

Problem Set 1: Solutions

Ghina Al Shdaifat - gha2009 - 004 (Collaborated with Molshree Singhal - ms12483)

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)

The fundamental problem of causal inference is that we can never observe both potential outcomes - the observed and counterfactual reality - under the same state of the world. In more other words, for any individual, at a given point in time, we can observe either the treatment effect or the control effect, but not both. This makes it challenging to directly measure the causal effect of a treatment.

Example: A patient who takes a new drug and recovers from an illness -> We can never truly know for certain whether the patient would have recovered had they not taken the drug as we are unable to observe this counterfactual reality.

2. Why are experiments important? (5 points)

Experiments address the fundamental problem of causal inference and obey the requirements of ATE estimation: positivity (where each unit has a positive probability of receiving any treatment) and ignorability (where the probability of receiving treatment is independent of potential outcomes). Experiments, especially randomized controlled experiments, allow for randomization in assigning participants to either the treatment or control group, which prevents biases of the treatment effect estimate.

Example: In testing the effectiveness of a new vaccine, if participants are randomly assigned to either receive the vaccine or a placebo, any difference in outcomes between the two groups can be attributed to the vaccine, not to pre-existing differences between the groups.

3. What does ignorability mean? (5 points)

Ignorability, often referred to as “unconfoundedness” or “exchangeability,” means that given a set of observed covariates, the assignment to the treatment is independent of the potential outcomes.

Example: Studying the effect of tutoring on academic performance -> ignorability is satisfied if, after controlling for factors like prior academic achievement, motivation, and socio-economic status, the decision to receive tutoring is random with respect to potential academic outcomes.

4. What is SUTVA? (5 points)

SUTVA stands for Stable Unit Treatment Value Assumption. It is a consistency assumption, in which it assumes that the potential outcome of the observed outcome is the same as the observed outcome. In other words, the assumption states that:

- The potential outcome for any unit (e.g., an individual) does not depend on the treatment assignment of other units.
- There is only one version of treatment and one version of control, meaning the treatment effect is consistent -> no interference or spillover

Example: A vaccination campaign in a community -> If vaccinated individuals can prevent disease transmission to unvaccinated individuals, then the health outcome of an unvaccinated person might depend on whether others in their community are vaccinated. Such an example would be a violation of SUTVA.

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)

$\mathbb{E}[Y_i|D_i = 0]$ refers to the expected value (or average) of the outcome Y_i (whether the bed nets are used or not) given that the treatment D_i was 0 (i.e., women received a 90% subsidy instead of getting the bed net for free), where $D_i = 0$ is receiving a partial subsidy, and $D_i = 1$ is receiving a full subsidy.

In other words, $\mathbb{E}[Y_i|D_i = 0]$ represents the probability of bed net usage among women who received the bed net at a 90% subsidy.

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

$\mathbb{E}[Y_i(1)]$ represents the expected value of the potential outcome ($Y_i(1)$) if the treatment (D_i) were set to 1 for all individuals, regardless of their actual treatment assignment.

In the context of the study, $\mathbb{E}[Y_i(1)]$ is the expected (or average) number of bed net usage if all women were given the bed nets for free (i.e., if every woman were treated with $D_i = 1$), even those who actually received a 90% subsidy in the study.

This concept is a counterfactual: for the group that did not receive the bed nets for free (those with $D_i = 0$), we do not actually observe $Y_i(1)$. However, the idea is to imagine what would have happened on average if everyone had been given the bed nets for free.

In the study, researchers would look for the empirical average of bed net usage in the group that received the bed nets for free to approximate $\mathbb{E}[Y_i(1)]$.

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

$\mathbb{E}[Y_i(1)|D_i = 0]$ represents a specific counterfactual scenario. It refers to the expected value (or average) of the potential outcome $Y_i(1)$ if all individuals in the group that received the 90% subsidy (i.e., $D_i = 0$) had instead been given the bed nets for free.

$\mathbb{E}[Y_i(1)|D_i = 0]$ represents the fundamental problem of causation as we do not actually observe this scenario in the data. In other words, the women who received the 90% subsidy did not actually receive the bed nets for free. However, if the sample size is large enough and there are no violations of the randomization process, $\mathbb{E}[Y_i(1)|D_i = 0]$ would be similar to $\mathbb{E}[Y_i(1)|D_i = 1]$ because the groups are on average statistically similar in all aspects, except for the treatment they received.

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)

Given that the treatment was randomized at the clinic level but the outcomes are measured at the individual level, there is a potential for SUTVA violation due to intra-clinic interactions or spillovers. If one woman sees another using the bed net or discusses its benefits with her peers, it might influence her own likelihood of

using the bed net. This means that individual outcomes might not be entirely independent of the treatment assignments of others in the same clinic.

To further validate a SUTVA violation or not, one would need to identify specific details of the study design and context, such as the potential existence of social ties between individuals that might lead to influence.

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 on¹:

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 |
|-------------------------|---|
| <code>race</code> | Student's race (White = 1, Black = 2, Asian = 3, Hispanic = 4, Native American = 5, Others = 6) |
| <code>classtype</code> | Type of kindergarten class (small = 1, regular = 2, regular with aid = 3) |
| <code>g4math</code> | Total scaled score for math portion of fourth grade standardized test |
| <code>g4reading</code> | Total scaled score for reading portion of fourth grade standardized test |
| <code>yearssmall</code> | Number of years in small classes |
| <code>hsgrad</code> | 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  1.9.2      v tidyr      1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## 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
```

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

```

#Read in the dataset
STAR <- read.csv("STAR2.csv")

#Recode classtype into 'kinder'
STAR$kinder[STAR$classtype == "1"] <- "small"
STAR$kinder[STAR$classtype == "2"] <- "regular"
STAR$kinder[STAR$classtype == "3"] <- "regular with aid"
STAR$kinder <- as.factor(STAR$kinder)

#Re-code the 'race' variable
STAR$race[STAR$race=="1"] <- "white"
STAR$race[STAR$race=="2"] <- "black"
STAR$race[STAR$race=="4"] <- "hispanic"
STAR$race[STAR$race=="3" | STAR$race=="5" | STAR$race=="6"] <- "others"
STAR$race <- as.factor(STAR$race)

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)

```

library(dplyr)

#mean reading and math scores for the small class
small_reading_mean <- mean(STAR$g4reading[STAR$kinder=="small"], na.rm = TRUE)
small_math_mean <- mean(STAR$g4math[STAR$kinder=="small"], na.rm = TRUE)

#mean reading and math scores for the regular class
regular_reading_mean <- mean(STAR$g4reading[STAR$kinder=="regular"],
                             na.rm = TRUE)
regular_math_mean <- mean(STAR$g4math[STAR$kinder=="regular"], na.rm = TRUE)

#differences in mean values
diff_in_reading_mean <- small_reading_mean - regular_reading_mean
diff_in_math_mean <- small_math_mean - regular_math_mean

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

#to determine whether the difference in mean scores are truly significant or due
#to chance, we will conduct a t-test

```

```

#Calculate t-test scores to compare the means in both groups
ttest_reading <- t.test(g4reading ~ kinder, data = STAR, subset =
                        kinder %in% c("small", "regular"), na.rm = TRUE)
ttest_math <- t.test(g4math ~ kinder, data = STAR, subset =
                   kinder %in% c("small", "regular"), na.rm = TRUE)
print(ttest_reading)

##
## Welch Two Sample t-test
##
## data: g4reading by kinder
## t = -1.3021, df = 1132.5, p-value = 0.1931
## alternative hypothesis: true difference in means between group regular and group small is not equal
## 95 percent confidence interval:
## -10.240297 2.070433
## sample estimates:
## mean in group regular mean in group small
## 720.4467 724.5316

cat("Difference in Mean - Reading:", diff_in_reading_mean, "\n")

## Difference in Mean - Reading: 4.084932

cat("Standard Deviation - Reading:", sd_reading, "\n")

## Standard Deviation - Reading: 52.33846

print(ttest_math)

##
## Welch Two Sample t-test
##
## data: g4math by kinder
## t = 0.2441, df = 1105.7, p-value = 0.8072
## alternative hypothesis: true difference in means between group regular and group small is not equal
## 95 percent confidence interval:
## -4.379702 5.624264
## sample estimates:
## mean in group regular mean in group small
## 709.6314 709.0091

cat("Difference in Mean - Math:", diff_in_math_mean, "\n")

## Difference in Mean - Math: -0.6222809

cat("Standard Deviation - Math:", sd_math)

## Standard Deviation - Math: 43.8088

```

Interpretation of results:

1. Reading: Students in small classes scored, on average, 4.0849 points higher in reading than those in regular classes with an overall average score of 724.5316. However, comparing the difference in reading scores to the standard deviation (52.3385) shows that the effect size is relatively small. Specifically, the difference is less than 10% of the standard deviation (approximately a 7.8% effect). Although there is a positive difference, the effect is relatively modest when considering the variability in reading scores.

The calculated t-value is -1.3021. A negative t-value indicates that the mean score of the first group (regular) is lower than the second group (small).

2. Math: Students in small classes score slightly lower, on average, than those in regular classes with a difference of -0.62. However, it is important to note that this difference is very small, especially when compared to the standard deviation of 43.81. In reality, the difference in math scores is negligible.

The t-value is 0.2441. A positive t-value indicates that the mean score of the first group (regular) is higher than the second group (small), but this value is close to zero, indicating that the difference might be negligible.

Overall, while there seems to be a slight advantage in reading performance for students in small classes, the benefits in terms of fourth-grade reading and math performance are not substantial.

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)

```
#calculate lower and upper quantile for math and reading scores
#regular class:
regular_math_quantile <- quantile(STAR$g4math[STAR$kinder == "regular"],
                                probs = c(0.33, 0.66), na.rm = TRUE)
regular_reading_quantile <- quantile(STAR$g4reading[STAR$kinder == "regular"],
                                    probs = c(0.33, 0.66), na.rm = TRUE)

#small class
small_reading_quantile <- quantile(STAR$g4reading[STAR$kinder == "small"],
                                  probs = c(0.33, 0.66), na.rm = TRUE)
small_math_quantile <- quantile(STAR$g4math[STAR$kinder == "small"],
                               probs = c(0.33, 0.66), na.rm = TRUE)

#store quantile values in a dataframe
quantile_df <- data.frame(
  Class_Type = c(rep("Small", 2), rep("Regular", 2), rep("Small", 2),
                rep("Regular", 2)),
  Subject = c(rep("Reading", 4), rep("Math", 4)),
  Quantile = rep(c("33rd", "66th"), 4),
  Value = c(
    small_reading_quantile[1], small_reading_quantile[2],
    regular_reading_quantile[1], regular_reading_quantile[2],
    small_math_quantile[1], small_math_quantile[2],
    regular_math_quantile[1], regular_math_quantile[2]
  )
)
print(quantile_df)
```


| ## | Class_Type | Subject | Quantile | Value |
|------|------------|---------|----------|--------|
| ## 1 | Small | Reading | 33rd | 708.00 |
| ## 2 | Small | Reading | 66th | 741.00 |
| ## 3 | Regular | Reading | 33rd | 705.00 |
| ## 4 | Regular | Reading | 66th | 740.00 |
| ## 5 | Small | Math | 33rd | 695.00 |
| ## 6 | Small | Math | 66th | 727.00 |
| ## 7 | Regular | Math | 33rd | 696.00 |
| ## 8 | Regular | Math | 66th | 724.64 |

Quantile Scores:

- Reading Scores:
 - For students in small classes, the 33rd percentile score (those performing below average) is 708, while it's 705 for those in regular-sized classes.
 - For students in small classes performing above average (66th percentile), the score is 741, while it's 740 for those in regular-sized classes.
- Math Scores:
 - For students in small classes at the 33rd percentile, the score is 695, while it's 696 for those in regular-sized classes.
 - At the 66th percentile, students in small classes have a score of 727, while those in regular classes score 725.

Interpretation of results:

- Reading:
 - Students in smaller classes who are at the 33rd percentile (below average) are performing slightly better than their peers in regular classes, but the difference is just 3 points.
 - Similarly, students in smaller classes who are at the 66th percentile (above average) are performing marginally better than their counterparts in regular classes by 1 point.
- Math:
 - Interestingly, students in smaller classes at the 33rd percentile are performing slightly worse by 1 point compared to their peers in regular classes.
 - However, students in smaller classes at the 66th percentile perform slightly better by 2 points compared to those in regular classes.

Although the differences are small in magnitude, this nuanced analysis provides a more detailed picture than the mean comparison, highlighting that the benefits (or lack thereof) of small classes can vary slightly across the performance spectrum. However, the key takeaway is that the differences, while present, are modest. This suggests that while there may be some benefits to smaller classes, they might not be substantial when considering the range of possible scores.

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)

```

#create variables to store all 'white' and 'minority' students
white <- subset(STAR, race=="white")
minority <- subset(STAR, race=="black" | race=="hispanic")

#calculate mean values for the regular class, sub-setting for both racial groups
mean_white_reading_regular <- mean(white$g4reading[white$kindergarten=="regular"],
                                   na.rm = TRUE)
mean_minority_reading_regular <- mean(minority$g4reading[minority$kindergarten=="regular"],
                                   na.rm = TRUE)
mean_white_math_regular <- mean(white$g4math[white$kindergarten=="regular"],
                                na.rm = TRUE)
mean_minority_math_regular <- mean(minority$g4math[minority$kindergarten=="regular"],
                                   na.rm = TRUE)

#calculate mean values for the small class, sub-setting for both racial groups
mean_white_reading_small <- mean(white$g4reading[white$kindergarten=="small"],
                                na.rm = TRUE)
mean_minority_reading_small <- mean(minority$g4reading[minority$kindergarten=="small"],
                                   na.rm = TRUE)
mean_white_math_small <- mean(white$g4math[white$kindergarten=="small"],
                              na.rm = TRUE)
mean_minority_math_small <- mean(minority$g4math[minority$kindergarten=="small"],
                                 na.rm = TRUE)

#calculate difference in means to compare achievement gaps within racial groups
diff_reading_reg = mean_white_reading_regular - mean_minority_reading_regular
diff_reading_small = mean_white_reading_small - mean_minority_reading_small
diff_math_reg = mean_white_math_regular - mean_minority_math_regular
diff_math_small = mean_white_math_small - mean_minority_math_small

#store mean values in a dataframe
mean_df <- data.frame(
  Race = c("White", "Minority", "White", "Minority", "White", "Minority",
           "White", "Minority"),
  Class_Type = c(rep("Regular", 2), rep("Small", 2), rep("Regular", 2),
                 rep("Small", 2)),
  Subject = c(rep(c("Reading", "Math"), 4)),
  Mean = c(mean_white_reading_regular, mean_minority_reading_regular,
            mean_white_reading_small, mean_minority_reading_small,
            mean_white_math_regular, mean_minority_math_regular,
            mean_white_math_small, mean_minority_math_small)
)

#store difference in mean values in a dataframe
diff_df <- data.frame(
  Class_Type = c("Regular", "Small", "Regular", "Small"),
  Subject = c("Reading", "Reading", "Math", "Math"),
  Difference_in_Mean = c(diff_reading_reg, diff_reading_small,
                        diff_math_reg, diff_math_small)
)

print(mean_df)

```

```
##      Race Class_Type Subject      Mean
```

```
## 1    White    Regular Reading 725.3104
## 2 Minority Regular   Math 692.2556
## 3    White      Small Reading 728.8264
## 4 Minority      Small   Math 699.2963
## 5    White    Regular Reading 711.4358
## 6 Minority Regular   Math 699.0889
## 7    White      Small Reading 711.1030
## 8 Minority      Small   Math 696.3375
```

```
print(diff_df)
```

```
##   Class_Type Subject Difference_in_Mean
## 1   Regular Reading           33.05486
## 2     Small Reading           29.53008
## 3   Regular   Math           12.34691
## 4     Small   Math           14.76550
```

Interpretation:

- In regular classes, the achievement gap (based on the difference in scores) between white and minority students is 33.05 points for reading and 12.35 points for math.
- In small classes, the achievement gap slightly narrows to 32.49 points for reading but slightly increases to 14.76 points for math.
- The STAR program, by way of implementing small classes, seems to have had a minor positive impact on reducing the racial achievement gap in reading, as the gap narrowed by about 0.56 points.
- However, for math, the achievement gap increased slightly by 2.41 points in small classes.

Overall, the STAR program appears to be somewhat beneficial in slightly narrowing the achievement gap in reading scores, but its impact on math scores is less clear. While the reduction in the reading gap is a positive outcome, the differences in scores, even in small classes, highlight that there still exists a substantial achievement gap that needs addressing.

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)

```
#calculate graduation rate by class size
gradrate_by_class_size <- tapply(STAR$hsgrad, STAR$kinder, mean, na.rm = TRUE)

#calculate graduation rate by years spent in a small class
gradrate_by_years_spent_small <- tapply(STAR$hsgrad, STAR$yearssmall, mean,
                                         na.rm = TRUE)

#Calculate racial gap in graduation rates

#compute the means of both races separately
grad_white <- tapply(white$hsgrad, white$kinder, mean, na.rm = TRUE)
grad_minority <- tapply(minority$hsgrad, minority$kinder, mean, na.rm = TRUE)

#difference in graduation rates by race (white - minority)
```

```
grad_by_race <- grad_white - grad_minority

#store values in a dataframe
df_gradrate_by_class_size <- data.frame(
  Class_Type = names(gradrate_by_class_size),
  Graduation_Rate = as.vector(gradrate_by_class_size))

df_gradrate_by_years_spent_small <- data.frame(
  Years_Spent_In_Small_Class = as.numeric(names(gradrate_by_years_spent_small)),
  Graduation_Rate = as.vector(gradrate_by_years_spent_small))

df_grad_by_race <- data.frame(
  Class_Type = names(grad_by_race),
  Racial_Gap_In_Graduation = as.vector(grad_by_race))

print(df_gradrate_by_class_size)
```

```
##           Class_Type Graduation_Rate
## 1           regular      0.8225214
## 2 regular with aid      0.8358586
## 3             small      0.8477306
```

```
print(df_gradrate_by_years_spent_small)
```

```
##  Years_Spent_In_Small_Class Graduation_Rate
## 1                        0      0.8250000
## 2                        1      0.7979798
## 3                        2      0.8518519
## 4                        3      0.8269231
## 5                        4      0.8833747
```

```
print(df_grad_by_race)
```

```
##           Class_Type Racial_Gap_In_Graduation
## 1           regular      0.1214189
## 2 regular with aid      0.1387456
## 3             small      0.1005110
```

Interpretation:

1. Racial Disparities: White students consistently have higher graduation rates than minority students across all class types. The racial gap in graduation rates is the smallest in small classes, suggesting that smaller environments might be more effective in reducing racial disparities in educational outcomes.
2. Class Size Impact: Graduation rates are generally higher for students in smaller classes.
3. Duration in Small Classes: Interestingly, students who spent all 4 years in small classes have the highest graduation rate of 88.34%. Although the rates fluctuate for those who spent between 0 to 3 years in small classes, the data suggests a potential cumulative benefit from being in a small class for a prolonged period of time.

Overall, the STAR program with its emphasis on smaller class sizes, appears to have positive implications for high school graduation rates. While it has led to higher graduation rates overall, especially for those spending more extended periods in smaller classes, the program has been somewhat effective in reducing—but not completely eliminating—the racial achievement gap. The remaining disparity between white and minority students' graduation rates indicates that further interventions might be necessary to achieve educational equity.

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 the incorporation of nature-oriented breaks during the workday improve employee mental well-being and productivity?
- Objective: To investigate the impact of short nature-oriented breaks on the mental well-being and productivity of employees compared to those who take regular breaks.
- Approach/Process: We will conduct a randomized controlled trial (RCT), adhering to principles like ATE (Average Treatment Effect) and SUTVA (Stable Unit Treatment Value Assumption). Participants will be categorized into the control group (those who take regular breaks) and the treatment group (those who participate in nature-oriented breaks).
- Participants: Recruit 500 employees aged 25-60 from various professional sectors and industries. To ensure no violation of SUTVA, participants would not be co-workers, work in the same office or be in close physical proximity. It would also be ensured that they do not have any known allergies or ailments that would make exposure to nature harmful.
- Random Assignment: Participants are randomly distributed between the treatment and control groups. Through randomization, selection bias is minimized, ensuring that any difference in outcomes can be attributed to the intervention.
 - Treatment Group: Participants will engage in a 15-minute nature break twice a day. This involves walking in a garden, park, any natural setting, or watching a nature-oriented video if physical nature spaces aren't accessible.
 - Control Group: Participants will continue with their regular breaks, which involves any activity of their choice not related to nature.
- Measurement: Mental well-being is assessed using standardized psychological questionnaires, while productivity is measured using self-reports and supervisor evaluations at the beginning and end of the study, spanning 8 weeks. Additionally, pre-treatment covariates are identified to account for potential imbalances or confounders. This includes a pre-survey of general mental well-being, age, job type, regular work hours, and prior exposure to nature.
- Outcomes:
 - Nature breaks have an effect: If, by the end of 8 weeks, the treatment group shows a statistically significant improvement in mental well-being and productivity levels compared to the control group, this would suggest that nature-oriented breaks have a positive impact.
 - Nature breaks have no effect: If no significant difference is observed between the two groups, it would indicate that nature-oriented breaks don't significantly affect well-being or productivity levels.
- Causal Treatment Effect: Randomized controlled trials ensure that bias is minimized, allowing us to observe a clearer and more accurate establishment of causality. Through randomization of treatment assignment and by obeying the concept of ignorability, we can determine whether nature-oriented breaks have a causal effect on an employee's well-being and productivity levels.

Why observational data might not work:

- Observational data does not account for confounding variables, such as personal interests or pre-existing mental health conditions. People predisposed to nature might inherently have better mental well-being. Additionally, observational data in the context of this study (e.g. surveys from existing breaks) may lead to a higher chance of inaccuracy and bias. In other words, memories are often faulty, and perceptions change based on mood, recent experiences and a myriad of other factors that may influence survey responses.
- It is often challenging to ascertain directionality with observational data. Productive individuals may be more inclined to go on nature breaks, yet on the other hand, taking nature-oriented breaks may increase individual productivity and well-being.