HIGH SCHOOL EDITION
Journal of Student Research

A Multidimensional Statistical Analysis of Lifestyle Determinants Affecting Sleep Quality in Adults

Ruiwen Mao

St. Clement's School, Canada

ABSTRACT

Sleep quality is a critical determinant of physical and mental health. However, the various lifestyle factors influencing sleep quality are not fully understood. This study investigates the relationships between sleep quality (measured on a scale of 1-10) and lifestyle metrics, including sleep duration, stress levels, body mass index (BMI), physical activity, and gender. Leveraging a secondary dataset comprising 374 participants aged 27 to 59, statistical analyses, including regression modeling and visualization techniques, were employed to elucidate these associations. The results demonstrated that stress levels significantly impacted sleep quality, with higher stress consistently associated with lower sleep quality across both genders. Interestingly, BMI demonstrated an inverse relationship with sleep quality, while physical activity showed only a weak correlation, suggesting that its influence on sleep may be moderated by other factors such as stress. Gender-specific differences were also observed, with females reporting higher average sleep quality than males. These findings highlight the importance of maintaining adequate sleep duration, stress management, and weight regulation in promoting better sleep quality. Despite limitations such as reliance on self-reported data and potential selection bias, this study offers evidence-based insights that could inform individual lifestyle adjustments and targeted public health interventions to improve sleep health.

Introduction

Sleep is an essential pillar of human health, playing a critical role in cognitive function, emotional well-being, and physical health. However, modern lifestyles often interfere with achieving optimal sleep, leading to widespread sleep disturbances and associated risks such as cardiovascular diseases, impaired cognitive performance, and mental health disorders. Despite growing awareness of its importance, many individuals struggle to improve their sleep quality. Understanding the factors that contribute to poor sleep quality is essential for developing effective interventions.

While previous research has highlighted the impact of individual factors such as sleep duration and stress levels, the interplay of multiple lifestyle metrics remains less explored. This study seeks to address this gap by examining the relationships between sleep quality and various factors, including sleep duration, physical activity, stress levels, BMI, and gender. Using data from a sample of 374 participants aged 27 to 59, the study aims to provide a comprehensive analysis of how these variables influence sleep quality.

Through statistical techniques such as regression modeling and visualizations, this study investigates trends and associations that could inform actionable strategies to enhance sleep health. By exploring these relationships, the findings may contribute to a deeper understanding of the complex dynamics that shape sleep quality, with implications for public health and personal well-being.

Methods

Population and Sample

The population for this study includes all individuals aged 27 to 59 who could potentially have been surveyed about their sleep quality and other lifestyle factors. From this population, a random sample of 374 participants was drawn.

Equation 1: Calculation of the Mean ($\bar{x}$), where n is the number of data points and x_i is the individual data points.

$$\bar{x} = \frac{\sum x_i}{n}$$

Equation 2: Calculation of the Standard Deviation (s), where $\bar{x}$ is the sample mean, n is the number of data points, and x_i is the individual data points.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

The sample's mean sleep quality was 7.313, according to Equation 1, with a standard deviation of 1.197, according to Equation 2. A 95% confidence interval for the population mean sleep quality was calculated to estimate the true average sleep quality using a one-sample *t*-interval.

The sample needs to meet all necessary conditions to ensure the calculation is valid. The sample was randomly selected, satisfying the randomness requirement. The sample size of 374 exceeded 30, meeting the condition for normal approximation. Additionally, the sample represented less than 10% of the population of adults aged 27 to 59, fulfilling the 10% condition. Since all conditions were satisfied, the confidence interval calculation could proceed.

Equation 3: Formula for calculating the confidence interval for the population mean, where $\bar{x}$ is the sample mean, s is the sample standard deviation, n is the sample size, and $t*$ is the critical value for 95% confidence.

$$\bar{x} \pm t * \frac{s}{\sqrt{n}}$$

Equation 4: Formula to determine the Critical $t*$-Value.

$$t *= Inverse\ t - distribution\ at\ \frac{\propto}{2}$$

Equation 5: Calculation for Significance Level

$$\propto = 1 - confidence\ level$$

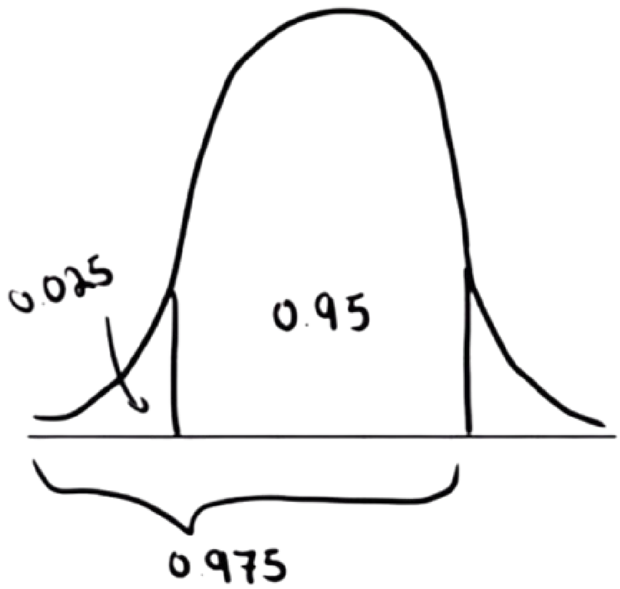

Figure 1. Visualization of the 95% Confidence Interval and Significance Levels

The diagram illustrates the 95% confidence interval, with a confidence level of 0.95 represented in the central portion of the t-distribution and a significance level (∝= 0.05) divided equally between the two tails (0.025 each).

Equation 6: Calculation of Degrees of Freedom

$$df = n - 1$$

Substituting appropriate values in the equations above yields a confidence interval of $[7.307, 7.319]$, indicating that, with 95% confidence, the true population mean sleep quality lies within this range.

This confidence interval establishes a range within which the true mean sleep quality for the population is likely to lie with 95% confidence. By providing a statistically reliable estimate of the population mean, the interval validates the representativeness of the sample and ensures that subsequent analyses and conclusions about sleep quality are grounded in robust statistical inference.

Data Collection

This study utilized secondary data obtained from a publicly available Kaggle dataset titled "Sleep Health and Lifestyle Dataset." The dataset was selected for its comprehensive inclusion of variables relevant to understanding sleep quality. As the data was pre-collected, no additional surveys or experiments were conducted. The dataset was cleaned and processed to ensure all analyses were based on valid and relevant entries.

Variables and Metrics

The response variable in this study is sleep quality, measured on a scale from 1 to 10. Explanatory variables include gender (categorical: male or female), BMI category (categorical: normal, overweight, obese), sleep duration (quantitative: hours of sleep per day), physical activity (quantitative: minutes of physical activity per day), and stress level (quantitative: perceived stress on a scale of 1-10)

Statistical Techniques and Visualizations

To explore the relationships between the variables, several statistical techniques and visualizations were employed. Descriptive statistics, including means, standard deviations, and a five-number summary (minimum, Q1, median, Q3, and maximum), were calculated for key variables to summarize the central tendency and variability in the data. Visualizations were created to illustrate these relationships and included bar graphs for displaying distributions and categorical comparisons (e.g., sleep quality by BMI category), scatterplots for examining correlations between quantitative variables (e.g., sleep duration and sleep quality), and box-and-whisker plots to compare sleep quality across gender groups. Grouped bar graphs were also to visualize interactions between stress level, gender, and sleep quality.

A linear regression model was used to quantify the strength and direction of relationships between sleep duration and sleep quality. Key metrics, including the correlation coefficient (r) and coefficient of determination (R^2), were used to assess the significance of the relationship.

Results

Bar Graph of Sleep Quality Scores

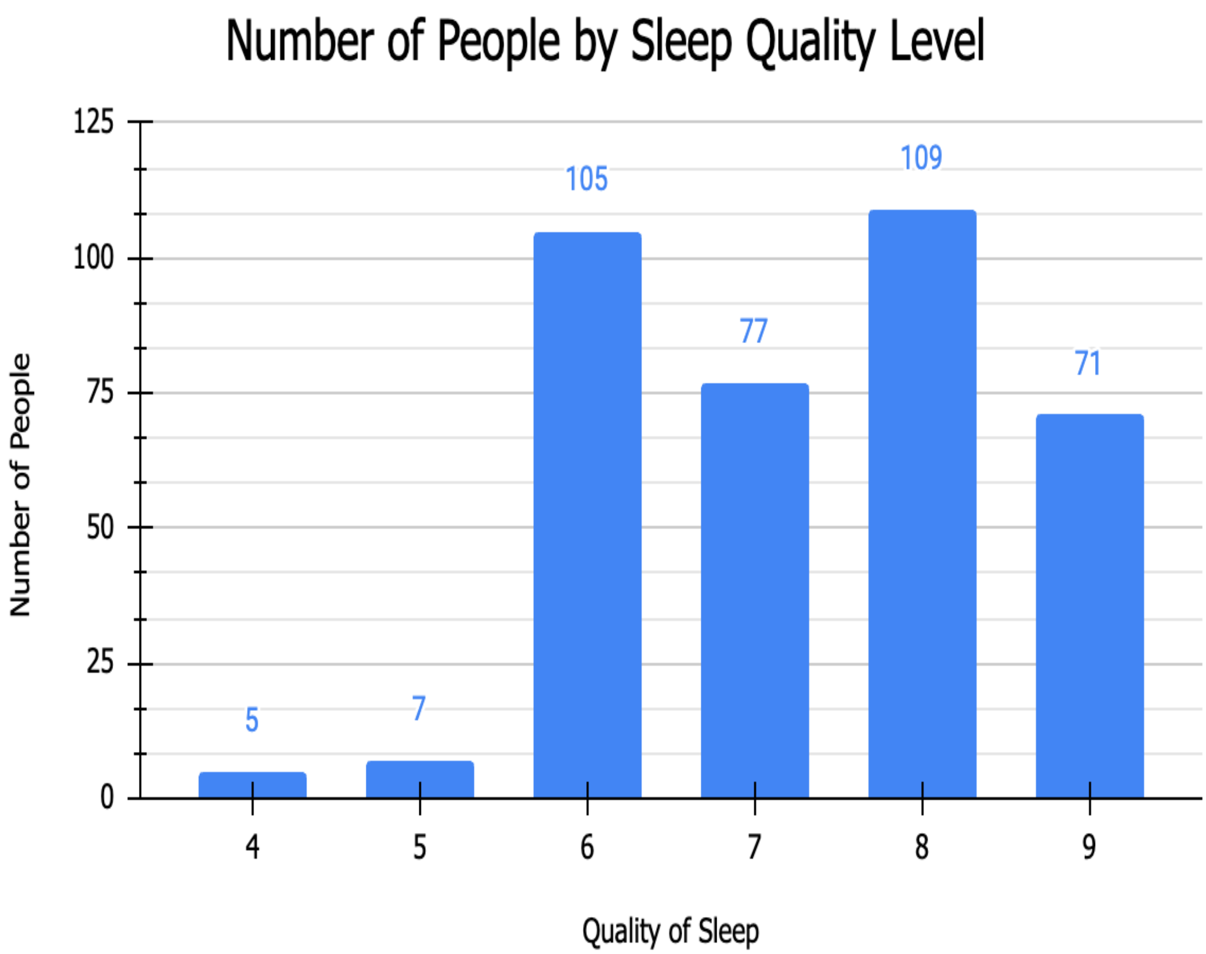

Figure 2. Distribution of Sleep Quality Scores. Each bar represents the number of participants reporting a given sleep quality score on a scale of 1-10.

The bar graph in Figure 2 highlights the frequency of each rating on a scale from 1 to 10. The distribution is left-skewed, with higher sleep quality ratings (6, 7, 8, and 9) occurring more frequently, while lower ratings (4 and 5) are less common. Most participants reported relatively high sleep quality, as indicated by the absence of ratings below 4 or at 10. The most common ratings were 6 and 8, with 105 and 109 responses, respectively, whereas the least common ratings were 4 and 5, reported by 5 and 7 participants, respectively.

The mean sleep quality was 7.313, with a standard deviation of 1.197 according to Equation 1 and Equation 2 respectively. This indicates that participants generally rated their sleep quality toward the higher end of the scale, with moderate variability. The standard deviation suggests that most responses fell within approximately 1.2 points of the mean, reflecting a relatively tight clustering of ratings around the central value.

Scatterplot of Sleep Duration vs. Sleep Quality

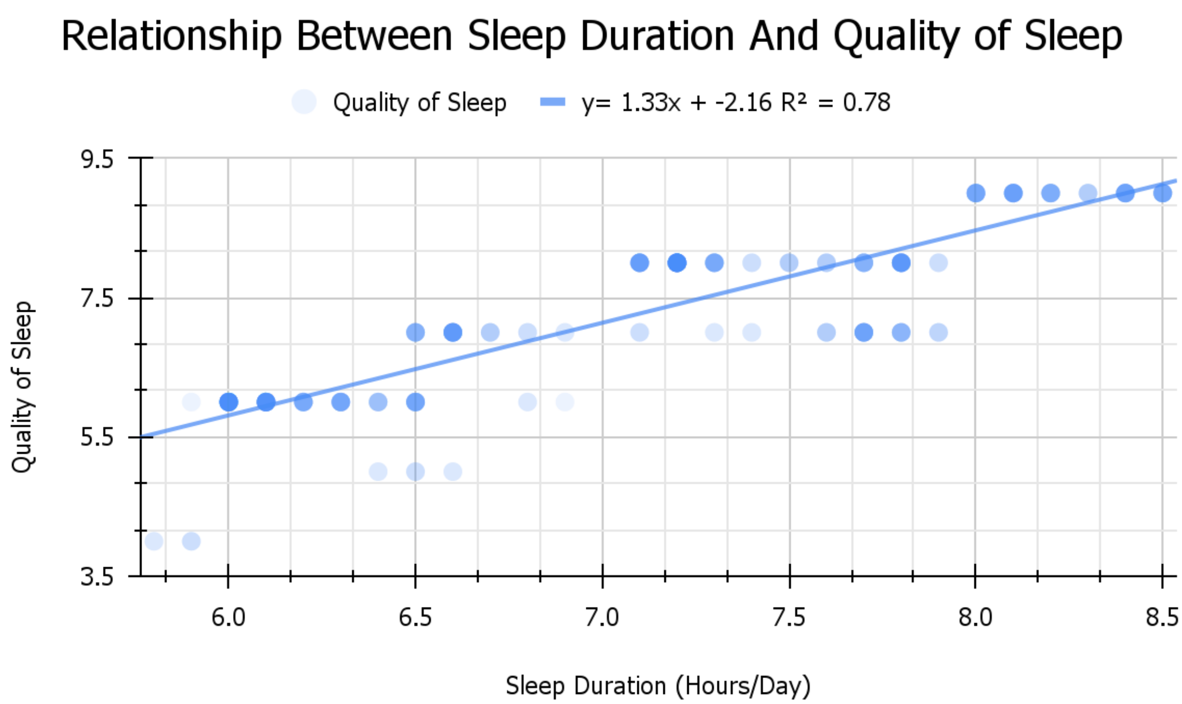

Figure 3. Relationship Between Sleep Duration and Sleep Quality. Each point represents an individual participant's sleep duration and corresponding sleep quality

A scatterplot was used to display the relationship between sleep duration (hours/day) and sleep quality. This visualization effectively illustrates the correlation between these two quantitative variables. Each point on the scatterplot represents an individual participant's sleep duration and corresponding sleep quality. The color intensity indicates the density of overlapping data points. Darker blue points highlight areas with higher concentrations of participants who reported similar sleep durations and sleep quality ratings, while lighter blue points show lower concentrations of participants with those specific values.

The scatterplot reveals clusters of data points around sleep durations of 6 hours, 7.25 hours, and 8 hours, suggesting that many participants sleep for these durations. The dense clustering at these durations indicates that a significant proportion of participants reported similar sleep quality ratings within these ranges. Negative outliers are visible around 7.75 hours and 6.5 hours of sleep, where the observed values lie below the regression line, indicating lower-than-expected sleep quality ratings based on the regression model.

Equation 7: Coefficient of Determination (R^2), where $SS_{residual}$ is the sum of the squared residuals (difference between observed and predicted values), and SS_{total} is the total sum of the squares (differences between observed values and mean).

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$$

The analysis generated an R^2 value of 0.78, using Equation 7. This coefficient of determination demonstrates that 78% of the variability in sleep quality can be explained by sleep duration. The high R^2 value signifies a strong positive relationship between these variables, confirming that sleep duration serves as a significant predictor of sleep quality.

Equation 8: Correlation Coefficient (r)

$$r = \pm\sqrt{R^2}$$

The correlation coefficient (r), calculated as the square root of R^2, is approximately 0.88. This value, close to 1, indicates a very strong positive relationship. The positive sign of the correlation coefficient shows that as sleep duration increases, sleep quality also tends to increase; conversely, shorter sleep durations correlate with lower sleep quality.

Equation 9: Regression Line Slope (b_1), s_y is the standard deviation of sleep quality, and s_x is the standard deviation of the sleep duration

$$b_1 = r \cdot \frac{s_y}{s_x}$$

The regression line's slope, calculated as 1.33 from Equation 9, further quantifies this relationship. For every additional hour of sleep, the average sleep quality score increases by 1.33 points. These findings underscore the importance of sleep duration as a key factor influencing sleep quality and provide robust evidence of a strong positive relationship between the two variables.

Bar Graph of Average Sleep Quality by BMI Category

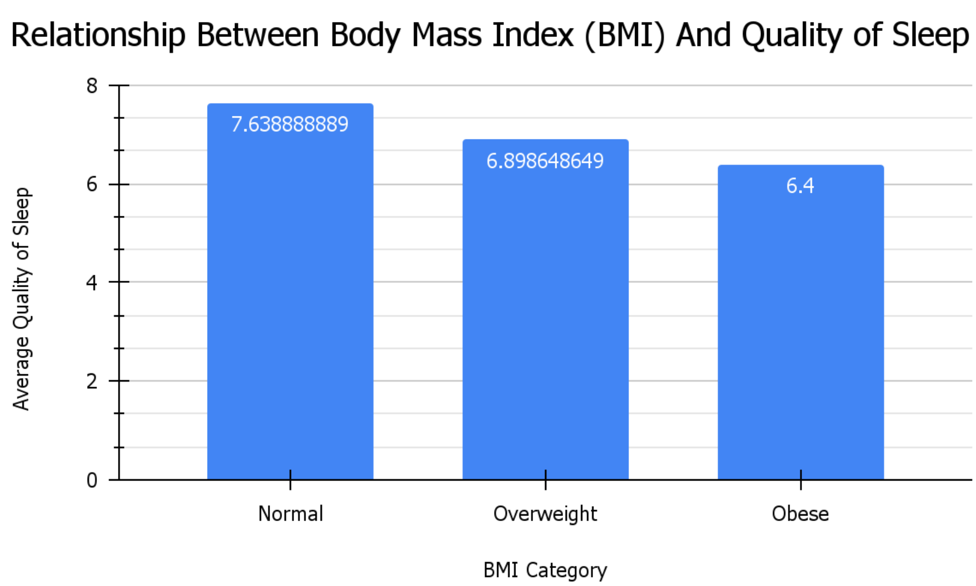

Figure 4. Average Sleep Quality by BMI Category. Each bar represents the mean sleep quality for participants within a specific BMI category.

The implications of body mass index (BMI) on sleep quality were analyzed by calculating the average sleep quality for three BMI categories: Normal, Overweight, and Obese. A bar graph was used to display the results, highlighting trends across these categories. The analysis revealed that participants in the Normal BMI category reported the highest average sleep quality of approximately 7.64. Those in the Overweight BMI category had an average sleep quality of 6.90, while participants in the Obese category reported the lowest average sleep quality of 6.40.

The bar graph illustrates a clear inverse relationship between BMI and sleep quality. As BMI increases from Normal to Overweight to Obese, the average sleep quality decreases. This trend suggests that maintaining a Normal BMI may be associated with better sleep quality, while higher BMI categories might contribute to poorer sleep outcomes.

These findings underscore the potential importance of body weight management in promoting better sleep quality. Further investigation could explore whether this trend reflects physiological factors associated with BMI, such as metabolic or hormonal changes, or external factors like lifestyle and stress, which may also influence sleep

Grouped Bar Graph of Gender, Stress Level, and Sleep Quality

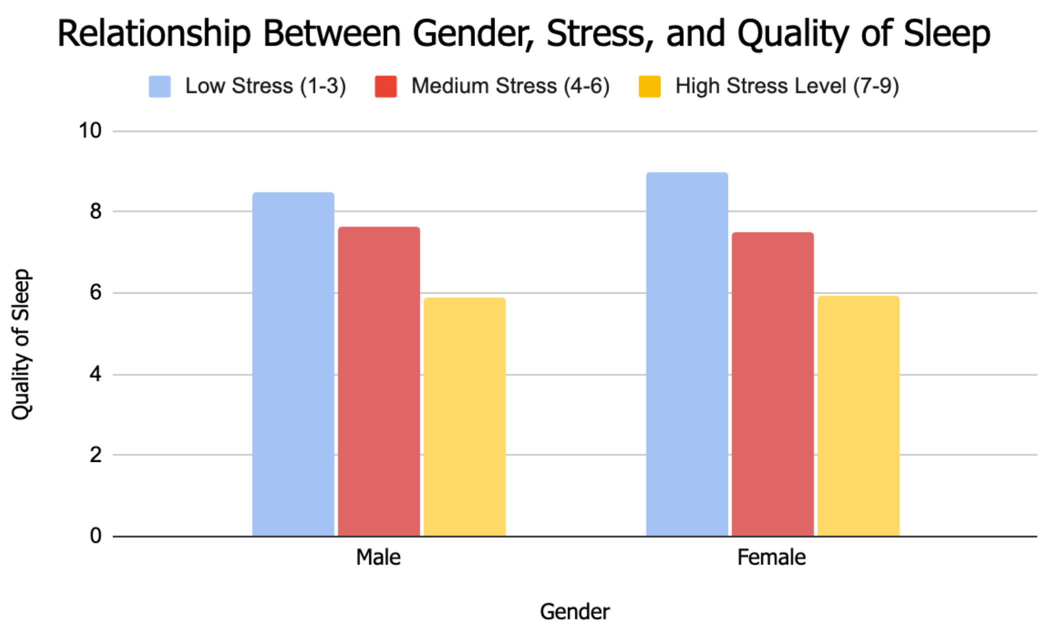

Figure 5. Relationship Between Gender, Stress Level, and Sleep Quality. Bars are grouped by stress levels and further divided by gender.

A grouped bar graph was created to examine the relationship between stress levels, sleep quality, and gender. Stress ratings were categorized into three groups: low stress (1-3), medium stress (4-6), and high stress (7-9).

The graph reveals a clear inverse relationship between stress levels and sleep quality for both genders. As stress levels increase from low to high, the average quality of sleep decreases consistently. Within each stress category, the differences in sleep quality between males and females are relatively small. However, females report slightly higher sleep quality than males in the low and high stress categories. In the medium stress category, males demonstrate marginally higher sleep quality than females.

Table 1. Conditional probability numerical summary between stress level and gender.

	Low Stress (1-3)	Medium Stress (4-6)	High Stress (7-9)	Total
Male	4	111	74	189
Female	67	72	46	185
Total	71	183	120	374

Equation 10: Conditional Probability Formula

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Where $P(A \mid B)$ is the probability of event A (e.g., a male experiencing low stress) given event B (e.g., being male), $\mathrm{P}(A \cap B)$ is the joint probability of events A and B occurring together (e.g., the proportion of males in the sample experiencing low stress), and $P(B)$ is the probability of event B occurring (e.g., the proportion of males in the sample).

Table 2. Conditional probability of stress levels by gender.

	Low Stress (1-3)	Medium Stress (4-6)	High Stress (7-9)
Male	0.021	0.587	0.392
Female	0.362	0.389	0.249

The conditional probabilities calculated with Equation 10 illustrated in Table 2 reveal significant differences in the distribution of stress levels between males and females, which help explain observed differences in sleep quality. Among males, the probability of experiencing medium stress is notably higher at 58.7%, compared to 2.1% for low stress and 39.2% for high stress. In contrast, females exhibit a more balanced distribution, with 36.2% experiencing low stress, 38.9% experiencing medium stress, and 24.9% experiencing high stress.

These differences in stress distribution suggest gender-specific patterns that could inform targeted interventions. For males, where medium and high stress levels are more prevalent, stress management strategies might prioritize reducing high stress and mitigating its effects. For females, maintaining low stress levels could be a key focus, alongside addressing medium and high stress levels as needed.

Recognizing these gender-specific patterns allows for the development of more effective and personalized approaches to stress management, which may subsequently improve sleep quality for both groups.

Scatterplot of Physical Activity vs. Sleep Quality

HIGH SCHOOL EDITION
Journal of Student Research

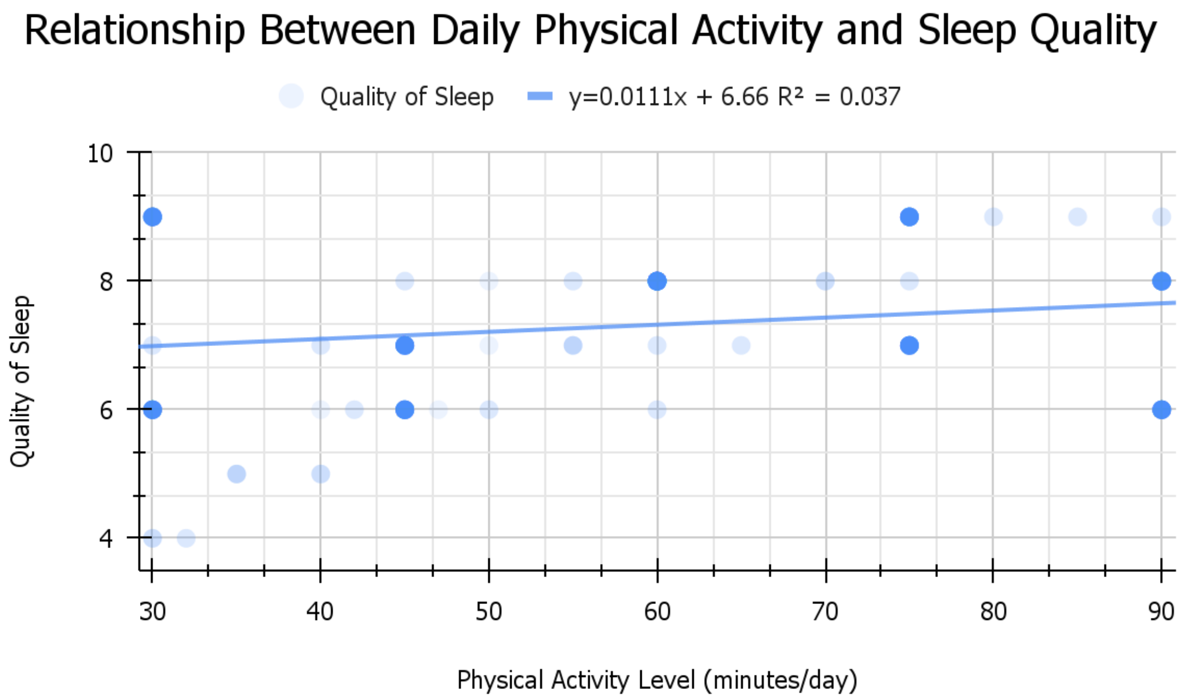

Figure 6. Scatterplot of Physical Activity vs. Sleep Quality. Each point represents an individual participant's physical activity level and corresponding sleep quality.

The scatterplot illustrates the relationship between physical activity (minutes per day) and sleep quality. The color intensity represents the density of overlapping data points, with darker blue areas indicating higher concentrations of participants with similar activity levels and sleep quality ratings, while lighter blue areas indicate lower concentrations. Clusters of data points are evident around common activity levels, such as 30, 45, 75, and 90 minutes per day. However, the scatterplot also reveals more negative outliers than positive ones, suggesting that factors such as stress, health conditions, or lifestyle habits may adversely affect sleep quality, even at higher levels of physical activity.

The coefficient of determination (R^2) was calculated as 0.037 with Equation 7, indicating that only 3.7% of the variability in sleep quality can be explained by physical activity duration. This low R^2 value demonstrates a weak relationship between these variables, suggesting that most of the variation in sleep quality is likely attributable to factors not captured in this graph. The correlation coefficient (r) was approximately 0.192 as calculated with Equation 8, further confirming the weak positive relationship.

The positive sign of the correlation coefficient and the small slope of the trendline (0.0111), calculated with Equation 9, indicate that as physical activity increases, sleep quality also tends to increase, but only minimally. These findings suggest that physical activity alone has a limited impact on sleep quality.

Box Plot of Gender vs. Sleep Quality

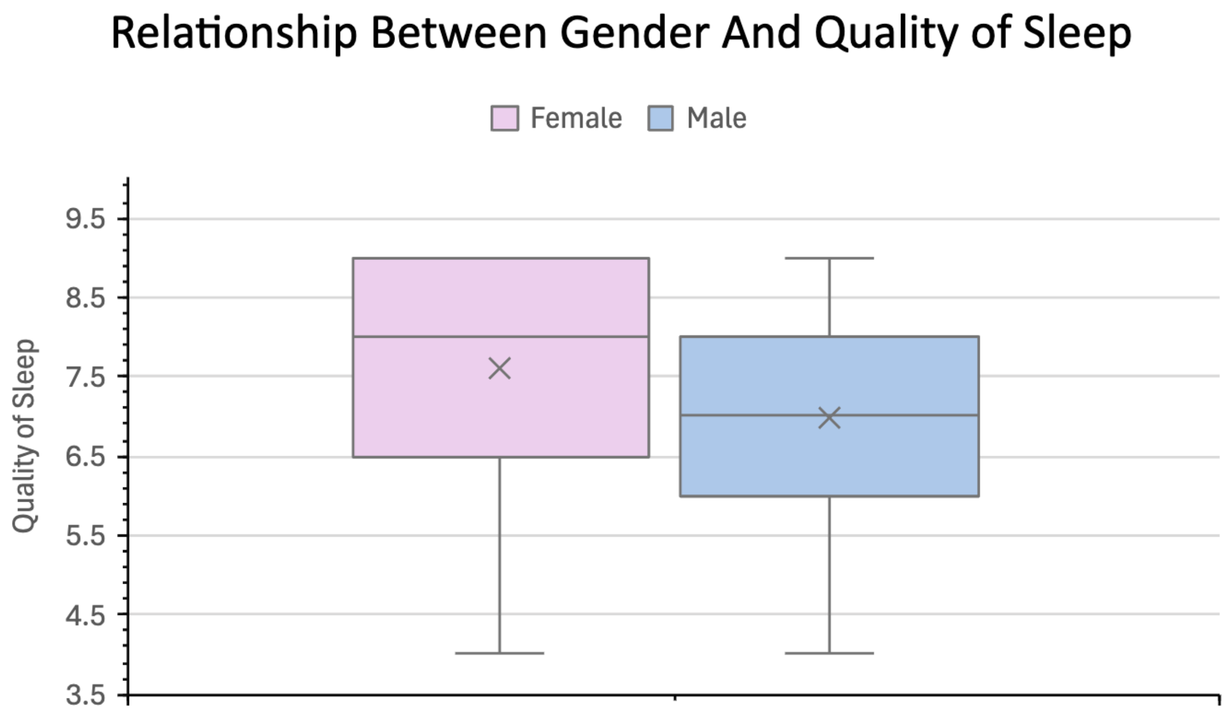

Figure 7. Scatterplot of Physical Activity vs. Sleep Quality. Each point represents an individual participant's physical activity level and corresponding sleep quality.

The analysis of sleep quality by gender reveals distinct differences in distribution and central tendencies. For females, the minimum reported sleep quality is 4, the maximum is 9, and the interquartile range (IQR) spans from 7 to 9. For males, the minimum and maximum sleep quality scores are also 4 and 9, but the IQR ranges from 6 to 8. The median sleep quality for females is 8, with a mean slightly lower at 7.67, while the median for males is 7, with a mean of 6.97. These metrics indicate that, on average, females report higher sleep quality than males, as reflected in both the higher median and mean values.

Equation 10: Interquartile Range Formula, where Q_3 represents the values below which 75% of the data falls, and Q_1 represents the values below which 25% of the data falls.

$$IQR = Q_3 - Q_1$$

While both genders have an IQR of 2, the higher position of the IQR for females (7 to 9) compared to males (6 to 8) reinforces the observation that females consistently report better sleep quality. Interestingly, although the minimum and maximum scores are identical for both genders, the distribution within these ranges differs. This disparity highlights gender-specific trends in sleep quality, with females tending toward higher scores.

The results from this boxplot align with the trends observed in Figure 5, which compared gender, stress levels, and sleep quality. Both visualizations suggest that females generally report better sleep quality than males, even when accounting for stress variations. This consistency strengthens the conclusion that gender plays a significant role in influencing sleep quality. Potential contributing factors could include differences in stress management strategies, hormonal influences, and variations in sleep hygiene practices.

An additional observation is that the boxplot for females lacks an upper whisker, indicating no outliers on the higher end of the scale. The third quartile and maximum value are both 9, suggesting a ceiling effect in female sleep quality scores. This capped distribution may reflect consistent high-quality sleep among females at the upper range.

Conclusion

Overall, this study provides several insights into the relationship between lifestyle factors, sleep metrics, and their impacts on sleep quality. Stress levels appear to have a significant effect on sleep quality, as shown in Figure 5, where high stress levels were consistently associated with lower sleep quality across both genders. Additionally, the Box and Whisker Plot in Figure 7 indicates that females tend to report higher sleep quality levels than males, suggesting that gender-specific factors may influence sleep quality.

The correlation between physical activity and sleep quality, shown in Figure 6, was weaker than expected. While a slight positive correlation suggests that increased physical activity might improve sleep quality, the variability in the data points indicates that other factors, such as stress levels, may play a more prominent role. This finding highlights the individualized nature of physical activity's effects on sleep quality and underscores the complexity of factors influencing sleep.

Figure 4 demonstrates that higher BMI is associated with lower sleep quality. Participants in the normal BMI category reported better sleep quality compared to those in the overweight and obese categories, suggesting that maintaining a healthy weight may support better sleep quality. Furthermore, Figure 3 reveals a strong positive correlation between sleep duration and sleep quality, with a high R^2 value of 0.78 indicating that 78% of the variability in sleep quality can be explained by sleep duration. This result supports existing recommendations that adults obtain at least seven hours of sleep per night to promote optimal health and well-being.

Limitations

Despite conclusions about some significant relationships between lifestyle factors/sleep metrics and sleep quality, there are still limitations that need to be taken into account. First, the secondary data used from the randomly collected data may not fully represent the broader population which limits the generalizability of the conclusions. For instance, the participants in the secondary data were mainly working in established career sectors such as healthcare, education, technology, or sales. The limited variety of occupations means that the findings might not apply to individuals in other fields or those with different work patterns or stressors, resulting in selection bias. Additionally, since all the data was recorded over a single point in time, it is difficult to infer causation between various lifestyle factors and sleep quality. Furthermore, inconsistencies in how participants interpret and respond to questions about sleep quality and lifestyle factors can introduce measurement bias which limits the generalizability of the conclusions. The study from which the secondary data was sourced may not have accounted for all potential confounding variables that could influence sleep quality, including medication use or pre-existing health conditions like insomnia.

Overall, despite some limitations, the large sample size of 374 participants enhances the statistical power of the results and the comprehensive analysis of multiple lifestyle factors and their relationship to sleep quality provides a holistic understanding of sleep health. Thus, there is relatively strong evidence for the conclusions drawn about the relationship between lifestyle factors/sleep metrics and sleep quality. Ultimately, the findings in this study could be used to guide individual lifestyle changes and develop targeted interventions aimed at improving sleep quality and overall health.

A follow-up study could analyze the relationship between additional factors in the secondary dataset, such as age and blood pressure. Future research could also focus on more diverse samples, including participants from a wide range of occupations and backgrounds, to enhance the generalizability of the findings. Additionally, exploring the effects of specific interventions, such as stress management programs or dietary changes, on sleep quality would provide valuable insights into strategies for improving individual sleep outcomes.

Acknowledgments

I would like to thank Ms. Festarini, my AP Statistics Teacher, for her guidance during the course of this project.

References

American Psychological Association. (2013). *Stress and sleep*. American Psychological Association. https://www.apa.org/news/press/releases/stress/2013/sleep

Benisek, A. (2023, February 16). *How do age, gender, and other social and economic factors affect sleep?*. WebMD. https://www.webmd.com/sleep-disorders/features/sleep-quality-socioeconomic-factors

Exercising for better sleep. (2024, June 20). Johns Hopkins Medicine. (2024). https://www.hopkinsmedicine.org/health/wellness-and-prevention/exercising-for-better-sleep

Fatima, Y., Doi, S. A., Najman, J. M., & Mamun, A. A. (2016). Exploring Gender Difference in Sleep Quality of Young Adults: Findings from a Large Population Study. *Clinical Medicine & Research*, 14(3–4), 138–144. https://doi.org/10.3121/cmr.2016.1338

Gupta, P., Srivastava, N., Gupta, V., Tiwari, S., & Banerjee, M. (2022). Association of sleep duration and sleep quality with body mass index among young adults. *Journal of Family Medicine and Primary Care*, 11(6), 3251–3256. https://doi.org/10.4103/jfmpc.jfmpc_21_21

Kalmbach, D. A., Anderson, J. R., & Drake, C. L. (2018, May 24). The impact of stress on sleep: Pathogenic sleep reactivity as a vulnerability to insomnia and circadian disorders. *Journal of Sleep Research, 27*(6). https://doi.org/10.1111/jsr.12710

Starnes, D. S., Yates, D., & Moore, D. S. (2010). *The Practice of Statistics*. Macmillan.

Tharmalingam, L. (2023, September 18). *Sleep health and lifestyle dataset*. Kaggle. https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset