Technical Report

Group 5

November 18, 2022

Introduction

In 2020, Universities were scrambling to deal with covid 19 and students were locked indoors, a completely new environment full of tension. To adapt, a lot of students were changing throughout Covid 19 and this study's objective is to determine whether smartphone usage and mental health was part of the changing factors. Specifically, whether the smartphone usage of an average STA304H5 Fall 2022 student before covid changed in comparison to during covid. Similarly, whether the mental health score of an average STA304H5 Fall 2022 student before covid changed in comparison to during covid. Our survey also recorded whether academic success, financial status were impacted by Covid 19 for STA304H5 Fall 2022 students. Readers can utilize or refer to this study as evidence for whether there was a change in smartphone usage, and mental health score for STA304H5 Fall 2022 students when comparing before covid and during covid.

Objectives

i. Determine whether STA304H5 students' smartphone usage was impacted by Covid 19.
ii. Determine whether STA304H5 students' mental health was impacted by Covid 19.

Variables

Variable	Definition
Gender	Male, Female, or Non-Binary
Smartphone Usage before Covid	Measured in hours spent per day (0 - 8)
Smartphone Usage during Covid	Measured in hours spent per day (0 - 8)
Stress / Anxiety Level before Covid (Inverted)	Indicates anxiety level from a range of 1 - 7, where 7 is very low stress/anxiety and 1 is very high
Stress / Anxiety Level during Covid (Inverted)	Indicates anxiety level from a range of 1 - 7, where 7 is very low stress/anxiety and 1 is very high
Quality of Sleep before Covid	Indicates quality of sleep from a range of 1-7, where 7 is very good quality of sleep and 1 is very low
Quality of Sleep during Covid	Indicates quality of sleep from a range of 1-7, where 7 is very good quality of sleep and 1 is very low
Social Interactions Rating before Covid	Indicates quality of social interaction from a range of 1-7, where 7 is very good and 1 is very low
Social Interactions Rating during Covid	Indicates quality of social interaction from a range of 1-7, where 7 is very good and 1 is very low
Mental Health Score before Covid	Calculated as : Stress / Anxiety Level (inverted) + Social Interaction Level + Sleep Quality Score (all before Covid). Also, the mental health score is calculated per Student.
Mental Health Score during Covid	Calculated as : Stress / Anxiety Level (inverted) + Social Interaction Level + Sleep Quality Score (all during Covid). Also, the mental health score is calculated per Student.

Sample size

For our study, we are looking at STA304 population size N = 200 . We are using the variance from the responses to our smartphone usage before Covid 19 questions for this computation and we are looking for a bound of error of estimation where $B = 0.35$. We are also sampling SRS without replacement.

$$n = \frac{N\sigma^2}{(N-1)D+\sigma^2} = \frac{200(3)}{(200-1)(0.35^2/4)+(3)} = \frac{600}{(199)(0.0306)+(3)} \simeq 66$$

Therefore, we should sample 66 students and we have done so successfully by distributing it through lectures/piazza/office hours.

Point Estimators

$$\overline{y}_{smartphone\ usage\ \ before\ covid} = \frac{\sum_{i=1}^{66} (smartphone\ usage\ during\ covid)}{n = 66} = \frac{272}{66} = 4.121$$

$$\overline{y}_{smartphone\ usage\ during\ covid} = \frac{\sum_{i=1}^{66} (smartphone\ usage\ during\ covid)}{n = 66} = \frac{401}{66} = 6.076$$

$$\overline{y}_{mental\ health\ before\ covid} = \frac{\sum_{i=1}^{66} (mental\ health\ score\ before\ covid)}{n = 66} = \frac{914}{66} = 13.848$$

$$\overline{y}_{mental\ health\ during\ covid} = \frac{\sum_{i=1}^{66} (mental\ health\ score\ during\ covid)}{n = 66} = \frac{791}{66} = 11.985$$

Hypothesis

To achieve the first objective (O1) i.e., determining whether there was any *real* change between average STA304 students smartphone usage before covid vs during covid. We have the following hypothesis:

Let μ_1 be the average smartphone usage in hours **before** covid 19

Let μ_2 be the average smartphone usage in hours **during** covid 19

Then, we have

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

To elaborate, the above statements means that we are testing, for example, whether the true mean of smartphone usage before covid 19 differs from true mean of smartphone usage during covid 19.

To achieve the second objective (O2) i.e., determining whether there was any *real* change between average STA304 students' mental health score before covid vs during covid. We consider the following hypothesis:

Let μ_1 be the average mental health score **before** covid 19

Let μ_2 be the average mental health score **during** covid 19

Then, we have

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

To elaborate, the above statements means that we are testing, for example, whether the true mean of the mental health score before covid 19 differs from the true mean of the mental health score during covid 19.

From our point estimators, we can generally expect there to be a difference since

$$|\overline{y}_{smartphone\ usage\ \ before\ covid} - \overline{y}_{smartphone\ usage\ during\ covid}| = 1.955\ (hrs)$$

$$|\overline{y}_{mental\ health\ before\ covid} - \overline{y}_{mental\ health\ during\ covid}| = 1.863$$

So from our point estimators we are getting the indication that we should be rejecting our null hypothesis, let's test it.

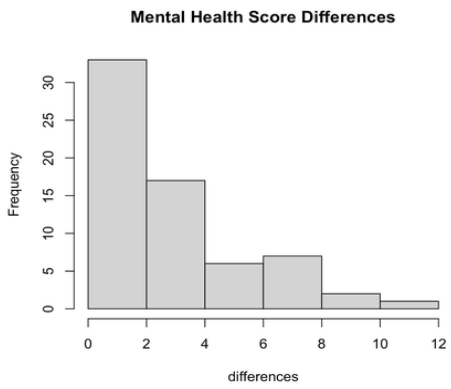
Assumptions

We are not able to use a two-sample t-test here since one of the assumptions states that the samples must be independent. However, the difference we are calculating here is comparing two variables i.e., smartphone usage before covid, smartphone usage during covid, that come from the same subject.

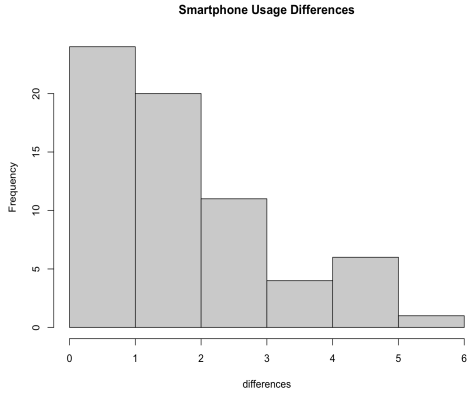
We are also unable to use paired t-tests for our hypothesis because both differences i.e., Smartphone Usage before Covid - Smartphone Usage during Covid and Mental health score before Covid - Mental health score during Covid, aren't normal. We test this using the Sharpiro-Wilk normality test. For the difference between smartphone usage, we obtain the test-statistic W = 0.89872 which gives a p-value of 5.601e-05 (< 0.05) so we cannot assume normality. Similarly, for mental health score differences, we obtain the test-statistic W = 0.88544 which gives us a p-value of 1.814e-05 (< 0.05) so again we cannot assume normality for the differences.

We also are not able to use Wilcoxon signed-rank test because as you can see in figure 1.0 and 1.1, for both sample differences we see an extremely skewed distribution,

hence, we expect the population differences to also be asymmetric. So we do not meet the population differences symmetry requirement for the Wilcoxon signed-rank test.



[Figure 1.0]



[Figure 1.1]

Therefore, we are left with the paired-sign test where asymmetry for the population differences isn't an issue and it's also a non-parametric test, hence we don't require our distributions to be normal. The assumptions for the paired-sign test are as follows:

- **Independent Subjects:** The responses we collected on the survey are independent of other responses, hence the pairs and differences are independents.
- **Continuity:** All samples are collected from the same continuous population.
- **Dependent Samples:** This is met since we will be using the Mental Health score and Smartphone usage and for both variables, comparing the before value to during Covid from the same subject.
- **Ordered:** All 4 of smartphone usage before covid, smartphone usage during covid, mental health score before covid and mental health score during covid were likert scale questions, hence they are ordinal and ordered already.
- **Random Sample**: Since we used SRS for the analysis, it holds that the sample collected constitutes a random sample

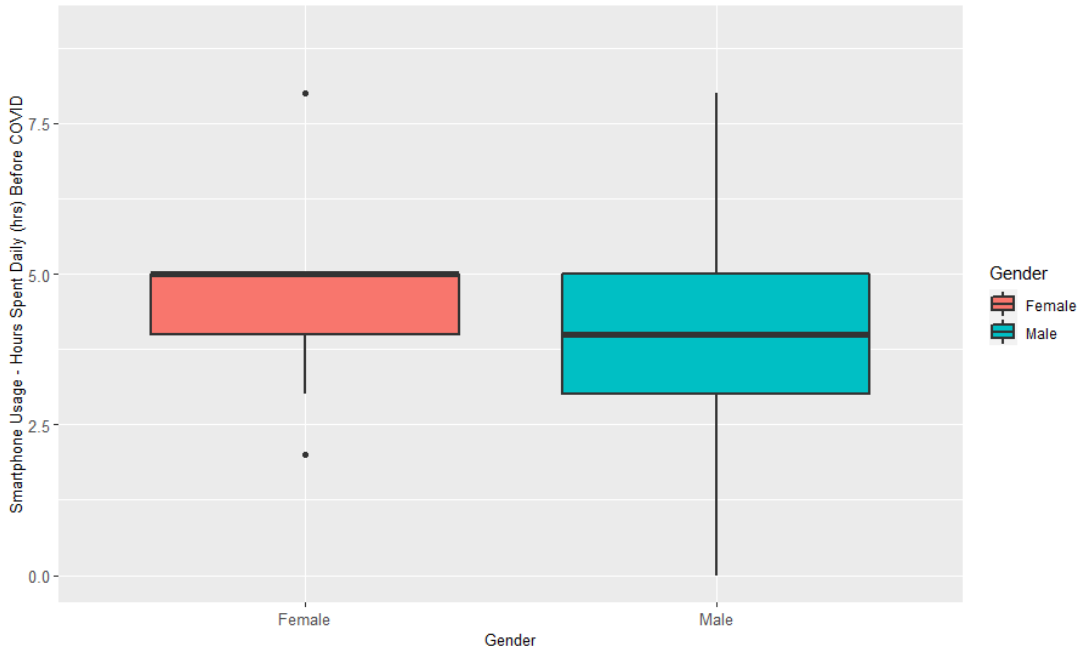
Paired-Sign Test Results

For the hypothesis for O1, where we are looking to check whether the true mean of smartphone usage before covid is different from the true mean smartphone usage during covid. We obtain the test-statistic S = 1 which gives a p-value of 3.482e-13. Therefore, we reject the null hypothesis at $\alpha = 0.05$.

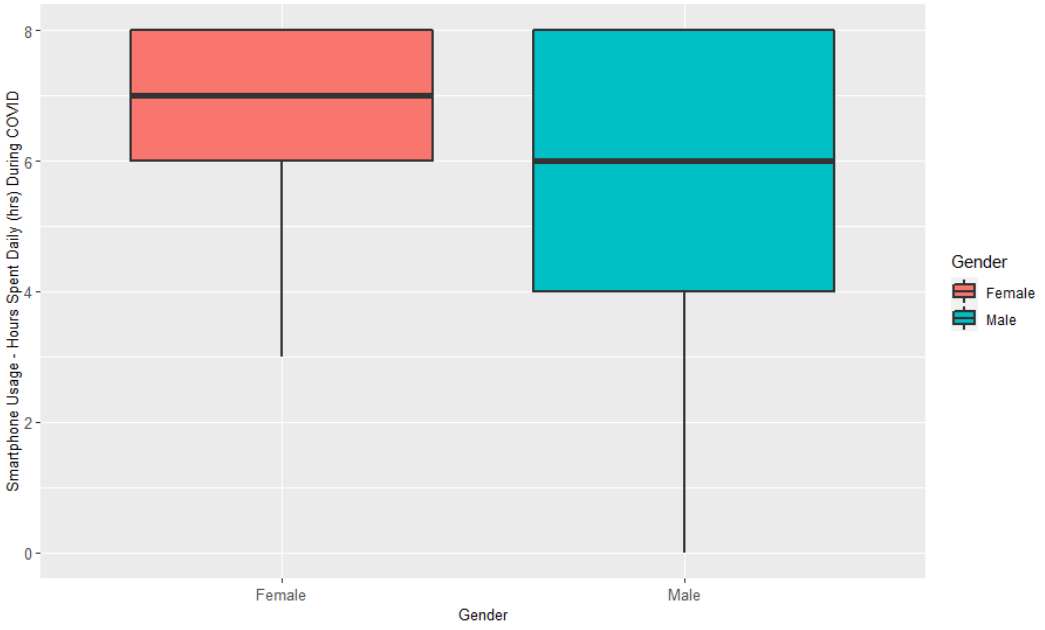
For our second hypothesis, relating to O2, where we check whether the true mean of mental health score before covid is different from the true mean of mental health score during covid. We obtain the test-statistic S = 42 which gives a p-value of 0.0008618. Therefore, we reject the null hypothesis at $\alpha = 0.05$.

Graphs

Let's explore our hypothesis results visually.



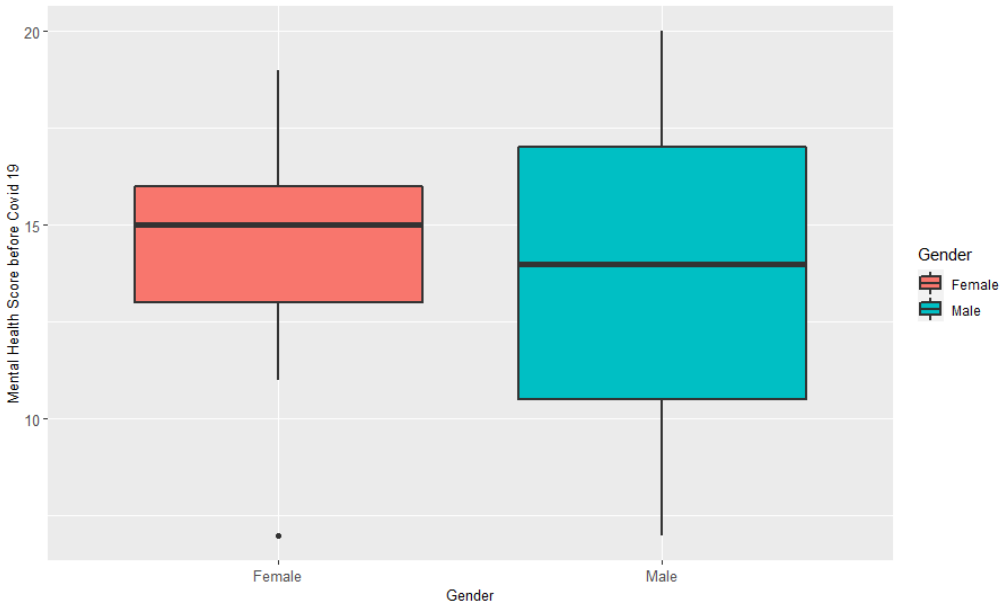
[Figure 1.2]



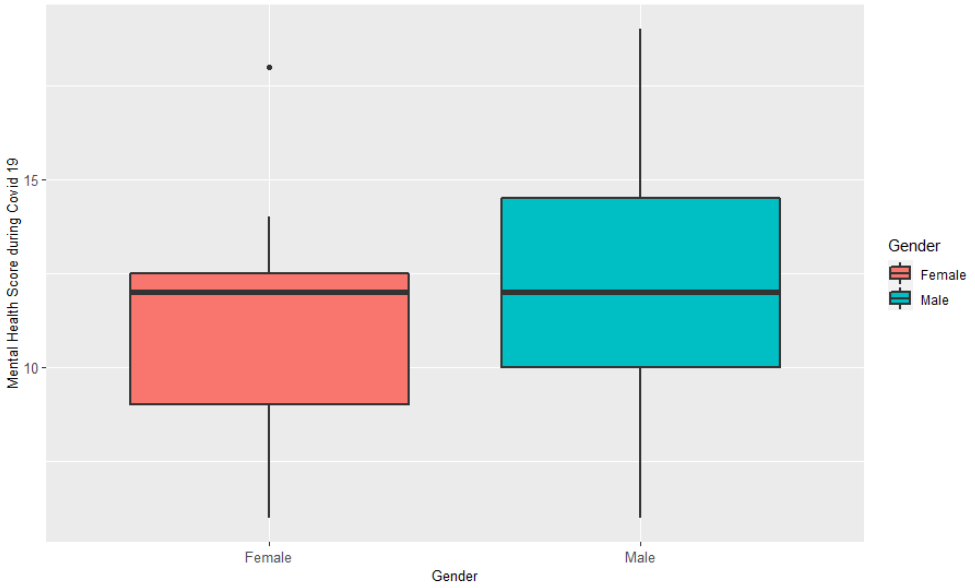
[Figure 1.3]

In Figure 1.2, we have a boxplot for smartphone usage before covid, and in Figure 1.3 we have a boxplot for smartphone usage during covid. We can clearly see the average/median smartphone usage changes from "before covid" to "during covid". For example, for females, the median line for smartphone usage before covid points at 5 hours, but during covid, it rises to 7. This shows quite a bit of change in pattern. Similarly for males (its proportional increase). Hence, this supports our hypothesis result.

Now on to investigating mental health score changes,



[Figure 1.4]



[Figure 1.5]

In Figure 1.4, we have a boxplot for mental health score before covid, and in Figure 1.5 we have a boxplot for mental health score during covid. We can clearly see the average/median mental health score changes from "before covid" to "during covid". We can see that, generally for both genders, the mental health score before covid 19 is

higher than during covid 19. Supporting our findings in our hypothesis testing that the true mean of mental health score before covid 19 is different from the true mean of mental health score during covid 19.

Advanced Methodologies

We calculated Cronbach's alpha for all entries (n=66) to measure the internal consistency of the questionnaires, which measures how closely correlated the questions are in a group. Specifically, we calculated the Cronbach's alpha test for questions relating to the mental health score (both before Covid and during Covid). These questions were asking for subjects' quality of sleep, social interaction rating and the inverted stress/anxiety level.

Cronbach's alpha for mental health score before COVID-19 was found to be $\alpha = 0.52$, which implies poor internal consistency demonstrating that our mental health parameters for before COVID-19 were poorly correlated and inconsistent.

During COVID-19, we found $\alpha = 0.2$ meaning the elements in this group were extremely poorly correlated as well.

Along with interrelatedness there exists a variety of reasons for the low cronbach alpha values. One of which is the number of questions, Cronbach's alpha is directly proportional to the number of questions examined and since only three questions were analyzed; the observed alpha may have been impacted and resulted in a smaller value. Along with this heterogenous constructs among the data may also cause alpha to be smaller.

Conclusion

The objective of this observational study was to examine whether Covid 19 had an impact on STA304H5 students' smartphone usage and their mental health. In particular,

we reject the null hypothesis for both smartphone usage means difference and mental health score mean difference with $\alpha = 0.5$. To elaborate, we can say that STA304H5 students' smartphone usage and mental health changed from before Covid to during Covid.

However, there were several limitations to this observational study that may have impacted the overall results. We mostly sampled students in lectures and office hours, so we missed students that never attended in-person, indicating that our study might be a victim of convenience sampling. Moreover, we noticed our Cronbach alpha for mental health questions was quite low, so our calculation of mental health score cannot be considered reliable. If we had provided more of an incentive to answer the questionnaire, we may have gotten more responses leading to an even better study. Better yet, we could have improved the length of our survey with more mental health related questions so that our Cronbach alpha isn't extremely low.

Appendix

Libraries

```
library(ggplot2)
library(ltm) # for cronbach.alpha
library(BSDA)
library(psych)
```

Data Processing

```
# load the data
dataset <- read.csv('c://FALL22/STA304/project/dataset.csv')
```

```
# High Stress/Anxiety => Bad
# High Q of sleep | High social interactions => Good

# Reverse Stress/Anxiety scores
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7. <-
dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.from.1.7.
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7. <-
dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.from.1.7.

dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 7] <- 1
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 6] <- 2
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 5] <- 3
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 4] <- 4
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
```

```
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 3] <- 5
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 2] <- 6
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fro
m.1.7. == 1] <- 7

dataset$Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.from.1.7.
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
om.1.7.

dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 7] <- 1
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 6] <- 2
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 5] <- 3
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 4] <- 4
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 3] <- 5
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 2] <- 6
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.[dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fro
m.1.7. == 1] <- 7

dataset$Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.from.1.7.
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7.
```

```
# Create mental health score column
dataset$mental.health.score.before.COVID.19 <-
dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.fr
```

```
om.1.7. + dataset$Quality.Of.Sleep.Before.COVID.19..scale.from.1.7. +
dataset$Rating.Social.Interactions.Before.COVID.19..scale.from.1.7.

dataset$mental.health.score.during.COVID.19 <-
dataset$Inversed.Stress.and.Anxiety.Level.Overall.During.COVID.19..scale.fr
om.1.7. + dataset$Quality.Of.Sleep.During.COVID.19..scale.from.1.7. +
dataset$Rating.Social.Interactions.During.COVID.19..scale.from.1.7.
```

Assumptions

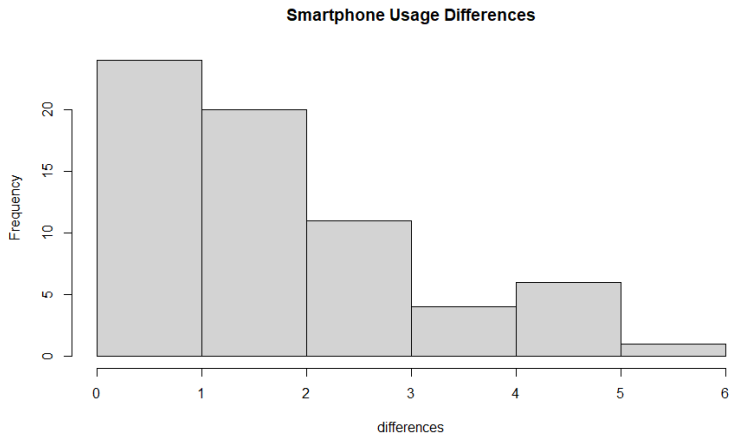
```
# Normality test on smartphone diff
smartphone_differences =
abs(dataset$Hours.Spent.Daily.On.Smartphone.Before.COVID.19..0.8.hours. -
dataset$Hours.Spent.Daily.On.Smartphone.During.COVID.19..0.8.hours.)
shapiro.test(smartphone_differences)
```

```
        Shapiro-Wilk normality test

data:  smartphone_differences
W = 0.89872, p-value = 5.601e-05
```

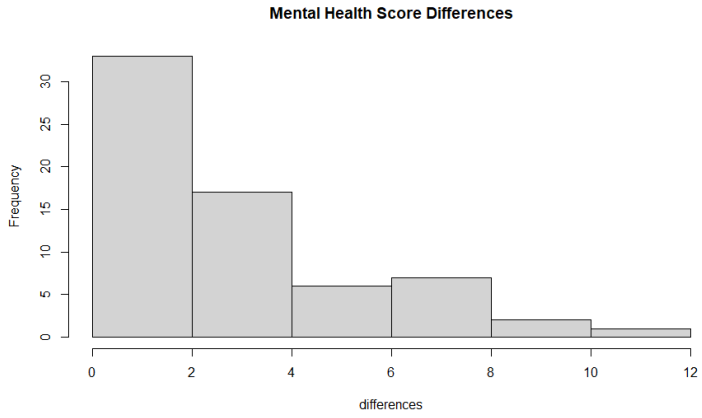
```
# Normality test on  mental health diff
mental_differences = abs(dataset$mental.health.score.before.COVID.19 -
dataset$mental.health.score.during.COVID.19)
shapiro.test(mental_differences)
```

```
        Shapiro-Wilk normality test

data:  mental_differences
W = 0.88544, p-value = 1.814e-05
```

```
# Graphs for differences distribution
smartphone_before=dataset$Hours.Spent.Daily.On.Smartphone.Before.COVID.19..
0.8.hours.
smartphone_during=dataset$Hours.Spent.Daily.On.Smartphone.During.COVID.19..
0.8.hours.
differences = abs(smartphone_before-smartphone_during)
hist(differences, breaks="Scott", main="Smartphone Usage Differences")
```




```
differences = abs(dataset$mental.health.score.before.COVID.19 -
dataset$mental.health.score.during.COVID.19)
hist(differences, breaks="Scott", main="Mental Health Score Differences")
```



Paired-Sign Test for Hypothesis

```
# Paired sign test for pairs diff means (two-sided, !=)
SIGN.test(x =
dataset$Hours.Spent.Daily.On.Smartphone.Before.COVID.19..0.8.hours.,
          y =
dataset$Hours.Spent.Daily.On.Smartphone.During.COVID.19..0.8.hours.,
          alternative = "two.sided",
          conf.level = 0.95)

SIGN.test(x = dataset$mental.health.score.before.COVID.19,
          y = dataset$mental.health.score.during.COVID.19,
```

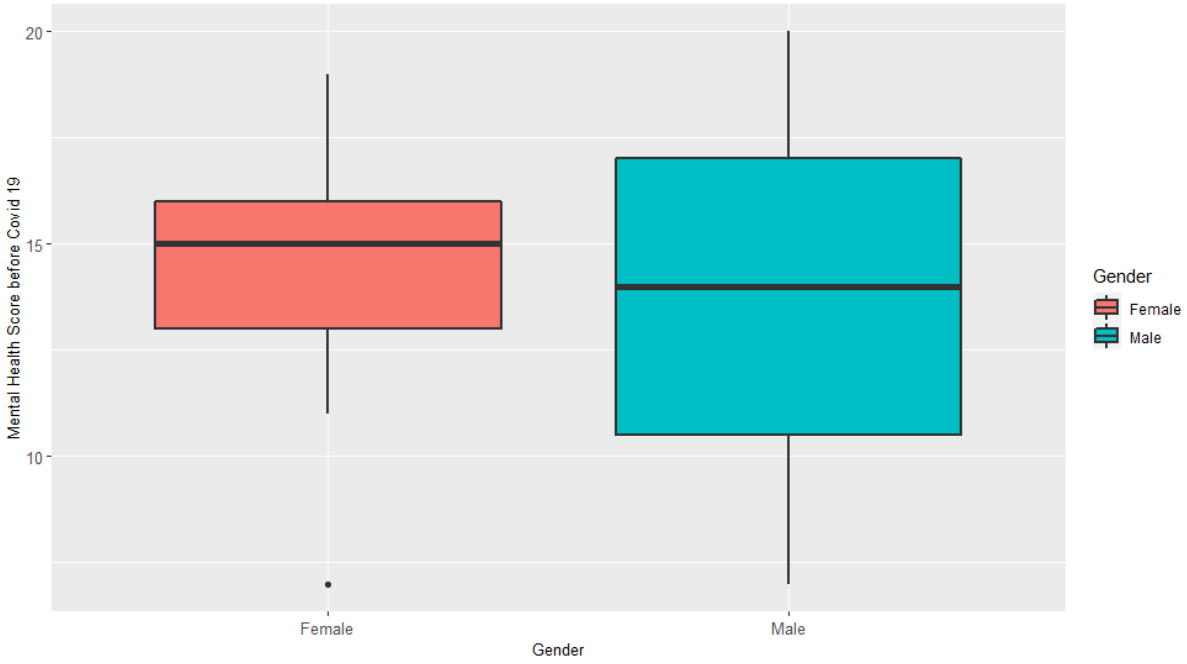
```
          alternative = "two.sided",
          conf.level = 0.95)
```

```
        Dependent-samples Sign-Test

data:  dataset$Hours.Spent.Daily.On.Smartphone.Before.COVID.19..0.8.hours. and
 dataset$Hours.Spent.Daily.On.Smartphone.During.COVID.19..0.8.hours.
S = 1, p-value = 3.482e-13
```

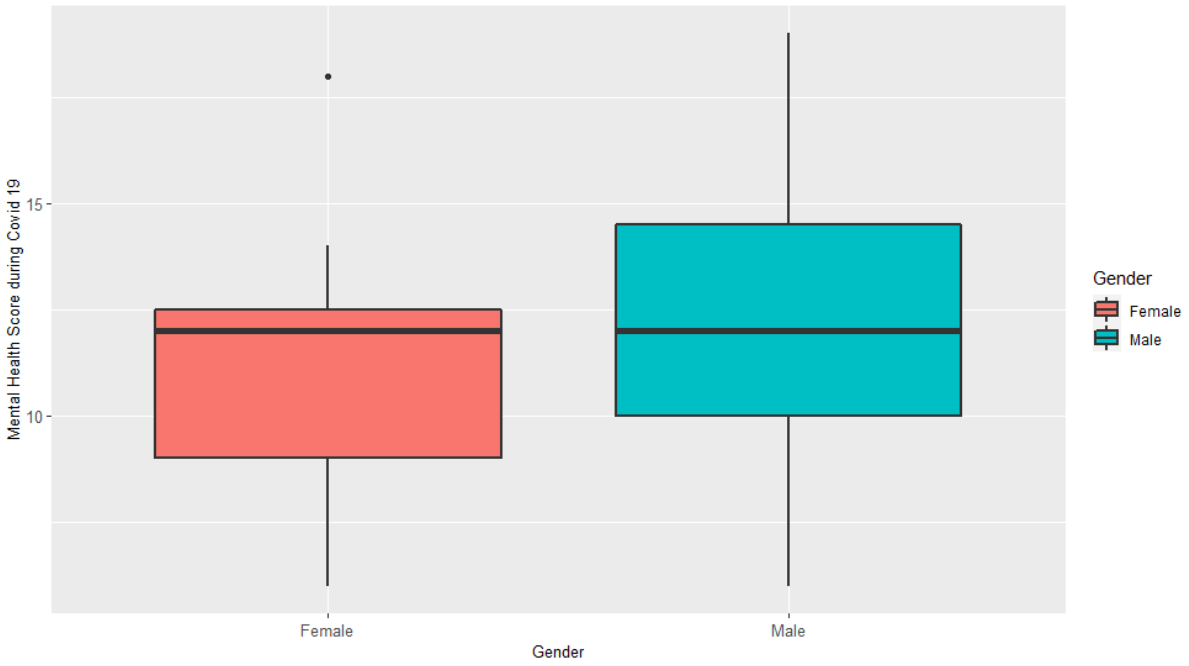
```
        Dependent-samples Sign-Test

data:  dataset$mental.health.score.before.COVID.19 and dataset$mental.health.sc
ore.during.COVID.19
S = 42, p-value = 0.0008618
```

Graphs

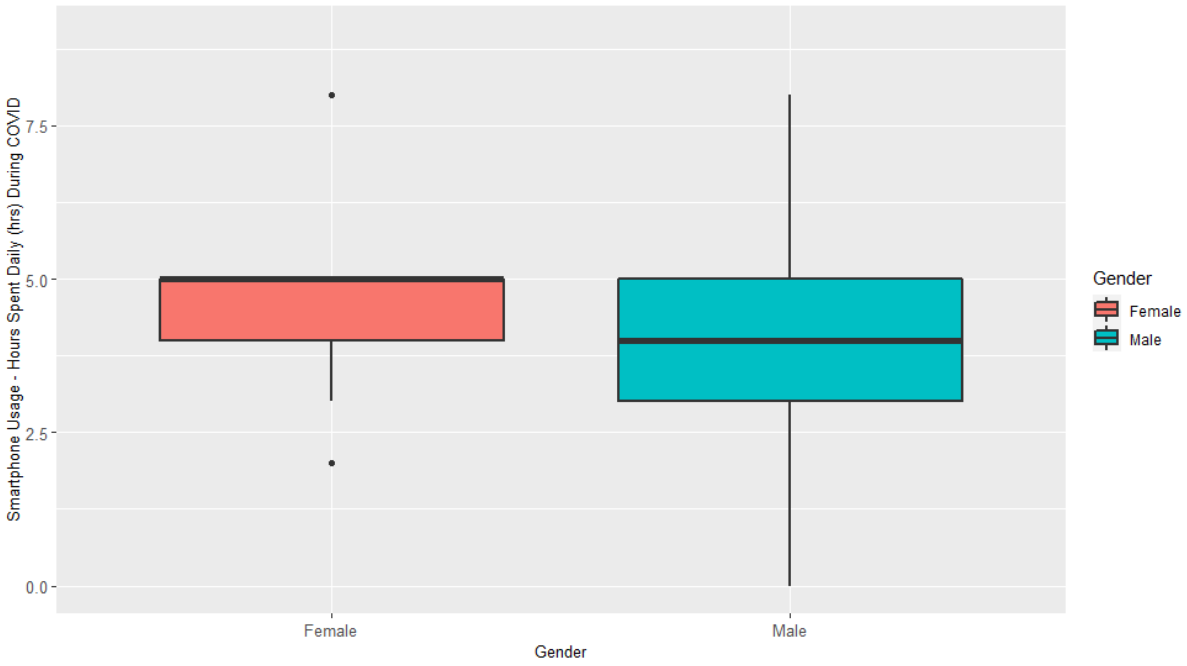
```
g_box_mental_bfr<- ggplot(dataset, aes(x=Gender,
y=mental.health.score.before.COVID.19, fill=Gender)) +
  geom_boxplot(lwd=1) +
  scale_fill_discrete("Gender", labels=c('Female', 'Male')) +
  labs(color = "Gender") +
  ylab("Mental Health Score before Covid 19") +
  theme(axis.title = element_text(size = 10.1))  +
  xlab("Gender") +
  theme(axis.text.x = element_text(size = 9.5)) +
  theme(axis.text.y = element_text(size = 9.5))

g_box_mental_bfr
```



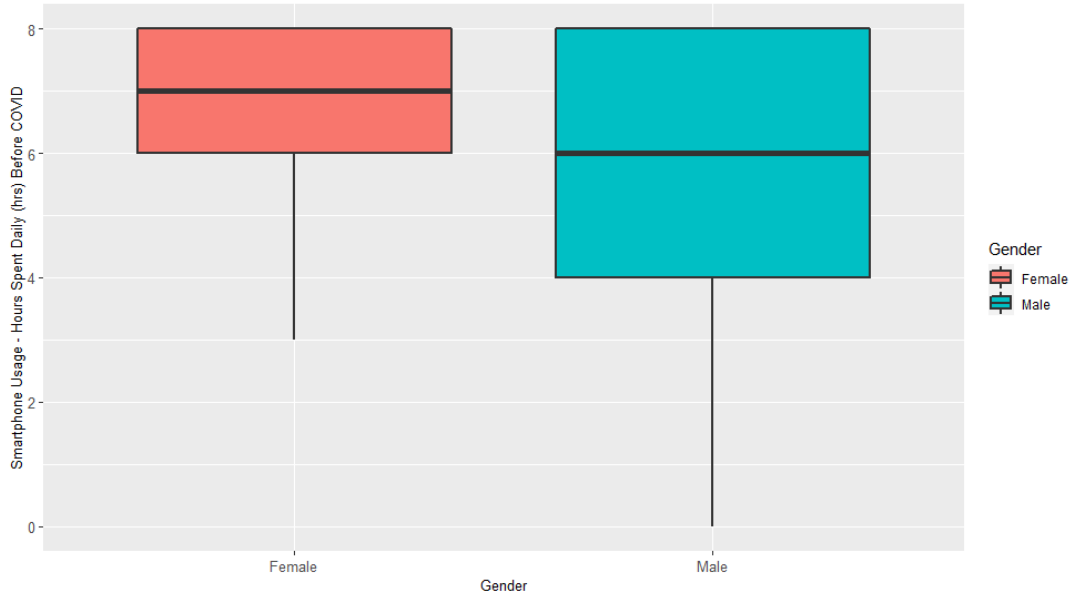

```
g_box_mental_dur <- ggplot(dataset, aes(x=Gender,
y=mental.health.score.during.COVID.19, fill=Gender)) +
  geom_boxplot(lwd=1) +
  scale_fill_discrete("Gender", labels=c('Female', 'Male')) +
  labs(color = "Gender") +
  ylab("Mental Health Score during Covid 19") +
  theme(axis.title = element_text(size = 10.1))  +
  xlab("Gender") +
  theme(axis.text.x = element_text(size = 9.5)) +
  theme(axis.text.y = element_text(size = 9.5))
  geom_boxplot()

g_box_mental_dur
```




```
g_box_smart_bfr <- ggplot(dataset, aes(x= Gender, y =
Hours.Spent.Daily.On.Smartphone.Before.COVID.19..0.8.hours., fill=Gender))
+ geom_boxplot(lwd=1) + ylim(c(0,9)) + scale_fill_discrete("Gender",
labels=c('Female', 'Male')) + labs(color = "Gender") + ylab("Smartphone
Usage - Hours Spent Daily (hrs) Before COVID") + theme(axis.title =
element_text(size = 10.1))  + xlab("Gender") + theme(axis.text.x =
element_text(size = 9.5)) + theme(axis.text.y = element_text(size = 9.5))

g_box_smart_bfr
```




```
g_box_smart_dur <- ggplot(dataset, aes(x= Gender, y =
Hours.Spent.Daily.On.Smartphone.During.COVID.19..0.8.hours., fill =
Gender)) + geom_boxplot(lwd=1) + scale_fill_discrete("Gender",
labels=c('Female', 'Male')) + labs(color = "Gender") + ylab("Smartphone
Usage - Hours Spent Daily (hrs) During COVID") + theme(axis.title =
element_text(size = 10.1))  + xlab("Gender") + theme(axis.text.x =
element_text(size = 9.5)) + theme(axis.text.y = element_text(size = 9.5))

g_box_smart_dur
```



Advanced Methodologies

```
mental_health_qs_before_covid <- data.frame(

stress_before_covid=dataset$Inversed.Stress.and.Anxiety.Level.Overall.Before.COVID.19..scale.from.1.7.,

q_of_sleep_before_covid=dataset$Quality.Of.Sleep.Before.COVID.19..scale.from.1.7.,

social_before_covid=dataset$Rating.Social.Interactions.Before.COVID.19..scale.from.1.7.

)

cronbach.alpha(mental_health_qs_before_covid)
```

```
Cronbach's alpha for the 'mental_health_qs_before_covid' data-set

Items: 3
Sample units: 66
alpha: 0.522
```

```
mental_health_qs_during_covid <- data.frame(

stress_before_covid=dataset$Inversed.Stress.and.Anxiety.Level.Overall.Durin
```

```
g.COVID.19..scale.from.1.7.,

q_of_sleep_before_covid=dataset$Quality.Of.Sleep.During.COVID.19..scale.fro
m.1.7.,

social_before_covid=dataset$Rating.Social.Interactions.During.COVID.19..sca
le.from.1.7.

)

cronbach.alpha(mental_health_qs_during_covid)
```

```
Cronbach's alpha for the 'mental_health_qs_during_covid' data-set

Items: 3
Sample units: 66
alpha: 0.2
```