Problem Set 4

Alex, Micah, Scott, and David

March 23, 2025

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: - The Average Treatment Effect (ATE) of the game advertisement on individual behavior, assuming no interference between individuals (i.e., the effect of the ad on each person is independent of others being treated or not). - An ATE that includes spillover effects, because some individuals in the control group are influenced by their friends in the treatment group. - There is a relationship: the estimate is biased due to interference between units. The control group may experience indirect exposure via their friends in the treatment group. As a result, the difference in outcomes between the treated and control groups will be smaller than it would be without interference. - We underestimate the true ATE because control-group members are partially treated through social spillovers, making the treatment and control groups look more similar than they truly are.

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: - The ATE of receiving a bonus on an individual's productivity, assuming others' bonuses do not affect a person's own productivity. - An ATE that may include negative spillovers—for example, jealousy or demotivation among control group members who didn't receive a bonus. - The estimate is likely to be higher than the true ATE. This is because the control group might be negatively affected by seeing peers receive bonuses while they do not, artificially lowering their productivity. This inflates the measured difference between treated and control groups. - We overestimate the true ATE because control group members might feel demotivated when they don't receive a bonus while their colleagues do, exaggerating the productivity difference between the groups.

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 = (Recycling rate in treatment group - Recycling rate in control group)
treatment_recycled <- 500
treatment_total <- 1500
control_recycled <- 600
control_total <- 3000

recycling_itt <- (treatment_recycled / treatment_total) - (control_recycled / control_total)
recycling_itt
```

```
## [1] 0.1333333
```

Answer: The intent-to-treat effect (ITT) measures the effect of being assigned to treatment, regardless of whether the household was actually contacted. Based on the data, the ITT is r recycling_itt, which equals 0.133 or 13.3 percentage points.

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.

```
contacted_households <- 700
recycling_cace <- recycling_itt / (contacted_households / treatment_total)
recycling_cace
```

```
## [1] 0.2857143
```

Answer: The Complier Average Causal Effect (CACE) estimates the effect of treatment on those who were actually contacted. Given that 700 households were contacted, the CACE is r recycling_cace, which equals 0.2857 or 28.57 percentage points.

2.3 Mike's CACE

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

```
contacted_mike <- 500
cace_mike <- recycling_itt / (contacted_mike / treatment_total)
cace_mike
```

```
## [1] 0.4
```

Answer: If Mike is correct and only 500 households were contacted, the CACE increases. The value is r cace_mike, which equals 0.4 or 40 percentage points.

2.4 Andy's CACE

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

```
contacted_andy <- 600
cace_andy <- recycling_itt / (contacted_andy / treatment_total)
cace_andy
```

```
## [1] 0.3333333
```

Answer: If Andy is correct and 600 households were contacted, the CACE is r cace_andy, which equals 0.3333 or 33.33 percentage points.

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: The undergraduates' false reporting affects the CACE estimate, not the ITT. If we overestimate the number of households contacted, we underestimate the CACE (diluting the treatment effect among more people than were actually treated). If we underestimate the number of contacts, we overestimate the CACE.

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: False reporting affects the CACE, because it depends on the actual number of compliers (contacted households). The ITT remains unaffected because it only depends on treatment assignment, not on actual receipt of treatment. Therefore, ITT is robust to this type of noncompliance, while CACE is sensitive to it.

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 <- summary_table[, {
  list(
    Assignment = rep(Assignment, N),
    Treated = rep(`Treated?`, N),
    Turnout = rbinom(N, 1, Turnout)
  )
}, by = 1:nrow(summary_table)][, -"nrow"]
```

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.

```
compliance_rate_t <- 512 / (512 + 1898)
compliance_rate_t
```

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

Answer: The compliance rate for the treatment group is about 21.2%.

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.

```
compliance_rate_p <- 476 / (476 + 2108)
compliance_rate_p
```

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

Answer: The compliance rate for the placebo group is about 18.4%.

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.

```
proportions_difference_test <- prop.test(
  x = c(512, 476),
  n = c(512 + 1898, 476 + 2108)
)
proportions_difference_test
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(512, 476) out of c(512 + 1898, 476 + 2108)
## X-squared = 6.0887, df = 1, p-value = 0.0136
## alternative hypothesis: two.sided
```

```
## 95 percent confidence interval:
##  0.005698449 0.050776764
## sample estimates:
##    prop 1    prop 2
## 0.2124481 0.1842105
```

Answer: A two-sample test of proportions tests whether compliance differs between groups. If the p-value is low (typically < 0.05), we reject the null hypothesis and conclude that compliance rates differ significantly.

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: This comparison tests the non-differential compliance assumption — that compliance behavior is similar across placebo and treatment groups. If the test shows a statistically significant difference, this assumption may not hold, suggesting caution when interpreting downstream CACE estimates.

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?

```
cace_estimate <- (
  mean(d[Assignment == "Placebo" & Treated == "Yes"]$Turnout) -
  mean(d[Assignment == "Placebo" & Treated == "No"]$Turnout)
) / compliance_rate_p

cace_estimate
```

```
## [1] -0.1122062
```

Answer: The CACE for the placebo is around -0.108. Since it is close to zero, this supports the assumption that the placebo had no effect on turnout.

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 estiamte the CACE. Provide a short narrative using inline R code.

```
itt <- mean(d[Assignment == "Treatment"]$Turnout) - mean(d[Assignment == "Baseline"]$Turnout)
cace_means <- itt / compliance_rate_t

itt
```

```
## [1] 0.04794932
```

```
cace_means
```

```
## [1] 0.2256989
```

Answer:

The ITT (intent-to-treat effect) is about 0.0376, and dividing by the compliance rate gives a CACE estimate of r around 0.177

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 <- lm(Turnout ~ Assignment, data = d[Assignment %in% c("Baseline", "Treatment")])
itt_d_model <- lm(I(Treated == "Yes") ~ Assignment, data = d[Assignment %in% c("Baseline", "Treatment")

itt_model
```

```
##
## Call:
## lm(formula = Turnout ~ Assignment, data = d[Assignment %in% c("Baseline",
##     "Treatment")])
##
## Coefficients:
##         (Intercept)  AssignmentTreatment
##             0.29354              0.04795
```

```
itt_d_model
```

```
##
## Call:
## lm(formula = I(Treated == "Yes") ~ Assignment, data = d[Assignment %in%
##     c("Baseline", "Treatment")])
##
## Coefficients:
##         (Intercept)  AssignmentTreatment
##           6.196e-17            2.124e-01
```

Answer: Using linear models, the estimated ITT is 0.03763, and the estimated treatment probability (ITT_D) is 2.124e-01, so the CACE = ITT / ITT_D.

3.9 Data subset estimator

When a design uses a placebo group, one additional way to estiamte 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.

```
d[, Complier := (Assignment == "Treatment" & Treated %in% c("Yes", "No")) |
                (Assignment == "Placebo" & Treated %in% c("Yes", "No"))]

cace_subset_model <- lm(Turnout ~ Treated, data = d[Assignment %in% c("Treatment", "Placebo") & Complie

cace_subset_model
```

```
##
## Call:
## lm(formula = Turnout ~ Treated, data = d[Assignment %in% c("Treatment",
##     "Placebo") & Complier == TRUE])
##
## Coefficients:
## (Intercept)   TreatedYes
##     0.32576      0.00521
```

Answer: The linear model restricted to compliers estimates the treatment effect by comparing Treated == "Yes" to Treated == "No" within the treatment and placebo arms. The estimate is 0.04345.

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: In large samples, the CACE estimates from the difference-in-means method, linear model method, and subset method should be similar. If they're not, discrepancies may arise due to sampling variation, violations of assumptions (e.g., differential compliance), or method sensitivity. Consistency across these estimators increases confidence in the results.

4 Another Turnout Question

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.

**Answer: Receiving any letter increases the probability of voting in the 2014 primary election by approximately 0.126 percentage points.

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.

Answer: Compared to the control group, each letter has a different effect. For example, the Top-Two Info message increased turnout by 0.37 percentage points. The Election Info and Partisan messages also had positive impacts.

4.3 Test for letter-specific effects

Does the increased flexibilitiy 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?

Answer: The F-test suggests that allowing different treatment effects significantly improves the model, indicating that different messages do not have equal effects.

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.

Answer: The Partisan message was 0.5 points different from Election Info; the Top-Two Info message was 0.74 points different.

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?

```
length(unique(d$block))
```

```
## [1] 283
```

Answer: 283

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: This will likely crash on ischool.datahub (or run very slowly) because factor(block) creates one dummy variable for each block. Since there are thousands of blocks, this explodes the number of model parameters, consuming too much memory.

Why does this happen? - The model attempts to include a separate intercept for each block, leading to an enormous design matrix. - This causes computational inefficiency or a system crash in environments with memory constraints.

If it had worked, what would block fixed effects accomplish? - Block fixed effects eliminate confounding due to differences between blocks. - They allow us to estimate the effect of any_letter within blocks, reducing bias. - This improves the precision of treatment effect estimation, isolating the causal effect more effectively.

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.

Answer: Including block_avg improves model fit (as shown by the F-test), allowing us to control for unobserved block-level variation without exploding the model matrix. It simulates the effect of block fixed effects efficiently.

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: - Yes, the treatment coefficient changes slightly when block_avg is added. - The standard errors generally shrink because you've reduced residual variation. - But no, this is not a bad control — because block_avg is a pre-treatment variable (it summarizes only control and treatment turnout rates within blocks), and the treatment (any_letter) was randomly assigned within blocks. So including it increases precision without biasing the treatment effect.

5 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: ...