Unveiling Patterns: Analyzing Sleep Habits and Lifestyle Choices

Rupakheti Sijal
Masters in Health Informatics
Kathmandu University
Kavre, Nepal
rupakhetisijal7@gmail.com

***Abstract*—There have been limited studies showing the relationship between sleep health and lifestyle. Even in the available studies, the relationship between physical activity and sleep health have not been properly described. This study aims to fill that gap by examining the correlation between physical activity and sleep health. The dataset was extracted from Kaggle and was preprocessed and cleaned using the 'pandas' library in python. Exploratory data analysis was done and relevant visualizations were done using 'matplotlib' and 'seaborn' libraries in python. T-test was done using the 'scipy' library of python. A correlation coefficient of 0.21 was seen between physical activity level and sleep duration, indicating a weak positive correlation. Machine learning algorithms like linear regression models were trained for prediction of sleep duration according to physical activity level. Mean squared error was found to be 0.59 with coefficients of 0.005 and intercept of 6.82, indicating accurate prediction of sleep duration using the linear regression model.**

Keywords—sleep, physical activity, body mass index

I. Introduction

Understanding the relationship between sleep and physical activity has received significant attention in recent times in the research field, particularly due to their deep implications for overall health and wellbeing across diverse populations. The literature review adds up findings from various studies to explore the multi-dimensional associations between sleep habits and lifestyle choices, focusing on university students, adults, children, and specific occupational groups[1],[3],[4]. While some studies suggest a positive correlation between moderate-to-high intensity physical activity and improved sleep quality, other studies do not have clear conclusions and some even have contradicting findings[1],[2],[5],[7]. In addition, the impact of physical activity on sleep extends beyond duration or intensity, with influencing factors such as stress level and overall quality of life[6]. This review highlights the need for continued research on the relationship between physical activity, stress and lifestyle choices with sleeping habits. There is also a need for prediction of sleeping habits including sleeping duration and sleeping disorders based on the physical activity level, stress level and age of the individuals. So this study aims to answer all these questions and provide a comprehensive view on the relationship between lifestyle choices and sleep habits, including training models for prediction of sleep duration based on available data.

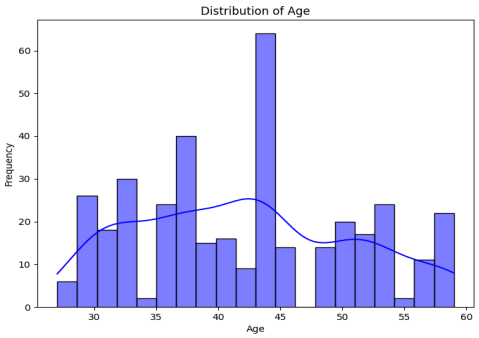
II. Methodology

A thorough browsing of the web was done for selecting relevant dataset. Sources like official WHO, CDC websites, data hosting platforms like Kaggle, etc were browsed thoroughly. Finally, a dataset titled 'Sleep health and Lifestyle' was chosen. The data was pre-processed thoroughly using the 'pandas' library of the python programming language. The pre-processing steps included cleaning of the data, making the values consistent across rows and columns, normalizing and standardizing data,

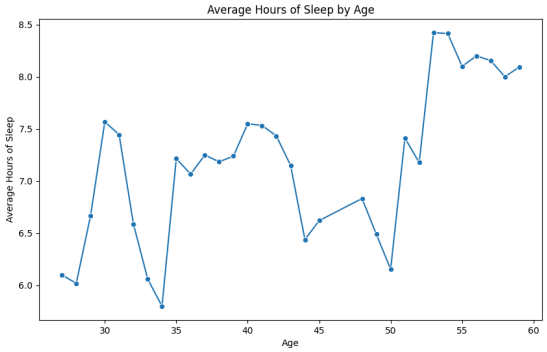
transforming the data, filling in the null values and making it ready for analysis. After this processing, necessary data visualizations were done using the 'matplotlib' and 'seaborn' libraries of python. A null hypothesis was formulated stating 'Physical activity is not related to sleep habit'. T-test was carried out to test for the hypothesis using the 'ttest_ind' function from the statistics module of the 'scipy' library. Machine learning algorithms like the 'Linear Regression' model were trained for prediction of sleep duration based upon the physical activity level.

III. Results

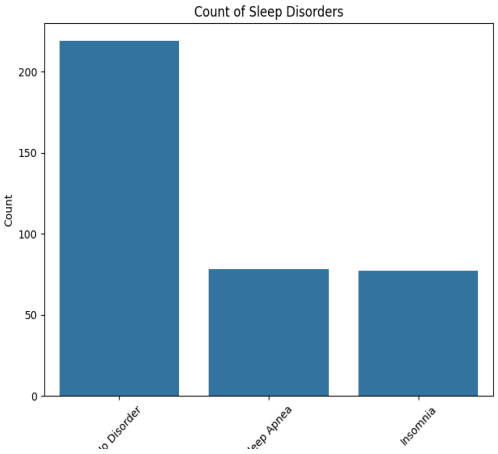
The study included the working population with age ranging from 25-60. The highest number of people included in the study were within the age range of 35-45 years.



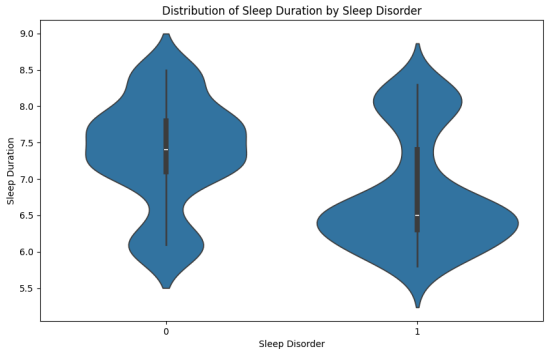
A line plot where age was plotted against the average sleep duration showed that people in their 50s had the most hours of sleep followed by people in their 30s and 40s.



A bar diagram of prevalence of sleep disorder showed that most of the people in the study didn't have any forms of sleep disorder. The remaining population were diagnosed equally with either insomnia or sleep apnea.

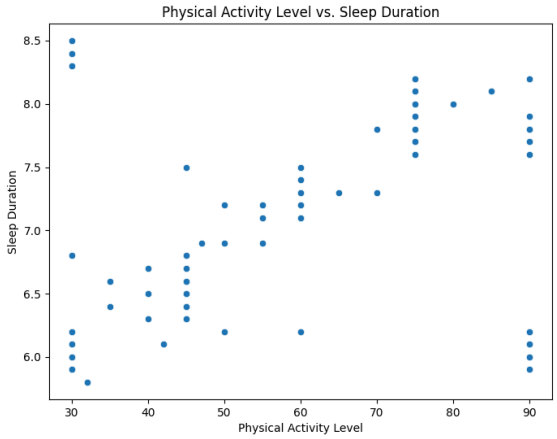


A violin plot which plotted sleep disorder against number of sleep hours showed that most of the people with no sleep disorder had, on an average, 7-8 hours of sleep, whereas those with sleep disorder had a wide range of values ranging from 5.5 hours- 7 hours. This can be generalized saying, the less sleep people get, the more likely they are to have a sleep disorder.



A scatter plot which compared physical activity level with sleep duration showed an interesting trend. The more physically active people were, the more likely

they were to sleep for longer duration in comparison to the less physically active people.



Although this data is not consistent throughout all populations, it seems to be just enough to establish a trend.

Hypothesis Testing

Hypothesis were formulated as;

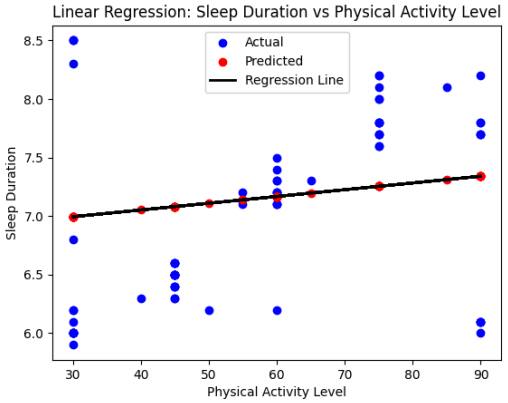
Null Hypothesis (H0): There is no significant relationship between sleep and physical activity among the studied populations.

Alternate Hypothesis (H1): There is a significant relationship between sleep and physical activity among the studied populations.

T-test was done to test for the hypotheses using the 'ttest_ind' function from the statistics module of the 'scipy' library. Results showed a correlation coefficient of 0.21 indicating a weak positive correlation between physical activity level and sleep duration. This suggested that as physical activity level increases, sleep duration tends to increase slightly, but the relationship is not very strong.

Furthermore, the T-statistic, which measures the size of the difference in mean sleep duration between the high and low physical activity groups relative to the variability within each group was calculated to be 6.53 with P-value 0.0000001 which showed that there is a statistically significant difference in sleep duration between individuals with high and low physical activity levels in your dataset. Specifically, those with high physical activity levels tend to have longer sleep durations compared to those with low physical activity levels.

The Linear Regression Model was trained for predicting sleep duration based on physical activity level. The Mean Squared Error(MSE) of the linear regression model was calculated to be 0.586. This value is a measure of the average squared difference between the actual and predicted values in the regression model. The value of 0.586 is a positive indicator, suggesting that the model's predictions are closer to the actual values, implying better model performance and indicating that the trained model is making reasonably accurate predictions.



The co-efficient, which refers to the weight assigned to each independent variable, was calculated for 'Physical Activity Level', which is the independent variable (feature) in our model. The value was calculated to be 0.00576912. This value implies that, for every unit increase in physical activity level, the predicted sleep duration increases approximately by 0.00576912 hours or about 21 seconds. Finally, intercept(bias) was calculated for the linear regression model and was found to be 6.82. This value is significant for providing baseline prediction, meaning when all other independent variables (Physical Activity Level in our case) are set to zero, the model predicts a baseline sleep duration of approximately 6.82 hours.

Overall, the machine learning model trained i.e. the linear regression model is fairly accurate in terms of prediction of the sleep duration in accordance with the physical activity level.

IV. Discussion

Findings from this study suggest a positive correlation between physical activity level and sleep duration among the individuals. This means increase in sleep quality among the individuals is directly proportional to increase in physical activity level. Contrary to previous studies conducted in similar topics[1],[3],[5], this study actually establishes a value corresponding to the impact the physical activity level has on the sleep quality of the individuals. Although the correlation coefficient of 0.21 is deemed to be low, the value at least gives us a direction of relationship between physical activity level and sleep quality.

The value of T-statistic, calculated to be 6.53 with p-value of 0.0000001, obtained from the t-test, signifies that that there is a statistically significant difference between sleep duration between the high and low physical activity groups and the results obtained are not merely due to random chance. This finding is in the direction of what we expected as higher levels of physical activity, in general, do tend to result in better sleep quality and longer duration of sleep.

The linear regression model was developed by providing 80% of training data and 20% of testing data. The MSE value of the model was calculated to be 0.586, which implies that the model is fairly accurate in terms of predicting the sleep duration based on the level of physical activity. In general, the closer the values of MSE are to zero, the more accurate the predictions of the models are. In our case, the finding of 0.586 shows that the predicted value of sleep duration according to the physical activity level is in line with the actual values of sleep duration. This comparison is also plotted in a scatter plot, which shows the data points of predicted values are very close to the data point of the actual values of sleep duration.

V. Conclusion

In conclusion, physical activity level is positively correlated with duration of sleep, although the correlation coefficient is low (0.21). This finding suggests that sleep duration is influenced not only by physical activity level but other factors too. Machine learning algorithms such as the linear regression model which were developed in the study for predicting sleep duration based on physical activity levels were accurate in predicting the values. Further studies are suggested to be conducted exploring the various factors that influence sleep quality other than physical activity.

References

[1] Memon AR, Gupta CC, Crowther ME, Ferguson SA, Tuckwell GA, Vincent GE. Sleep and physical activity in university students: A systematic review and meta-analysis. Sleep medicine reviews. 2021 Aug 1;58:101482.

[2] Semplonius T, Willoughby T. Long-term links between physical activity and sleep quality. Med. Sci. Sports Exerc. 2018 Dec 1;50:2418-24.

[3] Atoui S, Chevance G, Romain AJ, Kingsbury C, Lachance JP, Bernard P. Daily associations between sleep and physical activity: A systematic review and meta-analysis. Sleep Medicine Reviews. 2021 June 1;57:101426.

[4] Antczak D, Lonsdale C, Lee J, Hilland T, Duncan MJ, del Pozo Cruz B, Hulteen RM, Parker PD, Sanders T. Physical activity and sleep are inconsistently related in healthy children: A systematic review and meta-analysis. Sleep Medicine Reviews. 2020 June 1;51:101278.

[5] Mead MP, Baron K, Sorby M, Irish LA. Daily associations between sleep and physical activity. International journal of behavioral medicine. 2019 Oct;26:562-8.

[6] Atkinson G, Davenne D. Relationships between sleep, physical activity and human health. Physiology & behavior. 2007 Feb 28;90(2-3):229-35.

[7] Al Nawwar MA, Alraddadi MI, Algethmi RA, Salem GA, Salem MA, Alharbi AA. The effect of physical activity on sleep quality and sleep disorder: a systematic review. Cureus. 2023 Aug 16;15(8).

[8] Kline CE. The bidirectional relationship between exercise and sleep: implications for exercise adherence and sleep improvement. American journal of lifestyle medicine. 2014 Nov;8(6):375-9.

[9] Arslan SS, Alemdaroğlu İ, Karaduman AA, Yilmaz ÖT. The effects of physical activity on sleep quality, job satisfaction, and quality of life in office workers. Work. 2019 Jan 1;63(1):3-7.

[10] Martins AJ, Vasconcelos SP, Skene DJ, Lowden A, de Castro Moreno CR. Effects of physical

activity at work and life-style on sleep in workers from an Amazonian Extractivist Reserve. Sleep Science. 2016 Oct 1;9(4):289-94.