Problem Set #4

Experiment Design: Atit Wongnophadol

```r
# load packages
library(foreign)
library(data.table)
library(stargazer)
```

```
##
## Please cite as:

##   Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##   R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
library(magrittr)
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
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
# removes all objects from the current workspace (R memory)
rm(list = ls())
```

1. Potential Outcomes

a. Make up a hypothetical schedule of potential outcomes for three Compliers and three Never-Takers where the ATE is positive but the CACE is negative. By ATE, we mean the average treatment effect for the entire population, including both compliers and never-takers. Note that we can never compute this ATE directly in practice, because we never observe both potential outcomes for any individual, especially for never-takers. That's why this question requires you to provide a complete table of hypothetical potential outcomes for all six subjects.

```
make_data <- function(n=6) {

  set.seed(5)

  ## require(data.table)

  ## assign two quantities that are not related to one another
  ##  - Z is your treatment assignment
  ##  - X is a covaraite
  dt <- data.table(id = 1:n)

  dt[ , Z := rep(0:1, each=n/2)]
  dt[ , Y0 := runif(.N, min = 0, max = 10)]
  dt[ , Y1 := ifelse(Z==1, Y0 + rnorm(n = 1, mean = 2, sd = 1), Y0 - rnorm(n = 1, mean = 0, sd = 1))]
  dt[ , Type := ifelse(Z==1,"Complier","Never-Taker")]

  return(dt[ , .(Z, Y0, Y1, Type)])
  }

d <- make_data(6)

library(dplyr)

ate <- d %>%
        group_by(Type) %>%
        summarise(mean = mean(Y1-Y0, na.rm = T), count = n())

cace <- d %>%
        summarise(mean = mean(Y1-Y0, na.rm = T), count = n())

d

##    Z       Y0        Y1       Type
## 1: 0 2.002145 0.2907037 Never-Taker
## 2: 0 6.852186 5.1407451 Never-Taker
## 3: 0 9.168758 7.4573169 Never-Taker
## 4: 1 2.843995 4.9141373    Complier
## 5: 1 1.046501 3.1166440    Complier
## 6: 1 7.010575 9.0807174    Complier
```

   b. Suppose that an experiment were conducted on your pool of subjects. In what ways would the estimated CACE be informative or misleading?

Answer: CACE can be informative for a study that requires high degree of precision in estimating the treatment effect. For example, a study of the impact on a drug. But CACE can be misleading as well as a program that works well among Compliers may not produce satistfactory results when applied to a broard population which include both Compliers and Never-Takers. For example, a drug that is highly effective among Compliers may not work so well with Never-Takers.

Reference: FE Chapter 5 Page 147-148.

   c. Which population is more relevant to study for future decision making: the set of Compliers, or the set of Compliers plus Never-Takers? Why?

Answer: Given the data in this study it appears that each set on its own –Compliers and Never-Takers–would be relevant to the study for future decision making. The same treatment applied to Compliers results in a positive effect, but would have resulted in a negative effect to Never-Takers. The implication is that the treatment yields an opposite effect, and a decision maker should study the impact of the treatment for each set of population, so he could provide conclusion / recommendation that is specific to each population. Combining both sets in a study would not be appropriate, as the outcome would point to one direction or another, if not null, which can be misleading in understanding the true impact of the treatment.

2. Turnout to Vote

Suppose that a researcher hires a group of canvassers to contact a set of 1,000 voters randomly assigned to a treatment group. When the canvassing effort concludes, the canvassers report that they successfully contacted 500 voters in the treatment group, but the truth is that they only contacted 250. When voter turnout rates are tabulated for the treatment and control groups, it turns out that 400 of the 1,000 subjects in the treatment group voted, as compared to 700 of the 2,000 subjects in the control group (none of whom were contacted).

    a. If you believed that 500 subjects were actually contacted, what would your estimate of the CACE be?

```r
# proportion of Compliers
Alpha <- 500/1000

# Intent to treat effect (ITT)
ITT <- (400/1000) - (700/2000)

# Calculate CACE
CACE <- ITT/Alpha

# Print CACE
sprintf("If 500 subjects were actually treated, the CACE would be %.3f", CACE)
```

```
## [1] "If 500 subjects were actually treated, the CACE would be 0.100"
```

    b. Suppose you learned that only 250 subjects were actually treated. What would your estimate of the CACE be?

```r
# proportion of Compliers
Alpha_250 <- 250/1000

# Intent to treat effect (ITT)
ITT_250 <- (400/1000) - (700/2000)

# Calculate CACE
CACE_250 <- ITT_250/Alpha_250

# Print CACE
sprintf("If 250 subjects were actually treated, the CACE would be %.3f", CACE_250)
```

```
## [1] "If 250 subjects were actually treated, the CACE would be 0.200"
```

    c. Do the canvassers' exaggerated reports make their efforts seem more or less effective? Define effectiveness either in terms of the ITT or CACE. Why does the definition matter?

ANSWER: The exaggerated reports make their efforts seem less effective. ITT (intent to treat effect) is the overall effectiveness of the canvassing program, whereas CACE (complier average causal effect) is the average treatment effects among the compliers. The definitions matter depending on the objective of the study.

3. 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.

```r
library(foreign)
d <- read.dta("./data/Guan_Green_CPS_2006.dta")

# rename the columns
names(d) <- c('turnout','contacted','dorm','assigned')

head(d)
```

```
##   turnout contacted    dorm assigned
## 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
## 6       0         1 1010103        1
```

a. Using the data set from the book's website, estimate the ITT. First, estimate the ITT using the difference in two-group means. Then, 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.*

```r
# remove rows with NAs
d <- na.omit(d)

# summarize turnout rate by groups
turnout_rate <- d %>%
    group_by(assigned) %>%
    summarise(mean = mean(turnout, na.rm = T), count = n())

# calculate ITT
ITT_3 <- turnout_rate$mean[2] - turnout_rate$mean[1]

# Print ITT from the difference in two-group means
sprintf("ITT using the difference in two-group means is %.3f", ITT_3)
```

```
## [1] "ITT using the difference in two-group means is 0.132"
```

```r
# fit a linear model
mod3a <- lm(turnout ~ assigned, data = d)
```

```r
# Print ITT from a linear regression model
sprintf("ITT using a linear regression is %.3f", summary(mod3a)$coefficients[2,1])
```

```
## [1] "ITT using a linear regression is 0.132"
```

b. Use randomization inference to test the sharp null hypothesis that the ITT is zero for all observations, taking into account the fact that random assignment was clustered by dorm room. Interpret your results.

```r
# calculate the treatment and control ratios
ratio_treatment <- sum(d$assigned)/nrow(d)
ratio_control <- 1-ratio_treatment

# structure the data into the data table
d <- data.table(d)

# create function for randomization
ri <- function(){

  # find unique dorm id
  dorm <- data.table(d %>%
    group_by(dorm) %>%
    summarise())

  # number of dorms
  n_dorm <- nrow(dorm)

  # randomly assign treatment to the dorms
  dorm <- dorm[ , clustered_treatment :=
              sample(c(rep(0,round(ratio_control*n_dorm)),
                       rep(1,round(ratio_treatment*n_dorm))))]

  # assign to the treatment to individuals in the dorms accounting for clustering
  d <- merge(x = d, y = dorm, by = "dorm", all.x = TRUE)

  # calculate ITT
  d[ , .(mean_turnout = mean(turnout)), by = .(sample(clustered_treatment))
    ][ , mean_turnout[sample==1] - mean_turnout[sample==0]]

}

# simulate ITT
res <- replicate(10000, ri())

# plost histogram of ITT
hist(res)
```
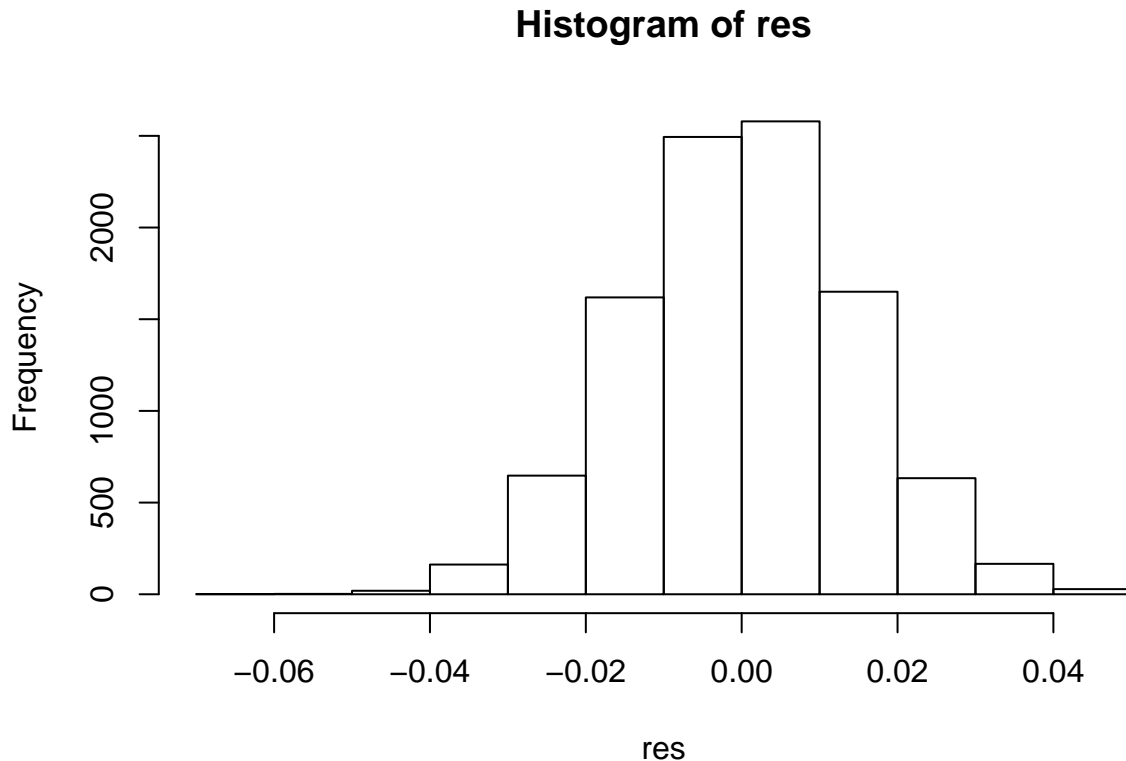
Histogram of res


```
# Print the one-sided p-value
sprintf("P-value given clustered randomization inference is %.5f", mean(res >= ITT_3))
```

```
## [1] "P-value given clustered randomization inference is 0.00000"
```

ANSWER: Given the p-value that is practically zero, the null hypothesis that ITT is 0 is rejected. This means that ITT is statistically significante given the clustered design. It is unlikely that this ITT is by chance.

c. Assume that the leaflet had no effect on turnout. Estimate the CACE. Do this in two ways: First, estimate the CACE using means. 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 the leaflet had any causal effect among compliers.

```
### using means to estimate CACE

# proportion of compliers
ratio_compliers <- sum(d$contacted)/sum(d$assigned)

# CACE
CACE_3c <- ITT_3 / ratio_compliers

# Print CACE of the model
sprintf("CACE from mean calculation is %.3f", CACE_3c)
```

```
## [1] "CACE from mean calculation is 0.149"
```

```
### using linear regression to estimate CACE

# fit a linear model to predict "contacted" given "assigned"
lm_contacted <- lm(data = d, contacted ~ assigned)
```

```
d <- d[ , pred_contacted:= predict(lm_contacted)]

lm_cace <- lm(data = d, turnout ~ pred_contacted)

# Print CACE from the linear regression model
sprintf("CACE from the linear model is %.3f", summary(lm_cace)$coefficients[2,1])

## [1] "CACE from the linear model is 0.149"
```

4. Why run a placebo?

Nickerson describes a voter mobilization experiment in which subjects were randomly assigned to one of three conditions: a baseline group (no contact was attempted); a treatment group (canvassers attempted to deliver an encouragement to vote); and a placebo group (canvassers attempted to deliver an encouragement to recycle). Based on the results in the table below answer the following questions

Treatment Assignment	Treated ?	N	Turnout
Baseline	No	2572	31.22%
Treatment	Yes	486	39.09%
Treatment	No	2086	32.74%
Placebo	Yes	470	29.79%
Placebo	No	2109	32.15%

First Use the information to make a table that has a full recovery of this data. That is, make a `data.frame` or a `data.table` that will have as many rows a there are observations in this data, and that would fully reproduce the table above. (*Yes, this might seem a little trivial, but this is the sort of "data thinking" that we think is important.*)

```
# count of each group
base <- 2572
d_treated <- 486
d_no_treated <- 2086
p_treated <- 470
p_no_treated <- 2109

# total
N <- base + d_treated + d_no_treated + p_treated + p_no_treated

# turnout rate
turnout_baseline <- 0.3122
turnout_d_treated <- 0.3909
turnout_d_no_treated <- 0.3274
turnout_p_treated <- 0.2979
turnout_p_no_treated <- 0.3215


# create dataframe with ID
d4 <- data.table(id = sample(1:N, replace = F))

# create data
d4[ , turnout:= c(rep(0, round(base*(1-turnout_baseline)))),
```

```
                    rep(1, round(base*turnout_baseline)),
                    rep(0, round(d_treated*(1-turnout_d_treated))),
                    rep(1, round(d_treated*turnout_d_treated)),
                    rep(0, round(d_no_treated*(1-turnout_d_no_treated))),
                    rep(1, round(d_no_treated*turnout_d_no_treated)),
                    rep(0, round(p_treated*(1-turnout_p_treated))),
                    rep(1, round(p_treated*turnout_p_treated)),
                    rep(0, round(p_no_treated*(1-turnout_p_no_treated))),
                    rep(1, round(p_no_treated*turnout_p_no_treated)))
    ][ , treated:= c(rep(0, base),
                    rep(1, d_treated),
                    rep(0, d_no_treated),
                    rep(1, p_treated),
                    rep(0, p_no_treated))
    ][ , baseline:= c(rep(1, base),
                    rep(0, N - base))
    ][ , treatment:= c(rep(0, base),
                    rep(1, d_treated + d_no_treated),
                    rep(0, p_treated + p_no_treated))
    ][ , placebo:= c(rep(0, N - p_treated - p_no_treated),
                    rep(1, p_treated + p_no_treated))
    ][ , type:= c(rep("Baseline", base),
                    rep("Treatment", d_treated + d_no_treated),
                    rep("Placebo", p_treated + p_no_treated))
    ]

# reorder by ID
d4 <- d4[order(id),]

# display the table
d4 %>%
  group_by(type, treated) %>%
  summarise(count = n(), turnout_rate = mean(turnout))
```

```
## # A tibble: 5 x 4
## # Groups:   type [?]
##   type       treated count turnout_rate
##   <chr>        <dbl> <int>        <dbl>
## 1 Baseline         0  2572        0.312
## 2 Placebo          0  2109        0.321
## 3 Placebo          1   470        0.298
## 4 Treatment        0  2086        0.327
## 5 Treatment        1   486        0.391
```

a. Estimate the proportion of Compliers by using the data on the Treatment group. Then compute a second estimate of the proportion of Compliers by using the data on the Placebo group. Are these sample proportions statistically significantly different from each other? Explain why you would not expect them to be different, given the experimental design. (Hint: ITT_D means "the average effect of the treatment on the dosage of the treatment." I.E., it's the contact rate α in the async).

```
# proportion of compliers
treatment_prop_compliers <- d_treated / (d_treated + d_no_treated)
placebo_prop_compliers <- p_treated / (p_treated + p_no_treated)

sprintf("Proportion of treatment compliers is %.3f", treatment_prop_compliers)
```

```
## [1] "Proportion of treatment compliers is 0.189"
```

```r
sprintf("Proportion of placebo compliers is %.3f", placebo_prop_compliers)
```

```
## [1] "Proportion of placebo compliers is 0.182"
## statistical test using t-stat
```

```r
# first extract two sets of interest -- treatment vs placebo
d4_treatment <- d4[type=="Treatment", ]
d4_placebo <- d4[type=="Placebo", ]

# t-test on difference in "treated"
t.test(d4_treatment$treated, d4_placebo$treated)
```

```
##
##  Welch Two Sample t-test
##
## data:  d4_treatment$treated and d4_placebo$treated
## t = 0.61987, df = 5147.6, p-value = 0.5354
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01452611  0.02795977
## sample estimates:
## mean of x mean of y
## 0.1889580 0.1822412
```

ANSWER: the proportions of compliers between the treatment and the placebo groups are not statistically significant given the t-statistic of 0.62 and p-value of 0.5354.

   b. Do the data suggest that Never Takers in the treatment and placebo groups have the same rate of turnout? Is this comparison informative?

ANSWER: the data suggests that Never Takers in the treatment and placebo groups seem to have the same rate of turnout, 32.15% vs 32.74%. They are not practically significant. This comparison is informative because the same turnout rates suggest that the research design (including the way in which subjects are assigned, treated - "non-interference") does not influence the behavior of the assigned groups that did not end up receiving any treatment. Had the turnout rates differed, then it should be investigated as to the cause of the difference.

   c. Estimate the CACE of receiving the placebo. Is this estimate consistent with the substantive assumption that the placebo has no effect on turnout?

```r
## calculate ITT
placebo_ITT <- data.table(d4 %>% filter(type=="Placebo") %>%
                            summarise(turnout_rate = mean(turnout)) -
                            d4 %>% filter(type=="Baseline", treated==0) %>%
                            summarise(turnout_rate = mean(turnout)))

## calculate proportion of compliers & CACE
placebo_ITT_d <- data.table(d4 %>% filter(type=="Placebo") %>%
                              summarise(complier = mean(treated)))
sprintf("The placebo CACE is %.3f", placebo_ITT/placebo_ITT_d)
```

```
## [1] "The placebo CACE is 0.027"
```
```r
# use 2SLS to determine significance of CACE

mod4c_first <- d4[type!="Treatment", lm(treated ~ type)]
mod4c_second <- d4[type!="Treatment", lm(turnout ~ predict(mod4c_first))]

summary(mod4c_second)
```
```
##
## Call:
## lm(formula = turnout ~ predict(mod4c_first))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3172 -0.3172 -0.3122  0.6828  0.6878
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.312208   0.009159  34.089   <2e-16 ***
## predict(mod4c_first) 0.027265   0.071024   0.384    0.701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4645 on 5149 degrees of freedom
## Multiple R-squared:  2.862e-05,  Adjusted R-squared:  -0.0001656
## F-statistic: 0.1474 on 1 and 5149 DF,  p-value: 0.7011
```

ANSWER: The placebo CACE is 2.7% with SE 0.07 and p-value of 0.7. Based on this information, the placebo doesn't have a statistical significance on the turnout rate.

    d. Estimate the CACE of receiving the treatment using two different methods. First, use the conventional method of dividing the ITT by the ITT_{D}. (This should be a treatment vs. control comparison.)

```r
## calculate ITT
treatment_ITT <- data.table(d4 %>% filter(type=="Treatment") %>%
                            summarise(turnout_rate = mean(turnout)) -
            d4 %>% filter(type=="Baseline", treated==0) %>%
                summarise(turnout_rate = mean(turnout)))

## calculate proportion of compliers & CACE
treatment_ITT_d <- data.table(d4 %>% filter(type=="Treatment") %>%
                            summarise(complier = mean(treated)))

sprintf("The treatment CACE is %.3f", treatment_ITT/treatment_ITT_d)
```
```
## [1] "The treatment CACE is 0.144"
```
```r
# use 2SLS to determine significance of treatment CACE

mod4d_first <- d4[type!="Placebo", lm(treated ~ type)]
mod4d_second <- d4[type!="Placebo", lm(turnout ~ predict(mod4d_first))]

summary(mod4d_second)
```
```
##
```

```
## Call:
## lm(formula = turnout ~ predict(mod4d_first))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3394 -0.3394 -0.3122  0.6606  0.6878
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.312208   0.009239  33.791   <2e-16 ***
## predict(mod4d_first) 0.144033   0.069150   2.083   0.0373 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4686 on 5142 degrees of freedom
## Multiple R-squared:  0.000843,   Adjusted R-squared:  0.0006487
## F-statistic: 4.339 on 1 and 5142 DF,  p-value: 0.03731
```

e. Then, second, compare the turnout rates among the Compliers in both the treatment and placebo groups. Interpret the results.

```
## CACE based on direct comparison between Treatment and Placebo
CACE_direct <- data.table(d4 %>% filter(type=="Treatment", treated==1) %>%
                            summarise(turnout_rate = mean(turnout)) -
              d4 %>% filter(type=="Placebo", treated==1) %>%
                summarise(turnout_rate = mean(turnout)))

sprintf("The treatment CACE is %.3f", CACE_direct)
```

```
## [1] "The treatment CACE is 0.093"
```

```
# use a linear regression model excluding "Baseline", so the slop coefficient would reflect
# the difference between the two -- Placebo vs Treatment
mod4e <- d4[type!="Baseline" & treated==1, lm(turnout ~ type)]

summary(mod4e)
```

```
##
## Call:
## lm(formula = turnout ~ type)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3910 -0.3910 -0.2979  0.6090  0.7021
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.29787    0.02185  13.634  < 2e-16 ***
## typeTreatment  0.09307    0.03064   3.037  0.00245 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4736 on 954 degrees of freedom
## Multiple R-squared:  0.009579,   Adjusted R-squared:  0.00854
## F-statistic: 9.226 on 1 and 954 DF,  p-value: 0.00245
```

f. Based on what we talked about in class – that the rate of compliance determines whether one or another

design is more efficient – given the compliance rate in this study, which design *should* provide a more efficient estimate of the treatment effect? If you want to review the specific paper that makes this claim, check out this link. Does it?

ANSWER: given a low compliace rate of about 19%, the placebo design would provide a more efficient estimate of the treatment effect. As shown in the results from (d) and (e) the standard error from the placebo design (i.e., 0.03064) is less than that of the conventional design (0.06915).

5. Tetris FTW?

A doctoral student conducted an experiment in which she randomly varied whether she ran or walked 40 minutes each morning. In the middle of the afternoon over a period of 26 days she measured the following outcome variables: (1) her weight; (2) her score in Tetris; (3) her mood on a 0-5 scale; (4) her energy; and (5) whether she got a question right on the math GRE.

```
d5 <- read.dta("./data/Hough_WorkingPaper_2010.dta")
d5 <- data.table(d5)
head(d5)
```

```
##    day run weight tetris mood energy appetite gre
## 1:   1   1     21  11092    3      3        0   1
## 2:   2   1     21  14745    3      1        2   0
## 3:   3   0     20  11558    3      3        0   1
## 4:   4   0     21  11747    3      1        1   1
## 5:   5   0     21  14319    2      3        3   1
## 6:   6   1     19   7126    3      2        0   1
```

a. Suppose you were seeking to estimate the average effect of running on her Tetris score. Explain the assumptions needed to identify this causal effect based on this within-subjects design. Are these assumptions plausible in this case? What special concerns arise due to the fact that the subject was conducting the study, undergoing the treatments, and measuring her own outcomes?

ANSWER: In this study the control condition is "walk" (i.e., not run) and the treatment is "run". The fundamental assumption that is needed is whether the decision between the "walk" and the "run" is randomly assigned each day. In a particular day, if she doesn't know ahead of time whether she would run or walk (i.e., flip a coin the moment before she starts taking one of the two activities), then this assumption of random assignment is met.

Now there are there are two additional assumption in order to define ATE. First, "No Anticipation" assumption which says that potential outcomes are unaffected by treatments that are administered in the future. Second, "No Persistence" assumption which requires that potential outcomes in one period are unaffected by treatments administered in prior periods.

Looking at "No Anticipation" assumption, she would have met this assumption if she wouldn't know what her future administration would be (i.e., a coin flip). Looking at "No Persistence" assumption, it is also likewise unlikely that there is any violation to this assumption given that the administration (i.e., decision between "walk" and "run") is random in nature. It may be argued that if she is administered with the treatment consecutively many days in a row, then there may be a persistence effect spilled over from one day to another; however,

this concern is overcome by the randomization of the administration. For example, if she is to be administered with "walk" consecutively many days in a row, in effect that would be a "washout period" before a future "run" is to occur; and by following this randomization procedure over time, the spillover effect should be mitigated (i.e., cancelled out).

b. Estimate the effect of running today on Tetris score. What is the ATE?

```
head(d5)
```

```
##     day run weight tetris mood energy appetite gre
## 1:    1   1     21  11092    3      3        0   1
## 2:    2   1     21  14745    3      1        2   0
## 3:    3   0     20  11558    3      3        0   1
## 4:    4   0     21  11747    3      1        1   1
## 5:    5   0     21  14319    2      3        3   1
## 6:    6   1     19   7126    3      2        0   1
```

```
ate_tetris <- d5[run==1, .(mean_tetris = mean(tetris, na.rm = T))] -
  d5[run==0, .(mean_tetris = mean(tetris, na.rm=T))]
```

```
sprintf("The effect of running today on Tetris score is %.2f", ate_tetris)
```

```
## [1] "The effect of running today on Tetris score is 13613.10"
```

c. One way to lend credibility to with-subjects results is to verify the no-anticipation assumption. Construct a regression using the variable `run` to predict the `tetris` score *on the preceding day*. Presume that the randomization is fixed. Why is this a test of the no-anticipation assumption? Does a test for no-anticipation confirm this assumption?

```
# create a lag variable
d5 <- d5 %>% mutate(tetris_lag = lag(tetris, 1))

# fit a linear model
mod5c.omit <- lm(data = d5, tetris_lag ~ run, na.action = na.omit)
mod5c.coef <- coeftest(mod5c.omit, vcovHC(mod5c.omit))

## use stargazer to print formatted tables
stargazer(mod5c.omit, type = "text",
          omit.stat=c("LL","ser","f","adj.rsq","rsq"),
          se = list(mod5c.coef[3:4]),
          align = TRUE, no.space=TRUE,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## ============================================
##                     Dependent variable:
##                 ----------------------------
##                            tetris_lag
## ------------------------------------------
## run                        645.621
##                           (5,118.211)
## Constant                18,903.830***
##                           (2,964.078)
## ------------------------------------------
## Observations                  23
## ============================================
## Note:           *p<0.05; **p<0.01; ***p<0.001
```

ANSWER: based on the regression outcome, there appears to be a violation on the "No Anticipation" assumption because the coefficient of the run on 1-day lag tetris doesn't show a statistical significance. This means that (i) either there is no whatsoever anticipation of the upcoming activity, hence no impact at all on on the tetris scoring or (ii) even if there is an anticipation that a run would happen the next day, that anticipation doesn't have an impact on tetris scoring.

    d. Now let's use regression to put a standard error on our ATE estimate from part (b). Regress Tetris score on the the variable `run`, this time using the current rather than the future value of `run`. Is the impact on Tetris score statistically significant?

```
# fit a linear model
mod5d.omit <- lm(data = d5, tetris ~ run, na.action = na.omit)
mod5d.coef <- coeftest(mod5d.omit, vcovHC(mod5d.omit))

## use stargazer to print formatted tables
stargazer(mod5c.omit, mod5d.omit, type = "text",
          omit.stat=c("LL","ser","f","adj.rsq","rsq"),
          se = list(mod5c.coef[3:4], mod5d.coef[3:4]),
          align = TRUE, no.space=TRUE,
          star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## =============================================
##                     Dependent variable:
##               -------------------------------
##                  tetris_lag         tetris
##                     (1)              (2)
## -------------------------------------------
## run               645.621       13,613.100**
##                  (5,118.211)    (4,376.427)
## Constant        18,903.830***  12,806.400***
##                  (2,964.078)    (1,549.003)
## -------------------------------------------
## Observations         23              24
## =============================================
## Note:           *p<0.05; **p<0.01; ***p<0.001
```

ANSWER: The impact of run on the tetris scoring is now statistically significant at the 1% significant level. The treatment effect of run on tetris scoring is both statsitically significant and practical.

    e. If Tetris responds to exercise, one might suppose that energy levels and GRE scores would as well. Are these hypotheses borne out by the data?

```
# let's fit a linear model to test this assumption.
mod5e_gre.omit <- lm(data = d5, gre ~ run, na.action = na.omit)
mod5e_gre.coef <- coeftest(mod5e_gre.omit, vcovHC(mod5e_gre.omit))

mod5e_energy.omit <- lm(data = d5, energy ~ run, na.action = na.omit)
mod5e_energy.coef <- coeftest(mod5e_energy.omit, vcovHC(mod5e_energy.omit))

## use stargazer to print formatted tables
stargazer(mod5e_gre.omit, mod5e_energy.omit, type = "text",
          omit.stat=c("LL","ser","f","adj.rsq","rsq"),
          se = list(mod5e_gre.coef[3:4], mod5e_energy.coef[3:4]),
```

```
        align = TRUE, no.space=TRUE,
        star.cutoffs = c(0.05, 0.01, 0.001))
```

```
## 
## =========================================
##                  Dependent variable:
## 
##                 -------------------------------
##                   gre            energy
##                   (1)             (2)
## ---------------------------------------------
## run             -0.175          0.071
##                 (0.188)         (0.446)
## Constant        0.818***        3.000***
##                 (0.128)         (0.314)
## ---------------------------------------------
## Observations      25              24
## =========================================
## Note:           *p<0.05; **p<0.01; ***p<0.001
```

ANSWER: From the regression results, the run activity does not have a statistically significant impact on either the GRE scoring or the energy level. If there is anything at all, the GRE score tends lower a bit, while energy tilts up a bit. But neither appears to have a practical significant.

f. Suppose the student decides to publish her results on Tetris, since she finds those most interesting. In the paper she writes, she chooses to be concise by ignoring the data she collected on energy levels and GRE scores, since she finds those results less interesting. How might you criticize the student's decision? What trap may she have fallen into?

ANSWER: Her practice sounds very much like an expedition and fish for a variable that does result in a statsitical significace. It is apparently a p-value hacking technique, in which her significant outcome might be just a result by chance. Determining and stating the variables of interest upfront would lessen this concern.

g. After submitting her paper to a journal, the student thinks of another hypothesis. What if running has a relatively long-lasting effect on Tetris scores? Perhaps both today's running and yesterday's running will affect Tetris scores. Run a regression of today's Tetris score on both today's `run` variable and yesterday's `run` variable. How does your coefficient on running today compare with what you found in part (d)? How do you interpret this comparison?

```
## fit a linear model for both run and lag-run

# create a lag variable for run
d5 <- d5 %>% mutate(run_lag = lag(run, 1))

mod5g.omit <- lm(data = d5, tetris ~ run + run_lag, na.action = na.omit)
mod5g.coef <- coeftest(mod5g.omit, vcovHC(mod5g.omit))

## use stargazer to print formatted tables
stargazer(mod5c.omit, mod5d.omit, mod5g.omit, type = "text",
        omit.stat=c("LL","ser","f","adj.rsq","rsq"),
        se = list(mod5c.coef[3:4], mod5d.coef[3:4], mod5g.coef[4:6]),
        align = TRUE, no.space=TRUE,
        star.cutoffs = c(0.05, 0.01, 0.001))
```

```
## 
## =========================================================
## 		             Dependent variable:
## 		      ---------------------------------------
## 		       tetris_lag            tetris
## 		          (1)           (2)           (3)
## ---------------------------------------------------------
## run               645.621    13,613.100**  15,025.950**
## 		       (5,118.211)   (4,376.427)   (4,827.689)
## run_lag                                     1,688.645
## 		                                   (5,339.613)
## Constant     18,903.830*** 12,806.400*** 11,793.210**
## 		       (2,964.078)   (1,549.003)   (3,760.861)
## ---------------------------------------------------------
## Observations        23            24            23
## =========================================================
## Note:                        *p<0.05; **p<0.01; ***p<0.001
```

ANSWER: Based on the regression outcomes, it appears that tetris scoring is more influenced by whether she ran on the same day (i.e., the coefficient on the run variable is statistically significant at 1% level). Whether or not she ran the day before doesn't seem to statistically influence her tetris scoring on the current day. What this tells us is that there is not much evidence of the "long-lasting" effect. It also suggests that the "No Persistence" assumption is held.

h. (optional) Note that the observations in our regression are not necessarily independent of each other. An individual might have serially correlated outcomes, regardless of treatment. For example, I might find that my mood is better on weekends than on weekdays, or I might find that I'm terrible at playing Tetris in the few days before a paper is due, but I get better at the game once my stress level has lowered. In computing standard errors for a regression, OLS assumes that the observations are all independent of each other. If they are positively serially correlated, it's possible that OLS will underestimate the standard errors.

To check this, let's do randomization inference in the regression context. Recall that the idea of randomization inference is that under the sharp null hypothesis, we can re-randomize, recompute the ATE, and get approximately the right answer (zero) for the treatment effect. So, returning to the regression we ran in part (g), please generate 1000 new randomizations of the run variable, use those to replace the current and lagged values of run in your dataset, then run the regression again. Record the coefficient you get on the contemporaneous value of run, and repeat this re-randomization exercise 1000 times. Plot the distribution of beta. What are the 2.5% and 97.5% quantiles? How do they compare with the width of the 95% confidence interval you got for your main run coefficient in the regression in part (g)?

```
# create function for randomization

ri5 <- function(){
  # randomly assign run

  d5 <- data.table(d5)

  d5 <- d5[ , run:= sample(run, replace = F)]

  # calculate ATE
  as.numeric(d5[run==1, .(mean_tetris = mean(tetris, na.rm = T))] -
             d5[run==0, .(mean_tetris = mean(tetris, na.rm=T))])
```
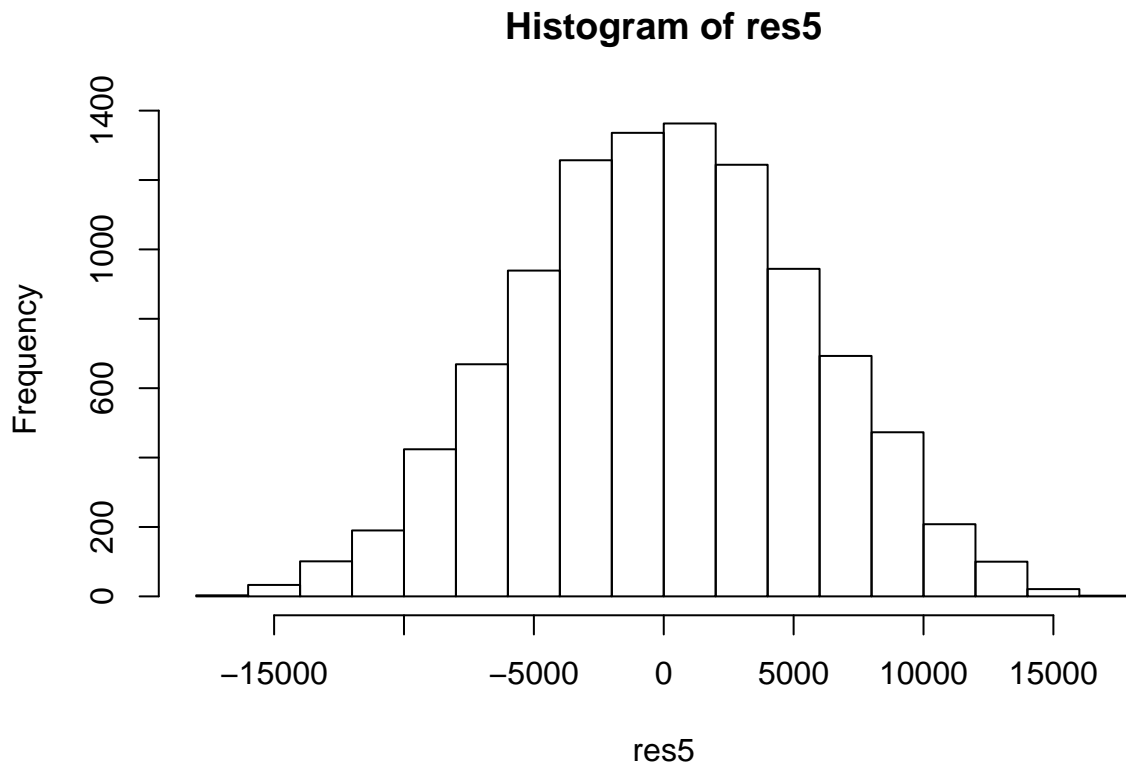
```
}

# simulate ATE
res5 <- replicate(10000, ri5())

# plost histogram of ITT
hist(res5)
```

Histogram of res5


```
sprintf("The 2.5 pct and 97.5 pct quantiles from randomization inference are %.2f and %2.f respectively
        quantile(res5, probs = c(0.025, 0.975))[1],
        quantile(res5, probs = c(0.025, 0.975))[2])
```

[1] "The 2.5 pct and 97.5 pct quantiles from randomization inference are -10622.64 and 10654 respecti

```
sprintf("The 2.5 pct and 97.5 pct quantiles from the regression model are %.2f and %2.f respectively",
        confint(mod5d.omit)[2 ,][1], confint(mod5d.omit)[2 ,][2])
```

[1] "The 2.5 pct and 97.5 pct quantiles from the regression model are 3543.15 and 23683 respectively

ANSWER: The outcome of the randomization inference suggests that the impact of run on tetris scoring from the study does have a statistical significance as it is very unlikely that the effect would be as high as 13,600 as found in the study. Based on the histogram most ATEs fall somewhere around zero, and the interval ranges from negative 10,000 to positive 10,000 approximately with the shape that is almost symmetrical in shape. The ETA found from the study deviates from this histogram significantly, and thus we drew a conclusion that it was statistically significant. The linear regression outcomes shows statistical significance of the run treatment on tetris scoring with its confidence interval of 95% covering the positive effects [3543.15, 23683].