Problem Set - Non-Compliance

Your Name

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.

data_treatcon <- fread('noncompliance_treat_small.csv')
nrow(data_treatcon)
```

```
[1] 293412
```

There are 293,412 observations.

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
# Note, 'hetero' here changes the standard errors to be heteroskedasticity robust.
reg2 <- feols(voted_aug2008 ~ treatment_attempt_turnout_call,
data = data_treatcon, se = 'hetero')

# Use the function 'etable' or 'modelsummary' to output it nicely.
etable(reg2)
```

```
                                              reg2
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
```

The intent-to-treat effect is 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.

```
alpha = nrow(subset(data_treatcon,treatment_attempt_turnout_call == 1 & contacted == 1))/n
alpha
```

```
[1] 0.57409274
```

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

```
ITT = 0.012
ITT_std = 0.0045
CACE = ITT/ alpha
CACE_std = ITT_std / alpha
CACE
```

```
[1] 0.020902546
```

```
CACE_std
```

```
[1] 0.0078384548
```

Assumption: if the randomization is done properly,the share of non-compliers is the same in the treatment and in the control group.

Interpretation: voted proportion increases by 0.02 due to answering the phone call about voting for those who comply.

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(outcome ~ 1 | actually_treated ~ treatment_assignment,
# data = this_data, se = 'hetero')
```

```
reg4 <- feols(voted_aug2008 ~ 1 | contacted ~ treatment_attempt_turnout_call,
data = data_treatcon, se = 'hetero')

etable(reg4)
```

```
                              reg4
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
```

As we can see,the CACE from feols is 0.0209. Notice that there's slight difference in standard error given by hand and directly from feols.(0.0078 vs 0.0079).

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?

```
data_treatplb <- fread('noncompliance_placebo.csv')
nrow(data_treatplb)
```

```
[1] 47540
```

```
percentage = (293412-47540)/293412
percentage
```

```
[1] 0.83797527
```

The placebo sample size is 83.8% percentage smaller than the treatment_control dataset.

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.

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.

```
placebo <- subset(data_treatplb,treatment_attempt_turnout_call == 0)
reg6 <- feols(voted_aug2008 ~ contacted,
data = placebo, se = 'hetero')

etable(reg6)
```

```
                              reg6
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 result is significant at 0.1% level. At 5% significant level,we have strong evidence against the null hypothesis that the coefficient of 'contacted' is 0. That is to say, we reject the null hypothesis that the vote rate is the same within two groups (contacted& non-contacted).

This just indicates that people contacted in placebo group tend to vote at a higher rate. This is not related to placebo since it compares the difference of whether a person is contacted,not whether a person is using placebo.

7a. 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_treatplb,contacted == 1)
reg7 <- feols(voted_aug2008 ~ treatment_attempt_turnout_call,
data = contacted, se = 'hetero')
```

```r
etable(reg7)
```

```
                                              reg7
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
```

7b. How do we interpret these estimates?

The vote proportion increases by 0.0306 due to answering the phone call about voting for those who comply.

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 is 0.0051,which is smaller than the standard error that we have in 3(c) (i.e,0.0079).

Possible explanation is that treatment is a significant factor in treatment_placebo dataset with those who are contacted. Smaller standard error indicates smaller variability around the estimate of the regression slope. It is not merely determined by sample size.

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

```r
# Reminder, we can add an interaction effect as follows:
reg8<- feols(voted_aug2008 ~ treatment_attempt_turnout_call*voted_nov2002, data = data_tre
etable(reg8)
```

```
                                                           reg8
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
```

The constant value is 0.1156.The voting proportion of control group who didn't vote in 2002 is 0.1156. The result is significant.

The coefficient of treatment is -0.0086. The voting proportion of treatment group who didn't vote in 2002 is 0.0086 lower than those in placebo group who didn't vote.(i.e,0.107.).The coefficient is only significant at 10% significance level.At 5% significance level,we fail to reject that the coefficient of treatment is 0.

The coefficient of voting is 0.1059. The voting proportion of those who voted in 2002 is 0.1059 higher than those who didn't vote in 2002 in the placebo group. The result is significant at 0.1%level.

The coefficient of voting*treatment is 0.0327.The voting proportion of those who voted in 2002 is (0.1059+0.0327) = 0.1386 higher than those who didn't vote in 2002 in the treatment group.The result is significant at 0.1% level.

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

Assumption: the placebo has no effect on the outcome.Hence, we need to compare the proportion of those who voted in control group and those who voted in placebo group.

```r
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:data.table':

    between, first, last


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
control_pr <- data_treatcon %>%
  filter(treatment_attempt_turnout_call == 0) %>%
  pull(voted_aug2008)

placebo_pr <- data_treatplb %>%
  filter(treatment_attempt_turnout_call == 0) %>%
  pull(voted_aug2008)

# Compute t-test
res <- t.test(control_pr, placebo_pr) # assume equal variance
res
```

```
	Welch Two Sample t-test

data:  control_pr and placebo_pr
t = 0.25834, df = 27875.5, p-value = 0.79615
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0044742490  0.0058327386
sample estimates:
 mean of x  mean of y 
0.18699295 0.18631371 
```

Let μ_1 denote the proportion of voted people in control group.

Let μ_2 denote the proportion of voted people in placebo group.

Null hypothesis : $\mu_1 = \mu_2$ Alternative hypothesis : $\mu_1 \neq \mu_2$

p-value = 0.79 as above.At 5% significance level, we fail to reject the null hypothesis.

Hence,the proportion of voting in treatment and control group is same. The assumption is satisfied.

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

2h.