Bayesian Hierarchical Linear Regression Model for Inferencing Sleep Quality

Jimin Heo and Minal Walvekar

STATS 205

Abstract

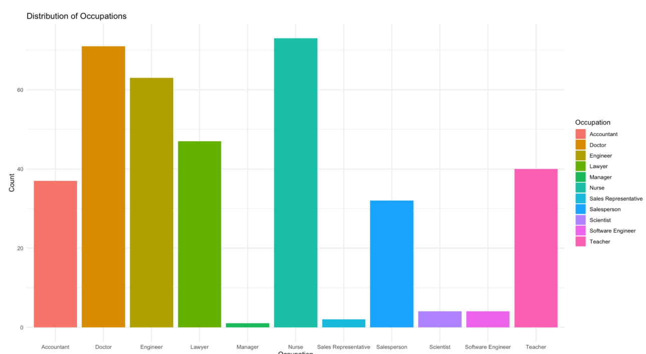
The goal of this project is to explore how our parameter of interest, stress level, impacts the sleep quality of individuals categorized by occupation. To achieve this, we implemented a hierarchical linear regression model.

Introduction

Sleep problems are on the rise, leading to substantial healthcare costs totaling $94.9 billion annually [1]. Previous research has established that stress disrupts sleep patterns by activating specific neural activity in the brain [2]. However, stress manifests in various forms, necessitating a deeper understanding of its effects on sleep. In this study, we aim to explore how stress impacts the likelihood of sleep quality across different occupations.

Our analysis utilizes data sourced from Kaggle, comprising 374 individuals of various demographics, along with the health observations of each person. Demographic columns included age, gender, and occupation, while the health explanatory variables included physical activity, stress level, blood pressure, sleep duration, quality of sleep, and sleep disorder.

Exploratory Data Analysis and Feature Selection



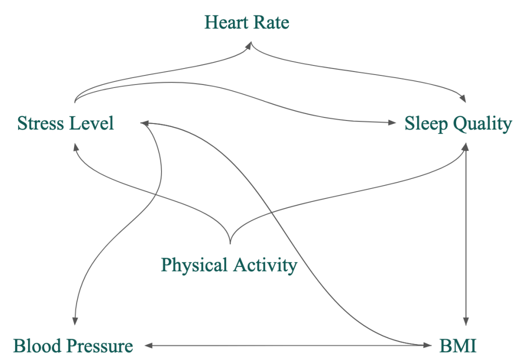


Figure 1: Count of Individuals per Occupation
Figure 2: Relationship Between Features

When exploring the raw data, there were initially 374 observations from 11 different occupations. The data was incredibly imbalanced as few of the occupations, such as Manager and Sales Representative, only had 1 and 2 observations respectively as seen in Figure 1 above. Because of the lack of data, we concluded that occupations would not contribute meaningfully to our model, so we excluded them from our model and analysis.

With "sleep quality" as the main focus of our model, we selected four total features to include in the analysis with stress levels being a key covariate due to its strong association with sleep quality. Previous studies have shown that high levels of stress are associated with an increased level of obesity, with variations in BMI linked to changes in acute stress reactivity [3]. Research has also shown that BMI levels and stress quality have an influence on each other [4]. Physical activity is also known to be associated with lower BMI and a higher quality of sleep [5]. Higher heart rates have been shown to be linked to a lower quality of sleep, and stress has been shown to affect both heart rate and blood pressure [6]. It has also been determined that blood pressure influences BMI and vice versa [7]. However, we chose to only incorporate stress level and physical activity as the health variables used in our model, as the stress level is our parameter of interest and physical activity is the only confounder between stress and sleep disorders.

Methods and Modeling

We chose to implement a Bayesian Linear Regression Hierarchical model to make inferences on the effect of stress level on sleep quality. A hierarchical model was suitable for our purpose because the features of our data needed to be assigned priors specific to their categories, so with a hierarchical model we were able to calculate variances and coefficients more accurately. By using a hierarchical approach, we can account for different sources of variability, in our case we applied different variations for demographic and health data.

Denote the N-th observed value as y_n for n=1 . . . N where N is the total number of observations. Let D represent the number of features and L represent each unique occupation. Let τ represent the array of priors for the feature variances. Let the τ's subscripted o, d, h, i represent the overall, demographic, health, and intercept τ's respectively.

$$\mathbf{y_n \sim N(x_n \cdot \beta_{ln}, \sigma_y)} \qquad \mathbf{\tau_o \sim Inv\text{-}\chi^2(1, 0.05)}$$
$$\mathbf{\sigma_y \sim Inv\text{-}\chi^2(1, 0.05)} \qquad \mathbf{\tau_d \sim Inv\text{-}\chi^2(1, \tau_o)}$$
$$\mathbf{\beta_{l,d} \sim N(0, \tau_d)} \qquad \mathbf{\tau_h \sim Inv\text{-}\chi^2(1, \tau_o)}$$
$$\mathbf{\tau_i \sim Inv\text{-}\chi^2(1, \tau_o)}$$

Seen above is the hierarchical model, with an "overall" posterior variance (set with a non-informative prior) that we will use as a prior for the middle level of the hierarchy. As we move on to the second level, we have two distributions, one that encapsulates the "demographic" variables, such as age and gender, and another that encapsulates the "health" variables, such as stress level and physical activity. These distributions use the "overall" posterior variance as their prior variances. On the level below that, we estimate the variance for all 4 feature variances as well as the intercept variance is estimated. As the intercept belongs to neither the demographic category nor the health category, we used the "overall" posterior variance as its prior variance. Since the age and gender variables are a part of the individual's demographic information, we used the "demographic" posterior variance as the prior, and since the stress level and physical activities are a part of the "health" information of each person, we used the posterior variance of the health category as the prior.

Evaluation & Result

When evaluating the quality of our model, the MCMC trace plots were generally all full and thick, which suggests proper convergence. To further assess our inference model, we simulated data from the posterior distributions to see how the means differ between the observed data versus the simulated data. As seen in Figure 4 where the vertical line represents the observed mean and the density plot represents the distribution of the means, we can see that the distribution of the simulated mean is centered around the observed mean. This suggests that our model is reliable for inferencing the influence of stress level on sleep quality.

Figure 3: Trace Plots of Betas

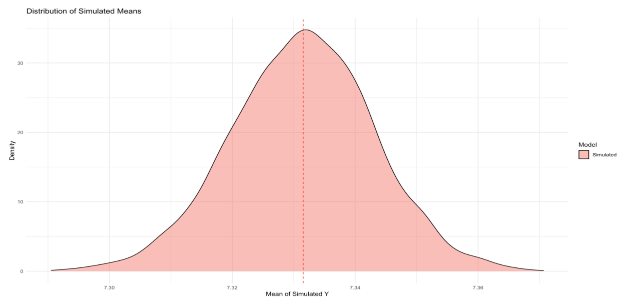


Figure 4: Density of simulated means vs. Observed Mean

Table 1: Beta Coefficients

	Intercept	**Gender**	**Age**	**Physical Activity**	**Stress Level**
Accountant	5.073649	0.067014806	-0.039055542	0.061508722	0.1718501
Doctor	13.238302	-0.556056762	-0.022448066	-0.009481743	-0.7117226
Engineer	9.166831	-0.861186277	0.015920460	0.018041154	-0.5161558
Lawyer	9.642321	0.513256293	0.076094063	-0.005925517	-0.9524627
Nurse	8.304651	-0.008727197	0.016765313	0.022763172	-0.6474387
Salesperson	4.426639	0.067576651	0.001256229	0.006026539	0.1685951
Scientist	7.568786	-0.035153290	0.012058892	0.038476946	-0.6502777
Software Engineer	6.949696	0.159088316	0.050626050	0.046844920	-0.7398651
Teacher	18.475556	-1.929031045	-0.170642529	-0.039545558	-0.5161605

Shown in Table 1, the beta coefficients table, the coefficients for stress levels on sleep quality vary across different occupations. Increased stress is associated with decreased sleep quality for doctors, engineers, lawyers, nurses, and teachers. In contrast, accountants and salespersons show a slight improvement in sleep quality with increased stress, although the effects are relatively minor. In the case of an accountant, a one-unit increase in stress level is associated with a slight increase of 0.178 in sleep quality. In the case of a doctor, a one-unit increase in stress level is associated with a slight increase of -0.7117226 in sleep quality. Other beta covariates were -0.5161558, -0.9524627, -0.6474387, 0.1685951, -0.6502777, -0.7398651, and -0.5161605 for engineer, lawyer, nurse, salesperson, scientist, software engineer, and teacher respectively.

Discussion

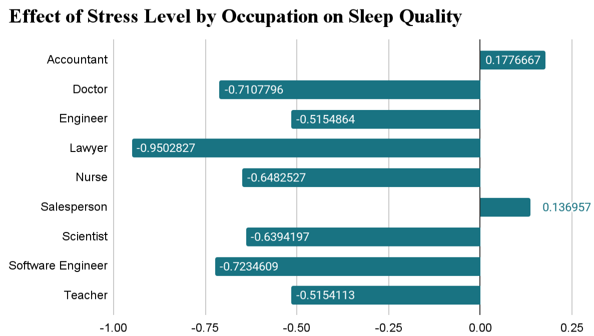


Figure 5: Effect of Stress Level by Occupation on Sleep Quality

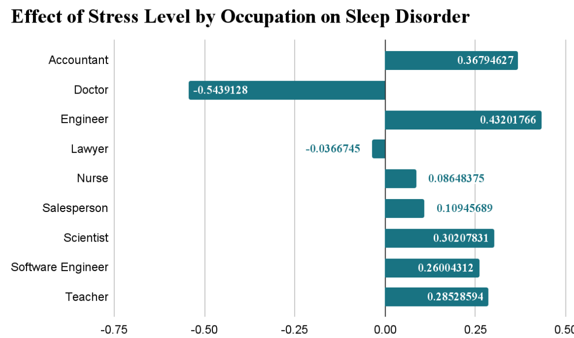


Figure 6: Effect of Stress Level by Occupation on Sleep Disorder

As mentioned earlier, stress disrupts sleep patterns, so we were curious why for accountants and salespeople stress doesn't affect sleep quality, and even has minor positive effects. This might be incurred due to an unknown confounder leading to Simpton's Paradox. However, we were intrigued to find that understanding the inherent characteristics of each occupation helped us understand the results. Accountants, for instance, experience high workloads mainly during the tax season or quarterly reports. Once these peak seasons pass, accountants can experience accomplishment and rest, potentially improving sleep quality. Similarly, for salespeople, the satisfaction of closing big deals can positively influence sleep quality in the short term. Therefore, our analysis suggests that the duration and nature of stress play pivotal roles in sleep quality.

However, we discovered that having low sleep quality doesn't necessarily relate to sleep disorder. (A one-unit increase in stress level increases approximately 0.3679 in the log odds of experiencing a sleep disorder.) The above graph shows how stress level affects sleep disorders categorized by occupation. Across different occupations, individuals experience varying types of stress [8]. Doctors and lawyers, for instance, may endure burnout due to high job demands and critical environments but have stable careers, resulting in a reduced likelihood of sleep disorders. Conversely, engineers and scientists may encounter stress from fear-based stress related to meeting deadlines, or job searching, which increases their risk of sleep disorders.

In conclusion, we observed that the effect of stress varies by occupation, and we can infer that sleep quality is primarily affected by the duration and nature of stress. However, occupations with seasonal workloads or jobs with high accomplishment may experience a positive effect on sleep quality in the face of stress. Additionally, when considering the likelihood of sleep disorders, individuals experiencing fear-based stress are at a higher risk. It is important to note that our results have limitations due to potential unknown confounders that might be affecting the observed outcomes. Further studies are needed to explore these confounders and validate our findings.

Reference

[1] Massachusetts Eye and Ear Infirmary. (2021, May 7). Sleep disorders tally $94.9 billion in health care costs each year. ScienceDaily. Retrieved June 2, 2024, from https://www.sciencedaily.com/releases/2021/05/210507160008.htm.

[2] University of Pennsylvania. (2023, December 1). Research shows how stress activates neurons that disrupt sleep. *Penn Medicine News Release*. Retrieved June 2, 2024, from https://www.pennmedicine.org/news/news-releases/2023/december/research-shows-how-stress-activates-neurons-that-disrupt-sleep.

[3] Kühnel, A., Hagenberg, J., Knauer-Arloth, J. et al. Stress-induced brain responses are associated with BMI in women. Commun Biol 6, 1031 (2023). https://doi.org/10.1038/s42003-023-05396-8

[4] Vgontzas AN, Bixler EO, Basta M. Obesity and sleep: a bidirectional association? Sleep. 2010 May;33(5):573-4. doi: 10.1093/sleep/33.5.573. PMID: 20469796; PMCID: PMC2864869.

[5] Gupta P, Srivastava N, Gupta V, Tiwari S, Banerjee M. Association of sleep duration and sleep quality with body mass index among young adults. J Family Med Prim Care. 2022 Jun;11(6):3251-3256. doi: 10.4103/jfmpc.jfmpc_21_21. Epub 2022 Jun 30. PMID: 36119287; PMCID: PMC9480680. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9480680/

[6] Sleep Foundation. "What Is a Normal Sleeping Heart Rate?" Sleep Foundation, www.sleepfoundation.org/physical-health/sleeping-heart-rate. Accessed 12 June 2024.

[7] Song, L., Li, J., Yu, S. *et al.* Body Mass Index is Associated with blood pressure and vital capacity in medical students. *Lipids Health Dis* **22**, 174 (2023). https://doi.org/10.1186/s12944-023-01920-1

[8] Amiri, M. (2023, September 7). Types of Occupational Stress in the Workplace. Retrieved June 2, 2024, from https://www.linkedin.com/pulse/types-occupational-stress-workplace-mohammad-amiri/