Statistics Project
Sleeping Habits of Students

Presented By:

Tanakala Yadunandana – MA22BTECH11018

Kallu Rithika – AI22BTECH11010

Talasani Sri Varsha – AI22BTECH11028

Kunche Aishwarya – AI22BTECH11015

Kota Dhanalakshmi – AI22BTECH11012

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:**

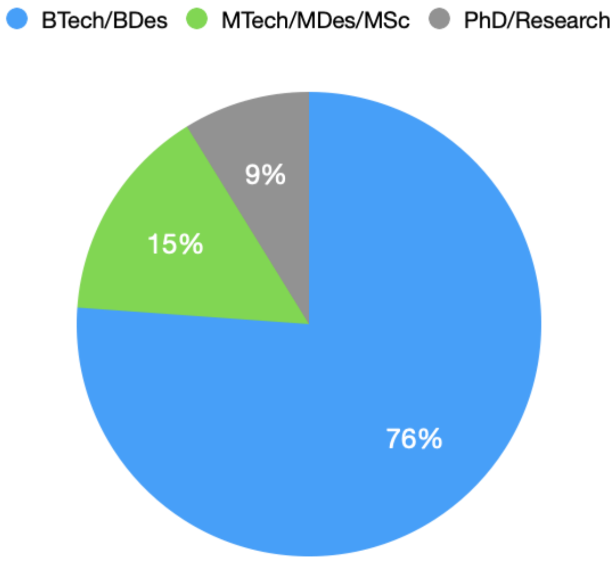
  Impact of irregular sleep on academics

- **Sleep Satisfaction:**

  Consistency of sleep schedule , Students sleep satisfaction

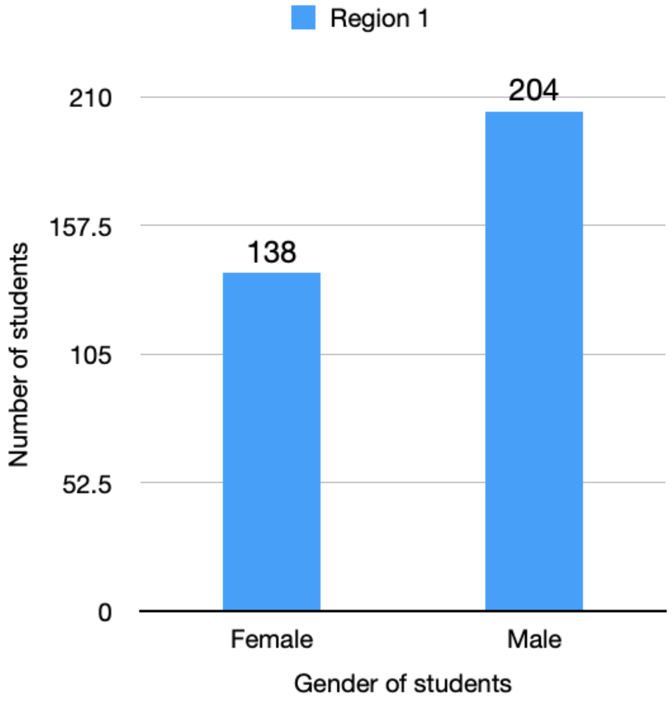
Background Information

Degree Being Pursued

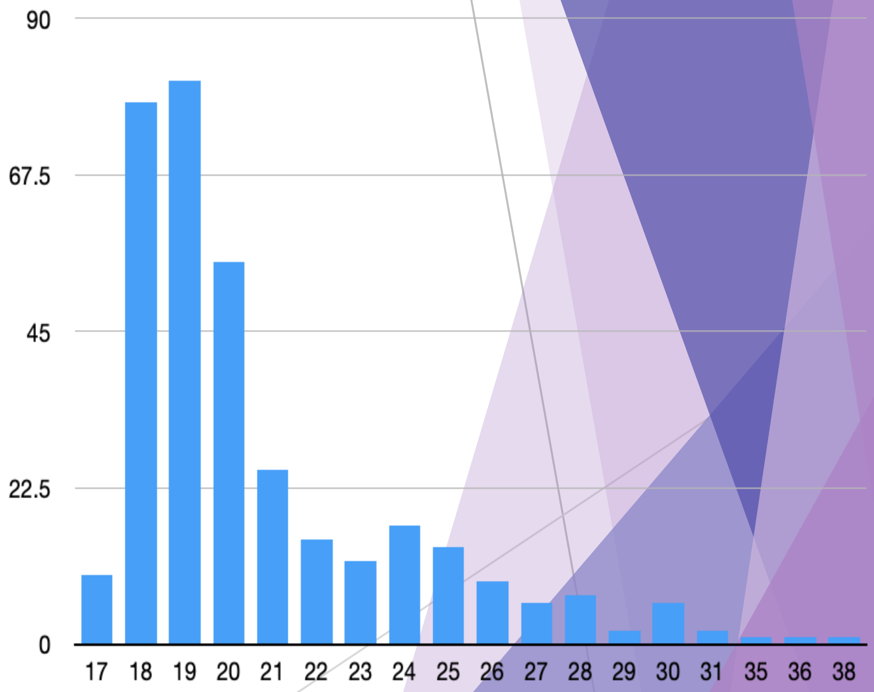


Percentage of students according to degree

Gender

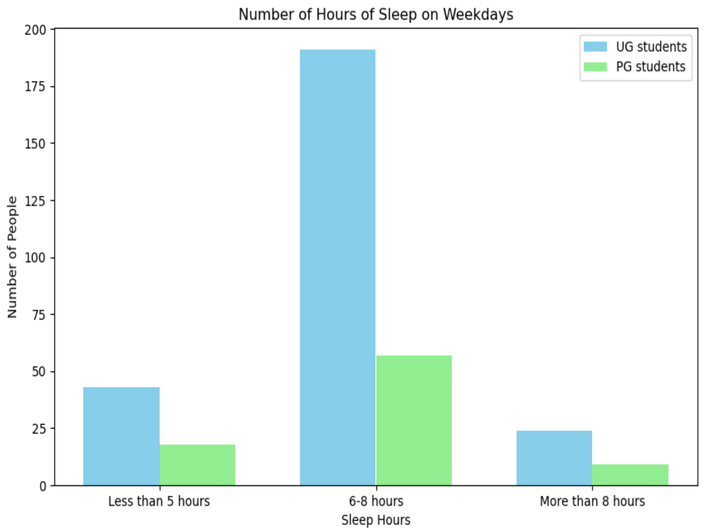


Age

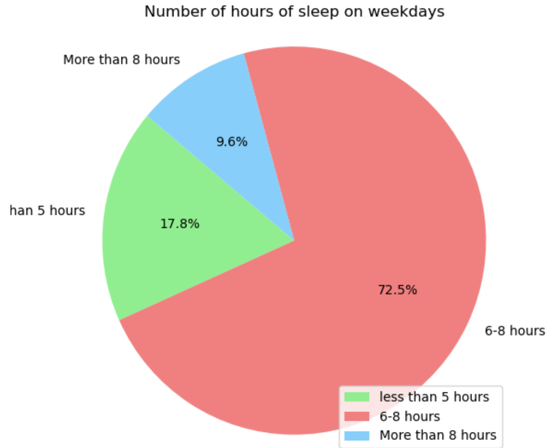


Frequency of sleeping hours on weekdays:

Data Visualization:



α



Confidence Interval Estimation:

(1-α)% CI for difference of proportions p1-p2 is given by $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Here, p1 = proportion of UG students who sleep more than 8-hours
p2 = proportion of PG students who sleep more than 8-hours

For p_1-p_2 , there is 95% confidence that it lies in interval is [-0.163, -0.011]
Therefore 95 % CI for p_1-p_2 is [-0.163, -0.011].

Hypothesis Testing:

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:

H0 = p1-p2 ≥ 0   Ha = p1-p2 < 0

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:

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

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

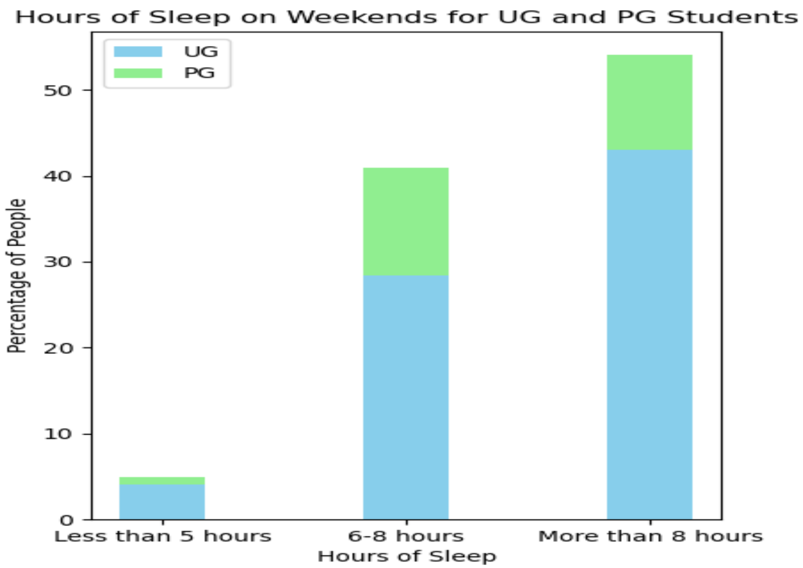
Hence, we can’t reject H0.

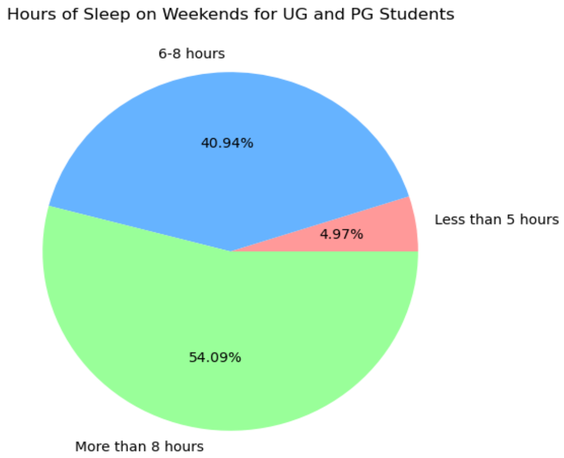
Conclusion :

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:

Data Visualization:





Confidence Interval:

CI for population proportions p with (1-α)100% confidence is given $\hat{p} \pm z_{a/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Here, 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

Considering significance level $\alpha=0.05$,

For p_1 ,there is 95% confidence that it lies in interval is [0.027,0.073]
For p_2 ,there is 95% confidence that it lies in interval is [0.357,0.461]
For p_3 ,there is 95% confidence that it lies in interval is [0.488,0.594]

Hypothesis Testing:

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:
H0 = p1-p2 $\leq$ 0   Ha = p1-p2 > 0

Test statistic:

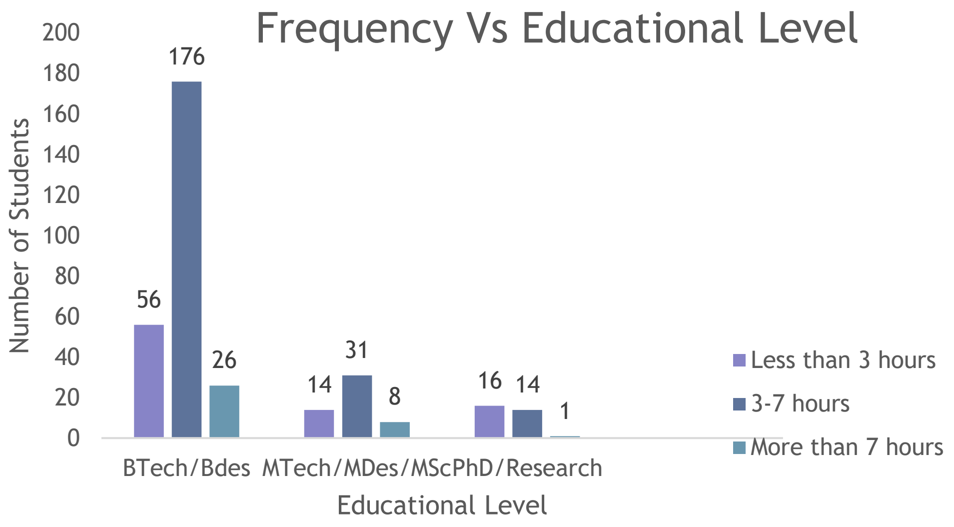
$$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n2}}}$$

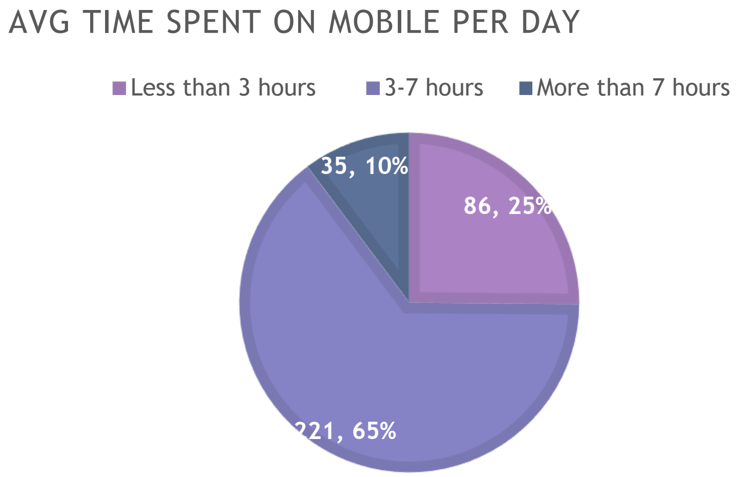
P-value = P(Z>Z*) = 0
For α=0.05, p<α
Hence, we can reject H0.

Conclusion:

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:





Point Estimation:

- 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.
- The point estimates of population proportions based on method of moments of students who use mobile less than 3 hours,3-7 hours, more than 7 hours per day are 0.25,0.65,0.10 respectively.

Confidence Interval:

- Confidence Interval for the difference in two population proportions p1-p2 is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- Using the formula for confidence interval estimation for population proportion, where confidence level is 95% ,we get **[0.227, 0.273],[0.624, 0.676],[0.084, 0.116]**.

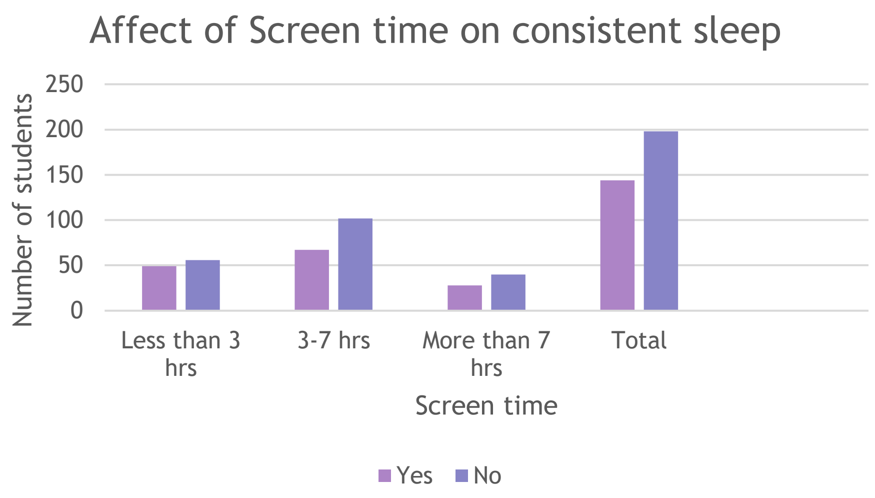
Hypothesis testing:

- Right tailed Hypothesis: Ho=p1-p2≤0, Ha=p1-p2>0, where
- p_1 - Proportion of undergraduate students who use phone for 3-7 hours a day=0.68.
- p_2 - Proportion of postgraduate students who use their phones for 3-7 hrs a day=0.54.
- $Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = 2.27$, $Z_{0.05} = 1.645$
- Since, $Z^* > Z_{0.05}$ we reject Ho at the confidence level of 95%.

Conclusion:

- 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.

Affect of screen time on sleep consistency:



<u>Chi square test of independence:</u>

- 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.

- Degrees of freedom =(1)(2).
- Considering α = 0.05, Critical value =5.991
- Calculated test statistic χ^{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.

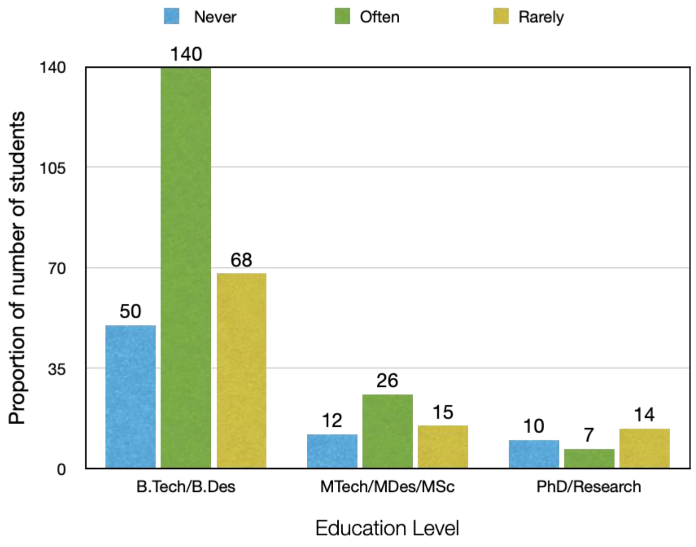
P-value:

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

Conclusion:

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

Frequency of missing meals



- p_1 - Proportion of BTech/BDes students who miss meals often due to irregular sleep = 0.542
- p_2 - Proportion of MTech/MDes/MSc students who miss meals often due to irregular sleep = 0.491

Confidence Interval :

Confidence interval for difference in proportion is $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

95% CI for estimation of p_1-p_2 is [-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:

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

Test statistic $Z^* = \dfrac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = 0.676$ $\quad$ $Z_{0.05} = 1.645$

$Z^* < Z_{0.05}$

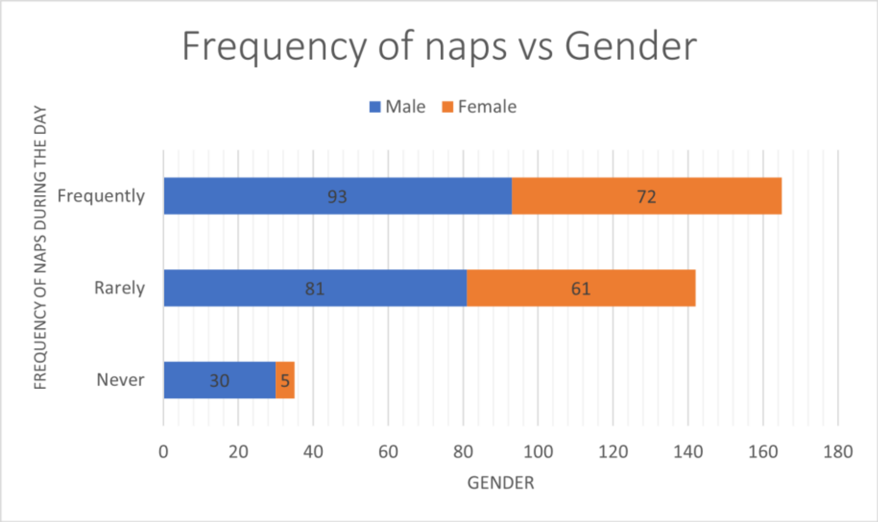
So, we can't reject the null hypothesis

Conclusion:

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

Visualization of Data



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

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

- $\hat{p}_1$=0.397 , n1= 204
- $\hat{p}_2$=0.442 , n2 = 138

Confidence Interval Estimation:

- $(1\text{-}\alpha)100\%$ Confidence Interval for the difference in two population proportions p1-p2 is $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
- 95% CI for estimation of p1-p2 is [-0.1516 , 0.0616]

Hypothesis Testing

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

$H_0 : p_1 - p_2 \leq 0$

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

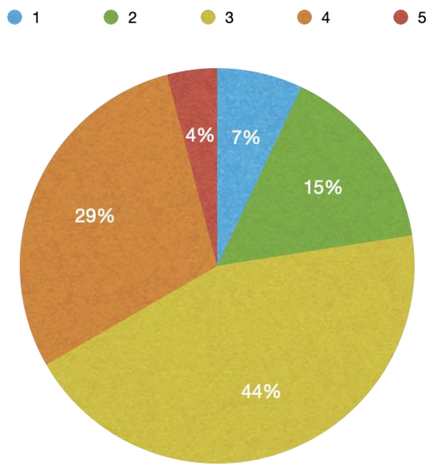
$H_a : p_1 - p_2 > 0$

- By using Right Tailed test,
- Test statistic is $Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$
- Rejection Region Approach:
- The value of test statistic is $Z^* = -0.387$. 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.

Conclusion :

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



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

Here,

- Sample mean $\bar{x}$ = 3.078
- Sample Variance S^2 = 0.941
- Sample size = 342

Confidence Interval:

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

For μ 95% CI is [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

Right tailed test for mean:

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

Test statistic $t^* = \frac{\bar{x} - p_0}{S/\sqrt{n}} = 1.346$

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

p-value = $0.0896 > 0.05 = \alpha$

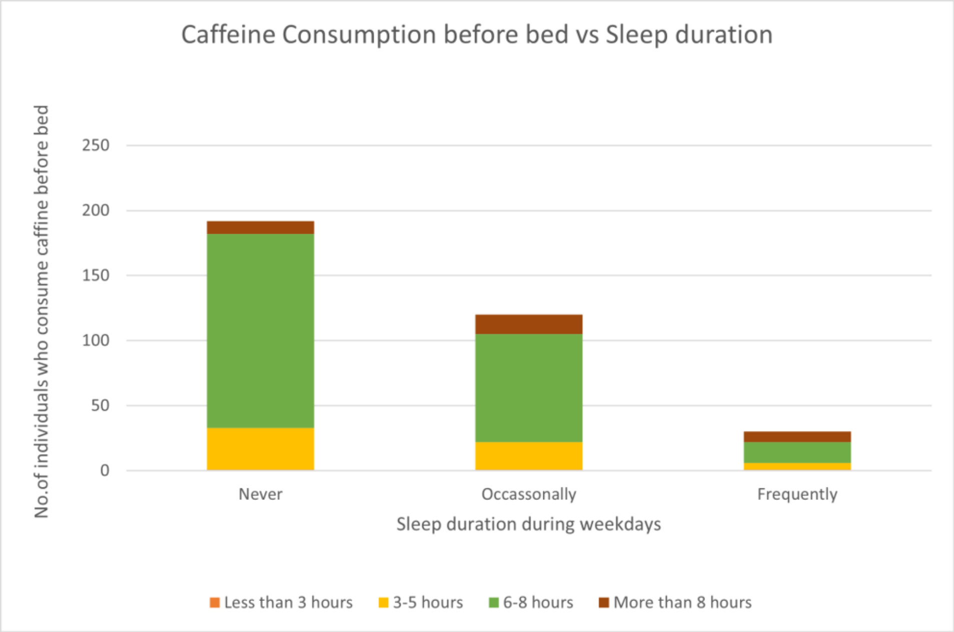
Hence, we fail to reject H_0

Conclusion:

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

Does caffeine consumption affect weekday sleep?

Data Visualization



We are performing chi square test of independence if there is significant association between categorical variables.

χ^2 Test Statistic is :

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

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.
- Table with expected frequencies:

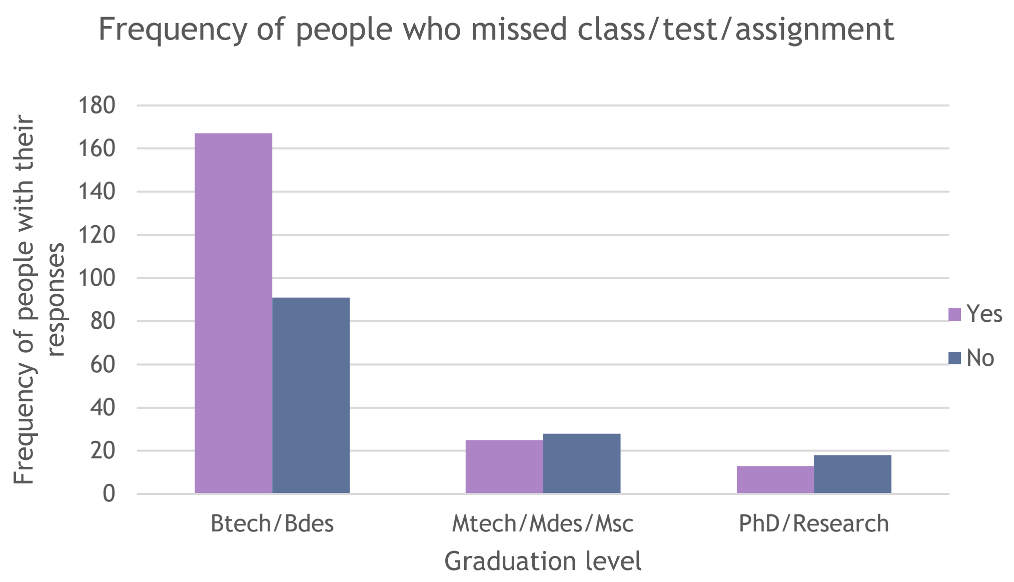
Caffeine consumption Before Bed	Less than 3 hrs	3-5 hrs	6-8 hrs	More than 8 hrs
Never	1.68	32.56	139.22	18.52
Occassonally	1.052	20.35	87.017	11.57
Frequently	0.263	5.087	21.75	2.89

- From the above data , χ^2 = 18.73
- Critical Value: Since our test statistic χ^2(18.73) is larger than the critical value 12.592, we reject the null hypothesis of our test.
- 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.

Conclusion :

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:



Confidence Interval:

Confidence Interval for population proportion : $\hat{p} \pm z_{a/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Let p_1: proportion of people who missed a class/test/assignment
p_2: proportion of people who did not miss anything

The confidence interval for the proportion of people who said Yes is [0.547, 0.651]

The confidence interval becomes for the proportion of people who said No is [0.349, 0.453]

Hypothesis Testing:

Null Hypothesis(H_0): 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: $p_1 \le p_2 \Rightarrow H_0$: $p_1 - p_2 \le 0$

Alternate Hypothesis (H_a): 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: $p_1 > p_2 \Rightarrow H_a$: $p_1 - p_2 > 0$

For the above hypothesis, we consider using Right-Tailed Test with the rejection region approach as:

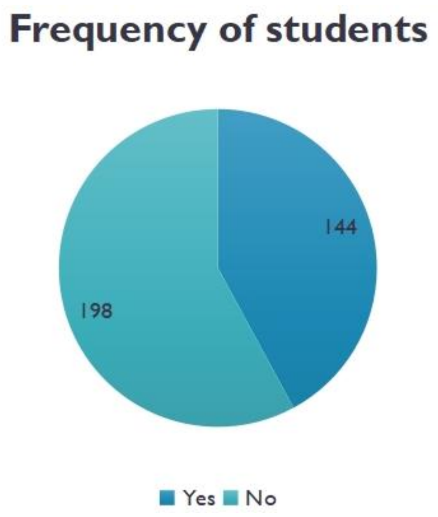
$Z^* \ge Z_\alpha$ and formula as $Z^* = \dfrac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$

$Z^* = 3.151$ and $Z_{0.05} = 1.645$
Here, $Z^* \ge Z_\alpha \Rightarrow$ We can reject null hypothesis.

Conclusion:

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

Consistency of sleep schedule:



Point Estimation:

- p_1 - Proportion of students whose response is Yes =0.42
- p_2 - Proportion of students whose response is No=0.58

Confidence Interval:

Confidence Interval for population proportion $\hat{p} \pm z_{a/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Using the formula for confidence interval estimation for population proportion, where confidence level is 95% ,we get **[**0.393, 0.447**],[**0.553, 0.607**]** for yes, no responses respectively.

Hypothesis testing:

<u>Left tailed Test</u>

Null hypothesis: Ho=p≥0.50,

Alternate Hypothesis: Ha=p<0.50, where

p_1 - Proportion of students whose response is Yes.

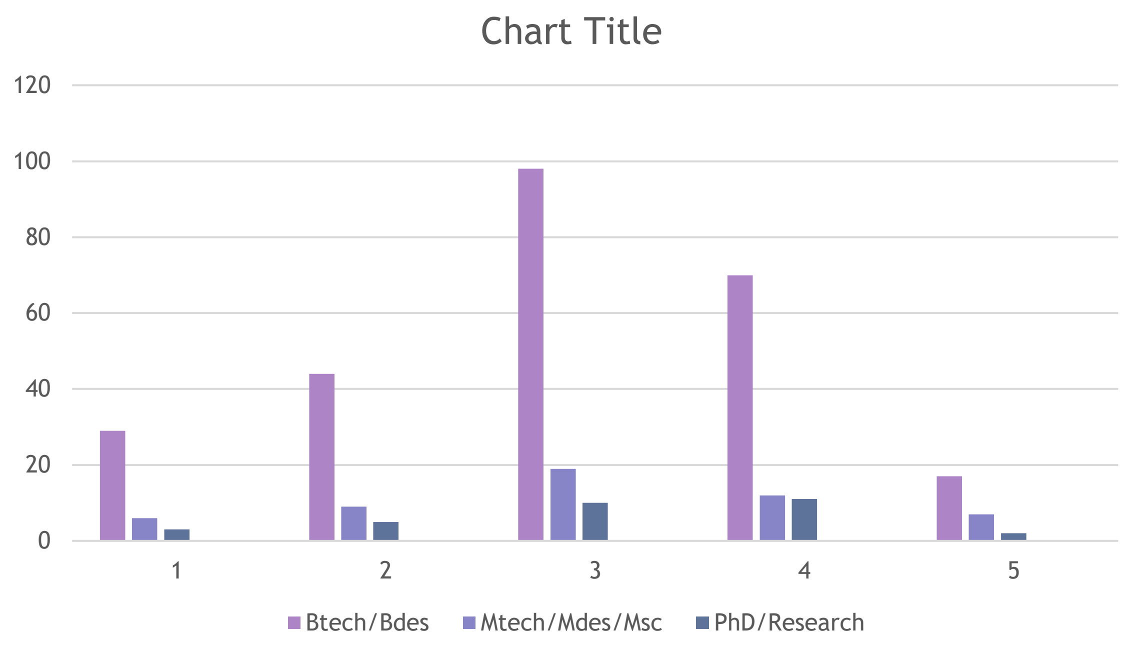
$Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ = -2.99, $-Z_{0.05}$ = -1.645

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

Conclusion:

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

Student Sleep Satisfaction:



Confidence Interval:

We calculated the confidence interval using population variance by the formula:

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

The parameters are taken as a = $\chi^2_{0.975,341}$ = 291.735 and b = $\chi^2_{0.025,341}$ = 394.051

So, the confidence interval becomes $[919.927, 504.210]$

Hypothesis Testing:

$H_0 : \sigma^2 \geq (500)^2$ and Ha: $\sigma^2 < (500)^2$

$\alpha = 0.05$ and S = 479.16

specified value of $\sigma_0 = 500$

Now by using the Right-Tailed Test we get the test statistic as:

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

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

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

Conclusion:

Therefore, there is strong evidence that our claim Ha: $\sigma^2 < (500)^2$ is correct as p-value is greater than the α which is taken as 0.05

Thanking you