



NOVEMBER 15, 2020

STA304H5F : GROUP 10

TECHNICAL PROGRESS REPORT

The impact of COVID-19 on the student's mental health

UNIVERSITY OF TORONTO, MISSISSAUGA

CONTENTS

1. INTRODUCTION	3
2. OBJECTIVE	3
3. DATA GATHERING TOOL	4
4. DATA SUMMARY	4
5. SAMPLING AND RESEARCH METHODOLOGY	5
6. RESULTS AND CONCLUSION	7
7. CONCLUSION	9
8. REFERENCES	9
9. APPENDIX	10
R-CODE USED	14

1. INTRODUCTION

Pandemics such as the COVID-19 raise questions about the state of a population's mental health. The pandemic dominates the headlines, leading to stress and anxiety among the public. Conscientiousness is necessary to stop the spread, but excessive worry may wreck other aspects of a person's health. It is normal to feel stressed and anxious during the COVID-19 outbreak. In particular, we want to monitor the pandemic's effect on mental health, which constitutes the mentioned characteristics of worry above.

Mental health is defined as, "the condition of being sound mentally and emotionally that is characterized by the absence of mental illness and by adequate adjustment especially as reflected in feeling comfortable about oneself, positive feelings about others, and the ability to meet the demands of daily life" (Merriam-Webster dictionary). As part of our study, we try to define more variables that can contribute to students' mental health. This seemingly abstract condition of mental health can be assessed more concretely by measuring three (negative) emotional states-- stress, anxiety, and depression-- since these are recognizable reactions to difficult scenarios that may require treatment. **Stress** is "a state of mental tension and worry caused by problems in your life, work" (Merriam-Webster). **Anxiety** is "a strong desire sometimes mixed with doubt, fear, or uneasiness" (Merriam-Webster). **Depression** is "a mood disorder marked especially by sadness, inactivity, difficulty in thinking and concentration" (Merriam-Webster).

2. OBJECTIVE

The purpose of this study is to analyze how the pandemic has affected the mental health of STA304 students at University of Toronto Mississauga attending the Fall 2020 semester. The classes that these students attend have been moved online, and social-distancing advisories have furthered the isolation; depending on one's personality, these may affect their mental health positively or negatively. This study also aims to determine the correlation between the three characteristics, namely stress, anxiety, and depression, and overall mental state, as well as the particular factors that may cause these feelings to be exhibited.

3. DATA GATHERING TOOL

We designed an anonymous survey with 15 questions where each question refers to a variable (a survey-item) that attributes to students' mental health. The questions were made to be short and concise as possible to avoid any ambiguity. Next, the survey was posted on the course discussion board called Piazza by instructor and TA's for students to fill out the questionnaire.

The questionnaire consisted of the following items (variables to be include in the analysis): (1) students' gender ; (2) state of their mental health before Covid-19; (3) state of their mental health during Covid-19; (4) the extent to which they felt stressed before Covid-19; (5) the extent to which they felt stressed during Covid-19; (6) the extent to which they felt anxious before Covid-19; (7) the extent to which they felt anxious during Covid-19; (8) the extent to which they felt depressed before Covid-19; (9) the extent to which they felt depressed during Covid-19; (10) the extent to which they were scared of contracting Covid-19; (11) the extent to which they were scared of their family or loved ones contracting Covid-19; (12) did they contract Covid-19; (13) did any of their family member contract Covid-19; There were two optional questions which asked for the reasons behind their betterment/ worsening of mental health due to Covid-19. We used a Likert Scale from 1-9 as our options for questions regarding variables mentioned above from (2) to (11), where 1 stands for "strongly disagree" and 9 stands for "strongly agree" The completion of the survey was optional for students. A copy of the questionnaire can be found at <https://forms.gle/xnBruX2pUvFF4hF87>.

4. DATA SUMMARY

A total of 47 students completed the survey in the fall semester of the academic year 2020-2021. Students who completed the survey were all enrolled in the STA304 course.

Before analyzing the survey results, we had to ensure that the questions in the survey measure the research topic consistently and reliably, and are inter-related to each other. To determine the reliability and consistency of the test items in the survey, Cronbach's alpha was run on the entire sample size of 47 students. Internal consistency reliability was assessed using Cronbach's alpha to measure the internal consistency of the scale, with preferred values between 0.7 and 1. Initial obtained

Cronbach's alpha value of the study-questionnaire is 0.88, which indicates that the survey-items are internally consistent and hence the questionnaire is reliable. However Cronbach's alpha analysis suggested removing an existing variable, namely "the extent to which they were scared of contracting Covid-19", and removing that variable gave Cronbach's alpha value of 0.90, hence that variable was ignored for further analysis.

5. SAMPLING AND RESEARCH METHODOLOGY

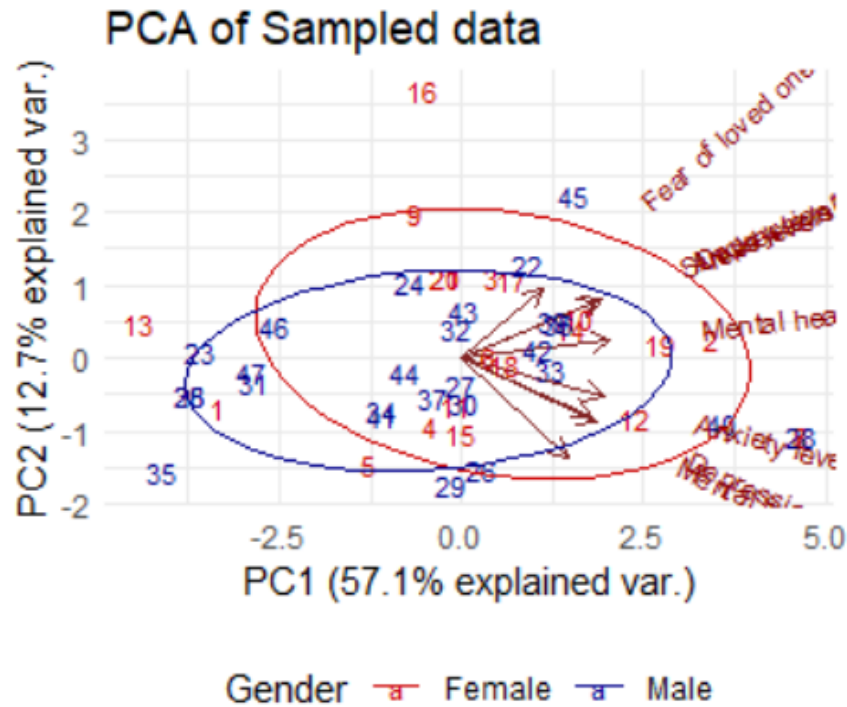
We used simple random sampling as our sample technique. This was survey based study and filled out by students at their convenience. Simple random sampling technique was easier to use in this case. Before carrying out the analysis for difference in means of each variable, we ensured that our sample size was close to the actual sample size needed to perform simple random sampling. To estimate the sample size, following R code was used.

```
> N = 270
> sigma = 2
> B = 0.55
> D = (B ** 2) / 4
> n = (N * (sigma ** 2)) / ((N - 1) * D + (sigma ** 2))
> print(n)
[1] 44.36571
```

Value of sigma was estimated using the formula $\frac{Max - Min}{4}$, where min and max values were 1 and 9.

Actual sample size with the bound on error of estimation of 0.55 was found to be 45, which is closer to our sample size 47.

Before analysing the significance of individual variables, principal component analysis was applied to check which variables preserve most of the information given by their variances. This also helps us to check any clustering in the data by different groups. Figure 1 in below shows the result of the principal component analysis.



It is noticeable in the above plot that most of the variation in the data was captured by PC1 and PC2 together. These two referred to student's mental health before and during Covid-19.

Boxplots were used for visualizing each variable. Generally, the median was higher for levels of all three variables during the Covid. The figures representing the levels of stress/anxiety/depression can be found in the appendix section, figure 2-4.

Furthermore, our sample size of 47 was greater than 30. Using the central limit theorem, this ensures that the sampling distribution of differences between variables during and before Covid-19 follows approximately normal distribution.

To test the significance of difference of each variable, we employed one sided two sample paired t-test for difference in means with the significance level of 5%. The hypothesis test below was conducted for each item. We let d denote the difference of two samples, before and during the covid for each variable of interest. We consider the following hypothesis:

$$H_0 : p = 0.5 \text{ vs } H_1 : p > 0.5$$

As expected, the stress, anxiety, and depression level of student's have risen during Covid-19. This is evident from the significant test of item 2,3,4 in Table 1 in which the p-value is notably less than 0.05 (significance level).

Table 1 summarizes the significance of each item's test along with their corresponding p-values reported.

Item	P-value	Test
Difference in the mental health during and before Covid-19.	0.002585	Significant
Difference in the level of stress during and before Covid-19	8.687e-05	Significant
Difference in the level of anxiety during and before Covid-19	0.001238	Significant
Difference in the level of depression during and before Covid-19	1.16e-06	Significant

Table 1. p-values of one-sided two sample paired t tests.

6. RESULTS AND CONCLUSION

Two sample paired T-tests were conducted to investigate the mean difference of variables, which characterizes mental health, before and during Covid-19. Based on the results, there is strong evidence to conclude that the mean difference in the levels of stress/anxiety/depression before and during Covid-19 was above 0. This means that there has been an increase in the levels of all of these three variables. Similar results hold about mental health overall.

Table 2 shows the mean, median, and standard deviation scores of overall mental health levels.

Overall	Before	During
Mean	4.064	4.894
Median	4.000	5.000

Standard Deviation	2.444197	2.664044
--------------------	----------	----------

Table 2. Overall mental health scores.

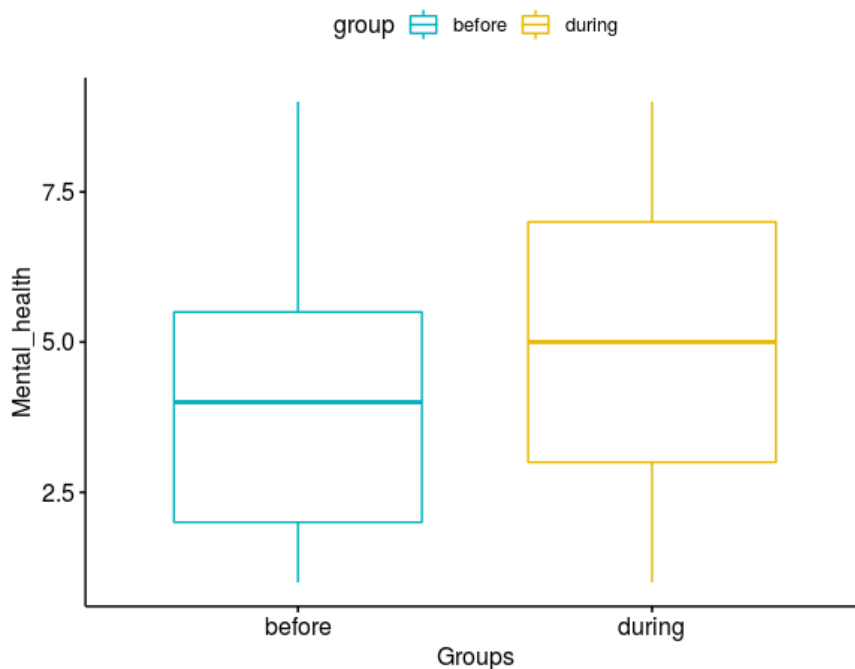


Figure 1. This plot illustrates the data from **Table 1**.

Before the pandemic, we estimated that the level of overall mental health mean is 4.064. This result is from the scale of 1 to 10 as mentioned earlier. Using the Simple Random Sampling formula for the estimated variance of mean, we found the estimated variance and bound to be 0.1049822 and 0.6480191. This bound is a bit on the higher side than our predicted bound of 0.55. The 95% confidence interval for the mean of the level of mental health before Covid-19 is (3.4160, 4.7120).

During the pandemic, we estimated that the overall mental health mean is 4.894. This result is again from the scale of 1 to 10. The estimated variance and bound was found to be 0.1247171 and the bound of error is 0.7063061, which is on the higher side of our predicted bound of 0.55. This might be due to the increase in the variance of data during Covid-19. The 95% confidence interval for the mean of the level of mental health during Covid-19 is (4.187694, 5.600306).

From our two sample paired t-test analysis on the mental health variable, the test statistic was found to be 0.8297872 and 95% confident was (0.2609841, 1.3985904).

Table 3 shows the correlation matrix between each category. It can be noted that Anxiety levels before Covid-19 were most correlated with Mental health before Covid-19. And, Depression levels during Covid-19 was most correlated with Mental health during Covid-19. Since this was a survey based study, causal relationships can not be established.

Table 3 shows the correlation matrix between each category.

	Mental health before Covid-19	Mental health during Covid-19	Stress levels before Covid-19	Stress levels during Covid-19	Anxiety levels before Covid-19	Anxiety levels during Covid-19	Depression levels before Covid-19	Depression levels during Covid-19
Mental health before Covid-19	1.0000000	0.7155262	0.5811242	0.4433055	0.7622344	0.4156523	0.5888319	0.4566728
Mental health during Covid-19	0.7155262	1.0000000	0.3485269	0.6234079	0.6580649	0.6477096	0.6086239	0.7324502
Stress levels before Covid-19	0.5811242	0.3485269	1.0000000	0.4256744	0.5552607	0.3530237	0.5943289	0.2008230
Stress levels during Covid-19	0.4433055	0.6234079	0.4256744	1.0000000	0.5909046	0.8299686	0.3744309	0.6146527
Anxiety levels before Covid-19	0.7622344	0.6580649	0.5552607	0.5909046	1.0000000	0.6605307	0.5636616	0.4811350
Anxiety levels during Covid-19	0.4156523	0.6477096	0.3530237	0.8299686	0.6605307	1.0000000	0.4996236	0.6819592
Depression levels before Covid-19	0.5888319	0.6086239	0.5943289	0.3744309	0.5636616	0.4996236	1.0000000	0.6617122
Depression levels during Covid-19	0.4566728	0.7324502	0.2008230	0.6146527	0.4811350	0.6819592	0.6617122	1.0000000

Table 3.

7. CONCLUSION

Based on the sampled data, it appears that the median and mean of the Likert Scale scores are higher for poor mental health levels; students agree more to the statement of “Currently, my mental health is poor” than the statement of “Before COVID-19, my mental health was poor.” Thus, it is reasonable to conclude that the mental health of Fall 2020 STA304 students has decreased since the beginning of this pandemic.

8. REFERENCES

Anxiety. (n.d.). <https://www.dictionary.com/browse/anxiety?s=t>

Depression. (n.d.).<https://www.merriam-webster.com/dictionary/depression>

Mental Health. (n.d.).<https://www.merriam-webster.com/dictionary/mental health>

Stress. (n.d.).<https://www.merriam-webster.com/dictionary/stress>

9. APPENDIX

EXTRA FIGURES

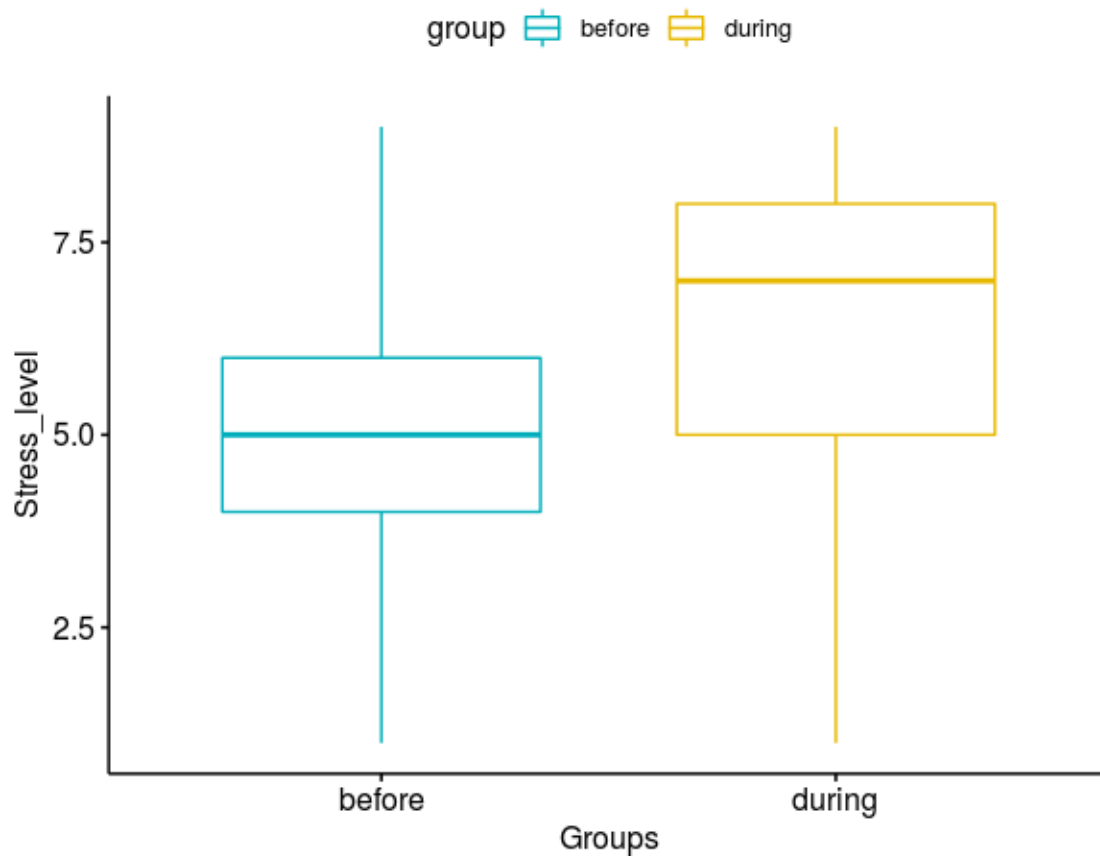


Figure 2. Represents the stress levels before and during the pandemic.

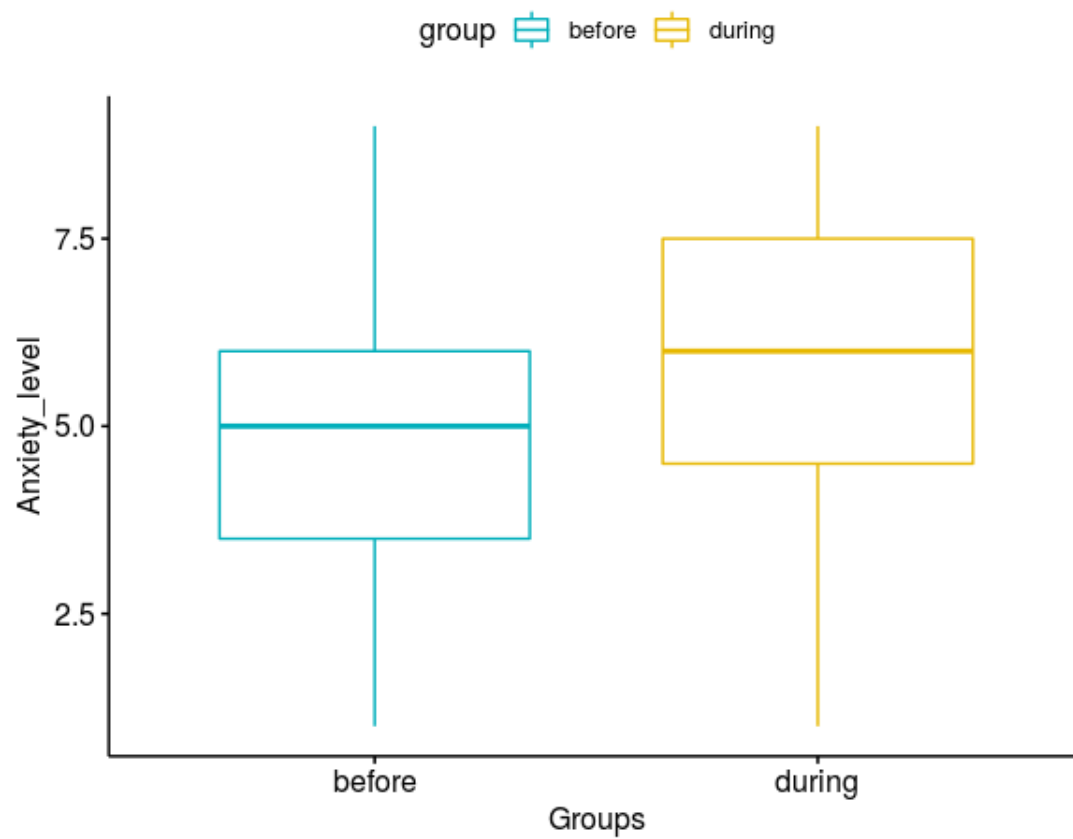


Figure 3. represents the anxiety levels before and during the pandemic.

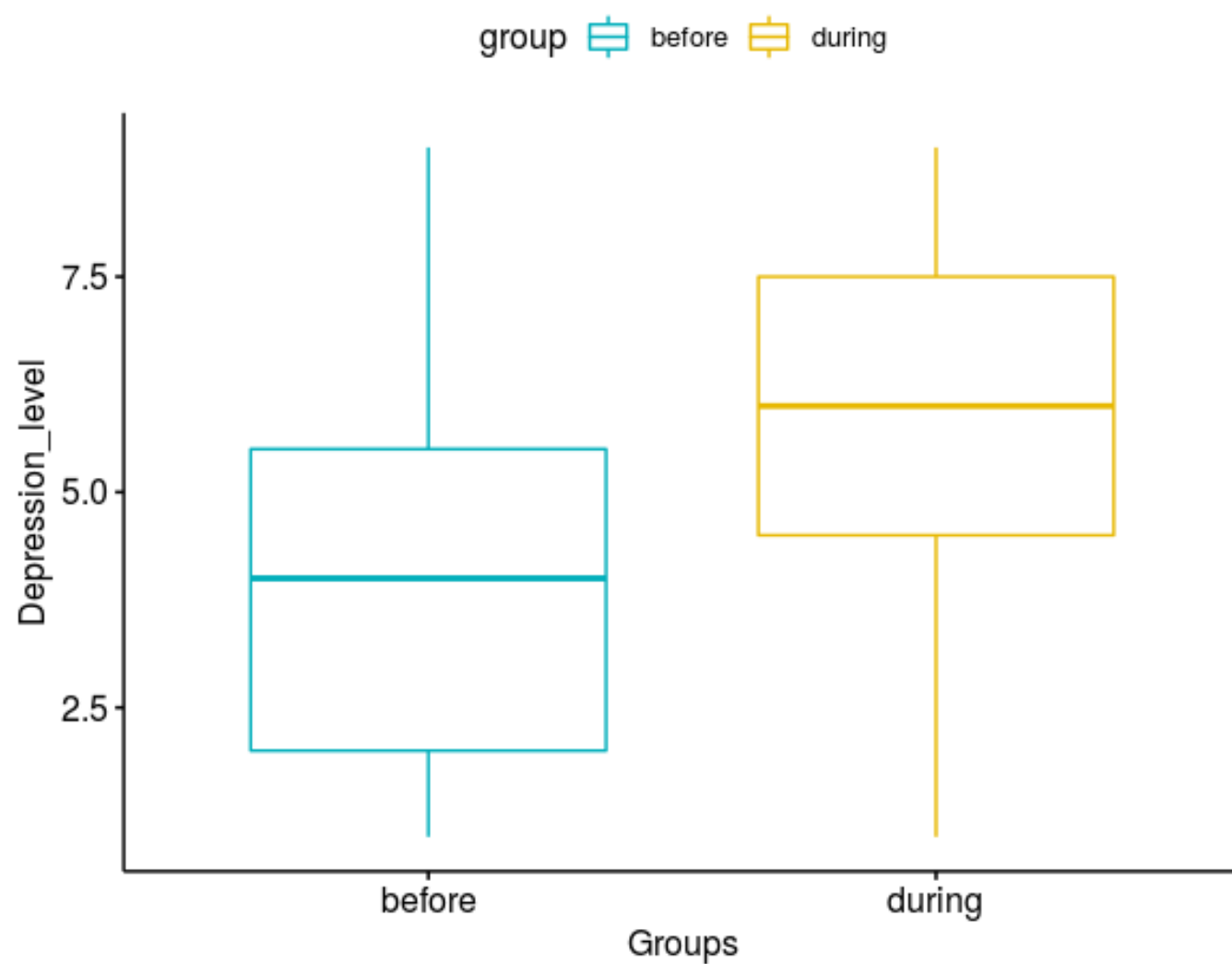
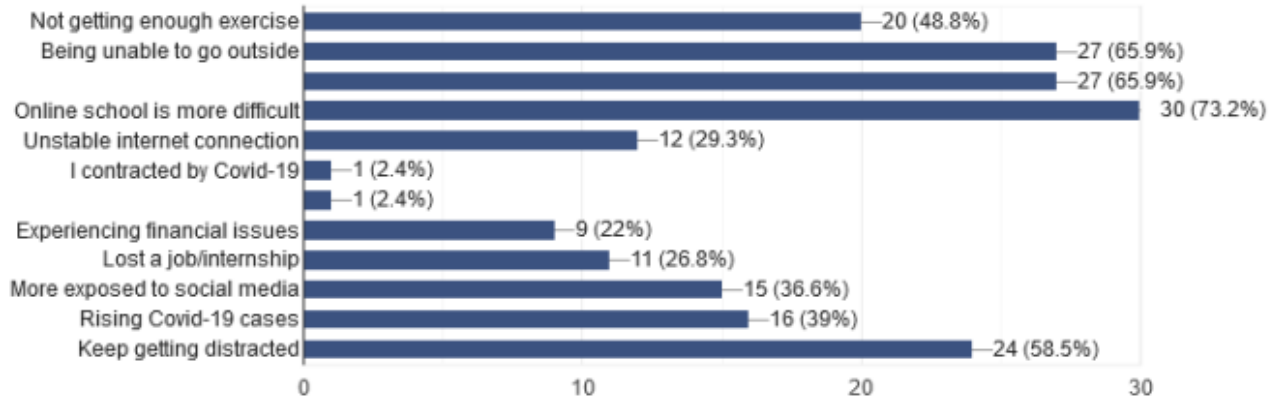


Figure 4. represents the depression levels before and during the pandemic.

If you suspect that you are feeling ***more*** stressed/anxious/depressed since the beginning of the pandemic, please select the following reasons that you relate to.

41 responses



If you suspect that you are feeling ***less*** stressed/anxious/depressed since the beginning of the pandemic, please select the following reasons that you relate to.

32 responses

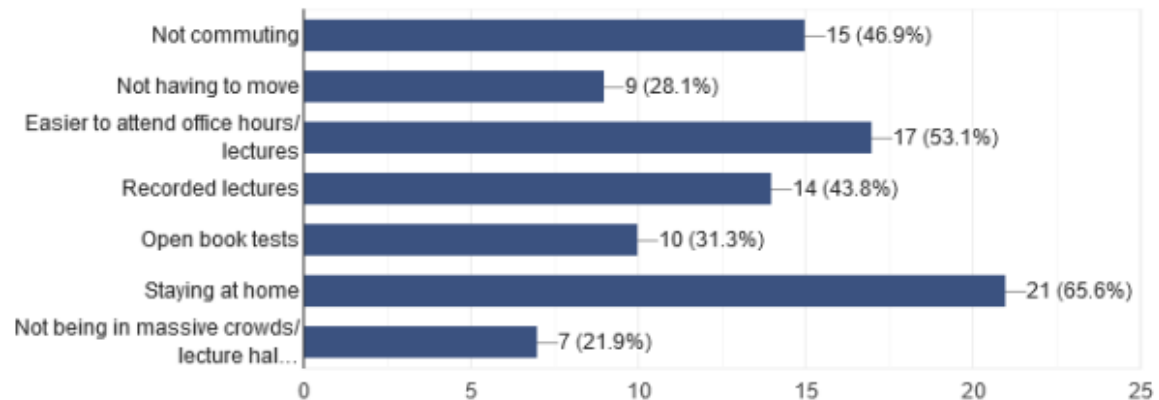


Figure 5. Showing the responses for two optional questions

R-CODE USED

Mean, Median, Standard Deviation

```
> summary(STA304)
```

```

Timestamp          Consent          Gender
Mental.health.before.Covid.19
Length:47          Length:47          Length:47          Min.
:1.000
Class :character    Class :character    Class :character    1st
Qu.:2.000
Mode  :character    Mode  :character    Mode  :character
Median :4.000
                                     Mean
:4.064
                                     3rd
Qu.:5.500
                                     Max.
:9.000
Mental.health.during.Covid.19 Stress.levels.before.Covid.19
Stress.levels.during.Covid.19
Min.    :1.000          Min.    :1.000
Min.    :1.00
1st Qu.:3.000          1st Qu.:4.000
1st Qu.:5.00
Median :5.000          Median :5.000
Median :7.00
Mean   :4.894          Mean   :4.936
Mean   :6.34
3rd Qu.:7.000          3rd Qu.:6.000
3rd Qu.:8.00
Max.    :9.000          Max.    :9.000
Max.    :9.00

```

Anxiety.levels.before.Covid.19	Anxiety.levels.during.Covid.19
Depression.levels.before.Covid.19	
Min. :1.000	Min. :1.000
Min. :1.0	
1st Qu.:3.500	1st Qu.:4.500
1st Qu.:2.0	
Median :5.000	Median :6.000
Median :4.0	
Mean :4.957	Mean :5.894
Mean :4.0	
3rd Qu.:6.000	3rd Qu.:7.500
3rd Qu.:5.5	
Max. :9.000	Max. :9.000
Max. :9.0	
Depression.levels.during.Covid.19	
Min. :1	
1st Qu.:2	
Median :6	
Mean :5	
3rd Qu.:7	
Max. :9	

```

> sd(STA304$Mental.health.before.Covid.19)
[1] 2.444197
> sd(STA304$Mental.health.during.Covid.19)
[1] 2.664044
> sd(STA304$Stress.levels.before.Covid.19)
[1] 2.099733
> sd(STA304$Stress.levels.during.Covid.19)
[1] 2.286792
> sd(STA304$Anxiety.levels.before.Covid.19)
[1] 2.44022
> sd(STA304$Anxiety.levels.during.Covid.19)

```



```
[1] 2.424818
> sd(STA304$Depression.levels.before.Covid.19)
[1] 2.386557
> sd(STA304$Depression.levels.during.Covid.19)
[1] 2.629267
```

Calculating Variance and Bound

Mental Health, Before Pandemic

```
> n = 47
> mean = 4.064
> s = 2.444197
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound
```

Mental Health, During Pandemic

```
> mean = 4.894
> s = 2.664044
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound
```

Stress Levels, Before Pandemic

```
> mean = 4.936
> s = 2.099733
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound
```

Stress Levels During pandemic

```
> mean = 6.340
```

```

> s = 2.286792
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound

```

Anxiety Levels Before Pandemic

```

> mean = 4.957
> s = 2.44022
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound

```

Anxiety Levels During Pandemic

```

> mean = 5.894
> s = 2.424818
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound

```

Depression Levels Before Pandemic

```

> mean = 4.000
> s = 2.386557
> est_var = (1 - (n / N)) * ((s ** 2) / n)
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound

```

Depression Levels During Pandemic

```

> mean = 5.000
> s = 2.629267
> est_var = (1 - (n / N)) * ((s ** 2) / n)

```

```
> bound = 2 * (est_var ** (1/2))
> min = mean - bound
> max = mean + bound
```

Everything Else (Barplots, Correlation Matrix, Cronbach's Alpha)

```
library(readxl)
install.packages("ggpubr")
library("ggpubr")
install.packages("corrplot")
source("http://www.sthda.com/upload/rquery_cormat.r")
require(dplyr)
install.packages("umx")
install.packages("psych")
library(psych)
library(devtools)
install_github("vqv/ggbiplot")
library(ggbiplot)

data = read_excel("STA304_Group_10_Responses.xlsx")

# Processing the data. Removing first two columns.
df <- data[-c(1,2)]

# Ordering the data by Gender
df = df[order(df$Gender),]
View(df)

# Performing Cronbach's alpha to check internal consistency.
alpha(df[c(2:11)]) #raw_alpha = 0.88. This also tells us to remove
the column "Fear of contracting covid-19"

#Removing the column gives us raw_alpha = 0.90
```

```

df = df[c(-10)]
alpha(df[c(2:10)]) # Raw alpha of 0.9
View(df)

# Using PCA(Prinipal Component Analysis) to factor out the variables
which explains the most variance.
df.pca = prcomp(df[,c(2:10)], center = TRUE,scale. = TRUE)
summary(df.pca)
df.group <- c(rep("Female", 21), rep("Male",26))

ggbiplot(df.pca,ellipse=TRUE,obs.scale = 1, var.scale = 1,
labels=rownames(df), groups=df.group) +
  scale_colour_manual(name="Gender", values= c("red3", "dark blue"))+
  ggtitle("PCA of Sampled data")+
  theme_minimal()+
  theme(legend.position = "bottom")

# Using Boxplots to visualize the difference between two time periods
of all the variables.
# Plot each variable by group and color by group

var_plot <- function(before,during,var_name){
  my_data <- data.frame(
    group = rep(c("before", "during"), each = 47),
    var_vec = c(before, during)
  )

  ggboxplot(my_data, x = "group", y = "var_vec",
    color = "group", palette = c("#00AFBB", "#E7B800"),
    order = c("before", "during"),
    ylab = var_name, xlab = "Groups")
}

```

```

result <- function(during,before){
  res <- t.test(during, before, paired = TRUE,alternative =
"greater")
  CI = t.test(during, before, paired = TRUE,conf.level = 0.95)
  return (list(res,CI))
}
# n > 30 so using CLT, we can assume that sampling distribution of
differences of the variables before and during COVID follow Normal
distribution
# Since we can assume normality for differences, we can use two
sample paired t-test to analyze the data.
# d = var_during_covid - var_before_covid.
# Ho = d_bar = 0. Ha: d_bar > 0.

# Mental health
before = df$`Mental health before Covid-19`
during = df$`Mental health during Covid-19`
var_plot(before,during,"Mental_health")

# check for significance and CI
res = result(during,before)

# Stress Level
before = df$`Stress levels before Covid-19`
during = df$`Stress levels during Covid-19`
var_plot(before,during,"Stress_level")

# check for significance and CI
res = result(during,before)

# Anxiety Level
before = df$`Anxiety levels before Covid-19`

```

```

during = df$`Anxiety levels during Covid-19`
var_plot(before,during,"Anxiety_level")

# check for significance and CI
res = result(during,before)

# Depression Level
before = df$`Depression levels before Covid-19`
after = df$`Depression levels during Covid-19`
var_plot(before,during,"Depression_level")

# check for significance and CI
res = result(during,before)
res

# Contracting Covid-19
personal_fear = df$`Fear of contracting COVID-19.`
family_fear = df$`Fear of loved ones contracting COVID-19.`
my_data <- data.frame(
  group = rep(c("personal_fear", "family_fear"), each = 47),
  var_vec = c(personal_fear, family_fear)
)
ggboxplot(my_data, x = "group", y = "var_vec",
  color = "group", palette = c("#00AFBB", "#E7B800"),
  order = c("personal_fear", "family_fear"),
  ylab = "Fear_level", xlab = "Groups")

#correlation matrix
reduced_data = df[c(2:9)]
mymatrix = cor(reduced_data)
View(mymatrix)

#Partition the data by gender

```

```

gender_data = df[c(1:9)]
View(gender_data)
#gender_data = df[c(1:11)]
f_data = gender_data[gender_data$Gender=="Female",]
m_data = gender_data[gender_data$Gender=="Male",]
f_updated_data = f_data[c(2:9)]
m_updated_data = m_data[c(2:9)]
m_matrix = cor(m_updated_data)
f_matrix = cor(f_updated_data)
View(m_matrix)
View(f_matrix)

# Partition data by people whose family members contracted Covid.

family_covid = df[df$`Family members contracted COVID-19` == "Yes",]
family_covid_data = family_covid[c(2:9)]
cor_mat = cor(family_covid_data)
View(cor_mat)

family_no_covid=df[df$`Family members contracted COVID-19` == "No",]
family_no_covid_data = family_no_covid[c(2:9)]
cor_matrix = cor(family_no_covid_data)
View(cor_matrix)

```