Stats 102A - Homework 6 - Output File

Daren Sathasivam – 306229580

Homework questions and prompts copyright Miles Chen, Do not post, share, or distribute without permission.

Academic Integrity Statement

By including this statement, I, **Daren Sathasivam**, declare that all of the work in this assignment is my own original work. At no time did I look at the code of other students nor did I search for code solutions online. I understand that plagiarism on any single part of this assignment will result in a 0 for the entire assignment and that I will be referred to the dean of students.

Part 1

Explain what a p-value is and how a researcher should use it. (150 words or less)

- The p-value is a measure of strength against the null hypothesis when someone conducts statistical hypothesis testing. However, p-values can be manipulated into what the researcher would prefer to observe. As in "Science Isn't Broken", by Christie Aschwanden, states that people should "think of the p-value as an index of surprise" instead. Researchers should utilize p-values as a tool to assess the compatibility of their data with a specified model rather than definitive proof of a certain hypothesis. Over-reliance on the p-value can lead to misunderstandings and misinterpretation of data that many people may rely on. Aschwanden concludes, "We should make the best decisions we can with the current evidence and take care not to lose sight of its strength and degree of certainty", when new data arises. Thus, researchers should utilize the p-value to inform rather than dictate certain scientific conclusions and complement other findings of the research.

Part 2

Randomization test for numeric data

```
# Data credit: David C. Howell
# `no_waiting` is a vector that records the time it took a driver to leave the
# parking spot if no one was waiting for the driver
no_waiting <- c(36.30, 42.07, 39.97, 39.33, 33.76, 33.91, 39.65, 84.92, 40.70, 39.65,
39.48, 35.38, 75.07, 36.46, 38.73, 33.88, 34.39, 60.52, 53.63, 50.62) # 20 obs

# `waiting` is a vector that records the time it takes a driver to leave if
# someone was waiting on the driver
waiting <- c(49.48, 43.30, 85.97, 46.92, 49.18, 79.30, 47.35, 46.52, 59.68, 42.89,
49.29, 68.69, 41.61, 46.81, 43.75, 46.55, 42.33, 71.48, 78.95, 42.06) # 20 obs

mean(waiting)
```

```
## [1] 54.1055
```

```
mean(no_waiting)
```

```
## [1] 44.421
```

```
obs_dif <- mean(waiting) - mean(no_waiting) # 9.6845
# obs_dif
```

Randomization test

Conduct a randomization test

```
set.seed(1)
wait_time <- c(no_waiting, waiting)
differences <- rep(NA, 10000)
for (i in seq_along(differences)) {
  randomized <- sample(wait_time)
  groupA <- randomized[1:20]
  groupB <- randomized[21:40]
  differences[i] <- mean(groupA) - mean(groupB)
}
summary(differences)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -16.09850  -3.31600  -0.00850  -0.03937   3.22300  17.46050
```

```
p_value <- mean(differences >= obs_dif) # estimated empirical p-value after 10000 repetitions
cat("Observed differences: ", obs_dif, "\n")
```

```
## Observed differences:  9.6845
```

```
cat("Estimated p-value after 10000 repetitions: ", p_value, "\n")
```

```
## Estimated p-value after 10000 repetitions:  0.0193
```

- The empirical p-value < 0.05 provides sufficient evidence to reject the null hypothesis that there is no difference in average waiting time between the two groups. Therefore, it can be observed that drivers who have someone waiting tend to take longer to leave the parking spot compared to those who do not have someone waiting.

Comparison to traditional t-test

Conduct a traditional two-sample independent t-test.

```
t_test <- t.test(no_waiting, waiting, var.equal = TRUE)
t_test
```

```
##
##  Two Sample t-test
##
## data:  no_waiting and waiting
## t = -2.1496, df = 38, p-value = 0.03802
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.8049004  -0.5640996
## sample estimates:
## mean of x mean of y
##   44.4210   54.1055
```

```r
cat("p-value from traditional t-test: ", t_test$p.value, "\n")
```

p-value from traditional t-test: 0.03801646

```r
cat("p-value from randomization tests: ", p_value, "\n")
```

p-value from randomization tests: 0.0193

- Since the p-value from the t-test is also smaller than 0.05, the same interpretation can be said for the t-test's null hypothesis. This is consistent with the randomization test as they both reject the null hypothesis.

Part 3

Another Randomization test for numeric data
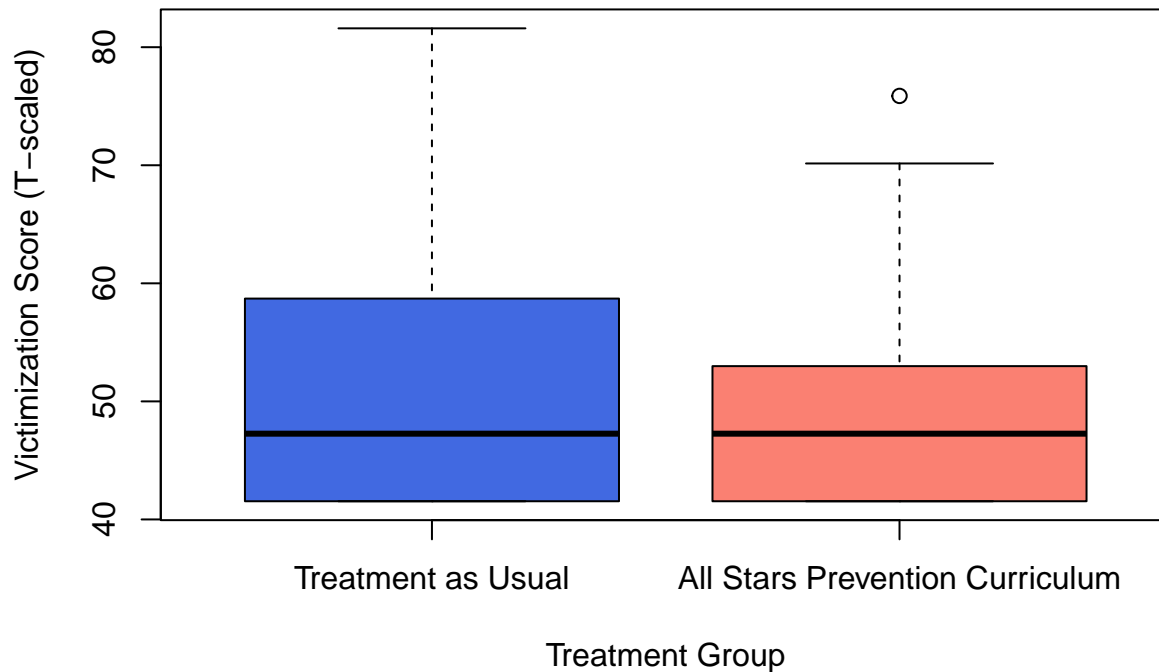
Exploratory Analysis

```r
library(ggplot2)
school_data <- read.csv("AfterSchool.csv")
# school_data

# ID: ID Number
# Treatment: 0 = treatment as usual; 1 = all stars prevention curriculum
# Aggress(measure of student's aggresion): T-scaled(M=50, SD=10)
# Delinq(Measure of student's delinquent behavior): T-scaled(M=50,SD=10)
# Victim(Measure of victimization): T-scaled(M=50, SD=10)
str(school_data)
```

```
## 'data.frame':    356 obs. of  5 variables:
##  $ ID       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Treatment: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Aggress  : num  63.2 51.8 74.5 40.5 56.4 ...
##  $ Delinq   : num  44.5 76.8 50.9 44.5 44.5 ...
##  $ Victim   : num  64.4 64.4 41.5 41.5 53 ...
```

```r
# summary(school_data)
boxplot(school_data$Victim ~ school_data$Treatment,
        main = "Victimization Scores by Treatment Group",
        xlab = "Treatment Group",
        ylab = "Victimization Score (T-scaled)",
        names = c("Treatment as Usual", "All Stars Prevention Curriculum"),
        col = c("royalblue", "salmon"))
```

Victimization Scores by Treatment Group

- **Exploratory Analysis** Based on the boxplots above, it can be observed that the students who received treatment as usual had a larger interquartile range in victimization scores. Both have a similar central tendency, however the victimization score for the all stars prevention curriculum has a smaller IQR and the highest value being around the 70s besides the outlier whereas the highest value for the treatment as usual group is a little above 80. Observations based on the boxplot can suggest that the All Stars Prevention Program does have a slight impact on victimization scores with the IQR and max value being relatively smaller in comparison to the control group.

Randomization Test

```r
treatment <- school_data$Victim[school_data$Treatment == 1] # 169 obs; mean=49.29484; sd = 8.837291
control <- school_data$Victim[school_data$Treatment == 0] # 187 obs; mean=50.5987; sd=10.88845
victim_data <- c(treatment, control)

obs_dif <- mean(treatment) - mean(control) # -1.303856
obs_dif
```

```
## [1] -1.303856
```

```r
set.seed(1)
differences <- rep(NA, 10000)
for (i in seq_along(differences)) {
  randomized <- sample(victim_data)
  groupA <- randomized[1:length(treatment)]
  groupB <- randomized[(length(treatment) + 1):length(victim_data)]
  differences[i] <- mean(groupA) - mean(groupB)
}
summary(differences)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
```

```
## -3.88224 -0.72372 -0.01467 -0.02361  0.69439  4.17520
```

```r
p_value <- mean(differences >= obs_dif) # estimated empirical p-value after 10000 repetitions
cat("Observed differences: ", obs_dif, "\n")
```

```
## Observed differences:  -1.303856
```

```r
cat("Estimated p-value after 10000 repetitions: ", p_value, "\n")
```

```
## Estimated p-value after 10000 repetitions:  0.8865
```

- Null Hypothesis H_0 : There is no effect of the after school program on victimization.
- Alternative Hypothesis H_a : There is an effect of the after school program on victimization.
- Conclusion: The observed difference in mean victimization is -1.303856 indicating a slightly lower average victimization score compared to the control group. However, after conducting a randomization test of 10,000 repetitions, the empirical p-value is 0.8865. Since it is greater than 0.05, we fail to reject the null hypothesis, meaning we do not have convincing evidence to suggest that the after school program as a statisitically significant effect on victimization.

Part 4

Randomization test

```r
# TRUE = live; FALSE = die
cmt_outcome <- c(rep(TRUE, 6), rep(FALSE, 4)) # CMT: 4 deaths, 6 alive ; mean=0.6
ecmo_outcome <- c(rep(TRUE, 28), rep(FALSE, 1)) # ECMO: 1 death, 28 alive ; mean =0.9655172

obs_dif <- mean(ecmo_outcome) - mean(cmt_outcome)
obs_dif
```

```
## [1] 0.3655172
```

```r
outcomes <- c(ecmo_outcome, cmt_outcome)
set.seed(1)
differences <- rep(NA, 10000)
for (i in seq_along(differences)) {
  randomized <- sample(outcomes)
  groupA <- randomized[1:length(ecmo_outcome)]
  groupB <- randomized[(length(ecmo_outcome) + 1):length(outcomes)]
  differences[i] <- mean(groupA) - mean(groupB)
}
summary(differences)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.172414 -0.037931 -0.037931 -0.002871  0.096552  0.500000
```

```r
p_value <- mean(differences >= obs_dif)

cat("Observed differences: ", obs_dif, "\n")
```

```
## Observed differences:  0.3655172
```

```r
cat("Empirical p-value: ", p_value, "\n")
```

```
## Empirical p-value:  0.0111
```

- Null Hypothesis H_0: The ECMO treatment is not as effective as CMT treatment in treating severe respiratory failure.

- Alternative Hypothesis H_a: The ECMO treatment is more effective than CMT treatment in treating severe respiratory failure.
- Conclusions: The observed difference is 0.3655172, suggesting ECMO is more effective in the observed test and control group. After a randomization test of 10,000 repetitions, the empirical p-value is 0.0111, which is less than 0.05. Thus, rejecting the null hypothesis suggests that there is statistically significant evidence to support the claim that the ECMO(extracorporeal membrane oxygenation) treatment is more effective than CMT(conventional medical therapy) in saving the lives of newborn babies with severe respiratory failure.

Comparison to Fisher's Exact Test

Use R's `fisher.test()`

```
outcome_table <- matrix(c(28, 1, 6, 4), nrow = 2, byrow = TRUE,
                        dimnames = list(c("Live", "Die"), c("ECMO", "CMT")))
outcome_table
```

```
##      ECMO CMT
## Live   28   1
## Die     6   4
```

```
fisher_result <- fisher.test(outcome_table)
fisher_result
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  outcome_table
## p-value = 0.01102
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##     1.366318 944.080411
## sample estimates:
## odds ratio
##    16.78571
```

```
cat("Fisher test result p-value: ", fisher_result$p.value, "\n")
```

```
## Fisher test result p-value:  0.01101506
```

- Null Hypothesis H_0: The ECMO treatment is not as effective as CMT treatment in treating severe respiratory failure.
- Alternative Hypothesis H_a: The ECMO treatment is more effective than CMT treatment in treating severe respiratory failure.
- Conclusion: Both the Fisher test and Randomization test result in a p-value around 0.011. Given that both test results in similar p-values, the Fisher test can provide evidence to reject the null-hypothesis and that there is statistically significant evidence of the ECMO treatment as more effective at saving newborns' lives than CMT.

Part 5

Comparing Groups, Chapter 7, Exercise 7.1

Non-parametric bootstrap test

- Variance of science scores between public and provate school students. $H_0 : \sigma^2_{Public} = \sigma^2_{Private}$

```r
hsb <- read.csv("HSB.csv")
# schtyp: Type of school: 0=Private, 1=Public
# science: Student's science achievement score: T-scaled(M=50,SD=10)
# head(hsb)

public <- hsb$Sci[hsb$Schtyp == 1] # 168 obs
private <- hsb$Sci[hsb$Schtyp == 0] # 32
obs_var_diff <- var(public) - var(private) # 36.89112
obs_var_diff
```

```
## [1] 36.89112
```

```r
all_students <- c(public, private)

set.seed(1)
differences <- rep(NA, 10000)
for (i in seq_along(differences)) {
  groupA <- sample(public, length(public), replace = TRUE)
  groupB <- sample(private, length(private), replace = TRUE)
  differences[i] <- var(groupA) - var(groupB)
}
summary(differences)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -25.69   27.07   38.27   38.31   49.69   92.37
```

```r
p_value <- mean(abs(differences) >= abs(obs_var_diff)) # 0.5363

cat("Observed differences in variances: ", obs_var_diff, "\n")
```

```
## Observed differences in variances:  36.89112
```

```r
cat("Non-parametric bootstrap with 10000 repetitions empirical p-value: ", p_value, "\n")
```

```
## Non-parametric bootstrap with 10000 repetitions empirical p-value:  0.5357
```

- Assumption: The science scores are normally distributed within both the private and public school groups.
- Null Hypothesis $H_0 : \sigma^2_{Public} = \sigma^2_{Private}$
- Alternative Hypothesis: $H_a : \sigma^2_{Public} \neq \sigma^2_{Private}$
- Conclusion: The observed variance between public and private schools is 36.89112, indicating that the public school has a higher variance in science scores when compared to the scores of private school students. After performing a non-parametric bootstrap of 10000 repetitions, the empirical p-value is 0.5363 which is greater than 0.05. Thus we reject to fail the null hypothesis, indicating that the data does not provide statistically significant evidence against science achievement scores among public and private schools.

Parametric bootstrap test

```r
mean_public <- mean(public)
sd_public <- sd(public)
mean_private <- mean(private)
sd_private <- sd(private)

differences <- rep(NA, 10000)
for (i in seq_along(differences)) {
```

```
  groupA <- rnorm(length(public), mean_public, sd_public)
  groupB <- rnorm(length(private), mean_private, sd_private)
  differences[i] <- var(groupA) - var(groupB)
}
summary(differences)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -58.22   23.91   38.10   37.02   51.23  108.31
```

```
p_value <- mean(abs(differences) >= abs(obs_var_diff)) # 0.5238

cat("Observed differences in variances: ", obs_var_diff, "\n")
```

```
## Observed differences in variances:  36.89112
```

```
cat("Parametric bootstrap with 10000 repetitions empirical p-value: ", p_value, "\n")
```

```
## Parametric bootstrap with 10000 repetitions empirical p-value:  0.5224
```

- Null Hypothesis $H_0 : \sigma^2_{Public} = \sigma^2_{Private}$
- Alternative Hypothesis: $H_a : \sigma^2_{Public} \neq \sigma^2_{Private}$
- Conclusion: After conducting a bootstrap test of 10000 repetitions, the resulting emprical p-value is 0.5238. Similar to non-parametric's emprical p-value, we fail to reject the null hypothesis, suggesting that there is not statistically significant evidence that points to the difference of science scores between public and private schools.

Part 6

```
light <- c(28, 26, 33, 24, 34,-44, 27, 16, 40, -2, 29, 22, 24, 21, 25, 30, 23, 29, 31, 19,
           24, 20, 36, 32, 36, 28, 25, 21, 28, 29, 37, 25, 28, 26, 30, 32, 36, 26, 30, 22,
           36, 23, 27, 27, 28, 27, 31, 27, 26, 33, 26, 32, 32, 24, 39, 28, 24, 25, 32, 25,
           29, 27, 28, 29, 16, 23)
```

Non-parametric bootstrap test

```
# modern accepted value is 33
observed_mean <- mean(light) # 26.21212
modern_val <- 33
adjustment <- modern_val - observed_mean
centered <- light + adjustment

bootstrap_means <- rep(NA, 10^5)
set.seed(1)
for(i in seq_along(bootstrap_means)) {
  bootstrap_sample <- sample(centered, length(light), replace = TRUE)
  bootstrap_means[i] <- mean(bootstrap_sample)
}
summary(bootstrap_means)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.36   32.18   33.12   33.00   33.95   37.20
```

```
p_value <- mean(abs(bootstrap_means - modern_val) >= abs(observed_mean - modern_val))

cat("Observed mean: ", observed_mean, "\n")
```

```
## Observed mean:  26.21212
```
```
cat("Non-parametric bootstrap with 100,000 repetitions emprical p-value:", p_value, "\n")
```

```
## Non-parametric bootstrap with 100,000 repetitions emprical p-value: 5e-05
```

- Null Hypothesis: $H_0 : \mu = 33$
- Alternative Hypothesis: $H_a : \mu \neq 33$
- Conclusion: After 100,000 repetitions of non-parametric bootstrap, the empirical p-value is 5e-5. Since it is less than 0.05, we reject the null hypothesis indicating that it is highly unlikely that Newcomb's measurements would have such a significant negative deviation from the modern accepted value of 33 nanoseconds by chance alone.

Non-parametric bootstrap test with outliers removed

Perform the bootstrap test again after removing the two negative outliers (-2, and -44)

```r
new_light <- light[light >= 0]
new_observed_mean <- mean(new_light) # 27.75
new_adjustment <- modern_val - new_observed_mean
new_centered <- new_light + new_adjustment

new_bootstrap_means <- rep(NA, 10^5)
set.seed(1)
for (i in seq_along(new_bootstrap_means)) {
  new_bootstrap_sample <- sample(new_centered, length(new_light), replace = TRUE)
  new_bootstrap_means[i] <- mean(new_bootstrap_sample)
}
summary(new_bootstrap_means)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   30.34   32.58   33.00   33.00   33.42   35.88
```

```r
p_value <- mean(abs(new_bootstrap_means - modern_val) >= abs(new_observed_mean - modern_val))

cat("New observed mean: ", new_observed_mean, "\n")
```

```
## New observed mean:  27.75
```
```
cat("Non-parametric bootstrap with 100,000 repetitions empirical p-value:", p_value, "\n")
```

```
## Non-parametric bootstrap with 100,000 repetitions empirical p-value: 0
```

- The p-value becomes even smaller after removing the outliers although the new mean is closer to the accepted value of 33. Removing the extreme negative outliers tightens the distributions of means from the non-parametric bootstrap resamples. This means that there are even fewer resampled datasets with means as low or lower than the new observed mean(27.75). Consequently, the p-value decreases further, indicating stronger statistical evidence against the null hypothesis of $H_0 : \mu = 33$.