BACS HW WEEK 5

109090046 helped by 109090035

2023-03-16

Question 1)

The following problem's data is real but the scenario is strictly imaginary The large American phone company Verizon had a monopoly on phone services in many areas of the US. The New York Public Utilities Commission (PUC) regularly monitors repair times with customers in New York to verify the quality of Verizon's services. The file `verizon.csv` has a recent sample of repair times collected by the PUC.

```
library(readr)

## Warning: 套件 'readr' 是用 R 版本 4.2.2 來建造的

verizon <- read_csv("D:/下載/verizon.csv")

## Rows: 1687 Columns: 2
## ── Column specification ─────────────────────────────────────────────
## ───────────
## Delimiter: ","
## chr (1): Group
## dbl (1): Time
##
## ℹ Use `spec()` to retrieve the full column specification for this da
ta.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
this message.
```

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

```
library(ggplot2)

## Warning: 套件 'ggplot2' 是用 R 版本 4.2.2 來建造的

ggplot(verizon, aes(x = Time)) +
  geom_density(alpha = 0.8, fill = "lavender", linewidth = 0.8) +
  geom_vline(aes(xintercept = mean(Time)), color = "blue", linewidth =
1, linetype = "longdash") +
```

```
  labs(x = "Repair Time (mins)", y = "Density", title = "Distribution o
f Verizon's Repair Times") +
  theme_minimal()
```


Distribution of Verizon's Repair Times
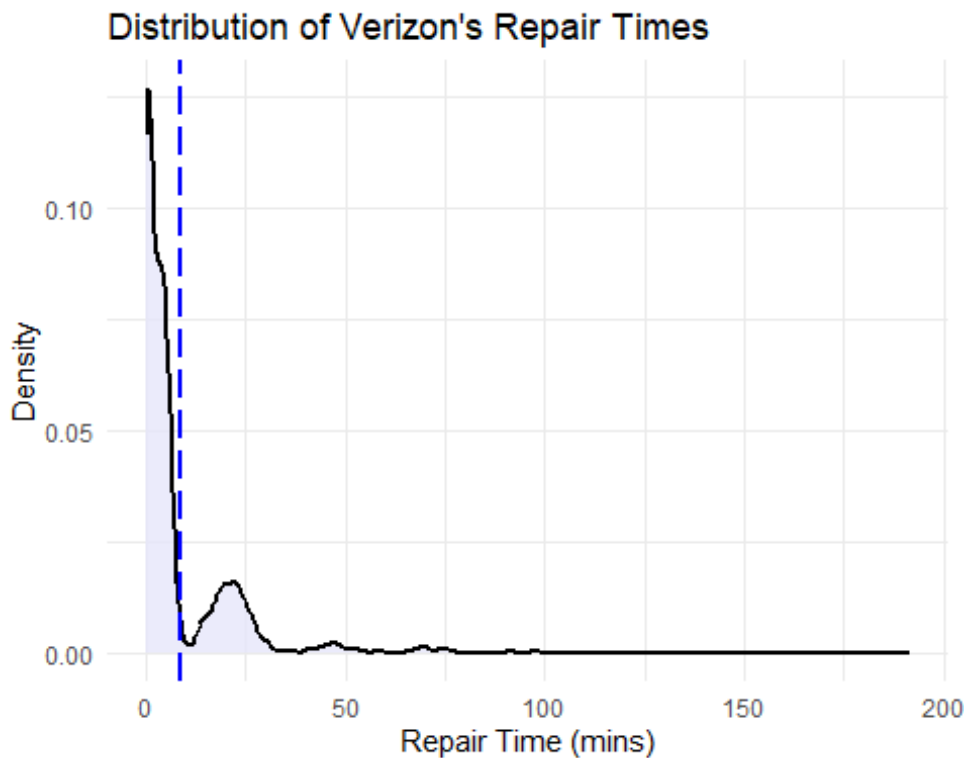
ii. Given what the PUC wishes to test, how would you write the hypothesis? (not graded)

Ans:

Null hypothesis (H0): The true population mean repair time is equal to 7.6 minutes. Alternative hypothesis (Ha): The true population mean repair time is different from 7.6 minutes.

If the p-value is less than the significance level (α), reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis. Otherwise, fail to reject the null hypothesis and conclude that there is not enough evidence to support the alternative hypothesis.

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

```
pop_mean <- mean(verizon$Time)
t_result <- t.test(verizon$Time, mu = 7.6, conf.level = 0.99, alternati
ve="greater")
lb <- t_result$conf.int[1]
ub <- t_result$conf.int[2]
```

```
## The population mean is 8.522009 and the 99% confidence interval (CI)
 is [ 7.683604 , Inf ]
```

iv. Find the t-statistic and p-value of the test

```
t_result$statistic
t_result$p.value

## t-statistic is 2.560762 and p-value is 0.005265342
```

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

Ans:

The t-statistic is the number of standard errors that the sample mean is away from the hypothesized population mean under the null hypothesis. It indicates that the sample mean is significantly different from the hypothesized population mean of 7.6 minutes.

Because the p-value is much smaller than the significance level (α = 0.01), we reject the null hypothesis and conclude that there is sufficient evidence to support the alternative hypothesis. This means that the true population mean repair time is different from 7.6 minutes.

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

```
# critical value
qt(0.99, df=length(verizon$Time)-1)

## [1] 2.328562
```

If the calculated t-statistic exceeds the critical value, we can conclude that it is large and that the sample mean is significantly different from the hypothesized population mean at the chosen level of significance. Based on the t-statistic of 2.56 > critical value of 2.328562, we can conclude that there is evidence against the company's claim that they take 7.6 minutes to repair phone services for its customers on average.

---

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

```
library(boot)

## Warning: 套件 'boot' 是用 R 版本 4.2.2 來建造的
```

```
set.seed(1234)

# function to calculate the mean of each bootstrap sample
boot_mean <- function(sample0, i) {
  mean(sample0[i])
}

# resampling with 10000 times
boot_means <- boot(verizon$Time, boot_mean, R = 10000)

t_result <- t.test(boot_means$t, mu = 7.6, conf.level = 0.99,alternativ
e = "greater")

t_result

##
##   One Sample t-test
##
## data:  boot_means$t
## t = 257.09, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 7.6
## 99 percent confidence interval:
##   8.515195       Inf
## sample estimates:
## mean of x
##   8.523553

## Bootstrapped 99% CI of population mean is [ 8.515195 , Inf ]
```

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

```
boot_diff_means <- boot_means$t - 7.6

t_result <- t.test(boot_diff_means, mu = 0, conf.level = 0.99,alternati
ve = "greater")

t_result

##
##   One Sample t-test
##
## data:  boot_diff_means
## t = 257.09, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 99 percent confidence interval:
##   0.9151948       Inf
## sample estimates:
## mean of x
## 0.9235531
```

```
## Bootstrapped 99% CI of the difference between the sample mean and th
e hypothesized mean: [ 0.9151948 ,   Inf ]
```

iii. Plot distribution the two bootstraps above on two separate plots.

```
boot_means_df <- data.frame(value = boot_means$t)

mean(boot_means$t)

## [1] 8.523553

ggplot(boot_means_df, aes(x = value)) +
  geom_density(color = "black", linewidth=0.8, fill="lavender", alpha=0.
8) +
  geom_vline(xintercept = mean(boot_means$t), linetype = "longdash", co
lor = "blue", linewidth=1)  +
  labs(title = "Bootstrap Distribution of Population Mean",
       x = "Time", y = "Density") +
  theme_bw()
```
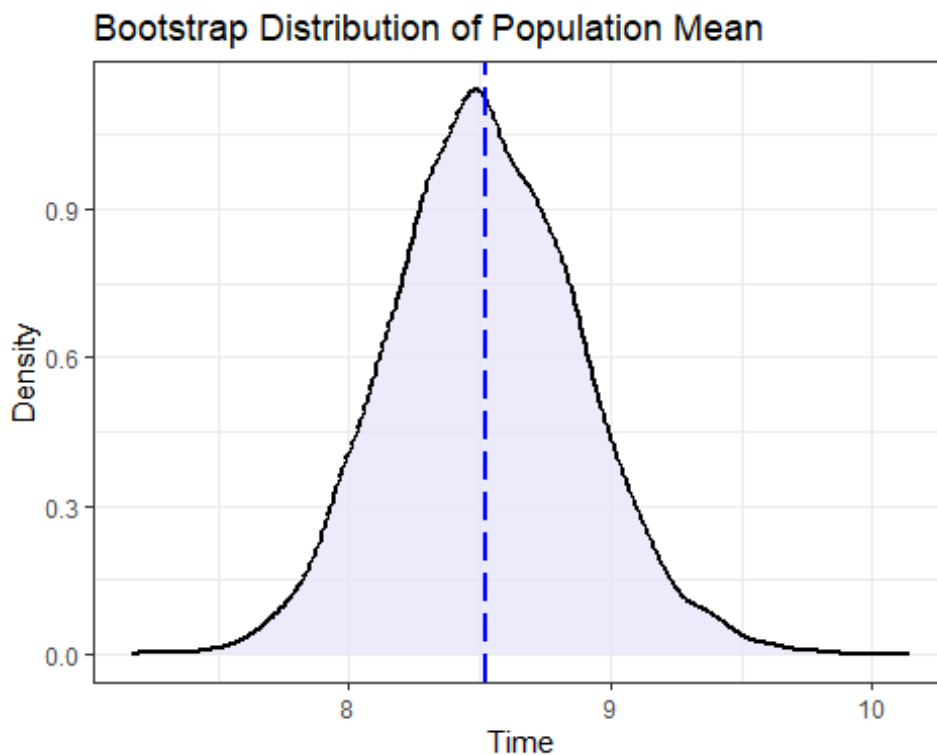


```
boot_diff_df <- data.frame(value = boot_diff_means)

mean(boot_diff_means)

## [1] 0.9235531

ggplot(boot_diff_df, aes(x = value)) +
  geom_density(color = "black", linewidth=0.8, fill="lavender", alpha=0.
```

```
8) +
  geom_vline(xintercept = mean(boot_diff_means) , linetype = "longdash",
 color = "blue", linewidth=1)  +
  labs(title = "Bootstrap Distribution of Difference of Means",
       x = "Value(Means - 7)", y = "Density") +
  theme_bw()
```

Bootstrap Distribution of Difference of Means
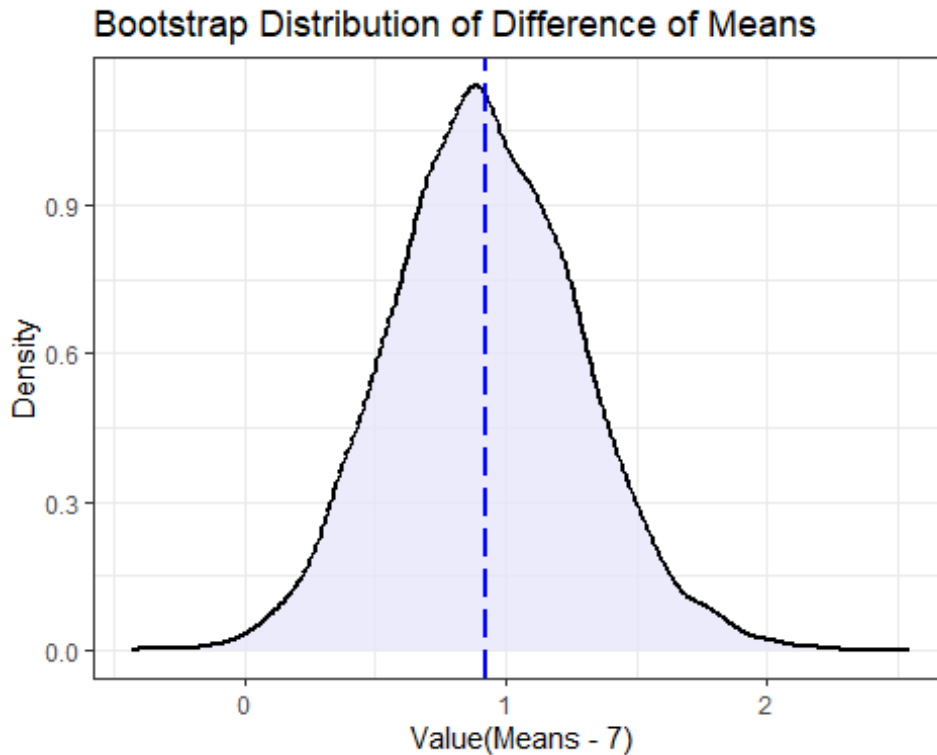
iv. Does the bootstrapped approach agree with the traditional t-test in part [a]?

Ans: The bootstrapped approach result is consist with the traditional t-test in part [a]. Traditional method concluded that we cannot reject H0. However, in the bootstrapped part, since the interval of bootstrapped means and means different does not include zero ,which we can conclude that the sample mean is different from the hypothesized mean at the 99% CI, but we still can't reject H0.

---

c.

Finally, imagine that Verizon notes that the distribution of repair times is highly skewed by outliers, and feel that testing the mean in 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

```
boot_median <- function(sample0, i) {
  median(sample0[i])
}

boot_meds <- boot(verizon$Time, boot_median, R = 10000)

# Calculate the 99% confidence interval
t_test <- t.test(boot_meds$t, mu = 3.5, conf.level = 0.99,alternative =
 "greater")

t_test

##
##   One Sample t-test
##
## data:  boot_meds$t
## t = 79.105, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 3.5
## 99 percent confidence interval:
##   3.610571       Inf
## sample estimates:
## mean of x
##   3.613922

## Bootstrapped 99% CI of population median is [ 3.610571 , Inf ]
```

ii. *Bootstrapped Difference of Medians*: What is the 99% CI of the bootstrapped difference between the sample median and the hypothesized median?

```
boot_diff_meds <- boot_meds$t - 3.5

t_test <- t.test(boot_diff_meds, mu = 0, conf.level = 0.99, alternative
 = "greater")

t_test

##
##   One Sample t-test
##
## data:  boot_diff_meds
## t = 79.105, df = 9999, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 0
## 99 percent confidence interval:
##   0.1105712       Inf
## sample estimates:
## mean of x
##   0.113922

## Bootstrapped 99% CI of the difference between the sample median and
the hypothesized median: [ 0.1105712 ,  Inf ]
```

iii. Plot distribution the two bootstraps above on two separate plots.

```
boot_meds_df <- data.frame(value = boot_meds$t)

mean(boot_meds$t)

## [1] 3.613922

ggplot(boot_meds_df, aes(x = value)) +
  geom_density(color = "black", linewidth=0.8, fill="lavender", alpha=0.
8) +
  geom_vline(xintercept = mean(boot_meds$t), linetype = "longdash", col
or = "blue", linewidth=1)  +
  labs(title = "Bootstrap Distribution of Population Median",
       x = "Time", y = "Density") +
  theme_bw()
```
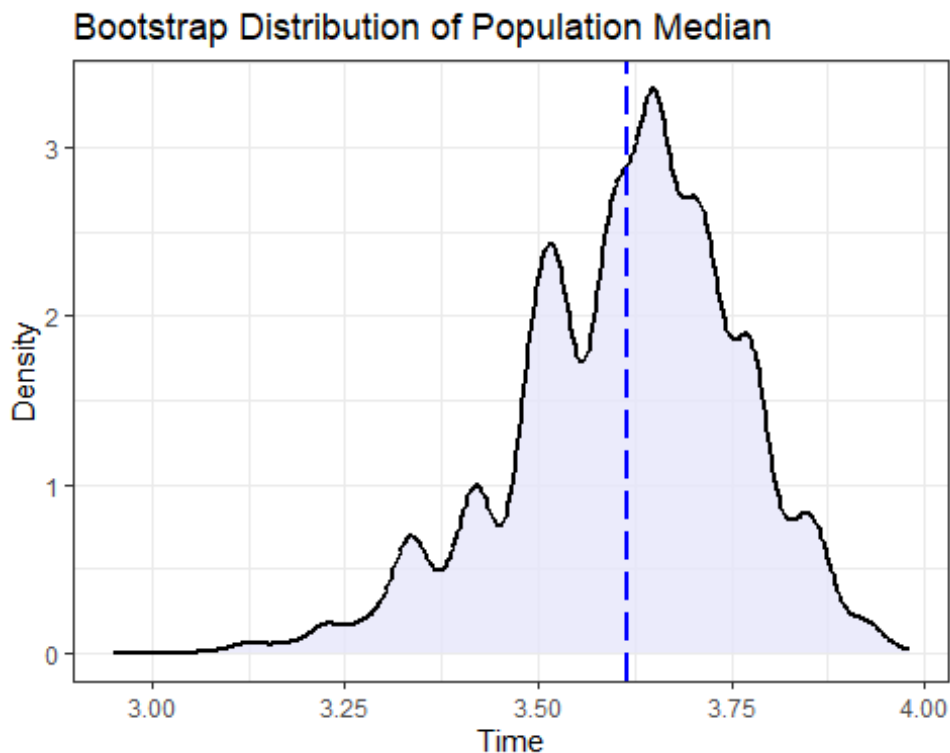


```
boot_diff_meds_df <- data.frame(value = boot_diff_meds)

mean(boot_diff_meds)

## [1] 0.113922

ggplot(boot_diff_meds_df, aes(x = value)) +
  geom_density(color = "black", linewidth=0.8, fill="lavender", alpha=0.
8) +
  geom_vline(xintercept = mean(boot_diff_meds), linetype = "longdash",
color = "blue", linewidth=1)  +
```
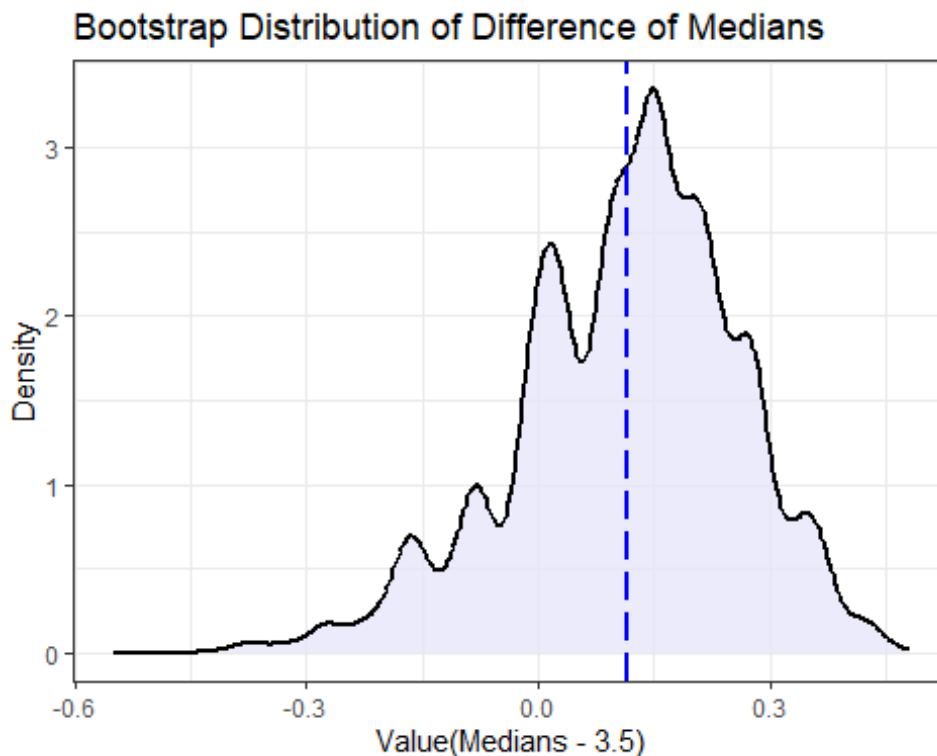
```
labs(title = "Bootstrap Distribution of Difference of Medians",
     x = "Value(Medians - 3.5)", y = "Density") +
theme_bw()
```


Bootstrap Distribution of Difference of Medians

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

Ans:

Based on the result from bootstrapped analysis, the interval [3.610571, infinity) does not overlap with the value 3.5, which means that we can reject Verizon's claim that the median repair time is no more than 3.5 minutes at 99% confidence. In other words, the result suggests that the true median repair time is likely to be greater than 3.5 minutes, with a high degree of certainty (99% confidence). This conclusion is based on the fact that the bootstrapped confidence interval does not include the value 3.5, which means that the probability of observing a median repair time less than or equal to 3.5 is very low, given the repair time data.

Question 2)

Your colleague, a data analyst in your organization, is working on a hypothesis test where he has sampled product usage information from customers who are using a new smartwatch. He wishes to test whether the mean (xi) usage time is higher than the usage time of the company's previous smartwatch released two years ago (o):
Hnull: The mean usage time of the new smartwatch is the same or less than for the

previous smartwatch. Halt: 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 diff=0.3 and sd=2.9. Your colleague believes 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:

```
library(compstatslib)
#interactive_t_test()
```

Image of n=50. diff=0.3, sd=2.9, alpha=0.5
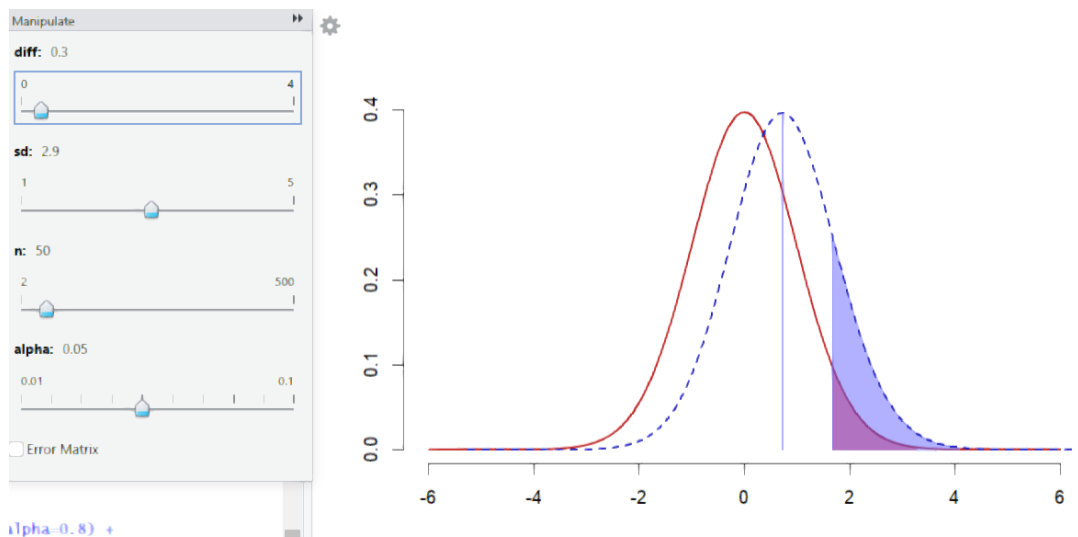
Since the significant level(red shadow) < p-value(blue shadow), we cannot reject H0.

Scenario 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.

Manipulate

diff: 0.1

0   4

sd: 2.9

1   5

n: 50

2   500

alpha: 0.05

0.01   0.1

☐ Error Matrix
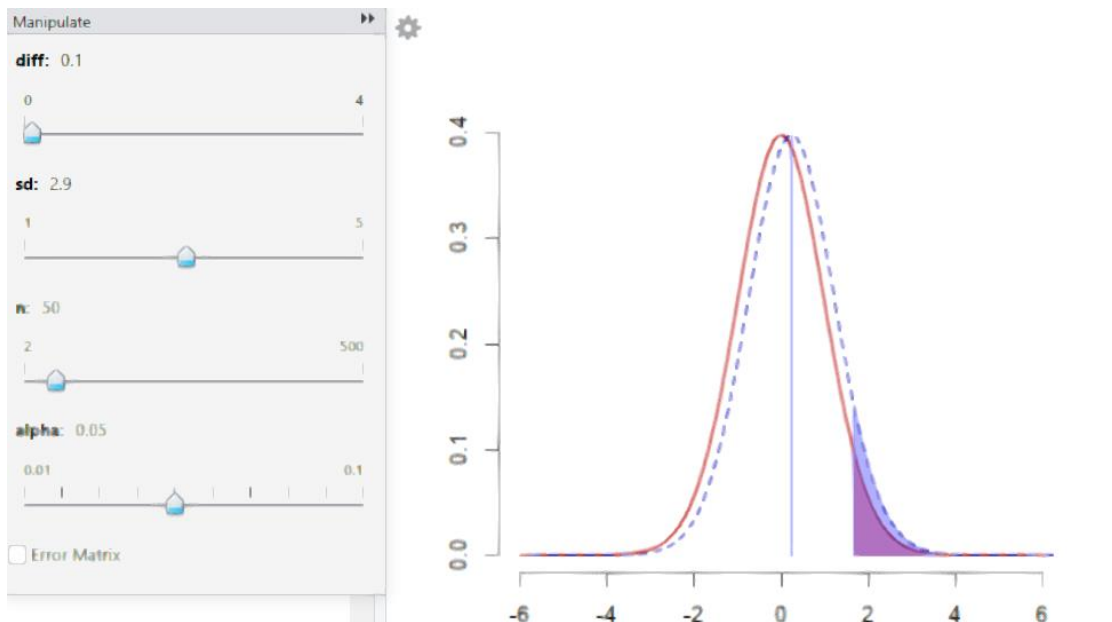
Adjusted: `dif=0.3` to `diff=0.1`, since it may be less diff if all the data are from young man.

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

Ans: This scenario is likely to create systematic error because the data collection method used by the colleague was biased towards young consumers and did not include a representative sample of the entire population of Taiwanese users.

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

Ans: The t-statistic and significance would be affected by this scenario because the sample used to calculate these statistics is not representative of the entire population. The difference in means (diff), standard deviation (sd), and sample size (n) may all be affected by the biased sample, which could impact the validity and reliability of the statistical analysis.

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

Ans: The power to reject the null hypothesis may decrease because the biased sample may not accurately represent the entire population, leading to a higher likelihood of a Type II error (failing to reject the null hypothesis when it is actually false).

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

Ans: This scenario increases the likelihood of a Type II error because the biased sample may not accurately represent the entire population, leading to a higher probability of failing to reject the null hypothesis when it is actually false. In other words, the statistical analysis based on this biased sample may not detect a significant difference between the groups even if there is a true difference in usage

patterns between young and old consumers. As a result, there is a higher probability of failing to reject the null hypothesis (i.e., concluding that there is no difference between the groups) even when there is a true difference.

Scenario 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.

Adjusted: n=50 to n=30.

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

Ans: These errors are random because the 20 respondents are not systematically different from the others in the sample, but the error arises from a random misreporting of data. However, if the 20 respondents who reported data from the wrong device were systematically different from the others in some way (e.g. they all had a particular demographic characteristic), then removing them could introduce systematic errors in the sample.

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

Ans: The sample size (n) and the sample standard deviation (sd) would be affected by removing the 20 respondents from the data. The difference between the means (diff) and the significance level (alpha) would not be affected.

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

Ans: Removing 20 respondents could potentially decrease the power to reject the null hypothesis, as the sample size would be reduced, resulting in less statistical power. However, it could also increase the power to detect a true effect, as removing data points that are not representative of the population could reduce noise and increase the accuracy of the sample.

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

Ans: This scenario does not necessarily increase the likelihood of a Type I or Type II error, but it can affect the accuracy and reliability of the results. Removing 20 respondents who reported data from the wrong wearable device could potentially increase the accuracy of the sample and reduce measurement error.

Scenario C

A very annoying professor visiting your company has criticized your colleague's "95% confidence" criteria, and has suggested relaxing it to just 90%.
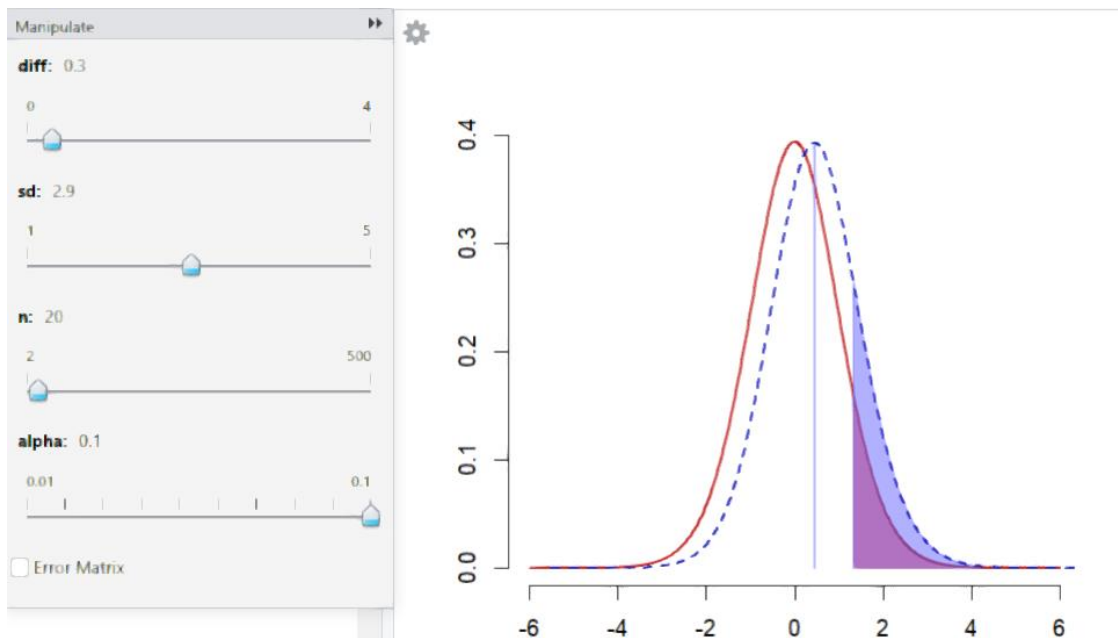
Adjusted: `alpha=0.05` to `alpha=0.1`.

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

Ans: Relaxing the confidence level from 95% to 90% would not necessarily introduce systematic errors, but it could introduce random errors into the analysis. The error would be random because it arises from a change in the statistical analysis procedure, rather than from any systematic difference in the sample or the population.

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

Ans: The part of the t-statistic or significance that would be affected by changing the confidence level is the critical value that is used to determine whether the test statistic is in the rejection region or not. The critical value is based on the chosen alpha level (0.05 for a 95% confidence interval, and 0.1 for a 90% confidence interval), the degrees of freedom (n-1), and the direction of the alternative hypothesis (one-tailed or two-tailed).

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

Ans: By reducing the confidence level from 95% to 90%, the critical value will become smaller, making it easier to reject the null hypothesis. This means that the power to reject the null hypothesis will increase.

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

Ans: It is more likely to commit a Type I error (rejecting the null hypothesis when it is actually true). This is because the critical value used to reject the null hypothesis becomes smaller, increasing the probability of rejecting the null hypothesis even when it is true. However, by reducing the risk of a Type II error, which is failing to

reject the null hypothesis when it is actually false, the power to detect a true effect also increases. Therefore, changing the confidence level should be done thoughtfully, taking into consideration the specific research question and the potential costs of both types of errors.

Scenario 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.
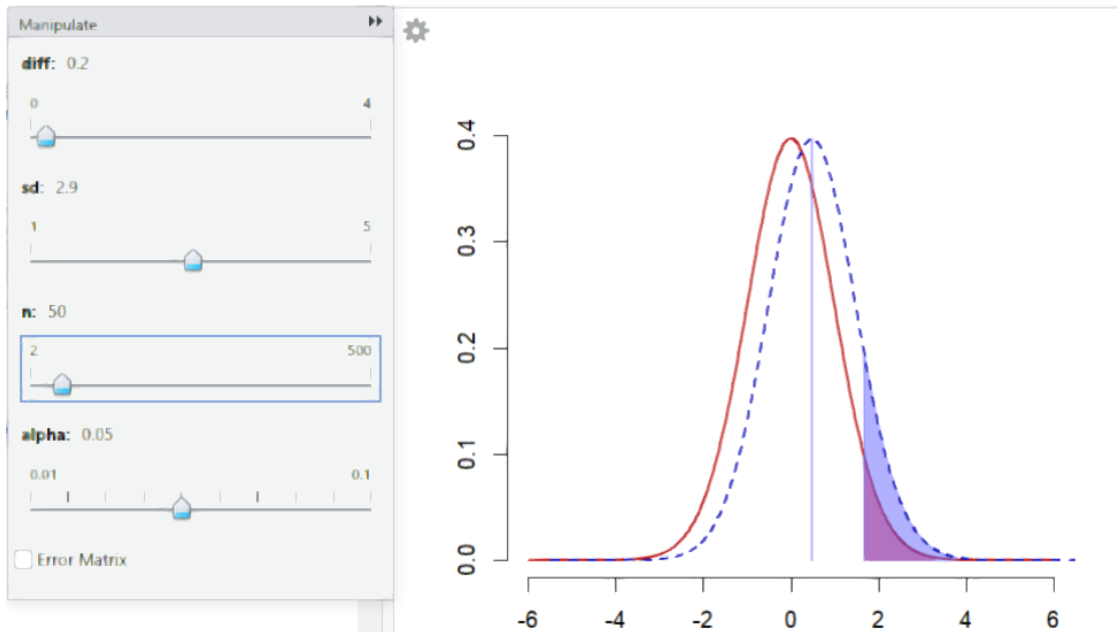
Adjusted: diff=0.3 to diff=0.2 because of underreports usage.

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

Ans: This scenario could create both systematic and random errors. By only measuring usage times on weekdays, the data is systematically biased against younger people who are more active on weekends. This is a systematic error because it is a consistent bias in the measurement process that affects a specific subgroup of the population. On the other hand, the over-reporting of usage times for older users could be a random error if it is due to individual variations in reporting rather than a consistent bias in the measurement process.

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

Ans: The part of the t-statistic or significance that would be affected depends on how the bias affects the data. If the bias leads to a systematic underestimation of the mean usage time, then the difference between the sample mean and the null hypothesis mean would be smaller, which would increase the p-value and reduce the power to detect a true effect. Conversely, if the bias leads to a systematic overestimation of the mean usage time, then the difference between the sample

mean and the null hypothesis mean would be larger, which would decrease the p-value and increase the power to detect a true effect.

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

Ans: This scenario would likely decrease the power to reject the null hypothesis because the biased data would make it harder to detect a true effect.

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

Ans: This scenario would increase the risk of a Type II error (failing to reject the null hypothesis when it is actually false). The bias in the data towards older users would make it more difficult to detect a true effect among younger users who are more active on weekends. As a result, the null hypothesis may not be rejected even if the new smartwatch does, in fact, have a higher mean usage time than the previous one.