Problem Set #4

Experiments and Causality

```
#install.packages("tinytex")
#tinytex:::install_prebuilt()
#install.packages("ivreg")
```

```
# load packages
library(foreign)
library(data.table)
library(knitr)
library(lmtest)
library(ivreg)
library(sandwich)
library(MASS)
```

For INFO 290, Fall 2020:

- Let's streamline this problem set somewhat.
- Everyone can **optionally** skip questions 1 & 2
- Everyone can **optionally** skip question 5.
- Everyone **has** to answer questions 3 & 4.
- Question 5 is, in my opinion, a nice challenge quesiton for people who want to work as data scientists. The data is of *just* the wrong size – it fits on your laptop but it is pretty clumsy. This is the type of question that I would give out on an interview data challenge; but, I'd probably ask you to do more in an interview.

Question 3: Fun with the placebo

The table below summarizes the data from a political science experiment on voting behavior. Subjects were randomized into three groups: a baseline control group (not contacted by canvassers), a treatment group (canvassers attempted to deliver an encouragement to vote), and a placebo group (canvassers attempted to deliver a message unrelated to voting or politics).

Assignment	Treated?	N	Turnout
Baseline	No	2463	0.3008
Treatment	Yes	512	0.3890
Treatment	No	1898	0.3160
Placebo	Yes	476	0.3002
Placebo	No	2108	0.3145

1. Construct a data set that would reproduce the table.

```
#count <- 2463 + 512 + 1898 + 476 + 2108 = 7457
df_3 <- data.frame(matrix(ncol = 3, nrow = 0))
colnames(df_3) <- c("Assignment", "Treated?","Turnout")

df_baseline <- data.frame(Assignment = c("Baseline"), Treated = c(0),
                          Voted = c(rep(1, round(2463*0.3008)), rep(0, round(2463*(1-0.3008)))))
df_treatment_yes <- data.frame(Assignment = c("Treatment"), Treated = c(1),
```

```
                                Voted = c(rep(1, round(512*0.3890)), rep(0, round(512*(1-0.3890)))))

df_treatment_no <- data.frame(Assignment = c("Treatment"), Treated = c(0),
                              Voted = c(rep(1, round(1898*0.3160)), rep(0, round(1898*(1-0.3160)))))
df_placebo_yes <- data.frame(Assignment = c("Placebo"), Treated = c(1),
                             Voted = c(rep(1, round(476*0.3002)), rep(0, round(476*(1-0.3002)))))
df_placebo_no <- data.frame(Assignment = c("Placebo"), Treated = c(0),
                            Voted = c(rep(1, round(2108*0.3145)), rep(0, round(2108*(1-0.3145)))))

df_3bind <- rbind(df_baseline, df_treatment_yes, df_treatment_no, df_placebo_yes, df_placebo_no)
#df_3bind

df_3dt <- data.table(df_3bind)
head(df_3dt)
```

```
##    Assignment Treated Voted
## 1:   Baseline       0     1
## 2:   Baseline       0     1
## 3:   Baseline       0     1
## 4:   Baseline       0     1
## 5:   Baseline       0     1
## 6:   Baseline       0     1
```

2. Estimate the proportion of compliers by using the data on the treatment group.

```
sum(df_3bind$Assignment == "Treatment" & df_3bind$Treated == 1) / sum(df_3bind$Assignment == "Treatment"
```

```
## [1] 0.2124481
```

0.2124481

3. Estimate the proportion of compliers by using the data on the placebo group.

```
sum(df_3bind$Assignment == "Placebo" & df_3bind$Treated == 1) / sum(df_3bind$Assignment == "Placebo")
```

```
## [1] 0.1842105
```

0.1842105

4. Are the proportions in parts (1) and (2) statistically significantly different from each other? Provide *a test* and an description about why you chose that particular test, and why you chose that particular set of data. **Since we are dealing with categorical variables and to test if the variables are independent, chi-test will be used. Since (1) and (2) is dealing between compliers and non-compliers, we look into Assignment and the Treated, with focuses on Treatment & Placebo as the variables to test.**

```
tbl <- table(df_3bind$Assignment, df_3bind$Treated)
tbl <- tbl[-1,]
tbl
```

```
##
##                0    1
##   Treatment 1898  512
##   Placebo   2108  476
```

```
chisq.test(tbl)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
```

```
##
## data:  tbl
## X-squared = 6.0887, df = 1, p-value = 0.0136
```

Looking at the p-value of 0.0136, we can conclude that we can regect the null hypothesis of no relationsihp at the 0.01 level or we have insufficient evidence to reject the null at the 0.01 level. Thus the conclusion is that (1) and (2) are significantly different from each other.

5. What critical assumption does this comparison of the two groups' compliance rates test? **The assumption that the chiq-test is testing is given that we've randomized, the fraction of the compliers in the treatment and the placebo group are the same(there is no independence between placebo and treatment group). Looking at the p-value, we can conclude that the test is statistically significant and thus reject the assumption(null hypothesis) that there is independence.**

6. Estimate the CACE of receiving the placebo. Is the estimate consistent with the assumption that the placebo has no effect on turnout? **The CACE of receiving placebo is 0.06007739. When tested on 2SLS test on all data points except those assigned to treatment. The p-value is 0.3945 which means it is not statistically significant and the hypothesis that the placebo has no effect on turnout cannot be rejected.**

```
df_3dt <- data.table(df_3bind)

itt_m <- df_3dt[Assignment == "Placebo", mean(Voted)] - df_3dt[Assignment=="Baseline",mean(Voted)]
CACE_3 <- itt_m / 0.1842105

CACE_3
```

```
## [1] 0.06007739
```

```
df_3d_wt <- df_3dt[Assignment != 'Treatment'] #all data without treatment
df_cace_p <- ivreg(Voted ~ Treated, ~ Assignment, data = df_3d_wt)
coeftest(df_cace_p, vcovHC(df_cace_p))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.300853   0.009245 32.5423   <2e-16 ***
## Treated     0.060077   0.070551  0.8515   0.3945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7. Estimate the CACE of receiving the treatment by first estimating the ITT and then dividing by ITT_D.

```
itt_7 <- df_3dt[Assignment == "Treatment", mean(Voted)] - df_3dt[Assignment=="Baseline",mean(Voted)]
itt_7d <- df_3dt[Assignment == "Treatment" & Treated == 1,.N]/df_3dt[Assignment == "Treatment",.N]

itt_7/itt_7d
```

```
## [1] 0.1444242
```

CACE = 0.1444242

8. Estimate the CACE of receiving the treatment by comparing the turnout rates among the compliers in both the treatment and placebo groups.

```
turnout_8 <- df_3dt[Assignment == "Treatment" & Treated == 1, mean(Voted)] - df_3dt[Assignment=="Placebo
turnout_8
```

```
## [1] 0.08825171
```

CACE = 0.08825171

9. Estimate the CACE of receiving the treatment using two stage least squares regression. You could either estimate two different models; or, you could use the `ivreg` package in the `AER` library.

```
df_3d_wt <- df_3dt[Assignment != 'Placebo'] #all data without treatment
df_cace_p <- ivreg(Voted ~ Treated, ~ Assignment, data = df_3d_wt)
df_cace_p
```

```
##
## Call:
## ivreg(formula = Voted ~ Treated | Assignment, data = df_3d_wt)
##
## Coefficients:
## (Intercept)      Treated
##      0.3009       0.1444
```

```
#coeftest(df_cace_p, vcovHC(df_cace_p))
```

CACE = 0.1444

10. We have a theroetical expectation that the results from 7, 8, and 9 should be the same. Are they? If so, does this give you confidence that these methods are working well. If not, what explains why these estimators are producing different estimates?

No, they are not all the same. This may be due to the difference in proportion of compliers being different in the treatment and placebo group that we assumed to be the same.

11. In class we discussed that the rate of compliance determines whether one or another design is more efficient. (You can review the textbook expectation on page 162 of *Field Experiments*)). Given the compliance rate in this study, which design *should* provide a more efficient estimate of the treatment effect?

```
comply_rate <- df_3dt[Assignment != 'Baseline', mean(Treated)]
comply_rate
```

```
## [1] 0.1978374
```

According to Field Experiments, conventional design is preferable when the compliance rate is more than 0.5. The compliance rate of this study is 0.1978374 which means placebo design should provide a more efficient estimate of the treatment effect.

12. When you apply what you've said in part (11) against the data that you are working with, does the {placebo vs. treatment} or the {control vs. treatment} comparison produce an estimate with smaller standard errors?

Standard error for placebo vs. treatment is 0.030124, while standard error for control vs treatment is 0.0626916 This supports the statement made above that placebo design provide more efficient results

```
#cluster by group
#placebo vs. treatment
df_3d_12 <- df_3dt[Treated == 1]
df_cace_121 <- lm(Voted ~ Assignment, data = df_3d_12)
coeftest(df_cace_121, vcovCL(df_cace_121))
```

```
##
## t test of coefficients:
##
```

```
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)         0.388672   0.021564 18.0239 < 2.2e-16 ***
## AssignmentPlacebo -0.088252   0.030124 -2.9296  0.003472 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#control vs. treatment
coeftest(df_cace_p, vcovCL(df_cace_p))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.3008526  0.0092422 32.5522  < 2e-16 ***
## Treated     0.1444242  0.0626916  2.3037  0.02128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 4: Turnout in Dorms

Guan and Green report the results of a canvassing experiment conduced in Beijing on the eve of a local election. Students on the campus of Peking University were randomly assigned to treatment or control groups. Canvassers attempted to contact students in their dorm rooms and encourage them to vote. No contact with the control group was attempted. Of the 2,688 students assigned to the treatment group, 2,380 were contacted. A total of 2,152 students in the treatment group voted; of the 1,334 students assigned to the control group, 892 voted. One aspect of this experiment threatens to violate the exclusion restriction. At every dorm room they visited, even those where no one answered, canvassers left a leaflet encouraging students to vote.

```
d <- fread('./data/Guan_Green_CPS_2006.csv')
d
```

```
##       turnout treated   dormid treatment_group
##    1:       0       0  1010101               0
##    2:       0       0  1010101               0
##    3:       0       0  1010101               0
##    4:       0       0  1010102               0
##    5:       0       0  1010102               0
##   ---
## 4020:       1       1 24033067               1
## 4021:       1       1 24033068               1
## 4022:       1       1 24033068               1
## 4023:       1       1 24033068               1
## 4024:       1       1 24033068               1
```

Here are definitions for what is in that data:

- `turnout` did the person turn out to vote?
- `treated` did someone at the dorm open the door?
- `dormid` a unique ID for the door of the dorm
- `treatment_group` whether the dorm door was assigned to be treated or not

1. Using the data set from the book's website, estimate the ITT. **First**, estimate the ITT using the difference in two-group means. **Second**, estimate the ITT using a linear regression on the appropriate subset of data. **Heads up**: There are two NAs in the data frame. Just na.omit to remove these rows so that we are all working with the same data.

```
d_clean <- na.omit((d))

itt_dm<- d_clean[treatment_group == 1, mean(turnout)] - d_clean[treatment_group == 0, mean(turnout)]
itt_dm
```

```
## [1] 0.1319296
```

```
itt_linear <- lm(turnout ~ treatment_group, d_clean)
summary(itt_linear)$coefficients[2,1]
```

```
## [1] 0.1319296
```

The ITT using the difference in two-groups means and ITT estimate using a linear regression is both 0.1319296

2. Use RI to test!

a. How many people are in treatment and control? Does this give you insight into how the scientists might have randomized?

```
treatment_g <- d_clean[treatment_group ==1, sum(treatment_group)]
treatment_g
```

```
## [1] 2688
```

```
control_g <- d_clean[treatment_group == 0 , sum(treatment_group + 1)]
control_g
```

```
## [1] 1334
```

The number of people in the treatment is 2688, while the number of people in the control group is 1334. This does not give ne insight to how the scientists might have randomized?..

b. Write an algorithm to conduct the Randomization Inference. Be sure to take into account the fact that random assignment was clustered by dorm room.

```
u_dormids <- d_clean[,unique(dormid)]

randomize_clustered <- function() {
  treat_dormid <- sample(x = u_dormids,
                         size = length(u_dormids)/2,
                         replace = FALSE)
  return(as.numeric(d_clean$dormid %in% treat_dormid))
}

est_ate <- function(outcome, treat) {
  mean(outcome[treat==1]) - mean(outcome[treat==0])
}
```

c. What is value that you estimate for the treatment effect?

```
#sim
ate5 <- est_ate(d_clean$turnout, randomize_clustered())
ate5
```

```
## [1] 0.01564348
```

d. What is the standard deviation of the RI sampling distribution?

```
distribution_under_sharp_null <- replicate(1000, est_ate(d_clean$turnout, randomize_clustered())) #est_
sqrt(mean((distribution_under_sharp_null - mean(distribution_under_sharp_null))^2))
```

```
## [1] 0.02082448
```

e. What is the p-value that you generate for the test: How likely is this treatment effect to have been generated if the sharp null hypothesis were true.

```
#p_values[power_loop] <- mean(abs(ri) > abs(ate))
ate <- d_clean[treated == 1, mean(turnout)] - d_clean[treated == 0 , mean(turnout)]
p_value5 <- mean(abs(distribution_under_sharp_null) > abs(ate))
p_value5
```

```
## [1] 0
```

The p-value is 0, which means there is a statistical significance and thus can reject the sharp null hypothesis. It is less likely that the treatment effect have been generated due to the sharp null hypothesis

3. Assume that the leaflet had no effect on turnout. Estimate the CACE. Do this in two ways:

a. First, estimate the CACE using means.

```
itt_mean <- d_clean[treatment_group==1,mean(turnout)] - d_clean[treatment_group==0,mean(turnout)]
itt_d <- d_clean[treatment_group==1 & treated==1,.N]/d_clean[treatment_group==1,.N]
CACE <- itt_mean / itt_d

CACE
```

```
## [1] 0.1489402
```

0.1489402 b. Second, use some form of linear model to estimate this as well. If you use a 2SLS, then report the standard errors and draw inference about whether contact had any causal effect among compliers.

```
cace_ag <- ivreg(turnout ~ treated, ~treatment_group, data = d_clean)
#summary(cace_ag)
#considering clusters
cace_coef <- coeftest(cace_ag, vcovCL(cace_ag, cluster = d_clean[ , dormid]))
cace_coef
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 0.668666   0.020239 33.0390 < 2.2e-16 ***
## treated     0.148940   0.026308  5.6614 1.607e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cace_se <- cace_coef[2,2]
cace_se
```

```
## [1] 0.02630817
```

In order to take into account of clusters(dormid) we do vcovCL. Standard error is 0.02630817,coefficient 0.148940. The p-value is 1.607e-08 which indicates that the contact causal effect among compliers is statistically significant.