Problem Set 4

Alex, Micah, Scott, and David

12/01/2021

Contents

1 Consider Designs

1.1 Game night!

Suppose that you're advertising a board-game or online game to try and increase sales. You decide to individually randomly-assign into treatment and control. After you randomize, you learn that some treatment-group members are friends with control-group members IRL.

- What is the causal quantity that you would have **liked** to estimate?
- What is the causal quantity that you have **in fact** estimated?
- Is there any relationship between the two? Do you think that what you have estimated will be higher, lower, or about the same effect as the causal quantity that you would have liked to estimate?

Answer: In this experiment, we wanted to see if there is a causal relationship between advertising a board-game or online game on sales. However, since we find out that some treatment group members are friends with control-group members, there is a positive spill-over going on in this experiment. This will result in an underestimation because as individuals in treatment group talk to their friends in control groups, they may let them know about the ads they saw polluting each other in terms of intended treatment that we wanted to create.

1.2 Bonus time!

As we're writing this question, end-of-year bonuses are being given out in people's companies. (This is not a concept we your instructors have in the program – each day with your smiling faces is reward enough – and who needs money anyways?)

Suppose that you're interested in knowing whether this is a good idea from the point of view of worker productivity and so you agree to randomly assign bonuses to some people.

- What is the causal quantity that you would have **liked** to estimate?
- What is the causal quantity that you have **in fact** estimated?
- Is there any relationship between the two? Do you think that what you have estimated will be higher, lower, or about the same effect as the causal quantity that you would have liked to estimate?

Answer: In this experiment, we wanted to see if there is a causal relationship between giving bonuses to instructors and worker productivity. Since we have no more information given about whether control or treatment groups will talk about their bonuses with each other, as long as those who received the bonus(treatment) and those who didn't (control) don't affect each others productivity, then the we wouldn't expect a overestimation or underestimation.

Now, if it turns out that instructors freely talked about whether they received a bonus or not with each other. Then we will be polluting the results in terms of intended treatment that we wanted to create. If individuals who were in treatment group told individuals who were in control group about their bonuses, this could go two ways. First, control group might work harder (more productive) in which case this would be a positive spill over, thus resulting in underestimation. Second, treatment group seeing that they received a bonus and their control group didn't even if the treatment group didn't work as hard as control group, then they might slack off, thus resulting in overestimation.

2 Noncompliance in Recycling Experiment

2.1 Intent to treat effect

What is the ITT? Do the work to compute it, and store it into the object `recycling_itt`. Provide a short narrative using inline R code, such as r inline_reference.

```
# ITT = E[Y|z=1] - E[Y|z=0]

recycling_itt <- 500 - 600
recycling_itt
```

```
## [1] -100
```

Answer: ITT is -100

2.2 Compliers average causal effect

What is the CACE? Do the work to compute it, and store it into the object `recycling_cace`. Provide a short narrative using inline R code.

```
#CACE = ITT/ITT_d
recycling_cace <- (-100)/(700/1500)
recycling_cace
```

```
## [1] -214.2857
```

Answer: CACE is -214.2857143

2.3 Mike's CACE

What is the CACE if Mike is correct? Provide a short narrative using inline R code.

```
#CACE = ITT/ITT_d
cace_mike <- (-100)/(500/1500)
cace_mike
```

```
## [1] -300
```

Answer: Mike's CACE is -300

2.4 Andy's CACE

What is the CACE if Andy is correct? Provide a short narrative using inline R code.

```
#CACE = ITT/ITT_d
cace_andy <- (-100)/(600/1500)
cace_andy
```

```
## [1] -250
```

Answer: Andy's CACE is -250‘

2.5 Effect of false reporting

What was the impact of the undergraduates's false reporting on our estimates of the treatment's effectiveness?

Answer: Per the instructions, we will suppose that Mike's CACE is correct with a value of -300 and Andy's false report generated a CACE of -250, which resulted in a difference of 50. Andy's false reporting had a higher percentage of compliers out of the total 1500 treatment group bringing CACE closer to ATE. As the number of compliers increases, the value of CACE gets closer to ATE.

2.6 Effect of false reporting... on what quantity?

Does your answer change depending on whether you choose to focus on the ITT or the CACE?

Answer: In this case, ITT remains the same, since Mike's or Andy's reporting does not effect the ITT because ITT is calculated based on the outcome between overall treatment and overall control groups.

3 Fun with the placebo

3.1 Make data

Construct a data set that would reproduce the table. (Too frequently we receive data that has been summarized up to a level that is not useful for our analysis. Here, we're asking you to "un-summarize" the data to conduct the rest of the analysis for this question.)

```
#d <- data.table('fill this in')

set.seed(2)
baseline <- data.frame(Assignment = c("Baseline"), Treated = c("No"),
                       Turnout=sample(c(0,1),prob=c(1-0.3008, 0.3008),
                                      size=2463, replace=TRUE))

treatment_1 <- data.frame(Assignment = c("Treatment"), Treated = c("Yes"),
                          Turnout=sample(c(0,1), prob=c(1-0.3890, 0.3890    ),
                                         size=512, replace=TRUE))

treatment_0 <- data.frame(Assignment = c("Treatment"), Treated = c("No"),
                          Turnout=sample(c(0,1),  prob=c(1-0.3160, 0.3160),
                                         size=1898, replace=TRUE))

placebo_1 <- data.frame(Assignment = c("Placebo"), Treated = c("Yes"),
                        Turnout=sample(c(0,1),  prob=c(1-0.3002, 0.3002),
                                       size=476,  replace=TRUE))

placebo_0 <- data.frame(Assignment = c("Placebo"), Treated = c("No"),
                        Turnout=sample(c(0,1),prob=c(1-0.3145, 0.3145),
                                       size=2108, replace=TRUE))


d <- rbind(baseline, treatment_1, treatment_0, placebo_1, placebo_0)
d <- data.table(d)
glimpse(d)
```

```
## Rows: 7,457
## Columns: 3
## $ Assignment <chr> "Baseline", "Baseline", "Baseline", "Baseline", "Baseline",~
## $ Treated    <chr> "No", "No", "No", "No", "No", "No", "No", "No", "No", "No",~
## $ Turnout    <dbl> 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0,~
```

3.2 Estimate the compliance rate using the treatment group

Estimate the proportion of compliers by using the data on the treatment group. Provide a short narrative using inline R code, such as r inline_reference.

```
#treatment_1/(treatment_1 + treatment_0)
compliance_rate_t <- (512)/(512 + 1898)
compliance_rate_t
```

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

Answer: The proportion of compliers is 0.2124481.

3.3 Estimate the compliance rate using the control group

C. Estimate the proportion of compliers by using the data on the placebo group. Provide a short narrative using inline R code.

```
#placeb_1/(placebo_1 + placebo_0)
compliance_rate_p <- (476)/(476 + 2108)
compliance_rate_p
```

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

Answer: The proportion of compliers is 0.1842105

3.4 Compare these compliance rates

Are the two compliance rates statistically significantly different from each other? Provide *a test* – this means that you cannot simply "look at" or "eyeball" the coefficients and infer some conclusion – and a description about why you chose that particular test, and why you chose that particular set of data.

```
# I will take a subset of the data without the baseline.
data_subset <- table(d$Assignment, d$Treated)
data_subset <- tail(data_subset, -1)
data_subset
```

```
##
##                No  Yes
##   Placebo    2108  476
##   Treatment 1898  512
```

```
#loading libraries
library(MASS)
library(dplyr)

#run a chi squared test

proportions_difference_test <- chisq.test(data_subset)
proportions_difference_test
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data_subset
## X-squared = 6.0887, df = 1, p-value = 0.0136
```

Answer: Since the variables are categorical, chi-squared test was appropriate for this analysis. The p-value of 0.0136 is less than 0.05 and therefore statistically significant. Since we are measuring compliance rate from T and Placebo groups, those are the groups that I chose. I tested whether the treated variable "Yes" and "No" for both Placebo and Treatment are statistically different. As it turns out, based on the p-value they are different.

3.5 Evaluate assumptions

What critical assumption does this comparison of the two groups' compliance rates test? Given what you learn from the test, how do you suggest moving forward with the analysis for this problem?

Answer: We have two main assumptions > Non-interference and exclusion restriction. Based on our assumptions, the compliance rate should be similar in treatment and placebo groups.

3.6 Compliers average treatement effect... of the placebo?

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

```
ITT   <- (((476)/(476+2108))*0.3002 + ((2108)/(476+2108))*0.3145 - 0.3008)

ITT_d <- (476)/(476+2108)

cace_estimate <- ITT / ITT_d
cace_estimate
```

```
## [1] 0.06007143
```

Answer: the CACE of receiving placebo is 0.0600714. CACE although very small, still seems to have some effect.

3.7 Diference in means estimator

Using a difference in means (i.e. not a linear model), compute the ITT using the appropriate groups' data. Then, divide this ITT by the appropriate compliance rate to produce an estimate the CACE. Provide a short narrative using inline R code.

```
itt   <- (((512)/(512+1898))*0.3890 + ((1898)/(512+1898))*0.3160 - 0.3008)
itt_d <- (512)/(512+1898)

cace_means <- itt /  itt_d
cace_means
```

```
## [1] 0.1445469
```

Answer: The CACE of receiving treatment is 0.1445469

3.8 Linear model estimator

Use two separate linear models to estimate the CACE of receiving the treatment by first estimating the ITT and then dividing by ITT_D. Use the `coef()` extractor and in line code evaluation to write a descriptive statement about what you learn after your code.

```
itt_model   <- d[Assignment != 'Placebo', lm(Turnout ~ Assignment)]
#summary(itt_model)

# First Stage
itt_d_model <- d[Assignment != 'Placebo', lm(Treated =='Yes' ~ Assignment)]
#summary(itt_d_model)

stargazer(
  itt_model, itt_d_model,
  type = 'text'
)
```

```
##
## ===============================================================
##                                    Dependent variable:
##                                ---------------------------
##                                  Turnout   Treated == "Yes"
##                                    (1)           (2)
## ---------------------------------------------------------------
## AssignmentTreatment               0.030**       0.212***
##                                   (0.013)       (0.008)
##
## Constant                         0.307***        -0.000
##                                   (0.009)       (0.006)
##
## ---------------------------------------------------------------
## Observations                       4,873         4,873
## R2                                 0.001         0.120
## Adjusted R2                        0.001         0.120
## Residual Std. Error (df = 4871)    0.467         0.288
## F Statistic (df = 1; 4871)        5.022**      664.140***
## ===============================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

```
cace_from_models <- coef(itt_model)[2]/coef(itt_d_model)[2]
cace_from_models
```

```
## AssignmentTreatment
##           0.1411484
```

Answer: CACE estimator by this method give us the above result.

3.9 Data subset estimator

When a design uses a placebo group, one additional way to estimate the CACE is possible – subset to include only compliers in the treatment and placebo groups, and then estimate a linear model. Produce that estimate here. Provide a short narrative using inline R code.

```
cace_subset_model <- d[Treated == 'Yes', lm(Turnout ~ Assignment)]
#summary(cace_subset_model)
coef(cace_subset_model)[2]
```

```
## AssignmentTreatment
##           0.1174009
```

Answer: CACE estimator by this method is as above.

3.10 Evaluate estimators

In large samples (i.e. "in expectation") when the design is carried out correctly, we have the 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?

Answer: The CACE in three cases are similar but not the same. 8 and 9 seem to have closer values than 7. Placebo design seems to be the most reliable of the three because it can deal with never takers in better way in control group.

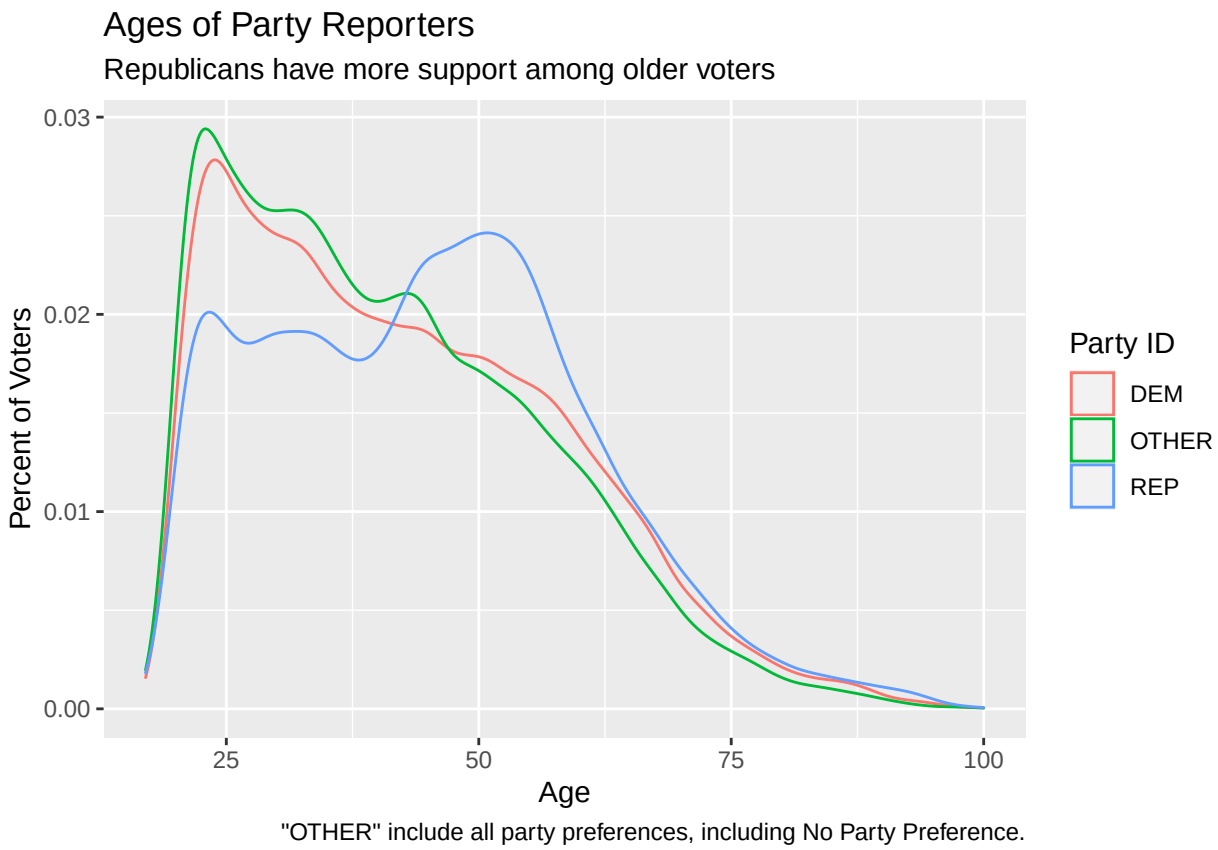
4 Another Turnout Question

```
#randomly sampling 100,000 data points from the total 4,000,000

#setting seed
set.seed(113)

cols <- fread("curl https://people.ischool.berkeley.edu/~d.alex.hughes/data/hill_kousser_analysisFile.cs
d <- fread("curl https://people.ischool.berkeley.edu/~d.alex.hughes/data/hill_kousser_analysisFile.csv

head(d)
```

```
##    LocalityCode age.bin party.bin in.toss.up.dist minority.dist vote.10.gen
## 1:            1       2         2               0             0           0
## 2:           36       5         4               0             0           0
## 3:           19       4         2               0             0           0
## 4:           19       2         3               0             1           0
## 5:           39       5         2               0             0           0
## 6:           37       5         3               1             0           1
##    vote.08.gen Party age.in.14 Gender Dist1 Dist2 Dist3 Dist4 Dist5 Dist6 Dist7
## 1:           1   DEM        35      M CG015 SA020 SE002 SS010    NA    NA    NA
## 2:           0    AI        68        CG008 SA033 SE001 SS021    NA    NA    NA
## 3:           1   DEM        54      M CG023 SA036 SE001 SS021    NA    NA    NA
## 4:           0    DS        30      M CG034 SA051 SE003 SS024    NA    NA    NA
## 5:           0   DEM        66      M CG009 SA013 SE001 SS005    NA    NA    NA
## 6:           1   NPP        61      F CG052 SA077 SE004 SS039    NA    NA    NA
##    Dist8 reg.date.pre.08 reg.date.pre.10 vote.12.gen vote.12.pre.pri
## 1:    NA               0               1           1               0
## 2:    NA               0               0           1               0
## 3:    NA               1               1           1               0
## 4:    NA               1               1           1               0
## 5:    NA               0               0           1               0
## 6:    NA               1               1           1               0
##    vote.10.gen.pri vote.08.pre.pri vote.08.gen.pri block.num leftover.case
## 1:               0               0               0       108             0
## 2:               0               0               0       331             0
## 3:               0               0               0       236             0
## 4:               0               0               0       127             0
## 5:               0               0               0       299             0
## 6:               0               0               0       326             0
##    treatment.assign yvar matched.to.post vote.14.gen
## 1:          Control    0               1           0
## 2:          Control    0               1           0
## 3:          Control    0               1           0
## 4:          Control    0               1           0
## 5:          Control    0               1           0
## 6:          Control    0               1           0
```

```
d %>%
  ggplot() +
  aes(x = age, color = three_party) +
  geom_density() +
```

```
  scale_x_continuous(limits = c(17, 100)) +
  labs(
    title = 'Ages of Party Reporters',
    subtitle = 'Republicans have more support among older voters',
    x = 'Age', y = 'Percent of Voters',
    color = 'Party ID',
    caption = '"OTHER" include all party preferences, including No Party Preference.'
  )
```



4.1 Simple treatment effect

Load the data and estimate a `lm` model that compares the rates of turnout in the control group to the rate of turnout among anybody who received *any* letter. This model combines all the letters into a single condition – "treatment" compared to a single condition "control". Report robust standard errors, and include a narrative sentence or two after your code using inline R code, such as r inline_reference.

```
# Use this block for your work
model_simple <- d[, lm(vote ~ any_letter)]
#summary(model_simple)
model_simple$vcovHC_ <- vcovHC(model_simple)
coeftest(model_simple, vcov. = model_simple$vcovHC_)
```

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

```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.09378967 0.00094046 99.7277   <2e-16 ***
## any_letterTRUE 0.00795302 0.00493134  1.6128   0.1068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: The p-value is greater than 0.05 is statistically not significant. Therefore, we cannot reject the null hypothesis that there is difference in outcome being control and treatment group.

4.2 Letter-specific treatment effects

Suppose that you want to know whether different letters have different effects. To begin, what are the effects of each of the letters, as compared to control? Estimate an appropriate linear model and use robust standard errors. Provide a short narrative using inline R code.

```
model_letter <- d[, lm(vote ~ treatment_f )]
#summary(model_simple)
model_letter$vcovHC_ <- vcovHC(model_letter)
coeftest(model_letter, vcov. = model_letter$vcovHC_)
```

```
##
## t test of coefficients:
##
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.09378967 0.00094046 99.7277  < 2e-16 ***
## treatment_fElection info 0.00298452 0.01046909  0.2851  0.77558
## treatment_fPartisan      0.00417470 0.00756104  0.5521  0.58086
## treatment_fTop-two info  0.01447804 0.00801985  1.8053  0.07103 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: Based on the p-values above, it doesn't look like any of the letters are statistically significant.

4.3 Test for letter-specific effects

Does the increased flexibility of a different treatment effect for each of the letters improve the performance of the model? Test, using an F-test. What does the evidence suggest, and what does this mean about whether there **are** or **are not** different treatment effects for the different letters?

```
anova(model_simple, model_letter, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: vote ~ any_letter
## Model 2: vote ~ treatment_f
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1  99998 8524.3
## 2  99996 8524.2  2   0.10722 0.6289 0.5332
```

Answer: Based on the F-test from above, it does not look like having letter specific effects improved the performance of the model.

4.4 Compare letter-specific effects

Is one message more effective than the others? The authors have drawn up this design as a full-factorial design. Write a *specific* test for the difference between the *Partisan* message and the *Election Info* message. Write a *specific* test for the difference between *Top-Two Info* and the *Election Info* message. Report robust standard errors on both tests and include a short narrative statement after your estimates.

```
#levels(d$treatment_f)
model_1 <- d[treatment_f %in% c("Election info","Partisan"), lm(vote ~ treatment_f)]

model_1$vcovHC_ <- vcovHC(model_1)
coeftest(model_1, vcov. = model_1$vcovHC_)
```

```
##
## t test of coefficients:
##
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.0967742  0.0104268  9.2813   <2e-16 ***
## treatment_fPartisan 0.0011902  0.0128453  0.0927   0.9262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: In both cases, we can't reject the null hypothesis.

4.5 Count the number of blocks

Blocks? We don't need no stinking blocks? The blocks in this data are defined in the `block.num` variable (which you may have renamed). There are a *many* of blocks in this data, none of them are numerical – they're all category indicators. How many blocks are there?

```
block_num <- length(unique(d$block))
block_num
```

```
## [1] 382
```

Answer: 382 blocks.

4.6 Add block fixed effects

SAVE YOUR CODE FIRST but then try to estimate a `lm` that evaluates the effect of receiving *any letter*, and includes this block-level information. What happens? Why do you think this happens? If this estimate *would have worked* (that's a hint that we don't think it will), what would the block fixed effects have accomplished?

Answer: Since we have too many blocks and we are factoring them in in our regression individually, I don't think it would have tell us much at all. If it actually worked, then each Block would have soaked up some of the error.

4.7 A clever work-around?

Even though we can't estimate this fixed effects model directly, we can get the same information and model improvement if we're *just a little bit clever*. Create a new variable that is the *average turnout within a block* and attach this back to the data.table. Use this new variable in a regression that regresses voting on `any_letter` and this new `block_average`. Then, using an F-test, does the increased information from all these blocks improve the performance of the *causal* model? Use an F-test to check.

```
d[ ,block_average := mean(vote), keyby = .(block)]

mod_average_block <- d[, lm(vote ~ any_letter + block_average)]
mod_average_block$vcovHC_ <- vcovHC(mod_average_block)
coeftest(mod_average_block, vcov. = mod_average_block$vcovHC_)
```

```
##
## t test of coefficients:
##
##                   Estimate  Std. Error t value Pr(>|t|)
## (Intercept)    -0.00027946  0.00171966 -0.1625   0.8709
## any_letterTRUE  0.00727577  0.00483198  1.5058   0.1321
## block_average   0.99995277  0.01965880 50.8654   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model_simple, mod_average_block, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: vote ~ any_letter
## Model 2: vote ~ any_letter + block_average
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1  99998 8524.3
## 2  99997 8133.1  1    391.19 4809.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: Block_average variable is statistically significant and after running F test between the causal model and the model with the covariate block_average, the p-value is also significant, indicating that adding the block_averages did improve the model.

4.8 Does cleverness create a bad-control?

Doesn't this feel like using a bad-control in your regression? Has the treatment coefficient changed from when you didn't include the `block_average` measure to when you did? Have the standard errors on the treatment coefficient changed from when you didn't include the `block_average` measure to when you did? Why is this OK to do?

Answer: The treatment coefficient did increase when we included the block_average. The standard errors seems to have remained the same. I think this is a bad control because it appears that they average_block variable may have been affected by any_letter.

5 3. Optional Turnout in Dorms

5.1 Use Linear Regressions

1. Estimate the ITT using a linear regression on the appropriate subset of data. Notice that there are two `NA` in the data. Just na.omit to remove these rows so that we are all working with the same data. Given the ways that randomization was conducted, what is the appropriate way to construct the standard errors?

```
dorm_model <- 'fill this in'
```

5.2 Use Randomization Inference

1. How many people are in treatment and control? Does this give you insight into how the scientists might have randomized? As usual, include a narrative sentence after your code.

```
n_treatment <- 'fill this in'
n_control   <- 'fill this in'
```

Narrative: ...

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

```
# Use this block for your work
```

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

```
dorm_room_cace <- 'fill this in'
```

Narrative: ...

4. What are the 2.5% and 97.5% quantiles of this distribution?

```
dorm_room_ci <- 'fill this in with a length-two vector; first number 2.5%, second number 97.5%'
```

Narrative: ...

5. 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_value <- 'fill this in'
```

Narrative: ...

6. Assume that the leaflet (which was left in case nobody answered the door) had no effect on turnout. Estimate the CACE either using ITT and ITT_d or using a set of linear models. What is the CACE, the estimated standard error of the CACE, and the p-value of the test you conduct?

```
dorm_room_cace <- 'fill this in'
```

Narrative: . . .

7. What if the leaflet that was left actually *did* have an effect? Is it possible to estimate a CACE in this case? Why or why not?

Narrative: . . .