Problem Set 1: Solutions

Leo Yao - yy3959 - Section Number

Due Oct 6th, 2023

This homework must be turned in on Brightspace by Oct 6th 2023. It must be your own work, and your own work only – you must not copy anyone's work, or allow anyone to copy yours. This extends to writing code. **You may consult with others, but when you write up, you must do so alone.**

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. You should submit:

1. A compiled PDF file named yourNetID_solutions.pdf containing your solutions to the problems.

2. A .Rmd file containing the code and text used to produce your compiled pdf named yourNetID_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part

2. Do not include written answers as code comments. Write out answers and explanations separately.

3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

Question 1. Definitions and Examples (20 points)

Answer the following questions. Be as specific and detailed as possible. Give examples.

1. What is the fundamental problem of causal inference? (5 points)

The fundamental problem of causal inference is that we never get to observe both potential outcomes under the same state of condition. For example, when we are trying to estimate the causal effect of taking aspirin on headaches, we cannot observe the potential outcomes of taking aspirin and not taking aspirin on the same person because the individual we are analyzing can either take it or not take it, never both.

2. Why are experiments important? (5 points)

Experiments are important because we cannot establish causation using observational data. Experiments allow us to collect data that satisfies the requirements of causal inference, SUTVA, ignorability and positivity. For example, we wouldn't have our industrialized world if scientists like Newton hadn't done experiments on gravity.

3. What does ignorability mean? (5 points)

Ignorability assumption assumes that researchers are able to ensure that the potential outcomes are independent of treatment assignment, meaning there is no selection bias. For example, if we are analyzing the causal effect of hospitalization on health condition, observational data cannot be used because people who choose to go to the hospital are generally more likely to have a poor health condition than those who don't. Here, the ignorability assumption is violated because of the self-selection.

4. What is SUTVA? (5 points)

SUTVA stands for Stable Unit Treatment Value Assumption. SUTVA is consist of two parts: no interference, which requires the potential outcomes for any unit do not vary with the treatment assigned to other units, and a single value of treatment, which requires the treatment group to have no different forms or versions of each treatment level which could lead to different potential outcomes. For example, for no interference, let's say we are analyzing the causal effect of fertilizer on land yield. The fertilizer we applied on the treatment group may infiltrate nutrients in the control group and affect the potential outcome if we do not separate different sections of land. For a single version of treatment, let's say if we are analyzing the causal effect of tutoring on students' grades. We must have the treatment group receive help from the same tutor for the same amount of time at the same time of the day so that the quality of tutoring can be of the same level for all individuals in the treatment group.

Question 2. Bed Nets and Malaria (20 points)

Article: Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment by Jessica Cohen and Pascaline Dupas

Some economists have argued that "cost-sharing" makes it more likely that a product will be used (versus giving it away for free). Cohen and Dupas partnered with 20 Kenyan prenatal clinics to distribute subsidized anti-malarial bed nets. For each clinic, they varied the extent of the subsidy: either full (free bed-nets, $D_i = 1$) or partial (90% cheaper bed-nets, $D_i = 0$). They measure (among other things) whether women who received bed nets used them (Y_i).

1. What is $\mathbb{E}[Y_i|D_i = 0]$? (5 points)

$\mathbb{E}[Y_i|D_i = 0]$ is the expected value of whether the woman (defined as unit i) use the bed net they received given that their subsidy is partial.

2. What is $\mathbb{E}[Y_i(1)]$? (5 points)

$\mathbb{E}[Y_i(1)]$ is the expected value of the potential outcomes if the woman (defined as unit i) received full subsidy, free bed-nets.

3. What is $\mathbb{E}[Y_i(1)|D_i = 0]$? (5 points)

$\mathbb{E}[Y_i(1)|D_i = 0]$ is the expected value of whether the woman (defined as unit i) will use the bed-nets that comes with full subsidy in a world in which the bed-nets received by the woman is gained by partial subsidy.

4. Cohen and Dupas randomized treatment at the level of the clinic, but the outcomes of interest are at the individual level. Is there a violation of consistency/SUTVA? Why or why not? Argue your case. (5 points)

From the information given, we can see that SUTVA is not violated. This is a cluster-randomized design. The reason is to avoid the spillover effect. If they decide to randomize treatment at the individual level instead of the clinic level, there could be spillover among women who goes to the same clinic. This does not violate SUTVA. It actually ensures that SUTVA holds.

Question 3. Application (Coding) (30 points)

The STAR (Student-Teacher Achievement Ratio) Project is a four year *longitudinal study* examining the effect of class size in early grade levels on educational performance and personal development.

This exercise is in part based on[1]:

Mosteller, Frederick. 1997. "The Tennessee Study of Class Size in the Early School Grades." *Bulletin of the American Academy of Arts and Sciences* 50(7): 14-25.

A longitudinal study is one in which the same participants are followed over time. This particular study lasted from 1985 to 1989 involved 11,601 students. During the four years of the study, students were randomly assigned to small classes, regular-sized classes, or regular-sized classes with an aid. In all, the experiment cost around $12 million. Even though the program stopped in 1989 after the first kindergarten class in the program finished third grade, collection of various measurements (e.g., performance on tests in eighth grade, overall high school GPA) continued through the end of participants' high school attendance.

We will analyze just a portion of this data to investigate whether the small class sizes improved performance or not. The data file name is `STAR.csv`, which is a CSV data file. The names and descriptions of variables in this data set are:

Name	Description
race	Student's race (White = 1, Black = 2, Asian = 3, Hispanic = 4, Native American = 5, Others = 6)
classtype	Type of kindergarten class (small = 1, regular = 2, regular with aid = 3)
g4math	Total scaled score for math portion of fourth grade standardized test
g4reading	Total scaled score for reading portion of fourth grade standardized test
yearssmall	Number of years in small classes
hsgrad	High school graduation (did graduate = 1, did not graduate = 0)

Note that there are a fair amount of missing values in this data set. For example, missing values arise because some students left a STAR school before third grade or did not enter a STAR school until first grade.

1. Create a new factor variable called `kinder` in the data frame. This variable should recode `classtype` by changing integer values to their corresponding informative labels (e.g., change 1 to `small` etc.). Similarly, recode the `race` variable into a factor variable with four levels (`white`, `black`, `hispanic`, `others`) by combining Asians and Native Americans as the `others` category. For the `race` variable, overwrite the original variable in the data frame rather than creating a new one. Recall that `na.rm = TRUE` can be added to functions in order to remove missing data. (5 points)

```
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3      v readr     2.1.4
## v forcats   1.0.0      v stringr   1.5.0
## v ggplot2   3.4.3      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------- tidyverse_conflicts() --
```

---

[1]I have provided you with a sample of their larger dataset. Empirical conclusion drawn from this sample may differ from their article.

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
STAR <- read.csv("STAR2.csv")
# create the "kinder" variable
STAR$kinder <- factor(STAR$classtype, levels = c(1, 2, 3),
                      labels = c("small", "regular", "regular with aid"))

# overwrite the 'race' variable
STAR$race <- recode(STAR$race,
    '1' = 'white',
    '2' = 'black',
    '3' = 'others',
    '4' = 'hispanic',
    '5' = 'others',
    '6' = 'others'
  )

# Convert the 'race' variable to a factor with the desired levels
STAR$race <- factor(STAR$race, levels = c("white", "black", "hispanic", "others"))

# show the first few rows:
head(STAR)
```

```
##    race classtype yearssmall hsgrad g4math g4reading           kinder
## 1 white         2          0     NA     NA        NA          regular
## 2 black         3          0      1     NA        NA regular with aid
## 3 white         1          4      1    707       719            small
## 4 white         1          2      1     NA        NA            small
## 5 black         2          0      0     NA        NA          regular
## 6 black         2          0     NA    674       571          regular
```

Here I didn't use na.rm = TRUE to remove missing data when creating the kinder variable and changing the race variable because I intended to keep the NA values as NA values in the new/updated columns. I can use na.rm = TRUE to eliminate them at any time later when I am doing statistical calculation on them.

2. How does performance on fourth grade reading and math tests for those students assigned to a small class in kindergarten compare with those assigned to a regular-sized class? Do students in the smaller classes perform better? Use means to make this comparison while removing missing values. Give a brief substantive interpretation of the results. To understand the size of the estimated effects, compare them with the standard deviation of the test scores. (10 points)

```r
# Create two new data frames for small class size (classtype = 1) and regular class size (classtype = 2)
small_class_grades <- STAR[STAR$classtype == 1, c("g4math", "g4reading")]
regular_class_grades <- STAR[STAR$classtype == 2, c("g4math", "g4reading")]

# Calculate means and standard deviations for math grades and reading grades in small class size
mean_math_small <- mean(small_class_grades$g4math, na.rm = TRUE)
mean_reading_small <- mean(small_class_grades$g4reading, na.rm = TRUE)

# Calculate means and standard deviations for math grades and reading grades in regular class size
mean_math_regular <- mean(regular_class_grades$g4math, na.rm = TRUE)
mean_reading_regular <- mean(regular_class_grades$g4reading, na.rm = TRUE)

# Calculate overall standard deviations for math and reading scores
sd_math <- sd(STAR$g4math, na.rm = TRUE)
sd_reading <- sd(STAR$g4reading, na.rm = TRUE)

# Print the results
cat("Math grades:\n")
```

```
## Math grades:
```

```r
cat("Mean in small class size:", mean_math_small, "\n")
```

```
## Mean in small class size: 709.0091
```

```r
cat("Mean in regular class size:", mean_math_regular, "\n")
```

```
## Mean in regular class size: 709.6314
```

```r
cat("Difference in Mean (small-regular):", mean_math_small-mean_math_regular, "\n")
```

```
## Difference in Mean (small-regular): -0.6222809
```

```r
cat("Standard Deviation:", sd_math, "\n")
```

```
## Standard Deviation: 43.8088
```

```r
cat("Reading grades:\n")
```

```
## Reading grades:
```

```r
cat("Mean in small class size:", mean_reading_small, "\n")
```

```
## Mean in small class size: 724.5316
```

```r
cat("Mean in regular class size:", mean_reading_regular, "\n")
```

```
## Mean in regular class size: 720.4467
```

```r
cat("Difference in Mean (small-regular):", mean_reading_small-mean_reading_regular, "\n")
```

```
## Difference in Mean (small-regular): 4.084932
```

```r
cat("Standard Deviation:", sd_reading, "\n")
```

```
## Standard Deviation: 52.33846
```

For math, students in regular class size perform better since the mean is higher than that of small class size. On average, regular class size students score 0.6 points higher. For reading, it's the opposite. Students in small class size perform better since the mean is higher than that of regular class size. On average, small class size students score 4 points higher.

To understand the size of the estimated effects, we compare the means with the standard deviations. We may conclude it as a substantial effect if the difference in means is larger than one standard deviation. For math, with a standard deviation of 43.8 and an estimated effect of -0.62, we can conclude there is no substantial difference between the means. Thus, students in regular classes are just performing slightly better than those in small classes. However, for reading, with a standard deviation of 52.3, the increase in average reading score of 4.08 is better compared to the difference in means they have in math scores, though still not a substantial effect yet.

3. Instead of comparing just average scores of reading and math tests between those students assigned to small classes and those assigned to regular-sized classes, look at the entire range of possible scores. To do so, compare a high score, defined as the 66th percentile, and a low score (the 33rd percentile) for small classes with the corresponding score for regular classes. These are examples of *quantile treatment effects*. Does this analysis add anything to the analysis based on mean in the previous question? (Hint: You will use the quantile() function in r.) (5 points)

```r
# Calculate percentiles for small class math scores
small_class_math_scores <- STAR$g4math[STAR$classtype == 1]
high_percentile_small_math <- quantile(small_class_math_scores, probs = 0.66, na.rm = TRUE)
low_percentile_small_math <- quantile(small_class_math_scores, probs = 0.33, na.rm = TRUE)

# Calculate percentiles for regular class math scores
regular_class_math_scores <- STAR$g4math[STAR$classtype == 2]
high_percentile_regular_math <- quantile(regular_class_math_scores, probs = 0.66, na.rm = TRUE)
low_percentile_regular_math <- quantile(regular_class_math_scores, probs = 0.33, na.rm = TRUE)

# Calculate percentiles for small class reading scores
small_class_reading_scores <- STAR$g4reading[STAR$classtype == 1]
high_percentile_small_reading <- quantile(small_class_reading_scores, probs = 0.66, na.rm = TRUE)
low_percentile_small_reading <- quantile(small_class_reading_scores, probs = 0.33, na.rm = TRUE)

# Calculate percentiles for regular class reading scores
regular_class_reading_scores <- STAR$g4reading[STAR$classtype == 2]
high_percentile_regular_reading <- quantile(regular_class_reading_scores, probs = 0.66, na.rm = TRUE)
low_percentile_regular_reading <- quantile(regular_class_reading_scores, probs = 0.33, na.rm = TRUE)

# Combine the percentile scores into a vector
high_scores <- c(
  high_percentile_small_math,
  high_percentile_regular_math,
  high_percentile_small_reading,
  high_percentile_regular_reading
)

low_scores <- c(
  low_percentile_small_math,
  low_percentile_regular_math,
  low_percentile_small_reading,
  low_percentile_regular_reading
)

# Print the percentile scores
cat("66th Percentile Scores:\n")
```

```
## 66th Percentile Scores:
```

```r
cat("(small math, regular math, small reading, regular reading)\n")
```

```
## (small math, regular math, small reading, regular reading)
```

```r
cat(high_scores, "\n")
```

```
## 727 724.64 741 740
```

```r
cat("33rd Percentile Scores:\n")
```

```
## 33rd Percentile Scores:
```

```r
cat("(small math, regular math, small reading, regular reading)\n")
```

```
## (small math, regular math, small reading, regular reading)
```

```r
cat(low_scores, "\n")
```

```
## 695 696 708 705
```

This time, we can see at the 66th percentile, for math, small class students score 2.36 points higher than regular class students, which is against our previous conclusion when only considering the difference in means. Specifically, our previous conclusion is that, on average, regular class size students score 0.6 points higher. This means that small class students will score better in the high score range and such difference is not only the opposite but also a lot larger than the average difference.

For reading, small class students score 1 point higher than regular class students, which aligns with our previous conclusion that, on average, small class students score 4.08 point higher. We can add another conclusion that the advantage of small class students in reading decreases in the high score range.

At the 33th percentile, for math, regular class students score 1 point higher than small class students, which aligns with our previous conclusion that, on average, regular class students score 0.6 point higher. We can add another conclusion that the advantage of regular class students in math remains roughly the same in the low score range.

For reading, small class students score 3 point higher than regular class students, which aligns with our previous conclusion that, on average, small class students score 4.08 point higher. We can add another conclusion that the advantage of small class students in math remains roughly the same in the low score range.

Overall, the difference in means is a better indicator for lower scores than higher scores because the lower scores difference pattern is the same to the difference in means but when it comes to higher scores, either the pattern is less obvious or is even reversed.

4. We examine whether the STAR program reduced the achievement gaps across different racial groups. Begin by comparing the average reading and math test scores between white and minority students (i.e., Blacks and Hispanics) among those students who were assigned to regular classes with no aid. Conduct the same comparison among those students who were assigned to small classes. Give a brief substantive interpretation of the results of your analysis. (5 points)

```r
# math score of white vs minority students in a regular class with no aid
no_aid_white_math <- STAR$g4math[STAR$classtype == 2 & STAR$race == 'white']
mean_no_aid_white_math <- mean(no_aid_white_math, na.rm = TRUE)
no_aid_minority_math <- STAR$g4math[STAR$classtype == 2 & (STAR$race == 'black' | STAR$race == 'hispani
mean_no_aid_minority_math <- mean(no_aid_minority_math, na.rm = TRUE)
no_aid_math_diff <- mean_no_aid_white_math - mean_no_aid_minority_math

# math score of white vs minority students in a small class
small_white_math <- STAR$g4math[STAR$classtype == 1 & STAR$race == 'white']
mean_small_white_math <- mean(small_white_math, na.rm = TRUE)
small_minority_math <- STAR$g4math[STAR$classtype == 1 & (STAR$race == 'black' | STAR$race == 'hispanic
mean_small_minority_math <- mean(small_minority_math, na.rm = TRUE)
small_math_diff <- mean_small_white_math - mean_small_minority_math

# reading score of white vs minority students in a regular class with no aid
no_aid_white_reading <- STAR$g4reading[STAR$classtype == 2 & STAR$race == 'white']
mean_no_aid_white_reading <- mean(no_aid_white_reading, na.rm = TRUE)
no_aid_minority_reading <- STAR$g4reading[STAR$classtype == 2 & (STAR$race == 'black' | STAR$race == 'h
mean_no_aid_minority_reading <- mean(no_aid_minority_reading, na.rm = TRUE)
no_aid_reading_diff <- mean_no_aid_white_reading - mean_no_aid_minority_reading


# reading score of white vs minority students in a small class
small_white_reading <- STAR$g4reading[STAR$classtype == 1 & STAR$race == 'white']
mean_small_white_reading <- mean(small_white_reading, na.rm = TRUE)
small_minority_reading <- STAR$g4reading[STAR$classtype == 1 & (STAR$race == 'black' | STAR$race == 'hi
mean_small_minority_reading <- mean(small_minority_reading, na.rm = TRUE)
small_reading_diff <- mean_small_white_reading - mean_small_minority_reading

# Combine the percentile scores into a vector
math_racial_diff  <- c(
  no_aid_math_diff,
  small_math_diff
)

reading_racial_diff <- c(
  no_aid_reading_diff,
  small_reading_diff
)
# Print the differences in scores
cat("Math mean difference(white-minority):\n")
```

```
## Math mean difference(white-minority):
```

```r
cat("(no aid, small)\n")
```

```
## (no aid, small)
```

```r
cat(math_racial_diff, "\n")
```

```
## 12.34691 14.7655
```

```r
cat("Reading mean difference(white-minority):\n")
```

```
## Reading mean difference(white-minority):
```

```r
cat("(no aid, small)\n")
```

```
## (no aid, small)
```

```r
cat(reading_racial_diff, "\n")
```

```
## 33.05486 29.53008
```

In regular classes with no aid, the average math score of white students is higher than the average score of minority students by 12.3. In small classes, the average math score of white students is higher than the average score of minority students by 14.7. We can see that the gaps in math scores across different racial groups are reduced with regular classes with no aid.

In regular classes with no aid, the average reading score of white students is higher than the average score of minority students by 33. In small classes, the average reading score of white students is higher than the average score of minority students by 29.5. We can see that the gaps in reading scores across different racial groups are reduced with small classes.

According to the data we collected, in reading scores of white and minority students, we can see the gap is reduced, which indicates that the STAR program is helpful in reducing achievement gaps across different racial groups in terms of reading scores. However, in math scores, the STAR program is harmful rather than helpful since regular class with no aid possesses a smaller gap.

5. We consider the long term effects of kindergarten class size. Compare high school graduation rates across students assigned to different class types. Also, examine whether graduation rates differ by the number of years spent in small classes. Finally, as done in the previous question, investigate whether the STAR program has reduced the racial gap between white and minority students' graduation rates. Briefly discuss the results. (5 points)

```
# high school graduation rates across students assigned to different class types
tapply(STAR$hsgrad, STAR$kinder, mean, na.rm = TRUE)
```

```
##           small         regular regular with aid
##       0.8477306       0.8225214        0.8358586
```

The high school graduation rate is 84.77% for students assigned to small classes, 82.25% for students assigned to regular classes with no aid, and 83.58% for students assigned to regular classes with aid. Students in small classes have the highest high school graduation rate. And, regular classes with aid have a higher high school graduation rate than regular classes without aid.

```
# graduation rates baesd on the number of years spent in small classses
tapply(STAR$hsgrad, STAR$yearssmall, mean, na.rm = TRUE)
```

```
##         0         1         2         3         4
## 0.8250000 0.7979798 0.8518519 0.8269231 0.8833747
```

The high school graduation rate is 82.5% for students who were assigned to small classes for zero years, 79.79% for those who were assigned to small classes for one year, 85.18% for those who were assigned to small classes for two years, 82.69% for those who were assigned for three years, and 88.33% for those who were assigned for four years. From the graduation rate we've calculated, we could observe that for students who are once in small classes, the longer they stayed in small classes, the higher the high school graduation rate is. Hence, we could conclude that the STAR program has helped increase the high school graduation rate.

```
# split data into different subsets to make our calculation easier
white <- subset(STAR, race=="white")
minority <- subset(STAR, race =="black" | race =="hispanic")
# the difference in high school graduation rate between white vs minority students across different cla
tapply(white$hsgrad, white$kinder, mean, na.rm = TRUE)-tapply(minority$hsgrad, minority$kinder, mean, n
```

```
##           small         regular regular with aid
##       0.1005110       0.1214189        0.1387456
```

The racial gap between white and minority students' graduation rates is the widest in regular classes with aid, with a difference of 13.87%. The difference in graduation rate between white students and minority students in regular classes without aid is 12.14%. And, the difference in graduation rate between white students and minority students is 10.05% in small classes. We can conclude that the STAR program does not reduce the racial gap between white and minority students' graduation rates. In fact, it increases the racial gap.

Question 4. Design Your Experiment (30 points)

Design your own experiment from start to finish. Choose an *interesting* question. Explain why observational data may give you the wrong answer. Detail the potential outcomes and a well-defined treatment. Explain the type of experiment (completely random, cluster-design, block/stratified). Will your design ensure a causal treatment effect? (Remember: Be as specific as possible and give examples.)

My experiment is on the causal effect of a new vaccine on a treatment of a disease. It's not observational data because it may give us the wrong answer because there are many factors that could affect the disease, such as the population, hygiene condition, wealth, physical and mental health, etc. Thus, we cannot derive causal inference from observational data since none of the other factors are controlled.

The potential outcome is the outcome for an individual under a potential treatment. In this case, we can define the potential outcome as the affect rate of patients under treatment or control.

I plan to implement a completely randomized experiment. In my experiment design, the unit of analysis is a group of people selected randomly. The treatment is taking the vaccine shot. The control is not taking any vaccine shots. We need to make sure that other factors are held constant. For example, these people should be same age, same race, same sex, etc with similar health conditions. Then, we randomly assign treatment and control to our units of analysis (using coin flips, tossing a head means the unit will receive treatment, and tossing a tail means the unit will receive control).

With this experiment design, all assumptions which are required in order to ensure a causal treatment effect are satisfied. SUTVA is satisfied because we have no interference as patients receive vaccine shots separately, which cannot affect the state of other patients, and we also have single value of treatment since we use vaccine from the same company. The quality of the treatment is at the same level for all patients that received treatment. Ignorability is satisfied because the random assignment makes the treatment assignment independent of the potential outcome. Also, the positivity assumption is met because we use coin flips to decide whether our units receive treatment. That means each unit has a 50% chance of receiving treatment and a 50% chance of receiving control.