STATISTICS PROJECT

SLEEPING HABITS OF STUDENTS

Tanakala Yadunandana –MA22BTECH11018
Kallu Rithika – AI22BTECH11010
Talasani Sri Varsha – AI22BTECH11028
Kunche Aishwarya – AI22BTECH11015
Kota Dhanalakshmi – AI22BTECH11012

Introduction:

The report analyses the sleep patterns and experiences of students. We conducted a survey of 342 students to get a better understanding of sleep habits and various factors affecting the sleep health. The survey investigated various factors affecting student sleep habits, including:

Background Information:

- Degree being pursued by student
- Gender
- Age

Sleep Patterns:

- Sleep Duration on weekdays
- Sleep Duration on weekends

Screen Time:

- Screen time habits of students
- Effect of screen time on consistency of sleep

Sleep Quality:

- Factors affecting sleep quality
- Frequency of missing meals due to irregular sleep

Daytime Functioning:

- Frequency of naps during the day
- Alertness or focus during the class
- Frequency of consuming caffeinated beverages before bed

Sleep and Academics:

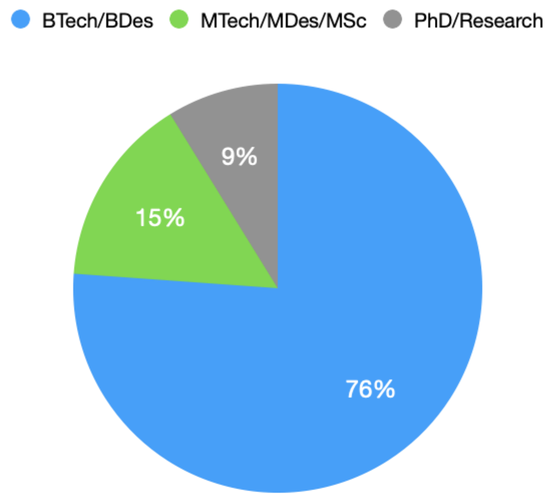
- Impact of irregular sleep on academics

Sleep Satisfaction:

- Consistency of sleep schedule
- Students sleep satisfaction

INTERPRETATION OF DATA:

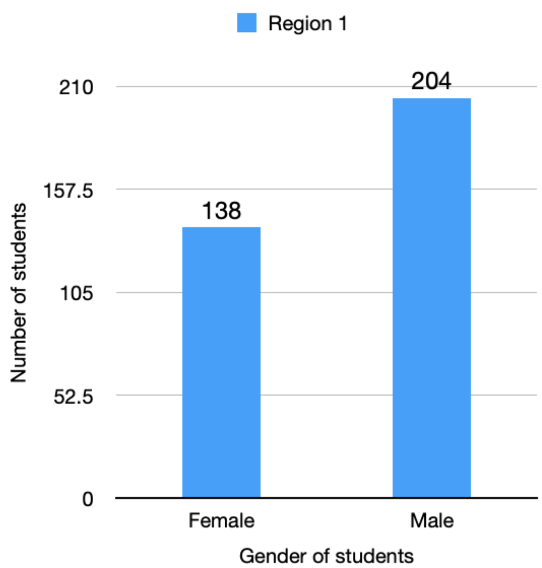
DEGREE BEING PURSUED:



Percentage of students according to degree

From the observations made from the survey ,76% of students are pursuing BTech/BDes, 9% of students are pursuing PhD/Research, 15% students are pursuing MTech/MDes/MSc.

GENDER:



From the graph we can see that, among the students responded 204 are male and 138 are female.

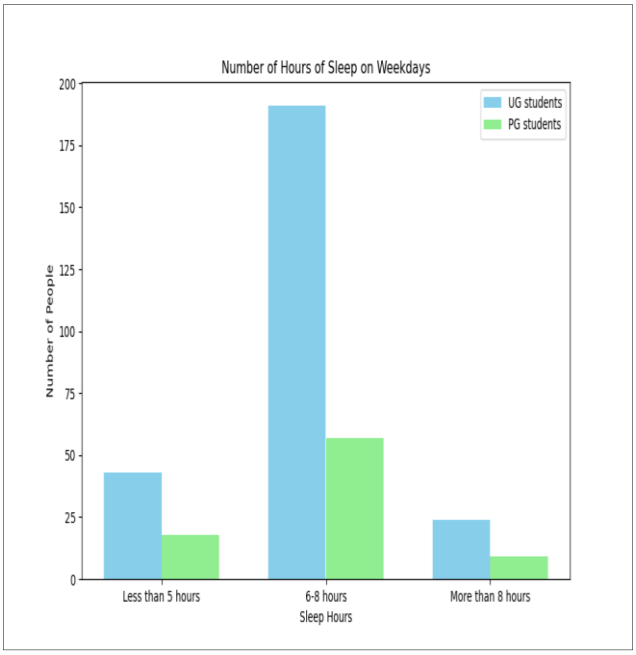
Frequency of Sleeping hours on weekdays:

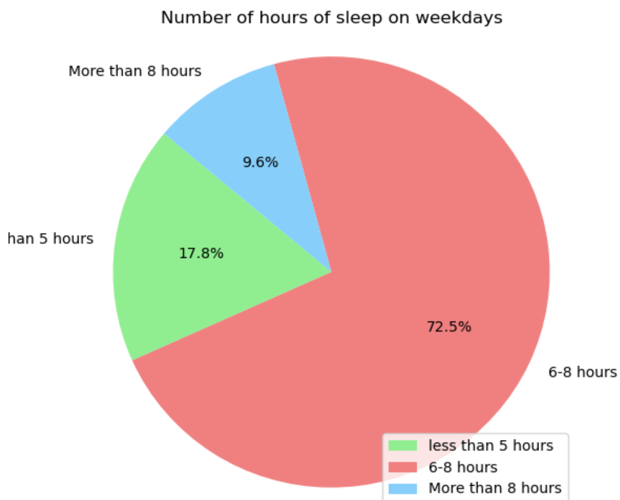
We begin our analysis on Sleeping habits of Students with no. of hours they sleep on weekdays.

Summarizing and Visualizing data:

No. of hours of sleep on weekdays	**Number of people**			**Percentage of People**
	UG students	PG students	Total	
Less than 5 hours	43	18	61	17.84
6-8 hours	191	57	248	72.51
More than 8 hours	24	9	33	9.65

Data Visualization:





Confidence Interval for difference of proportions:

CI for population proportion p_1-p_2 with $(1-\alpha)100\%$ confidence is given by

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}$$

Here,

p1 = proportion of UG students who sleep more than 8-hours

p2 = proportion of PG students who sleep more than 8-hours

From the data above,

$\hat{p}_1$=0.093 n_1 = 258

$\hat{p}_2$=0.11 n_2 = 84

Consider significance value α = 0.05

So, $Z_{\alpha/2}$ = 1.96

Note: Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

$$z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n2}} = 1.96\sqrt{\frac{0.093(1-0.093)}{258}+\frac{0.11(1-0.11)}{84}}$$

For p_1-p_2 95% CI is [-0.087 – 0.076, -0.087 + 0.076] = [-0.163, -0.011]

Hypothesis Testing:

In our sample data, the proportion of UG students who sleep more than 8-hours is lesser than the proportion of PG students who sleep more than 8-hours.

Now we would like to see how our hypothesis holds for the population,

Null Hypothesis: The proportion of UG students who sleep more than 8 hours is more than or equal to proportion of PG students who sleep more than 8 hours

Alternative Hypothesis: The proportion of UG students who sleep more than 8 hours is less than to proportion of PG students who sleep more than 8 hours

Left-tailed Test:

H_0 = p_1-p_2 ≥ 0 H_a = p_1-p_2 < 0

Here, $\hat{p}_1$ = 0.093 and n_1=258

$\hat{p}_2$ = 0.11 and n_2=84

Note: Here, n_1p_1 ≥ 5 , $n_1(1-p_1)$ ≥ 5 , n_2p_2 ≥ 5 and $n_2(1-p_2)$ ≥ 5 so we can proceed with the analysis.

Test – Statistic:

$$Z^* = \frac{(\hat{p}_1-\hat{p}_2)-p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}$$

p-value approach:

$$Z^* = \frac{(0.093-0.11)-0}{\sqrt{\frac{0.093(1-0.093)}{258}+\frac{0.11(1-0.11)}{84}}} = -0.44$$

$p = P(Z \leq Z^*) = 0.3306$

For $\alpha=0.05$, $p>\alpha$

Hence, we can't reject H_0.

Therefore, we do not have enough statistical evidence to conclude that proportion of UG students who sleep more than 8-hours on weekdays is more than proportion of PG students who sleep more than 8-hours on weekdays.

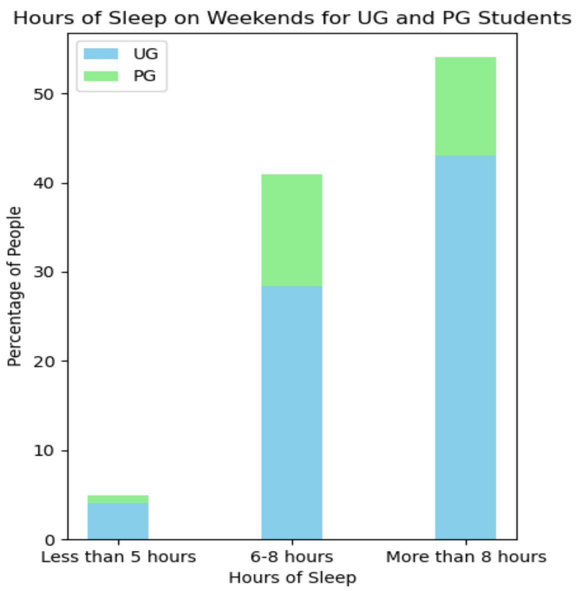
Frequency of Sleeping hours on weekends:

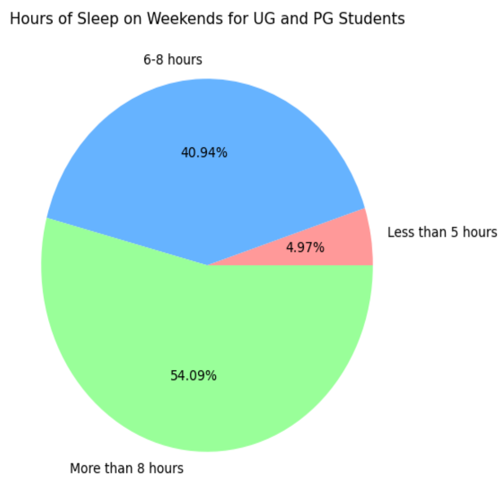
Another way to compare the sleep schedule of students with their busy weekdays is to analyse sleeping hours on weekends.

Summarizing and Visualizing data:

No. of hours sleeping on weekends	Number of people			Percentage of people
	UG	PG	Total	
Less than 5 hours	14	3	17	4.97
6-8 hours	97	43	140	40.94
More than 8 hours	147	38	185	54.09

Data Visualization:





Confidence Interval for proportion:

CI for population proportion p with (1-α)100% is given by,

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

Here n=342,

p_1 = Proportion of students who sleeps less than 5 hours

p_2 = Proportion of students who sleeps 6-8 hours

p_3 = Proportion of students who sleeps more than 8 hours

$\hat{p}_1$=0.05, $\hat{p}_2$=0.409, $\hat{p}_3$=0.541

Consider significance level α=0.05,

So, $Z_{\alpha/2}$ = 1.96

Note: Here $np \geq 5$ and $n(1-p) \geq 5$ for all proportions above and so we can proceed with the analysis.

For p_1 95% CI is [0.027,0.073]

For p_2 95% CI is [0.357,0.461]

For p_3 95% CI is [0.488,0.594]

Hypothesis Testing:

In our sample data, the proportion of UG students who sleep 6-8 hours on weekdays is more than UG students who sleep 6-8 hours on weekends.

Now we would like to see how our hypothesis holds for the population,

p1 = proportion of UG students who sleep 6-8 hours on weekdays

p2 = population of UG students who sleep 6-8 hours on weekends

Null Hypothesis: The proportion of UG students who sleep 6-8 hours on weekdays is less than or equal to UG students who sleep 6-8 hours on weekends.

Alternative Hypothesis: The proportion of UG students who sleep 6-8 hours on weekdays is more than UG students who sleep 6-8 hours on weekends.

Right-tailed test:

$H_0 = p_1 - p_2 \leq 0$ $\qquad$ $H_a = p_1 - p_2 > 0$

Here, $\hat{p}_1$= 0.74 and n1=258

$\hat{p}_2$= 0.376 and n2=258

Note: Here, $n_1p_1 \geq 5$, $n_1(1-p_1) \geq 5$, $n_2p_2 \geq 5$ and $n_2(1-p_2) \geq 5$ so we can proceed with the analysis.

Test Statistic:

$$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}$$

p-value Approach:

$$Z^* = \frac{(0.74-0.376)-0}{\sqrt{\frac{0.74(1-0.74)}{258}+\frac{0.376(1-0.376)}{258}}} = 8.947$$

p = $P(Z \geq Z^*) = P(Z \geq 8.947) = 0$

For $\alpha=0.05$, $p<\alpha$

Hence, we can reject H_0.

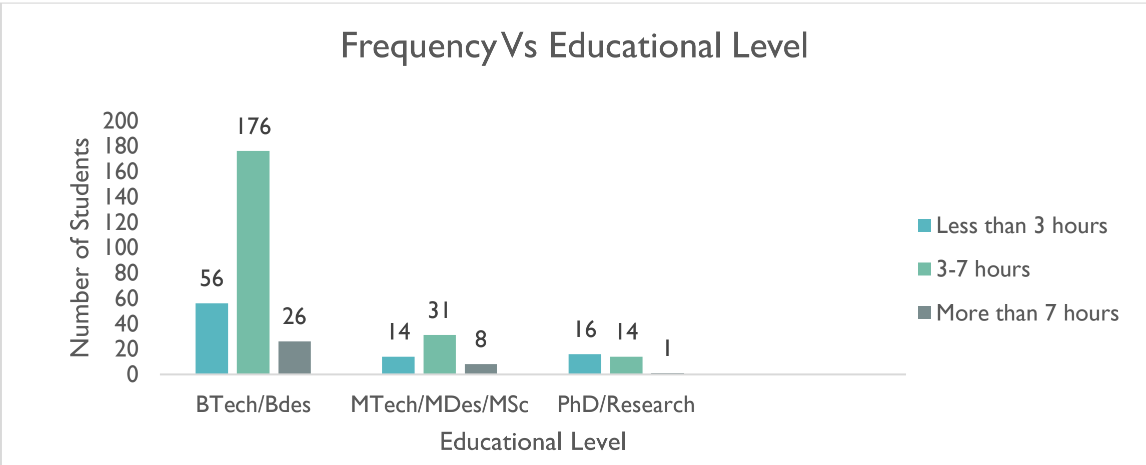
Therefore, we have sufficient evidence to show that proportion of UG students who sleep 6-8 hours on weekdays is more than proportion of UG students who sleep 6-8 hours on weekends.

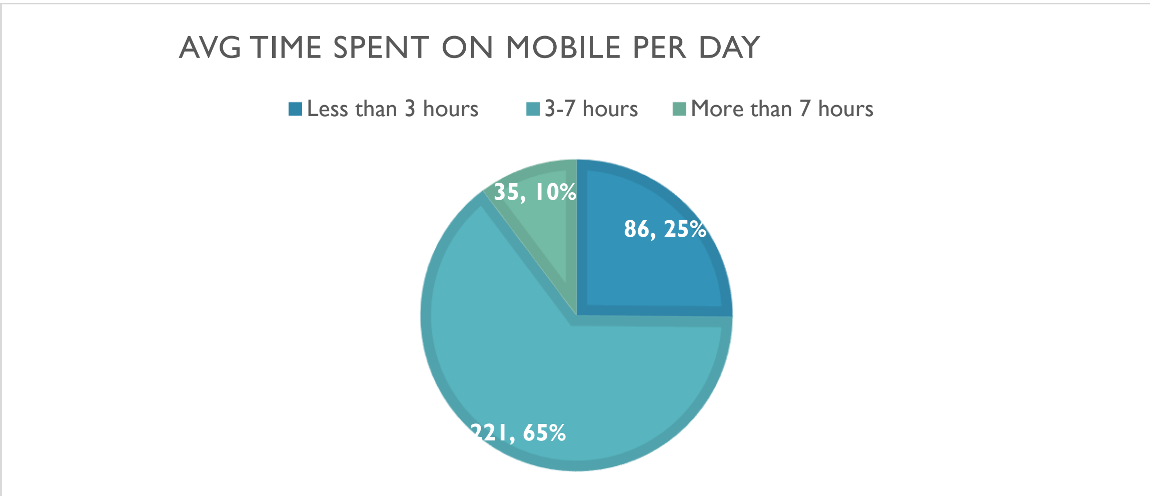
SCREEN TIME HABITS OF STUDENTS:

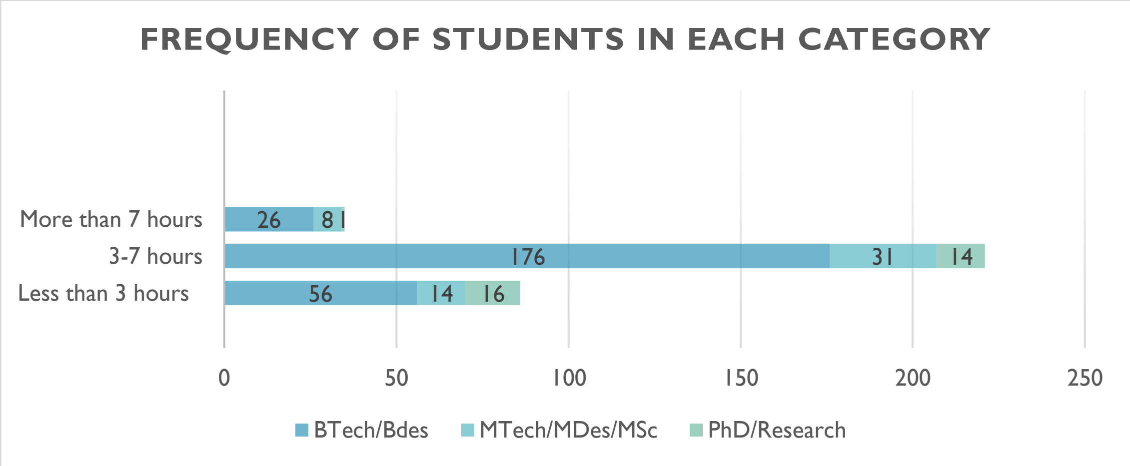
Excessive screen time has been linked to various health issues, including eye strain, sleep disturbances. In this section we analyse the screen time of students on their phones.

Interpretation and Visualisation of data:

Average time spent on mobile per day	Number of People			
	BTech/BDes	MTech/MDes/MSc	PhD/Research	Total
Less than 3 hours	56	14	16	86
3-7 hours	176	31	14	221
More than 7 hours	26	8	1	35
Total	258	53	31	342







Estimation of population proportions:

Based on the above graph we estimate the following:

p_1 - Proportion of students who use mobile less than 3 hours per day.

p_2 - Proportion of students who use mobile in between 3-7 hours per day.

p_3 – Proportion of students who use mobile more than 7 hours per day.

Here, Sample Size =342.

$\hat{p}_1 = \frac{86}{342} = 0.25, \; \hat{p}_2 = \frac{221}{342} = 0.65, \; \hat{p}_3 = \frac{35}{342} = 0.10$

Therefore,

25% of students use mobile less than 3 hours per day.

65% of students use mobile for 3-7 hours per day.

10% of students use mobile more than 7 hours per day.

Confidence Interval Estimation for proportion:

$(1-\alpha)100\%$ Confidence interval for proportion p is

$$\boxed{\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}}$$

Here, we calculate the Confidence interval for proportion for all three categories.

Take significance level $\alpha = 0.05$

So, $Z_{\alpha/2} = 1.96$, n=342.

Note: Here $np \geq 5$ and $n(1-p) \geq 5$ for all the proportions above and so we can proceed with the analysis.

For p_1 95% CI is $\left[0.25 - 1.96\sqrt{\frac{0.25(0.75)}{342}}, 0.25 + 1.96\sqrt{\frac{0.25(0.75)}{342}}\right]$

= [0.227, 0.273]

For p_2 95% CI is $\left[0.65 - 1.96\sqrt{\frac{0.65(0.35)}{342}}, 0.65 + 1.96\sqrt{\frac{0.65(0.35)}{342}}\right]$

= [0.624, 0.676]

For p_3 95% CI is $\left[0.10 - 1.96\sqrt{\frac{0.1(0.9)}{342}}, 0.10 + 1.96\sqrt{\frac{0.1(0.9)}{342}}\right]$

= [0.084, 0.116]

Hypothesis Testing:

In our sample data, the proportion of undergraduate students who use phone for 3-7 hours a day is greater than the postgraduate students who use their phones for 3-7 hours a day. So, we would like to see how likely this is to hold true for the population.

Here,

p_1 - Proportion of undergraduate students who use phone for 3-7 hours a day

p_2 - Proportion of postgraduate students who use their phones for 3-7 hrs a day

Right tailed Hypothesis:

$Ho = p1 - p2 \leq 0$ $Ha = p1 - p2 > 0$

Average time spent on mobile per day	Number of People		
	Under graduate	Post graduate	Total
Less than 3 hours	56	30	86
3-7 hours	176	45	221
More than 7 hours	26	9	35
Total	258	84	342

Based on the above table,

$\hat{p}_1$ = 176/258 = **0.68**, n1=258

$\hat{p}_2$ = 45/84 = **0.54**, n2=84

Since, $n_1p_1 \geq 5$, $n_1(1 - p_1) \geq 5$, $n_2p_2 \geq 5$ and $n_2(1 - p_2) \geq 5$ and so we can proceed with the analysis.

Consider $\alpha = 0.05$;

<u>Test Statistic:</u>

$$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}$$

<u>Rejection Region Approach:</u>

$$Z^* = \frac{0.68-0.54}{\sqrt{\frac{0.68(1-0.68)}{258} + \frac{0.54(1-0.54)}{84}}} = 2.27 \qquad Z_{0.05} = 1.645$$

Since, $Z^* > Z_{0.05}$ we reject H_o at the confidence level of 95%.

We therefore conclude that the proportion of undergraduate students who use phone for 3-7 hours a day is greater than the postgraduate students who use their phones for 3-7 hours a day.

DOES SCREEN TIME AFFECT THE SLEEP CONSISTENCY?

Continuing our analysis, in this section we examine how screen time affect the sleep consistency of students. we aim to ascertain if a meaningful relationship exists between these two variables by conducting the Chi-square test of independence.

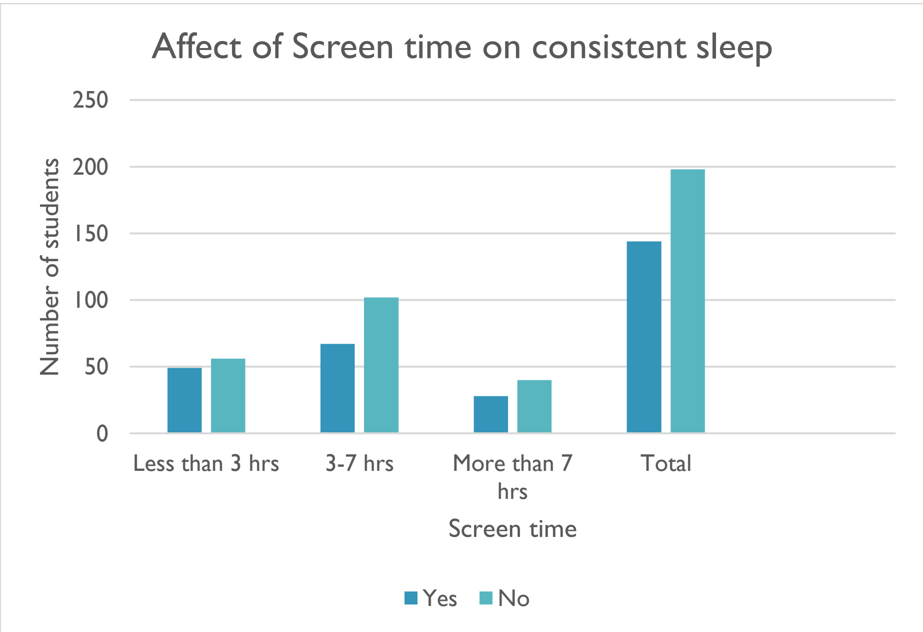
Interpretation and Visualisation of data:

Contingency table of sleep consistency and screen time:

Sleep consistency of students	Average screen time per day			
	Less than 3 hrs	3-7 hrs	More than 7 hrs	Total
Yes	49	67	28	144
No	56	102	40	198
Total	105	169	68	342

Contingency table based on overall total:

Sleep consistency of students	Average screen time per day			
	Less than 3 hrs	3-7 hrs	More than 7 hrs	Total
Yes	14.33%	19.59%	8.19%	42.11%
No	16.37%	29.82%	11.70%	57.89%
Total	30.70%	49.41%	19.89%	100%



Chi square test of independence:

Here,

H_0 : There is no association between screen time and sleep consistency of students.

H_a : There is an association between screen time and sleep consistency of students.

Expected frequency for each cell in a contingency table is calculated by:

$$E_{ij} = \frac{(\text{Row Total}_i) \times (\text{Column Total}_j)}{\text{Grand Total}}$$

$$E_{\text{yes, less than 3 hours}} = \frac{(Total\ of\ yes)\times(Total\ of\ less\ than\ 3\ hrs)}{Grand\ Total} = \frac{(144)\times 105}{342} = 44.21$$

$$E_{\text{yes , 3-7 hours}} = \frac{(Total\ of\ yes)\times(Total\ of\ 3-7\ hrs)}{Grand\ Total} = \frac{(144)\times(169)}{342} = 71.16$$

$$E_{\text{yes , more than 7 hours}} = \frac{(Total\ of\ yes)\times(Total\ of\ more\ than\ 7\ hrs)}{Grand\ Total} = \frac{(144)\times(68)}{342} = 28.63$$

$$E_{\text{no, less than 3 hours}} = \frac{(Total\ of\ no)\times(Total\ of\ less\ than\ 3\ hrs)}{Grand\ Total} = \frac{(198)\times(105)}{342} = 60.79$$

$$E_{\text{no , 3-7 hours}} = \frac{(Total\ of\ no)\times(Total\ of\ 3-7\ hrs)}{Grand\ Total} = \frac{(198)\times(169)}{342} = 97.84$$

$$E_{\text{no , more than 7 hours}} = \frac{(Total\ of\ no)\times(Total\ of more\ than\ 7\ hrs)}{Grand\ Total} = \frac{(198)\times(68)}{342} = 39.37$$

Table with expected frequencies:

Sleep consistency	Less than 3 hrs	3-7 hrs	More than 7 hrs
Yes	44.21	71.16	28.63
No	60.79	97.84	39.37

Here, 1)Number of rows r=2, number of columns c=3.

2)Degrees of freedom df = (r-1)(c-1)=(1)(2)= 2.

3)Considering α = 0.05, $\chi^2_{\alpha,(r-1)(c-1)}$ = 5.991. Critical value =5.991.

4) χ^2 Test Statistic is : $\boxed{\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}}$

where $O_1, O_2, \ldots O_6$: the observed counts for each cell.

$E_1, E_2, \ldots E_6$: the respective expected counts for each cell.

For our observed data, this calculation is:

$$\chi^{2^*} = \frac{(49-44.21)^2}{44.21} + \frac{(67-71.16)^2}{71.16} + \frac{(28-28.63)^2}{28.63} + \frac{(56-60.79)^2}{60.79} + \frac{(102-97.84)^2}{97.84} + \frac{(40-39.37)^2}{39.37}$$

$$\chi^{2^*} = \frac{22.94}{44.21} + \frac{17.31}{71.16} + \frac{0.40}{28.63} + \frac{22.94}{60.79} + \frac{17.31}{97.84} + \frac{0.40}{39.37}$$

$$\chi^{2^*} = 1.34$$

As1.34<5.991 that is the calculated test statistic χ^2(1.34) is less than the critical value 5.991, we fail to reject the null hypothesis.

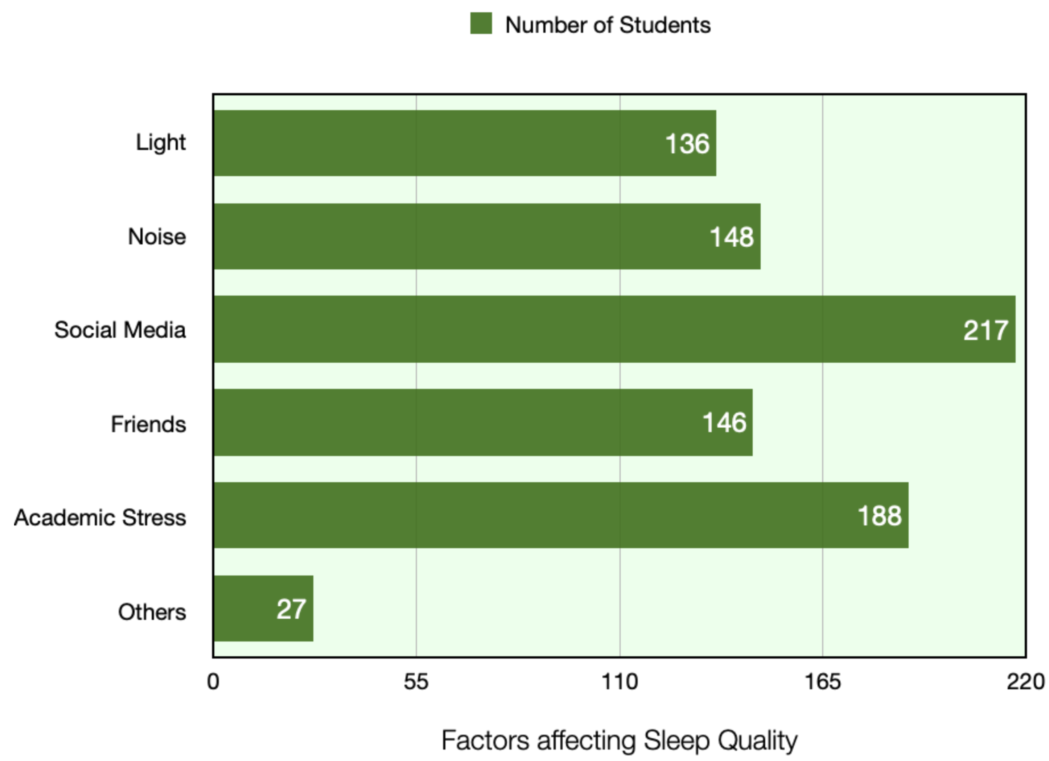
p-value:

p-value is 0.51.

Since the p-value (0.51) is greater than our alpha value (0.05), we fail to reject the null hypothesis of our test.

There is no significant evidence at the 5% significance level to suggest an association between screen time and sleep consistency of students.

FACTORS AFFECTING SLEEP QUALITY:



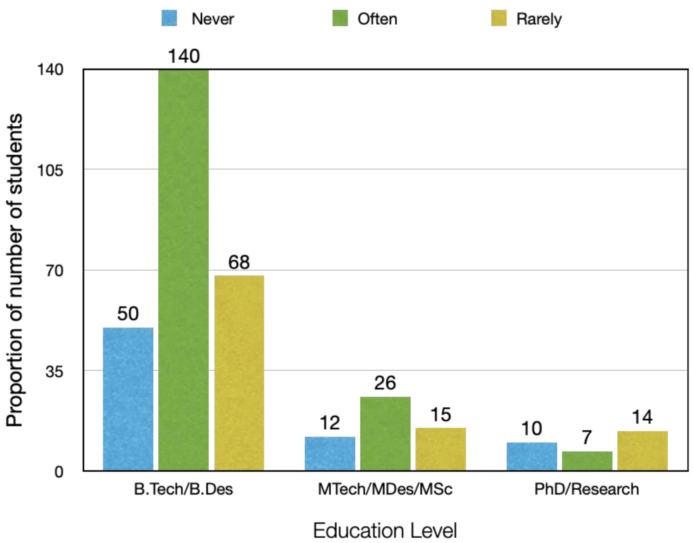
Among 342 students, Sleep quality of 217 students i.e. 63.4% of the students are affected by social media.
Sleep quality of 188 students i.e. 54.97% of students are affected by Academic stress.
Sleep quality of 148 students i.e. 43.27 % of students are affected by Noise.
Sleep quality of 146 students i.e. 42.69% of students are affected by Friends.
Sleep quality of 136 students i.e. 39.76% of students are affected by Light.
Sleep quality of 27 students i.e 7.89% of students are affected by other factors.

FREQUENCY OF MISSING MEALS:

In this section we will observe how irregular sleep affects the frequency of ~~of students~~ missing the meals by students.

Interpretation and Visualisation of data:

Frequency of Missing Meals due to irregular sleep	Number of People			
	BTech/BDes	MTech/MDes/MSc	PhD/Research	Total
Never	50	12	10	72
Often	140	26	7	173
Rarely	68	15	14	97
Total	258	53	31	342



Confidence Interval for proportion:

(1-α)100% Confidence interval for proportion $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}$$

Let p_1 = Proportion of BTech/BDes students who miss meals often due to irregular sleep

p_2 = Proportion of MTech/MDes/MSc students who miss meals often due to irregular sleep

$\hat{p}_1 = \frac{140}{258} = 0.542$ $n_1 = 258$

$\hat{p}_2 = \frac{26}{53} = 0.491$ $n_2 = 53$

Take significance level $\alpha = 0.05$

So, $Z_{\alpha/2} = 1.96$

$$z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n2}} = 1.96\sqrt{\frac{0.542(1-0.542)}{258}+\frac{0.491(1-0.491)}{53}} = 0.1476$$

Note: Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

95% CI for estimation of $p_1 - p_2$ is [0.05 – 0.1476 , 0.05 + 0.1476] = [-0.0966 , 0.1986]

Hypothesis Testing:

Null Hypothesis (H_0): The proportion of students pursuing BTech/BDes who miss meals often due to irregular sleep is less than or equal to the proportion of MTech/MDes/MSc students who miss meals often due to irregular sleep.

Alternate Hypothesis (H_a): The proportion of students pursuing BTech/BDes who miss meals often due to irregular sleep is more than the proportion of MTech/MDes/MSc students who miss meals often due to irregular sleep.

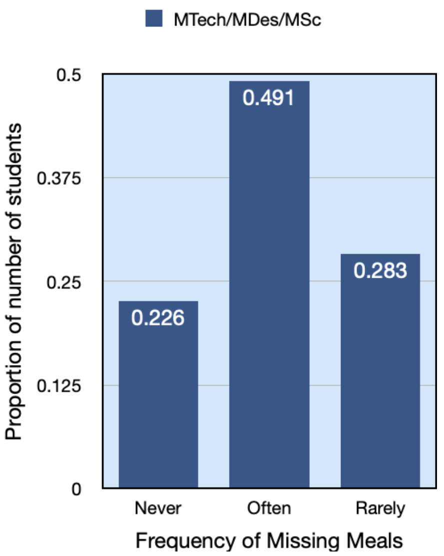
Right-tailed test for two proportions:

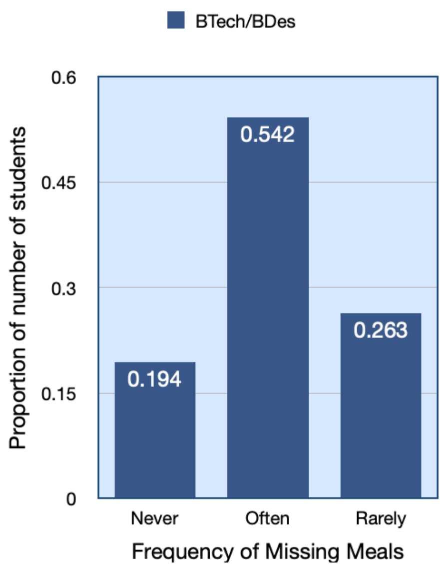
$H_0: p_1 - p_2 \leq 0$ $H_a: p_1 - p_2 > 0$

Note: Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

$\hat{p}_1 = \frac{140}{258} = 0.542$ $n_1 = 258$

$\hat{p}_2 = \frac{26}{53} = 0.491$ $n_2 = 53$





Considering $\alpha = 0.05$;

<u>Test Statistic:</u>

$$\boxed{Z^{*} = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}}$$

<u>Note:</u> Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

<u>Rejection Region Approach:</u>

$$Z^{*} = \frac{0.542 - 0.491}{\sqrt{\frac{0.542(1-0.542)}{258} + \frac{0.491(1-0.491)}{53}}} = 0.676 \qquad Z_{0.05} = 1.645$$

$$Z^{*} < Z_{0.05}$$

Since the value of test statistic (Z^{*}) does not fall in the rejection region, we fail to reject H_0.

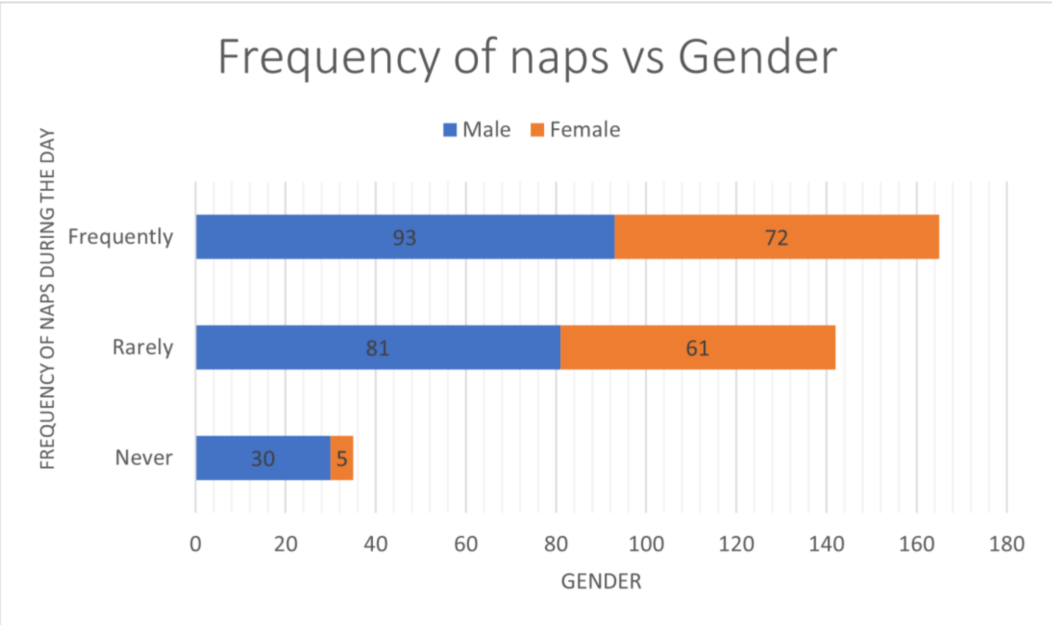
Therefore, we do not have sufficient statistical evidence to reject the null hypothesis. Thus, the sample data do not provide strong enough evidence to conclude that the proportion of students pursuing BTech/BDes who miss meals often due to irregular sleep is less than or equal to the proportion of MTech/MDes/MSc students who miss meals often due to irregular sleep.

FREQUENCY OF NAPS VS GENDER:

In this section we will analyse the frequency table for the data is as follows:

Nap Frequency	Gender		
	Male	Female	Total
Never	30	5	35
Rarely	81	61	142
Frequently	93	72	165
Total	**204**	**138**	**342**

Visualization of Data:



Confidence Interval Estimation:

Given Sample size n = 342

100(1-α)% Confidence Interval for the difference in two population proportions p1-p2 is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}$$

Consider α=0.05.

$Z_{0.025}$ = 1.96

From the data,

$\hat{p}_1 = \frac{81}{204} = 0.397$ $n_1 = 204$ $\hat{p}_2 = \frac{61}{138} = 0.442$ $n_2 = 138$

Note: Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

Substituting in the formula we get 95% confidence interval for p1-p2 ,

CI for estimation of p1-p2 $= (0.397 - 0.442) \pm 1.96\sqrt{\frac{0.397\times(0.603)}{204} + \frac{0.442\times(0.558)}{138}}$

$= (-0.045) \pm 0.1066 \quad = \quad [-0.1516 , 0.0616]$

Hypothesis Testing:

Let p_1 = Proportion of male who rarely nap during the day.

P_2 = Proportion of female who rarely nap during the day.

Considering α = 0.05

Null Hypothesis (H_0): The proportion of individuals who rarely nap is lower in males compared to females.

Alternative Hypothesis (H_a): The proportion of individuals who rarely nap is higher in males compared to females.

Right Tailed Test:

$H_0 : p_1\text{-}p_2 \leq 0$ $H_a : p_1\text{-}p_2 > 0$

From the data,

$\hat{p}_1 = \frac{81}{204} = 0.397$ $n_1 = 204$

$\hat{p}_2 = \frac{61}{138} = 0.442$ $n_2 = 138$

Test statistic:

$$\boxed{Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}}$$

p_0 is the specified value

Note: Here np ≥ 5 and n(1-p) ≥ 5 for all proportions above and so we can proceed with the analysis.

Rejection Region Approach:

The value of the test statistic is $Z^* = \frac{(0.397 - 0.442) - 0}{\sqrt{\frac{0.397\times(0.603)}{204} + \frac{0.442\times(0.558)}{138}}}$

$Z^* = -0.387$

For α = 0.05, we have $z_{0.05} = 1.645$, so the critical value is 1.645.

Since test statistic is less than the critical z-value, we fail to reject the null hypothesis.

Thus the sample data do not provide strong enough evidence to conclude that the proportion of individuals who rarely nap is lower in males compared to females.

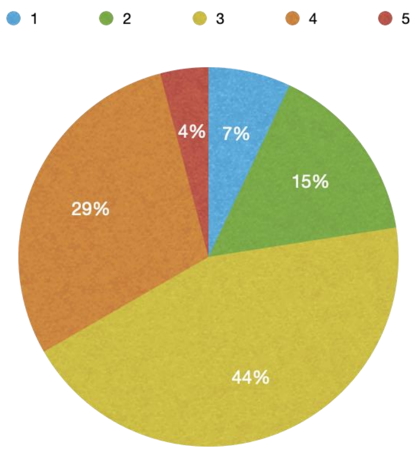
ALERTNESS OR FOCUS DURING THE CLASS:

In this section, we will observe the level of attention or focus students have during classes.

Let us observe the data categorized according to their education level.

Interpretation and Visualisation of data:

Education Level	Level of Attention					
	1	2	3	4	5	Total
BTech/BDes	23	45	118	64	8	258
MTech/MDes/Msc	1	6	24	18	4	53
PhD/Research	0	2	9	18	2	31
Total	24	53	151	100	14	342



Pie chart representing percentage of students with different level of attention during the class

Confidence Interval:

$(1-\alpha)100\%$ CI for population mean μ when σ is unknown is $\boxed{\overline{X} \pm t_{a/2}\left(\frac{S}{\sqrt{n}}\right)}$

$\overline{X}$ - Sample mean S – Sample Variance a - Significance level

Here, Sample size n = 342

Sample mean $\overline{x} = \frac{24+(53\times2)+(151\times3)+(100\times4)+(14\times5)}{342} = 3.078$

Sample variance $S^2 = 0.941$

Take significance level α = 0.05

In this case degrees of freedom df = n−1 = 341

So $t_{\alpha/2,n-1}$ = 1.9669

$t_{a/2,n-1}(\frac{S}{\sqrt{n}})$ = 1.9669 × $\frac{\sqrt{0.941}}{\sqrt{342}}$ = 0.103

For μ 95% CI is [3.078 – 0.103 , 3.078 + 0.103] = [2.975 , 3.181]

Hypothesis Testing:

Null Hypothesis(H_0): The mean of the level of attention of students in class is less than or equal to 3

Alternate Hypothesis(H_a): The mean of the level of attention of students during class is greater than 3

Let μ – Mean of the level of attention of students

Right tailed test for mean:

$H_0 : \mu \leq 3$ $\quad\quad$ $H_a : \mu > 3$

Here take significance value α = 0.05

$\bar{x}$ = 3.07

S = 0.97

n = 342

Test Statistic: $\boxed{t^* = \frac{\bar{x} - p_0}{S/\sqrt{n}}}$

p-value Approach:

Here p_0 = 3

$t^* = \frac{3.07 - 3}{0.97/\sqrt{342}} = \frac{0.07}{0.052}$ = 1.346

In this case,

p-value = $P(t \geq t^*) = P(t \geq 1.346)$ = 0.0896

p-value = 0.0896 > 0.05 =α

Hence, we fail to reject H_0

We do not have enough statistical evidence to conclude that mean of level attention/focus of students in class is greater than 3

DOES CAFFIENE CONSUMPTION AFFECT WEEKDAY SLEEP?

In this section we will be study the impact of caffeine consumption before bed (within 4 hours) on weekday sleep duration. We analyse data collected from individuals to determine whether there is association between these two factors by doing CHI SQUARE TEST OF INDEPENDENCE.

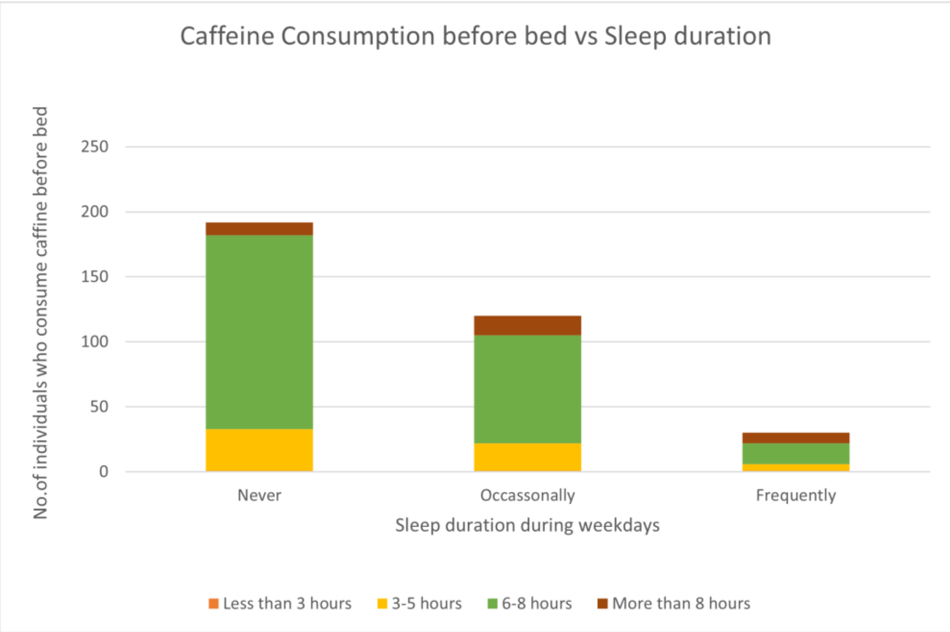
Contingency table from the data collected:

Caffeine consumption Before Bed	Sleep duration during weekdays				Total
	Less than 3 hours	3-5 hours	6-8 hours	More than 8 hours	
Never	1	32	149	10	192
Occassionally	1	21	83	15	120
Frequently	1	5	16	8	30
Total	**3**	**58**	**248**	**33**	342

Caffeineconsumption Before Bed	Sleep duration during weekdays				Total(%)
	Less than 3 hours(%)	3-5 hours(%)	6-8 hours(%)	More than 8 hours(%)	
Never	0.29%	9.36%	43.57%	2.92%	56.14%
Occassionally	0.29%	6.14%	24.27%	4.38%	35.08%
Frequently	0.29%	1.46%	4.67%	2.33%	8.75%
Total	**0.87%**	**16.96%**	**72.51%**	**9.63%**	**100%**

Chi square test of independence determines if there is a significant association between two categorical variables.

Visualization of data:



Hypothesis Testing:

Null Hypothesis (H0): There is no association between caffeinated beverage consumption before going to bed and weekday sleep duration.

Alternative Hypothesis (Ha): There is an association between caffeinated beverage consumption before going to bed and weekday sleep duration.

Expected Frequencies are calculated as:

$$E_{\text{Never , Less than 3 hours}} = \frac{(Total\ of\ never)\times(Total\ of\ less\ than\ 3\ hrs)}{Grand\ Total} = \frac{(192)\times(3)}{342} = 1.68$$

$$E_{\text{Never , 3-5 hours}} = \frac{(Total\ of\ never)\times(Total\ of\ 3-5\ hrs)}{Grand\ Total} = \frac{(192)\times(58)}{342} = 32.56$$

$$E_{\text{Never , 6-8 hours}} = \frac{(Total\ of\ never)\times(Total\ of\ 6-8\ hrs)}{Grand\ Total} = \frac{(192)\times(248)}{342} = 139.22$$

$$E_{\text{Never , More than 8 hours}} = \frac{(Total\ of\ never)\times(Total\ of\ more\ than\ 8\ hrs)}{Grand\ Total} = \frac{(192)\times(33)}{342} = 18.52$$

Table with expected frequencies:

Caffeine consumption Before Bed	Less than 3 hrs	3-5 hrs	6-8 hrs	More than 8 hrs
Never	1.68	32.56	139.22	18.52
Occassonally	1.052	20.35	87.017	11.57
Frequently	0.263	5.087	21.75	2.89

χ^2 Test Statistic is :

$$\boxed{\chi^2 = \sum_{i=1}^{12} \frac{(O_i - E_i)^2}{E_i}}$$

$O_1, O_2, \ldots O_{12}$: the observed counts for each cell.

$E_1, E_2, \ldots E_{12}$: the respective expected counts for each cell.

For our observed data, this calculation is:

$$c^2 = \frac{(1-1.68)^2}{1.68} + \frac{(32-32.56)^2}{32.56} + \frac{(149-139.22)^2}{139.22} + \frac{(10-18.52)^2}{18.52} + \frac{(1-1.052)^2}{1.052} + \frac{(21-20.35)^2}{20.35} + \frac{(83-87.017)^2}{87.017} + \frac{(15-11.57)^2}{11.57} + \frac{(1-0.263)^2}{0.263} + \frac{(5-5.087)^2}{5.087} + \frac{(16-21.75)^2}{21.75} + \frac{(8-2.89)^2}{2.89}$$

$$c^2 = \frac{0.462}{1.68} + \frac{0.3136}{32.56} + \frac{95.64}{139.22} + \frac{72.59}{18.52} + \frac{0.0027}{1.052} + \frac{0.4225}{20.35} + \frac{16.13}{87.017} + \frac{11.76}{11.57} + \frac{0.543}{0.263} + \frac{0.0075}{5.087} + \frac{33.06}{21.75} + \frac{26.12}{2.89}$$

$\chi^2 = 18.73$

Consider significance level alpha = 0.05

Degrees of freedom: df = (3-1) × (4-1) = 6

Critical Value:

Comparing the test statistic χ^2 of 18.73 to the critical value in the chi square distribution table. The critical value is the value in the table that aligns with a significance of 0.05 and a degree of freedom of 6. This number turns out to be 12.592.

Since our test statistic χ^2(18.73) is larger than the critical value 12.592, we reject the null hypothesis of our test.

There is significant evidence level to suggest an association between caffeine consumption before going to bed and sleep duration during weekdays at alpha level 0.05.

p-value:

p-value is 0.0046.

Since the p-value (0.0046) is less than our alpha value (0.05), we reject the null hypothesis of our test.

There is significant evidence at the 0.05 significance level to suggest an association between caffeine consumption before going to bed and sleep duration during weekday.

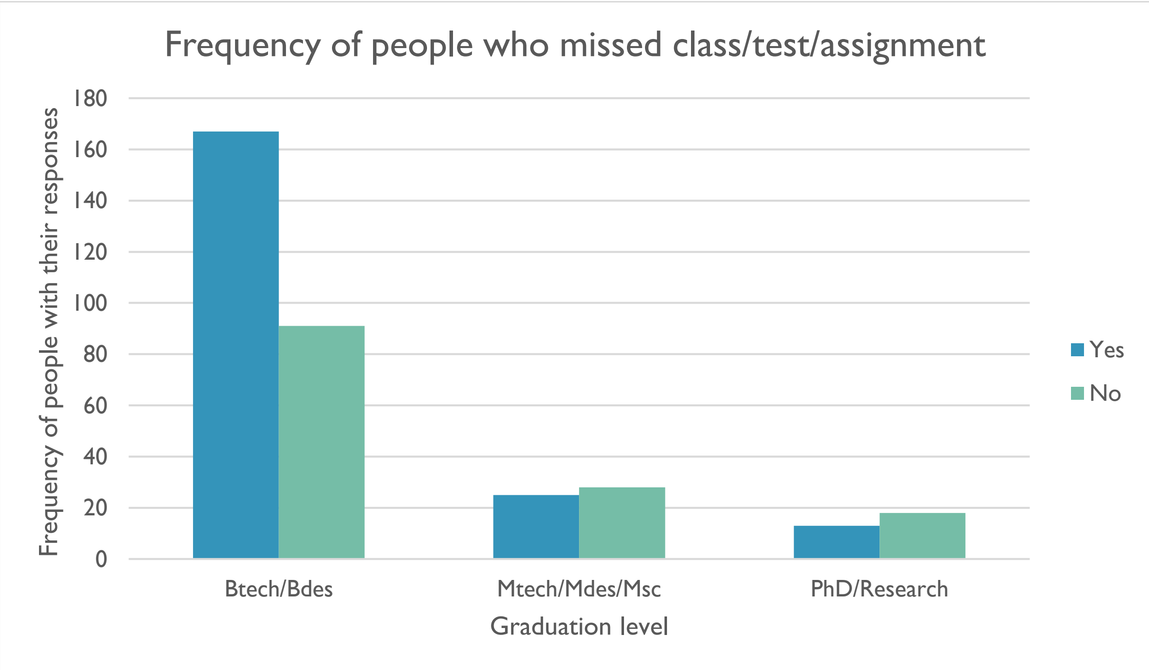
IMPACT OF IRREGULAR SLEEP ON ACADEMICS:

Here we are going to visualize the frequency of people who missed a class/test/assignment with respect to their graduate level.

The frequency table is below:

Graduate level of students	Frequency of people who missed a class/test/assignment		Total
	Yes	No	
BTech/BDes	167	91	258
MTech/MDes/MSc	25	28	53
PhD/Research	13	18	31
Total	205	137	342

Visualization of the above interpreted data:



Calculating the Confidence Interval:

Here we are going to calculate the CI using the proportion of the responses Yes and No.

The proportion table is as below:

Response	Number of people	Proportion
Yes	205	0.599
No	137	0.401

Formula for Confidence Interval with respect to proportions:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

Upon evaluating the confidence intervals for the proportion while taking $\alpha = 0.05$ and $\alpha/2$ becomes 0.025.

$Z_{0.025} = 1.96$

The proportion of people whose response is Yes = 0.599

The proportion of people whose response is No = 0.401

<u>Note:</u> Here $np \geq 5$ and $n(1-p) \geq 5$ for all proportions above and so we can proceed with the analysis.

Calculating CI:

- For proportion of people who said Yes is as follows:

  $\hat{p} = 0.599$ $Z_{0.025} = 1.96$ $n = 342$

  Now the interval becomes:

  $$\left[0.599 - 1.96\sqrt{\frac{0\cdot599(0.401)}{342}}, 0.599 + 1.96\sqrt{\frac{0.599(0.401)}{342}}\right]$$

  Now the interval becomes [0.547, 0.651]

- For proportion of people who said No is as follows:

  $\hat{p} = 0.401$ $Z_{0.025} = 1.96$ $n = 342$

  Now the interval becomes:

  $$\left[0.401 - 1.96\sqrt{\frac{0\cdot401(0.599)}{342}}, 0.401 + 1.96\sqrt{\frac{0.401(0.599)}{342}}\right]$$

  Now the interval becomes [0.349, 0.453]

<u>**Hypothesis Testing:**</u>

Let p_1: proportion of people who missed a class/test/assignment

p_2: proportion of people who did not miss anything

<u>Null Hypothesis(H_0):</u> The proportion of people who missed a class/test/assignment in BTech/BDes are less than or equal to the proportion of people who missed a class/test/assignment pursuing MTech/PhD/Research/MDes .

$H_0: p_1 \leq p_2 \Rightarrow H_0: p_1 - p_2 \leq 0$

<u>Alternate Hypothesis (H_a):</u> The proportion of people who missed a class/assignment/test in BTech/BDes are greater than the proportion of people who missed a class/test/assignment pursuing MTech/PhD/Research/MDes/MSc .

$H_a: p_1 > p_2 \Rightarrow H_a: p_1 - p_2 > 0$

Here, in our case

Considering two samples as BTech/BDes and MTech/PhD/Research/MDes/MSc

$n_1 = 258$, $n_2 = 84$

Proportion of people who missed a class/test/assignment in sample 1: $\hat{p}_1 = 0.647$
Proportion of people who missed a class/test/assignment in sample 2: $\hat{p}_2 = 0.452$

Consider the confidence interval as 0.95 ⇒ $\alpha = 0.05$
Finally, $Z_{0.05} = 1.645$
Note: Here $np \geq 5$ and $n(1\text{-}p) \geq 5$ for all proportions above and so we can proceed with the analysis.

Right-Hand Tailed Test:

H_0: $p_1 - p_2 \leq 0$ and H_a: $p_1 - p_2 > 0$

Test – Statistic:

$$\boxed{Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}}$$

$$z^* = \frac{(0.647 - 0.452) - 0}{\sqrt{\frac{0.647(0.353)}{258} + \frac{0.452(0.548)}{84}}}$$

Rejection Region Approach:

For a level α, we can reject H_0 if $Z^* \geq Z_\alpha$

$$z^* = 3.151 \text{ and } z_{0.05} = 1.645$$

Here, $Z^* \geq Z_\alpha$ ⇒ We can reject null hypothesis.

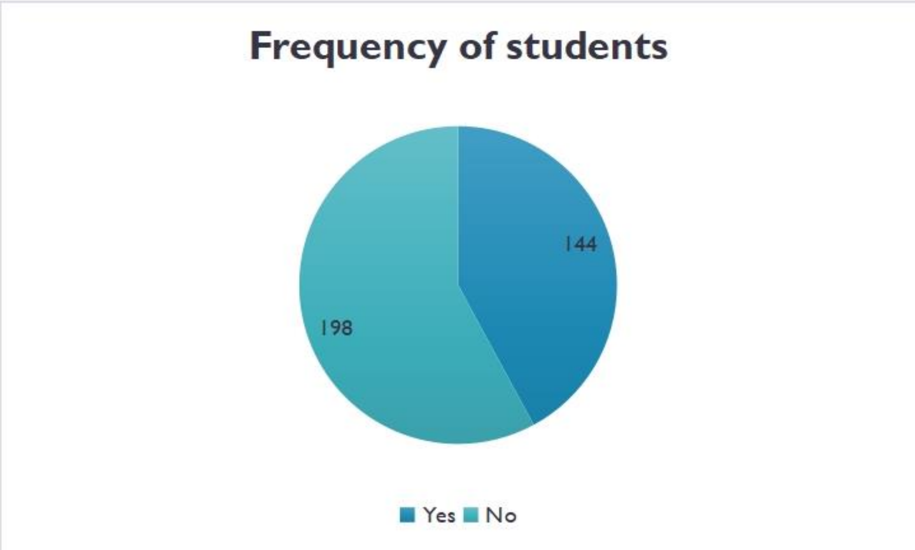
Hence there is strong evidence that most of the bachelors have missed their classes or tests or assignments.

CONSISTENCY OF SLEEPING SCHEDULE:

Here, we wish to calculate the proportion of students who have consistent sleep schedule.

Interpretation and Visualisation of data:

Response	Frequency
Yes	144
No	198



Point estimation of proportion:

p_1 - Proportion of students whose response is Yes
p_2 - Proportion of students whose response is No

Here, Sample Size =342.

$\hat{p}_1 = \frac{144}{342} = 0.42$, $\hat{p}_2 = \frac{198}{342} = 0.58$

Therefore,

42% of students have consistent sleep schedule.

58% of students don't have consistent sleep schedule.

Confidence Interval Estimation for proportion:

(1-α)100% Confidence interval for proportion p is

$$\boxed{\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}}$$

Here, we calculate the Confidence interval for proportion.
Take significance level α = 0.05

So, $Z_{\alpha/2}$ = 1.96, n=342.

Note: Here np≥5 and n(1-p) ≥ 5 for all the proportions above and so we can proceed with the analysis.

For p_1 95% CI is $\left[0.42 - 1.96\sqrt{\frac{0.42(0.58)}{342}}, 0.42 + 1.96\sqrt{\frac{0.42(0.58)}{342}}\right]$

= [0.393 , 0.447]

For p_2 95% CI is $\left[0.58 - 1.96\sqrt{\frac{0.58(0.42)}{342}}, 0.65 + 1.96\sqrt{\frac{0.58(0.42)}{342}}\right]$

= [0.553 , 0.607]

Hypothesis Testing:

In our sample data, proportion of students whose response is yes is 0.42. So, we would like to see if not even 50 percent of the population have a consistent sleep schedule.

Here, p_1 - Proportion of students whose response is Yes.

Left tailed test:

H_o: $p \geq 0.50$ H_a: $p < 0.50$

Test Statistic: $\boxed{Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n1}}}}$

Rejection Region Approach:

$Z^* = \frac{0.42-0.50}{\sqrt{\frac{0.42(1-0.42)}{342}}}$ = -2.99 $-Z_{0.05}$ = -1.645

Since, $Z^* < -Z_{0.05}$ we reject H_o at the confidence level of 95%.

Conclusion:

We therefore conclude not even 50 percent of the population have a consistent sleep schedule.

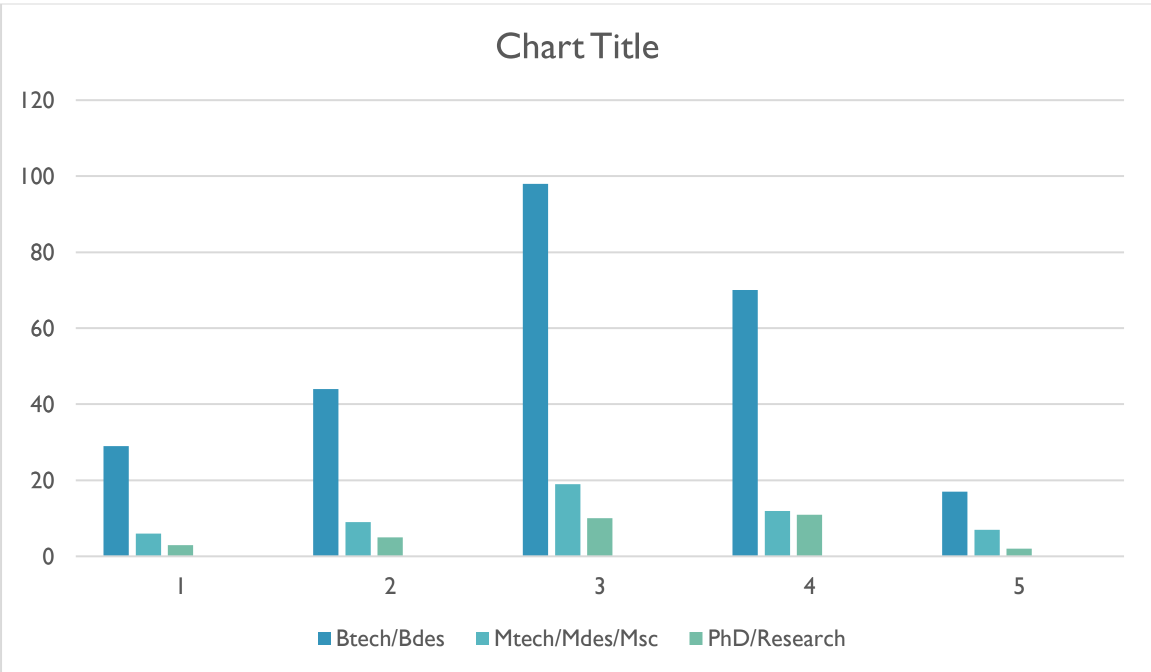
Students Sleep Satisfaction:

The above values are collected from a scale of numbers from 1 to 5 where 1denotes insufficient sleep and 5 denotes happy sleep.

The table shows the frequency of people and their respective selected scale factor:

Scaled factor	frequency of students
1	38
2	58
3	127
4	93
5	26

Graduation level	Frequency of people for their respective scale factor					Total
	1	2	3	4	5	
BTech/BDes	29	44	98	70	17	258
MTech/MDes/Msc	6	9	19	12	7	53
PhD/Research	3	5	10	11	2	31
Total	38	58	127	93	26	342



Confidence Interval:

Here we are going to calculate the confidence interval for the population variance

$n = 342$, α = 0.05

a = $\chi^2_{(1-\alpha/2),n-1}$ ⇒ a = $\chi^2_{0.975,341}$ = 291.735

b = $\chi^2_{\alpha/2,n-1}$ ⇒ b = $\chi^2_{0.025,341}$ = 394.051

CI: $\left[\frac{(n-1)S^2}{a^2}, \frac{(n-1)S^2}{a^2}\right]$

CI: $\left[\frac{(342-1)(479.16)^2}{(291.73)^2}, \frac{(342-1)(479.16)^2}{(394.05)^2}\right]$

CI: $[919.927, 504.210]$

So, the 95% confidence interval for σ^2 becomes $[919.927, 504.210]$

Hypothesis Testing:

Consider the α = 0.05, calculated S = 479.16

Using the Right-Tailed Test:

H0: $\sigma^2 \geq (500)^2$ and Ha: $\sigma^2 < (500)^2$

Here the specified value of σ_0 = 500

Test-Statistics:

$\chi^2_{computed} = \frac{(n-1)s^2}{\sigma_0^2}$

$\chi^2_{computed} = \frac{(342-1)(479.16)^2}{(500)\text{^}2}$

$\chi^2_{computed} = 313.17$

p-value approach:

P ($\chi^2 \geq \chi^2_{computed}$) = P ($\chi^2 \geq 313.17$) , n-1 = 341

p-value = 0.85778

0.858 > 0.05 and hence we fail to reject H_0.

Hence, we can say that our alternate hypothesis is observed correct from the above p-value approach.

Contributions:

Tanakala Yadunandana – MA22BTECH11018

- Frequency of missing meals - Pg 17-19
- Alertness or focus during the class - Pg 22-23

Kallu Rithika – AI22BTECH11010

- Frequency of naps vs gender- Pg 20-21
- Does caffeine consumption affect weekdays sleep - Pg 24-26

Talasani Sri Varsha – AI22BTECH11028

- Screentime habits of Students- Pg 9-12
- Does screentime effect the sleep consistency- Pg 13-15
- Consistency of sleep schedule – Pg 30-31

Kunche Aishwarya – AI22BTECH11015

- Impact of Irregular sleep on academics - Pg 27-29
- Student sleep satisfaction - Pg 32-33

Kota Dhanalakshmi – AI22BTECH11012

- Frequency of sleeping hours on weekdays - Pg 3-5
- Frequency of sleeping hours on weekends - Pg 6-8