Problem Set 1: Solutions

Anita Ye - yy3557 - 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 your-NetID_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)

Only one of the potential outcomes can be observed $Y_i(1)$ and $Y_i(0)$, and it is impossible to directly observe both in real world. Which means that we are always unable to observe the condition with or without the treatment. For example, we want to learn the causal relationship between alcohol consumption and liver disease. When we are trying to reveal the relationship, it is not possible to identify accurately what happened to the group of individuals who drinks if they had not been doing so. Such unboserved outcome is the counterfactual, meaning that we could never observe it in this world.

2. Why are experiments important? (5 points)

Experiments are important because the randomization allows it to test hypothesis with minimized bias, along with controlled conditions. These all ensures experiment are replicable, which guarantees randomness and reliability. This means that the treatment assignment is always controlled by the person conducting the experiment, who will make sure it satisfies positivity, ignorability, and SUTVA, also controlling the counfounding variables. It reveals causal relationships with probability by collecting data.

For example, we want to know if doing homework leads to better grades, then we can design and experiment. We will need to control the other variables, for example, the teacher who is teaching, and the materials learned. In this case, the students sample could be taken from ONE class period, for example. After controlling the confounding variables, we can randomly select students to split into two groups, one with homework and one not, then observe the results. Experiments leads us to causal relationships because of such randomness that limits any bias.

3. What does ignorability mean? (5 points)

Ignorability is the assumption indicating that given the observed variables, the treatment assignment is independent of the potential outcomes, then the observed and counterfactural results are the same. With the concept of ignorability in mind, we can identify if the outcomes are indepedent using random assignment. It is used in experience with randomly assigned treatment, and is important in identifying causal effects. We still need to be careful of the limitation that ignorability relies on assumption that all relevant confounders are fully discovered and measured.

For example, consider a new drug on covid-19, we can design and experiment to test out its effectiveness. Other variables, such as age, gender, health condition, are independent to the assignment of individuals receiving the new drug is independent of their potential outcomes. This means that the probability of receiving treatment should be independent of all the conditions, and such assignment is unrelated.

4. What is SUTVA? (5 points)

SUTVA (Stable Unit Treatment Value Assumption) implies that there is no spillover or interference effects, and emphasizes the isolation of individuals' potential outcomes regarding the outcomes of other individuals. The observed outcome is the potential outcome of the observed treatment, so $Y_i(d) = Y_i$ if $D_i = d$. Individuals are not influenced by the treatment assignment of other individuals, meaning tha there should not be any interference. This is important because it guarantees randomness and controlled groups that dedicates to an accurate estimation of each individual independently. Another key to SUTVA is consistency, which means a treatment is of single version, one and only outcome.

Consider the experiment in q2, if we are studying how effective doing homework is for students learning. SUTVA here indicates that the resulting outcome is determined only by doing this type of homework, and not affected by any other practice they are exposed to. Outcomes of each individual are independent of how other individuals are given, and they do not influence each other's result. The treatment effects are isolated as to each individuals.

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 outcomes of whether women who received bed nets used them, given that the subsidy is the 90% cheaper bed nets. This will lead us to the average use of the 90% cheaper bed nets.

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

$\mathbb{E}[Y_i(1)]$ is the expected outcomes of whether women received free bed nets used them if all get full subsidy of free bed nets, which lead to the average of the use of the 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 outcome of whether women will use the bed nets if those who purchased 90% chepaer bed nets are given free bed nets instead.

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)

On considering the outcomes of interest are at the individual level, since SUTVA assumes no spillover effect and that the potential outcome is independent of other individuals, and based on this statement, consistency is not violated. Since outcomes for individuals are independent of each other, it has no interference so it does not violate consistency . The treatment is also split into two, the 90% cheaper bed net, and the free bed nets. This means that it follows the single version rule, and therefore SUTVA is not violated. The assignment of such treatment is also independent of any other intervention.

However, it is still possible to violate. randomization at the clinic level may not be enough and the assumptions may not hold in some cases. It may be same for all inside a clinic, but different clinic may conduct it differently, and this could potentially make a difference. Individuals are also not limited to one clinic, and they may not get both treatment. In reality, individuals may consult among each other, and they could, for example, convince their friends to use the bed net, and then the friends becomes influenced and therefore we have spillover effects. Since we have not restricted such interaction and communication among individuals, it is possible that there is a violation.

In conclusion, it does not violate SUTVA on large extend, but we may need to examine more details and be more careful to make sure th experiment rules out any other factors that may influence SUTVA.

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)

```
#install.packages("tidyverse")
setwd('/Users/anita/Documents/Causal Inference/HW/HW1')
getwd()
```

```
## [1] "/Users/anita/Documents/Causal Inference/HW/HW1"
```

```
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")
library(dplyr)

# Recode classtype into kinder
STAR <- STAR %>%
  mutate(kinder = recode(classtype,
                         "1" = "small",
                         "2" = "regular",
                         "3" = "regular with aid"))

# Recode race into a factor variable with four levels
STAR <- STAR %>%
  mutate(race = factor(recode(race, "1" = "White", "2"= "Black", "4" = "Hispanic", "3" = "Others", "5" =
                       levels = c("White", "Black", "Hispanic", "Others")

    )
  )
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `race = factor(...)`.
## Caused by warning:
## ! Unreplaced values treated as NA as `.x` is not compatible.
## Please specify replacements exhaustively or supply `.default`.
```

```r
#print out the head of file
print(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
```

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 dataframes for small and regular class
small = STAR%>%
  filter(kinder == "small")
regular = STAR%>%
  filter(kinder == "regular")

#means of math
mean_small_math = mean(small$g4math, na.rm = TRUE)
mean_regular_math = mean(regular$g4math, na.rm=TRUE)
cat("mean of math in small class:", mean_small_math, "\n")
```

```
## mean of math in small class: 709.0091
```

```
cat("mean of math in regular class:" , mean_regular_math, "\n")
```

mean of math in regular class: 709.6314

```
#means of reading
mean_small_reading = mean(small$g4reading, na.rm=TRUE)
mean_regular_reading = mean(regular$g4reading, na.rm=TRUE)
cat("mean of reading in small class:",mean_small_reading, "\n")
```

mean of reading in small class: 724.5316

```
cat("mean of reading in regular class:",mean_regular_reading, "\n")
```

mean of reading in regular class: 720.4467

```
#sd of math and reading as a whole
sd_math = sd(STAR$g4math, na.rm=TRUE)
sd_reading = sd(STAR$g4reading, na.rm=TRUE)
cat("standard deviation of math grades", sd_math, "\n")
```

standard deviation of math grades 43.8088

```
cat("standard deviation of reading grades", sd_reading, "\n")
```

standard deviation of reading grades 52.33846

```
#sd of math
sd_small_math = sd(small$g4math, na.rm = TRUE)
sd_regular_math = sd(regular$g4math, na.rm=TRUE)
cat("standard deviation of math in small class:", sd_small_math, "\n")
```

standard deviation of math in small class: 45.3054

```
cat("standard deviation of math in regular class:" , sd_regular_math, "\n")
```

standard deviation of math in regular class: 40.81335

```
#sd of reading
sd_small_reading = sd(small$g4reading, na.rm = TRUE)
sd_regular_reading = sd(regular$g4reading, na.rm=TRUE)
cat("standard deviation of reading in small class:", sd_small_reading, "\n")
```

standard deviation of reading in small class: 51.48347

```
cat("standard deviation of reading in regular class:" , sd_regular_reading, "\n")
```

standard deviation of reading in regular class: 54.30644

Interpretation:

For the math scores, both class sizes are close in mean value while regular has sightly higher. Since the mean difference is less than that of standard deviation, it is likely that such difference is normal and within the variability.

For reading scores, small class has a higher than regular score than the regular sized class. However, since the mean difference is still less than the standard deviation, this difference is acceptable and within the variability.

The mean difference is 724.5316-720.4467 = 4.0849 for math, and 709.6314-709.0091 for reading, which is less than one. The effect size is also not significant enough. Since both the means are quite close and the difference falls within the range, we can conclude that the class size does not have an effect significant enough on the test scores.

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
#math scores
p_math_small = quantile(small$g4math, probs = c(0.33, 0.66), na.rm = TRUE)
p_math_regular = quantile(regular$g4math, probs = c(0.33, 0.66), na.rm = TRUE)

#quantiles for math scores
cat("33rd Percentile math Score for Small Class:", p_math_small[1], "\n")
```

```
## 33rd Percentile math Score for Small Class: 695
```

```r
cat("66th Percentile math Score for Small Class:", p_math_small[2], "\n")
```

```
## 66th Percentile math Score for Small Class: 727
```

```r
cat("33rd Percentile math Score for Regular Class:", p_math_regular[1], "\n")
```

```
## 33rd Percentile math Score for Regular Class: 696
```

```r
cat("66th Percentile math Score for Regular Class:", p_math_regular[2], "\n")
```

```
## 66th Percentile math Score for Regular Class: 724.64
```

```r
#reading scores
p_reading_small = quantile(small$g4reading, probs = c(0.33, 0.66), na.rm = TRUE)
p_reading_regular = quantile(regular$g4reading, probs = c(0.33, 0.66), na.rm = TRUE)

#quantiles for reading scores
cat("33rd Percentile Reading Score for Small Class:", p_reading_small[1], "\n")
```

```
## 33rd Percentile Reading Score for Small Class: 708
```

```r
cat("66th Percentile Reading Score for Small Class:", p_reading_small[2], "\n")
```

```
## 66th Percentile Reading Score for Small Class: 741
```

```r
cat("33rd Percentile Reading Score for Regular Class:", p_reading_regular[1], "\n")
```

```
## 33rd Percentile Reading Score for Regular Class: 705
```

```r
cat("66th Percentile Reading Score for Regular Class:", p_reading_regular[2], "\n")
```

```
## 66th Percentile Reading Score for Regular Class: 740
```

Interpretation:

The low score for small class in math is slightly higher than that of regular class, while the opposite for reading. For higher score, smaller class has higher for both, but not more than 5.

Most of the cases, smaller classses have a slightly higher score than regular sized classes for all percentile except from 33rd percentile for math score. It is tempting to say that the smaller class is more effective; however, the values are still too close to make a conclusion. Therefore, it does not add much to the previous analysis and remains similar results.

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
#regular no aid = regular

#regular class
white_regular = STAR %>%
  filter(race == "White", kinder == "regular")
mean_white_regular_math = mean(white_regular$g4math, na.rm=TRUE)
mean_white_regular_reading = mean(white_regular$g4reading, na.rm=TRUE)

minority_regular = STAR %>%
  filter(race == c("Black", "Hispanic"), kinder == "regular")
```

```
## Warning: There were 2 warnings in `filter()`.
## The first warning was:
## i In argument: `race == c("Black", "Hispanic")`.
## Caused by warning in `==.default`:
## ! longer object length is not a multiple of shorter object length
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```r
mean_minority_regular_math = mean(minority_regular$g4math, na.rm=TRUE)
mean_minority_regular_reading = mean(minority_regular$g4reading, na.rm=TRUE)

cat("Mean math score of white students in regular class:", mean_white_regular_math, "\n")
```

```
## Mean math score of white students in regular class: 711.4358
```

```r
cat("Mean math score of minority students in regular class:", mean_minority_regular_math, "\n")
```

```
## Mean math score of minority students in regular class: 704.3958
```

```r
cat("Mean reading score of white students in regular class:", mean_white_regular_reading, "\n")
```

```
## Mean reading score of white students in regular class: 725.3104
```

```r
cat("Mean reading score of minority students in regular class:", mean_minority_regular_reading, "\n")
```

```
## Mean reading score of minority students in regular class: 696
```

```r
white_small = STAR %>%
  filter(race == "White", kinder == "small")
mean_white_small_math = mean(white_small$g4math, na.rm=TRUE)
mean_white_small_reading = mean(white_small$g4reading, na.rm=TRUE)

minority_small = STAR %>%
  filter(race == c("Black", "Hispanic"), kinder == "small")
```

```
## Warning: There were 2 warnings in `filter()`.
## The first warning was:
## i In argument: `race == c("Black", "Hispanic")`.
## Caused by warning in `==.default`:
## ! longer object length is not a multiple of shorter object length
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```r
mean_minority_small_math = mean(minority_small$g4math, na.rm=TRUE)
mean_minority_small_reading = mean(minority_small$g4reading, na.rm=TRUE)

cat("Mean math score of white students in small class:", mean_white_small_math, "\n")
```

```
## Mean math score of white students in small class: 711.103
```
```
cat("Mean math score of minority students in small class:", mean_minority_small_math, "\n")
```

```
## Mean math score of minority students in small class: 702.7222
```
```
cat("Mean reading score of white students in small class:", mean_white_small_reading, "\n")
```

```
## Mean reading score of white students in small class: 728.8264
```
```
cat("Mean reading score of minority students in small class:", mean_minority_small_reading, "\n")
```

```
## Mean reading score of minority students in small class: 706.9444
```

Interpretation:

For regular classes, white students have higher scores in both math and reading grades than the minority students, by around 7 and more than 20. While such differences are not very prominent, the gap in reading scores are larger than that of math grades.

For small classes, white students also have a higher score in both math and reading grades than the minority students, by around 10 in math and over 20 for math. While the gap is still larger in reading, it is smaller than that in regular classes.

Comparing the effect of class sizes within the races, the differences are minor except reading scores in small classes for minority students are a bit higher than in regular classes. For other ones, the gaps are too small to determine the effect.

In conclusion, white students tend to perform better in both math or reading, in both smaller or regular classes compared to minority students. Such difference is slightly more obvious in reading scores compared to math scores. We notice that the mean score is higher when it comes to white students, but the trend does indicate a higher score overall. Such difference is much more obvious, while could still not be significant enough to make a conclusion or determine relationship since the gap is not large, and we may need further investigation.

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)

6. graduation rates by class type

```
graduation_rates_by_class <- STAR %>%
  group_by(kinder) %>%
  summarize(graduation_rate = mean(hsgrad, na.rm=TRUE))
print(graduation_rates_by_class)
```

```
## # A tibble: 3 x 2
##   kinder           graduation_rate
##   <chr>                      <dbl>
## 1 regular                    0.823
## 2 regular with aid           0.836
## 3 small                      0.848
```

Interpretation: The class size seemed that it does not impact the graducation rate, since the difference is quite small to be significant.

2. years spent in small classes

```
# Calculate graduation rates by years in small classes
graduation_rates_by_years_small <- STAR %>%
```

```
  group_by(yearssmall) %>%
  summarize(graduation_rate = mean(hsgrad, na.rm=TRUE))
print(graduation_rates_by_years_small)
```

```
## # A tibble: 5 x 2
##   yearssmall graduation_rate
##        <int>           <dbl>
## 1          0           0.825
## 2          1           0.798
## 3          2           0.852
## 4          3           0.827
## 5          4           0.883
```

Interpretation: It seemed that the year spent also has very minor effect on the graduation rate, as graduation rate fluctuates and does not show a clear growing or declining relationship as the year increase or decrease.

3.racial gap

```
grad_race_gap_class <- STAR %>%
  mutate(race_group = case_when(race == "White" ~ "White",
  race %in% c("Black", "Hispanic") ~ "Minority"
)) %>%
filter(race_group %in% c("White", "Minority")) %>%
filter(kinder %in% c("small", "regular")) %>%
group_by(kinder, race_group) %>%
summarise(graduation_rate = mean(hsgrad, na.rm = TRUE), .groups = "drop")
print(grad_race_gap_class)
```

```
## # A tibble: 4 x 3
##   kinder  race_group graduation_rate
##   <chr>   <chr>                <dbl>
## 1 regular Minority             0.735
## 2 regular White                0.856
## 3 small   Minority             0.771
## 4 small   White                0.872
```

```
grad_white_small = STAR %>%
filter(race == "White", kinder == "small")
```

```
grad_white_regular = STAR %>%
filter(race == "White", kinder == "regular")
```

```
grad_white_aid = STAR %>%
filter(race == "White", kinder == "regular with aid")
```

```
grad_minority_small = STAR %>%
filter(race == c("Black", "Hispanic"), kinder == "small")
```

```
## Warning: There were 2 warnings in `filter()`.
## The first warning was:
## i In argument: `race == c("Black", "Hispanic")`.
## Caused by warning in `==.default`:
## ! longer object length is not a multiple of shorter object length
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

10

```r
grad_minority_regular = STAR %>%
filter(race == c("Black", "Hispanic"), kinder == "regular")
```

```
## Warning: There were 2 warnings in `filter()`.
## The first warning was:
## i In argument: `race == c("Black", "Hispanic")`.
## Caused by warning in `==.default`:
## ! longer object length is not a multiple of shorter object length
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```r
grad_minority_aid = STAR %>%
filter(race == c("Black", "Hispanic"), kinder == "regular with aid")
```

```
## Warning: There were 2 warnings in `filter()`.
## The first warning was:
## i In argument: `race == c("Black", "Hispanic")`.
## Caused by warning in `==.default`:
## ! longer object length is not a multiple of shorter object length
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

```r
mean_grad_white_small = mean(grad_white_small$hsgrad, na.rm = TRUE)
print(mean_grad_white_small)
```

```
## [1] 0.8715953
```

```r
mean_grad_minority_small = mean(grad_minority_small$hsgrad, na.rm = TRUE)
print(mean_grad_minority_small)
```

```
## [1] 0.7272727
```

```r
print(mean_grad_white_small - mean_grad_minority_small)
```

```
## [1] 0.1443226
```

```r
mean_grad_white_regular = mean(grad_white_regular$hsgrad, na.rm = TRUE)
print(mean_grad_white_regular)
```

```
## [1] 0.8559322
```

```r
mean_grad_minority_regular = mean(grad_minority_regular$hsgrad, na.rm = TRUE)
print(mean_grad_minority_regular)
```

```
## [1] 0.7297297
```

```r
print(mean_grad_white_regular - mean_grad_minority_regular)
```

```
## [1] 0.1262025
```

```r
mean_grad_white_aid = mean(grad_white_aid$hsgrad, na.rm = TRUE)
print(mean_grad_white_aid)
```

```
## [1] 0.8739054
```

```r
mean_grad_minority_aid = mean(grad_minority_aid$hsgrad, na.rm = TRUE)
print(mean_grad_minority_aid)
```

```
## [1] 0.6981132
```

```r
print(mean_grad_white_aid - mean_grad_minority_aid)
```

```
## [1] 0.1757922
```

Interpretation: The overall average stayed constant with the previous conclusion that white students tend to do better, and it seemed that the program does not eliminate the racial gaps much. we can see that the differences are not large enough to conclude a large gap, yet it aligns with the previous analysis of white and minority groups that white students are better in performance. The third analysis is more obvious in data compared to the previous two, and we can possibly say that white students have higher graduation rate overall. The difference among the means of two groups in different class size, however, is still not large enough to conclude a significant relationship. Therefore, the racial gap remains and is not resolved by STAR program.

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.)

Question:

Does Meditation have an impact on Stress Reduction?

Design:

Recruit 100 participants randomly, ensuring that they have diverse background and conditions, such as ages and ethnithity. Randomly assign the participants into the control group and treatment group in the RCT. For the observational study, participants will choose themselves into either groups on their own preference.

(To showcase how the observational study is conducted, I also states what observational practice could be in parentencies, while observational study is not conducted since we should not have such multiple groups.)

Participants in the treatment group will conduct in a meditation session for 15 minutes every day, and such treatment continues for 5 weeks. The control group will not have any meditation session and just keeps their regular life. This is going to be a completely random experiment.

We will record stress levels using a standardized scale after the 5 weeks of meditation for the treatment group($Y(1)$), who receive the meditation treatment, and using the same standardized scale for the controlled group($Y(0)$)who do not receive them also after 5 weeks. The scale could be a small questionnare, asking participants to answer questions by rating their stress level.

(Participants in observational group will self-select whether to engage in meditation or maintain their regular routines.)

We will determine the effect by measuring the stress level in the beginning of the study with the same standardized scale, and then at the end of the 5 week treatment.

(Stress levels observed in the observational study for individuals who self-select into either the treatment or control group.)

Causal relationship

RCT ensures a causal effect because of the random assignment of participants into treatment and control groups. This eliminates the confounding variables. On the other hand, observational studies will not be as firm as the RCT treatments. However, in observational study, out finding is affected by biases and confounding variables that are not controlled, and the self-selection and a lack of randomness may result in reverse causality.

Why Observational Data May Give the Wrong Answer:

Observational data are affected by confounding variables and selection biases that has not been ruled out by randomness. Therefore, the true effect of the treatment is hindered. Self selection makes it hard to determine if the two groups started with different stress levels, which lead to them choosing one group over the other. Such selection is biased, and observational study does not eliminate the effect of it, and this will affect the conclusion of the causal relationship or the effectiveness of the treatment.