Mobile Phone Usage and Digital Wellbeing

MA4240 - Applied Statistics

Himanshu Jindal Naman Chhibbar Kaustubh Dandegaonkar
Deepinder Singh Shubham Vishwakarma Roshan Kumar
Devashish Chaudhari Karthik Dhanavath

May 3, 2023

Outline

Introduction

The impact of smartphone screen time on students' physical health, mental health, and academic performance 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 balancing a good lifestyle with other pursuits.

Points of interest

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?

Pre-Processing 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 with the median value in the case of numerical, and with modal values in the case of categorical variables.
3. Categorical data like "between n and n+1" is replaced with n+(1/2) to make it numerical.
4. String data like "hostel" is replaced with the corresponding numerical index to make calculations easier.

Data Visualization

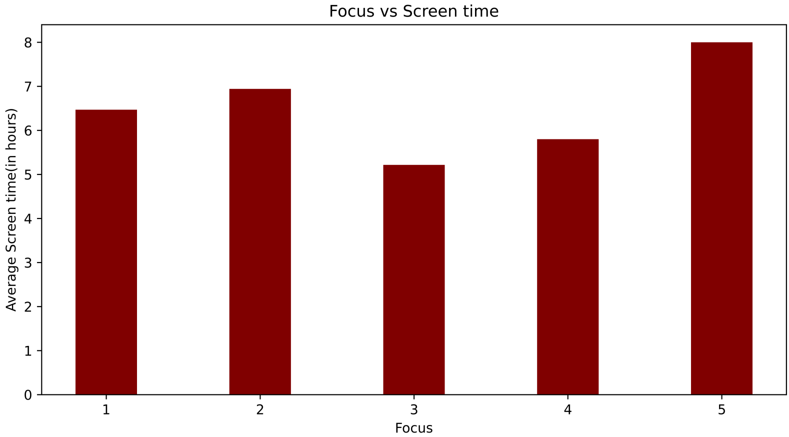
Analyzing the Uni-variate Numerical dataset

Table: What is your average screen time in a day?

Count	106(non-null)
Mean	5.96
Median	5
Mode	5
std	3.52
25%	4
50%	5
75%	6.75
95% Confidence Interval	(5.28,6.63)
99% Confidence Interval	(5.06,6.85)

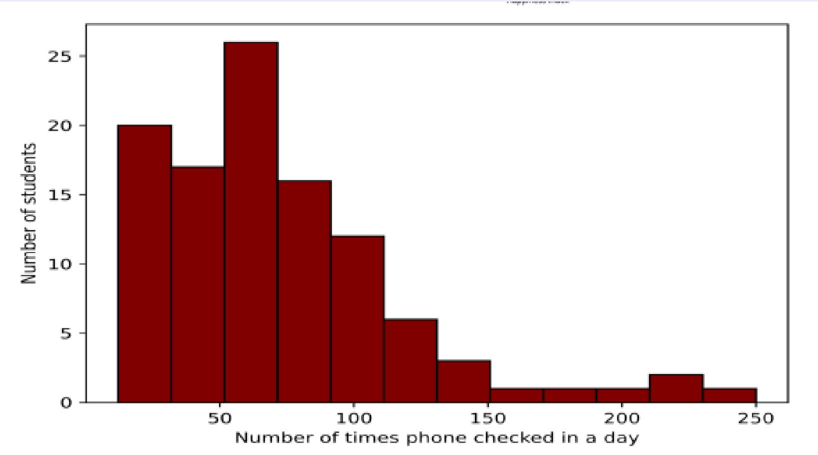
Focus v/s screen-time

Figure: Focus[1 to 5] v/s screen-time



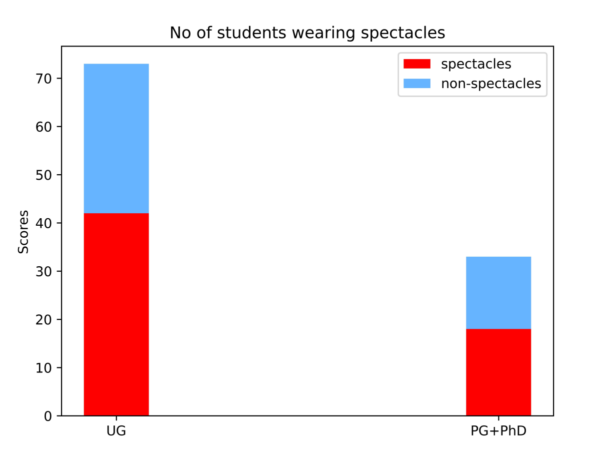
Number of times phone checked in a day

Figure: Distribution of number of times phone checked by students



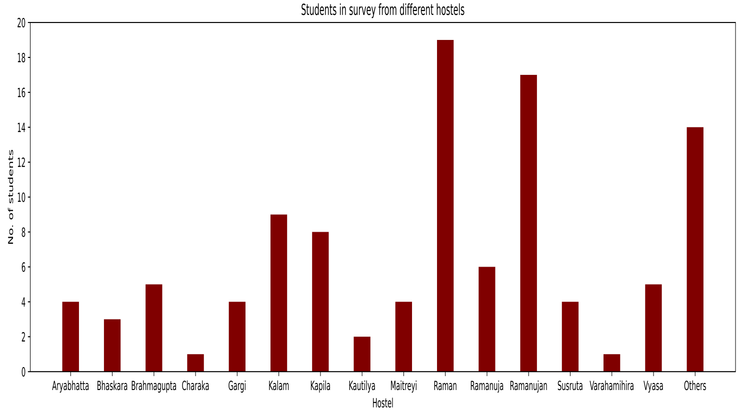
Data Visualization with Segmented Bar plots

Figure: Categorical variables in a segmented bar plot



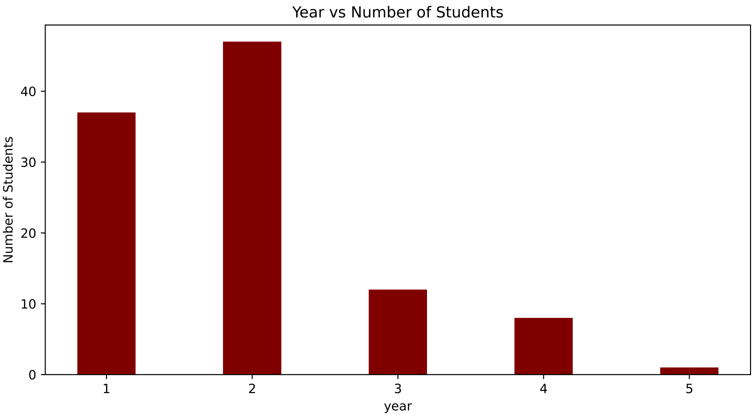
Students in survey from different hostels

Figure: Count of students from different hostel



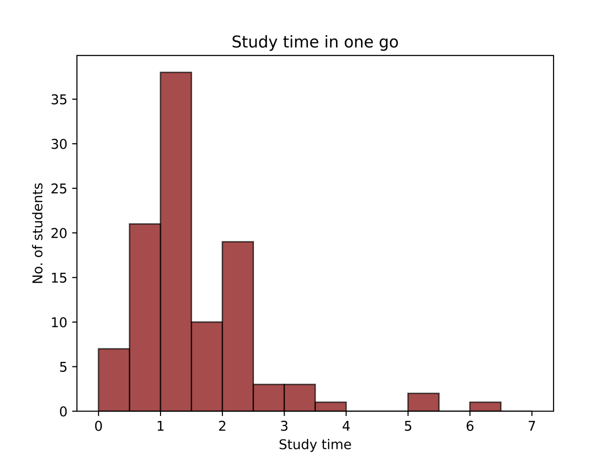
Students from different years

Figure: Students from different years



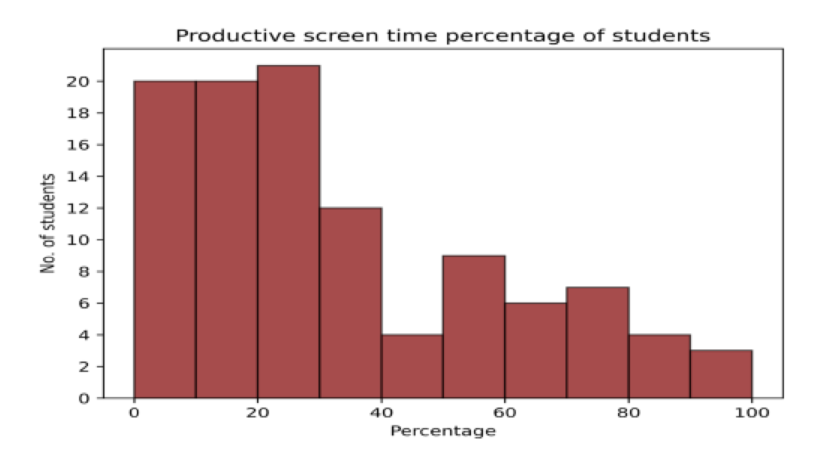
Study-time in one go v/s no. of students

Figure: study-time in one go[categorial groups] v/s no. of students in that category



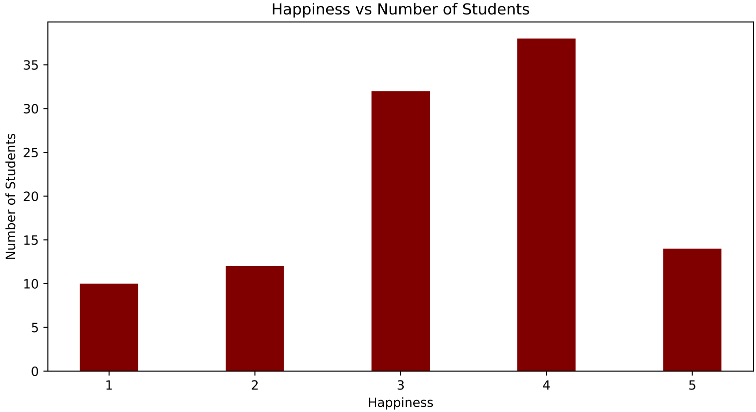
Productive screen time (percentage) distribution

Figure: Productive screen-time [categorial groups size = 10] v/s number of students in group



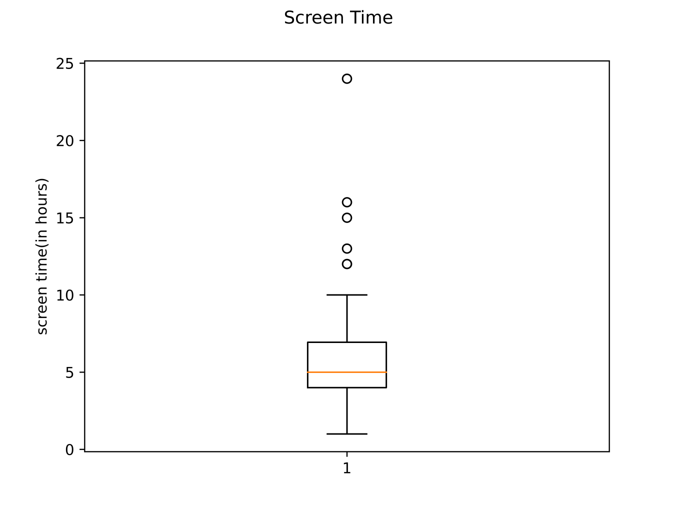
Happiness index v/s students

Figure: Happiness index[1 to 5] v/s students



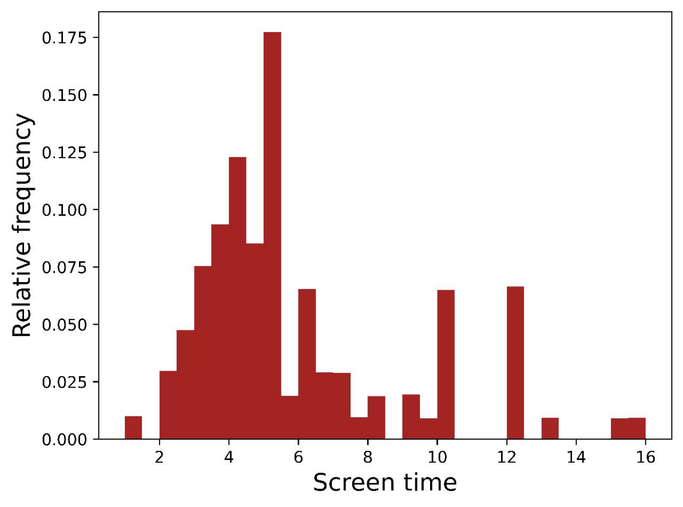
Box-plot for the screen-time of students

Figure: box-plot of screen-time of students



Distribution of screen-time among the students

Figure: Distribution of students v/s screen time



Inference

1. 70.75% people have screen-time between 2 to 6 hours.
2. While 28.3% people have screen-time more than 6 hours.
3. And a very few percentage of people, i.e., 0.94% have screen-time less than 2 hours.

Confidence Interval

Formula

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:

$$[\bar{x} - E, \bar{x} + E] \tag{1}$$

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

$$E = t_{\alpha, n-1}(\frac{S}{\sqrt{n}}) \tag{2}$$

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

Confidence Interval

Case 1 : People with specs

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

$$\bar{x} = 6.68 \text{ hours} \tag{3}$$

$$S = 3.99 \tag{4}$$

$$n = 60 \tag{5}$$

Where, E(Margin of Error) is given by,
For 95% Confidence Interval

$$\alpha = 0.05$$

$$E = t_{\alpha,n-1}\left(\frac{S}{\sqrt{n}}\right) = t_{0.05,59}\left(\frac{3.99}{\sqrt{60}}\right) = 2.0010\left(\frac{3.99}{\sqrt{60}}\right) = 1.030 \tag{6}$$

Width of the Confidence Interval is given by, $W = 2E = 2.06$.

Confidence Interval

Case 1 : People with specs

95% Confidence Interval

$$(\bar{x} - E, \bar{x} + E) = (5.65, 7.71) \tag{7}$$

Similarly, The 99% confidence interval is given by:

$$(5.31, 8.05) \tag{8}$$

The width of the interval is given by: 2.74

Therefore we can say with 95% confidence that people who wear specs have average screen time between 5.65 and 7.71 hours

And with 99% confidence that people who wear specs have average screen time between 5.31 and 8.05 hours

Confidence Interval

Case 2 : People without specs

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

$$\bar{x} = 5.03 \text{ hours} \tag{9}$$

$$S = 2.54 \tag{10}$$

$$n = 46 \tag{11}$$

Where, E(Margin of Error) is given by,
For 95% Confidence Interval

$$\alpha = 0.05$$

$$E = t_{\alpha,n-1}\left(\frac{S}{\sqrt{n}}\right) = t_{0.05,45}\left(\frac{3.99}{\sqrt{60}}\right) = 2.0141\left(\frac{5.03}{\sqrt{46}}\right) = 0.75 \tag{12}$$

Width of the Confidence Interval is given by, $W = 2E = 1.51$.

Confidence Interval

Case 2 : People without specs

95% Confidence Interval

$$(\bar{x} - E, \bar{x} + E) = (4.27, 5.78) \tag{13}$$

Similarly, The 99% confidence interval is given by:

$$(4.02, 6.03) \tag{14}$$

The width of the interval is given by: 2.01

Therefore we can say with 95% confidence that people who wear specs have average screen time between 4.27 and 5.78 hours

And with 99% confidence that people who wear specs have average screen time between 4.02 and 6.03 hours

Confidence Interval

Confidence Interval of Variance:

Let $\bar{x}$ = Sample Mean of target variable
Let S^2 = Sample Variance of target variable
let n= size of same
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 $(1-\alpha)100\%$ confidence interval for the population variance σ^2 is given by:

$$\left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right) \tag{15}$$

Confidence Interval

Case 3 : Confidence interval of Variance of Screen-time

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

$$S = 3.52 \tag{16}$$

$$n = 106 \tag{17}$$

$$a = \chi^2_{(1-\frac{\alpha}{2}, n-1)} = \chi^2_{(1-\frac{0.05}{2}, 105)} = 78.54 \tag{18}$$

$$b = \chi^2_{(\frac{\alpha}{2}, n-1)} = \chi^2_{(\frac{0.05}{2}, 105)} = 135.25 \tag{19}$$

Confidence Interval

Case 3 : Confidence interval of Variance of Screen-time

95% Confidence Interval

$$\left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right) \tag{20}$$

$$\left(\frac{(105)12.3904}{135.25}, \frac{(105)12.3904}{78.54}\right) \tag{21}$$

$$(9.62, 16.57) \tag{22}$$

Therefore we can say with 95% confidence that variance of average screen time of students lies between 9.62 and 16.57.

Confidence Interval

Confidence Interval of Proportion:

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 :

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \tag{23}$$

Confidence Interval

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

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

$$\hat{p} = 0.47 \tag{24}$$

$$n = 106 \tag{25}$$

Confidence Interval

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

95% Confidence Interval :

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \tag{26}$$

$$\left(0.47 - z_{0.025}\sqrt{\frac{0.47(0.53)}{106}}, 0.47 + z_{0.025}\sqrt{\frac{0.47(0.53)}{106}}\right) \tag{27}$$

$$(0.3767, 0.5667) \tag{28}$$

Therefore we can say with 95% confidence that the proportion of people using phone for 4 to 6 hours is between 0.3767 and 0.5667.

Hypothesis Testings

Case 1: Comparing the screen time of people who attend all classes and those who don't attend all the classes

Null hypothesis : People who attend all the classes have higher screen time compared to people who don't.

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

$$\bar{x_1} = 5.33 \text{ hours} \qquad \bar{x_2} = 6.67 \text{ hours}$$
$$S_1^2 = 2.90 \qquad S_2^2 = 3.95$$
$$n_1 = 56 \qquad n_2 = 50$$

Since $\dfrac{S_1^2}{S_2^2} = 0.734 > 0.25$, we can assume the population variances are nearly equal.

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$, and the pooled variance will be:

Hypothesis Testings

Case 1: Continued

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 11.8 \tag{29}$$

The test statistic t is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -1.995 \tag{30}$$

$$\tag{31}$$

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 less than -1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those who regularly go to class have less screen time on average than those who don't go to classes regularly.

Hypothesis Testings

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

Null hypothesis :people who wear specs have less screen time compared to people who don't wear specs.

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

$$\bar{x_1} = 5.027 \text{ hours} \quad \bar{x_2} = 6.683 \text{ hours}$$
$$S_1^2 = 6.453 \quad S_2^2 = 15.934$$
$$n_1 = 60 \quad n_2 = 46$$

Since $\dfrac{S_1^2}{S_2^2} = 0.40 > 0.25$, we can assume the population variances are nearly equal.

Hypothesis Testings

Case 2: Continued

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$, and the pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 10.5556 \tag{32}$$

The test statistic t is then given by:

$$t = -2.601 \tag{33}$$

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 $t = -2.601$ 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

Hypothesis Testings

Case 3: Comparing variance in screen time of UG and PG + PHD

Null hypothesis: The Variance in screen-time of UG is higher than Variance in screen-time of PG + PHD.

For Hypothesis Testing, we make the following statements -

$$H_0 : \sigma_1^2 \geq \sigma_2^2 \text{ and } H_a : \sigma_1^2 < \sigma_2^2$$
$$s_1^2 = 7.388 \quad s_2^2 = 22.279$$
$$n_1 = 73 \quad n_2 = 33$$
$$df_2 = n_2 - 1 \; df_1 = n_1 - 1$$

Hypothesis Testings

Case 3: Continued

$$\text{Test statistic: } F = \frac{s_1^2}{s_2^2}$$

$$F = 0.331$$

$$F_{1-\alpha, df_1, df_2} = 0.623$$

Rejection region: For a level α with degrees of freedom $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$, reject H_0 if $F \leq F_{1-\alpha, df_1, df_2}$.

$0.331 \leq 0.623$

So we are able to reject the null hypothesis. So we can say the variance in screen-time of PG+PHD is greater than variance in screen-time of UG.

Hypothesis Testings

Case 4: Comparing the screen time of people who are unhappy and those who are happy

Null hypothesis : people who are higher on happiness metric (≥ 3) have higher screen time than those who are lower on happiness metric (≤ 2)).
$H_0 : \mu_1 - \mu_2 \geq 0$ and $H_a : \mu_1 - \mu_2 < 0$. Now,

$$\begin{aligned} \bar{x_1} &= 5.457 \text{ hours} \quad & \bar{x_2} &= 7.897 \text{ hours} \\ S_1^2 &= 3.2928 \quad & S_2^2 &= 3.6139 \\ n_1 &= 84 \quad & n_2 &= 22 \end{aligned}$$

Since $\dfrac{S_1^2}{S_2^2} = 0.911 > 0.25$, we can assume the population variances are nearly equal.

Hypothesis Testings

Case 4: Continued

The degrees of freedom, $df = n_1 + n_2 - 2 = 104$, and the pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 3.357 \tag{34}$$

The test statistic t is then given by:

$$t = \frac{\bar{x_1} - \bar{x_2} - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -3.03 \tag{35}$$

Using the rejection region approach, we reject H_0 if $t \leq -t_{0.05,116}$, where $t_{0.05,116} = 1.658$. 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).

Takeaways from the analysis

From Confidence Intervals

1. Screentime vs Spectacles
   Mom was right, high screen time does make your eyesight weak. This was later confirmed by the Hypothesis testing.
2. Variance of Screentime data
   This gives us an idea of how much the data deviates from the mean.
3. Proportion of People who have Screentime between 4-6 hours
   Tells us that a sizable amount of the people use phone 4-6 hours daily.

Takeaways from the analysis

From Hypothesis Testings

1. Screentime and attendance
   Students who are regular in class have lower screen time than irregular students.
2. Screentime and Spectacles
   Students with spectacles were found to have higher average screen time than others.
3. Variance of Screentime data
   This gives us an idea of how much the data deviates from the mean. The variance of screen-time of PG/PHD is higher than that of UG.
4. Happiness and screen-time
   Students higher on happiness metric have lower average screen time than students on lower happiness metric.

THANK YOU

MA4240 - Applied Statistics

Himanshu Jindal Naman Chhibbar Kaustubh Dandegaonkar
Deepinder Singh Shubham Vishwakarma Roshan Kumar
Devashish Chaudhari Karthik Dhanavath

May 3, 2023