase in the probability of voting(statistically significant at ***/0.001 level) We can say that people who voted in 2002 have higher turn-out rate in 2008.

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

Placebo Analysis, The following assumptions can be made: 1. Placebo having no effect on the outcome. 2. Same placebo group compliance rate as of treatment and control group.

```r
turnout_contact_one = data_placebo[data_placebo$treatment_attempt_turnout_call ==1,
                                   contacted]
turnout_contact_nocall = data_placebo[data_placebo$treatment_attempt_turnout_call ==0,
                                      contacted]

t.test(turnout_contact_one,turnout_contact_nocall)
```

```
    Welch Two Sample t-test

data:  turnout_contact_one and turnout_contact_nocall
t = 1.3211, df = 47537.7, p-value = 0.18647
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0029065082  0.0149263169
sample estimates:
 mean of x  mean of y
0.56659237 0.56058247
```

Welch two-sample t-test The output is as - t = 1.3211 degrees of freedom are 47537.7 The p-value is 0.18647 indicates that we cannot reject the null hypothesis, that the true difference in means compliance rate between the treatment and placebo group is zero.

The 95% confidence interval for the difference in means ranges from -0.0029 to 0.0149, which includes zero, further supporting the conclusion that, there is no significant difference in means between the two groups.

Based on these results, we can assume that the placebo group is a valid control group for conducting a placebo analysis to test the assumptions of the causal effect of the treatment group.

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

comparatively difficult, took 1 day to understand and complete.