UTM UNIVERSITI TEKNOLOGI MALAYSIA RESEARCH UNIVERSITY	**UNIVERSITI TEKNOLOGI MALAYSIA**

Probability & Statistical Data Analysis (SECI 1143)

PROJECT 2

Mental Health and Digital Behaviour

Student Name : 1. Angela Ngu Xin Yi (A24CS0226)
2. Cheong Yi Shien (A24CS0058)
3. Tan Xin Tian (A24CS0198)
4. Teoh Xin Yee (A24CS0307)

Lecturer Name : Dr Shamini A/P Raja Kumaran

Section : 2

Table of Content

1.0 Introduction

Mental health has emerged as one of the most pressing public health challenges in recent years, particularly in the wake of rapid technological advancement and increased screen exposure. Following the COVID-19 pandemic, people's lives have become increasingly shaped by digital technology for work, study, social interaction and entertainment. Excessive screen time, frequent notifications and non-stop social media engagement are now common elements of daily life, particularly in young adults and students. While digital tools offer convenience and connectivity, there is growing concern about its potential adverse effects on mental health.

This study aims to investigate various aspects of digital behaviours, such as screen time, app switching frequency, social media usage, sleep duration and notification exposure. These behaviours relate to mental health indicators such as anxiety levels, mood, focus and overall digital wellbeing. By utilizing inferential statistical methods, we seek to uncover patterns and potential causative relationships between these variables that can inform better personal and public health decisions.

The dataset used in this study is the Mental Health and Digital Behavior (2020–2024) dataset, sourced from Kaggle and curated by Atharva Soundankar. This dataset comprises information from over 1000 individuals. It is collected during a crucial post-pandemic period when digital reliance and mental health concerns significantly increased. The dataset includes key behavioural and psychological metrics.

To analyze this data, we employ a combination of two-sample hypothesis tests, correlation analysis and regression, as well as ANOVA test and Chi Square Test. These techniques will allow us to determine whether significant relationships exist between digital behaviours and mental health outcomes. Moreover, we can identify which variables are the most predictive of negative or positive mental health states.

The primary objectives of this project are to apply and perform appropriate statistical test analyses on the selected dataset. Besides, it can determine the dependency and association relationships between digital habits and mental health indicators.

By achieving these objectives, this study aims to provide meaningful insights into how lifestyle in the digital age affects mental health. The findings could be valuable in shaping digital wellness programs, guiding mental health interventions and helping individuals make more informed choices about their technology use.

2.0 Dataset

2.1 Dataset Overview

The dataset used in this study is titled "Mental Health and Digital Behavior (2020–2024)", publicly available on Kaggle and curated by Atharva Soundankar. This dataset contains observations from over 500 individuals across a five-year period. It captures key indicators related to digital usage habits and mental health status. The dataset is particularly relevant due to its timing. For this reason, it reflects post-pandemic digital behaviour when screen reliance increased significantly for both professional and personal activities.

In this study, the population refers to all digitally active individuals, particularly students, young adults and working professionals. These individual groups' mental health may be influenced by their screen usage, social media activity and other digital behaviours. This includes individuals across different countries, backgrounds and professions who engage regularly with technology.

The sample comprises 500 participants whose behavioural and mental health data were collected and compiled in the Mental Health and Digital Behavior dataset. It is assumed that the sample reflects a diverse and representative group of individuals affected by digital engagement trends during the 2020–2024 period.

2.2 Selected Variables

Table 2.1: Description of Selected Variables

No.	Variables	Description	Type of Variables	Level of Measurement
1.	daily_screen_time_min	Total screen time in minutes per day	Numerical	Ratio
2.	sleep_hours	Average number of hours the individual sleeps per night	Numerical	Ratio
3.	notification_count	Total number of notifications received in a day	Numerical	Ratio
4.	social_media_time_min	Time spent on social media in minutes per day	Numerical	Ratio
5.	focus_score	Self-reported focus level on a 0-100 scale	Numerical (Score)	Interval
6.	mood_score	Self-reported mood level on a 0-100 scale	Numerical (Score)	Interval
7.	anxiety_level	Self-reported anxiety level	Numerical (Score)	Interval
8.	digital_wellbeing_score	Composite score measuring overall digital wellbeing	Numerical (Score)	Interval

The selection of variables for this study was guided by their relevance to common digital habits and their potential psychological impacts. Variables such as **daily_screen_time_min** and

social_media_time_min were chosen to reflect general patterns of screen usage. They have been frequently linked to mental health outcomes in existing literature. **Sleep_hours** was included because sleep quality is a critical factor in maintaining emotional balance and digital self-regulation. Notification_count was selected as it represents a measure of digital interruptions, which are often associated with increased stress or anxiety. The psychological indicators—mood_score, focus_score, anxiety_level, and digital_wellbeing_score—were included to capture different dimensions of mental health, ranging from emotional states to cognitive performance and overall wellbeing. These combinations of behavioral and psychological variables allow for a comprehensive investigation of how digital behaviors may influence mental health, making them suitable for hypothesis testing, correlation, regression, and ANOVA.

In the process of designing statistical tests, the variable **num_app_switches** is not used. For this reason, its interpretability and direct relationship with mental health outcomes is weaker if compared to the selected variable. This ensures the statistical tests conducted are relevant to research questions, statistically appropriate and interpretable for real-world conclusions.

2.3 Data Preprocessing

Before conducting the statistical analyses, several preprocessing steps were carried out to ensure the quality and consistency of the data. First, the variable types were reviewed to confirm that all relevant fields were in the appropriate numerical format, suitable for inferential methods such as correlation, regression and ANOVA. To facilitate group comparisons, the **daily_screen_time_min** variable was converted into a categorical variable named "Screen Time Group" by dividing it into two groups: Low (≤ 300 minutes) and High (> 300 minutes). Similarly, **social_media_time_min** was grouped into Low, Medium, and High usage categories based on quantile ranges to ensure balanced sample sizes across groups for ANOVA testing.

2.4 Statistical Test Analysis

Table 2.2: Statistical Test Analysis

No.	Selected Variables	Objectives	Test Analysis and Expected Outcome
1.	daily_screen_time_min mood_score	To determine whether there is a significant difference in mood scores between individuals who spend more time on screens and those who spend less	**Test Analysis:** Two-sample hypothesis test (Test on mean, variance unknown) **Possible Outcome:** By conducting a two-sample hypothesis test, we can compare the mean mood score between individuals with high and low screen time. If the p-value is significant (typically $p < 0.05$), it would suggest that there is a statistically significant difference in mood levels between the two groups. This would imply that screen time usage may have an influence on emotional wellbeing.
2.	sleep_hours digital_wellbeing_score	To examine the relationship between sleep duration and overall digital wellbeing	**Test Analysis:** Correlation Analysis **Possible Outcome:** The analysis is expected to reveal whether a significant linear relationship exists between sleep duration and digital wellbeing score, suggesting that individuals who sleep longer tend to have better digital wellbeing.
3.	notification_count anxiety_level	To determine whether the number of notifications a person receives per day can predict their anxiety level	**Test Analysis:** Simple Linear Regression **Possible Outcome:** If the p-value < 0.05, it indicates a statistically

			significant linear relationship between notification count and anxiety level. The regression equation will provide a quantitative estimate of how anxiety level is expected to change with notification count.
4.	social_media_time_min focus_score	To determine whether individuals with different levels of social media usage (Low, Medium, High) differ significantly in their average focus scores.	**Test Analysis:** One-way ANOVA with equal sample size **Possible Outcome:** By conducting a one-way ANOVA test, we can compare the mean focus scores across different social media usage groups. If the F-statistic yields a p-value that is significant ($p < 0.05$), it would suggest that there are statistically significant differences in focus levels between at least two of the groups. This would imply that the amount of time spent on social media may have an impact on an individual's ability to concentrate or maintain focus during daily tasks.

3.0 Data Analysis

3.1 Two-Sample Hypothesis Test

Variables **daily_screen_time_min** and **mood_score** are used in this two-sample hypothesis test. In order to perform this test, a new categorical variable named **Screen Time Group** is created by dividing participants into two groups based on their **daily_screen_time_min values**.

- Group 1 - Low Screen Time: ≤ 300 minutes
- Group 2 - High Screen Time: > 300 minutes

In this analysis, it will test whether mean mood score of high screen time users is different from that of low screen time users at the 95% confidence level, assuming that the variances are unequal. From the data, frequency (n), mean ($\bar{x}$) and standard deviation (s) are calculated.

Table 3.1.1: Frequency, mean and standard deviation of Screen Time Group

Group	Frequency (n)	Mean ($\bar{x}$)	Standard Deviation (s)
Low Screen Time	$n_1 = 76$	$\bar{x}_1 = 8.912$	$s_1 = 0.5256$
High Screen Time	$n_2 = 424$	$\bar{x}_2 = 9.002$	$s_2 = 0.5174$

```
> print(summary_table)
             Group Frequency   Mean     SD
1  Low Screen Time        76 8.9120 0.5260
2 High Screen Time       424 9.0024 0.5174
```

Figure 3.1.1: Frequency, mean and standard deviation of Screen Time Group using RStudio

1. **Hypothesis Statement**

$$\mathbf{H_0}: \mu_1 = \mu_2$$
$$\mathbf{H_1}: \mu_1 \neq \mu_2$$

Where μ_1 represents the mean mood score of low screen time users and μ_2 represents the mean mood score of high screen time users.

2. **Test Statistic**

Given 95% confidence level, $\alpha = 0.05$, the test statistics, $\mathbf{t_0 = -1.386}$ and **p-value = 0.1688** by using R-Studio.

```
t0          -1.38581411894815
```

Figure 3.1.2: Test Statistics using RStudio

```
p_value     0.168796882904207
```

Figure 3.1.3: P value using RStudio

3. **Degree of freedom**

By using RStudio, degree of freedom, $\mathbf{v = 102.762 \approx 103}$.

```
v           102.761827668907
```

Figure 3.1.4: Degree of Freedom from RStudio

Therefore, using $\alpha = 0.05$, we reject H0 if $t_0 < t_{-0.025,\,103}$ or $t_0 > t_{0.025,103}$.

4. Critical value

Critical value, $\mathbf{t_{-0.025,103} = -1.983}$, $\mathbf{t_{0.025,\,103} = 1.983}$ if according to RStudio.

```
t.alpha        -1.98349525856288
```

Figure 3.1.5: Critical Value from RStudio

5. Interpretation

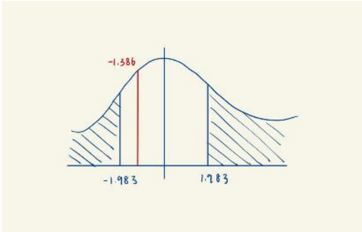

Figure 3.1.6: Test statistic location for correlation test

Since $t_0 > t_{-0.025,\,103}$ and $t_0 < t_{0.025,103}$, at the same time, $p > 0.05$, we **fail to reject** the null hypothesis. Hence, there is no sufficient evidence to conclude that the mean mood score between individuals with high and low screen time are different.

```
	Welch Two Sample t-test

data:  mood_score by screen_time_group
t = -1.3858, df = 102.76, p-value = 0.1688
alternative hypothesis: true difference in means between group Group1 and group Group2 is not equal to 0
95 percent confidence interval:
 -0.22005962  0.03902685
sample estimates:
mean in group Group1 mean in group Group2 
            8.911842             9.002358 
```

Figure 3.1.7: Two-Sample Hypothesis Result from RStudio

6. Conclusion

In short, there is no sufficient evidence that it exists a difference in mood scores between high and low screen time users, suggesting that screen time may not strongly impact emotional wellbeing. Mood may be influenced more by how screen time is used rather than how long one spends on screens. For example, someone who uses a screen to play games might be happier than someone who uses it for stressful work, even for the same amount of time.

3.2 Correlation Test

In this analysis, we used the variables **Sleep Hours and Digital Wellbeing Score** to determine whether there is a linear relationship between the amount of screen time and the mental health rating of a respondent. This analysis was conducted at a 95% confidence level using **Pearson's Product-Moment Correlation Coefficient** which suitable for two continuous, ratio-level variables. We aim to calculate the sample correlation coefficient, r.

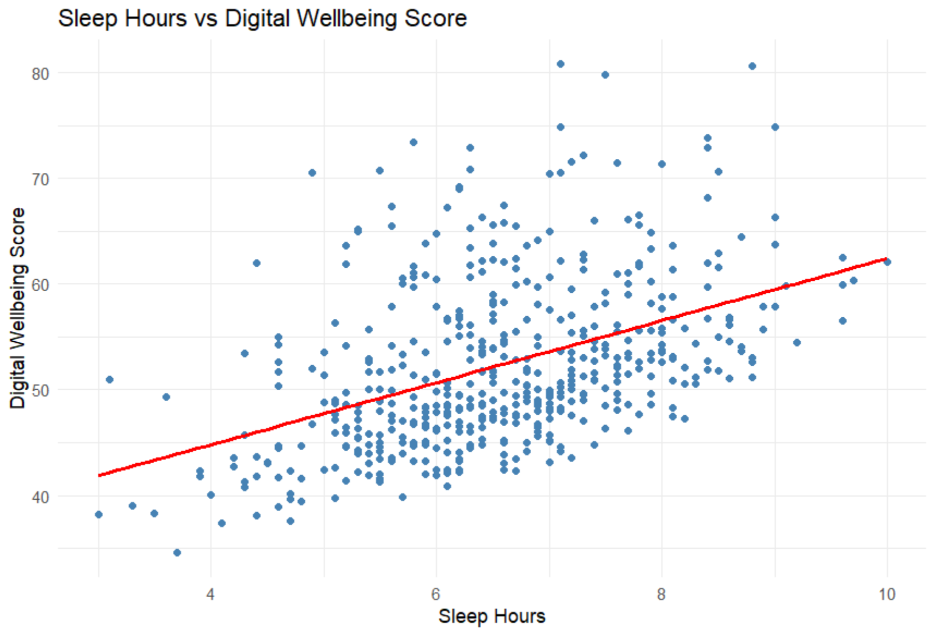

Figure 3.2.1: Scatter plot of screen time vs digital wellbeing score using RStudio

1. Sample correlation coefficient (Pearson's method)

```
> correlation <- cor(sleep_hours, digital_wellbeing_score, method = "pearson")
> print(paste("Pearson's correlation coefficient: ",correlation))
[1] "Pearson's correlation coefficient:  0.440425633129268"
```

Figure 3.2.2: Sample Correlation Coefficient using RStudio

By using RStudio, we can see that the sample correlation coefficient, r is 0.4404. This suggests a moderate positive linear correlation between sleep hours and digital wellbeing score.

2. Significance Test for Correlation

A) Hypothesis statement

H_0: $\rho = 0$ (no linear correlation)

H_1: $\rho \neq 0$ (linear correlation exists)

where ρ represents the population correlation coefficient.

B) Test Statistic by Test statistic formula, t

```
> r <- correlation
> n <- sum(complete.cases(mental_health_digital_behavior_data$sleep_hours,
+                         mental_health_digital_behavior_data$digital_wellbeing_score))
> t <- (r * sqrt(n - 2)) / sqrt(1 - r^2)
> print(paste("Test Statistic:",t))
[1] "Test Statistic: 10.9474490357277"
```

Figure 3.2.3: Test Statistic using RStudio

By using RStudio, we get the test statistic, **t = 10.9474**

```
> p_value <- 2 * pt(-abs(t), df)
> print(paste("p-value:", p_value))
[1] "p-value: 3.85933032712111e-25"
```

Figure 3.2.4: P value using RStudio

By using RStudio, we get the p-value = **3.85933032712111e-25**

Find critical value, using $\alpha = 0.05$, **df = n-2 = 498**

From t-table, since this is a two-tailed test, there are two critical values:

Lower tail critical value $-t_{0.025,498}$ = **-1.9647**

Higher tail critical value $t_{0.025,498}$ = **1.9647**

```
> alpha <- 0.05
> df <- n-2
> critical_value <-qt(1-alpha/2, df)
> print(paste("Critical Value at 95% confidence level: ",critical_value))
[1] "Critical Value at 95% confidence level:  1.96473898296729"
```

Figure 3.2.5: Critical Value using RStudio

From RStudio, the critical value that we obtained is 1.9647. Hence, if the test statistic > 1.9647 or the test statistic < -1.9647, we will reject the null hypothesis, H_0. Otherwise, we will fail to reject the null hypothesis, H_0.

3. Interpretation

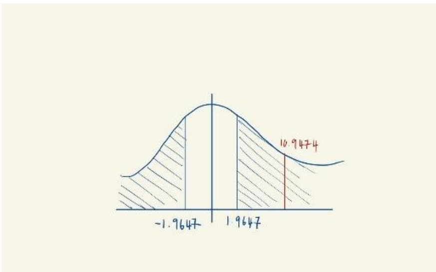

Figure 3.2.6: Test statistic location for correlation test

Since t = 10.9474 > 1.9647, the test statistic falls in the rejection region at the 0.05 significance level. Therefore, we **reject the null hypothesis**. This indicates that there is sufficient evidence to conclude a significant positive linear relationship between sleep hours and digital wellbeing score. Thus, at $\alpha = 0.05$, we conclude that **sleep hours significantly affect digital wellbeing score** in the sampled population.

```
> correlation_test <- cor.test(sleep_hours, digital_wellbeing_score, method = "pearson")
> print(correlation_test)

	Pearson's product-moment correlation

data:  sleep_hours and digital_wellbeing_score
t = 10.947, df = 498, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3669054 0.5084782
sample estimates:
      cor 
0.4404256 
```

Figure 3.2.7: Performing significance test for correlation using RStudio

4. **Conclusion**

Overall, the finding indicates that there is sufficient evidence to conclude a significant linear relationship between sleep hours and digital wellbeing score. From the correlation coefficient, r (0.4404), we can conclude that a moderate positive correlation was found. This suggests that individuals who sleep longer tend to report higher levels of better digital wellbeing. Adequate sleep controls mood, reduce stress and enhances attention which can improve how a person interacts with digital platform. On the other hand, poor sleep can lead to negative digital behavior such as digital fatigue.

3.3 Regression Test

This study aims to examine whether the number of notifications received per day can predict a person's anxiety level using a simple linear regression analysis. In this study, we treat notification count as an independent variable (x) and anxiety level as a dependent variable (y). If the regression coefficient is found to be significant ($p < 0.05$), it would imply that there is a linear relationship between number of notification and anxiety level and an increase in notification count is associated with changes in anxiety levels at significant level, $\alpha = 5\%$.

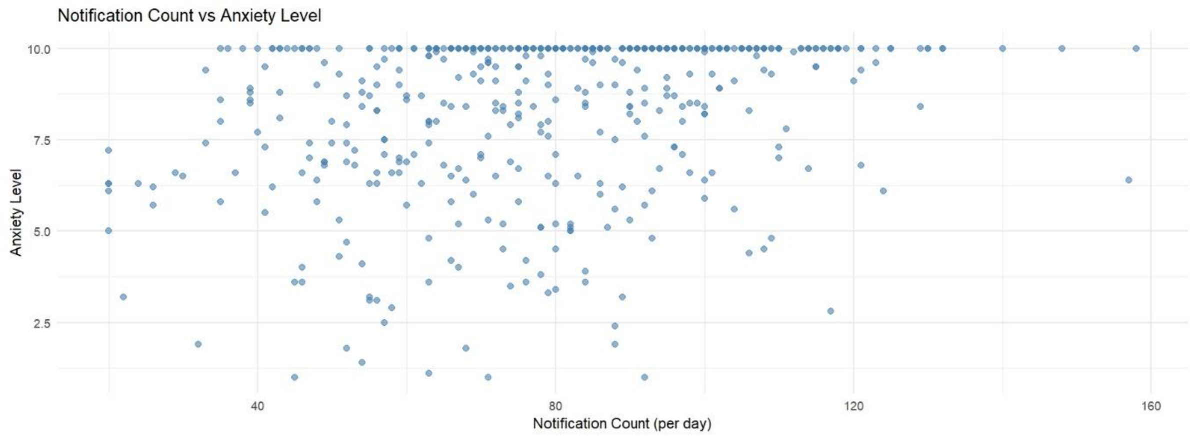

Figure 3.3.1: Scatter plot of notification count vs anxiety level using RStudio

To build the Population Linear Regression Model, we assume:

- Error values (ε) are statistically independent, and is normally distributed for any x
- The probability distribution of errors has a constant variance
- The underlying relationship between variable x and variable y is linear

1. **Estimated Regression Model**

```
> # Extract b₀ and b₁
> b0 <- coef(model)[1]  # Intercept
> b1 <- coef(model)[2]  # Slope
> cat("Intercept (b0):", b0, "\n")
Intercept (b0): 6.321495
> cat("Slope (b1):", b1, "\n")
Slope (b1): 0.02679225
```

Figure 3.3.2: Intercept, b_0 and slope b_1 in estimated regression model using RStudio

By using RStudio, we obtain $\mathbf{b_0 = 6.321495}$, $\mathbf{b_1 = 0.026792}$

By substitute the value b_0 and b_1 into the estimated regression model, we get

$$\hat{y}i = 6.321495 + 0.02679225x$$

```
Regression Equation: anxiety_level = 6.321 + 0.027 * notification_count
```

Figure 3.3.3: Regression Equation using RStudio

Based on the regression model equation above, b_0 = 6.321495 is the estimated average value of anxiety level when notification count is 0 while b_1 = 0.026792 is the estimated change in the average value of anxiety level as the result of one-unit change in notification count.

2. **Explained and Unexplained Variation**

Using RStudio, we obtain **SSE = 2067.357**, **SSR = 205.9718** and **SST = 2273.328**.

```
> # Print SSR, SSE, SST
> cat("SST =", sst, "\nSSR =", ssr, "\nSSE =", sse)
SST = 2273.328
SSR = 205.9718
SSE = 2067.357
```

Figure 3.3.4: Value of SSE, SSR, SST using RStudio

3. **Coefficient of Determination, R^2**

We get coefficient of determination, $\mathbf{R^2 = 0.0906}$ in RStudio.

```
> cat("R-squared:", round(r_squared, 5), "\n")
R-squared: 0.0906
```

Figure 3.3.5: R-squared value using RStudio

4. **Standard Error of Estimate, S_ε**

By using RStudio, we get the residual standard error, $\mathbf{S_\varepsilon = 2.037}$.

```
Residual standard error: 2.037 on 498 degrees of freedom
```

Figure 3.3.6: Value of residual standard error, S_ε

5. **Standard deviation of the Regression Slope, S_{b1}**

By using RStudio, we get the standard deviation of regression slope, $\mathbf{S_{b1} = 0.003804}$.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)        6.321495   0.315862  20.013  < 2e-16 ***
notification_count 0.026792   0.003804   7.044 6.25e-12 ***
```

Figure 3.3.7: Value of Standard deviation of the Regression Slope, S_{b1}

6. **Inference About the Slope: t-test**

A) Hypothesis statement

H_0: $\beta_1 = 0$ (no linear relationship)

H_1: $\beta_1 \neq 0$ (linear relationship does exists)

B) Critical Value

A two-tailed test is conducted, $\alpha/2 = 0.025$, degree of freedom, **df = n-2 = 498**, we obtain critical value, $\mathbf{t_{498,0.025} = +1.9647, -1.9647}$.

C) Test Statistic

We obtained test statistics, **$t = 7.044$**, the p-value, **$p = 6.25\times10^{-12}$**.

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        6.321495   0.315862  20.013  < 2e-16 ***
notification_count 0.026792   0.003804   7.044 6.25e-12 ***
```

Figure 3.3.8: Value of test statistic using RStudio

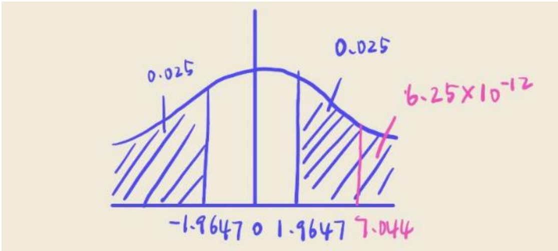

Figure 3.3.9: Test statistic location for regression test

D) Interpretation

Since $t = 7.044 > t_{498,0.025} = 1.9647$ and $p = 6.25\times10{-12} < \alpha/2 = 0.025$, the test statistic falls on the reject region at the 0.05 significant level. **The null hypothesis is rejected.** There is sufficient evidence to conclude that there is a linear relationship between the number of notifications and anxiety level, and notification count is associated with changes in anxiety levels.

E) Conclusion

In short, at $\alpha = 0.05$, we reject the null hypothesis and conclude that a significant linear relationship exists between the number of notifications and anxiety level. Thus, notification count is associated with changes in anxiety levels. The positive regression coefficient indicates anxiety level increases as number of notifications increase. These results highlight the psychological impact of digital engagement and suggest that managing notification exposure may benefit mental health.

F) Scatter plot with Regression Line

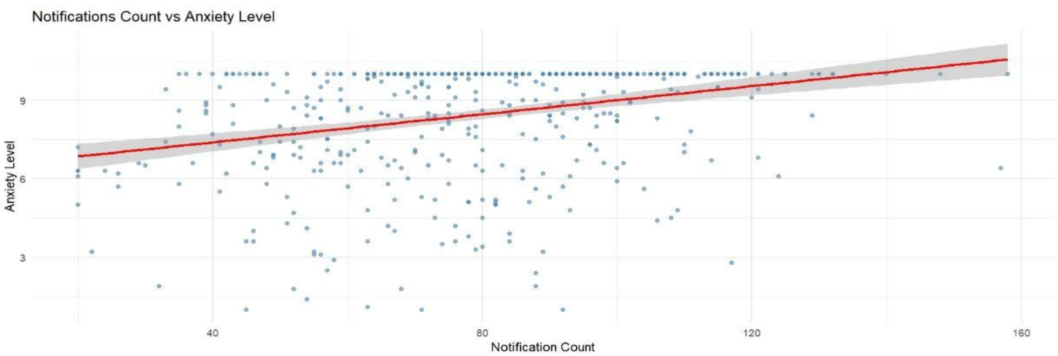

Figure 3.3.10: Scatter plot of notification count vs anxiety level with linear regression model using RStudio

3.4 Anova Test

The purpose of the ANOVA test we conducted is to determine whether there are **significant differences in mean focus scores between individuals grouped** by Low, Medium, and High levels of social media usage. Since the sample size in each group is the same, which is n equal to 166, we use a one-way ANOVA with equal sample sizes. We test this at a 0.05 significance level to evaluate the null hypothesis that all groups share the same average focus score.

1. **Hypothesis statement**

$$\mathbf{H_0:}\ \mu_1 = \mu_2 = \mu_3$$
$$\mathbf{H_1:}\ \text{at least one mean is different}$$

Where μ_1 represents the mean focus score for Low social media usage group, μ_2 represents the mean focus score for Medium social media usage group, μ_3 represents the mean focus score for High social media usage group.

2. **Test Statistic**

A significance level of **α = 0.05** is used to test the claim that the average focus scores are equal across the three groups. From the RStudio, we obtain mean of focus scores for Low, Medium and High social media usage group are 7.044, 6.984, 6.958. The variance between samples is 0.0019 and the variance within samples is 0.383. In addition, test statistic value, **F = 0.834**, degree of freedom for numerator is 2 and degree of freedom for denominator is 495. The probability value that we obtain is **P-value = 0.435.**

Table 3.4.1: Mean of Focus Group

Group	Low Focus	Medium Focus	High Focus
Mean	7.044	6.984	6.958

```
> mean(low_focus)
[1] 7.043976
> mean(med_focus)
[1] 6.984337
> mean(high_focus)
[1] 6.958434
```

Figure 3.4.1: Mean of Focus Scores for Low, Medium and High Social Media Usage Group using RStudio

```
> S2_mean
[1] 0.001924203
```

Figure 3.4.2: Variance between Samples using RStudio

```
> S2_pooled
[1] 0.382811
```

Figure 3.4.3: Variance within Samples using RStudio

```
> F_stat
[1] 0.8344004
```

Figure 3.4.4: Statistic Test Value using RStudio

```
> p_value
[1] 0.4347444
```

Figure 3.4.5: P-Value using RStudio

3. Interpretation

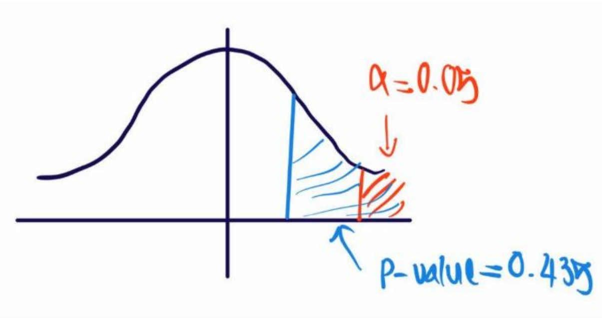

Figure 3.4. 6: Test statistic location for Anova test

Since $P-value > \alpha$ (0.435>0.05), we **fail to reject the null hypothesis H_0.** There is not enough evidence to suggest a significant difference in focus scores between Low, Medium, and High social media usage groups.

4. Conclusion

In conclusion, the one-way ANOVA test with equal sample sizes indicates that there is no significant difference in focus levels across social media usage groups. This suggests that social media time alone may not directly impact an individual’s ability to focus. Instead, attention may be influenced more by multitasking habits, digital boundaries, or personal media use strategies. For instance, someone who schedules dedicated time for social media might maintain better focus, whereas others who frequently switch between apps may experience reduced attention, regardless of their total usage time.

4.0 Conclusion and Discussion

This study explored how digital habits are related to mental health by applying different statistical tests such as two-sample hypothesis test, correlation, regression, and ANOVA. Throughout this project, we learned how to handle a real-world dataset from selecting suitable variables, performing pre-processing, and applying suitable statistical methods using RStudio. The project also improves our understanding on how to interpret statistical results and connect them to real-life behaviour patterns. By working with R environment, our confidence was strengthened in using data science tools to answer meaningful questions.

The results show that not all types of digital behaviour affect mental health in the same way. While screen time and social media usage showed no significant direct impact on mood score or focus level, this does not mean they are irrelevant. Instead, it suggests that how we use digital tools and in what situation may matter more than how long we use them.

On the other hand, two variables stood out with clear links to mental health outcomes. People who reported more sleep hours tended to have higher digital wellbeing scores, showing that good sleep supports mental balance and healthier digital use. Also, people who received more notifications tended to have higher levels of anxiety, suggesting that frequent interruptions can increase stress and affect emotional health.

These results highlight that mental health in the digital world is not only dependent by time spent on screen but also by the quality of attention and our ability to manage digital boundaries. Getting enough sleep and limiting unnecessary notifications may help protect mental health and improve our daily digital experiences.

In a nutshell, this project underscores a central message that digital habits can affect mental health, but the effects are multifaceted. Healthy routines, clear limits, and thoughtful use of technology are important steps toward better mental wellbeing in today’s connected world.

5.0 Appendices

1. Raw Dataset

Mental Health and Digital Behavior (2020–2024). (2025, May 13). Kaggle. https://www.kaggle.com/datasets/atharvasoundankar/mental-health-and-digital-behavior-20202024

2. Video Presentation Link

https://youtu.be/AbKrEBABzTA?si=ym0IVMNp7XVg2X7g