Mobile Phone Usage and Digital Wellbeing

Himanshu Jindal
MA21BTECH11007@iith.ac.in

Shubham Vishwakarma
MA21BTECH11020@iith.ac.in

Kaustubh Dandegaonkar
MA21BTECH11003@iith.ac.in

Roshan Kumar
MA21BTECH11013@iith.ac.in

Naman Chhibbar
MA21BTECH11011@iith.ac.in

Deepinder Singh
MA21BTECH11004@iith.ac.in

Devashish Chaudhari
MA21BTECH11005@iith.ac.in

Karthik Dhanavath
MA21BTECH11006@iith.ac.in

Abstract

Smartphones have become an indispensable part of our everyday lives, yet there is growing concern about the possible detrimental effects of excessive screen time. We surveyed students to investigate the consequences of smartphone use on physical health, mental health, academic performance, and social behavior. Our study concentrated on smartphone usage, which is currently the most popular gadget. We gathered information from students and analyzed it statistically.

1 Introduction

The impact of smartphone screen time on students' physical health, mental health, academic performance, and social behavior is investigated in our study. We performed a survey to gather information about students' smartphone screen time habits and perceived effects. The collected data is statistically analyzed to determine the association between smartphone screen time and its impacts. Our findings will provide vital insights into the possible hazards of excessive smartphone screen time and the significance of keeping a good lifestyle balance with other pursuits.

The survey consisted of the following questions:

1. What is your average screen time in a day?
2. How many times do you check your phone in a day?
3. How many notifications do you receive in a day?
4. What percentage of your screen time is productive?
5. What is the average time you study daily (outside college hours)?
6. How much do you usually study in one sitting? (Hours)
7. Which hostel are you staying in?
8. Which degree are you pursuing?
9. Which year are you currently in?
10. Gender of the student.
11. Do you wear spectacles?
12. Do you use the phone in class?
13. Do you attend classes?
14. How do you rate your focus?
15. How do you rate your happiness or mental well-being?

Based on the responses obtained to the questions mentioned above, we pre-processed the data and the data visualization is as follows:

2 Pre-processing and Visualisation of Data

The following steps were taken to pre-process the data:

1. We started by removing white spaces in columns containing string values.
2. Missing values (NaNs) were replaced by the median value in the case of numerical, and by modal in the case of categorical variables.
3. Categorical data like ”between n and n+1” is replaced by n+1/2 to make it numerical.
4. String data like ”hostel” was replaced by the corresponding numerical index to make calculations easier.

After pre-processing, the collected data was visualized as simple bar plots:

final.png

Figure 1. Visualizing the collected data

3 Analysis and Conclusions

After pre-processing the data, we had the data of 106 students from the survey, and we began to analyze the data. This involved the analysis of the one question involving a numerical variable:
“What is your average screen time in a day?”
The following results were obtained:

- Mean = 5.96 hours
- Median = 5 hours
- Mode = 5 hours

output3.1.png

Figure 2. Screen Time of Students (Box Plot)

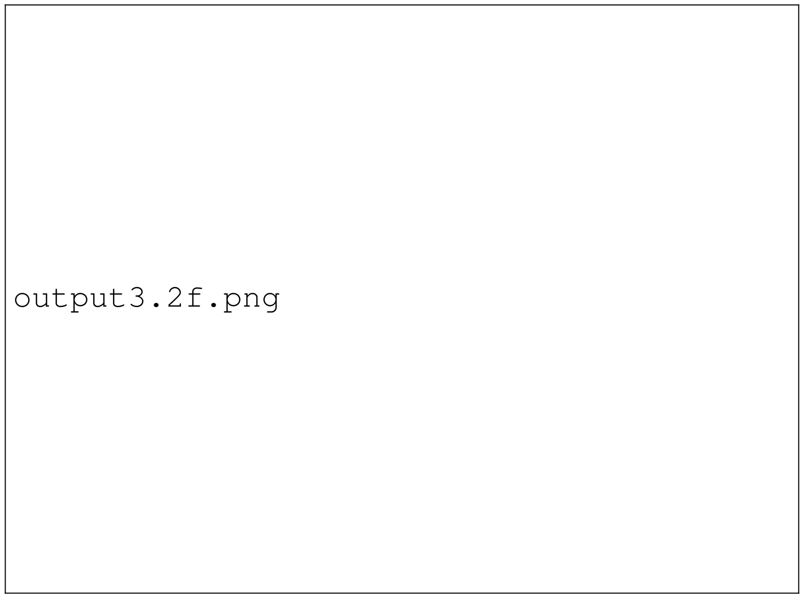

Figure 3. Screen Time of Students (distribution function) (Stem Plot)

After getting an idea of how the data stacks up, answers to some intriguing questions which involved the comparison of two questions were also given using bar charts:

Based on the Bar Plot: Figure 3, the following conclusions can be made:

1. Most of the students have their screen time lying between 2 to 6 hours

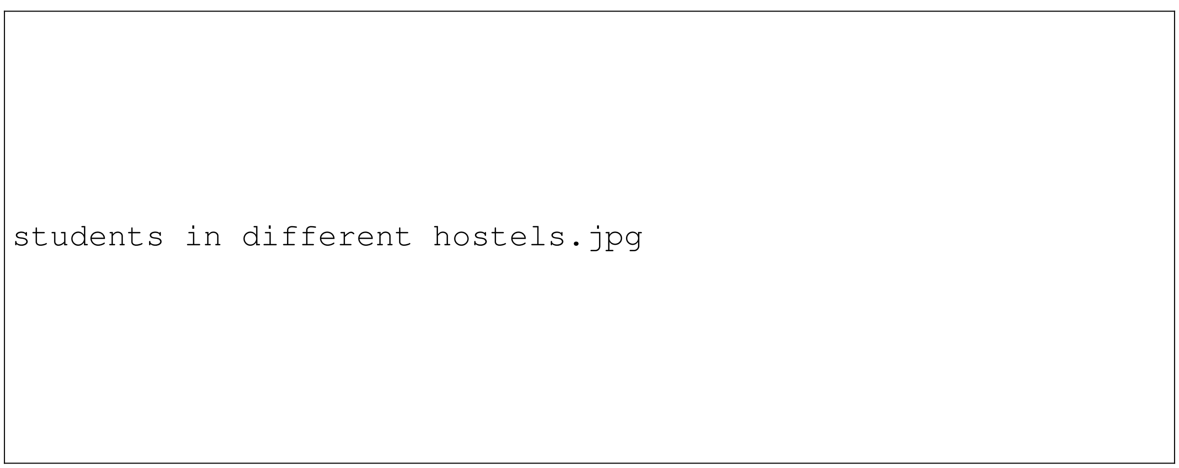

Figure 4. Students in different hostels

no_of_students wearing_specs.jpg

Figure 5. Students with spectacles

Based on the Segmented Bar Plot: Figure 5, the following conclusions can be made:

1. The proportion of UG students having spectacles is nearly same to that of PG+PhD.

Focus_vs_Screen_time.png

Figure 6. Focus index vs Average screen time

Based on the Bar Plot: Figure 6, the following conclusions can be made: Screen time of students is nearly same for each group of focus index

Happiness vs students (1).jpg

Figure 7. Happiness index vs No of students

Students in every year (2).png

Figure 8. Year of degree vs Students

Attend_class_vs_number_of_Students.png

Figure 9. No. of students who attend classes

Based on the Bar Plot: Figure 9, the following conclusions can be made:

A very small proportion of students don’t attend any classes.

4 Confidence Interval:

4.1 Confidence Interval for Mean Screen Time :

Let $\bar{x}$ = Sample Mean of target variable
Let S^2 = Sample Variance of target variable
let n= size of sample The Confidence interval is given by:

$$\bar{x} \pm E \tag{1}$$

also given by,

$$[\bar{x} - E, \bar{x} + E] \tag{2}$$

Where, E(Margin of Error) is given by,

$$E = t_{\alpha,n-1}(\frac{S}{\sqrt{n}}) \tag{3}$$

Width of the Confidence Interval is given by, $W = 2E$

4.1.1 Confidence interval for screen time of all students :

Let $\bar{x}$ = Sample Mean of screentime of students
Let S^2 = Sample Variance of screentime of students

Based on the sample selected, we have the following information-

$$\bar{x} = 5.96 \text{ hours} \tag{4}$$
$$S = 3.52 \tag{5}$$
$$n = 106 \tag{6}$$

We find the confidence interval of Screen time of students .
Confidence testing for the mean of the data gives us the following:

- The 95% confidence interval is given by: (5.28,6.63).
  The width of the interval is given by: 1.35
- The 99% confidence interval is given by: (5.06,6.85).
  The width of the interval is given by: 1.79

Increasing the confidence level, the width of the interval also increases.

4.1.2 Confidence interval for screen time of all male students :

Let $\bar{x}$ = Sample Mean of screentime of male students
Let S^2 = Sample Variance of screentime of male students

Based on the sample selected, we have the following information-

$$\bar{x} = 5.98 \text{ hours} \tag{7}$$
$$S = 3.6 \tag{8}$$
$$n = 86 \tag{9}$$
$$\tag{10}$$

We find the confidence interval of Screen time of students .
Confidence testing for the mean of the data gives us the following:

- The 95% confidence interval is given by: (5.21,6.79).
  The width of the interval is given by: 1.55
- The 99% confidence interval is given by: (4.95,7.01).
  The width of the interval is given by: 2.06

Increasing the confidence level, the width of the interval also increases.

4.1.3 Confidence interval for screen time of all female students :

Let $\bar{x}$ = Sample Mean of screentime of female students
Let S^2 = Sample Variance of screentime of female students

Based on the sample selected, we have the following information-

$$\bar{x} = 5.88 \text{ hours} \tag{11}$$
$$S = 3.21 \tag{12}$$
$$n = 20 \tag{13}$$
$$\tag{14}$$

We find the confidence interval of Screeen time of students .
Confidence testing for the mean of the data gives us the following:

- The 95% confidence interval is given by: (4.33,7.43).
  The width of the interval is given by: 3.1
- The 99% confidence interval is given by: (3.76,7.997).
  The width of the interval is given by: 4.237

Increasing the confidence level, the width of the interval also increases.

4.1.4 Confidence interval for screen time of all students who don't wear specs :

Let $\bar{x}$ = Sample Mean of screentime of students who don't wear specs
Let S^2 = Sample Variance of screentime of students who don't wear specs

Based on the sample selected, we have the following information-

$$\bar{x} = 5.03 \text{ hours} \tag{15}$$
$$S = 2.54 \tag{16}$$
$$n = 46 \tag{17}$$

We find the confidence interval of Screeen time of students .
Confidence testing for the mean of the data gives us the following:

- The 95% confidence interval is given by: (4.27,5.78).
  The width of the interval is given by: 1.51
- The 99% confidence interval is given by: (4.02,6.03).
  The width of the interval is given by: 2.01

Increasing the confidence level, the width of the interval also increases.

4.1.5 Confidence interval for screen time of all students who wear specs :

Let $\bar{x}$ = Sample Mean of screentime of students who wear specs
Let S^2 = Sample Variance of screentime of students who wear specs

Based on the sample selected, we have the following information-

$$\bar{x} = 6.68 \text{ hours} \tag{18}$$
$$S = 3.99 \tag{19}$$
$$n = 60 \tag{20}$$

We find the confidence interval of Screen time of students .
Confidence testing for the mean of the data gives us the following:

- The 95% confidence interval is given by: (5.65,7.71).
  The width of the interval is given by: 2.06
- The 99% confidence interval is given by: (5.31,8.05).
  The width of the interval is given by: 2.74

Increasing the confidence level, the width of the interval also increases.

4.2 Confidence Interval for Variance :

Let $\bar{x}$ be the sample mean of the target variable, S^2 be the sample variance of the target variable, and n be the sample size.

If $X_1, X_2, ..., X_n$ are normally distributed and $a = \chi^2_{(1-\frac{\alpha}{2}, n-1)}$ and $b = \chi^2_{(\frac{\alpha}{2}, n-1)}$, then a $(1-\alpha)\%$ confidence interval for the population variance σ^2 is given by:

$$\left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a} \right) \tag{21}$$

Similarly, a $(1-\alpha)\%$ confidence interval for the standard deviation σ is given by:

$$\left(\frac{\sqrt{(n-1)}S}{\sqrt{b}}, \frac{\sqrt{(n-1)}S}{\sqrt{a}} \right) \tag{22}$$

Note that the confidence interval formulas above are only applicable if the sample data is normally distributed.

4.2.1 Confidence interval for Variance of screen time of all students :

Let S^2 = Sample Variance of screentime of students
Based on the sample selected, we have the following information-

$$S = 3.52 \tag{23}$$
$$n = 106 \tag{24}$$
$$a = 78.54 \tag{25}$$
$$b = 135.25 \tag{26}$$

We find the confidence interval of Variance of Screen time of students .
Confidence testing for the Variance of the data gives us the following:

- The 95% confidence interval is given by: (9.62,16.57).
  The width of the interval is given by: 6.95

4.2.2 Confidence interval for Variance of screen time of all male students :

Let S^2 = Sample Variance of screentime of male students
Based on the sample selected, we have the following information-

$$S = 3.6 \tag{27}$$
$$n = 86 \tag{28}$$
$$a = 61.39 \tag{29}$$
$$b = 112.39 \tag{30}$$

We find the confidence interval of Variance of Screen time of students .
Confidence testing for the Variance of the data gives us the following:

- The 95% confidence interval is given by: (9.80,17.94).
  The width of the interval is given by: 8.14

4.2.3 Confidence interval for Variance of screen time of all female students :

Let S^2 = Sample Variance of screentime of female students
Based on the sample selected, we have the following information-

$$S = 3.21 \tag{31}$$
$$n = 20 \tag{32}$$
$$a = 8.91 \tag{33}$$
$$b = 32.85 \tag{34}$$

We find the confidence interval of Variance of Screen time of students .
Confidence testing for the Variance of the data gives us the following:

- The 95% confidence interval is given by: (5.96,21.98).
  The width of the interval is given by: 16.02

4.2.4 Confidence interval for Variance of screen time of all students who don't wear specs :

Let S^2 = Sample Variance of screentime of students who don't wear specs
Based on the sample selected, we have the following information-

$$S = 2.54 \tag{35}$$
$$n = 46 \tag{36}$$
$$a = 28.37 \tag{37}$$
$$b = 65.41 \tag{38}$$

We find the confidence interval of the Variance of Screen time of students.
Confidence testing for the Variance of the data gives us the following:

- The 95% confidence interval is given by: (4.44,10.23).
  The width of the interval is given by: 5.80

4.2.5 Confidence interval for Variance of screen time of all students who wear specs :

Let S^2 = Sample Variance of screentime of students who wear specs
Based on the sample selected, we have the following information-

$$S = 3.99 \tag{39}$$
$$n = 60 \tag{40}$$
$$a = 39.66 \tag{41}$$
$$b = 82.12 \tag{42}$$

We find the confidence interval of the Variance of Screen time of students who wear specs.
Confidence testing for the Variance of the data gives us the following:

- The 95% confidence interval is given by: (11.44,23.68).
  The width of the interval is given by: 12.24

4.3 Confidence Interval for Proportions :

Let $\hat{p}$ = Sample Proportion
Let n= size of sample size

For large Random samples, a 100% CI for the Population proportion p is given by :

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \tag{43}$$

4.3.1 Confidence interval for Proportion of students who use phone between 4 to 6 hours :

$$\hat{p} = 0.47 \tag{44}$$
$$n = 106 \tag{45}$$

We find the confidence interval of the Proportion of students who use phones between 4 to 6 hours.
Confidence testing for the Proportion of the data gives us the following:

- The 95% confidence interval is given by: (37.67,56.67).
  The width of the interval is given by: 20.00

5 Hypothesis Testing:

5.1 Case-1: Comparing the screen time for Undergraduates and Postgraduates

We compare screen time of Undergraduates and Postgraduates. We assume our null hypothesis to be that undergraduates have more screen time than postgraduates. Let $\alpha = 0.05$

Let $\bar{x_1}$ = Sample Mean of screen-time of Undergraduates
Let $\bar{x_2}$ = Sample Mean of screen-time of Postgraduates
Let S_1^2 = Sample Variance of screen-time of Undergraduates
Let S_2^2 = Sample Variance of screen-time of Postgraduates

For Hypothesis Testing, we make the following statements-

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ and } H_a : \mu_1 - \mu_2 < 0$$

We have the following information-

$$\bar{x_1} = 5.479 \text{ hours} \tag{46}$$
$$\bar{x_2} = 7.038 \text{ hours} \tag{47}$$
$$S_1^2 = 7.388 \tag{48}$$
$$S_2^2 = 22.279 \tag{49}$$
$$n_1 = 73 \tag{50}$$
$$n_2 = 33 \tag{51}$$

-

Since $\dfrac{S_1^2}{S_2^2} = 0.3316 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 11.9716 \tag{52}$$

The test statistic is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = -2.15 \tag{53}$$

Using the rejection region approach, we reject H_0 if $t \leq -t_{0.05,104}$, where $t_{0.05,104} = 1.659$. Because the observed value of $t = -2.15$ is less than -1.659, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, the postgraduates have more screentime in one go than the undergraduates on average.

Error Analysis: We might have committed a Type 1 error if the null hypothesis is actually true, and we have incorrectly concluded that undergraduates have less screen time than postgraduates .Therefore, there is a 5% chance that we have falsely rejected the null hypothesis and concluded that postgraduates have more screen time than undergraduates when in reality, there is no difference in their screen time.

5.2 Case-2: Comparing the screen time of people with specs and people with no specs.

We compare screen times of people who wear specs and people who do not. We assume our null hypothesis that people who wear specs have less screen time compared to people who don't wear specs.

Let $\bar{x_1}$ = Sample Mean of screen-time of people with no specs
Let $\bar{x_2}$ = Sample Mean of screen-time of people with specs
Let S_1^2 = Sample Variance of screen-time of people with no specs
Let S_2^2 = Sample Variance of screen-time of people with specs

For Hypothesis Testing, we make the following statements-

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ and } H_a : \mu_1 - \mu_2 < 0$$

$$\bar{x_1} = 5.027 \text{ hours} \tag{54}$$
$$\bar{x_2} = 6.683 \text{ hours} \tag{55}$$
$$S_1^2 = 6.453 \tag{56}$$
$$S_2^2 = 15.934 \tag{57}$$
$$n_1 = 60 \tag{58}$$
$$n_2 = 46 \tag{59}$$

Since $\dfrac{S_1^2}{S_2^2} = 0.40 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 10.5556 \tag{60}$$

The test statistic is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -2.601 \tag{61}$$

Using the rejection region approach, we reject H_0 if $t \leq -t_{0.05,116}$, where $t_{0.05,116} = 1.659$. Because the observed value of $t = -2.601$ is less than -1.659, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those with specs have more screen time on average than those with no specs

Error Analysis : Since the hypothesis testing resulted in rejecting the null hypothesis, we did not commit a Type 2 error. However, we may have committed a Type 1 error if the null hypothesis is actually true, and we have wrongly concluded that those with specs have more screen time. The probability of making a Type 1 error is represented by the level of significance, which in this case is 0.05.

5.3 Case-3: Comparing the productive screen time of Undergraduate students to that of Postgraduate students.

We compare the productive screen time of Undergraduate students and Postgraduate students! We assume our null hypothesis that Undergraduate students have more productive screen time compared to Postgraduate students .

Let $\bar{x_1}$ = Sample Mean of screen time of UG students
Let $\bar{x_2}$ =Sample Mean of screen time of PG students
Let S_1^2 = Sample Variance of screen time of UG students
Let S_2^2 = Sample Variance of screen time of PG students

For Hypothesis Testing we make the following statements-

$$H_0 : \mu_2 - \mu_1 \leq 0 \text{ and } H_a : \mu_2 - \mu_1 > 0$$

$$\bar{x_1} = 27.89 \text{ hours} \tag{62}$$
$$\bar{x_2} = 35.80 \text{ hours} \tag{63}$$
$$S_1^2 = 522.002 \tag{64}$$
$$S_2^2 = 887.061 \tag{65}$$
$$n_1 = 73 \tag{66}$$
$$n_2 = 33 \tag{67}$$

Since $\frac{S_1^2}{S_2^2} = 0.588 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 634.32 \tag{68}$$

The test statistic is then given by:

$$t = \frac{\bar{x_2} - \bar{x_1} - 0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = 1.513 \tag{69}$$

Using the rejection region approach, we reject H_0 if $t \geq t_{0.05,104}$, where $t_{0.05,104} = 1.658$. Because the observed value of $t = 1.513$ is less than 1.659, we fail to reject the null hypothesis, and thus, do not have enough evidence to say Undergraduate students have more productive screen-time compared to Postgraduate students .

Error Analysis : Type 1 error: Rejecting the null hypothesis that Undergraduate students have more productive screen-time compared to Postgraduate students when it is actually false.

Type 2 error: Failing to reject the null hypothesis that Undergraduate students have more productive screen-time compared to Postgraduate students when it is actually true.

5.4 Case-4: Comparing the screen time of people who attend all the classes and those who don't attend all the classes.

We compare screen time of people who attend all the classes and those who don't attend all the classes! We assume our null hypothesis that people who attend all the classes have higher screen time than those who don't.

Let $\bar{x_1}$ = Sample Mean of screen time of people who attend all the classes
Let $\bar{x_2}$ = Sample Mean of screen time of people who don't attend all the classes
Let S_1^2 = Sample Variance of screen time of people who attend all the classes
Let S_2^2 = Sample Variance of screen time of people who don't attend all the classes

For Hypothesis Testing we make the following statements-

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ and } H_a : \mu_1 - \mu_2 < 0$$

$$\bar{x_1} = 5.33 \text{ hours} \tag{70}$$
$$\bar{x_2} = 6.67 \text{ hours} \tag{71}$$
$$S_1^2 = 2.90 \tag{72}$$
$$S_2^2 = 3.95 \tag{73}$$
$$n_1 = 56 \tag{74}$$
$$n_2 = 50 \tag{75}$$

Since $\frac{S_1^2}{S_2^2} = 0.734 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 11.83 \tag{76}$$

The test statistic is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = -1.995 \tag{77}$$

Using the rejection region approach, we reject H_0 if $t \leq -t_{0.05,104}$, where $t_{0.05,104} = 1.658$. Because the observed value of $t = -1.995$ is lesser than -1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those who attend all the classes have lower screen time than those who don't attend all the classes.

Error Analysis : The calculated t-value of -1.995 is less than the critical value set at -1.658, so the null hypothesis is rejected.However, there is a risk of making a type I error if the null hypothesis is true, and a risk of making a type II error if it is false. To reduce the risk of a type II error, increasing the sample size or lowering the significance level can be considered.

5.5 Case-5: Comparing the screen time of people who are sad and those who are happy.

We compare screen time of people who are higher on the happiness metric and those who are lower on the happiness metric! We assume our null hypothesis that people who are higher on happiness metric (≥ 3) have higher screen time than those who are lower on happiness metric (≤ 2).

Let $\bar{x_1}$ = Sample Mean of screen time of people who have happiness (≥ 3)
Let $\bar{x_2}$ = Sample Mean of screen time of people who have happiness (≤ 2)
Let S_1^2 = Sample Variance of screen time of people who have happiness (≥ 3)
Let S_2^2 = Sample Mean of screen time of people who have happiness (≤ 2)

For Hypothesis Testing we make the following statements-

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ and } H_a : \mu_1 - \mu_2 < 0$$

$$\bar{x_1} = 5.457 \text{ hours} \tag{78}$$
$$\bar{x_2} = 7.897 \text{ hours} \tag{79}$$
$$S_1^2 = 3.2928 \tag{80}$$
$$S_2^2 = 3.6139 \tag{81}$$
$$n_1 = 84 \tag{82}$$
$$n_2 = 22 \tag{83}$$

Since $\frac{S_1^2}{S_2^2} = 0.911 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 3.357 \tag{84}$$

The test statistic is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = -3.03 \tag{85}$$

Using the rejection region approach, we reject H_0 if $t \leq -t_{0.05,104}$, where $t_{0.05,104} = 1.658$. Because the observed value of $t = -3.03$ is lesser than -1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those who are higher on the happiness metric (≥ 3) have lower screen time compared to people at lower happiness (≤ 2).

Type 1 error: Rejecting the null hypothesis that people who are higher on the happiness metric ($\geq$ 3) have higher screen time than those who are lower on the happiness metric ($\leq$ 2), when it is actually true.

Type 2 error: Failing to reject the null hypothesis that people who are higher on the happiness metric ($\geq$ 3) have higher screen time than those who are lower on the happiness metric ($\leq$ 2), when it is actually false.

5.6 Case-6: Comparing the screen time of students with phone checking habits vs other students

We compare screen time of people with frequent phone checking habits (Phone-checks/day ≥ 150) with others (Phone-check/day < 150)! We assume our null hypothesis that people who check phone ≥ 150 times have lower screen time than those who check phone < 150 times.

Let $\bar{x_1}$ = Sample Mean of screen time of people who check phone (≥ 150) times
Let $\bar{x_2}$ = Sample Mean of screen time of people who check phone (< 150) times
Let S_1^2 = Sample Variance of screen time of people who check phone (≥ 150) times
Let S_2^2 = Sample Mean of screen time of people who check phone (< 150) times

For Hypothesis Testing we make the following statements-

$$H_0 : \mu_1 - \mu_2 \leq 0 \text{ and } H_a : \mu_1 - \mu_2 > 0$$

$$\bar{x_1} = 8.75 \text{ hours} \tag{86}$$

$$\bar{x_2} = 5.674 \text{ hours} \tag{87}$$

$$S_1^2 = 9.250 \tag{88}$$

$$S_2^2 = 32.8099 \tag{89}$$

$$n_1 = 96 \tag{90}$$

$$n_2 = 10 \tag{91}$$

Since $\frac{S_1^2}{S_2^2} = 0.28 > 0.25$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 11.288 \tag{92}$$

The test statistic is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 2.755 \tag{93}$$

Using the p-value approach, we reject H_0 if P($t \geq t_{calculated}$), where $t_{calculated} = 2.755$, is less than 0.05. P($t \geq t_{calculated}$) = 0.003465 $<$ 0.05 we have enough statistical evidence to reject the null hypothesis, and thus, we can say, people who check their phones more frequently have higher screen time than people who don't. Type 1 error: Rejecting the null hypothesis when it is actually true. In this case, it would mean concluding that people who check their phone 150 times or more have higher screen time than those who check their phone less than 150 times, when in fact, the null hypothesis is true.

Type 2 error: Failing to reject the null hypothesis when it is actually false. In this case, it would mean failing to conclude that people who check their phone 150 times or more have higher screen time than those who check their phone less than 150 times, when in fact, the alternative hypothesis is true.

6 Contributions:

- **Devashish Chaudhari**
  - Ideation
  - Data Collection
  - Data Visualisation
  - Proofreading
- **Himanshu Jindal**
  - Ideation
  - Choice of Target Variable
  - Creation of Google Form
  - Data Collection
  - Confidence Interval (Computation and Latex Writing)
  - Hypothesis Testing(One Case)
  - Report Creation in LaTeX
- **Shubham Vishwakarma**
  - Ideation
  - Data collection
  - Hypothesis testing
  - Supervised and Delegation of duties throughout the project.
  - Checking values, and computations and making sure consistency of results.
- **Kaustubh Dandegaonkar**
  - Ideation
  - Data Collection
  - Visualizing Categorical Variables
  - Asisted Making Reports in LaTeX
  - Assisted in hypothesis testing
  - Computation of Confidence Interval
- **Roshan Kumar**

  - Ideation
  - Data collection
  - Ideation of Confidence interval
  - Ideation of Hypothesis testing(2 cases computation)
  - Helped in LaTeXcreation
- **Deepinder Singh**
  - Ideation
  - Hypothesis testing(2 cases testing and latex)
  - Data visualisation
  - Data collection
  - Data analysis
- **Naman Chhibbar**
  - Ideation
  - Data Pre-Processing data analysis
  - Data visualisation
  - Contributed to google form Ideation
- **Karthik Dhanavath**
  - Ideation