APPLIED STATISTICS

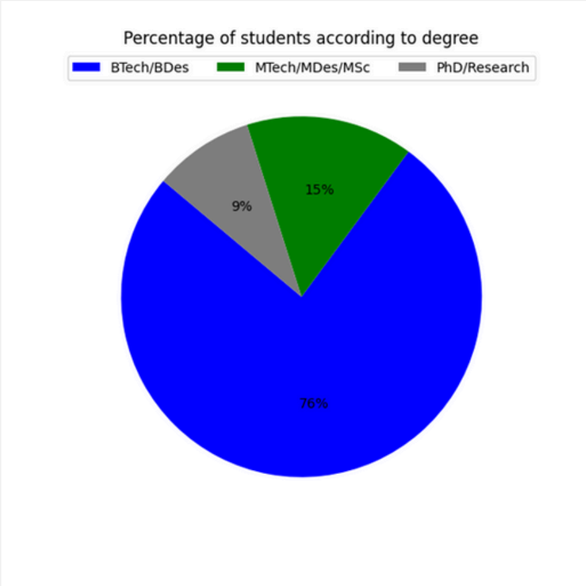
K Meenakshi
M Bhavya Kumari
R Roshni

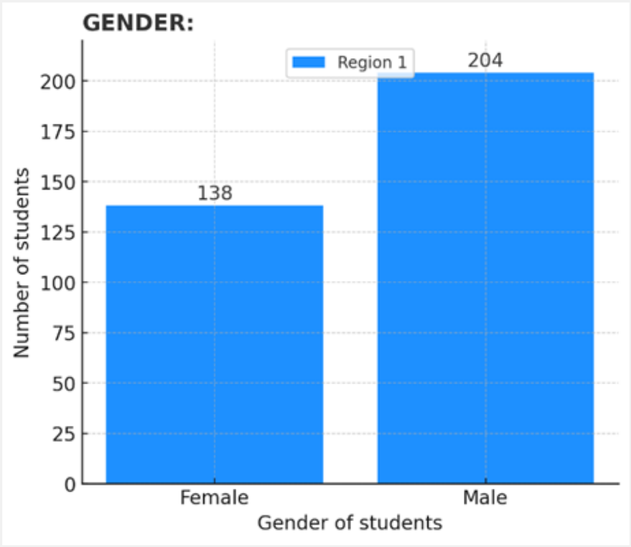
Computational Project 1

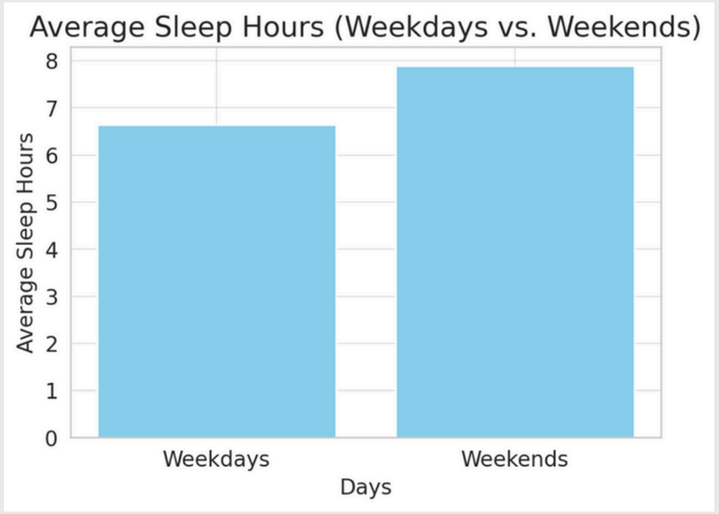
Introduction:

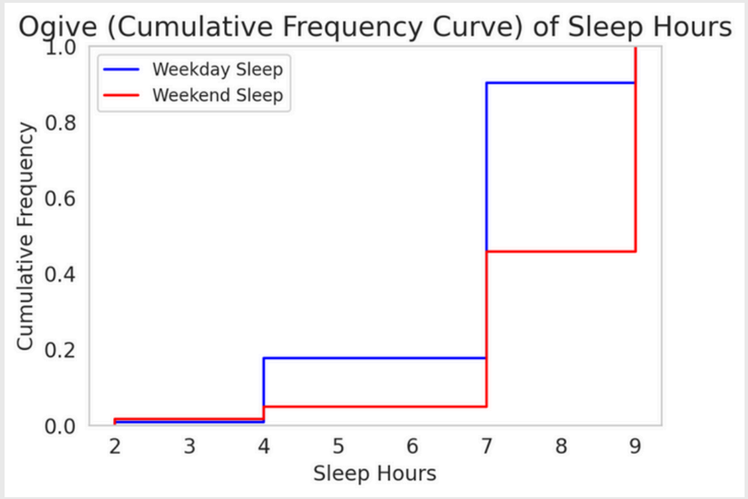
The report analyses the sleep patterns and experiences of students. We conducted a survey of 342 students to get a better understanding of sleep habits and various factors affecting the sleep health. The survey investigated various factors affecting student sleep habits, including:Background Information:

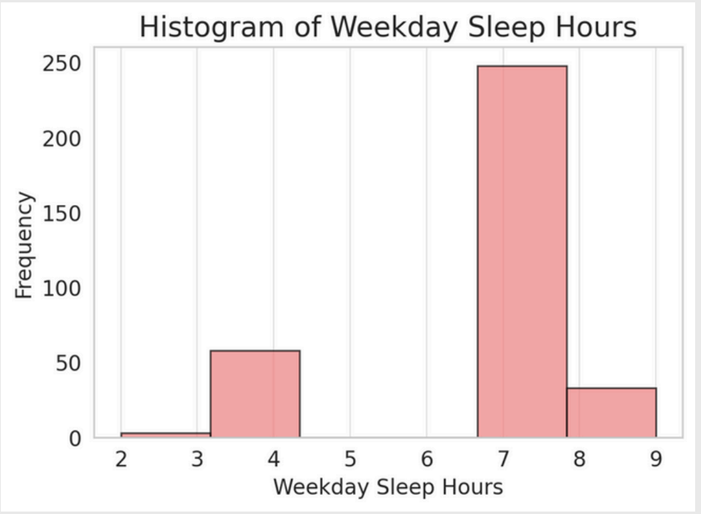
- Degree being pursued by student
- Gender
- Age

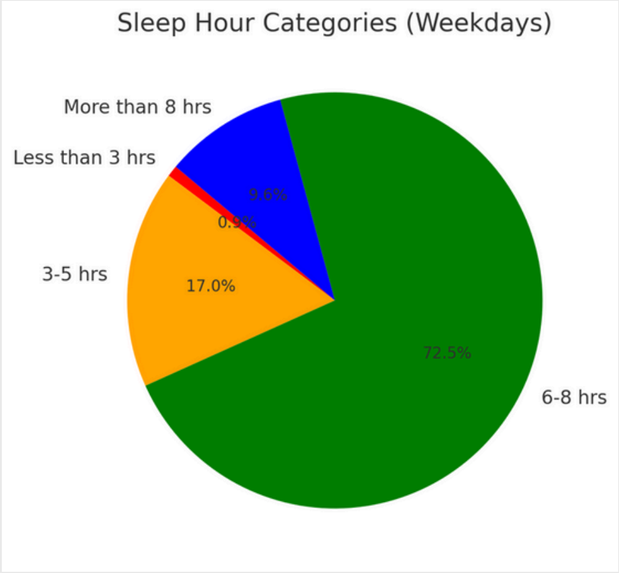


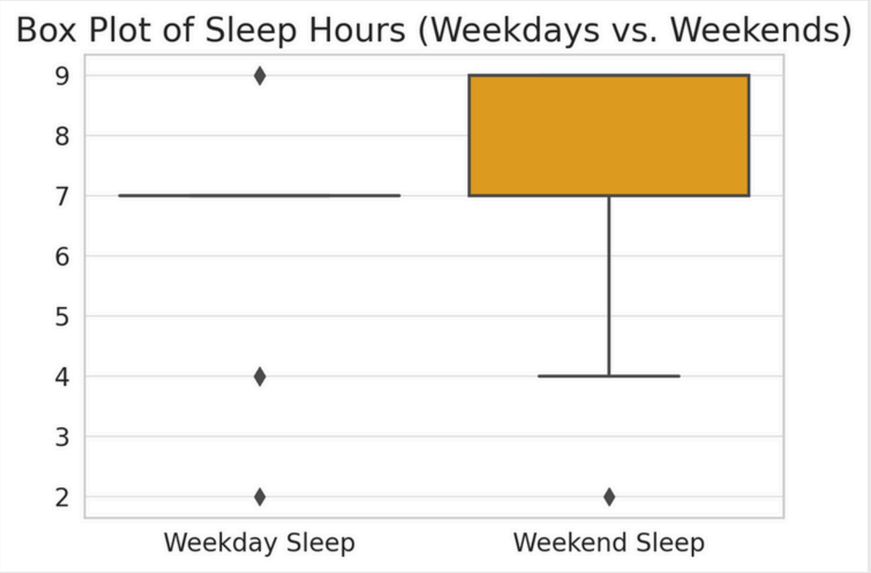



Average Sleep Hours (Weekdays vs. Weekends)
Average Sleep Hours
0
1
2
3
4
5
6
7
8
Weekdays
Weekends
Days


Ogive (Cumulative Frequency Curve) of Sleep Hours
Weekday Sleep
Weekend Sleep
Cumulative Frequency
0.0
0.2
0.4
0.6
0.8
1.0
2
3
4
5
6
7
8
9
Sleep Hours


Histogram of Weekday Sleep Hours
Frequency
0
50
100
150
200
250
2
3
4
5
6
7
8
9
Weekday Sleep Hours


Sleep Hour Categories (Weekdays)
More than 8 hrs
Less than 3 hrs
3-5 hrs
6-8 hrs
9.6%
0.9%
17.0%
72.5%


Box Plot of Sleep Hours (Weekdays vs. Weekends)
2
3
4
5
6
7
8
9
Weekday Sleep
Weekend Sleep

Descriptive Statistics Summary

1. Measures of Central Tendency

- Mean (Average Sleep Hours)
  - Weekdays: 6.64 hours
  - Weekends: 7.90 hours
  - Age: 20.62 years
- Median (Middle Value)
  - Weekdays: 7 hours
  - Weekends: 9 hours
  - Age: 19.5 years
- Mode (Most Frequent Value)
  - Weekdays: 7 hours
  - Weekends: 9 hours
  - Age: 19 years

2. Measures of Dispersion

- Standard Deviation (Spread of Data)
  - Weekdays: 1.42 hours
  - Weekends: 1.45 hours
  - Age: 3.24 years
- Variance (Square of Std Dev)
  - Weekdays: 2.01
  - Weekends: 2.09
  - Age: 10.47
- Interquartile Range (IQR)
  - Weekdays: 0.00 (Most values are concentrated around 7)
  - Weekends: 2.00 (More variation in sleep hours)
  - Age: 2.75 years

3. Data Range

- Minimum & Maximum Sleep Hours
  - Weekdays: Min = 2, Max = 9
  - Weekends: Min = 2, Max = 9
- Minimum & Maximum Age
  - Min = 17 years, Max = 38 years

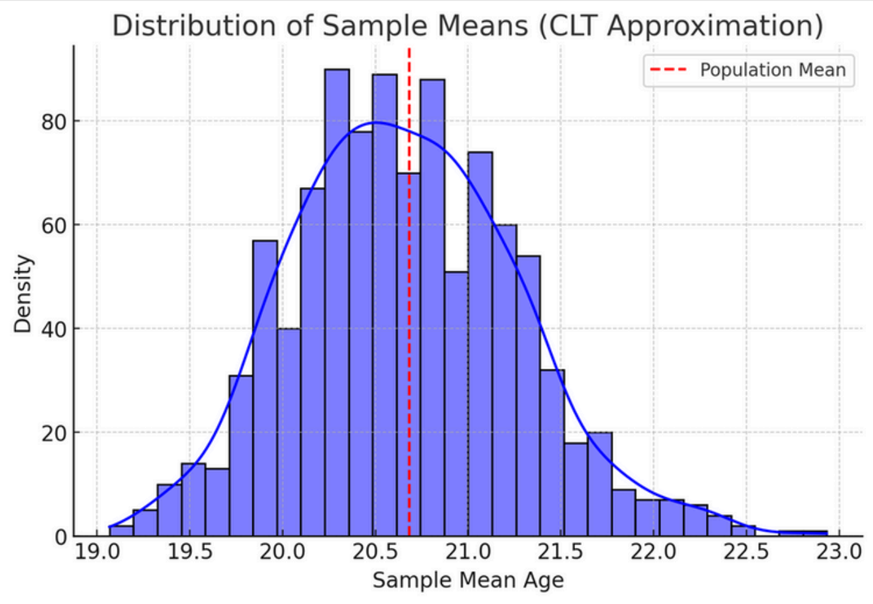
Insights

1. Sleep hours are higher on weekends than on weekdays. 2. Weekday sleep is more consistent, with an IQR of 0 (most students sleep around 7 hours).

3. Weekend sleep shows more variation, with an IQR of 2 hours, meaning students follow different sleep patterns.

CLT to approximate the distribution of the sample mean

For applying the Central Limit Theorem (CLT), we are using 'Age' since it's numerical and suitable for this analysis



Insights from the CLT Analysis:

1. Shape of the Distribution: The distribution of sample means closely follows a normal distribution, which aligns with the Central Limit Theorem (CLT) regardless of the original distribution of the 'Age' data.
2. Mean Alignment: The red dashed line represents the population mean, which matches the peak of the sample mean distribution. This further validates the accuracy of the approximation.
3. Reduced Variability: The sample means are more tightly clustered around the true mean than individual data points, showing the expected reduction in variability as sample size increases.

Summary

General Sleep Trends:

- Students sleep more on weekends than weekdays.
- Sleep duration is relatively stable on weekdays but more varied on weekends.

Factors Affecting Sleep Quality:

- Social media and academic stress are the most common factors disrupting sleep.
- Other notable factors include noise, light, and peer interactions.

Impact on Academics:

- A significant number of students (especially BTech/BDes students) have missed classes, assignments, or tests due to inadequate sleep.

Sleep Satisfaction & Consistency:

- Less than 50% of students maintain a consistent sleep schedule.
- The average sleep satisfaction rating is around 3, indicating moderate satisfaction but room for improvement.

Overall Implications:

- The findings suggest a need for better sleep hygiene among students.
- Awareness programs on sleep quality and time management could help improve students' academic performance and well-being.

Computational project 2

This report includes survey of 342 students to get a better understanding of sleep habits and various
factors affecting the sleep health.
In this project we shall explore the following two factors:

- Students sleep satisfaction rating 1 to 5(1Q,2Q,3Q)
- Have you ever missed a class / assignment / test due to being tired(4Q)

1(a) Method Of Moments(MOM):

For Gamma(a,b)

Mean=a/b

variance=a/b^2

Equationg sample moments and population moments,we get:

a_mom = mean^2 / var

b_mom = var / mean

For the data of 342 students,we get:

Method of Moments Estimates:->

Sample Mean: 3.0322

Sample Variance: 1.1925

MoM Estimate of a:7.7098

MoM Estimate of b:0.3933

(b)Maximum Likelihood Estimates(MLE):

we get MLE equation as:

$\psi(a) - \log(a) = \text{mean\_log\_x} - \log(\bar{x})$

Equating above equation to zero,we get the value of a_mle.

Thereafter,we can compute value of b_mle using:

b_mle = x_bar / a_mle

For the data of 342 students,we get:

Maximum Likelihood Estimation(Gamma):->

Sample mean ($\bar{x}$): 3.0322

Mean of log(x): 1.0248

MLE Estimate for shape (a): 6.0838

MLE Estimate for scale (b):0.4984

Conclusion:

The MLE fit indicates a slightly lower shape and higher rate, meaning the MLE estimates a distribution more concentrated around the mean, which may better reflect the central tendency and skewness of the actual data. Since MLE is generally more robust, we would prefer the MLE estimates for further modeling.

2. 95% confidence interval for variance:

alpha=0.05

Confidence interval for Variance:((n-1)var/chi_upper,(n-1)var/chi_lower)

For the data of 342 students,we get:

95% Confidence Interval for Variance:->

Sample Variance(s^2):1.1925

Confidence Interval:(1.0320,1.3939)

Conclusion:

This range tells us that the variability in sleep satisfaction is moderate. Since the entire interval is above 1, it suggests that students' satisfaction ratings are not tightly clustered
around the mean — there’s noticeable spread, possibly due to differences in lifestyle, stress,
or sleep duration.

3. 95% confidence interval for difference in means:

Here data 1 represents sleep satistfaction of female students and data 2 represents sleep satisfaction of male students.

Confidence interval for difference in means=((mean1-mean2)-t_crit*SE,(mean1-mean2)+t_crit*SE)

For the data of 342 students,we get:

Mean of data1:3.0145

Mean of data2: 3.0441

Difference(data1-data2): -0.0296

Confidence Interval: (-0.2640,0.2047)

Conclusion:

The small difference in mean sleep satisfaction between male and female students (−0.0296) is not statistically significant, as the 95% confidence interval includes zero.

4. Comprehensively test the hypothesis at a level of significance of 0.05

This is based the number of agrees and disagrees on if they missed a class/assignments/test due to being tired.

sample proportion:

p_hat = x / n

Standard error:

se = math.sqrt(p0 * (1 - p0) / n)

Z-statistic:

z = (p_hat - p0) / se

p-value for one-tailed test (H1: p > 0.5)

p_value = 1 - norm.cdf(z)

Based on the survey conducted:

Number of YES:222

Number of NO:120

For the data of 342 students,we get:

Z-statistic: 5.5155

P-value: 0.0000

Reject the null hypothesis (H0). There is significant evidence that p > 0.5.

Conclusion:

There is strong statistical evidence that more than half of the students have missed a class/test due to being tired.This highlights the potential academic consequences of poor sleep patterns among college students, emphasizing the need for better sleep management.