

# Problem Set #4

## *Experiments and Causality*

```
# load packages
library(foreign)
library(data.table)
library(knitr)
```

## 1. Potential Outcomes

Consider the following hypothetical schedule of potential outcomes.

- Amy, Brian, and Chris are *compliers*. This means they actually get the treatment if they are assigned to the treatment group. Their potential outcomes in the untreated state of the world are 11, 10, and 11 respectively.
  - David, Erin, and Felipe are never-takers. (I.e. they do not get the treatment even if they are assigned to the treatment group.) Their potential outcomes in the untreated state of the world is 3, 2, and 4 respectively.
- Make up a set of potential outcomes in the treated state of the world (i.e.  $Y_i(1)$  for each of the individuals listed above) that would make both the ATE and the CACE positive.

```
d <- data.table(name = c("Amy", "Brian", "Chris", "David", "Erin", "Felipe"))
d[, complier := c(1, 1, 1, 0, 0, 0)]
d[, y0 := c(11, 10, 11, 3, 2, 4)]
d[, y1_1 := c(16, 14, 17, 10, 9, 10)]

ate_1 = mean(d[, y1_1-y0])
cace_1 = mean(d[complier == 1, y1_1-y0])
print(ate_1)
```

```
## [1] 5.833333
```

```
print(cace_1)
```

```
## [1] 5
```

- Make up a set of potential outcomes in the treated state of the world that would make the ATE positive but the CACE *negative*.

```
d[, y1_2 := c(10, 9, 8, 8, 10, 7)]

ate_2 = mean(d[, y1_2-y0])
cace_2 = mean(d[complier == 1, y1_2-y0])
print(ate_2)
```

```
## [1] 1.833333
```

```
print(cace_2)
```

```
## [1] -1.666667
```

- Suppose that you are conducting a trial for a new feature to be released in a product. From a limited point of view, if you are the person who wrote the *creative* content that is in the new feature, do you

care more about the CACE or the ATE?

4. Suppose that you are conducting a trial for a new feature to be released in the same product. From a limited point of view, compared to when you wrote the creative, if you are the product manager, do you care relatively **more** about the CACE or the ATE than before?

## 2. Noncompliance in Recycling Experiment

Suppose that you want to conduct a study of recycling behavior. A number of undergraduate students are hired to walk door to door and provide information about the benefits of recycling to people in the treatment group. 1,500 households are assigned to the treatment group. The undergrads tell you that they successfully managed to contact 700 households. The control group had 3,000 households (not contacted by any undergraduate students). The subsequent recycling rates (i.e. the outcome variable) are computed and you find that 500 households in the treatment group recycled. In the control group, 600 households recycled.

1. What is the ITT?

```
itt <- 500/1500 - 600/3000
```

ITT is 0.1333333

2. What is the CACE?

```
itt_d = 700/1500
```

CACE = ITT/ITT\_D = 0.2857143

3. There appear to be some inconsistencies regarding how the undergraduates actually carried out the instructions they were given. One of the students, Mike, tells you that they actually lied about the the number of contacted treatment households. The true number was 500. Another student, Andy, tells you that the true number was actually 600.

- a. What is the CACE if Mike is correct?

```
itt_d_m = 500/1500
```

CACE = 0.4

- b. What is the CACE if Andy is correct?

```
itt_d_a = 600/1500
```

CACE = 0.3333333

4. Suppose that Mike is correct.

- a. What was the impact of the undergraduates's false reporting on our estimates of the treatment's effectiveness? If fewer people actually received treatment, it means that the estimated CACE should be even bigger in magnitude. The rationale is that the same outcome was achieved with less treatment actually administered. Therefore, we could see that CACE estimate is the biggest when Mike is correct.
- b. Does your answer change depending on whether you choose to focus on the ITT or the CACE? Yes. Since ITT depends solely on treatment assignment but not administration, it remains the same regardless how many people are actually treated.

### 3. Fun with the placebo

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

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

1. Construct a data set that would reproduce the table.
2. Estimate the proportion of compliers by using the data on the treatment group.
3. Estimate the proportion of compliers by using the data on the placebo group.
4. Are the proportions in parts (1) and (2) statistically significantly different from each other? Provide a *test* and an description about why you chose that particular test, and why you chose that particular set of data.
- e. What critical assumption does this comparison of the two groups' compliance rates test?
- f. Estimate the CACE of receiving the placebo. Is the estimate consistent with the assumption that the placebo has no effect on turnout?
- g. Estimate the CACE by first estimating the ITT and then dividing by  $ITT_D$ .
- h. Estimate the CACE by comparing the turnout rates among the compliers in both the treatment and placebo groups. Interpret the results.
- i. In class we discussed that the rate of compliance determines whether one or another design is more efficient. (You can review the paper [here](#)). Given the compliance rate in this study, which design *should* provide a more efficient estimate of the treatment effect?
- j. Does it?

### 4. Turnout in Dorms

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.

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

```
##      turnout treated  dormid treatment_group
##    1:         0      0 1010101              0
```

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

Here's what is in that data:

- **turnout** did the person turn out to vote?
  - **treated** did someone at the dorm open the door?
  - **dormid** a unique ID for the door of the dorm
  - **treatment\_group** whether the dorm door was assigned to be treated or not
1. Using the data set from the book's website, estimate the ITT. First, estimate the ITT using the difference in two-group means. 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.*
  2. 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 – in particular, are you surprised at your result when you compare it to the p-value in part (1)? (This is a 2 point question, because there's quite a bit of work here.)
  3. Assume that the leaflet had no effect on turnout. Estimate the CACE. Do this in two ways:
    - a. First, estimate the CACE using means.
    - b. Second, use some form of linear model to estimate this as well. If you use a 2SLS, then report the standard errors and draw inference about whether contact had any causal effect among compliers.

## 5. Another Turnout Question

We're sorry; it is just that the outcome and treatment spaces are so clear!

Hill and Kousser (2015) report that it is possible to increase the probability that someone votes in the California *Primary Election* simply by sending them a letter in the mail. This is kind of surprising, because who even reads the mail anymore anyways? (Actually, if you talk with folks who work in the space, they'll say, "We know that everybody throws our mail away; we just hope they see it on the way to the garbage.")

Can you replicate their findings? Let's walk through them.

```
## d <- fread('http://ischool.berkeley.edu/~d.alex.hughes/data/hill_kousser_analysisFile.csv')
## head(d)
```

You'll note that this takes some time to download. Probably best to save a copy locally, and keep from reading this off the internet. In your project structure, create a folder called `./data/raw/` and write this file to the folder. Data that is in this raw folder *never* gets modified; instead, any changes that you make should be reflected into either an `./data/interim/` or `./data/analysis/` folder. You might consider using the function `fwrite` from `data.table`.

Here's what is in that data.

- **age.bin** a bucketed version of the **age.in.14** variable
- **party.bin** a bucketed version of the **Party** variable
- **in.toss.up.dist** whether the voter lives in a close race

- `minority.dist` whether the voter lives in a majority minority district
- `Gender` voter file reported gender
- `Dist1-8` congressional and data districts
- `reg.date.pre.08` whether the voter has been registered since before 2008
- `vote.xx.gen` whether the voter voted in the `xx` general election
- `vote.xx.gen.pri` whether the voter voted in the `xx` general primary election
- `vote.xx.pre.pri` whether the voter voted in the `xx` presidential primary election
- `block.num` a block indicator for blocked random assignment.
- `treatment.assign` either “Control”, “Election Info”, “Partisan Cue”, or “Top-Two Info”
- `yvar` the outcome variable: did the voter vote in the 2014 primary election

These variable names are horrible. Do two things:

- Rename the smallest set of variables that you think you might use to something more useful
- For the variables that you think you might use; check that the data makes sense;

Then, save this data to `./data/analysis/`.

Well, while you’re at it, you might as well also modify your `.gitignore` to ignore the data folder. Because you’re definitely going to have the data rejected when you try to push it to github.

1. **A Simple Treatment Effect:** Load the data from `./data/analysis/` and estimate a model that compares the rates of turnout in the control group to the rate of turnout among *anybody* who received a letter. Report robust standard errors.
2. **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? Report robust standard errors on a linear model.
3. Then, test, using an F-test, whether the increased flexibility of the model estimated in part (2) has improved the performance of the model over that estimated in part (1). What does the evidence suggest?
4. **More Specific Treatment 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.
5. **Blocks?** There are a *many* of blocks in this data. How many?
6. Create a new indicator that is the *average turnout within a block* and attach this back to the `data.table`. Use this new indicator in a regression that predicts the difference between Control and Any Letter. 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.
7. **HTES?** Do you think that there are features of the data that might systematically predict that people will respond strongly or weakly to the treatment effect? List two that you think might be there, in the order that you would like to test them. Then, test for these heterogeneities. What do you learn? What is the right way to adjust your p-values, given that you’re testing twice?
8. Summarize these results in a short paragraph that includes inline reports from your estimated models. (This can be integrated into your last response, if that works better for you.)
9. Cheating? Suppose that you didn’t write down your testing plan. How risky is the false discovery problem in this data set?