The Analysis of the Sleep Efficiency

Submitted by: Ziad, Yehia, Omar, Mohamed

26/5/2024

Introduction

Sleep is crucial for maintaining overall health and well-being. Numerous factors, such as lifestyle habits, can significantly influence sleep quality and efficiency. This analysis investigates how smoking, caffeine intake, alcohol consumption, and physical activity affect different sleep stages—REM sleep, light sleep, and deep sleep—using a dataset of individuals' sleeping habit.

Research Question

How do smoking, caffeine intake, alcohol consumption, and physical activity influence the efficiency and stages of sleep (REM, light, and deep sleep)?

Hypothesis

1. ***Smoking Hypothesis***

   Null Hypothesis (H_0): Smoking has no effect on sleep efficiency or the duration of light and deep sleep stages.
   Alternative Hypothesis (H_1): Smoking negatively affects sleep efficiency and reduces the duration of deep sleep stage.

2. ***Alcohol Hypothesis***

   Null Hypothesis (H_0): Alcohol consumption has no effect on sleep efficiency or the duration of light and deep sleep stages.
   Alternative Hypothesis (H_1): Alcohol consumption negatively affects sleep efficiency, decreases deep sleep, and increases light sleep duration.

3. ***Exercise Hypothesis***

   Null Hypothesis (H_0): Exercise frequency has no effect on sleep efficiency or the duration of light and deep sleep stages.
   Alternative Hypothesis (H_1): Regular exercise positively affects sleep efficiency, increasing the duration of deep sleep.

Population of Interest:

The population of interest consists of randomly picked individuals from a dataset available on Kaggle, which includes various metrics related to their sleep patterns and lifestyle habits.

Sampling Method:

The dataset was randomly sampled from Kaggle. The random sampling helps ensure that the analysis captures a diverse range of individuals, reducing selection bias.

Bias Identification:

Potential biases in the study include:

- **_Self-Reported Data:_** Recall bias and social desirability bias may affect the accuracy of self-reported lifestyle habits.
- **_Non-Representative Sample:_** If the dataset predominantly includes certain demographics, it may not be fully representative of the general population.
- **_Confounding Variables:_** Other unmeasured factors, such as stress or underlying health conditions, may influence sleep patterns.
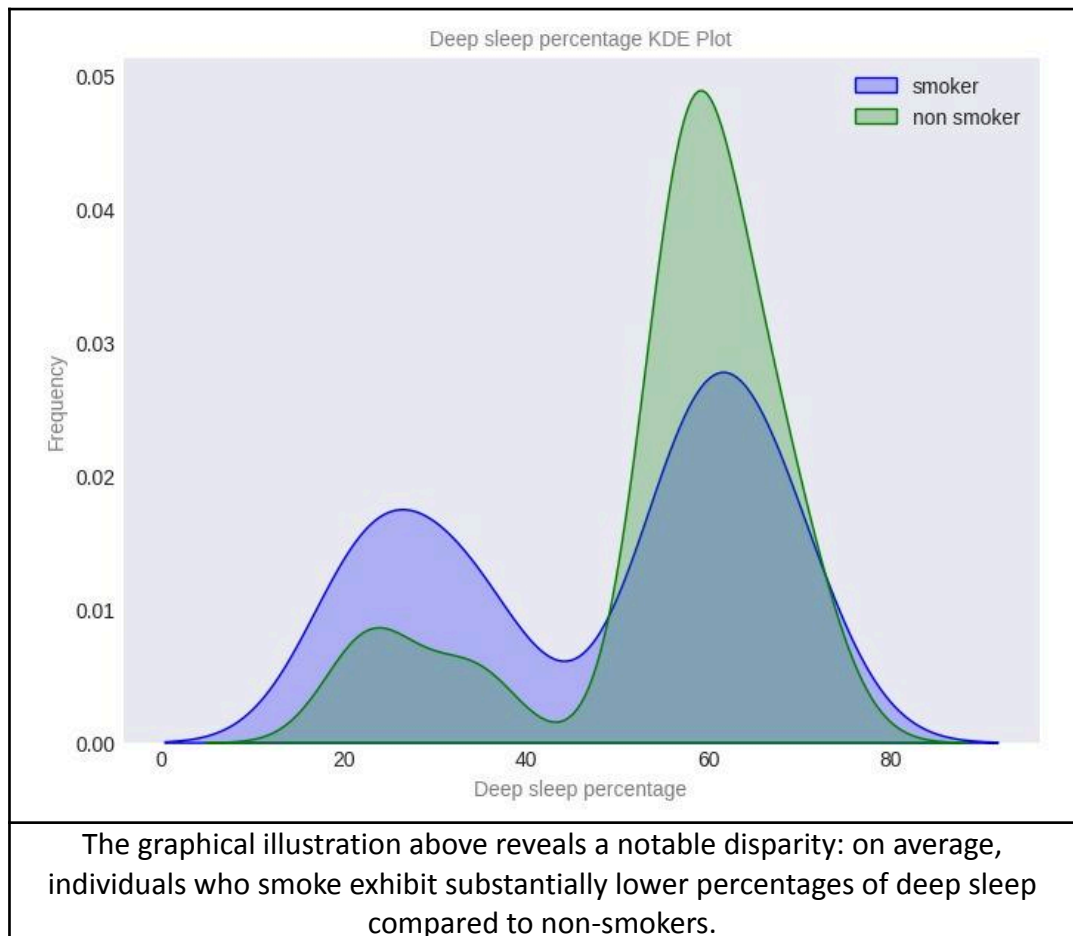
Survey Questions/Collected Data/Dataset:

Below is a brief explanation of each feature:

- **_Age:_** Age of the individual in years.
- **_Gender:_** Gender of the individual.
- **_Bedtime_**: Time the individual goes to bed.
- **_Wakeup Time:_** Time the individual wakes up.
- **_Sleep Duration:_** Total hours of sleep.
- **_Sleep Efficiency:_** Percentage of time spent asleep while in bed.
- **_REM Sleep Percentage:_** Percentage of sleep spent in REM sleep.
- **_Deep Sleep Percentage:_** Percentage of sleep spent in deep sleep.
- **_Light Sleep Percentage:_** Percentage of sleep spent in light sleep.
- **_Awakenings:_** Number of times the individual wakes up during the night.
- **_Caffeine Consumption:_** Amount of caffeine consumed daily (mg).
- **_Alcohol Consumption:_** Amount of alcohol consumed daily (units).
- **_Smoking Status:_** Indicates if the individual smokes.(Yes or No)
- **_Exercise Frequency:_** Frequency of physical activity per week.

Number of samples used: **_452_**

Analysis:

Deep sleep percentage KDE Plot

The graphical illustration above reveals a notable disparity: on average, individuals who smoke exhibit substantially lower percentages of deep sleep compared to non-smokers.

Conclusion: Smokers are less likely to be in the deep sleep stage than the non-smokers, hence rejecting the null hypothesis.

P-value: 1.0534925598548739e-08

Reject the null hypothesis: Smoking affects sleep efficiency
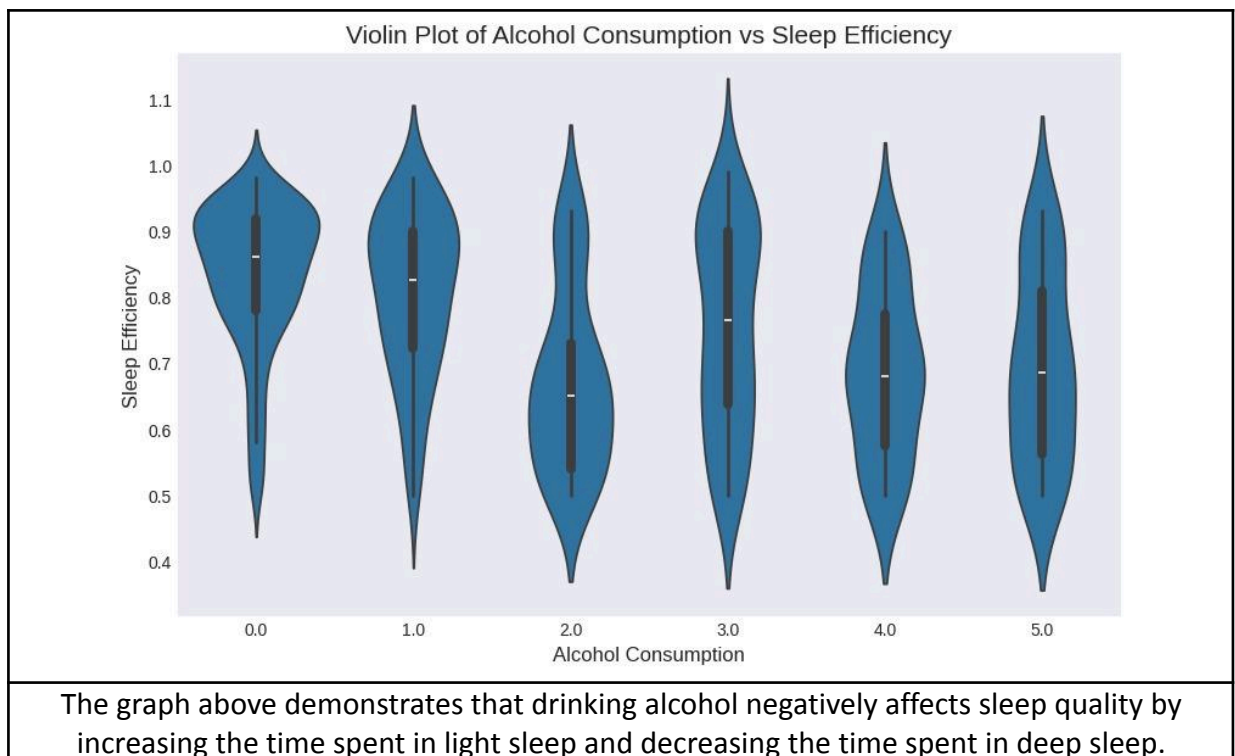
Exercise Frequency vs Sleep Efficiency
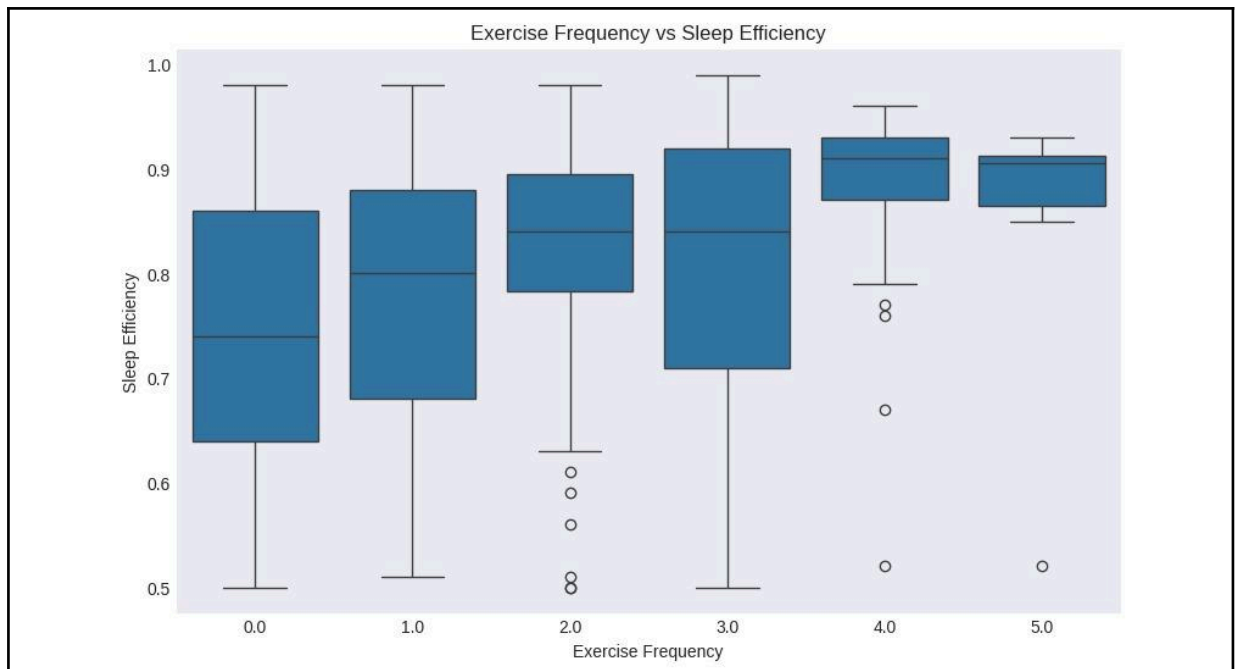
The graph above indicates that when people exercise more often, they tend to sleep better.

Conclusion: Individuals who participate in physical exercise are more likely to sleep better and more efficiently than the people who don't engage in physical activity, hence rejecting the null hypothesis.

P-value: 1.5151001199864567e-05

Reject the null hypothesis: Exercise affects sleep efficiency.

The graph above demonstrates that drinking alcohol negatively affects sleep quality by increasing the time spent in light sleep and decreasing the time spent in deep sleep.

Conclusion: Alcohol drinkers are less likely to be in the deep sleep stage than the non drinkers, although alcohol tends to help sleep faster but less efficiently; hence, rejecting the null hypotheses.

P-value: 3.0751582149379786e-05

Reject the null hypothesis: Exercising affects sleep efficiency.

Hypothesis Testing Steps

- *Formulate Hypotheses:* Clearly define the null and alternative hypotheses for each factor.
- *Select Significance Level:* Choose an appropriate alpha level (0.05) for hypothesis testing.
- *Conduct Tests:* Use statistical tests to evaluate the hypotheses.
- *Analyze Results:* Compare p-values to the significance level to accept or reject the null hypotheses.
- *Interpret Findings:* Discuss the implications of the results in the context of sleep efficiency and stages.

Conclusion

This analysis provides insights into how smoking, caffeine intake, alcohol consumption, and physical activity affect different stages of sleep. The results indicate significant effects of these lifestyle choices on sleep efficiency,

particularly on REM and deep sleep stages. Understanding these relationships can help individuals make informed decisions to improve their sleep quality.

Any potential issues

Data Quality: The data set would've been more accurate if it had more rows.

Confounding Variables: Other factors not accounted for in the dataset may influence sleep patterns, such as stress, physical activity, or underlying health conditions.

Generalizability: The findings may not be applicable to all populations if the dataset is not representative.

Incomplete Data: There was some missing data, we used the mode to fill missing values based on each certain metric.

Data Volatility: There was limited numerical data to analyze and make a logical conclusion our of their correlations together.