Proposal for Investigating the Effect of the Quality and Quantity of Sleep on Productivity

An estimation of the effect of sleep on productivity through linear regression

Onon Burentuvshin Edric Shi

February 20, 2025

1 Introduction

A basic biological activity, sleep is essential for memory consolidation, cognitive function, and overall wellness. However, widespread sleep deprivation has been brought on by the modern work culture, which is characterized by long hours, high levels of stress, and increasing digital engagement. This has raised concerns about the influence of sleep deprivation on productivity. Knowing how sleep affects productivity is essential for companies and organizations trying for improved efficiency. In order to address the research question, "How are sleep and productivity related?" This study intends to investigate the relationship between sleep duration, quality, and productivity at work.

Prior studies have demonstrated a strong relationship between sleep and productivity at work. According to Ishibashi & Shimura (2020), presenteeism is caused by daytime dysfunction, and short sleep duration (less than five hours) and poor sleep quality considerably diminish workplace productivity. According to their research, while assessing performance at work, both subjective sleep quality and sleep duration should be taken into consideration. In a comparable manner, Gingerich et al. (2017) examined data from almost 600,000 workers in multiple industries and observed a U-shaped relationship between productivity loss and sleep duration, with both excessive sleep (>9 hours) and insufficient sleep (<6 hours) linked to higher presenteeism and absenteeism as they also highlighted fatigue as a strong predictor of productivity loss. Furthermore, a 2023 Nature article highlighted how sleep deprivation impairs cognitive functions, reaction times, and decision-making abilities, similar to the effects of alcohol consumption. The article also emphasized how outside variables, including technology use and work-related stress, can lead to unhealthy sleep patterns, which in return could affect productivity at work.

Given the established data associating between sleep and productivity, this study uses multiple linear regression (MLR) to quantify these relationships. Linear regression is a useful statistical method because it allows researchers to investigate how numerous predictors, such as sleep duration, sleep quality, fatigue levels, and work schedule, affect a continuous response variable, productivity. Furthermore, regression analysis allows for the discovery of potential non-linear trends (e.g., U-shaped relationships) as well as predictor interaction effects. Applying MLR, this study will provide empirical insights into how sleep-related characteristics affect workplace efficiency, potentially informing policies to increase employee well-being and performance.

2 Data Description

Our analysis uses a dataset of 5,000 unique observations tracking sleep habits and their relationship to productivity, mood, and stress. The data was sourced from Kaggle, uploaded by a user under the title "Sleep Cycle Productivity Dataset." For each observation, the original data set measures 15 unique variables. For the purposes of our analysis, three variable were dropped as it does not contribute to the depth of the analysis: Date of observation, Person_ID(identifier for each individual), and sleep start and end times.

2.1 Data Cleaning

Exercise Categorization Justification:

The original numerical exercise minutes variable was converted to low/medium/high categories using CDC activity guidelines (Centers for Disease Control and Prevention, 2023):

- Low: 0–30 mins/day (below recommended minimum)
- Medium: 30–60 mins/day (meets baseline recommendation)
- High: 60+ mins/day (exceeds recommendations)

With such categorization, it allows our analysis to capture stronger effects of exercise and test whether adherence to these guidelines creates heterogeneous effects of hours of sleep on productivity.

2.2 Data Collection Methodology

While the original curator did not explicitly document their methodology, the dataset's structure suggests it was collected through a combination of wearable sleep trackers (for variables like total sleep hours) and daily self-reported surveys (for productivity scores, mood, caffeine intake, and exercise). Participants likely recorded metrics over a 30-day period, as indicated by date stamps in the raw data.

2.3 Variables included in our analysis

2.3.1 Response Variable

Productivity Score:

- **Description:** A discrete metric (0–10) where participants rated their daily work/output efficiency.
- **Distribution**:

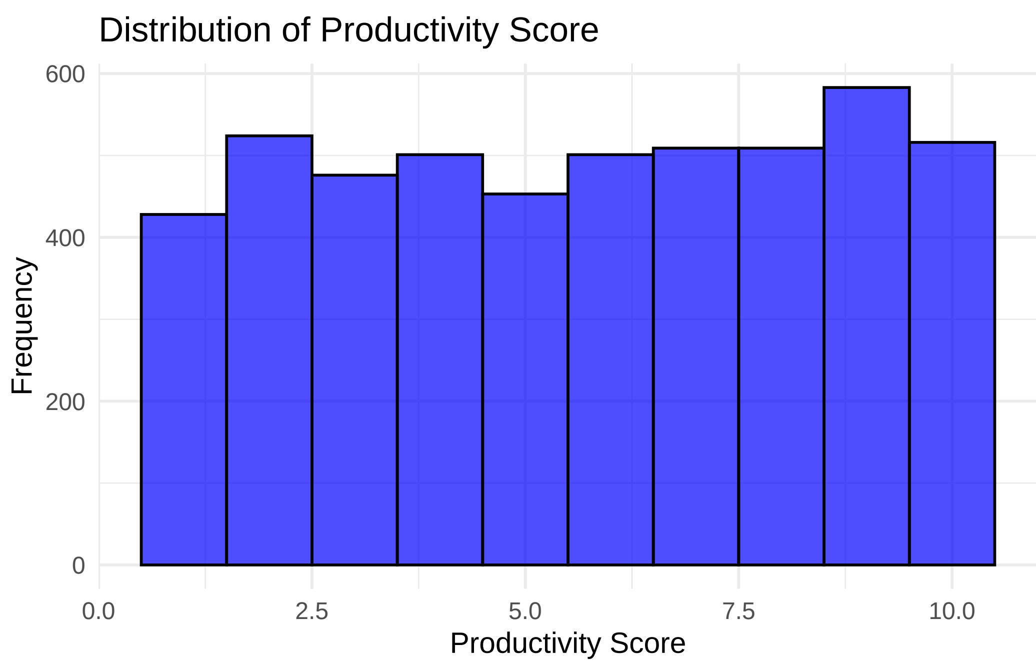

Figure 1

- **Linear Model Suitability**: The variable is roughly uniformly distributed. Consistent with our ideal of our data containing adequate range, there is variation that can be exploited for linear regression, however, the discrete nature of this response variable introduces issues to our model. A further discussion will be presented later in this proposal.

2.3.2 Predictor Variable

An overview of our predictor variables can be observed from Table 1. Variables such as sleep quality, screen time, work hours, and stress allow us to test the efficacy of literature discussed in the Introduction. As we can observe from Figure 2, every self-report 1-10 scale variable is uniformly distributed much like our response variable.

Table 1: Description of Predictor Variables

Variable	Type	Relevance
Total Sleep Hours	Continuous	Key predictor
Sleep Quality	Discrete (1-10)	Ishibashi & Shimura (2020) predictor on subjective sleep quality
Exercise Level	Categorical	Controls heterogeneous effects of different levels of exercise
Caffeine Intake (mg)	Continuous	Potential alertness and stress modulator on productivity
Screen Time Before Bed	Continuous	Confounder as longer screen times may result in less hours of sleep
Hours of Work per day	Continuous	Confounder as longer hours may reduce sleep time
Mood Score	Continuous (1-10)	Confounder as Mood may affect sleep quality and hours
Stress Level	Continuous (1-10)	Confounder as Stress may affect sleep quality and hours
Age	Continuous	Controls for the differing effects of Age
Gender	Categorical	Controls for the differing effects of Gender

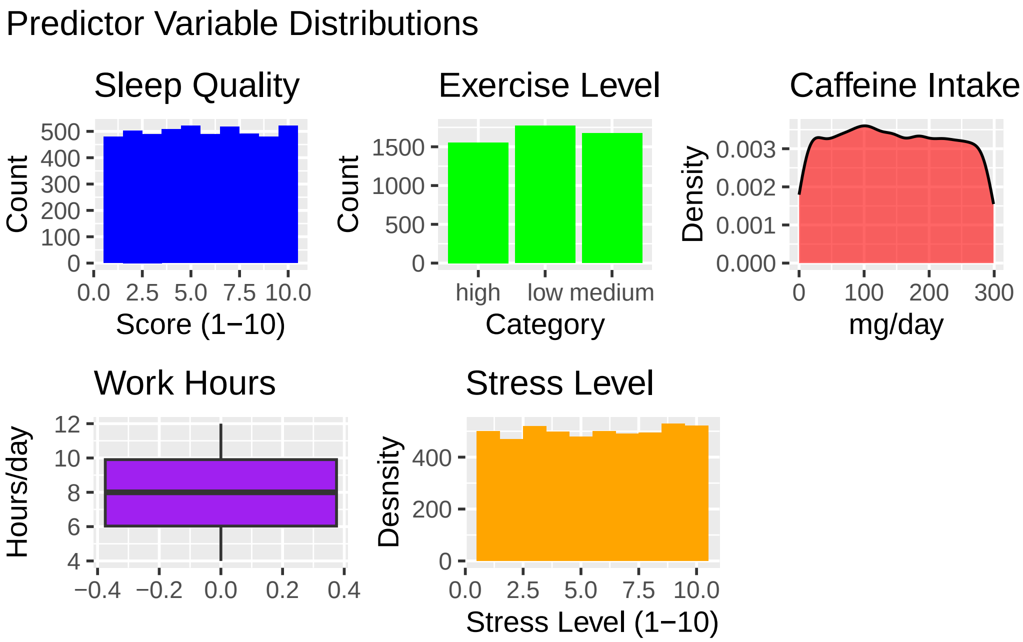

Figure 2: Distribution of Predictor Variables

3 Preliminary Results

3.1 Regression Results

Our initial analysis reveals key limitations in model specification and data structure (Table 2). Sleep duration and quality show negligible effects. Exercise demonstrates paradoxical impacts - productivity decreases by 0.18 points when moving from high to medium/low activity. Our preliminary findings refute previous literature on the positive effects of sleep on productivity. In fact, according to our preliminary model, it suggests that sleep quality and amount of sleep have almost no effect on self-reported productivity.

Table 2: Regression Results

Term	Estimate	Std. Error	t value	P-value
(Intercept)	5.586	0.335	16.688	0.000
total_sleep_hours	0.037	0.028	1.314	0.189
sleep_quality	-0.001	0.014	-0.101	0.919
exercise_catlow	0.179	0.100	1.795	0.073
exercise_catmedium	0.179	0.101	1.772	0.077

Table 2: Regression Results

Term	Estimate	Std. Error	t value	P-value
genderMale	-0.124	0.099	-1.258	0.208
genderOther	-0.115	0.100	-1.141	0.254
age	-0.002	0.003	-0.529	0.597
caffeine_intake_mg	0.000	0.000	0.328	0.743
screen_time_before_bed_mins	0.000	0.001	0.243	0.808
work_hours_hrs_day	-0.020	0.018	-1.137	0.255
stress_level	-0.001	0.014	-0.095	0.925
mood_score	-0.007	0.014	-0.481	0.631

3.2 Residual Analysis

Figure 3 highlights the residual plot of our model. Although the errors seem to have a constant variance and points are scattered around a conditional mean of 0 (linearity satisfied), there is a systemic pattern of parallel lines inhabiting our plot (violating normality of errors). This suggests that our response variable is a strictly continuous variable, making linear regression dubious. Furthermore, Figure 4 highlights that our model is explaining little variation around the mean. Our model data is clustered around the productivity values between 5.4 and 6, while the actual values span from 0 to 10. Due to extreme deviations from the 45 degree line, it challenges the linearity of our model. Moreover, Figure 5 shows the QQ-plot of our model standardized errors: it displays a step pattern between the 5.4 to 6 range due to the discrete nature of our response, violating normality of errors.

3.3 Further Work

Currently, our model explains variation in productivity scores weakly and our model diagnostics expose problems of model misspecification: linearity and normality assumption violations.

1. Transforming our response variable: Convert 0-10 productivity scores to binary outcomes (high/low) through validated thresholds, enabling logistic regression for our analysis.

2. Colinearity and Predictor Correlation: Predictors like screen_time_before_bed, caffeine_intake, and sleep_quality are theoretically linked, risking biased estimates. To resolve for this, we can employ a two-step regression instrumenting our main predictor of sleep hours on sleep quality, caffeine consumption, and screen time before bed.

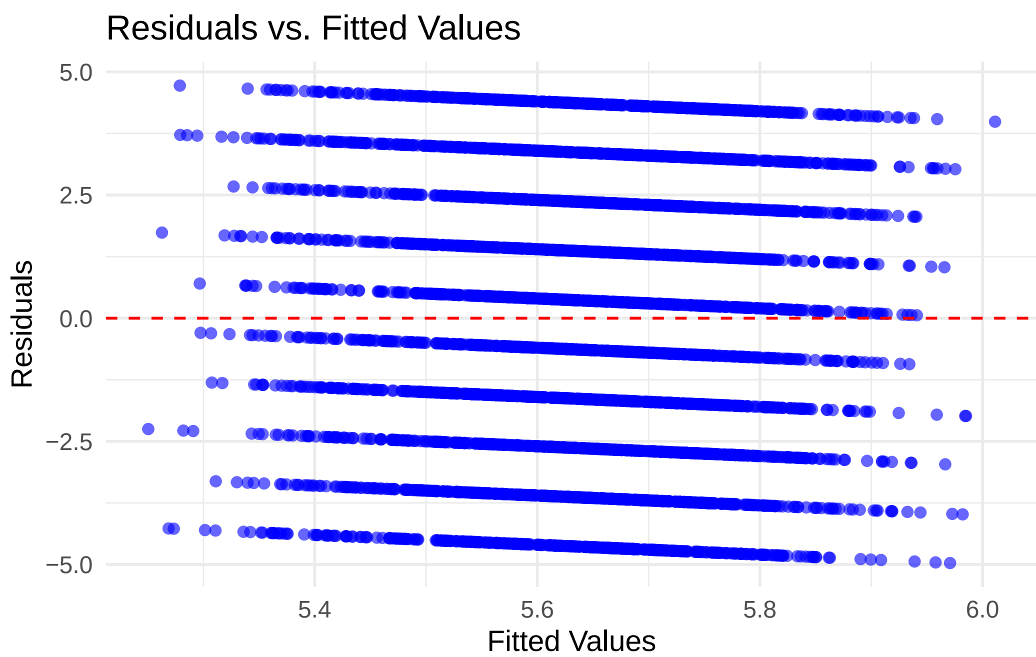

Figure 3

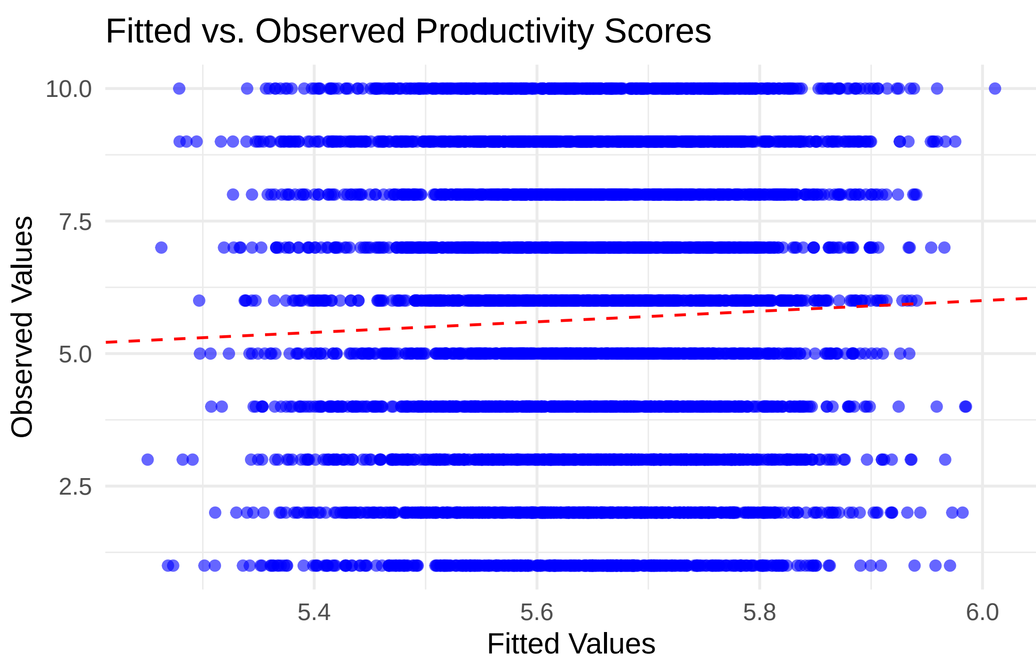

Figure 4: Figure of Fitted and Actual Values

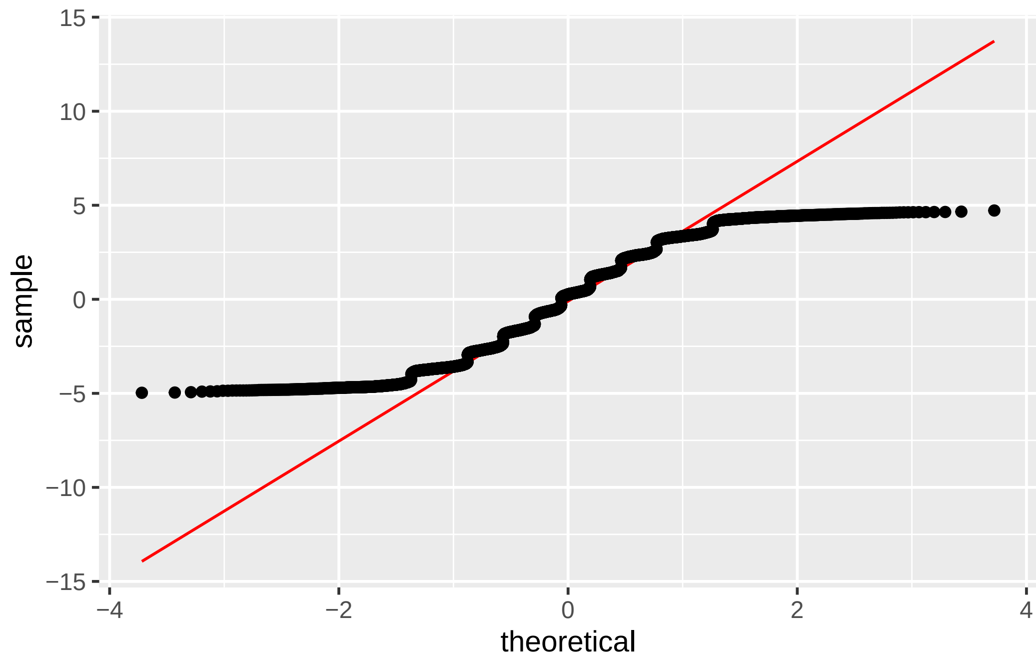

Figure 5: Q-Q Plot of Regression Residuals

4 References

1. Centers for Disease Control and Prevention. What counts as physical activity for adults. (2023, December 6). Physical Activity Basics. https://www.cdc.gov/physical-activity-basics/adding-adults/what-counts.html

2. Forrester, N. (2023). How better sleep can improve productivity. Nature (London), 619(7970), 659–661. https://doi.org/10.1038/d41586-023-02307-z

3. Gingerich, S. B., Seaverson, E. L. D., & Anderson, D. R. (2018). Association Between Sleep and Productivity Loss Among 598 676 Employees From Multiple Industries. American Journal of Health Promotion, 32(4), 1091–1094. https://doi.org/10.1177/0890117117722517

4. Ishibashi, Y., & Shimura, A. (2020). Association between work productivity and sleep health: A cross-sectional study in Japan. Sleep Health, 6(3), 270–276. https://doi.org/10.1016/j.sleh.2020.02.016