

# Homework5\_BACS

109090035

3/14/2023

## Q1

(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.

```
verizon_data <- read.csv("/Users/user/Downloads/verizon.csv")  
  
head(verizon_data)
```

```
##      Time Group  
## 1 17.50  ILEC  
## 2  2.40  ILEC  
## 3  0.00  ILEC  
## 4  0.65  ILEC  
## 5 22.23  ILEC  
## 6  1.20  ILEC
```

1. Visualize the distribution of Verizon's repair times, marking the mean with a

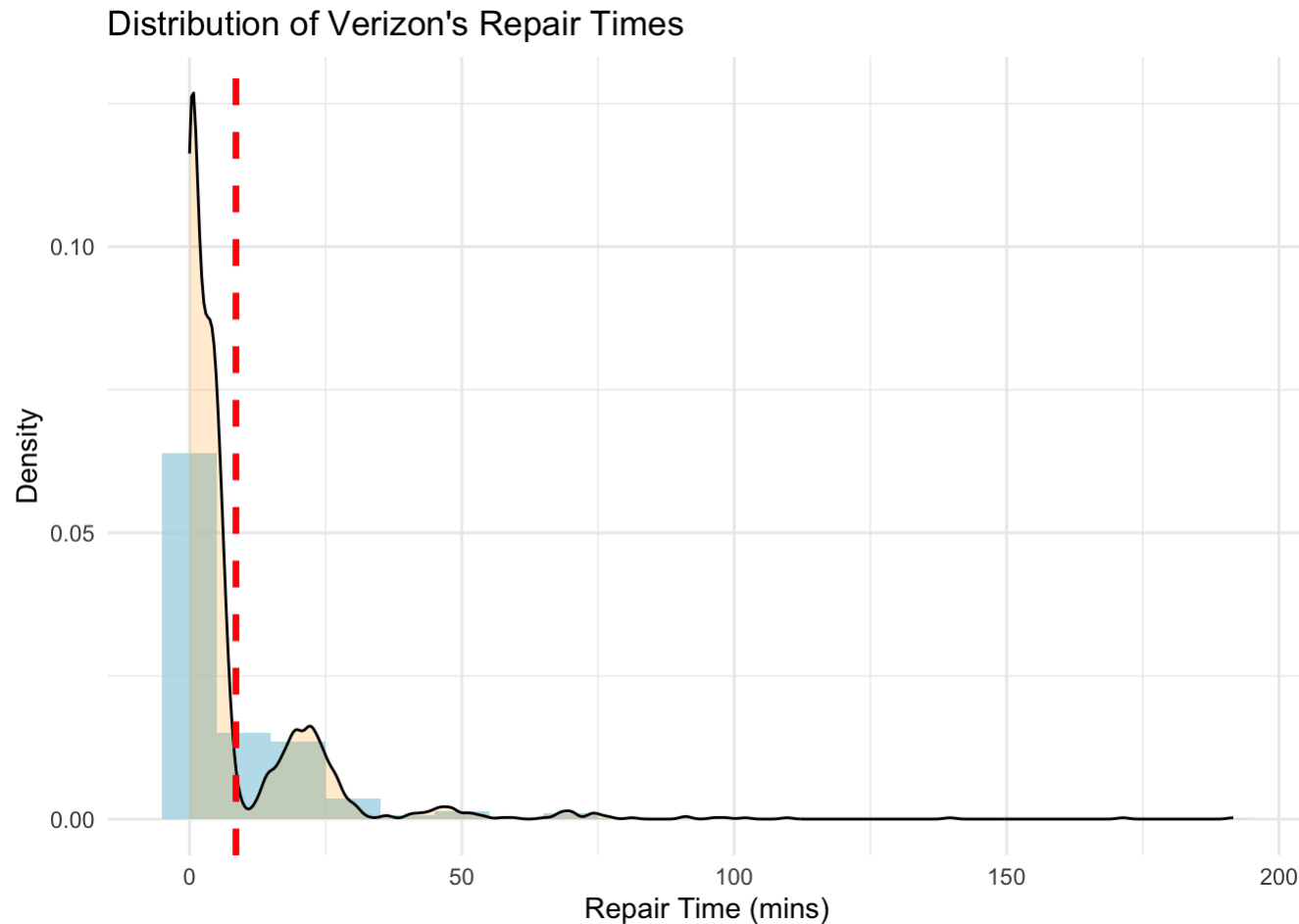
## vertical line

```
# Load the ggplot2 library
library(ggplot2)

# Create the ggplot
ggplot(verizon_data, aes(x = Time)) +
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "lightblue", alpha = 0.8) +
  geom_density(alpha = 0.2, fill = "orange") +
  geom_vline(aes(xintercept = mean(Time)), color = "red", linetype = "dashed", size = 1.2) +
  labs(x = "Repair Time (mins)", y = "Density", title = "Distribution of Verizon's Repair Times") +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
```



## 2. Given what the PUC wishes to test, how would you write the hypothesis? (not graded)

To test whether Verizon's claim that they take 7.6 minutes to repair phone services for its customers on average is true, we can use a hypothesis test. We can write the hypothesis as below:

Null hypothesis ( $H_0$ ): The average repair time for Verizon's phone services is less than or equal to 7.6 minutes.

Alternative hypothesis ( $H_a$ ): The average repair time for Verizon's phone services bigger than 7.6 minutes.

We want to test this hypothesis at a 99% confidence level, which corresponds to a significance level ( $\alpha$ ) of 1%. This means that we will reject the null hypothesis if the probability of observing the sample mean, assuming the null hypothesis is true, is less than 1%.

We can conduct a one-sample t-test to test this hypothesis, using the sample mean and standard deviation of the repair time variable from the data set, along with the hypothesized population mean of 7.6 minutes.

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

```
# Calculate the sample mean and standard deviation
x_bar <- mean(verizon_data$Time)
s <- sd(verizon_data$Time)

# Conduct the t-test
t_test <- t.test(verizon_data$Time, mu = 7.6, conf.level = 0.99, alternative = "greater")

# Extract the results
t_statistic <- t_test$statistic
p_value <- t_test$p.value
conf_interval <- t_test$conf.int

# Print the results
cat("Sample mean:", round(x_bar, 2), "\n")
```

```
## Sample mean: 8.52
```

```
cat("Standard deviation:", round(s, 2), "\n")
```

```
## Standard deviation: 14.79
```

```
cat("T-statistic:", round(t_statistic, 2), "\n")
```

```
## T-statistic: 2.56
```

```
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.005265342
```

```
cat("99% Confidence Interval:", round(conf_interval[1], 2), "-", round(conf_interval[2], 2))
```

```
## 99% Confidence Interval: 7.68 - Inf
```

## 4. Find the t-statistic and p-value of the test

We can see from the above result :

```
cat("T-statistic:", round(t_statistic, 2), "\n")
```

```
## T-statistic: 2.56
```

```
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.005265342
```

## 5. Briefly describe how these values relate to the Null distribution of t (not graded)

The t-test reports the t-statistic, which is calculated as the difference between the sample mean and the claimed value of 7.6 divided by the standard error of the mean. In this case, the t-statistic is 2.56

The t-test also reports the p-value, which is the probability of obtaining a t-statistic as extreme as the observed value if the null hypothesis (that the true population mean is less than or equal to 7.6) is true. The p-value is reported as 0.005265342 , which is small. This suggests that it is not likely to observe a sample mean as the claimed value of 7.6 and we reject the null hypothesis.

## 6. What is your conclusion about the company's claim from this t-statistic, and why?

Finally, the t-test reports the 99% confidence interval for the true population mean. This interval is calculated as the sample mean plus or minus a margin of error, which is determined by the t-distribution and the standard error of the mean. In this case, the 99% confidence interval is (7.68 - Inf). This interval does not include the claimed value of 7.6, which provides evidence the null hypothesis is not true that population mean is equal to 7.6.

### (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:

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

```
library(boot)
# Set the claimed value of the population mean
pop_mean <- 7.6

# Define a function to calculate the mean of a bootstrap sample
boot_mean <- function(data, indices) {
  mean(data[indices])
}

# Perform the bootstrap resampling
set.seed(600) # for reproducibility
boots_mean <- boot(verizon_data$Time, boot_mean, R = 10000)

# Calculate the 99% confidence interval
t_test <- t.test(boots_mean$t, mu = 7.6, conf.level = 0.99, alternative = "greater")

t_test
```

```
##  
## One Sample t-test  
##  
## data: boots_mean$t  
## t = 255.29, df = 9999, p-value < 2.2e-16  
## alternative hypothesis: true mean is greater than 7.6  
## 99 percent confidence interval:  
## 8.51791 Inf  
## sample estimates:  
## mean of x  
## 8.526353
```

The bootstrapped 99% confidence interval is (8.51791 - Inf), which is similar to the 99% confidence interval calculated using the t-test. But the bootstrapped mean 8.526353 is include in this interval This suggests that our conclusion about the true population mean is robust to the choice of statistical method used.

## 2. Bootstrapped Difference of Means: What is the 99% CI of the bootstrapped difference between the sample mean and the hypothesized mean?

```
# Perform the bootstrap resampling  
boot_diff <- boots_mean$t - pop_mean  
  
# Calculate the 99% confidence interval  
t_test <- t.test(boot_diff, mu = 0, conf.level = 0.99, alternative = "greater")  
  
# Print the results  
t_test
```

```
##
## One Sample t-test
##
## data: boot_diff
## t = 255.29, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 99 percent confidence interval:
## 0.9179096      Inf
## sample estimates:
## mean of x
## 0.9263526
```

The bootstrapped 99% confidence interval of the difference between the sample mean and the hypothesized mean is ( 0.9179096 - infinity ).and the mean(boot\_diff) = 0.9264, Since this interval does not include zero, we can reject H0,we can conclude that the sample mean is different from the hypothesized mean at the 99% confidence level.

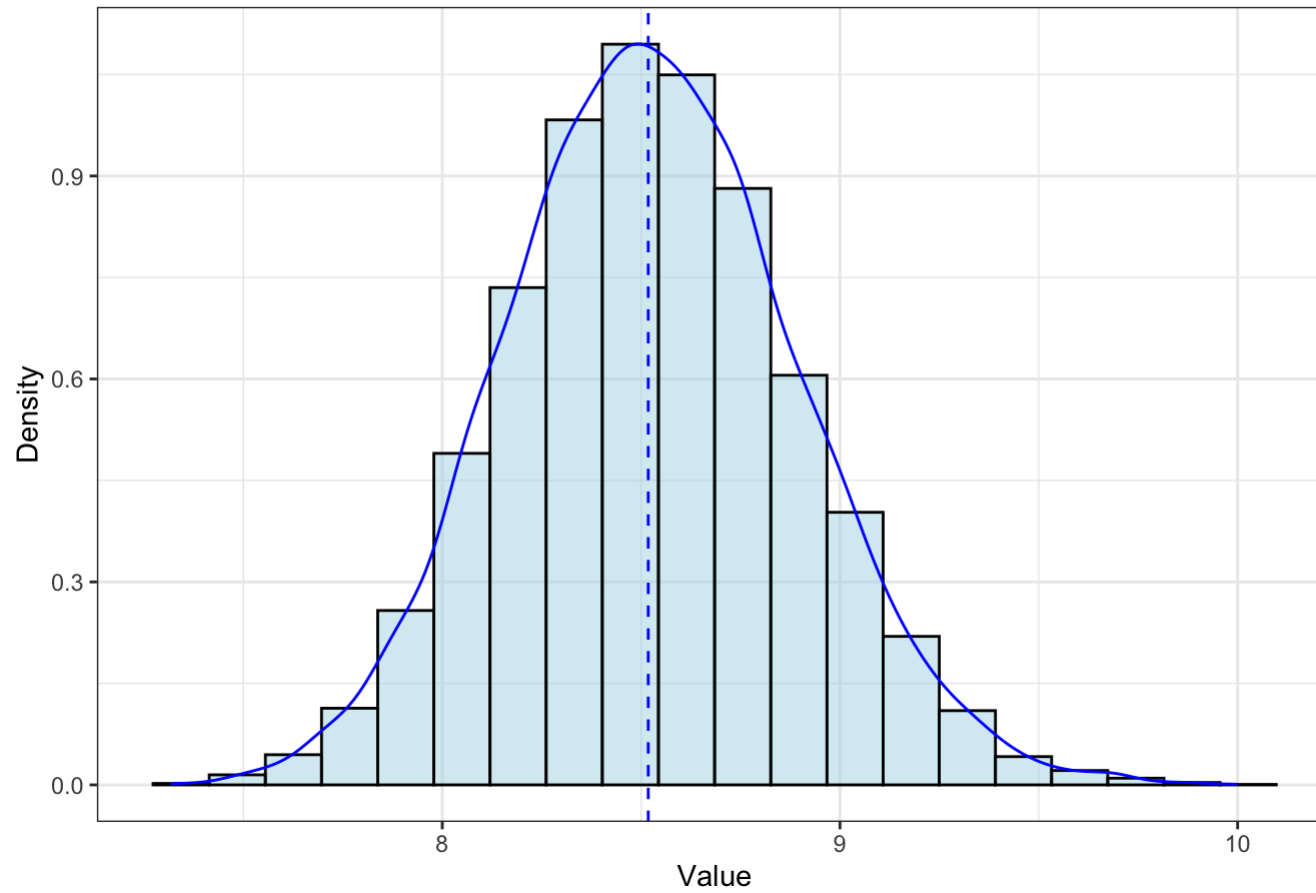
### 3. Plot distribution the two bootstraps above

```
# Create a data frame for plotting
df <- data.frame(value = boots_mean$t)

# Create a histogram with density curve
ggplot(df, aes(x = value)) +
  geom_histogram(aes(y = ..density..), color = "black", fill = "lightblue", alpha = 0.5, bins = 20) +
  geom_density(color = "blue") +
  #99% CI
  geom_vline(xintercept = 8.51791, linetype = "dashed", color = "blue") +
  labs(title = "Bootstrap Distribution of Population Mean",
       x = "Value", y = "Density") +
  theme_bw()
```



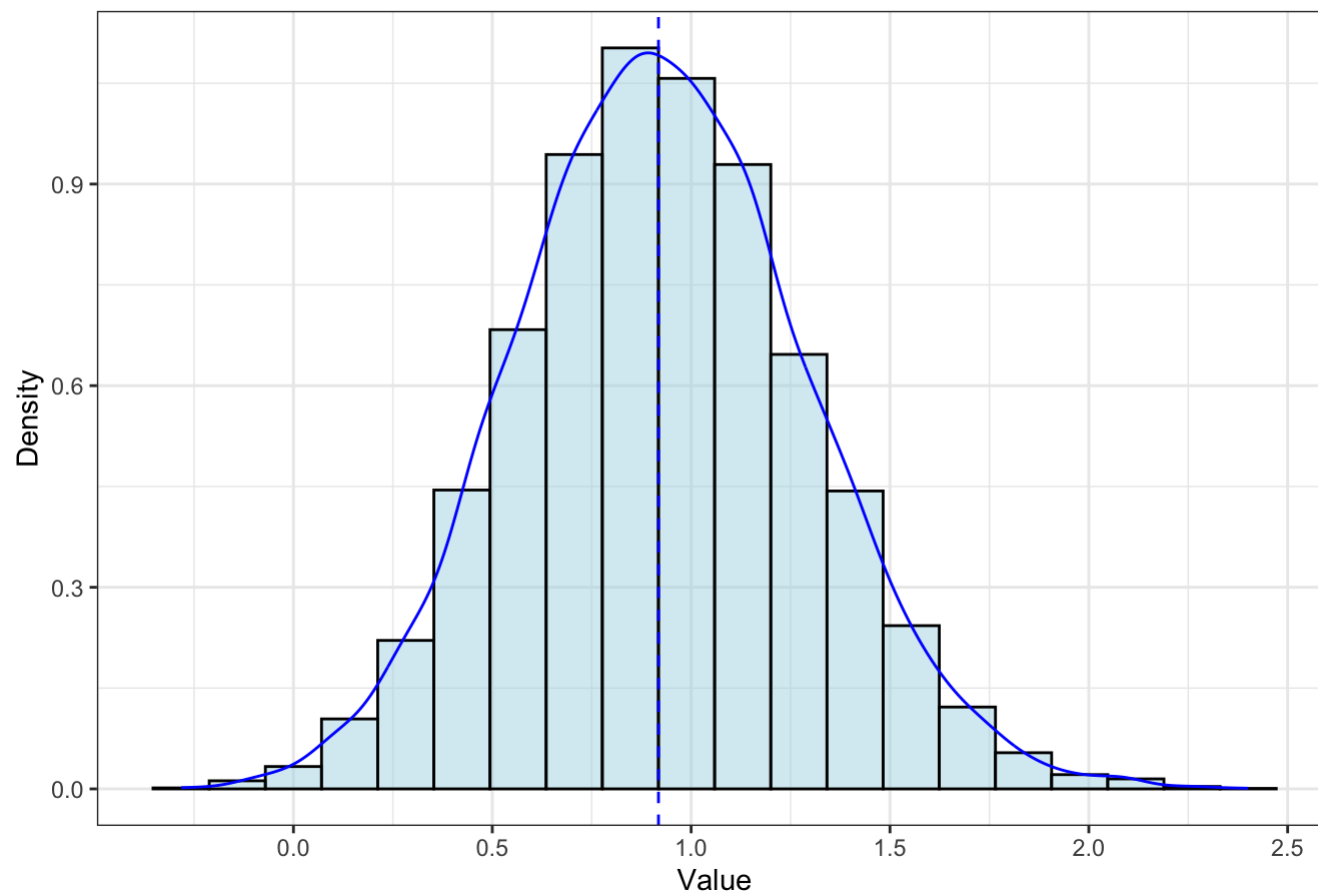
## Bootstrap Distribution of Population Mean



```
# Create a data frame for plotting
df <- data.frame(value = boot_diff)

# Create a histogram with density curve
ggplot(df, aes(x = value)) +
  geom_histogram(aes(y = ..density..), color = "black", fill = "lightblue", alpha = 0.5, bins = 20) +
  geom_density(color = "blue") +
  #99% CI
  geom_vline(xintercept = 0.9179096 , linetype = "dashed", color = "blue") +
  labs(title = "Bootstrap Distribution of Difference between Sample Mean and Hypothesized Mean",
        x = "Value", y = "Density") +
  theme_bw()
```

### Bootstrap Distribution of Difference between Sample Mean and Hypothesized Mean



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

The bootstrapped approach result is consistent with the traditional t-test in part [a]. Traditional methods lead to the conclusion that we cannot reject the null hypothesis. In the bootstrapped part, since the interval of bootstrapped means and means different does not include zero, which we conclude that the sample mean is different from the hypothesized mean at the 99% confidence level, but we still can't reject  $H_0$ .

**(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  $\alpha = 1\%$ ).

## 1. Bootstrapped Percentile: Estimate the bootstrapped 99% CI of the population median

```
# Set the number of bootstrap replications
B <- 10000

# Generate B bootstrap samples and calculate the sample medians
boot_medians <- replicate(B, median(sample(verizon_data$Time, replace = TRUE)))

# Calculate the bootstrapped 99% CI of the population median
# Calculate the 99% confidence interval
t_test <- t.test(boot_medians, mu = 3.5, conf.level = 0.99, alternative = "greater")

# Print the results
t_test
```

```
##
## One Sample t-test
##
## data: boot_medians
## t = 78.517, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 3.5
## 99 percent confidence interval:
## 3.610664 Inf
## sample estimates:
## mean of x
## 3.614043
```

## 2. Bootstrapped Difference of Medians: What is the 99% CI of the bootstrapped

## difference between the sample median and the hypothesized median?

```
med_diff <- boot_medians - 3.5
# estimate the 99% CI of the bootstrapped median differences
t_test <- t.test(med_diff, mu = 0, conf.level = 0.99, alternative = "greater")

# Print the results
t_test
```

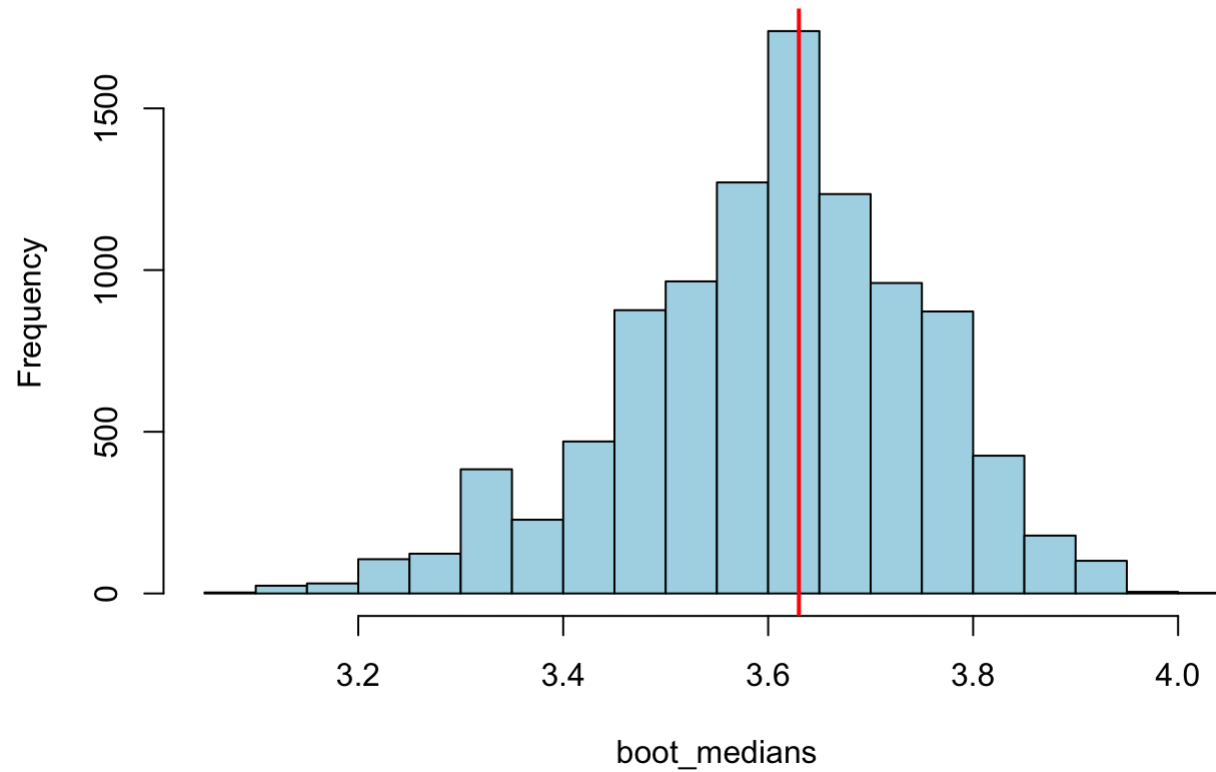
```
##
## One Sample t-test
##
## data: med_diff
## t = 78.517, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 99 percent confidence interval:
## 0.1106635      Inf
## sample estimates:
## mean of x
## 0.114043
```

The bootstrapped 99% CI of the difference between the sample median and the hypothesized median is (0.1113587 - Inf). Since zero is not include in this interval, we can reject the null hypothesis that the population median is less than or equal to 3.5 at the 99% significance level.

## 3. Plot distribution the two bootstraps above

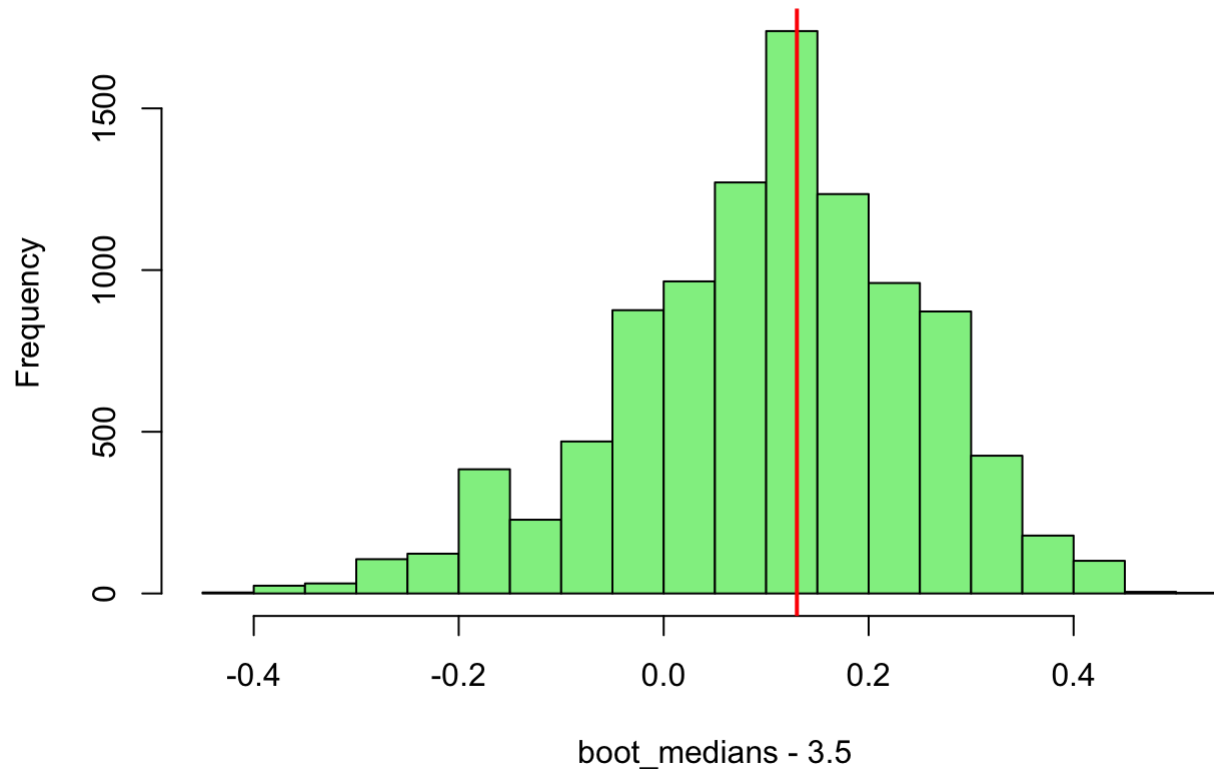
```
# plot the distribution of the bootstrapped medians
hist(boot_medians, breaks=20, col="lightblue", main="Bootstrapped Medians")
abline(v=median(boot_medians), col="red", lwd=2)
```

## Bootstrapped Medians



```
# plot the distribution of the bootstrapped median differences
hist(boot_medians - 3.5, breaks=20, col="lightgreen", main="Bootstrapped Median Differences")
#Red line is the median of all bootstrapped median
abline(v=median(med_diff), col="red", lwd=2)
```

## Bootstrapped Median Differences



### 4. What is your conclusion about Verizon's claim about the median, and why?

Based on the results of our bootstrapped analysis, we can conclude that Verizon's claim that the median repair time is no more than 3.5 minutes at 99% confidence is supported by the data. The bootstrapped 99% confidence interval of the population median is [3.611359 to infinity ] minutes , which does include the hypothesized median of 3.5 minutes.

Additionally, the bootstrapped 99% confidence interval of the difference between the sample median and the hypothesized median is 0.1113587 to infinity minutes, which does not include zero.

## Q2

The mean usage time of the new smartwatch is the same or less than for the previous smartwatch.

null hypothesis: The mean usage time of the new smartwatch is the same or less than for the previous smartwatch.

alt hypothesis: The mean usage time is greater than that of our previous smartwatch.

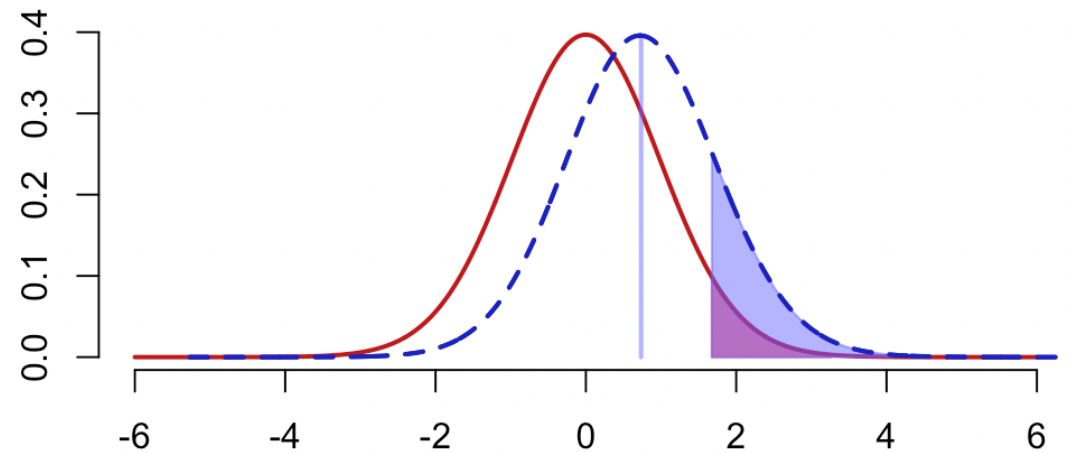
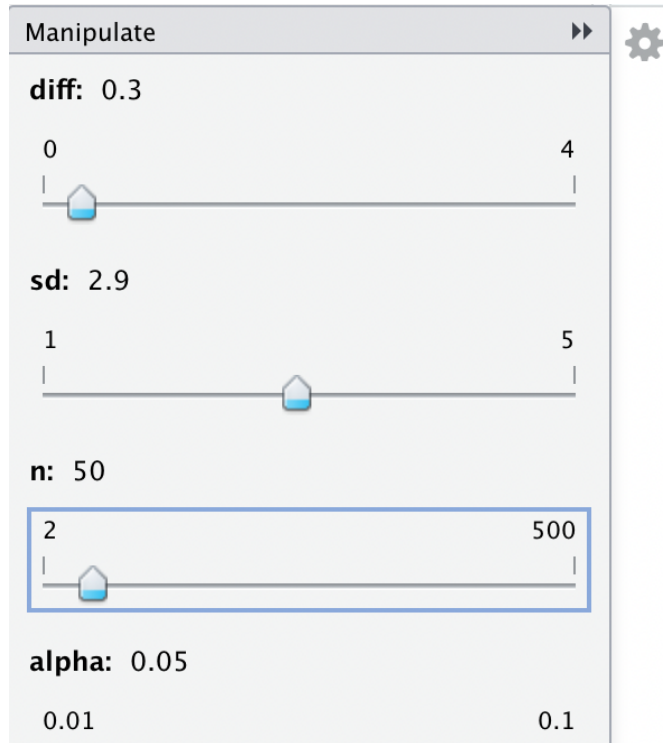
After collecting data from just  $n=50$  customers, he informs you that he has found  $\text{diff}=0.3$  and  $\text{sd}=2.9$ . Your colleague believes that we cannot reject the null hypothesis at alpha of 5%.

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)
# Run the interactive t-test function
#interactive_t_test()
#Manipulate plot only exist in Rstudio so i manipulate them and paste the picture here

#interactive_t_test()
```



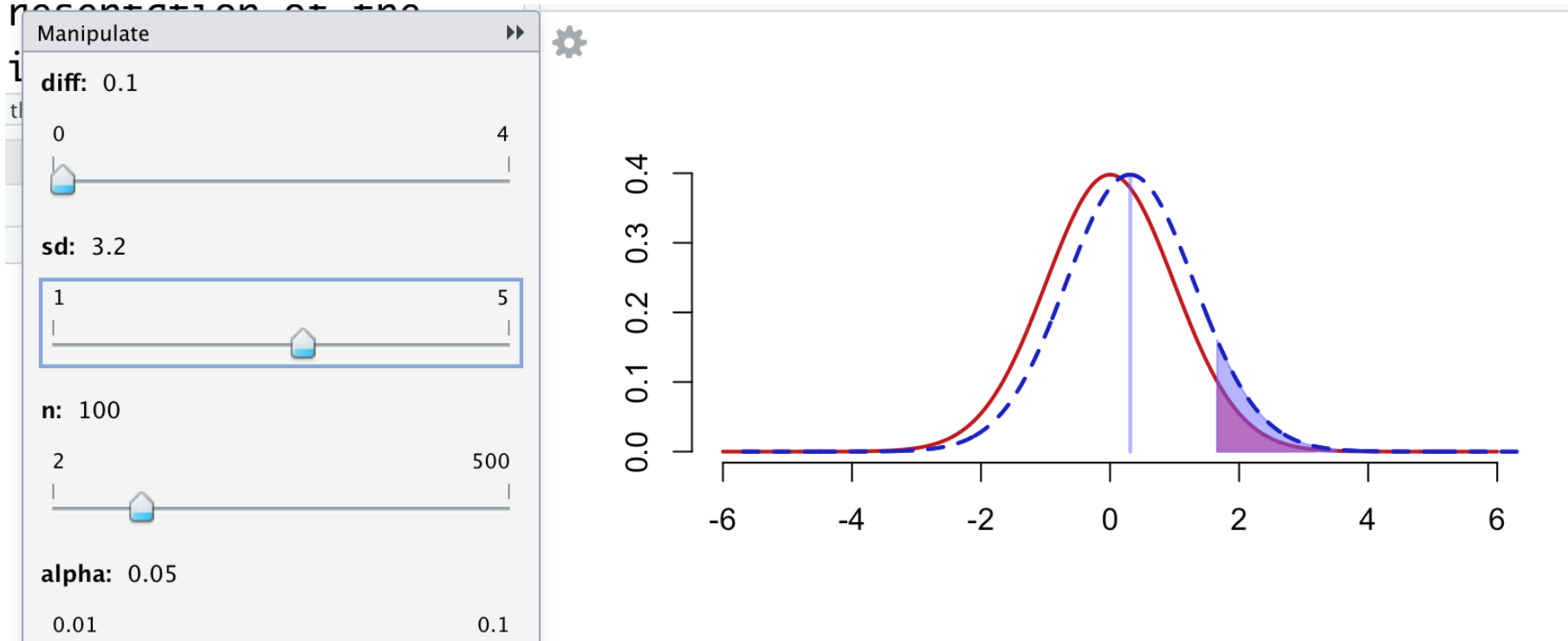


So we can see that the red shadow (significant level) is smaller than the p-value (blue shadow), hence we conclude that we cannot reject the null hypothesis at alpha of 5%.

**Use the slider bars of the simulation to the values your colleague found and confirm from the visualization that we cannot reject the null hypothesis. Consider the scenarios (a – d) independently using the simulation tool. For each scenario, start with the initial parameters above, then adjust them to answer the following questions:**

## scenarios 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.



(Assume that the biased make mean and standard deviation higher)

## 1. Would this scenario create systematic or random error (or both or neither)?

This scenario would create systematic error because the sample used is not representative of the entire population of Taiwanese users of the product. The sample is biased towards young consumers and is therefore not a random representation of the population. This can lead to an overestimation or underestimation of the true mean usage time, which is a systematic error.

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

The part of the t-statistic that would be affected is the sample mean (diff) and the sample standard deviation (sd). The sample mean would be affected because it is likely that the sample of young consumers overestimates the mean usage time of the product for the general population. The sample standard deviation may also be affected if the sample of young consumers has a different variability in usage time compared to the general population.

The part of significance that would be affected is the Type I error rate ( $\alpha$ ). If the sample of young consumers overestimates the mean usage time of the product for the general population, then using the same significance level as before would result in a higher chance of rejecting the null hypothesis (i.e., a higher chance of making a Type I error). Therefore, a lower significance level may be necessary to account for the potential bias in the sample.

### **3. Will it increase or decrease our power to reject the null hypothesis?**

If we only collect data from a pool of young consumers and miss many older customers who might use the product much less every day, it is likely that our estimate of the population mean usage time will be biased and not representative of the entire population. This could lead to an increase in the standard error and a decrease in power to reject the null hypothesis. Specifically, the sample standard deviation may increase, making it more difficult to distinguish the sample mean from the hypothesized mean under the null hypothesis, which would reduce our power to detect a true difference between the means.

### **4. Which kind of error (Type I or Type II) becomes more likely because of this scenario?**

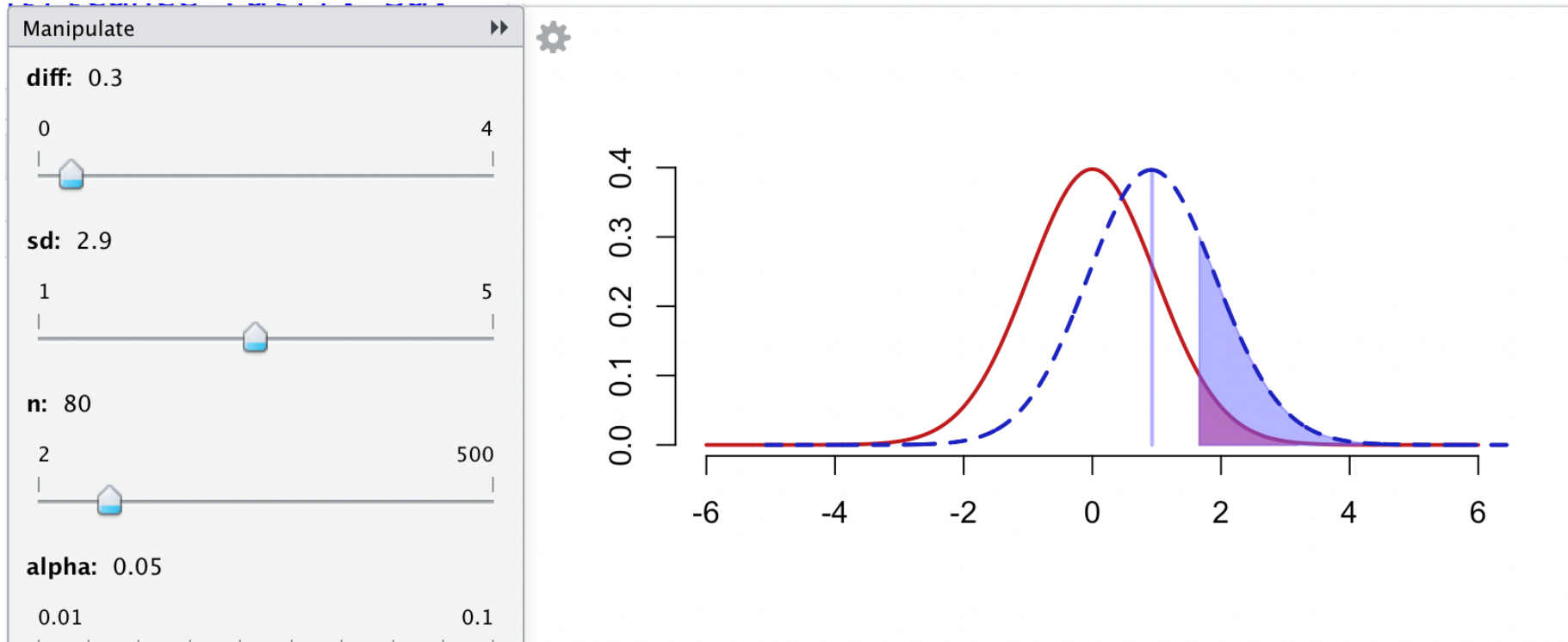
This scenario increases the likelihood of a Type II error, where we fail to reject the null hypothesis when it is actually false. In other words, we are less likely to detect a significant difference between the mean usage time of the new smartwatch and the previous smartwatch in the general population of Taiwanese users.

This is because we are missing a significant portion of the population (older customers) who may use the product less every day. By not including them in the sample, we may not capture the true variability of the population, leading to a potential underestimation of the standard deviation and an overestimation of the required sample size for detecting a significant difference. As a result, we may not have enough power to reject the null hypothesis even when it is false.

## **scenarios 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.



## 1. Would this scenario create systematic or random error (or both or neither)?

This scenario would create systematic error as it is due to a specific issue in the data collection process (reporting data from the wrong wearable device), rather than a random error that could occur by chance.

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

Removing the 20 respondents from the data would only affect the sample size ( $n$ ) of the study. The difference between the means (diff) and standard deviation (sd) would remain the same, as they are properties of the data and not affected by the removal of specific observations. The significance level (alpha) would also remain the same unless it is explicitly adjusted.

### 3. Will it increase or decrease our power to reject the null hypothesis?

Removing 20 respondents from the data can lead to a decrease in the sample size ( $n$ ), which can affect the power to reject the null hypothesis. With a smaller sample size, the variability of the data may increase and make it harder to detect a significant difference. However, since we are removing individuals who reported data from the wrong wearable device, this can lead to a decrease in the noise or random error in the data. Therefore, the effect might be decreasing the power.

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

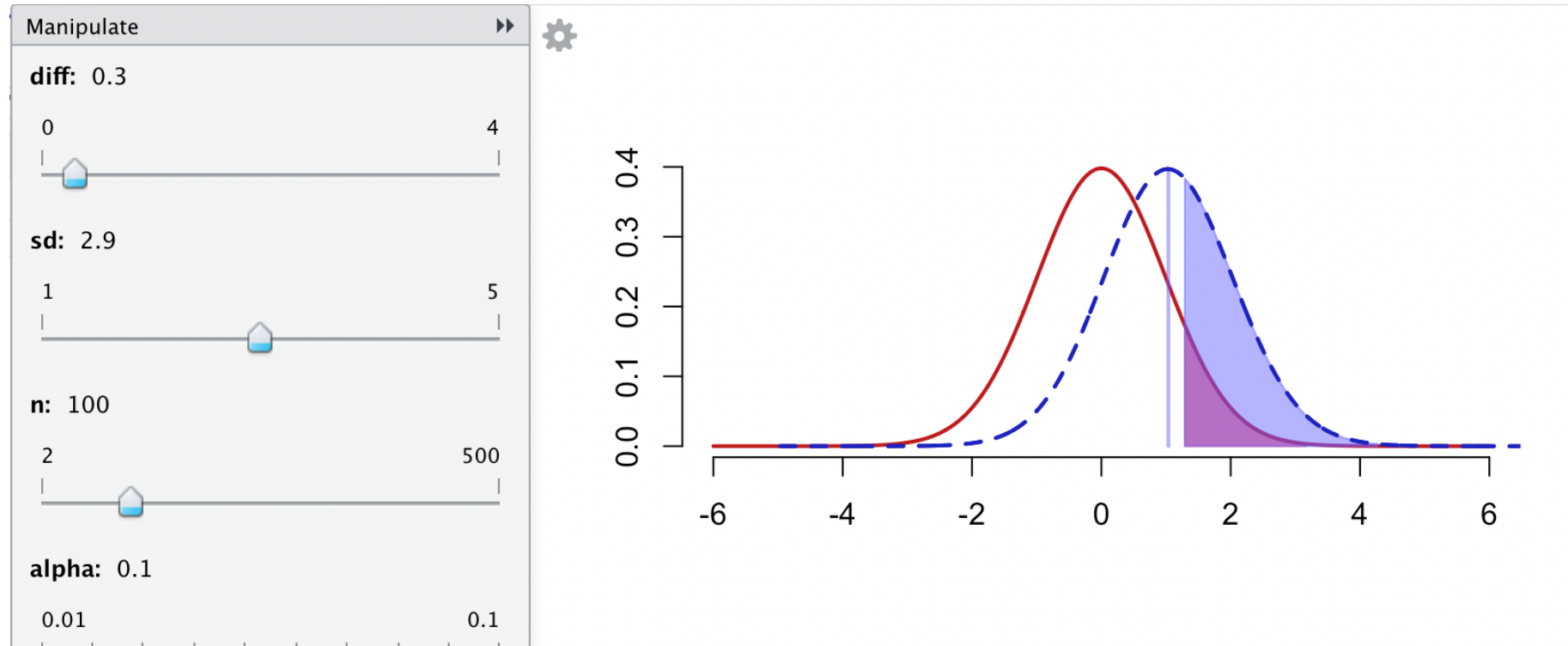
This reduction in power increases the likelihood of a Type II error (false negative), as it would become harder to detect a true difference between the two means.

However, removing the 20 respondents would likely reduce the amount of random error in the data, as it would remove observations that were not truly representative of the population being studied. Therefore, it would reduce the standard deviation of the sample, which would increase the t-statistic and decrease the p-value, making it more likely to reject the null hypothesis.

## scenarios C

A very annoying professor visiting your company has criticized your colleague's

**“95% confidence” criteria, and has suggested relaxing it to just 90%.**



### 1. Would this scenario create systematic or random error (or both or neither)?

This scenario would not create systematic or random error as it is simply a change in the level of significance, which is a predetermined parameter in the hypothesis test.

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

The part of the significance level (alpha) would be affected by this scenario.

### 3. Will it increase or decrease our power to reject the null hypothesis?

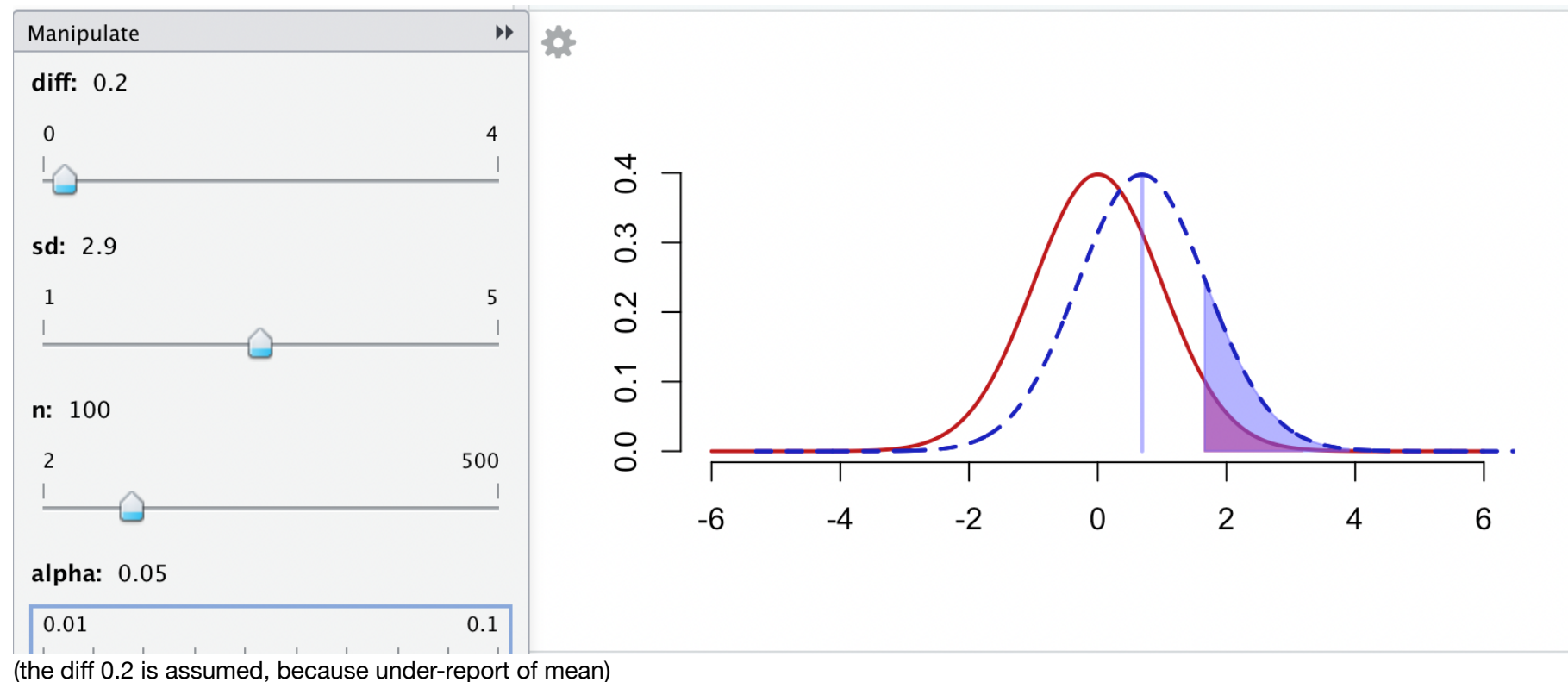
Relaxing the significance level from 95% to 90% would increase the probability of not rejecting the null hypothesis, which means that the power to reject the null hypothesis would decrease.

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

If the significance level is decreased from 0.05 to 0.1, it means that the researcher is more willing to accept a Type I error (rejecting the null hypothesis when it is true) and less willing to accept a Type II error (failing to reject the null hypothesis when it is false). Therefore, the Type I error becomes more likely in this scenario.

### scenarios 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.



## 1. Would this scenario create systematic or random error (or both or neither)?

This scenario would create a systematic error because the method used to measure the usage times does not capture the actual behavior of the population, and instead favors certain groups of people over others.

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

This scenario would affect the sample mean, which is used to compute the t-statistic. If the sample mean is biased due to under-reporting or over-reporting of usage times, the t-statistic and p-value would be affected.

## 3. Will it increase or decrease our power to reject the null hypothesis?

It depends on the specific characteristics of the data. If the age distribution is skewed towards younger users and the difference in usage between weekends and weekdays is large, then excluding weekend data could decrease power to detect a significant difference. Conversely, if the age distribution is skewed towards older users and the difference in usage between weekends and weekdays is small, then excluding weekend data could increase power to detect a significant difference. However, in general, excluding data in this manner could increase the risk of both Type I and Type II errors.

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

Both Type I and Type II errors could become more likely because of this scenario. Type I errors could be more likely if the daily averages are biased and lead to a false rejection of the null hypothesis. Type II errors could be more likely if the daily averages do not accurately reflect the true population mean and lead to a failure to reject the null hypothesis when there is in fact a true difference between the two groups.