

Research on Sleep Quality and Its Correlated Factors

Xi Wang, Li Zhi Liao, Zhonghan Deng, Jenish V. Dhaduk

Abstract—Given the concerns about people’s health is highly impacted by their sleep quality, this research is conducted through a survey questionnaire to collect data from people that are easily accessible. The research consists of 14 factors (“eat_habits,” “coffee_consumption,” “stress_level,” “age,” etc.) that might have an impact on people’s sleep quality. The survey is open for a total of 9 days for people in the University of the Pacific’s Data Science Program, and people who are friends of the researchers. Total data collected for this research is 171 rows. These 14 factors are tested individually with sleep quality by using the Chi-Squared test in R. The 4 significant variables (“eat_habits,” “dreams,” “feelings,” “exercise”) are combined in all possible ways into the ordinal logistic models. Out of the total 11 models, model 10 (“exercise” + “dreams”) shows the lowest AIC value, which indicates that it has the highest accuracy on predicting people’s sleep quality. The research concludes that people’s sleep quality is significantly correlated to their frequency of exercise and dreams.

Key words: sleep quality, exercise, dreams.

I. INTRODUCTION

The topic about Sub-health is slowly becoming more and more popular in this century, as many people are having more stress, which leads to difficulties in living a balanced life. Among many reasons that prevents people from living a healthy lifestyle, sleep quality of a person is definitely one of the biggest issues. A national data survey reports that around 50 million American adults are suffering from about 80 different sleeping disorders and over 20 million people in the United States are suffering from occasional sleeping problems in a yearly basis. The report also shows an increasing trend of sleeping problems within the total population over the past 30 years. According to the national report, sleeping problem is also highly correlated to some unnecessary public health and safety issues [3]. A journal published in 2015 indicated that good sleep quality is extremely critical to good health and overall quality of life [4]. Looking at the academic performance will provide more insight on this topic. A study posted in Oct. 2019 has reported a highly correlated relationship between academic performance and sleep quality, indicating that good sleep quality is associated with better performance on academic scores [5]. Since sleep can significantly affect one’s personal life and health, this has brought up the researchers’ interest in finding out what factors

could affect people’s sleep quality. By doing an appropriate data analysis, the researchers are hoping to provide constructive suggestions on how to improve people’s sleep quality.

Many researches related to sleep quality have been conducted within the last few decades. One research related to this scope involved Pittsburgh Sleep Quality Index (PSQI), which used an interval scale as the measurement for the dependent variable sleep quality. The research sent out a survey containing 24 questions to random participants for data collection. Since both the independent variables and sleep quality were measured on an interval scale, an ANOVA regression was conducted in this research to find out which variables were highly correlated to sleep quality. In addition, the research also studied two covariate variables, age and sex ratio, by using Ancova and found out that age was negatively correlated to participants’ sleep quality [2].

Another research was conducted in 2017 and the purpose of the study was to figure out what factors could possibly highly correlate to people’s sleep quality. The research collected sample data by using Pittsburg Sleep Quality Index (PSQI) and conducting a 11-questions survey focusing on these fields: age, gender, semester, area of residence, the existence of chronic diseases, caffeine consumption level, smoking habits. The sample population with 223 volunteered nursing students derived from Uludag University Faculty of Health Sciences Department of Nursing. The study displayed a regression analysis on the data collected and concluded that students’ sleep quality is greatly affected by smoking habits, total sleep hours, resting status in the morning and average daily coffee consumption [1].

The objective of this research is to somehow compare to the research mentioned above with different perspectives on looking at the independent variables and conducting regression analysis. To be more specific, this research purpose is to answer the question to what factors are used and how these factors can be used to predict a person’s sleep quality. This research is going to perform an ordinal logistic regression analysis to find out the correlations between different factors and sleep quality, and then build a regression model to predict sleep quality. In this research report, different factors will be analyzed to figure out whether they are positively, negatively or not correlated to sleep quality. The null hypothesis of this study will be that there is no significant relationship between the independent variables selected in this research and sleep quality. In contrast, the alternative hypothesis will be that there is a significant relationship between the independent variables selected in this research and sleep quality.

II. DATASET

Our independent variables are 1) number of awakenings per night, 2) number of cups of coffee per day, 3) number of exercises per week, 4) stress level, 5) age, 6) number of dependents in the household, 7) acceptable noise level, 8) eating habits, 9) type of medication(s) taken before bed, 10) frequency of sexual activity per week, 11) hours of sleep per day, 12) average number of dreams in one night, 13) the amount of time it takes to fall asleep and 14) emotional state after awakening. The dependent variable is sleep quality. Variables #8, 9, 14 are nominal and the rest of the variables are ordinal.

The survey was created using Google Forms. The survey consists of 15 questions, each question corresponding to a variable. Most questions are multiple choice, with 3 to 6 options, with the exception of the question regarding stress level, which is based on a 1 to 10 scale.

The sampling technique used is convenience sampling for practical considerations. Participants are not chosen randomly; they are chosen based on accessibility. The researchers initially planned to send the survey to the entire class, however, fewer than 20 responses are received. Thus, the researchers sent the survey to their friends and asked them to share it with their friends. The survey was open for 9 days and received a total of 171 rows of data. However, 48 rows are completely empty responses. There are also 3 rows that have missing value(s) in one or more columns. These 51 rows are dropped in R before performing any analysis.

When the survey was conducted, the questions were worded in detail to avoid ambiguity. Yet, it is not recommended to have long column names. Thus, the 15 column names are renamed for simplicity reason. After renaming, the 14 independent variable column names are “wake_up,” “coffee,” “exercise,” “stress,” “age,” “dependents,” “noise,” “eat_habits,” “medicine,” “sexual_activity,” “hours,” “dream,” “time_to_fall_asleep” and “feeling.” The dependent variable column name is renamed as “quality.”

III. METHODOLOGY

A. Dependent Variable

For this research, sleep quality is the dependent variable. It is an ordinal variable with 5 levels: excellent, good, fair, okay and terrible. Excellent means that the participant is having no complaints at all. Good means that the participant’s sleep quality is well. Fair means that the participant’s sleep quality needs to be improved. Okay means that the sleep quality is just fine and poor means that the sleep quality is terrible.

B. Independent Variables

The 14 independent variables in the dataset determine whether the relationship between the dependent variable and each independent variable is significant. Chi-Squared Test is

used to see if there are any correlations. There are a total of 14 Chi-Squared tests performed. The results show that “exercise” and “quality,” “eat_habits” and “quality,” “dream” and “quality,” and “feeling” and “quality” have significant relationships. The p-values are summarized in the table below in the Results and Discussion section.

C. Statistics

Since the dependent variable, sleep quality, contains 5 ordinal levels, Ordinal Logistic Regression will be used to build models for this dataset. First, the entire dataset is split into training and testing datasets with a percentage ratio of 90%, 10%. Based on the results from the Chi-Squared Test, the independent variables with significant p-values in the training dataset are combined in all possible ways and placed into 11 different models. Then, among these 11 different models, the best model with the lowest AIC (an indicator to measure how accurate a model is) will be selected.

After the best prediction model is built, the assumption of proportional odds, also known as parallel line, will be tested to see if the effects of any explanatory variables are consistent or proportional across the different thresholds. Failing to reject the null hypothesis ($p\text{-value} > 0.05$), the slope coefficients in the model are the same across response categories, which indicates that there is no violation of this assumption [6]. If there is a violation of the assumption, another regression approach such as categorical logistic regression, will be used instead.

Finally, the prediction model is used on the testing dataset to predict the dependent variable and see how good it fits the dataset. 10% sample from the clean dataset will be compared with the prediction results from running the model. This process will test how approximately accurate the best model is.

After the best prediction model is built, the assumption of proportional odds, also known as parallel line, will be tested to see if the effects of any explanatory variables are consistent or proportional across the different thresholds. Failing to reject the null hypothesis ($p\text{-value} > 0.05$), which states that slope coefficients in the model are the same across response categories, indicates that there is no violation of this assumption [6]. If there is a violation of the assumption, another regression approach such as categorical logistic regression, will be used instead.

Finally, the prediction model is used on the testing dataset to predict the dependent variable and see how good it is to fit the dataset. The actual responses in the test dataset will be compared with the prediction results from running the model. This process will test how approximately accurate the best model is.

IV. RESULTS AND DISCUSSION

The table below summarizes the p-value for the Chi-Squared Test between “quality” and each independent variable:

TABLE I
CHI-SQUARED TEST RESULTS

Quality and each IV	P-value
wake_up	0.5057
coffee	0.6372
exercise	0.002999
age	0.6112
dependents	0.1644
noise	0.6052
eat_habits	0.03848
medicine	0.06447
sexual_activity	0.2519
hours	0.1889
dream	0.02149
time_to_fall_asleep	0.08096
feeling	0.0004998
stress	0.1969

The table below summarizes the AIC values for each model:

TABLE II
AIC RESULTS

Model Name	Independent Variables	AIC
Model1	Exercise + Dream + Feeling + Eat_Habits	358.408
Model2	Exercise + Feeling + Eat_Habits	359.7224
Model3	Dream + Feeling + Eat_Habits	358.5382
Model4	Exercise + Dream + Eat_Habits	354.6474
Model5	Eat_Habits + Feeling	363.3166
Model6	Exercise + Eat_Habits	356.2518
Model7	Exercise + Dream + Feeling	354.8616
Model8	Exercise + Feeling	356.0763
Model9	Dream + Feeling	355.7818
Model10	Exercise + Dream	351.1082
Model11	Eat_Habits + Dream	354.5575

Among all 11 models built, model 10 is the best model for prediction. This model is used to predict people's sleep quality in the test dataset with 12 total samples. The result shows that there are 8 predictions that match the actual responses, which indicates that model 10 has about 66% accuracy rate in predicting a person's sleep quality through his or her exercise and dream situations.

Since "exercise" and "dream" can significantly affect sleep quality. Below are two tables containing results about sleep quality versus dream and sleep quality versus exercise. These tables interpret the correlations more intuitively between these two independent variables and sleep quality. The numbers shown in the tables are directly derived from the clean dataset.

People who are satisfied with their sleep quality are those who said they had either excellent or good sleep quality, which are in blue font. People who are not satisfied with their sleep quality are those who said they had either okay, fair or poor sleep quality, which are in red font. The tables below indicate that among people who are satisfied with their sleep, 92% of them are found to be doing some sort of exercise during the week, and 66% of them are having 3 to 4 dreams per night on average. In addition, among the people who are not satisfied with their sleep, there are 88% of them not doing any exercise and 65% of them are having 1 to 2 dreams per night on average.

TABLE III
QUALITY VERSUS EXERCISE RESULTS

	How People Felt About Their Sleep Quality				
Exercise Times Per Week	<i>Excellent</i>	<i>Good</i>	<i>Okay</i>	<i>Fair</i>	<i>Poor</i>
<i>None</i>	2	1	9	12	1
<i>1 – 2</i>	1	18	16	8	8
<i>3 – 4</i>	2	9	12	5	0
<i>More than 4</i>	3	5	2	5	1
Total	8	33	39	30	10

TABLE IV
QUALITY VERSUS DREAM RESULTS

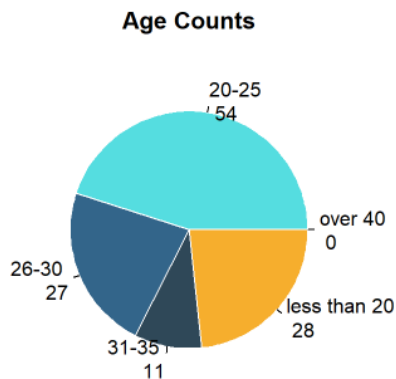
	How People Felt About Their Sleep Quality				
Number of Dreams Per Night	<i>Excellent</i>	<i>Good</i>	<i>Okay</i>	<i>Fair</i>	<i>Poor</i>
<i>None</i>	1	5	10	7	3
<i>1 – 2</i>	3	15	17	16	3
<i>3 – 4</i>	2	2	1	1	0
<i>5 – 6</i>	1	0	0	0	0
<i>More than 6</i>	0	0	0	0	1
<i>Do not Remember</i>	1	11	11	6	3
Total	8	33	39	30	10

V. CONCLUSION

In conclusion from the results of this research, people who are having 3 to 4 dreams per night on average and having some sort of daily exercise are satisfied with their sleep quality. Different from the other research mentioned in the introduction, the dependent variable of this research is measured in an ordinal scale. This research not only found the correlations between the independent variables and the dependent variable, but also built a statistically verified model that predicts the sleep quality of the participants, using a different approach from the research mentioned in the introduction.

There are some limitations in this research. For example, there are confounding extraneous variables that can affect the accuracy of our model, such as the type of exercise and the amount of protein intake for those people who have regular exercise. It also has ecological validity as this survey was done among people who live in urban areas, and therefore the conclusions cannot be generalized in people who live in rural

areas. The age of the participants is 40 and under (shown in the pie chart below), so the conclusion from the research cannot be generalized across all age groups. Furthermore, there were questions in the survey that were prone to seasonal changes, so responses might have been different if we asked in a different season, such as in a hot summer. Thus, the results from this research also cannot be generalized across other time periods.



This work done in this research might be able to help unravel the hidden relationship of some habitual routines with the sleeping quality of the person. “Exercise” and “dream” are the factors that have significantly affected the sleep quality of this research group. However, there are many limitations in drawing the conclusion regarding the generalized parameters that affect the sleep quality of the people in this research, such as the ones mentioned above.

In further research, more parameters should be considered. One possible change would be to either enlarge the sample size or include more participants from different areas. The sleep quality might vary from person to person, age to age, or location to location. Since it is hard to keep track of the generalized parameter, that can affect the sleep quality of an individual. In order to help people identify the factors that can influence their sleep quality, more local survey should be conducted because the side effects of bad sleep quality can be life threatening.

REFERENCES

- [1] D. Yilmaz, F. Tanrikulu, and Y. Dikmen, “Research on Sleep Quality and the Factors Affecting the Sleep Quality of the Nursing Students,” *Current health sciences journal*, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6286721/>. [Accessed: 07-Dec-2019].
- [2] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, “The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research,” *Psychiatry Research*, vol. 28, no. 2, pp. 193–213, 1989. [Accessed: 07-Dec-2019].
- [3] “The State of SleepHealth in America,” *SleepHealth*. [Online]. Available: <https://www.sleephealth.org/sleep-health/the-state-of-sleephealth-in-america/>. [Accessed: 07-Dec-2019].
- [4] *American Journal of Respiratory and Critical Care Medicine*. [Online]. Available: <https://www.atsjournals.org/doi/full/10.1164/rccm.201504-0767ST>. [Accessed: 07-Dec-2019].
- [5] K. Okano, J. R. Kaczmarzyk, N. Dave, J. D. E. Gabrieli, and J. C. Grossman, “Sleep quality, duration, and consistency are associated with better academic performance in college students,” *Nature News*, 01-Oct-2019. [Online]. Available: <https://www.nature.com/articles/s41539-019-0055-z>. [Accessed: 07-Dec-2019].
- [6] “5.3 Key Assumptions of Ordinal Regression,” 5.3 Key Assumptions of Ordinal Regression. [Online]. Available: <http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod5/3/index.html>. [Accessed: 07-Dec-2019].
- [7] “HOME,” IDRE Stats. [Online]. Available: <https://stats.idre.ucla.edu/spss/output/ordered-logistic-regression/>. [Accessed: 07-Dec-2019].