

HW 4

113078506

2025-03-13

- Gen AI Usage: I use Gen AI to refine my grammar and verify my reasoning.

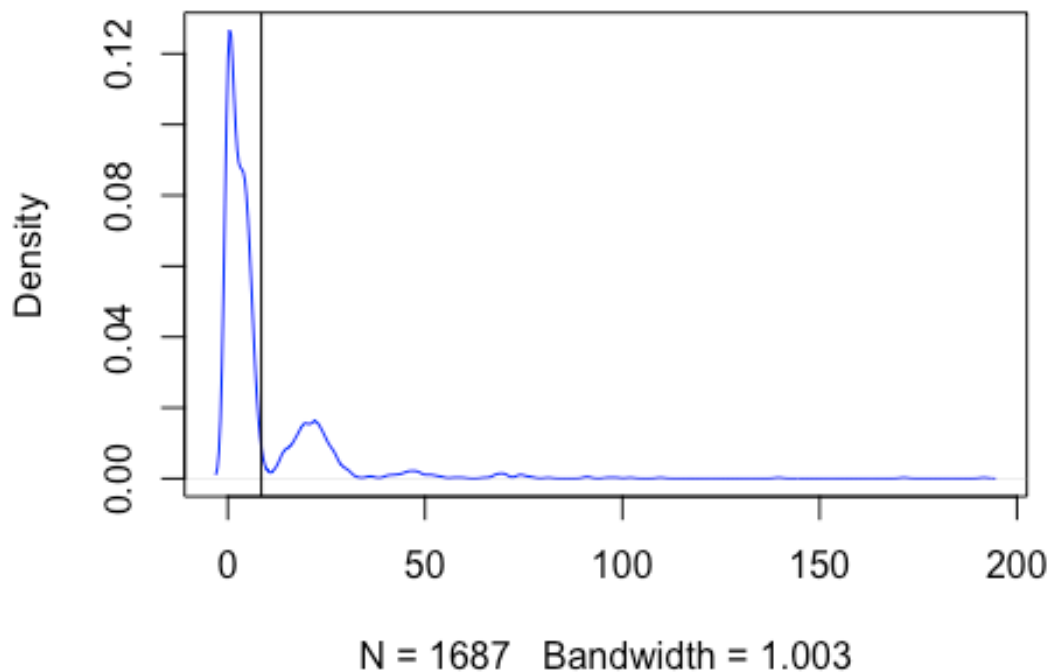
Question 1

(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

```
verizon <- read.csv("verizon.csv")  
plot(density(verizon$Time), main="Distribution of Repair Times", col = "blue")  
abline(v = mean(verizon$Time))
```

Distribution of Repair Times



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

- $H_0: \mu = 7.6$ (mins)
- $H_1: \mu \neq 7.6$ (mins)

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

```
sample_mean <- mean(verizon$Time)
sample_sd <- sd(verizon$Time)
n <- length(verizon$Time)
hyp_mean <- 7.6

# Standard Error
se <- (sample_sd / sqrt(n))
```

```

#99 CI sample_mean + or - t* se
ci_lower <- sample_mean - qt(1 - 0.01/2, df=n-1) * se
ci_upper <- sample_mean + qt(1 - 0.01/2, df=n-1) * se

cat("Sample Mean:", sample_mean, "\n")

## Sample Mean: 8.522009

cat("99% Confidence Interval:", ci_lower, "-", ci_upper, "\n")

## 99% Confidence Interval: 7.593524 - 9.450495

```

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

```

# T-statistic
t <- (sample_mean - hyp_mean) / se

# p-value: Probability of t (two-tailed test)
df <- n - 1
p <- 2 * (1 - pt(abs(t), df=n-1))

cat("T-statistic:", t, "\n")

## T-statistic: 2.560762

cat("p-value:", p, "\n")

## p-value: 0.01053068

```

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

- T-statistic (2.560762) represent the distance in standard error between sample mean and hypothesized mean. If $H_0 = \text{True}$, the t-distribution centered at 0, forming the Null distribution. The p-value indicates the

probability that the observed data is obtained in Null distribution,
therefore if $p\text{-value} < 0.01$ (99% CI), H_0 will be rejected.

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

- The t-statistic (2.560762) stays in the interval of 99% confidence [-2.58, 2.58], the p-value (0.01053068) is also bigger than the alpha (0.01), which is failed to reject H_0 . In conclusion, under 99 % confidence interval, we cannot reject the hypothesis "population mean (μ) is equal to 7.6 (mins)"

(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:

- $H_0: \mu \leq 7.6$
- $H_1: \mu > 7.6$

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

```
set.seed(1234)

sample_statistic <- function(stat_function, sample0) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  stat_function(resample)
}

sample_means <- replicate(2000, sample_statistic(mean, verizon$Time))

ci_sample_means <- quantile(sample_means, probs = c(0.01,1)) # on
e-tailed test only check lower bound

cat("99% Confidence Interval of Bootstrapped Means:", ci_sample_means, "\n")
```

```
## 99% Confidence Interval of Bootstrapped Means: 7.704504 9.870243
```

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

```
# Bootstrapping the 99% CI of the Difference of Means
boot_mean_diffs <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  return( mean(resample) - mean_hyp )
}

set.seed(1234)
mean_diffs <- replicate(
  2000,
  boot_mean_diffs(verizon$Time, hyp_mean)
)
diff_ci_99 <- quantile(mean_diffs, probs=c(0.01,1))
cat("Bootstrapped 99% CI for Difference:", diff_ci_99, "\n")

## Bootstrapped 99% CI for Difference: 0.1045037 2.270243
```

iii. Bootstrapped t-statistic: What is the 99% CI of the bootstrapped t-statistic of the sample mean versus the hypothesized mean?

```
boot_t_stat <- function(sample0, mean_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  diff <- mean(resample) - mean_hyp
  se <- sd(resample)/sqrt(length(resample))
  return( diff / se )
}

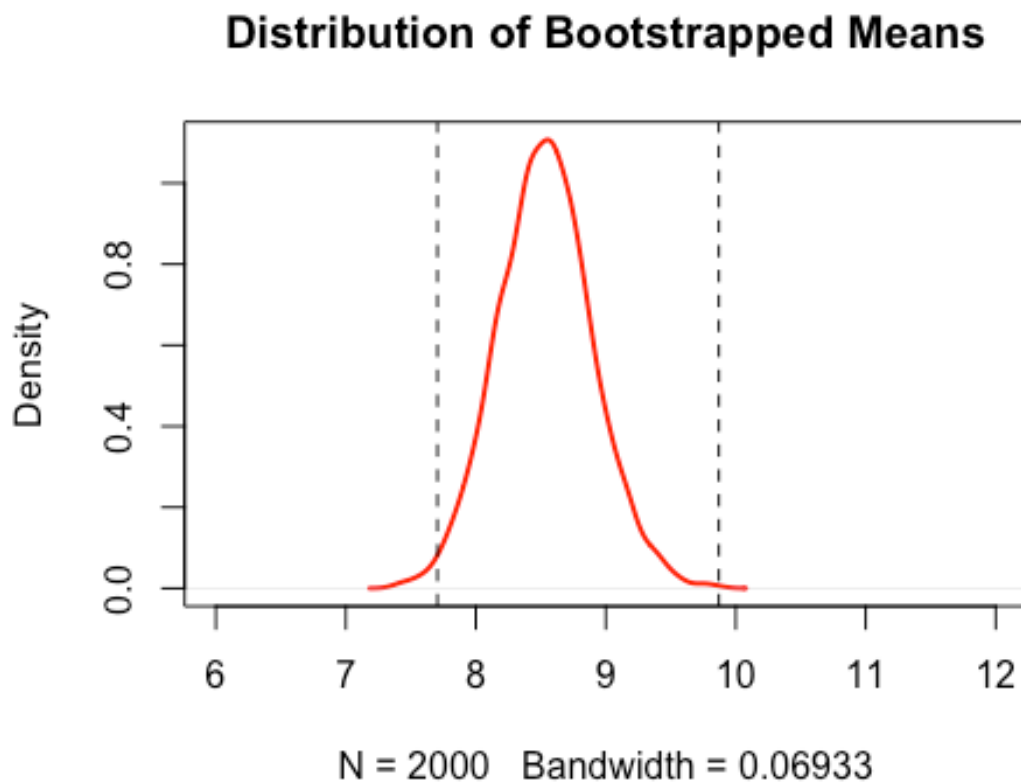
set.seed(1234)
t_boots <- replicate(2000, boot_t_stat(verizon$Time, hyp_mean))
```

```
t_ci_99 <- quantile(t_boots, probs=c(0.01,1))
cat("Bootstrapped 99% CI of t-statistic:", t_ci_99, "\n")

## Bootstrapped 99% CI of t-statistic: 0.3388869 5.400557
```

iv. Plot the distribution of the three bootstraps above on separate plots; draw vertical lines showing the lower/upper bounds of their respective 99% confidence intervals.

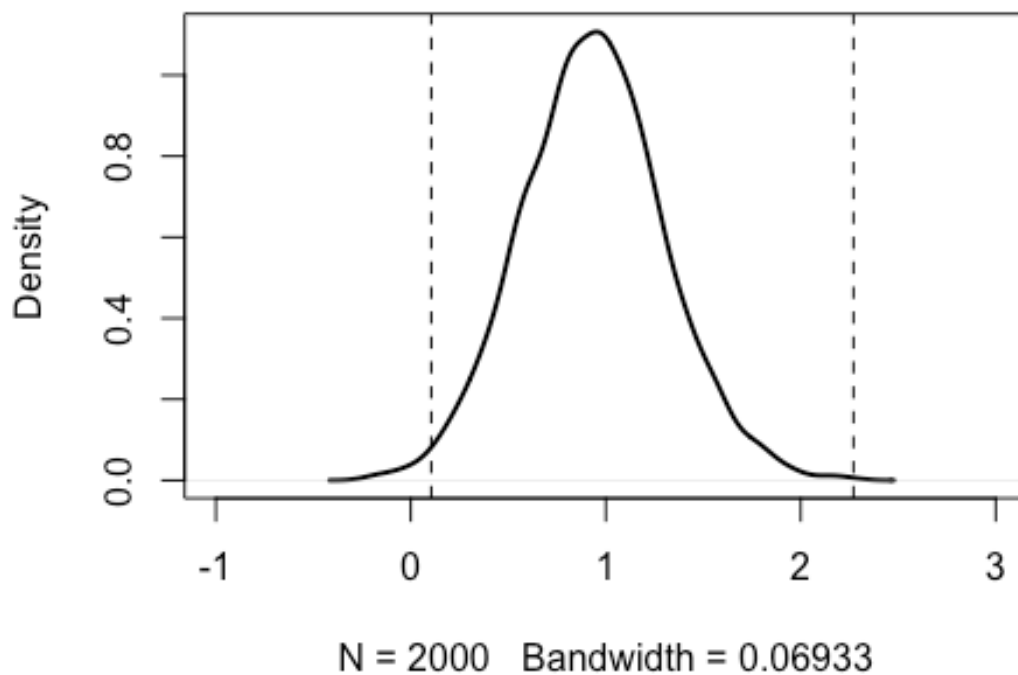
```
# Bootstrapped Means
plot(density(sample_means), xlim=c(6,12), col="red", lwd=2, main
= "Distribution of Bootstrapped Means")
abline(v=ci_sample_means, lty="dashed")
```



```
# Bootstrapped Difference of Means
plot(density(mean_diffs), xlim=c(-1,3), col="black", lwd=2, main
```

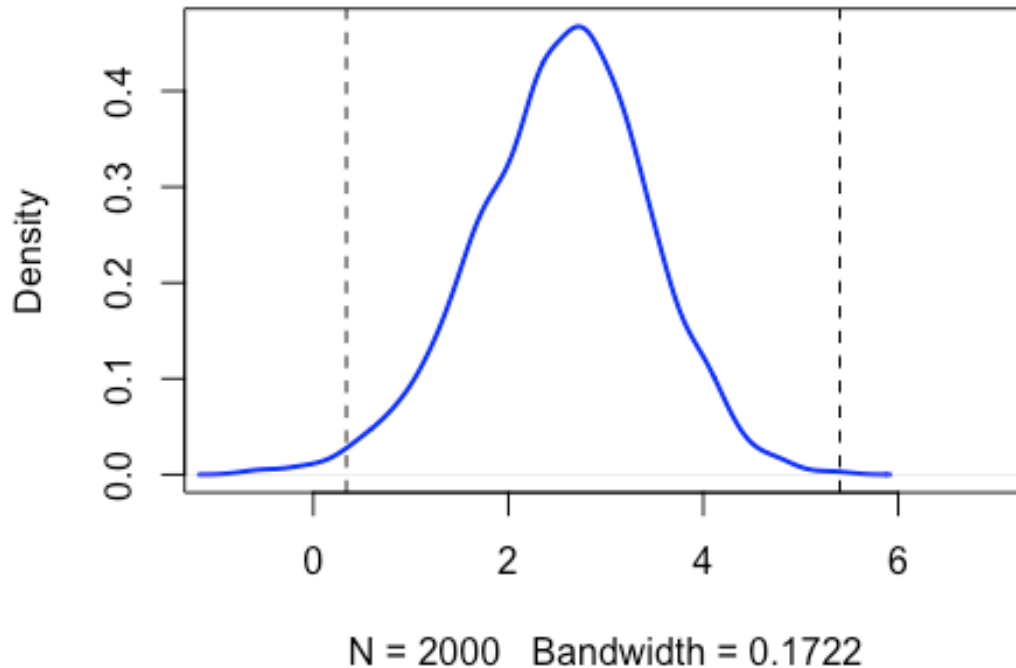
```
= "Distribution of Bootstrapped Difference of Means")  
abline(v=diff_ci_99, lty="dashed")
```

Distribution of Bootstrapped Difference of Means



```
# Bootstrapped T-statistic  
plot(density(t_boots), xlim=c(-1,7), col="blue", lwd=2, main = "D  
istribution of Bootstrapped T-statistic")  
abline(v=t_ci_99, lty="dashed")
```

Distribution of Bootstrapped T-statistic



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

- No, both the 99% CI of bootstrapped T-statistic and bootstrapped difference of means do not contain 0, showing that we have significant evidence to tell that the difference between population mean and hypothesized mean is not equal to 0. In addition, the 99% CI of bootstrapped mean also do not contain 7.6, meaning that 99% of sample means are greater than 7.6. This is a different result from the result in part [a], we can actually reject H_0 by doing the bootstrapped approach.

(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 argue that the median is a fairer test, and claim that the median repair time is no more than 3.5 minutes at 99% confidence (i.e., significance $\alpha = 1\%$).

- H_0 : population median ≤ 3.5
- H_1 : population median > 3.5

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

```
set.seed(1234)
sample_medians <- replicate(2000, sample_statistic(median, verizon$Time))
ci_sample_medians <- quantile(sample_medians, probs = c(0.01, 1))
# one-tailed test
cat("99% Confidence Interval of Bootstrapped Medians:", ci_sample_medians, "\n")

## 99% Confidence Interval of Bootstrapped Medians: 3.23 3.95
```

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

```
hyp_median = 3.5
# Bootstrapping the 99% CI of the Difference of Medians
boot_median_diffs <- function(sample0, median_hyp) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  return( median(resample) - median_hyp )
}

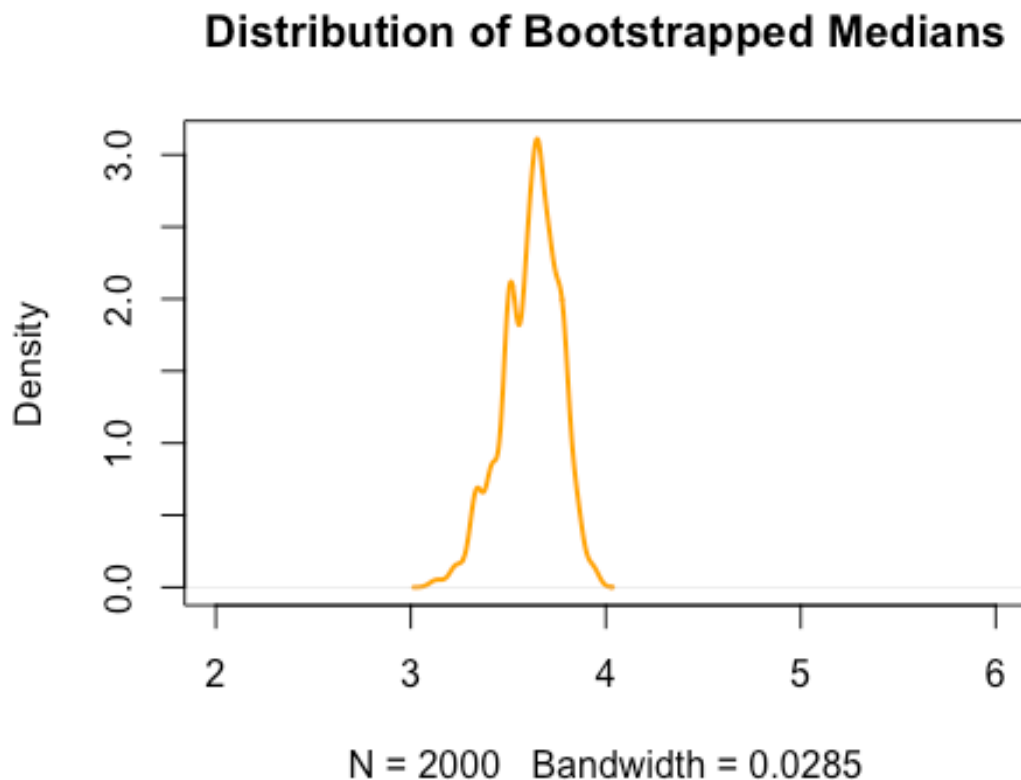
set.seed(1234)
median_diffs <- replicate(
  2000,
```

```
boot_median_diffs(verizon$Time, hyp_median)
)
diff_ci_99_me <- quantile(median_diffs, probs=c(0.01, 1))
cat("Bootstrapped 99% CI for Median Difference:", diff_ci_99_me,
"\n")

## Bootstrapped 99% CI for Median Difference: -0.27 0.45
```

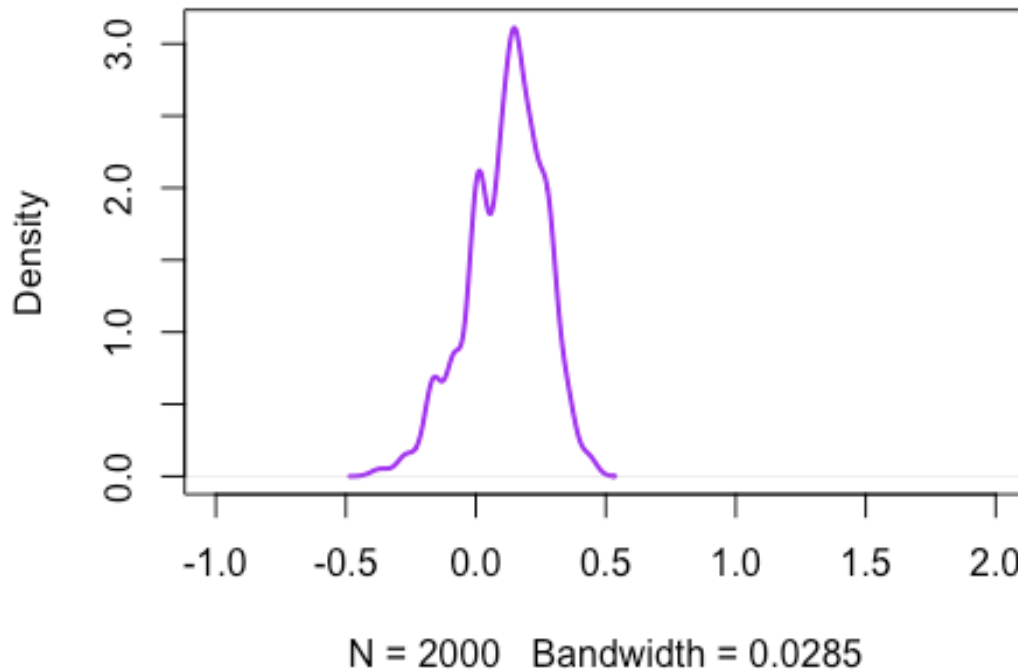
iii. Plot distribution the two bootstraps above on two separate plots (note that we are not using a test statistic for the median).

```
# Bootstrapped Medians
plot(density(sample_medians), xlim=c(2,6), col="orange", lwd=2, m
ain = "Distribution of Bootstrapped Medians")
```



```
# Bootstrapped Difference of Medians
plot(density(median_diffs), xlim=c(-1,2), col="purple", lwd=2, main = "Distribution of Bootstrapped Difference of Medians")
```

Distribution of Bootstrapped Difference of Median



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

- The 99% confidence interval for the difference of medians contains 0, meaning that the data does not provide strong evidence against the null hypothesis. Therefore, we fail to reject H_0 and cannot conclude that the population median is significantly different from 3.5.

Question 2

- H_0 : The mean usage time of the new smart watch is the \leq previous smart watch.
- H_1 : The mean usage time $>$ previous smart watch.

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

(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 scenario creates systematic error, because the collected data might not fully cover the true population of using the smart watch, the older customers are left out. The data are all from the pool of young customers also indicate that this result is favoring a certain group.

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

- The difference (diff) and standard deviation (sd) will be affected because younger customers may use the smartwatch more frequently than older customers, influencing the sample mean. If older customers are included in the sample, the mean usage time may decrease, altering the difference (diff). Additionally, since older customers represent a different user segment with potentially more diverse usage patterns, their inclusion would increase variability in the data, leading to a higher standard deviation (sd).

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

- Younger customers tend to use the smartwatch more frequently, which may result in a larger difference (diff) between the sample mean and the hypothesized mean. This increases the likelihood of rejecting the null hypothesis.

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

- Type I error is more likely in this scenario because the sample may overestimate the true usage duration, leading to a higher observed mean than the actual population mean. As a result, we may incorrectly reject the null hypothesis when it is actually true.

(b) You find that 20 of the respondents are reporting data from the wrong wearable device, and should not have been in the sample. 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)?

- This creates random error, because it provides wrong information in the experiment, which makes random fluctuation and increases noise to the data.

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

- The sample mean will be miscalculated by wrong reports, leading to an affected difference (diff). The standard deviation (sd) will be overestimated due to increased variation.

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

- Based on the t-statistic formula, an increase in standard deviation (sd) leads to a smaller t-statistic, which reduces the power to reject the null hypothesis.

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

- Type II errors become more likely because our power to reject H_0 is lower than reality, so the error of false negative will occur more easily.

(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)?

- Neither. Systematic errors occurred when the sample data are skewed from the population, changing the confident interval does not change the data. Random error comes from the variation from the original sample data, which is not able to change by only adjusting the confident interval. The adjustment like this only increases the risk of making wrong decision because the intervals are less strict.

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

- The alpha is increase from 0.05 to 0.1

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

- Our power is increased because now it's easier to reject H_0 with narrower confident intervals.

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

- Type I error is more likely to occurred, because we are rejecting H_0 more easily, which might cause us miss rejecting null hypothesis.

(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)?

- This will occur systematic error. The segmentation of time might inflate the behavior from certain customer group, e.g.: young people are more active on weekends.

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

- Similar to (a), if we exclude certain time periods, such as weekends, the sample mean may be underestimated due to missing data in weekend from younger customers, who tend to have higher usage. This affects the calculated difference (diff). Additionally, the standard deviation (sd) will also be impacted since excluding specific time periods alters the variability, making it different from the true population variation.

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

- The power decreases because excluding weekends likely underestimates the sample mean (reducing effect size) and distorts variability, making it harder to detect a true effect.

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

- Type II error becomes more likely because younger customers tend to have higher usage on the weekend, which might lead us to underestimate the sample mean. Correspondingly, we might more likely to accept H_0 when we should reject it, lead to false negative error (Type II Error).