

Intricacies of the Study Environment at IIT Hyderabad

Prakhar Patni

MA20BTECH11014@iith.ac.in

K N Vardhan

MA20BTECH11006@iith.ac.in

Varunaditya Singhal

MA20BTECH11021@iith.ac.in

Tanmay Goyal

AI20BTECH11021@iith.ac.in

Tanay Yadav

AI20BTECH11026@iith.ac.in

Sujal

AI20BTECH11020@iith.ac.in

Abstract

As we come out of the pandemic, studies in the classroom are almost a bygone methodology, and the introduction of online lectures and classes have seen newer and newer ways of studying, and newer ideas to boost efficiency and productivity. Binging lectures like never before, and newer techniques of retaining and understanding information continue to evolve on a daily basis. Through this project, we aim to derive conclusions on various study environment related habits and choices that exist among students at IITH.

1. Introduction

This project is based on the behavioral patterns of students studying at IITH. The vision was to see the changes that happened in the academic structure of our college after the strike of the pandemic, as well as how the students are dealing and coping with it. No. of study hours, preference of place, and material of studies were a few of the most common queries we asked. We use Statistics to deduce conclusions from the given data, assuming that the data is a random population sample.

The survey consisted of the following questions:

1. Where do you like to study?
2. Gender of the student?
3. Do you prefer to study alone?
4. How many hours do you study in one go?

5. Do you prefer snacks while studying?
6. In which block you stay?
7. Do you prefer to study on bed or study table?
8. Which program are you enrolled in?
9. Which department are you in?
10. Do you prefer live lectures or recording?
11. Do you prefer to study with lecture recording or lecture slides?

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

2. Pre-processing and Visualisation of Data

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

1. We began by removing white-spaces and dropping columns not required for further analysis.
2. Since the names of the Hostel Blocks were to be an important variable for our comparisons, all entries not having the Hostel Block name were dropped.
3. Any existing NaNs were replaced with the modal values of the specific column, since we do not have any model to predict the unentered values.
4. The *Hostel Block* names were replaced with their letters and the *Department* names with their departmental codes.
5. Finally, to make the data more interpretable, we assumed that any person studying between n and $n + 1$ hours would be studying $n + \frac{1}{2}$ hours on average.

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

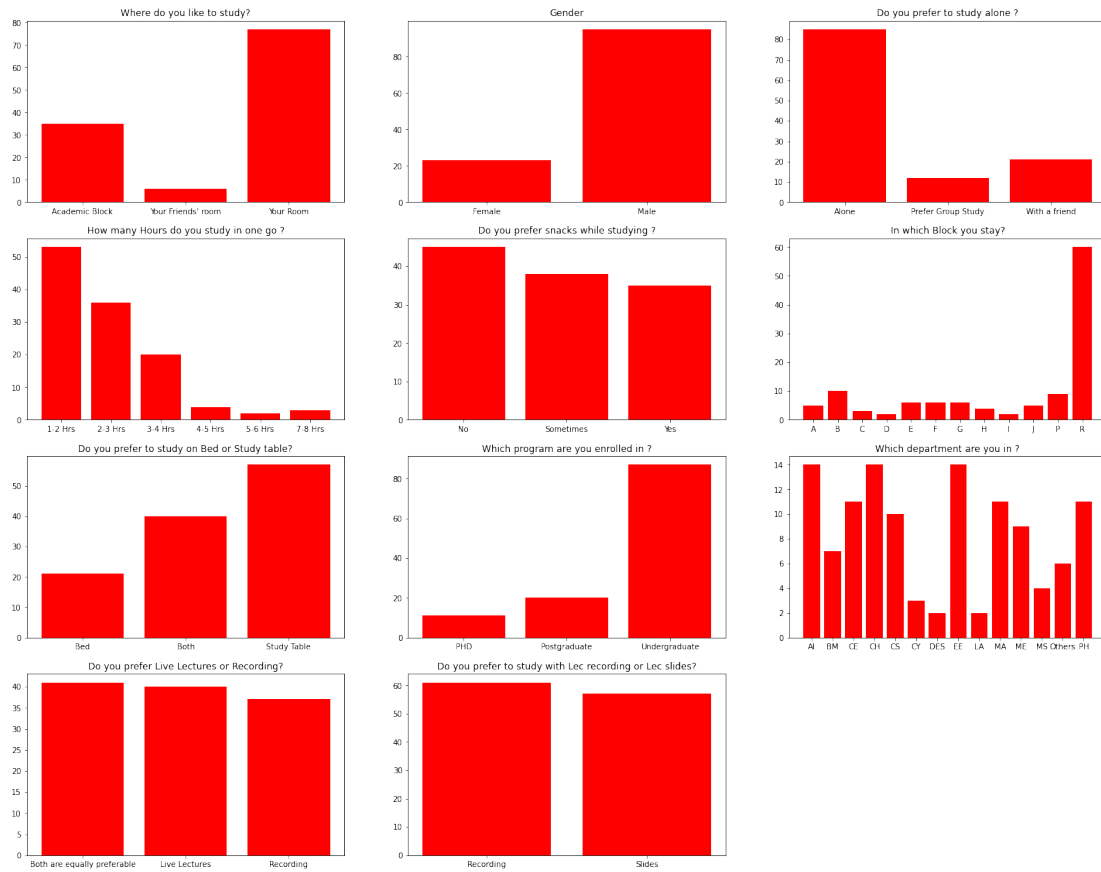


Figure 1. Visualizing the collected data

3. Analysis and Conclusions

After the pre-processing of the data, we had the data of 118 students from the survey, and we began to analyze the data. This involved the analysis of the one question that involved a numerical variable, that is:

“How many hours do you study in one go?”

The following results were obtained:

- Mean = 2.466 hours
- Median = 2.5 hours
- Mode = 2.5 hours

From the Box Plot: Figure 2, as well as the statistics, since the mean, median and mode are all approximately equal, we can say the data approximately follows a normal distribution.

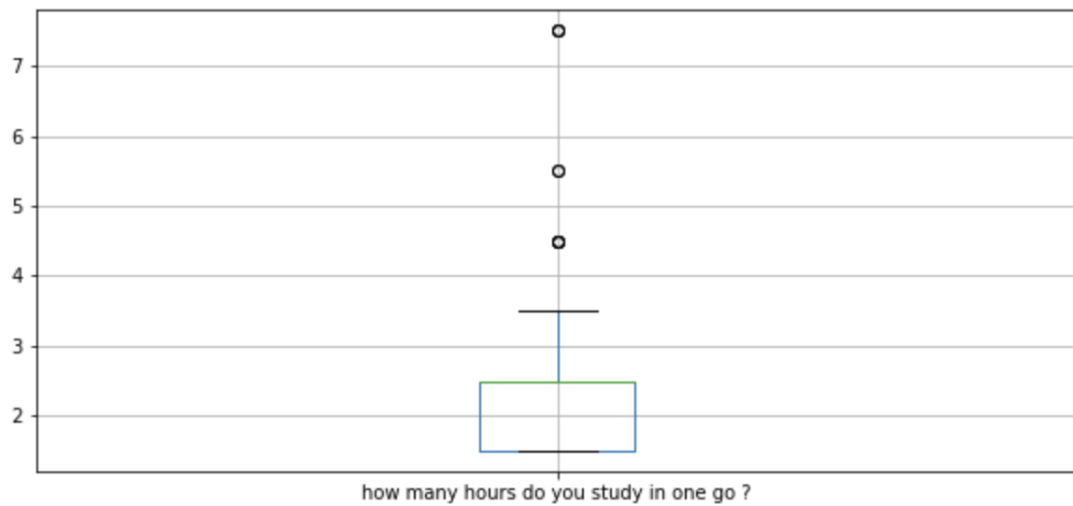


Figure 2. How many hours do you study in one go? (Box Plot)

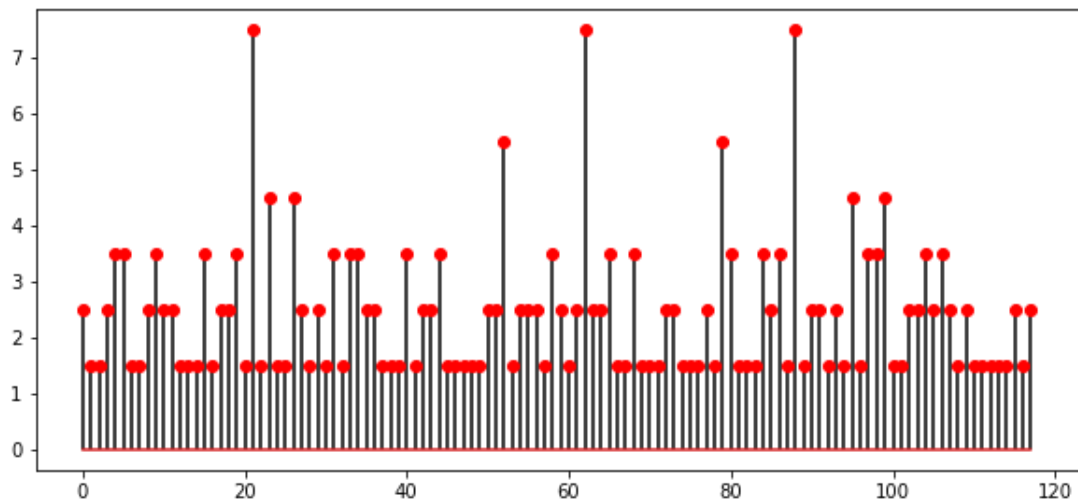


Figure 3. How many hours do you study in one go? (Stem Plot)

Doing some confidence testing for the mean of the data gives us the following:

- A 95% confidence interval is given by: (2.2388, 2.6934).
The width of the interval is given by: 0.4546
- A 99% confidence interval is given by: (2.1656, 2.7666).
The width of the interval is given by: 0.6010

As expected, as we increase the confidence level, the width of the interval also increases.

A contingency table made from the data is as follows:

do you prefer to study alone ?	where do you like to study?	Alone	Prefer Group Study	With a friend	All
0	Academic Block	18.64	2.54	8.47	29.66
1	Your Friends' room	1.69	1.69	1.69	5.08
2	Your Room	51.69	5.93	7.63	65.25
3	All	72.03	10.17	17.80	100.00

Figure 4. Contingency table for comparison of multiple categorical variables

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 segmented bar charts:

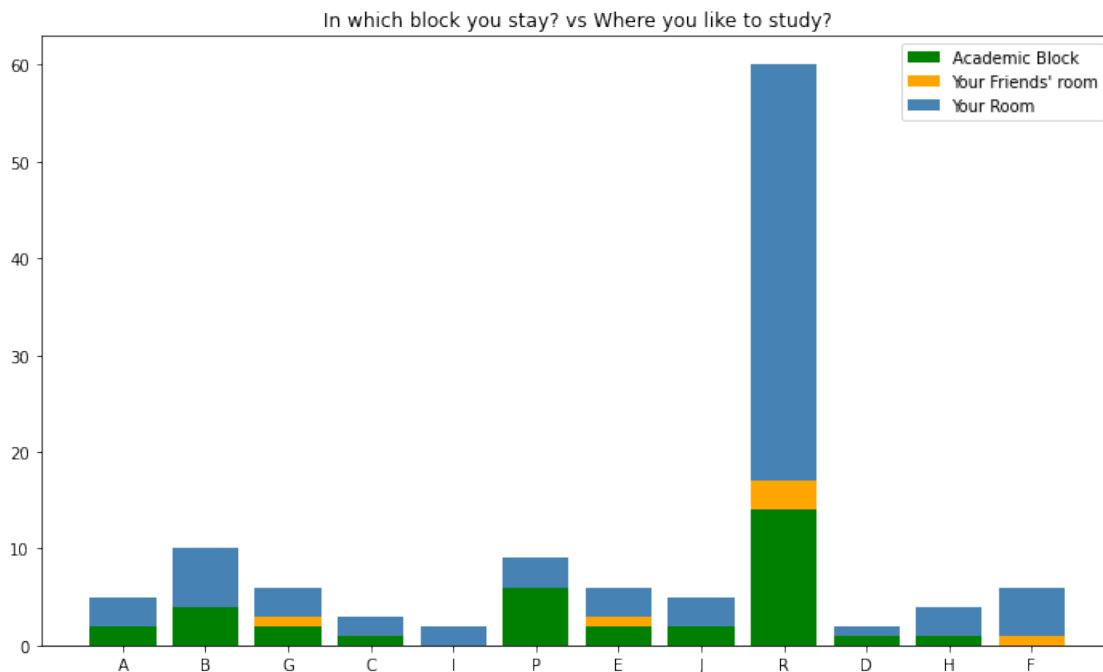


Figure 5. Residing Block vs Study room

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

1. A majority of students from the R block prefer to study in their rooms alone, whereas a majority of students from the P block (mainly involving Masters and Research students) prefer to go to the academic blocks.
2. Very few prefer to study in their friends' room. More or less, from every block, people want to study in their room itself.

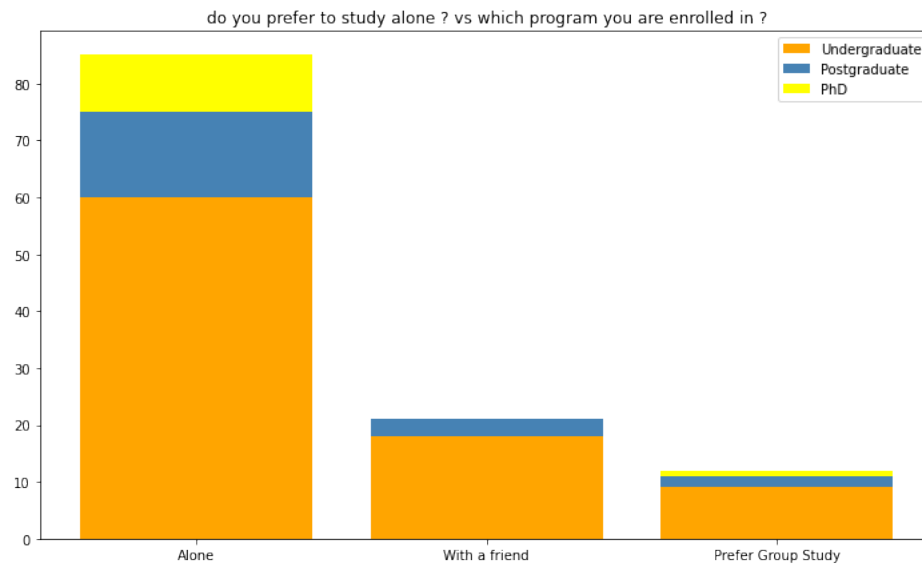


Figure 6. Preference towards studying alone vs Program

Based on the Segmented Bar Plot: Figure 6, the following conclusion can be made: Undergraduate students are open to studying alone, with a friend or even involving themselves in group studies. However, as we progress to Postgraduate and PhD students, they prefer studying alone in general.

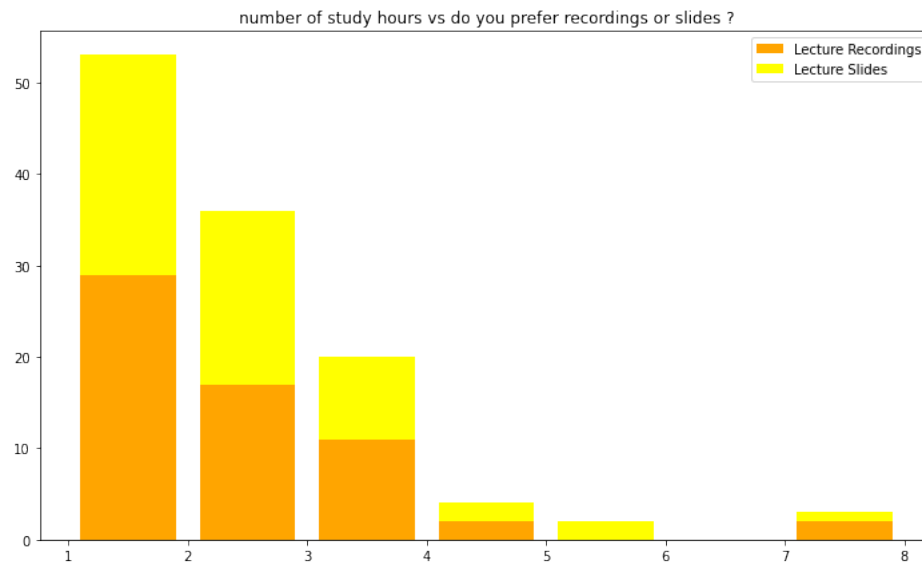


Figure 7. Hours studied vs Slides or Recordings

Based on the Segmented Bar Plot: Figure 7, the following conclusion can be made: More or less, equal number of people prefer lecture recordings and/or slides. However, most

people who prefer recordings study for the minimum time at one go, and as this time tends to increase, people prefer to use slides to study.

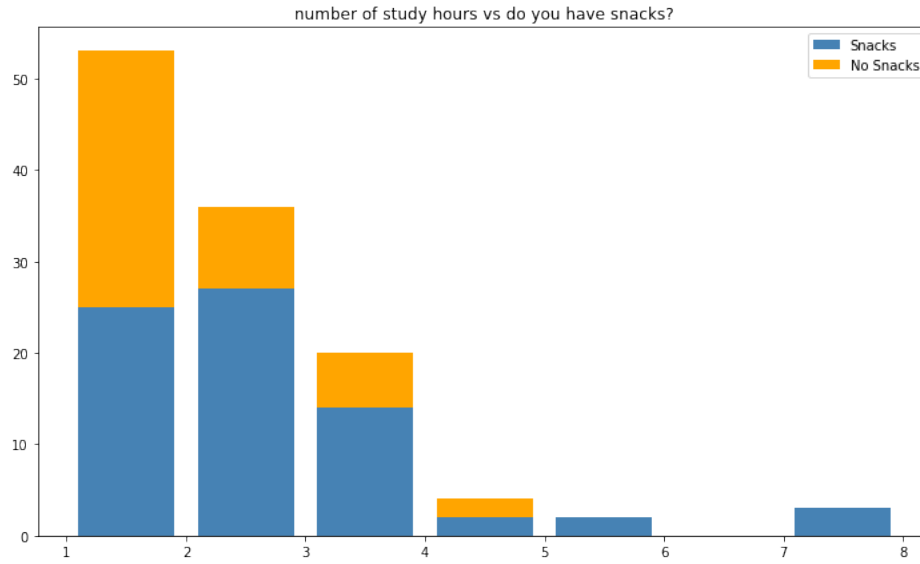


Figure 8. Study Hours vs Snacks Consumptions

Based on the Segmented Bar Plot: Figure 8, the following conclusion can be made: A larger proportion of people studying for longer hours prefer having snacks with them as compared to people studying just for 1 – 2 hours

4. Hypothesis Testing:

4.1. Case-1: Comparing the study hours for Undergraduates and Postgraduates

We compare study hours of Undergraduates and Postgraduates. We assume our null hypothesis to be that undergraduates study for longer duration than postgraduates. Let $\alpha = 0.05$

Let \bar{x}_1 = Sample Mean of study of Undergraduates

Let \bar{x}_2 = Sample Mean of study of Postgraduates

Let S_1^2 = Sample Variance of study of Undergraduates

Let S_2^2 = Sample Variance of study 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 will now select a random sample of 18 entries for the Undergraduates and 18 entries for the Postgraduates from our data. Based on the sample selected, we have the following

information-

$$\bar{x}_1 = 2.50 \text{ hours} \quad (1)$$

$$\bar{x}_2 = 2.37 \text{ hours} \quad (2)$$

$$S_1^2 = 1.219 \quad (3)$$

$$S_2^2 = 1.335 \quad (4)$$

$$n_1 = 81 \quad (5)$$

$$n_2 = 37 \quad (6)$$

$$(7)$$

Since $\frac{S_1^2}{S_2^2} = 1.095 < 4$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 116$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 1.255 \quad (8)$$

The test statistic is then given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 0.584 \quad (9)$$

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 $t = 0.584$ is less than 1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, the postgraduates study more in one go than the undergraduates on average.

4.2. Case-2: Comparing the study hours of people who study alone and who study in groups.

We compare study hours of people studying alone and in groups. We assume our null hypothesis that people studying alone study for longer time.

Let \bar{x}_1 = Sample Mean of study hours of people studying alone

Let \bar{x}_2 = Sample Mean of study hours of people studying in groups

Let S_1^2 = Sample Variance of study hours of people studying alone

Let S_2^2 = Sample Variance of study hours of people studying in groups

For Hypothesis Testing we make the following statements-

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

$$\bar{x}_1 = 2.488 \text{ hours} \quad (10)$$

$$\bar{x}_2 = 2.409 \text{ hours} \quad (11)$$

$$S_1^2 = 0.535 \quad (12)$$

$$S_2^2 = 0.647 \quad (13)$$

$$n_1 = 85 \quad (14)$$

$$n_2 = 33 \quad (15)$$

Since $\frac{S_1^2}{S_2^2} = 1.209 < 4$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 116$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.5658 \quad (16)$$

The test statistic is then given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 0.512 \quad (17)$$

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 $t = 0.512$ is less than 1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those who study in groups study more on average in one go than those who study alone.

4.3. Case-3: Comparing the study hours of people who study while eating snacks and who study without eating snacks.

We compare study hours of people studying while eating snacks and those who prefer not to! We assume our null hypothesis that people studying while eating snacks study for longer time.

Let \bar{x}_1 = Sample Mean of study hours of people studying while eating snacks

Let \bar{x}_2 = Sample Mean of study hours of people studying without eating snacks

Let S_1^2 = Sample Variance of study hours of people studying while eating snacks

Let S_2^2 = Sample Variance of study hours of people studying without eating snacks

For Hypothesis Testing we make the following statements-

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

$$\bar{x}_1 = 2.671 \text{ hours} \quad (18)$$

$$\bar{x}_2 = 2.379 \text{ hours} \quad (19)$$

$$S_1^2 = 0.91 \quad (20)$$

$$S_2^2 = 0.4 \quad (21)$$

$$n_1 = 35 \quad (22)$$

$$n_2 = 83 \quad (23)$$

Since $\frac{S_1^2}{S_2^2} = 2.275 < 4$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 116$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.55 \quad (24)$$

The test statistic is then given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.9685 \quad (25)$$

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 $t = 1.9685$ is greater than 1.658, we fail to reject the null hypothesis, and thus, do not have enough evidence to say that those who eat snacks while studying study for longer at one go than those who do not.

4.4. Case-4: Comparing the study hours of people who study from Lecture Slides and those who study from Lecture Recordings.

We compare study hours of people studying from lecture slides and those who study from lecture recordings! We assume our null hypothesis that people studying from slides study for longer time than those who study from lecture recordings.

Let \bar{x}_1 = Sample Mean of study hours of people studying from slides
Let \bar{x}_2 = Sample Mean of study hours of people studying from recordings
Let S_1^2 = Sample Variance of study hours of people studying from slides
Let S_2^2 = Sample Variance of study hours of people studying from recordings

For Hypothesis Testing we make the following statements-

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

$$\bar{x}_1 = 2.5 \text{ hours} \quad (26)$$

$$\bar{x}_2 = 2.43 \text{ hours} \quad (27)$$

$$S_1^2 = 0.5 \quad (28)$$

$$S_2^2 = 0.629 \quad (29)$$

$$n_1 = 57 \quad (30)$$

$$n_2 = 61 \quad (31)$$

Since $\frac{S_1^2}{S_2^2} = 1.258 < 4$, we can assume the population variances would be equal. Thus, we can say:

The degrees of freedom, $df = n_1 + n_2 - 2 = 116$

The pooled variance will be:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.5764 \quad (32)$$

The test statistic is then given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 0.5 \quad (33)$$

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 $t = 0.5$ is lesser than 1.658, we have enough statistical evidence to reject the null hypothesis, and thus, we can say, those who study from recordings study more on average in one go than those who study from slides.

5. Contributions:

- **Varunaditya Singhal**

- The idea of the Project.
- Making Questions along with the google form.
- Data collection for the project
- pre-processing of data.
- Hypothesis testing of 3 scenarios along with the code for same.
- Making report in \LaTeX

- **Prakhar Patni**

- Ideation for the project
- Data collection for the project
- Ideation for Hypothesis testing
- Making report in \LaTeX

- **Tanmay Goyal**

- Ideation for the project
- Helping with the pre-processing of data
- \LaTeX report and drawing conclusions
- Assisted with Hypothesis Testing ideation, and code for the same.
- Collab Notebook organisation and clean-up

- **K N Vardhan**

- Ideation for the project
- Analyzing the Uni-variate Numerical dataset and drawing conclusions
- Visualizing few of the categorical datasets with segmented bar charts
- Assisted in hypothesis testings
- Making presentation slides

- **Tanay Yadav**

- Ideation for the project
- Pre-processing the data using Pandas in Python
- Visualizing few of the categorical datasets with segmented bar charts

- Making report in \LaTeX

- **Sujal**

- Ideation for the project
- Made presentation