Analyzing Sleep Patterns of IITH Students

Prajwaldeep Kamble Krishna Teja Chilukuri Nikhil Kongara
Karthik Kurugodu VKS Deepak Reddy Tata Sai Manoj
Venkata Raghav Ambati Anurag Gopi Sumanth NR

Indian Institute of Technology Hyderabad

Project No. 12

April 26, 2022

Structure of the Project

The **objective** of the project is to identify various factors which affect our sleep and the way they affect us.

The **variables of interest** for this study are :

Response variables : -

- Number of hours of sleep
- Start time of sleep

Explanatory variables : -

- Whether a person consume caffeine
- How many times caffeinated drinks are consumed per day?
- How and at when do people watch lectures
- How and when do people submit assignments?
- Start time of study
- Time spent on recreational activities
- Time spent on sports and physical activities

As our interest is to analyze natural sleeping patterns, we perform an observational study. We assume the effect of any **confounding variables** to be negligible as we are concentrating on sleeping patterns of student life.

We used sampling, more specifically **volunteer sampling** for collection of data by sending mail to every student at IITH but only a few of them volunteered to respond. To get the maximum number of responses, we also sent frequent reminders. The data was collected from 132 people, and it is diverse with data from all years of UG, PG and Phd students.

Presentation Overview I

Presentation Overview II

Central Tendencies for Sleep Hours

	How much sleep do you get in a day?
count	132.000000
mean	6.641714
std	1.256754
min	3.199589
25%	6.080173
50%	6.472127
75%	6.926733
max	9.958100

Figure: Central Tendencies for Sleep Hours

Sampling Distribution of Sleep Hours

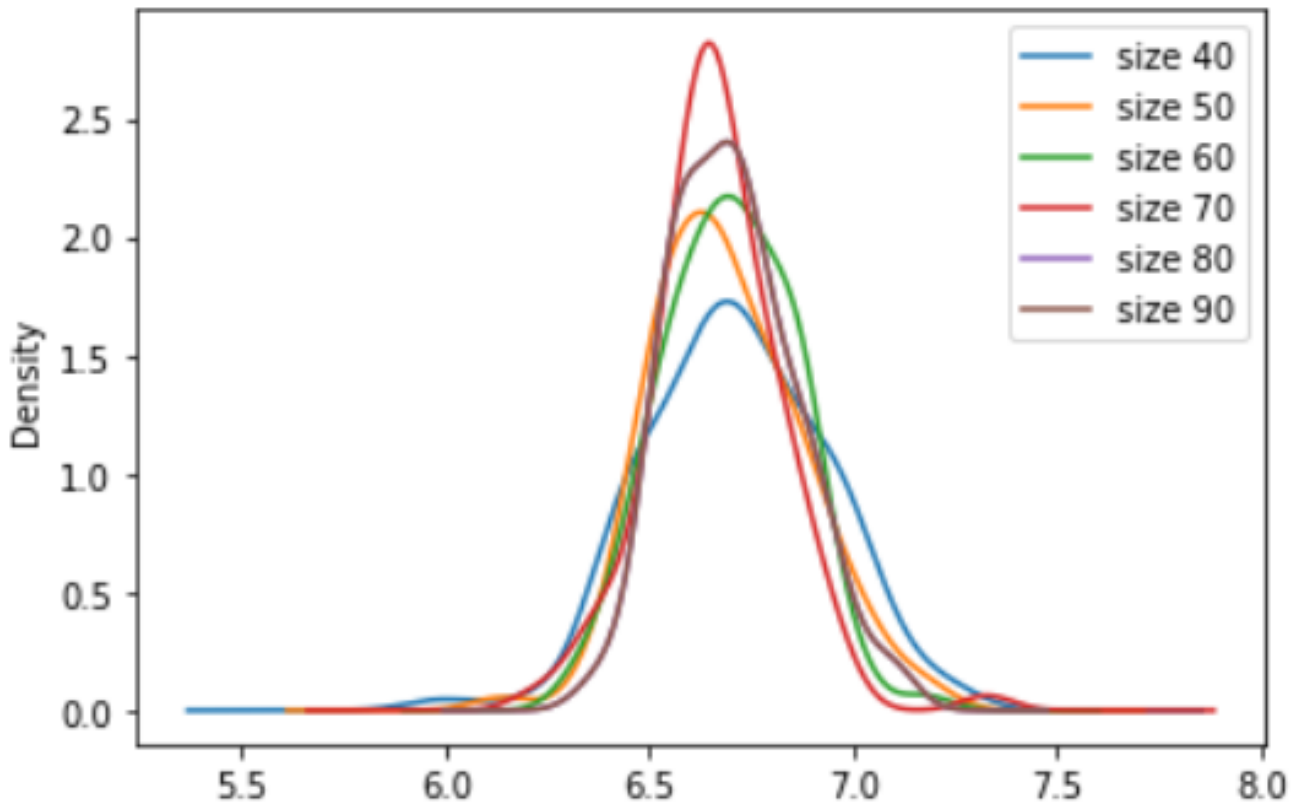

Figure: Sampling Distribution of Means of Sleep Hours

Sampling Distribution of Means tends to a normal distribution as the value of n increases.

Sampling Mean Distribution

We also plotted the sampling mean distribution for sleep time by selecting random samples of varying sizes greater than 30. We can observe that as the random sample size increases the distribution is tending to be normal. From this we can observe that our data is not highly skewed and good representative of the population.

Hypothesis Testing

Hypothesis

Students of IITH sleep at least 6 hrs on average.

Testing Hypothesis

Calculation

Null Hypothesis

$$(H_0) : \mu - \mu_0 \leq 0 \tag{1}$$

Alternate Hypothesis

$$(H_a) : \mu - \mu_0 > 0 \tag{2}$$

$\overline{X} = 6.641713$

$S = 1.256754$

$n = 132, df = 131$

Testing Hypothesis

Test Statistic

$$t = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{6.641713 - 6}{\frac{1.256754}{\sqrt{132}}} = 5.844220$$

$$t_{\alpha,df} = t_{0.01,131} = 2.355150$$

Since $t > t_{\alpha,df}$, we can reject H_0.

Conclusion

So, we can conclude that students of IITH sleep at least 6 hrs on average.

Central Tendencies for Sleep Times

	normalisedTime
count	130.000000
mean	5.635385
std	2.967924
min	2.000000
25%	4.000000
50%	5.000000
75%	6.000000
max	18.000000

Figure: Values of Central Tendencies (Time normalised about 8*PM*)

The central tendencies are as follows :

Count : Refers to the number of students who have responded to this question

Mean : The mean time of sleep (Since it is normalized around 8 PM, the mean time of sleep would be $5.63 + 20(8PM) = 25.63$ (or) 1.63, which is approximately $1:40$ AM).

Std : The standard deviation of sleep times

Min : The earliest normalised sleep time (corresponds to $22:00$)

25% : The 1^{st} quartile

50% : The 2^{nd} quartile or median

75% : The 3^{rd} quartile

Max : The last normalized sleep time (corresponds to $14:00$)

Sleep Time Analysis

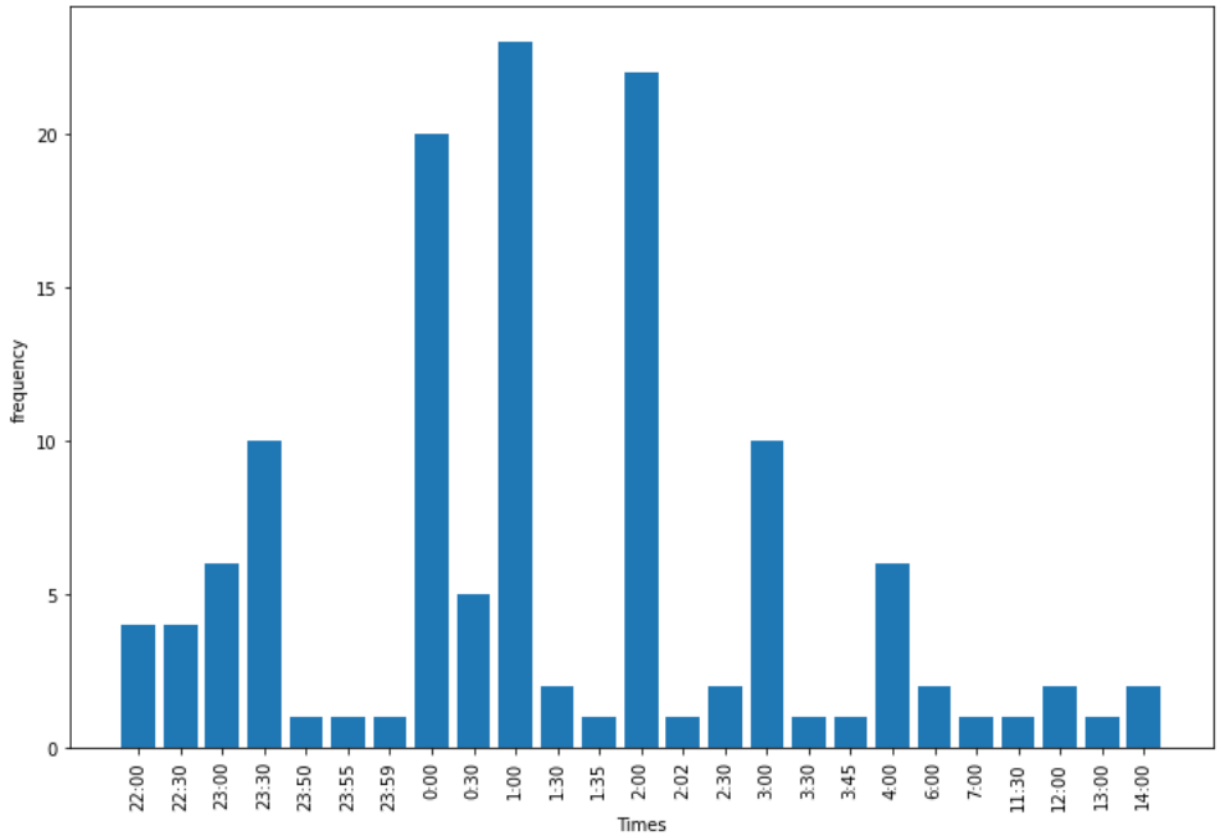

Figure: The frequency of students sleeping at certain times

We can see that it is a **slightly right-skewed** and **unimodal**.

Sampling Mean Distribution

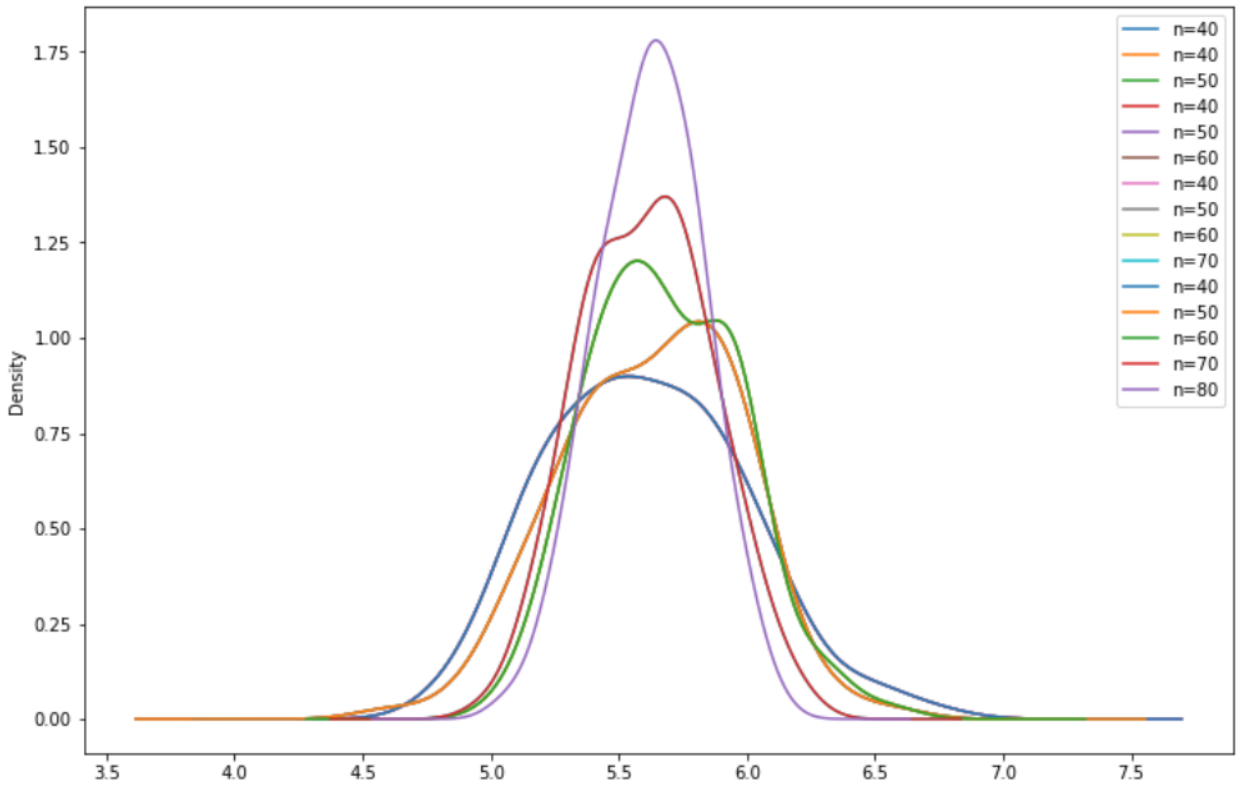

Figure: Data is consistent with Central Limit Theorem

Sampling Distribution of Means tends to a normal distribution as the value of n increases.

Central Tendencies for Study Times

	normalisedTime
count	130.000000
mean	11.152436
std	5.478415
min	1.000000
25%	6.000000
50%	11.750000
75%	16.375000
max	21.000000

Figure: Values of central Tendencies (Time normalised about 5*AM*)

The central tendencies are as follows :

Count : Refers to the number of students who have responded to this question

Mean : The mean time of study (Since it is normalized around 5 AM, the mean time of sleep would be $11.15 + 5 = 16.15$, which is approximately $4:10PM$).

Std : The standard deviation of study times

Min : The earliest normalised study time (corresponds to $6:00AM$)

25% : The 1^{st} quartile

50% : The 2^{nd} quartile or median

75% : The 3^{rd} quartile

Max : The last normalized study time (corresponds to $2:00AM$)

Study Time Analysis

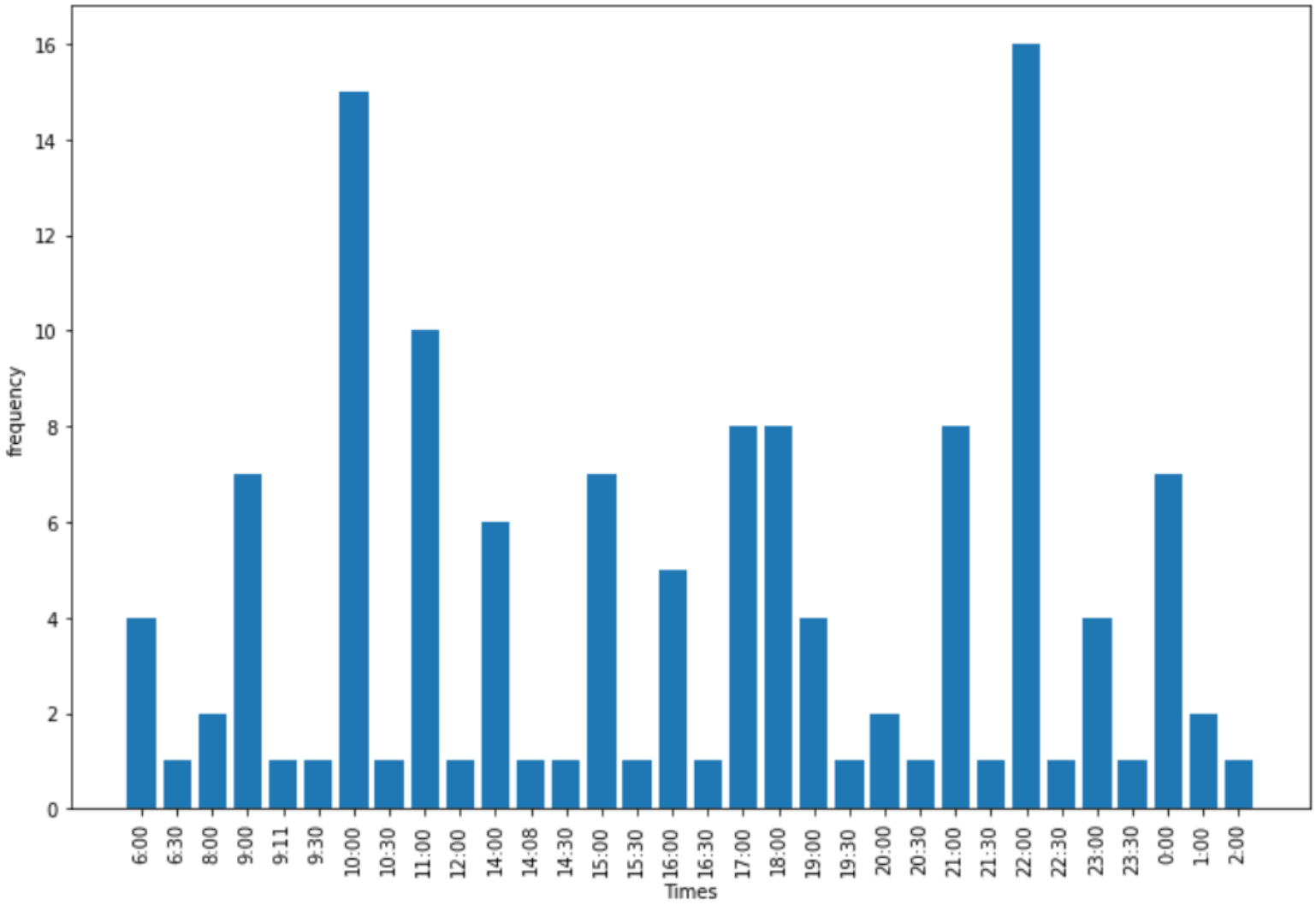

Figure: The frequency of students studying at certain times

Study Time Analysis

From the plot, we can see that study times is near-symmetric, and a bimodal data set.

One major reason for this occurrence is due to the fact that there are two groups of students, those who study early in the morning (before classes) and those who study at night (after classes).

This leads to two peaks, which we can see at 10 AM and 10 PM.

Relation between Sleep Hours and Study Times

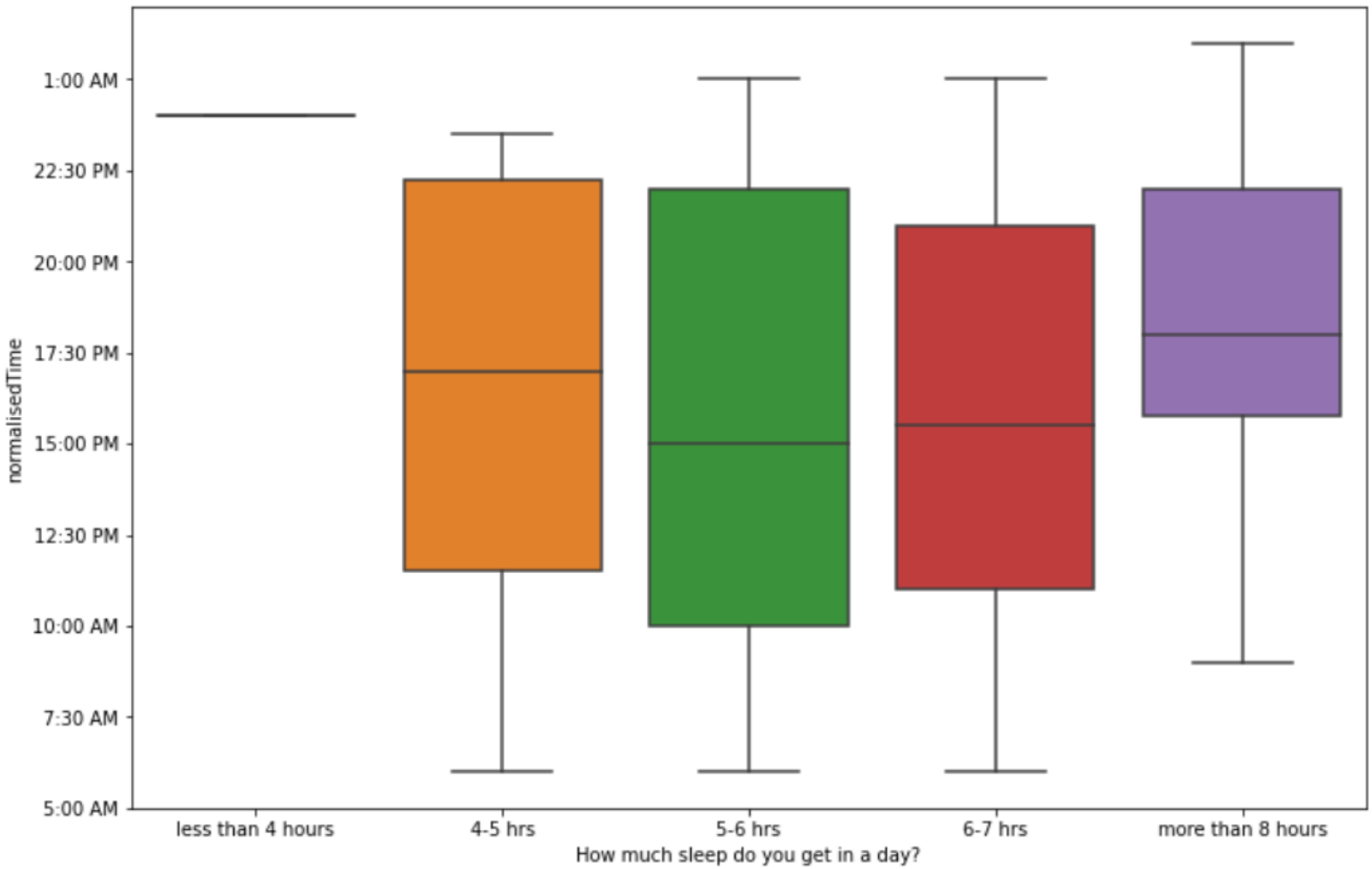

Figure: Plot of Study Time and Hours of Sleep

Inference from Box Plot

Shape: Slightly right-skewed, so most people tend to start studying early.

Center: People with moderate sleep tend to start studying earlier than people with extreme sleep habits.

Spread: The IQR of people who sleep more is less, so they consistently start studying in the evenings.

Confidence Interval Estimation

We can say that with 95% confidence that the population mean time at which a person sleeps lies between (1 : 07 AM, 2 : 09 AM).

We can say that with 95% confidence that the population mean time at which a person studies lies between (15 : 12 PM, 17 : 06 PM).

Caffeine

General Information about Effects of Caffeine

Caffeine is one of the most consumed substances on a day-to-day basis. It is mainly consumed because of its profound effect on sleep and other cognitive functions.

Known for its affect of sharpening the senses, it is majorly consumed by students to stay awake for longer periods of time, particularly for academic activities.

Studies have shown that caffeine dependence develops at relatively low daily doses and after short periods of regular daily use.

The risks to sleep and alertness of regular caffeine use are greatly underestimated by both the general population and physicians.

Relation between Caffeine intake and Sleep Times

	How much sleep do you get in a day?	TOS
count	46.000000	46.000000
mean	6.798549	15.278623
std	1.455183	5.180778
min	4.209851	10.000000
25%	5.942871	11.500000
50%	6.524550	13.291667
75%	8.097821	15.750000
max	9.819538	24.500000

Figure: DF0 - No Caffeine intake

	How much sleep do you get in a day?	TOS
count	39.000000	39.000000
mean	6.465055	15.526496
std	1.212115	5.046817
min	3.199589	2.000000
25%	6.112352	13.000000
50%	6.497537	14.000000
75%	6.897980	15.500000
max	9.811279	24.500000

Figure: DF1 - One Caffeinated drink

	How much sleep do you get in a day?	TOS
count	34.000000	34.000000
mean	6.578082	14.367647
std	0.937139	4.885521
min	4.878099	1.000000
25%	6.214373	13.000000
50%	6.430458	14.000000
75%	6.738214	15.875000
max	9.958100	24.000000

Figure: DF2 - Two Caffeine intake

	How much sleep do you get in a day?	TOS
count	8.000000	8.000000
mean	7.114196	15.218750
std	1.704296	4.203395
min	4.712343	11.000000
25%	6.074998	12.625000
50%	6.773000	14.250000
75%	8.386601	16.312500
max	9.463062	24.000000

Figure: DF3 - More than Three Caffeinated drinks

The central tendencies are as follows :

Count : Size of the sample

Mean : Mean in the first column is mean of sleep hours in a day, and in the second column, it is mean time of sleep (normalized around 12 PM). So, mean time of sleep is $12 + 15.27 = 27.27$ (or) 3.27, which is around $3 : 20$ AM.

Std : The standard deviation of study times

Min : Column - 1 : The minimum number of sleep hours is 4.2 hours.

Column - 2 : The earliest time of sleep is $10 + 12 = 22$ (or) 10 PM.

25% : The 1^{st} quartile

50% : The 2^{nd} quartile or median

75% : The 3^{rd} quartile

Max : The last normalized study time (corresponds to $2 : 00$ AM)

Relation between Caffeine intake and Sleep Hours

Boxplots

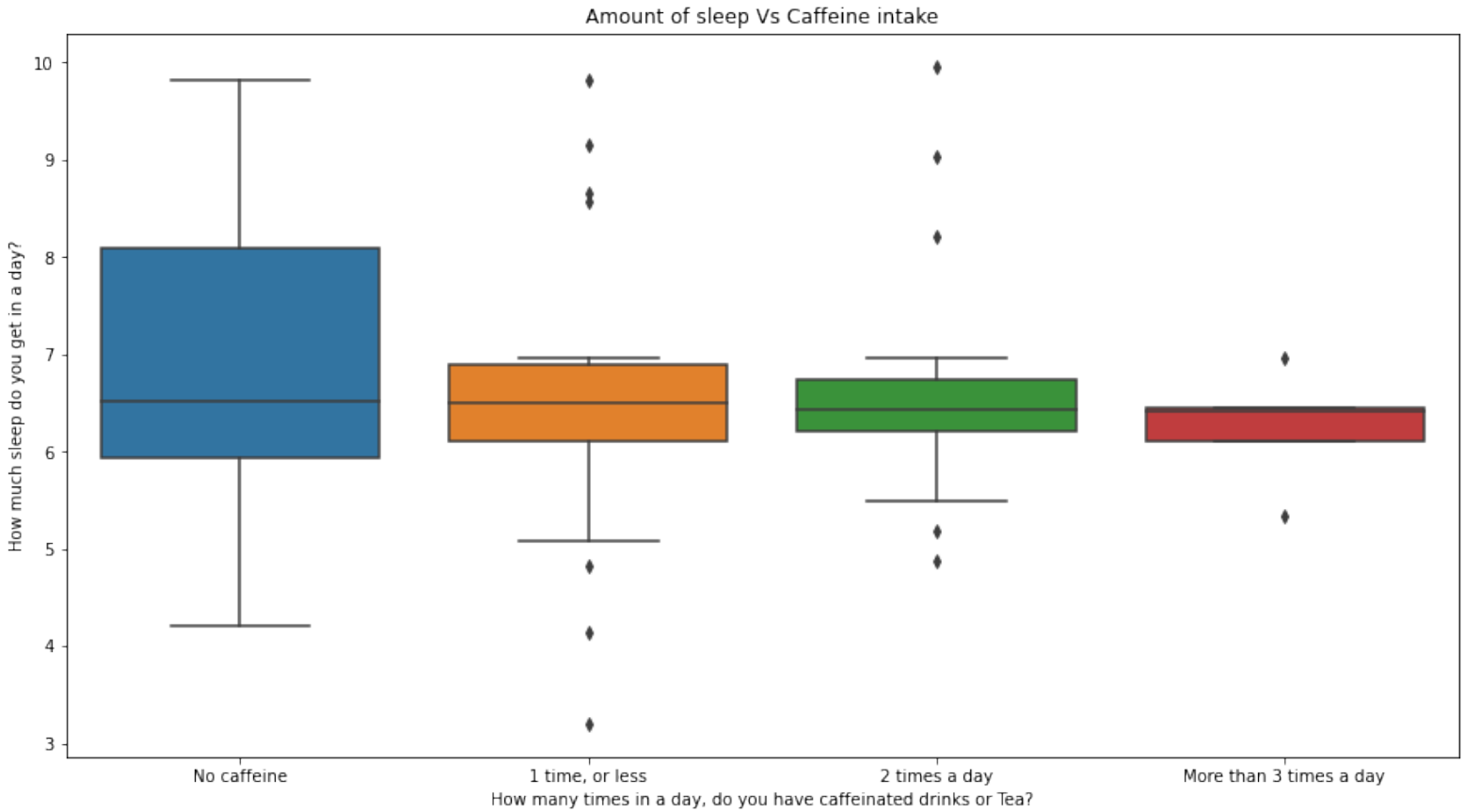

Figure: Caffeine effect on Sleep Hours

Relation between Caffeine intake and Time of Sleep

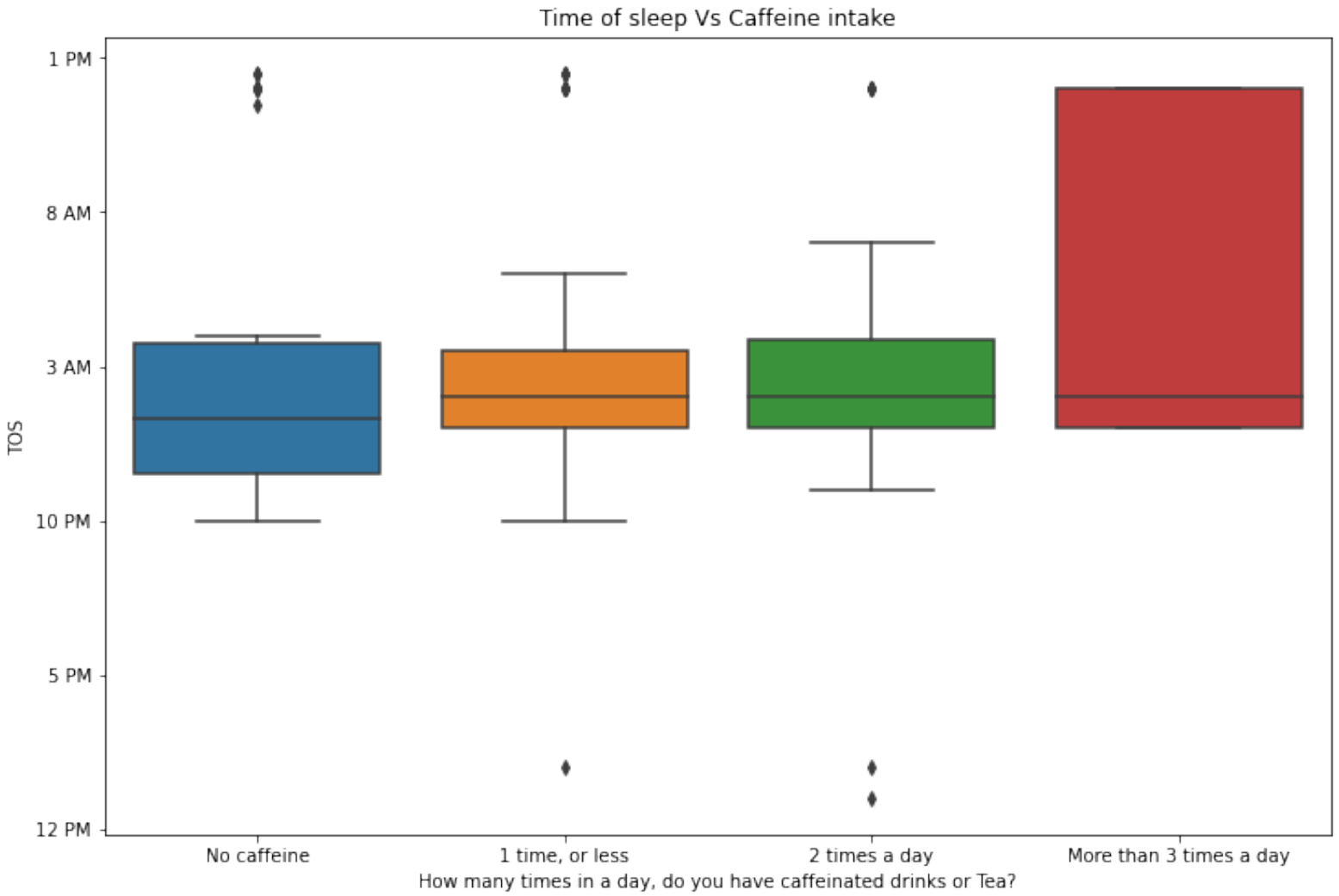

Figure: Caffeine effect on Time of Sleep

Inference from Box Plot - 1

Shape: For people who don't consume caffeine, the data is right-skewed so they tend to sleep less. But this value is greater than the number of sleep hours of other categories which consume caffeine.

Center: All categories have nearly the same median.

Spread: We can observe that IQR decreases as caffeine consumption increases, so people who consume caffeine consistently sleep for the same amount of time.

Outliers: There are a considerable number of outliers among the moderate caffeine consuming categories.

Inference from Box Plot - 2

Shape: The box plot for high caffeine consumers is right skewed. More people in this category sleep early but this is late when compared with other categories.

Center: All categories have nearly the same mean.

Spread: We can observe that IQR of high caffeine consumers is more when compared with other categories, so they donot sleep consistently at the same time.

Outliers: There are a few outliers among less or no caffeine consuming categories.

Confidence Interval Calculation

Confidence Interval for Number of Hours of Sleep

For a 95% confidence interval, the value of confidence coefficient is $\alpha = 0.05$.

$$t_{\frac{\alpha}{2}} = 1.97823853 \tag{3}$$

$$\overline{X} - (t_{\frac{\alpha}{2},df})(\frac{S}{\sqrt{n}}) \leq \mu \leq \overline{X} + (t_{\frac{\alpha}{2},df})(\frac{S}{\sqrt{n}}) \tag{4}$$

$$6.641 - (1.978)(\frac{1.251}{\sqrt{n}}) \leq \mu \leq 6.641 + (1,978)(\frac{1.251}{\sqrt{n}}) \tag{5}$$

Confidence Interval Conclusions

1. We can say with *95 %* confidence that the number of hours of sleep is in the interval (6.42, 6.85)

2. We can say with *95 %* confidence that the number of hours of sleep of a person who ***does not take*** any caffeine daily is in the interval (6.37, 7.22)

3. We can say with *95 %* confidence that the number of hours of sleep of a person who take ***one*** caffeinated drink a day is in the interval (6.07, 6.85)

Confidence Interval Conclusions

4. We can say with *95 %* confidence that the number of hours of sleep of a person who take ***two*** caffeinated drinks a day is in the interval (6.25, 6.90)

5. We can say with 95 % confidence that the number of hours of sleep of a person who take ***three*** caffeinated drinks a day is in the interval (5.78, 8.44)

6. We can say with *95 %* confidence that the number of hours of sleep of a person who take ***greater than three*** caffeinated drinks a day is in the interval (5.58, 6.92)

Hypothesis

Students who take caffeine sleep less than those who don’t take caffeine.

Testing Hypothesis

Calculation

Null Hypothesis (H_0) : $\mu_c - \mu_{nc} \geq 0$

Alternate Hypothesis (H_a) : $\mu_c - \mu_{nc} < 0$

$\overline{X}_c = 6.798548$

$n_c = 46$

$S_c = 1.455182$

$\overline{X}_{nc} = 6.544580$

$n_{nc} = 52$

$S_{nc} = 1.259272$

Test Statistic

$$t = \frac{(\overline{X}_c - \overline{X}_{nc}) - (0)}{\sqrt{\frac{(n_c - 1)S_c^2}{n_c + n_{nc} - 2} + \frac{(n_{nc} - 1)S_{nc}^2}{n_c + n_{nc} - 2}}\left(\sqrt{\frac{1}{n_c} + \frac{1}{n_{nc}}}\right)} = 0.918048$$

$$df = n_c + n_{nc} - 2 = 96$$

Rejection Region Approach

Reject H_0 if $t \leq -t_{\alpha,df}$ (let α = 0.05)

$t = 0.918048$, $t_{0.05,96} = 1.6609612$

Hence, the hypothesis H_0 is not rejected.

Conclusion

The evidence is insufficient to conclude that those who take caffeine sleep less than those who don't take caffeine.

Academics

Central Tendencies for Lecture Watching Patterns

	binge watching or reading	regular but recordings	live	both live and recordings
count	65.000000	19.000000	19.000000	27.000000
mean	6.768753	7.060560	6.993222	6.976894
std	0.995108	0.793951	0.848921	0.614738
min	3.653054	5.605172	5.702607	5.334842
25%	6.533755	6.805897	6.378027	6.671317
50%	6.874540	7.023833	7.045851	6.951411
75%	7.351985	7.496349	7.375875	7.161050
max	8.587655	8.476130	8.591787	8.410634

Figure: Lecture Watching Patterns

The central tendencies are as follows :

Count : Size of the sample (Here, number of students who binge watch lectures / who watch lectures live)

Mean : Mean in each column represents the mean of the hours of sleep the students of that category are getting.

Std : The standard deviation of sleep times

Min : The least hours of sleep a student of that category is getting.

25% : The 1^{st} quartile

50% : The 2^{nd} quartile or median

75% : The 3^{rd} quartile

Max : The maximum hours of sleep a student of that category is getting.

Central Tendencies for Assignment Submission Patterns

	A day or more before deadline	On the last day of deadline	Last minute
count	15.000000	79.000000	36.000000
mean	6.817429	6.843700	7.012588
std	0.991628	0.738656	1.099612
min	4.616059	4.611331	3.653054
25%	6.503229	6.599587	6.690900
50%	6.907687	6.958055	7.123132
75%	7.202889	7.251794	7.738544
max	8.583296	8.591787	8.587655

Figure: Assignment Submission Patterns

Box Plots

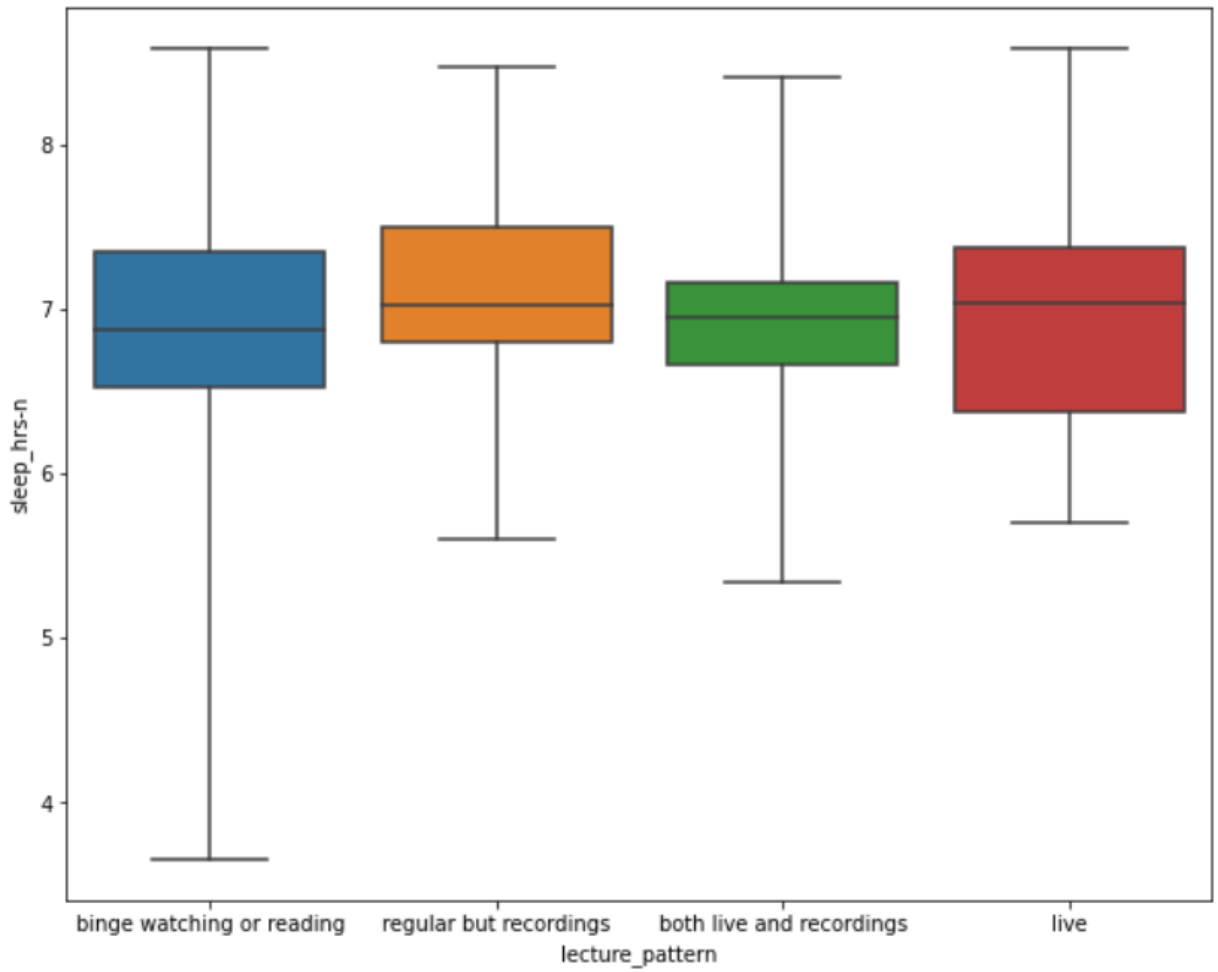

Figure: Box Plot for Modes of Lecture Watching

Box Plots

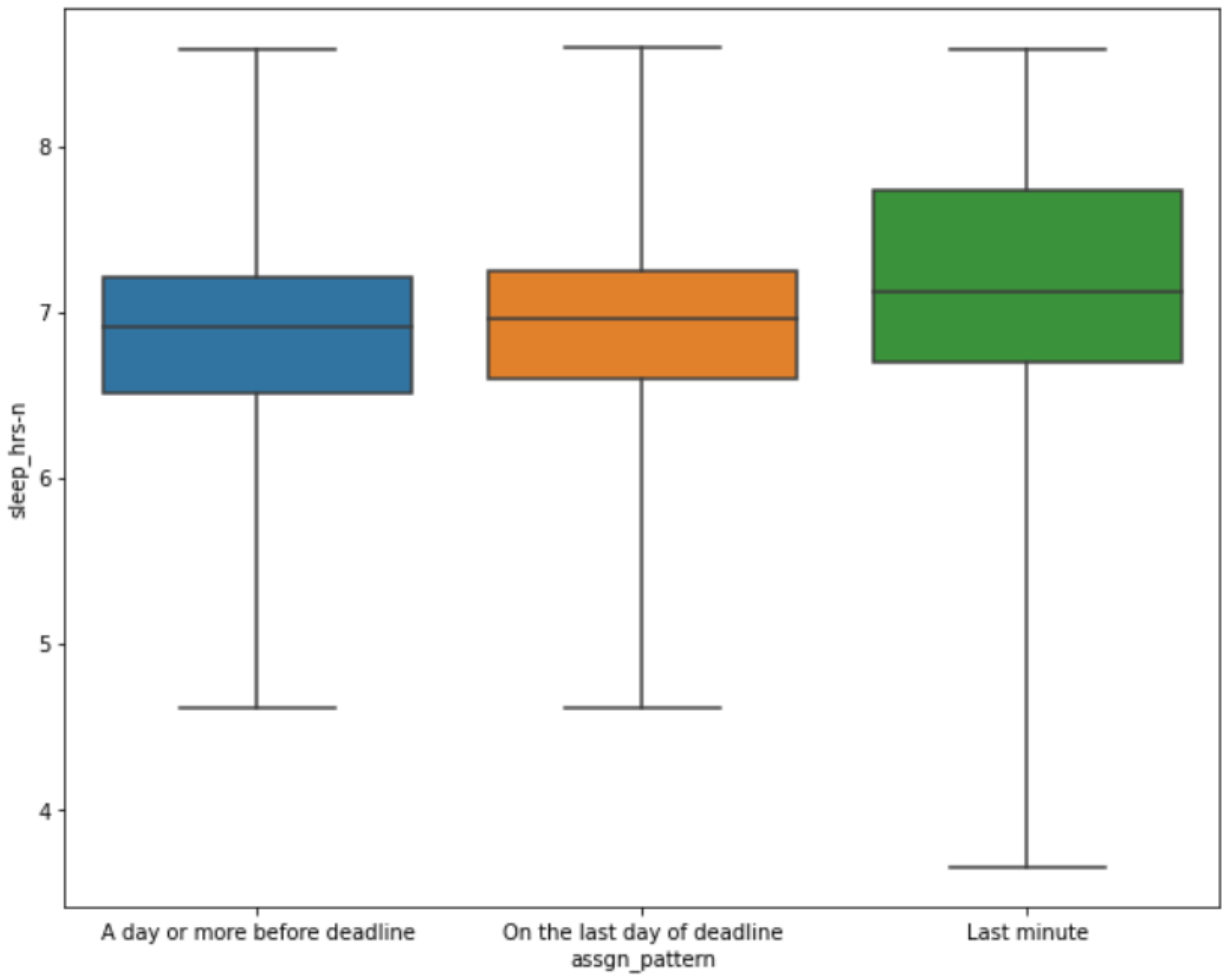

Figure: Box Plot for Assignment Submission Patterns

Inference for Box Plot - 1

Shape: Category with left-skewness is people who watch live lectures and both live and recordings, so they tend to sleep for more time among themselves, which is different when compared with other categories.

Center: The average amount of sleep for the category of people watching recordings regularly is slightly higher than other categories.

Spread: The IQR of the category of people who watch both live lectures and recordings is the least, so they sleep more consistently when compared with other categories. Also, the range of sleep hours for the category of people who binge watch lectures is very high when compared with others, so the number of hours they sleep is very unpredictable.

Inference for Box Plot - 2

Shape: Only category with right-skewness is people who submit the assignment last minute, so they tend to sleep for less time among themselves.

Center: The average amount of sleep for the category of people is slightly increasing based on how close to the deadline they are submitting.

Spread: The IQR of the category of people who submit the assignment last minute is highest, so they sleep less consistently when compared with other categories.

Segmented Bar Chart

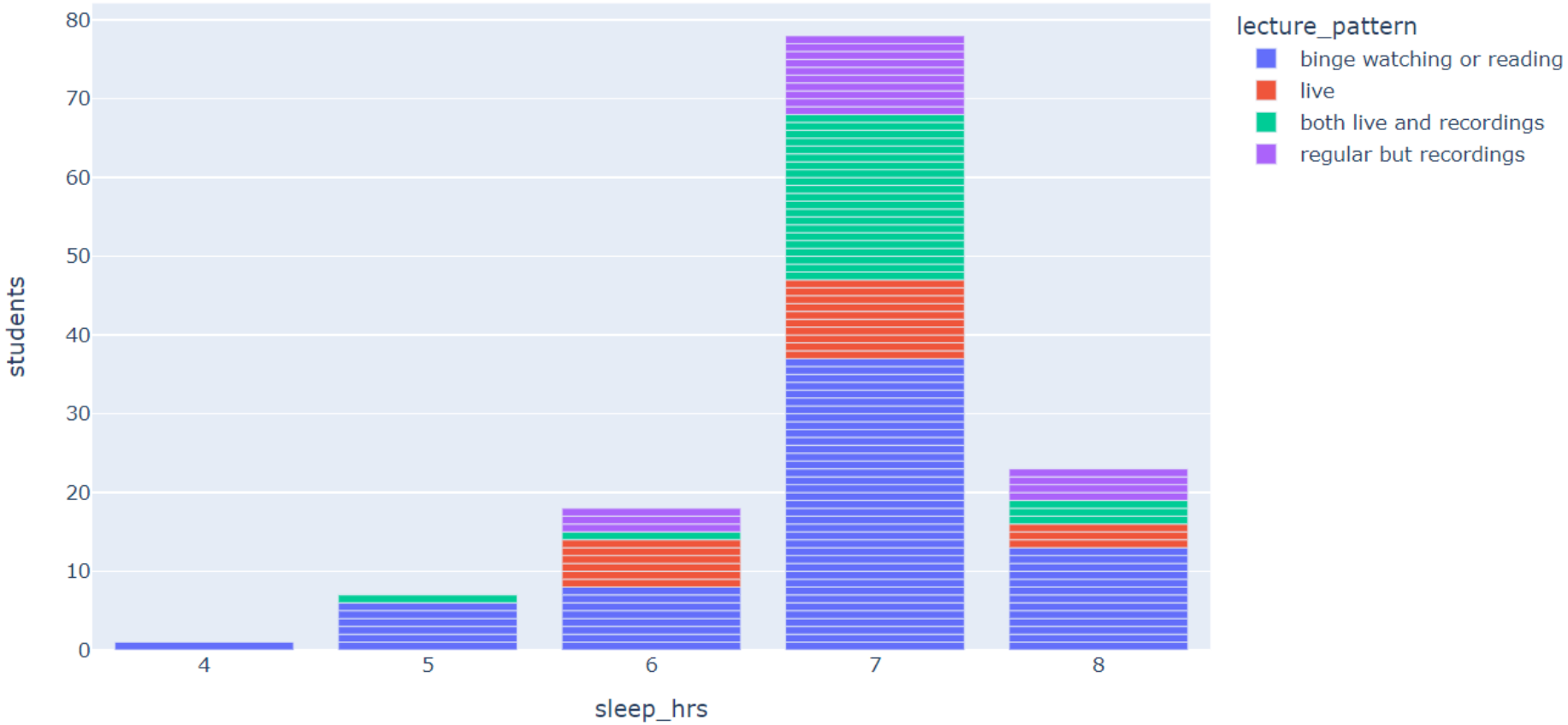

Figure: Segmented Bar Chart for Lecture Watching Patterns

Segmented Bar Chart

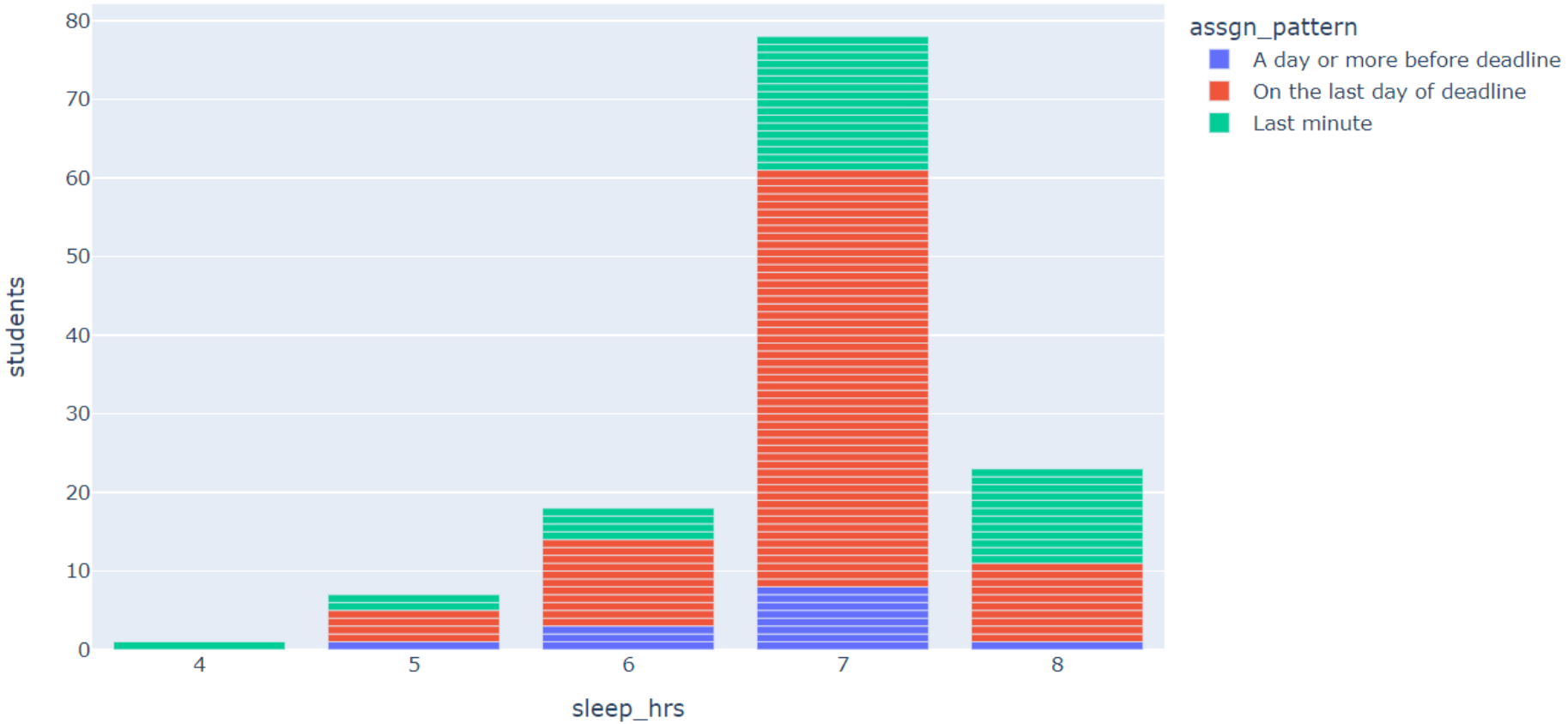

Figure: Segmented Bar Chart for Assignment Submission Patterns

Inferences from Segmented Bar Charts

Inference from Segmented Bar Chart - 1

A higher proportion of students who watch lectures live, live and recordings sleep over 6 hours.

Inference from Segmented Bar Chart - 2

For any quantity of sleep, there are more students who submit assignments on the last day than those who submit earlier.

Sports and Social Activities

Sports and Social Activities

Participating in sports activities can help people enhance their physical and mental health.

For young students, participation in sports activities is crucial to avoid developing negative habits of spending leisure time.

Common sports activities include playing football, basketball, badminton, cricket etc. and physical activities commonly include yoga, cycling, jogging etc.

Studies in Foreign Colleges show that 29 out of the 34 conducted studies concluded that exercise improved sleep duration.

Central Tendencies for Recreational Activities

	How much sleep do you get in a day?
count	39.000000
mean	6.604026
std	1.170625
min	4.871000
25%	6.065000
50%	6.511000
75%	6.744500
max	9.857000

	How much sleep do you get in a day?
count	43.000000
mean	6.526488
std	1.106792
min	4.172000
25%	6.049500
50%	6.411000
75%	6.727500
max	9.419000

Figure: Effect of 1-2 hours, 2-3 hours of Recreational activities on Sleep Hours

	How much sleep do you get in a day?
count	33.000000
mean	7.077273
std	1.554323
min	4.499000
25%	6.149000
50%	6.491000
75%	8.514000
max	9.980000

	How much sleep do you get in a day?
count	17.000000
mean	6.395412
std	1.410636
min	3.667000
25%	5.895000
50%	6.481000
75%	6.959000
max	9.280000

Figure: Effect of 4-5 hours, 6 or more hours of Recreational activities on Sleep Hours

Central Tendencies for Sports and Social Activities

	How much sleep do you get in a day?
count	64.000000
mean	6.406016
std	0.982130
min	4.499000
25%	6.095000
50%	6.311000
75%	6.613000
max	9.419000

	How much sleep do you get in a day?
count	20.000000
mean	6.929300
std	1.592823
min	4.956000
25%	5.903250
50%	6.234500
75%	8.438500
max	9.748000

Figure: Effect of 1-2 hours, 2-3 hours of Sports and Gym activities on Sleep Hours

	How much sleep do you get in a day?
count	5.000000
mean	6.530400
std	0.350119
min	6.005000
25%	6.378000
50%	6.635000
75%	6.728000
max	6.906000

	How much sleep do you get in a day?
count	43.000000
mean	6.959186
std	1.554718
min	3.667000
25%	6.406000
50%	6.716000
75%	8.069000
max	9.980000

Figure: Effect of 4 or more hours, and no Sports and Gym activities on Sleep Hours

Box Plot for Recreational Activities

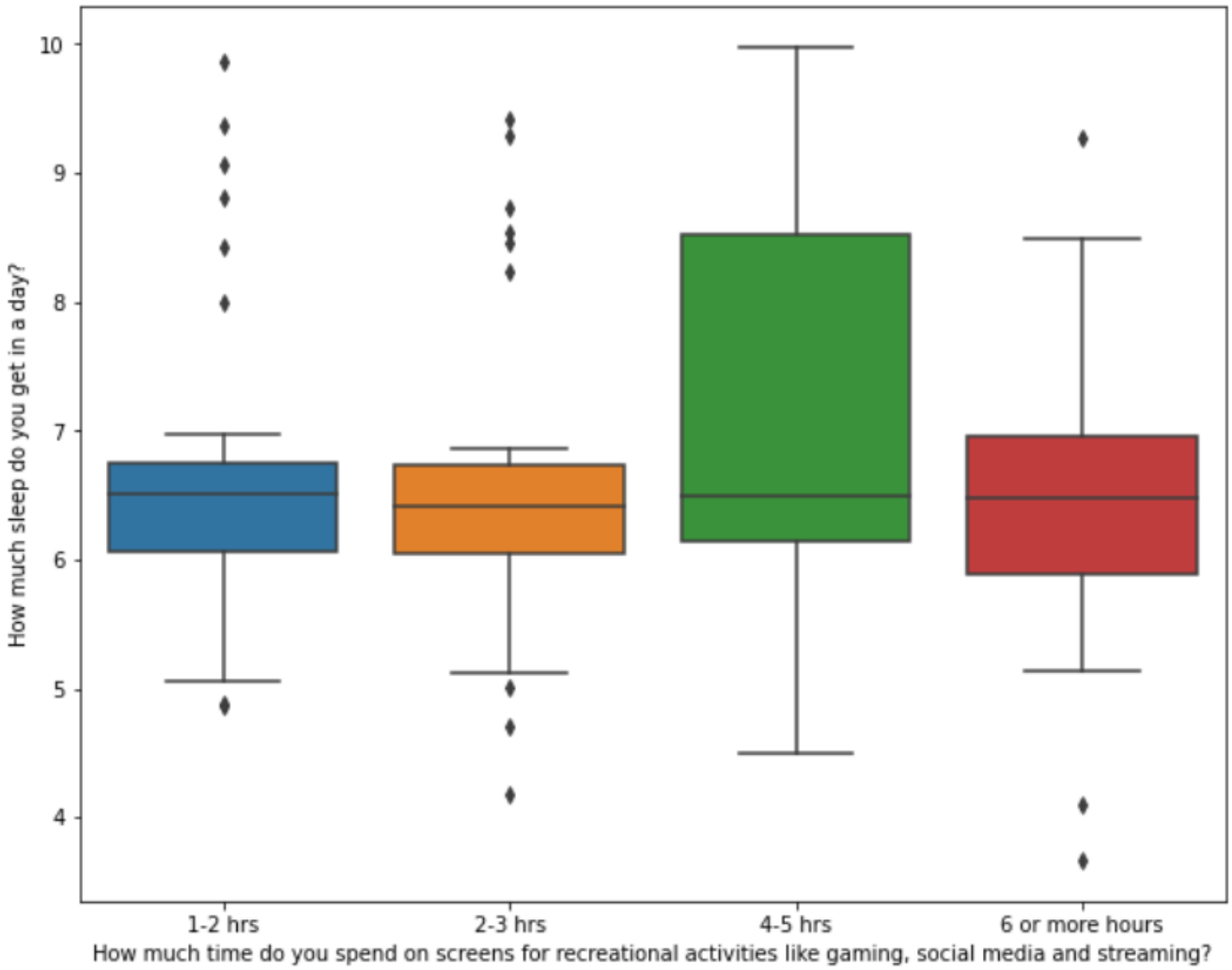

Figure: Effect of Recreational activities on Sleep Hours

Box Plot for Sports and Social Activities

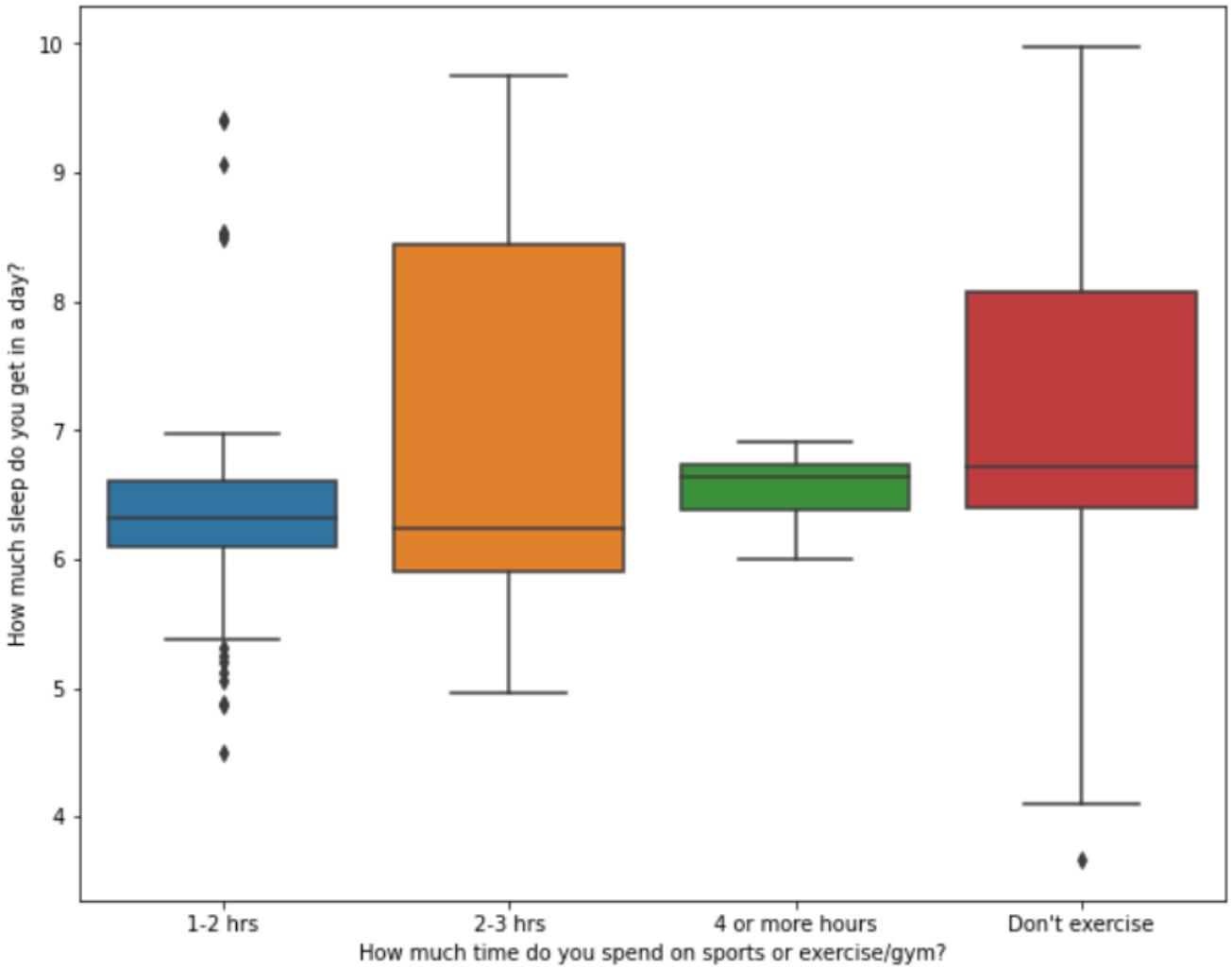

Figure: Effect of Sport activities on Sleep Hours

Inference from Box Plot - 1

Shape: Only category with right skewness is people who spend $4-5$ hours on recreation, so they tend to sleep for less time among themselves.

Center: The median of all the categories are nearly the same.

Spread: The IQR of the category of people who spend $4-5$ hours on recreation is highest, so they sleep less consistently when compared with other categories.

Inference from Box Plot - 2

Shape: Only category with left skewness is people who spend 4 or more hours on sports activities, so they tend to sleep for less time among themselves and exactly the opposite applies for the remaining categories.

Center: The median of people who do not exercise is slightly higher than other categories. The people who do not play sports tend to sleep more than others.

Spread: The IQR of the category of people who spend $4 - 5$ hours on recreation is least, so they sleep more consistently when compared with other categories.

Confidence Interval Estimation

1. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 1-2 hrs on recreational activities (6.10, 6.96)

2. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 2-3 hrs on recreational activities (6.15, 6.84)

3. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 4-5 hrs on recreational activities (6.56, 7.72)

4. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 6 or more hours on recreational activities (5.66, 7.30)

Confidence Interval Estimation

5. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 1-2 hrs on sports/gym (6.09, 6.65)

6. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 2-3 hrs on sports/gym (6.17, 7.80)

7. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends 4 or more hours on sports/gym (6.13, 6.75)

8. We can say with *95 %* confidence that the number of hours of sleep is in the interval who spends Don't exercise on sports/gym (6.48, 7.47)

Hypothesis Testing

Hypothesis

A student that spends 1-3 hours on recreational activities a day on average sleeps less than a student that spends more than 4 hours on the same activities a day.

Calculation

Null Hypothesis

$$(H_0) : \mu_{1-3} - \mu_{>4} \geq 0 \tag{6}$$

Alternate Hypothesis

$$(H_a) : \mu_{1-3} - \mu_{>4} < 0 \tag{7}$$

$$\overline{X}_{1-3} = 6.565374$$

$$n_{1-3} = 39 + 43 = 82$$

$$S_{1-3} = \sqrt{\frac{1.170625^2}{n_{1-2}} + \frac{1.106792^2}{n_{2-3}}} = 0.212241$$

$$\overline{X}_{>4} = 6.7363425$$

$$n_{>4} = 33 + 17 = 50$$

$$S_{>4} = \sqrt{\frac{1.554323^2}{n_{4-5}} + \frac{1.410636^2}{n_{>6}}} = 0.4361906$$

Test Statistic

$$t = \frac{(\overline{X}_{1-3} - \overline{X}_{>4}) - (0)}{\sqrt{\frac{S^2_{1-3}}{n_{1-3}} + \frac{S^2_{>4}}{n_{>4}}}} = \frac{6.565374 - 6.7363425}{\sqrt{0.0045811}} = -2.525969$$

$$c = \frac{\frac{S^2_{1-3}}{n_{1-3}}}{\frac{S^2_{1-3}}{n_{1-3}} + \frac{S^2_{>4}}{n_{>4}}} = 0.16937213$$

$$df = \frac{(n_{1-3} - 1)(n_{>4} - 1)}{(1-c)^2(n_{1-3} - 1) + (c^2)(n_{>4} - 1)} = 69.27787972 \sim 69$$

Rejection Region Approach

Reject H_0 if $t \leq -t_{\alpha,df}$ (let α = 0.05)

$t = -2.525969$, $t_{0.05,69} = 1.6673$

Hence, the hypothesis H_0 is rejected.

Conclusion

We can conclude that a student that spends 1-3 hours on recreational activities a day on average sleeps less than a student that spends more than 4 hours on the same activities a day.

Hypothesis Testing

Hypothesis

A student who does not exercise sleeps more on average than a student who exercises.

Calculation

Null Hypothesis

$$(H_0) : \mu_e - \mu_{de} \geq 0 \tag{8}$$

Alternate Hypothesis

$$(H_a) : \mu_e - \mu_{de} < 0 \tag{9}$$

$$\overline{X}_e = 6.6219053$$

$$n_e = 64 + 20 + 5 = 89$$

$$S_e = \sqrt{\frac{0.982130^2}{n_{1-2}} + \frac{1.592823^2}{n_{2-3}} + \frac{0.350119^2}{n_{>4}}} = 0.40797361$$

$$\overline{X}_d = 6.959186$$

$$n_d = 43$$

$$S_d = 1.554718$$

Test Statistic

$$t = \frac{(\overline{X}_e - \overline{X}_{de}) - (0)}{\sqrt{\frac{S_e^2}{n_e} + \frac{S_{de}^2}{n_{de}}}} = \frac{6.621905333 - 6.959186}{\sqrt{0.0580829}} = -1.399482137$$

$$c = \frac{\frac{S_e^2}{n_e}}{\frac{S_e^2}{n_e} + \frac{S_{de}^2}{n_{de}}} = 0.032197782$$

$$df = \frac{(n_e - 1)(n_{de} - 1)}{(1 - c)^2(n_e - 1) + (c^2)(n_{de} - 1)} \sim 45$$

Rejection Region Approach

Reject H_0 if $t \leq -t_{\alpha,df}$ (let α = 0.05)

$t = -1.399482137$, $t_{0.05,42} = 1.682$

Hence, the hypothesis H_0 is not rejected.

Conclusion

Insufficient evidence to conclude that a student who does not exercise sleeps more on average than a student who exercises.

Contributions of each team member

Prajwaldeep Kamble - Effect of Academics on Response Variables + Google Forms Design

Anurag Gopi - Effect of Academics on Response Variables

Krishna Teja Chilukuri - Effect of Caffeine on Response Variables + Hypothesis Testing of Caffeine

Venkata Raghav Ambati - Effect of Caffeine on Response Variables + Google Forms Distribution

Contributions of each team member

Nikhil Kongara - Study of Response Variables

Karthik Kurugodu - Effect of Sports and Recreational Activities on Response Variables + Hypothesis Testing of Recreational Activities

VKS Deepak Reddy - Effect of Sports and Recreational Activities on Response Variables + Topic of Study

Tata Sai Manoj - Latex Report + Hypothesis Testing of Sport Activities + Image Handling

Sumanth NR - Latex Report + Refining Data + Study Time Analysis

References

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5385214/
2. Research Gate article