

HW5

111078513

Helped By: 111078503, 111078505, 111078511

Question 1)

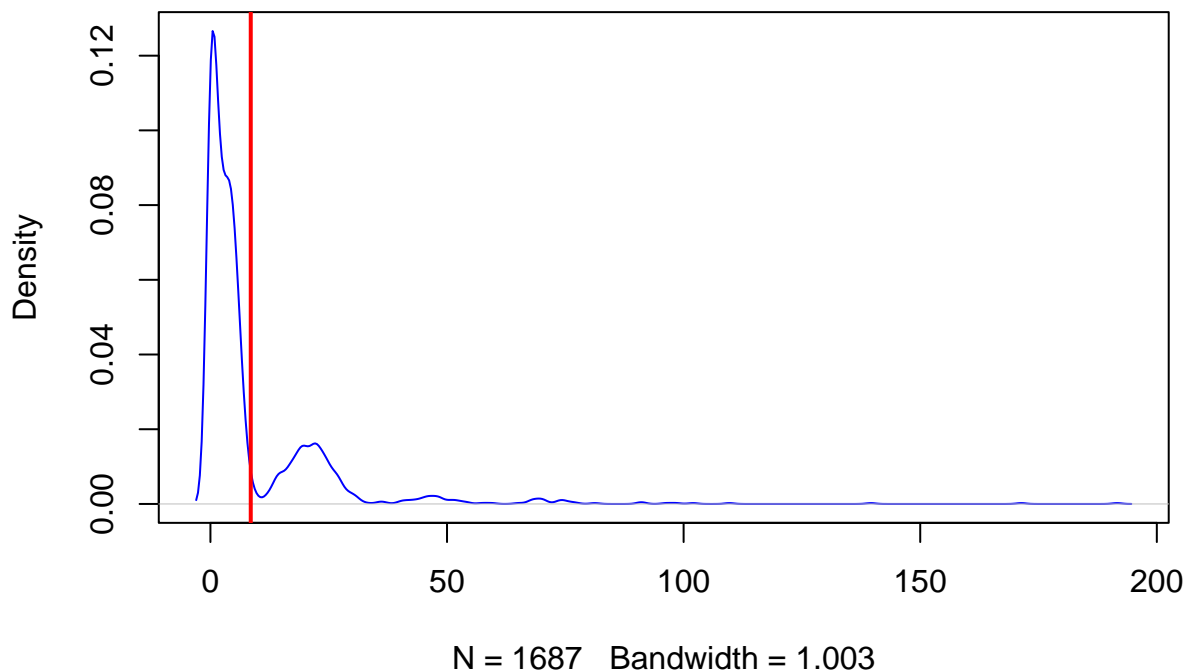
a. Imagine that Verizon claims that they take 7.6 minutes to repair phone services for its customers on average. The PUC seeks to verify this claim at 99% confidence (i.e., significance $\alpha = 1\%$) using traditional statistical methods.

(i) Visualize the distribution of Verizon's repair times, marking the mean with a vertical line

```
# load data
data <- read.csv("verizon.csv", header = TRUE, sep = ",")
rp_time <- data$Time
# Visualization

plot(density(rp_time), lty = 1, col = "blue")
abline(v = mean(rp_time), col = "red", lwd = 2)
```

density.default(x = rp_time)



(ii) Given what the PUC wishes to test, how would you write the hypothesis?

- H_0 : average repairment time = 7.6 min
- H_1 : average repairment time \neq 7.6 min

(iii) Estimate the population mean, and the 99% confidence interval (CI) of this estimate.

```
sample_size <- length(rp_time)
sample_mean <- mean(rp_time)
sample_sd <- sd(rp_time)
se <- (sample_sd /sqrt(sample_size))
cat("Estimated population mean => ", sample_mean, sep = "")

## Estimated population mean => 8.522009
cat("Estimated population 99% CI => ", sample_mean - 2.58*se
    , " ~ ", sample_mean + 2.58*se, sep = "")

## Estimated population 99% CI => 7.593073 ~ 9.450946
```

(iv) Find the t-statistic and p-value of the test

```
t <- (sample_mean - 7.6) / se ; t

## [1] 2.560762
df <- sample_size - 1
p <- 2 * (1 - pt(t, df)) ; p

## [1] 0.01053068
```

(v) Briefly describe how these values relate to the Null distribution of t

The larger the t-value, the more significant the difference between the sample mean and the null hypothesis. At the same time, the p-value will be smaller, indicating stronger evidence to reject the null hypothesis. Conversely, when the t-value is smaller, it indicates a smaller difference between the sample mean and the null hypothesis, and the p-value will be larger, indicating stronger evidence to accept the null hypothesis.

(vi) What is your conclusion about the company's claim from this t-statistic, and why?

Based on the conclusion we obtained in the t-test, since the P-value = 0.01053068 > 0.01, we tend to reject H_1 , which means that there is not enough evidence to prove that the repair time is not equal to 7.6 minutes.

b. Let's re-examine Verizon's claim that they take no more than 7.6 minutes on average, but this time using bootstrapped testing:

(i) Bootstrapped Percentile: Estimate the bootstrapped 99% CI of the population mean

```
boost_rp_time <- replicate(3000, mean(sample(rp_time, length(rp_time), replace = TRUE)))
boost_rp_time_mean <- mean(boost_rp_time)
boost_rp_time_quantile <- quantile(boost_rp_time, c(0.005, 0.995))
cat("Estimated population 99% CI => ", boost_rp_time_quantile[1],
    " ~ ", boost_rp_time_quantile[2], sep = "")

## Estimated population 99% CI => 7.66748 ~ 9.467576
```

(ii) Bootstrapped Difference of Means:

```

boost_rp_time_diff <- replicate(3000, mean(sample(rp_time, length(rp_time), replace = TRUE))-7.6)
boost_rp_time_diff_quantile <- quantile(boost_rp_time_diff, c(0.005, 0.995))
cat("99% CI of the bootstrapped difference between
    the sample mean and the hypothesized mean => "
    ,boost_rp_time_diff_quantile[1]
    , " ~ ", boost_rp_time_diff_quantile[2], sep = "")

## 99% CI of the bootstrapped difference between
##    the sample mean and the hypothesized mean => 0.04377905 ~ 1.897293

```

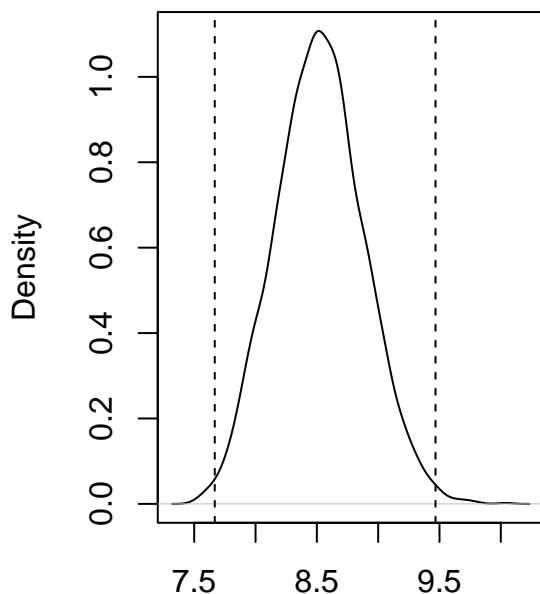
(iii) Plot distribution the two bootstraps above

```

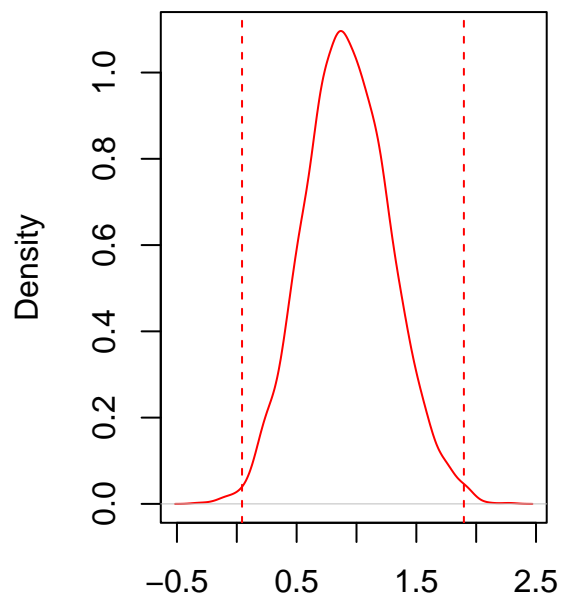
par(mfrow = c(1, 2))
plot(density(boost_rp_time))
abline(v = boost_rp_time_quantile, lty = 2)
plot(density(boost_rp_time_diff), col = "red")
abline(v = boost_rp_time_diff_quantile, lty = 2, col = "red")

```

density.default(x = boost_rp_time) **density.default(x = boost_rp_time_diff)**



N = 3000 Bandwidth = 0.06449



N = 3000 Bandwidth = 0.06512

(iv) Does the bootstrapped approach agree with the traditional t-test in part [a]?

In the traditional t-test, we rejected H_a (that is, the maintenance time is significantly different from 7.6 minutes); but from the graph above, it is easy to see that the bootstrapped estimated sample means are mostly around 8.5. From the above reasoning, we can conclude that the results of the traditional t-test and the bootstrap are not consistent.

c. Finally, imagine that Verizon notes that the distribution of repair times is highly skewed by outliers, and feel that testing the mean is not fair because the mean is sensitive to outliers. They claim that the median is a more fair test, and claim that the median repair time is no more than 3.5 minutes at 99% confidence (i.e., significance $\alpha = 1\%$).

(i) Bootstrapped Percentile: Estimate the bootstrapped 99% CI of the population median

```
boost_rp_time_median_sample <-
  replicate(3000, median(sample(rp_time, length(rp_time), replace = TRUE)))
boost_rp_time_median_quantile <- quantile(boost_rp_time_median_sample, c(0.005, 0.995))
cat("Estimated population 99% CI => ", boost_rp_time_median_quantile[1],
    " ~ ", boost_rp_time_median_quantile[2], sep = "")
```

```
## Estimated population 99% CI => 3.17 ~ 3.95
```

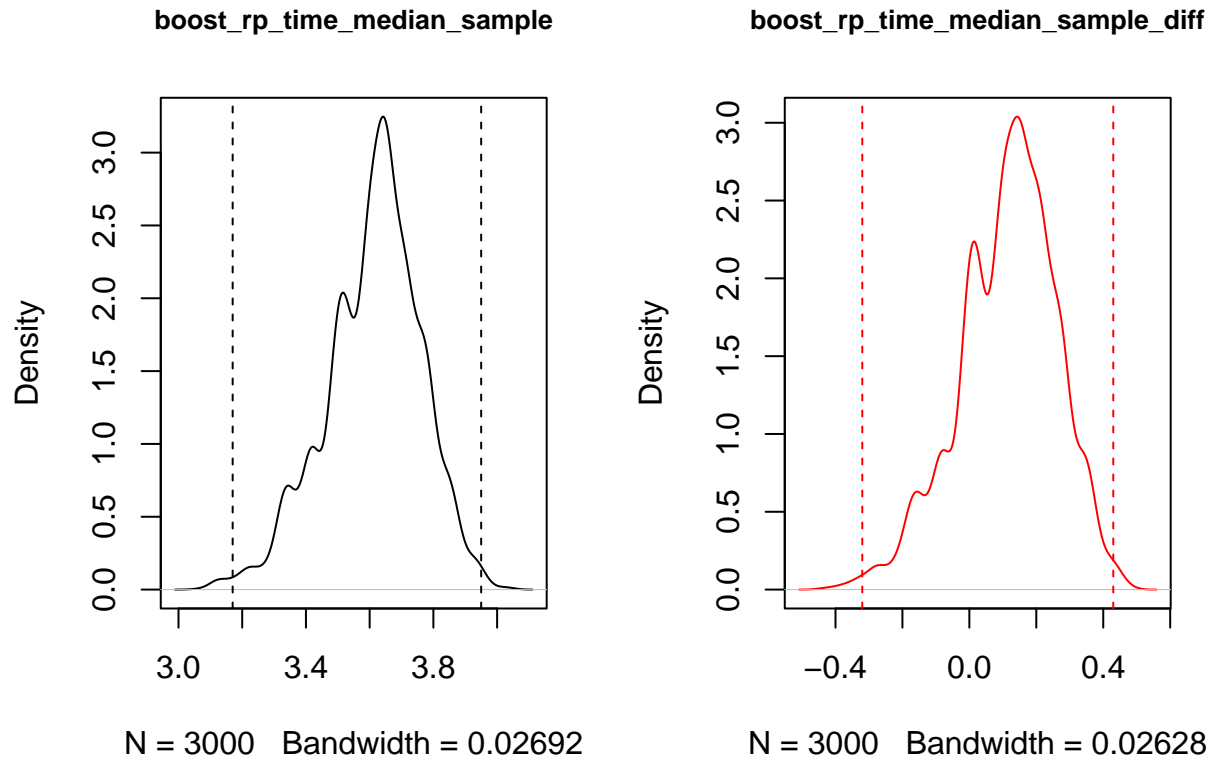
(ii) Bootstrapped Difference of Medians

```
boost_rp_time_median_sample_diff <-
  replicate(3000
    , median(sample(rp_time, length(rp_time), replace = TRUE)) - 3.5)
boost_rp_time_diff_median <- median(boost_rp_time_median_sample_diff)
boost_rp_time_median_diff_quantile <- quantile(boost_rp_time_median_sample_diff
  , c(0.005, 0.995))
cat("Estimated population 99% CI => ", boost_rp_time_median_diff_quantile[1],
    " ~ ", boost_rp_time_median_diff_quantile[2], sep = "")
```

```
## Estimated population 99% CI => -0.32 ~ 0.43
```

(iii) Plot distribution the two bootstraps above

```
par(mfrow = c(1, 2))
plot(density(boost_rp_time_median_sample)
  , main = "boost_rp_time_median_sample", cex.main = 0.8)
abline(v = boost_rp_time_median_quantile, lty = 2)
plot(density(boost_rp_time_median_sample_diff)
  , main = "boost_rp_time_median_sample_diff"
  , col = "red", cex.main = 0.8)
abline(v = boost_rp_time_median_diff_quantile, lty = 2, col = "red")
```



(iv) What is your conclusion about Verizon's claim about the median, and why?

Verizon's approach of using the median to estimate the population median is appropriate and can address the bias caused by right-skewed data. According to the results of the bootstrap, the population median will be included in the calculated confidence interval at a 99% confidence level. Therefore, there is not much problem with the median of the repair service time provided by Verizon, which is about 3.5 minutes.

Question 2) Load the `compstatslib` package and run `interactive_t_test()`. You will see a simulation of null and alternative distributions of the t-statistic, along with significance and power. If you do not see interactive controls (slider bars), press the gears icon on the top-left of the visualization.

```
library(compstatslib)
```

a. You discover that your colleague wanted to target the general population of Taiwanese users of the product. However, he only collected data from a pool of young consumers, and missed many older customers who you suspect might use the product much less every day.

(i) Would this scenario create systematic or random error (or both or neither)?

This sampling behavior will cause systematic bias, and this experimental design can intuitively reveal that a large lack of elderly sample in the sampling will lead to biased sampling results.

(ii) Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?

In theory, the time spent by elderly people using this smartwatch is expected to be lower, so the lack of elderly samples implies that the diff value will be larger.

(iii) Will it increase or decrease our power to reject the null hypothesis?

Based on the `interactive_t_test()` package, we can observe that as the value of “diff” increases, the resulting t-value is more likely to fall into the rejection region, leading to the rejection of the null hypothesis H_0 .

(iv) Which kind of error (Type I or Type II) becomes more likely because of this scenario?

As it is more likely to fall into the rejection region, the possibility of committing a Type I error (“rejecting the null hypothesis when it is true”) also increases. Therefore, it is more likely to occur when the “diff” value increases.

b. You find that 20 of the respondents are reporting data from the wrong wearable device, so they should be removed from the data. These 20 people are just like the others in every other respect.

(i) Would this scenario create systematic or random error (or both or neither)?

In this scenario, the incorrect use of the device by a participant is a type of experimental error, which leads to systematic error in the results.

(ii) Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?

Since the data from participants who used the device incorrectly will not be included, the sample size (n) will decrease.

(iii) Will it increase or decrease our power to reject the null hypothesis?

From the `interactive_t_test()` package, we can observe that as the sample size decreases, the t-value will move further away from the rejection region, leading to a lower probability of rejecting H_0 .

(iv) Which kind of error (Type I or Type II) becomes more likely because of this scenario?

The decreasing probability of rejecting the null hypothesis means that even if the null hypothesis is false, we may be more likely to fail to reject H_0 . As a result, the probability of committing a Type II error will increase.

c. A very annoying professor visiting your company has criticized your colleague’s “95% confidence” criteria, and has suggested relaxing it to just 90%.

(i) Would this scenario create systematic or random error (or both or neither)?

Adjusting the confidence interval is part of the experimental design, so it may be systematic errors.

(ii) Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?

The adjustment of the confidence interval primarily affects the alpha value, which in turn affects the interpretation of the P-value.

(iii) Will it increase or decrease our power to reject the null hypothesis?

As the alpha value increases, the rejection region will become larger, which may increase the probability of rejecting the null hypothesis.

(iv) Which kind of error (Type I or Type II) becomes more likely because of this scenario?

When the rejection region that leads to rejecting H_0 increases, the probability of Type I error will also increase.

d. Your colleague has measured usage times on five weekdays and taken a daily average. But you feel this will underreport usage for younger people who are very active on weekends, whereas it over-reports usage of older users.

(i) Would this scenario create systematic or random error (or both or neither)?

The adjustment of the sample acquisition source and method as part of experimental design can lead to systematic errors. Additionally, the underestimation and overestimation resulting from this sampling can also be considered as a form of random error.

(ii) Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?

Underestimating the number of adolescent users and overestimating the number of elderly users can both affect the value of diff. In addition, the SD will also be affected due to the fluctuation of the received values.

(iii) Will it increase or decrease our power to reject the null hypothesis?

Although an increase in the diff value makes it more likely for the t-value to enter the rejection region, an increase in the sd value will move the t-value away from the rejection region. Therefore, based on the current clues, it is impossible to determine whether the strength of rejecting H_0 will increase or decrease.

(iv) Which kind of error (Type I or Type II) becomes more likely because of this scenario?

Based on the current information available, both type I and type II errors are possible.