```
---
title: "HW"
author: '110078509'
date: '20220319'
output:
  html_document: default
  pdf_document: default
  word_document: default
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(ggplot2)
rm(list=ls())
```

Question 1

---
i. Would this scenario create systematic or random error (or both or neither)?
ii. Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?
iii. Will it increase or decrease our power to reject the null hypothesis?
iv. Which kind of error (Type I or Type II) becomes more likely because of this scenario?
---

a. 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.*systematic error*

---

ii.*sd, diff, n are affected.*

Explains:

Because we missed many older customers.
Therefore, the alternative distribution we obtained is too narrow compared to the real one. Hence, the sd is affected.

And the sample size we got is fewer than it shall be if we want to maintain a randomized selected sample assumption. Hence, n is affected.

The diff is (x mean -μo), in this scenario, the x means is not accurate anymore, because we missed plenty of samples lower than the average. Hence the diff is affected also.

---

iii. It decreases our power to reject the null hypothesis.

Explains:

Manipulate the sliding window to figure out.
---

iv. TypeII error

Explains:

It also called false negative.
The data of elder people should be considered in, however, they are not.Therefore, it Type II error.
```

---

b.Find that 20 of the respondents are reporting data from the wrong wearable device, so they should be removed from the data.

---

i. *random error*

---

ii. *n would be lower. *

Explains:

Because these noisy data shall not be considered in.

---

iii. *It decrease the power to reject the null hypothesis*

Reason:

Manipulate the sliding window to figure out.

---

iv. *Type I Error*

Explains:

It also called false positive. In this scenario, we considered the error into our analysis, however, we shall not. Therefore, it false positive.

---

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. *neither.*

Explains:

It just the change of the confidence level based on the professor's suggestion.

---

ii. *alpha*

Explains:

From 0.05 to 0.1

---

iii. *Increase the power to reject the null hypothesis.*

Explains:

Because the critical point of the right tail shifted leftward.

---

iv. *Type I*

Explains:
Type 1 errors has a probability of "α" correlated to the level of confidence we set. Originally, 95% confidence level means that there is a 5% chance of getting a type I error. According to the suggestion, we set the confidence level to 90% , the

chance of getting a type 1 error increase (10%).

---

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. *systematic error*

---

ii.*diff and mean will be affected.*

Explains:

Because the population means we want to inference includes the user behavior from Monday to Sunday. If we only choose the workday data as our sample, we will overemphasize the behavior of elder users and neglect the younger people who are very active on weekends. Therefore, the mean sample mean is underestimated. According to the formula of diff(mean - mu 0), as the mean is affected and the mu0 is still, the diff would be affected for sure.

---

iii. *it decrease our power to reject the null hypothesis.*

Explains:

H0 :the mean usage time of the new smartwatch is the same or less than for the previous smartwatch.

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

And this error results in underestimating frequent users among the young. Therefore, it makes us harder to reject H0, which decreases the power to reject the null hypothesis.

---

iv. *Type II*

Explains:
Because we should count weekend users in, but we don't. And false negative means that we should consider specific targets is negative, but it doesn't.

Explains:

#####

```{r 2}
verizon <- read.csv("verizon.csv")
time <- verizon$Time
```

Ans: