

# Problem Set 4

*Pavan Kurapati*

*July 17, 2018*

```
# load packages
library(foreign)
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

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

library(sandwich)
```

## FE exercise 5.2

- 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.

```
Observation = seq(1:6)
Y0 = c(6,5,7,4,3,1) # Control Outcome
Y1 = c(4,3,5,7,6,4) # Treatment Outcome
dz0 = c(rep(0,times=6)) # d(z=0). We assume no treatment in control group
dz1 = c(1,1,1,0,0,0) # d(z=1). Compliers will get treatment. Never-Takers won't
comply = c(rep("Complier",times=3),rep("NeverTaker",times=3))
df = data.frame(Y0,Y1,dz0,dz1,comply)
df
```

```
##      Y0 Y1 dz0 dz1      comply
## 1  6  4  0  1  Complier
## 2  5  3  0  1  Complier
## 3  7  5  0  1  Complier
## 4  4  7  0  0 NeverTaker
## 5  3  6  0  0 NeverTaker
## 6  1  4  0  0 NeverTaker
```

```
#ATE is the simple mean of Treatment and Control
ATE = mean(df$Y1) - mean(df$Y0)
#CACE is the average treatment among compliers
df_s = subset(df,df$dz1==1)
CACE = mean(df_s$Y1) - mean(df_s$Y0)
paste("ATE is ",ATE)
```

```
## [1] "ATE is 0.5"
```

```
paste("CACE is ", CACE)
```

```
## [1] "CACE is -2"
```

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

CACE is informative when all we care about is the impact of actual treatment received. Take for example a case where a new pill is designed to lower blood sugar levels for type-2 diabetes. I will only be interested in those who take the pill than those who are assigned for treatment. However, a CACE with placebo is even better in this case.

For the hypothetical results I presented in the above case, the CACE is misleading because the treatment effect is falsely determined as negative for CACE. Never takers include the control group who never got the treatment. For example, the experiment in “a” is about the effect of motivational speech on employee performance. If we do incorrect blocking or randomizing, we may end up picking the set of people for treatment who under-perform anyway and ignore the ones who would perform better if they were part of treatment. In this case, ATE gives us a better estimate. So the treatment effect is grossly underestimated with CACE

- c. In addition, please also answer this question: Which population is more relevant to study for future decision making: the set of Compliers, or the set of Compliers plus Never-Takers? Why?

It depends on the nature of the study. CACE estimates are specific to compliers. For experiments such as impact of showing advertisements on sales, or for the impact of epipens or other medicines, we may only care about compliers and not about never-takers.

However, we may want the information from both compliers plus never-takers in cases such as voter turnout with phone calls. As explained in the async example for voter turnout rates (Gerber and Green 2005), it would be a mistake to have just taken a CACE based on compliers in this case. Ignoring never-takers would have grossly over-estimated the turnout effect with treatment of phone calls. A placebo experiment with a canvasser asking for “blood donation” will help in determining the potential compliers in control group. This will greatly improve the precision of the result.

## FE exercise 5.6

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?

If 500 subjects were believed to be contacted, then the alpha ( $ITT_D$ ) is 0.5.

Treatment Average = 0.4 (400/1000)

Control Average = 0.35 (700/2000)

ATE = 0.4 - 0.35 = 0.05

$$CACE = \frac{ITT}{ITT_D} = \frac{0.05}{0.5} = 0.1$$

The CACE estimate is 0.1 or 10%

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

If 250 subjects were believed to be contacted, then the alpha ( $ITT_D$ ) is 0.25

$$CACE = \frac{0.05}{0.25} = 0.2$$

**The CACE estimate is 0.2 or 20%**

- 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?

**The efforts appear less effective here because they only contacted 250 people and the person analyzing the result considered 500, which gave a lesser CACE. From CACE, it is evident that the voter turn out increases by 20% when contacted. Here, the effectiveness indicate the actual treatment effect and not the intent to treat. The definition matters because without the right measurement, the decision to fund future canvassing would have been stopped due to lower estimated turnout than what is actual.**

## FE exercise 5.10

Guan and Green report the results of a canvassing experiment conducted 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.

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

```
##   turnout contact  dormid treat2
## 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.*

```
d = na.omit(d)
d_t = subset(d,d$treat2==1)
c_t = subset(d,d$treat2==0)
T_mean = mean(d_t$turnout)
C_mean = mean(c_t$turnout)
ITT = T_mean - C_mean

paste("ITT with difference of two means is ",ITT)
```

```
## [1] "ITT with difference of two means is 0.131929570928821"
```

```
model1 = lm(turnout~treat2,data=d)
ITT_lm = summary(model1)$coef[2,1]
paste("ITT with linear regression model is ",ITT_lm)
```

```
## [1] "ITT with linear regression model is  0.131929570928821"
```

**ITT with both methods are same and is 0.13192**

- 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.

```
# Get the number of dorms
n_dorms = length(unique(d$dormid))
n = nrow(d)
treat = vector()
# Create a cluster randomization function. We have clusters of different sizes.
# So, first create a random assignment to clusters
# Then, we loop with each element and make sure every element inside a cluster receives treatment/control

randomize <- function() {
  # Randomize clusters to treatment or control
  # We need to maintain same ratio. Treatment is 2/3 and control is 1/3
  dorms_t = round(n_dorms*2/3,0)
  dorms_c = n_dorms-dorms_t
  s = sample(c(rep(0,dorms_c),rep(1,dorms_t)))
  j = 1
  prev = ""
  # Loop through each element and assign treatment
  for (i in 1:n) {
    if (i == 1) {
      treat = append(treat,s[j])
    } else if (prev == d$dormid[i]) {
      # This element belongs to same cluster as previous
      treat = append(treat,s[j])
    } else {
      # This belongs to different cluster
      j = j + 1
      treat = append(treat,s[j])
    }
    prev = d$dormid[i]
  }
  return(treat)
}

treatment_vector = randomize()
```

```
# Randomization Inference
```

```
#Assign our outcomes
```

```
outcomes <- d$turnout * treatment_vector + d$turnout*(1-treatment_vector)
```

```
# Function for ate estimation based on treatment and outcomes
```

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

```

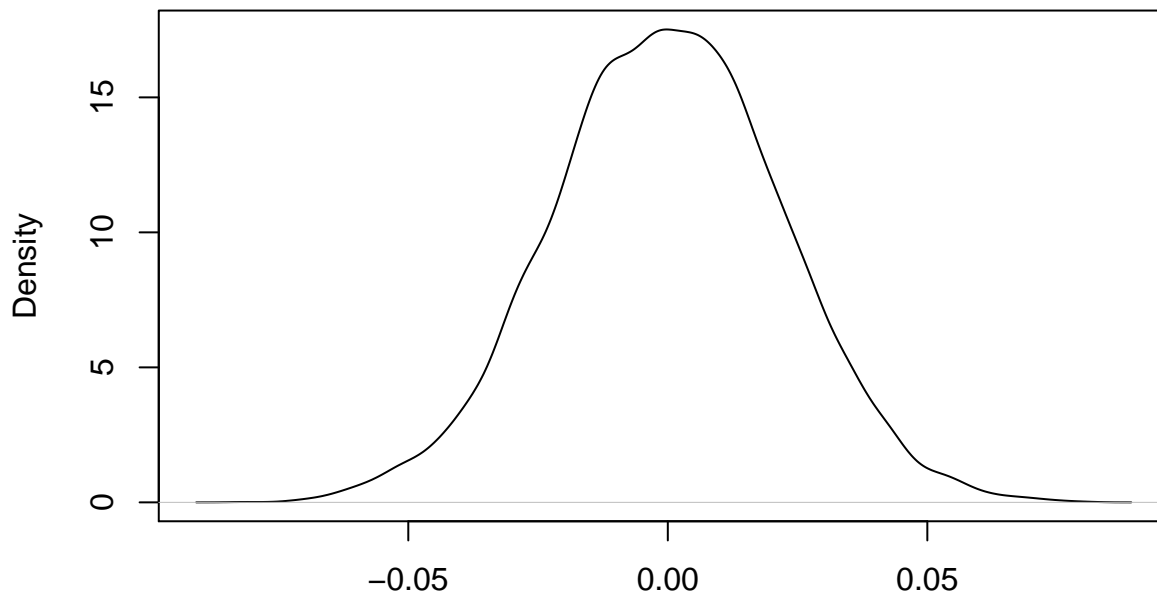
}

# Conduct 10000 simulated random estimations

sharp_null_distribution <- replicate(10000, ate_estimation(outcomes,
plot(density(sharp_null_distribution),main="Distribution of ATE under sharp NULL")
abline(v=ITT)

```

## Distribution of ATE under sharp NULL



N = 10000 Bandwidth = 0.003147

```

n_g = sum(sharp_null_distribution >= ITT)
p_value = n_g / length(sharp_null_distribution)
paste("The p-value observed Sharp Null Hypothesis Test is:", p_value)

```

```
## [1] "The p-value observed Sharp Null Hypothesis Test is: 0"
```

Sharp null hypothesis can be rejected based on the observed p-value. Even with using linear regression in part a, we got an extremely low p-value for treatment variable. From randomization inference, we can reject the hypothesis that there was no treatment effect.

- 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.

```

#Create a subset of those who were contacted (actually treated)
data_treated = subset(d, d$contact == 1)
alpha = nrow(data_treated) / nrow(d_t)
CACE = ITT / alpha
paste("CACE using means is ", CACE)

```

```
## [1] "CACE using means is 0.14894022959121"
```

```
# Now let us use a 2 Stage Least Square.
# First let us regress the treatment variable 'contact' with
# our Instrumental Variable 'treat2' (Assigned to Treatment)
```

```
model_stage1 = lm(contact~treat2,data=d)
contact_hat = fitted.values(model_stage1)
```

```
# Now run the 2nd stage with contact_hat
model_cace = lm(d$turnout~contact_hat)
cace_2sls = summary(model_cace)$coefficients[2,1]
paste("CACE with 2SLS is ",cace_2sls)
```

```
## [1] "CACE with 2SLS is 0.148940229591212"
```

```
summary(model_cace)$coefficients
```

```
##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 0.6686657 0.01162463  57.521479 0.000000e+00
## contact_hat 0.1489402 0.01605296   9.278051 2.747624e-20
```

```
confint(model_cace,level=0.95)
```

```
##              2.5 %    97.5 %
## (Intercept) 0.6458750 0.6914564
## contact_hat 0.1174675 0.1804129
```

ANSWER: CACE with both the methods is 0.1489402. The standard error coefficient with 2SLS is 0.016053. The 95% confidence interval of CACE overlaps with ITT. This means that CACE and ITT might be close to being same. Since the 95% CI of CACE overlaps with ITT, it implies that leaving a leaflet to the treatment group must have had an impact equivalent to having actually contacted. Based on that, I think leaving leaflet does seem to have an impact and caused indirect treatment effect.

- d. *SKIP*
- e. *SKIP*
- f. *SKIP*

## FE exercise 5.11

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 as 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.)

```
# Prepare the data frame

#Baseline
Treatment_Assignment <- rep("Baseline",2572)
Treated <- rep("No",2572)
Turnout <- c(rep(1,round(0.3122*2572,0)),rep(0,2572-(round(0.3122*2572,0))))
df_baseline <- data.frame(Treatment_Assignment,Treated,Turnout)

# Compliers
Treatment_Assignment <- rep("Treatment",486)
Treated <- rep("Yes",486)
Turnout <- c(rep(1,round(0.3909*486,0)),rep(0,486-(round(0.3909*486,0))))
df_Treatment_comp <- data.frame(Treatment_Assignment,Treated,Turnout)

# Treatment group with never takers
Treatment_Assignment <- rep("Treatment",2086)
Treated <- rep("No",2086)
Turnout <- c(rep(1,round(0.3274*2086,0)),rep(0,2086-(round(0.3274*2086,0))))
df_Treatment_nt <- data.frame(Treatment_Assignment,Treated,Turnout)

# Placebo treated
Treatment_Assignment <- rep("Placebo",470)
Treated <- rep("Yes",470)
Turnout <- c(rep(1,round(0.2979*470,0)),rep(0,470-(round(0.2979*470,0))))
df_Placebo_t <- data.frame(Treatment_Assignment,Treated,Turnout)

# Placebo untreated
Treatment_Assignment <- rep("Placebo",2109)
Treated <- rep("No",2109)
Turnout <- c(rep(1,round(0.3215*2109,0)),rep(0,2109-(round(0.3215*2109,0))))
df_Placebo_nt <- data.frame(Treatment_Assignment,Treated,Turnout)

df = rbind(df_baseline,df_Treatment_comp,df_Treatment_nt,df_Placebo_t,df_Placebo_nt)
#nrow(df)
```

- a. We are rewriting part (a) as follows: “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  $\alpha$  in the async).

```
treated_sum = sum(df$Treatment_Assignment=="Treatment" & df$Treated=="Yes")
untreated_sum = sum(df$Treatment_Assignment=="Treatment" & df$Treated=="No")
assigned_treatment = sum(df$Treatment_Assignment=="Treatment")

compliers_prop = treated_sum/assigned_treatment
paste("Proportion of Compliers are ",compliers_prop)

## [1] "Proportion of Compliers are 0.18895800933126"

placebo_treated = sum(df$Treatment_Assignment=="Placebo" & df$Treated=="Yes")
placebo_untreated = sum(df$Treatment_Assignment=="Placebo" & df$Treated=="No")
```

```

placebo_assigned = sum(df$Treatment_Assignment=="Placebo")
compliers_prop_placebo = (placebo_treated)/placebo_assigned
paste("Proportion of Compliers in Placebo are ",compliers_prop_placebo)

## [1] "Proportion of Compliers in Placebo are  0.182241178751454"
# Check for statistical significance of proportions

df_t_p <- matrix(c(treated_sum,placebo_treated,untreated_sum,placebo_untreated),ncol=2)
colnames(df_t_p) <- c('Treated','Untreated')
rownames(df_t_p) <- c('Treatment','Placebo')

df_t_p

##           Treated Untreated
## Treatment      486      2086
## Placebo        470      2109

prop.test(df_t_p)

##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  df_t_p
## X-squared = 0.34122, df = 1, p-value = 0.5591
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.01490526  0.02833893
## sample estimates:
##      prop 1      prop 2
## 0.1889580 0.1822412

```

Proportion of compliers is 0.188958. Proportion of copliers with placebo is 0.1822412. They are not very different from each other. I would not expect them to be different. From the prop test, the hypothesis is that they are same. There is no statistical significance to reject the NULL hypothesis. The confidence interval indicates the range of difference in which 0 overlaps. Hence, I am more inclined to think that the two proportions are same.

Conventional design is preferable only when compliers make up at least 1/2 of the sample size. In this experiment, the compliers are less than 1/2. So a placebo design is preferable. In this experiment, the groups are equally divided into Baseline, Treatment and Placebo. I would expect the same ratio of compliers between the treatment and placebo groups in a good experimental design.

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

Yes, the data suggests that never takers in the treatment and placebo groups have similar rate of turnout. It is informative because it proves one assumption that never taker outcomes are same. i.e  $Y(d(1)=0) = Y(d(0)=0)$

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

```

placebo_untreated = sum(df$Treatment_Assignment=="Placebo" & df$Treated=="No")
placebo_total = placebo_untreated + placebo_treated
placebo_ittd = placebo_treated/placebo_total

```



```

placebo_mean = (placebo_treated/placebo_total)*0.2979 + (placebo_untreated/placebo_total)*0.3215
control_mean = 0.3122
ITT_P = placebo_mean - control_mean
CACE_Placebo = ITT_P/placebo_ittd
paste("CACE of receiving the placebo is ",CACE_Placebo)

```

```
## [1] "CACE of receiving the placebo is 0.0274312765957448"
```

**CACE of receiving the placebo is 0.0274313. Placebo does seem to have an impact on turnout, albeit it is very small compared to the actual treatment effect. I would say “No”, it is not consistent with the substantive assumption that placebo has no effect on turnout.**

- d. Estimate the CACE of receiving the treatment using two different methods. First, use the conventional method of dividing the ITT by the ITT<sub>{D}</sub>.

```

treatment_compliers = sum(df$Treatment_Assignment=="Treatment" & df$Treated=="Yes")
treatment_total = sum(df$Treatment_Assignment=="Treatment")
treatment_nt = treatment_total - treatment_compliers
ittd = treatment_compliers/treatment_total

treatment_mean = (treatment_compliers/treatment_total)*0.3909 + (treatment_nt/treatment_total)*0.3274
ITT = treatment_mean - control_mean
CACE_Treatment = ITT/ittd
paste("CACE of receiving the treatment is ",CACE_Treatment)

```

```
## [1] "CACE of receiving the treatment is 0.143941152263375"
```

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

```

placebo_complier_mean = 0.2979
treatment_complier_mean = 0.3909
CACE = treatment_complier_mean - placebo_complier_mean
paste("CACE of receiving the treatment is ",CACE)

```

```
## [1] "CACE of receiving the treatment is 0.093"
```

**CACE of receiving the treatment is 0.1439412. This would imply that the treatment of contacting has an impact of 14% increase in turnout. However, the true treatment effect is between the compliers of treatment and placebo groups which brings the actual treatment effect down to 9.3% and this result has better precision.**

## FE exercise 8.10

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.

```

d <- read.dta("./data/Hough_WorkingPaper_2010.dta")
head(d)

```

```

##   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?

Two main assumptions are needed for within-subject designs. No-anticipation assumption (A treatment in future has no impact on the current potential outcome) and No-Persistence assumption (A treatment given in the past has no impact on the current potential outcome). In this case, the treatment is “Running in the morning”. In order to measure the causal effect, we need to assume that the tetris score being measured today has no impact on the preparations that goes towards running in the next period (next day) which sounds like a reasonable assumption. However, we also need to assume that the treatment of running the previous day has no impact on outcome today. This assumes that the washout period of 1 day is sufficient. What if she is tired of running the previous day and that has an impact on tetris score, and energy levels?

Another concern is that tetris score might be related to many other factors that coincides with treatment. Any personal reasons can cause the score to be high or low on a given day irrespective of the treatment. Also, since the same person is assigning the treatments and recording the scores, the experiment might be highly biased. The experimenter might knowingly/unknowingly want to make treatment work.

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

```
d_n = na.omit(d)
d_t = subset(d_n,d_n$run==1)
d_c = subset(d_n,d_n$run==0)
ate_run = mean(d_t$tetris) - mean(d_c$tetris)
paste("Effect of running today on tetris score is ",ate_run)
```

```
## [1] "Effect of running today on tetris score is 13613.1"
```

- 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?

```
# Let us create a new data frame and add a column of preceding day score.
# We will copy previous day's score. Initialize day 1 with 0
```

```
d_mod = d
d_mod$preceding_tetris = 0
for (i in 1:nrow(d)) {
  if (i==1) {
    d_mod$preceding_tetris[i] = 0
  } else {
    d_mod$preceding_tetris[i] = d$tetris[i-1]
  }
}
# Remove day 1 as we do not have preceding day for day 1
d_mod <- d_mod[-c(1),]
d_mod
```

```
##    day run weight tetris mood energy appetite gre preceding_tetris
```

```
## 2 2 1 21 14745 3 1 2 0 11092
## 3 3 0 20 11558 3 3 0 1 14745
## 4 4 0 21 11747 3 1 1 1 11558
## 5 5 0 21 14319 2 3 3 1 11747
## 6 6 1 19 7126 3 2 0 1 14319
## 7 7 0 20 16067 3 4 0 0 7126
## 8 8 0 20 3939 3 2 0 1 16067
## 9 9 1 21 28230 4 2 0 0 3939
## 10 10 0 21 17396 4 4 1 1 28230
## 11 11 1 20 36152 1 4 0 0 17396
## 12 12 0 20 16567 4 4 1 1 36152
## 13 13 0 20 NA NA NA NA NA 16567
## 14 14 1 18 11853 4 2 0 1 NA
## 15 15 1 18 20433 4 2 2 1 11853
## 16 16 1 18 20701 3 4 0 0 20433
## 17 17 0 20 NA NA NA NA 1 20701
## 18 18 1 19 17509 3 3 1 1 NA
## 19 19 0 21 9779 3 3 1 0 17509
## 20 20 0 22 18598 3 3 1 1 9779
## 21 21 1 20 36665 2 3 0 1 18598
## 22 22 0 21 8094 4 3 1 1 36665
## 23 23 1 19 48769 2 5 0 0 8094
## 24 24 1 20 22601 4 4 1 1 48769
## 25 25 1 19 37950 4 4 0 1 22601
## 26 26 1 20 56047 4 4 0 1 37950
```

```
model2 = lm(preceding_tetris~run,data=d_mod)
summary(model2)
```

```
##
## Call:
## lm(formula = preceding_tetris ~ run, data = d_mod)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15610  -7521  -2337    2424   29220
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18903.8     3335.8   5.667 1.26e-05 ***
## run           645.6     4823.5   0.134   0.895
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11560 on 21 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.0008524, Adjusted R-squared:  -0.04673
## F-statistic: 0.01792 on 1 and 21 DF, p-value: 0.8948
```

No anticipation assumption implies that a treatment in future has no impact on current. In this case, we are measuring the impact of running on previous day results. As can be seen from the report, the impact of running next day is 645 points today and it has NO statistical significance. So, there does not appear to be any effect of future treatment, thereby confirming the no-anticipation assumption

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?

```
model_run = lm(tetris~run,data=d)
summary(model_run)

##
## Call:
## lm(formula = tetris ~ run, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19294  -6707  -1154   4890  29628
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12806      3708    3.453  0.00226 **
## run          13613      4856    2.804  0.01035 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11730 on 22 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.2632, Adjusted R-squared:  0.2297
## F-statistic:  7.86 on 1 and 22 DF,  p-value: 0.01035
```

The answer is same as ‘b’. ATE is 13613 points and it is statistically significant with a p-value of 0.01.

- 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?

```
model_energy = lm(energy~run,data = d)
model_gre = lm(gre~run,data=d)

# Find robust standard errors
#coefest(model_energy, vcov = vcovHC(model_energy, type="HC1"))
#coefest(model_gre, vcov = vcovHC(model_gre, type="HC1"))

summary(model_energy)$coefficients

##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 3.00000000  0.3366180  8.9121790 9.396462e-09
## run         0.07142857  0.4407364  0.1620664 8.727328e-01

summary(model_gre)$coefficients

##              Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  0.8181818  0.1384649  5.9089477 5.048218e-06
## run        -0.1753247  0.1850315 -0.9475396 3.532139e-01
```

Energy levels seem to increase by 0.07 while GRE score has a drop by 0.175 with running. Both do not have statistical significance.

- 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?

I think this is part of publication bias where the experimenter has a tendency to publish only the findings which is statistically significant and to throw away the rest in their file drawer. She is also fishing here for statistical significance. This will lead to false conclusions.

- 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?

```
# Let us create a new data frame and add a column of previous day run.
# We will copy previous day's score. Initialize day 1 with 0
```

```
d_mod2 = d
d_mod2$prev_run = c("NA",d$run[1:25])
#for (i in 1:nrow(d)) {
#  if (i==1) {
#    d_mod2$prev_run[i] = 0
#  } else {
#    d_mod2$prev_run = d$run[i-1]
#  }
#}
# Remove day 1 as we do not have preceding day for day 1
d_mod2 <- d_mod2[-c(1),]

model3 = lm(tetris~run+prev_run,data=d_mod2)
summary(model3)
```

```
##
## Call:
## lm(formula = tetris ~ run + prev_run, data = d_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19693.2  -7830.5   -46.2   5359.5  27539.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11793      4754   2.481  0.02211 *
## run           15026      4986   3.014  0.00686 **
## prev_run1      1689      4948   0.341  0.73643
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11740 on 20 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.3125, Adjusted R-squared:  0.2437
## F-statistic: 4.545 on 2 and 20 DF,  p-value: 0.0236
```

Here, the effect of running today increased the tetris score (15026) compared to what we obtained in part 'd'. This indicates that omitting the previous day run had a downward bias on treatment 'run' there by underestimating the coefficient Running previous day has negative correlation to tetris score, and hence omitting it has underestimated the today's run coefficient. Previous day run has no statistical significance on the tetris score.

In order to check the negative correlation of previous day run, let us run a regression only on

previous run to see the result

```
# Measure the impact of previous day run on tetris score
model4 = lm(tetris~prev_run,data=d_mod2)
summary(model4)$coefficients[2,1]
```

```
## [1] -360.3485
```

- 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)?