Effect of smoking, sleep, and exercise habits on depression prevalence in Southwest America

By

Duaa Zaid Alsaaideh

Supervisor

Sammy Allaw

Paper Airplanes

Women in Tech

Business Analytics Level 2

January 2021

Table of Contents

Introduction

Context

Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest, and it affects how we feel, think, and behave and can lead to a variety of emotional and physical problems.[1] Because of this, there is great interest in identifying the factors that are associated with higher prevalence of depression in certain populations.

Smoking is considered a major health issue, but it is often overlooked by health professionals for its effects on mental health. Although most are familiar with the physical health effects of smoking, few consider the effects smoking may have on our mental and emotional health. In this paper, however, we discover a moderate association between smoking and depression.

Another essential health factor that has been the focus of a lot of research in recent years is sleep, particularly the effects of disrupted sleep patterns on both physical and mental health. It is established that the recommended sleep duration for younger adults (18-24) ranges from 7 to 9 hours, with this duration having a host of positive health benefits. Sleep below this recommended duration of 7-9 hours a day, after a long period of time, results in something known as sleep disruption. Sleep disruptions exist when sleep is insufficient to support adequate alertness, performance, and health, either because of reduced total sleep time or fragmentation of sleep by brief arousal.[2] Further, hypersomnia (sleeping too much) is considered a disruptive sleep disorder. In this paper we find that, in addition to smoking, sleep habits are another influence on prevalence of depression diagnosis.

Moreover, we find that exercise habits may influence depression onset in individuals. Habits are behaviors an individual repeats over and over again, but also include behaviors that are started over and over again, with intermittent pauses.[3]

Literature Review

This research aims to understand the effect of smoking, sleep, and exercise habits on depression prevalence. Although a lot of prior research has investigated these three factors, no studies analyzed them together and in the American Southwest (defined here as the states of Arizona, Texas, California, Nevada, New Mexico, Colorado, and Utah). For example, one study investigated whether sleep problems associated with depression affect overall health compared with sleep problems independent of depression.[5] Also, the study examined the relationship in adults aged ≥18 years old, in a very broad geographic list of 46 low-and middle-income countries. By doing so, it sheds the light on the importance of educating the clinicians that the co-occurrence of sleep problems and depression is associated with a variety of adverse health outcomes in low- and middle-income countries.[6] On the other hand, another study was conducted on 2,631 university freshmen in southern Japan, to identify bedtime, wake time, sleep duration, and sleep quality effects on depressive symptoms and suicidal ideation. The study demonstrated that late bedtime and poor sleep quality have a highly significant association with depressive symptoms. Furthermore, it showed sleep quality may prevent the development of depressive symptoms and reduce the possibility of suicidal ideation.[7] Moreover, a review was applied to a general college student population and young adults. The main purpose was to evaluate the current awareness around the association between sleep with depression. This study

concluded that applying the procedures regarding sleep hygiene, in some cases, may improve the quality of sleep and prevent the improvement of depression in other cases.[8] Regarding Smoking, a systematic review evaluated the association of different aspects of smoking behavior with depression and anxiety. The result differed for each category with confirmation for positive associations in both directions (smoking to later mental health and mental health to later smoking) as well as null findings. Overall, almost half the studies informed that baseline depression/anxiety was associated with some type of later smoking behavior, while over a third found evidence that smoking exposure was associated with depression and anxiety.[9] And to investigate the effect of exercise on depression levels a study was applied to 136 inactive and 469 active persons in the city center of Samsun, Turkey. They were aged 18-45. The active group did cycling, walking, running, jogging, swimming, weight lifting, and other things. As a result, the non-exercising group showed higher depressive symptoms than the exercising group.[10] Likewise, my research is investigating the effect of smoking, sleep, and exercise habits on depression, however, it focuses specifically on analyzing data that contains information about young adults' behavior and depression, in American Southwest.

Question & Hypotheses

Given the existing literature on different populations, it was hypothesized that smoking, less exercising, and sleep disruptions are all significant predictors for depression onset for populations in our target age group of 18-34 years in the Southwest United States. Therefore, we define our independent variables for each individual as a categorical variable

on whether they smoke, the individual's sleep duration, and a categorical "exercise habits" variable defined below. The dependent variable is depression diagnosis

Methodology

For the purpose of this research, existing data from Kaggle was used. The dataset is from the Behavioral Risk Factor Surveillance System (BRFSS), a collaborative project between all states in the United States, including participating US territories, and the Centers for Disease Control and Prevention (CDC). There are hundreds of measured columns, ranging from exercise habits, to technology access, to education levels. For this project, a few specific columns were selected and filtered by state (our definition of the Southwest US), the sleeping time, smoking, exercising, and depression diagnosis to satisfy the research objectives.

Quantitative research was held to show the cause and effect relationship between sleeping, smoking, exercising, and depression. The main characteristic of quantitative research is that it is most appropriate for big samples, and its outcomes are measurable and quantifiable. Before analysis, the dataset was checked for missing data and outliers. Then the data were analyzed using the R software. A logistic regression was built to test the hypothesis.

Quantitative Analysis

From the 688 observations, 64 were omitted due to lack of data on the majority of the variables used in this study, Ages between 18-34 were selected, and specific American

Southwest states were chosen, Arizona, Texas, California, Nevada, New Mexico, Colorado, and Utah. Thus, the final sample consisted of 624 observations. In addition, the headers label was inappropriate, therefore headers were renamed. Depression, smoking, and exercising columns were of type categorical with answers yes and no, they were converted to 0 and 1 in order to analyze them with R regression models.

The logistic regression variables were evaluated for predictive power by p-value testing to evaluate the null and alternative hypotheses for each coefficient. In addition, multiple linear regression using two variables was conducted to assess the association between sleep duration and depression. Another model was built which is the logistic linear regression on all of the independent variables, sleeping, smoking, and exercising with the dependent variable as depression.

Finally, a heat map was made to help visualize the correlation between every two variables, and a frequency box plot was visualized to provide an overview of the distribution of individuals within both the depression-diagnosed and non-depression-diagnosed independent variable groups.

Discussion

Figure 1 below shows a heatmap which indicates that there is a high correlation between depression and smoking whereas there is little/no correlation between depression and exercising.

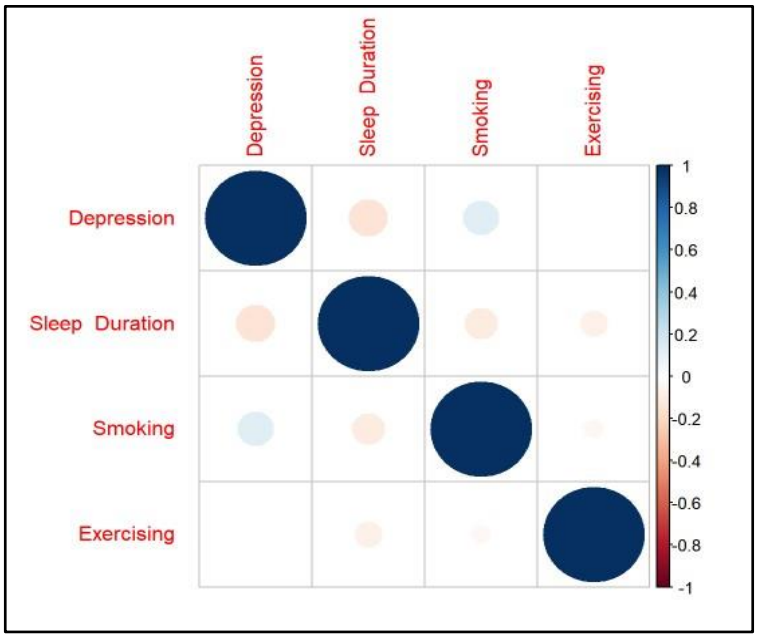

Figure 1: Heatmap of Variable Correlations

A more robust statistical test on each of the dependent variables was conducted by analyzing the p-values, which shows sleeping and smoking are strong indicators ($p<.01$) as shown in table 1 below. On the other hand, exercising within our dataset is not a good indicator ($p>.05$).

Table 1: P-value

	Est.	S.E.	z val.	p
(Intercept)	0.42	0.70	0.60	0.55
Sleep_DurationX	-0.30	0.09	-3.36	0.00
SmokingX	0.72	0.27	2.65	0.01
ExercisingX	-0.07	0.30	-0.23	0.82

As seen in Table 1, the coefficients for Sleep Duration, Smoking, and Exercising were -0.30, 0.72, and -0.07 respectively.

With this, our linear logistic regression model ultimately rejects the null-hypotheses for the coefficients of the sleep duration (negative correlation) and smoking dependent variable (positive correlation), while accepting the null-hypothesis (no predictive power or coefficient = zero) for the exercising habits dependent variable.

The following linear regression model is only for visualization purposes, as our response variable is categorical and therefore needs to be assessed with a logistic regression. For visualization purposes, our depression response variable was converted into a predictor variable to predict sleep duration. Figure 2 shows this multiple linear regression, and provides insight into the reversibility of predictors and responses. Further statistical analysis shows our two dependent variables to be statistically significant in predicting our response variable sleep duration ($p<0.01$).

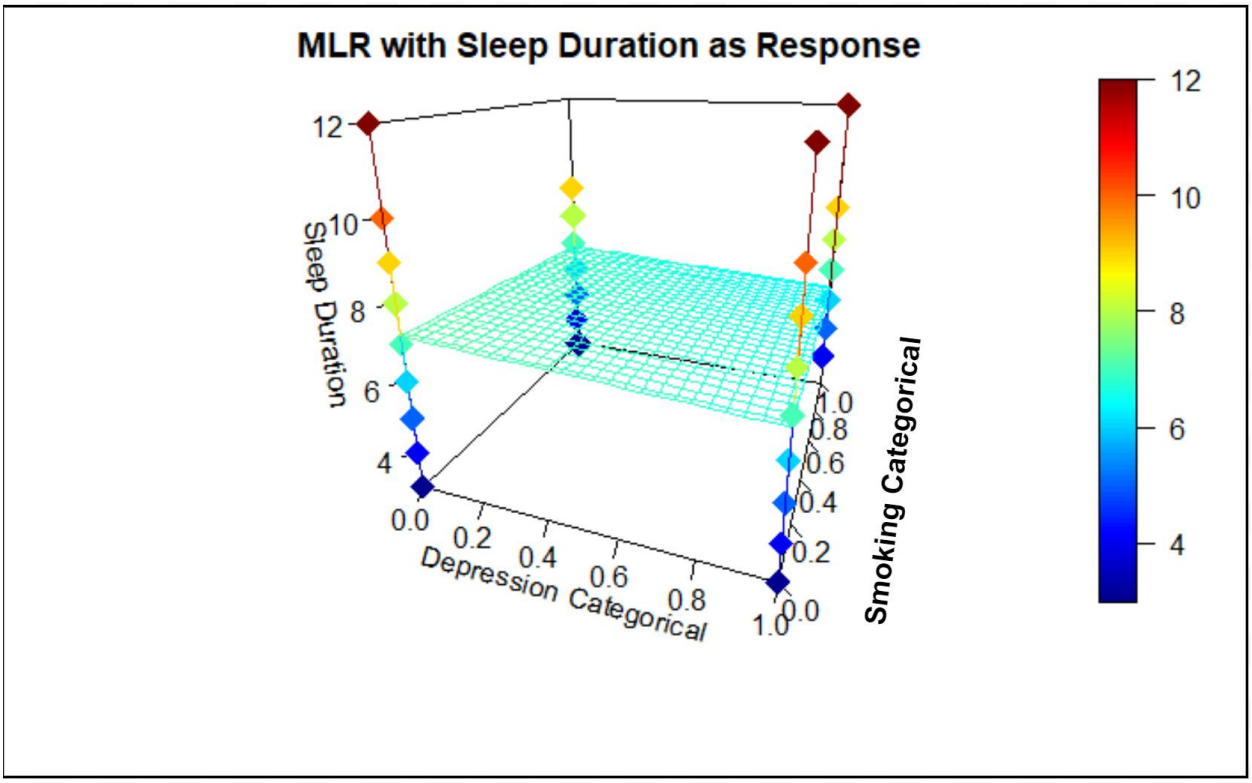

Figure 2: Multiple Linear Regression

Going back to the logistic regression model, Figure 3 visualizes the model predicting Depression with the sleep, smoking, and exercise habits variables. The "Effects" x-axis is simply a staggering of the individuals in the dataset, and this plot shows the relative distribution of probabilities for depressed and not depressed individuals. We can see the model predicts a vast majority of the individuals to have a 10-20% chance of having depression diagnosis.

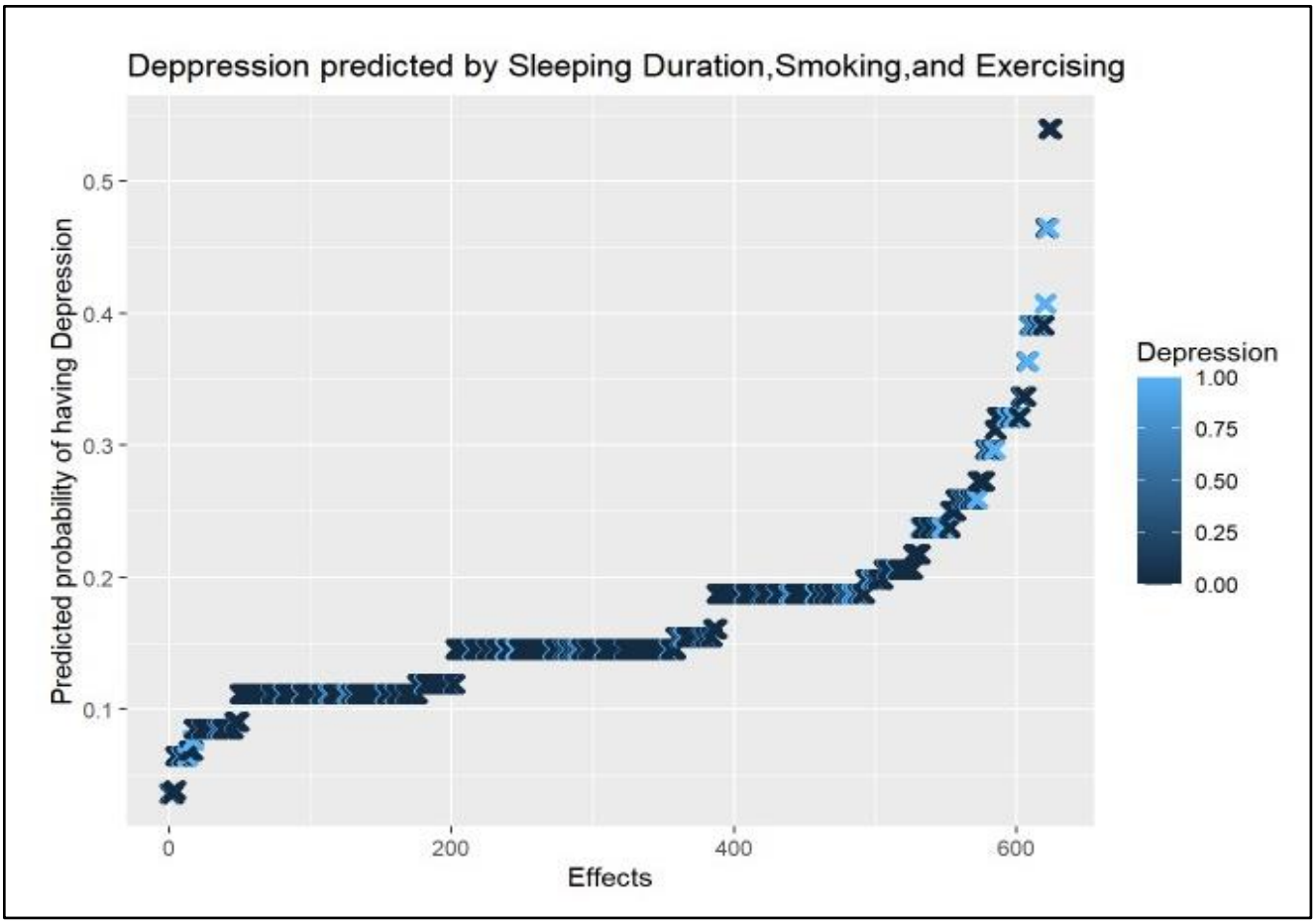

Figure 3: Logistic Linear Regression

The logistic regression was fitted with an ordinary linear regression, noticing that the response variable (depression) is a numeric variable that only takes values 0 and 1. Next, we performed the logistic regression with each individual dependent variable predicting depression for visualization purposes. As shown in figure 4, the probability to suffer from depression *decreases* as the sleep duration increases and exceeds the normal, recommended range of sleeping hours (7-9 hours a day).

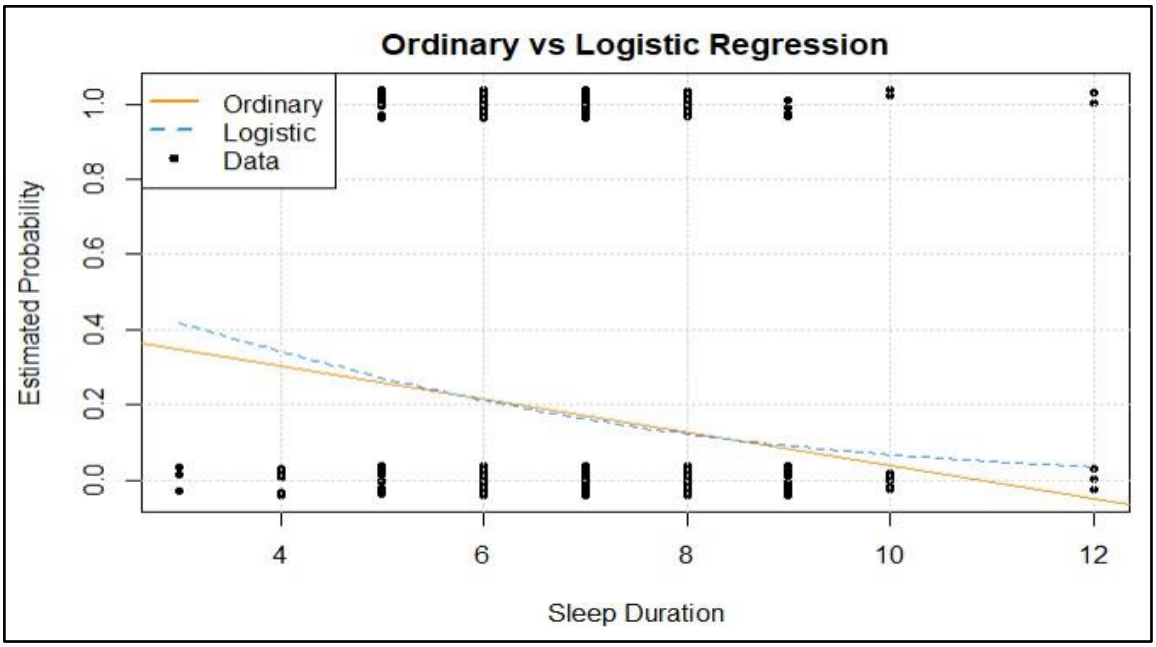

Figure 4: Ordinary vs logistic Regression

Figure 5 below shows another model built predicting to show that smoking behaviors are positively correlated with depression diagnosis.

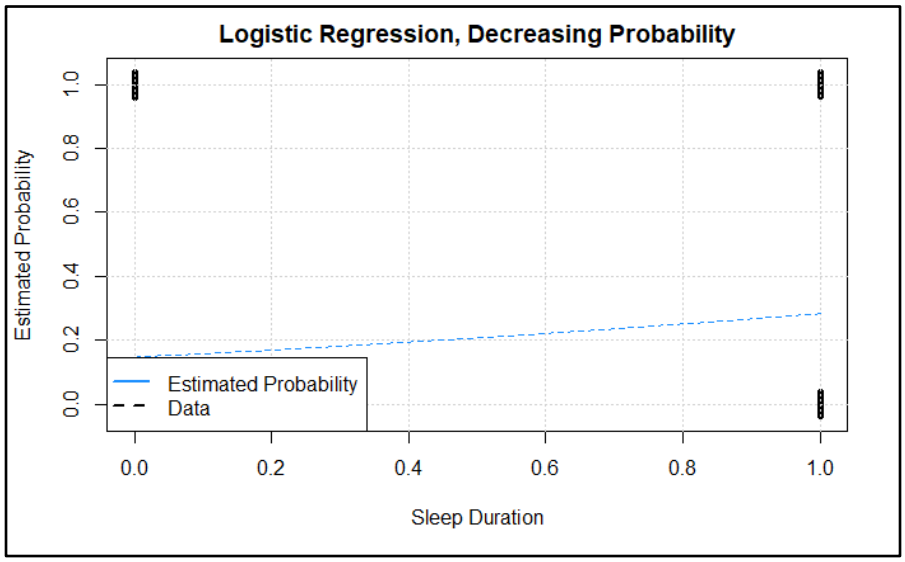

Figure 5: Logistic Regression with decreasing Probability

A quadratic relationship was applied in Figure 6 and it seems like this model is a better choice for this dataset, as it captures both ends of the sleep duration distribution. Those with very low sleep durations are almost exclusively in the depression category, and there is a non-zero chance of

depression for very large sleep durations. In fact, higher sleep durations than the mean have a higher prediction percentage for depression probability.

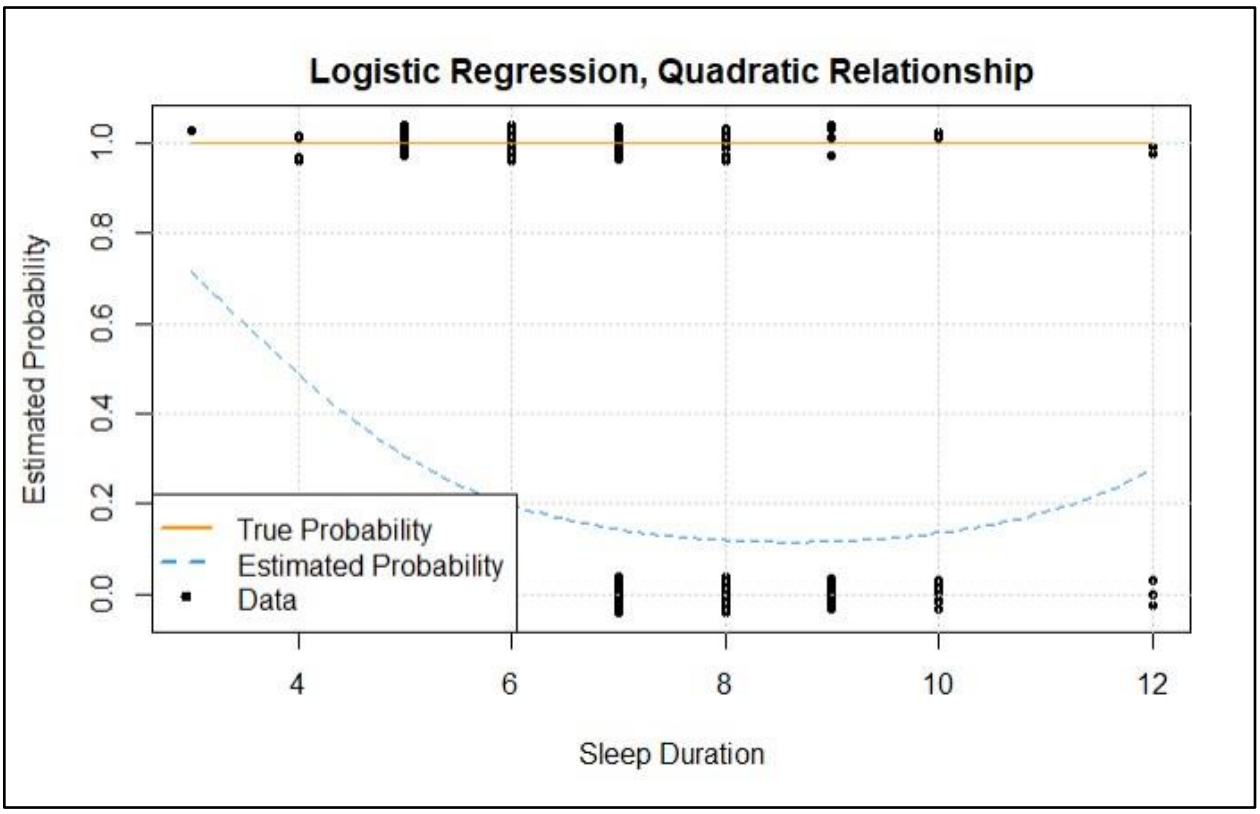

Figure 6: Logistic Regression with Quadratic Relationship

Box plots show the distribution of numerical data and skewness through displaying the data quartiles (or percentiles) and averages.[11] As shown in figure 7 below, a box plot was visualized to display the sleep patterns over the American Southwest states. We can see that the sleep duration in most of the states ranges between 6-8 hours (25 to 75 percentile) and with a median around 7.5 hours. This means that most individuals within the study dataset sleep in the recommended, normal range of hours.

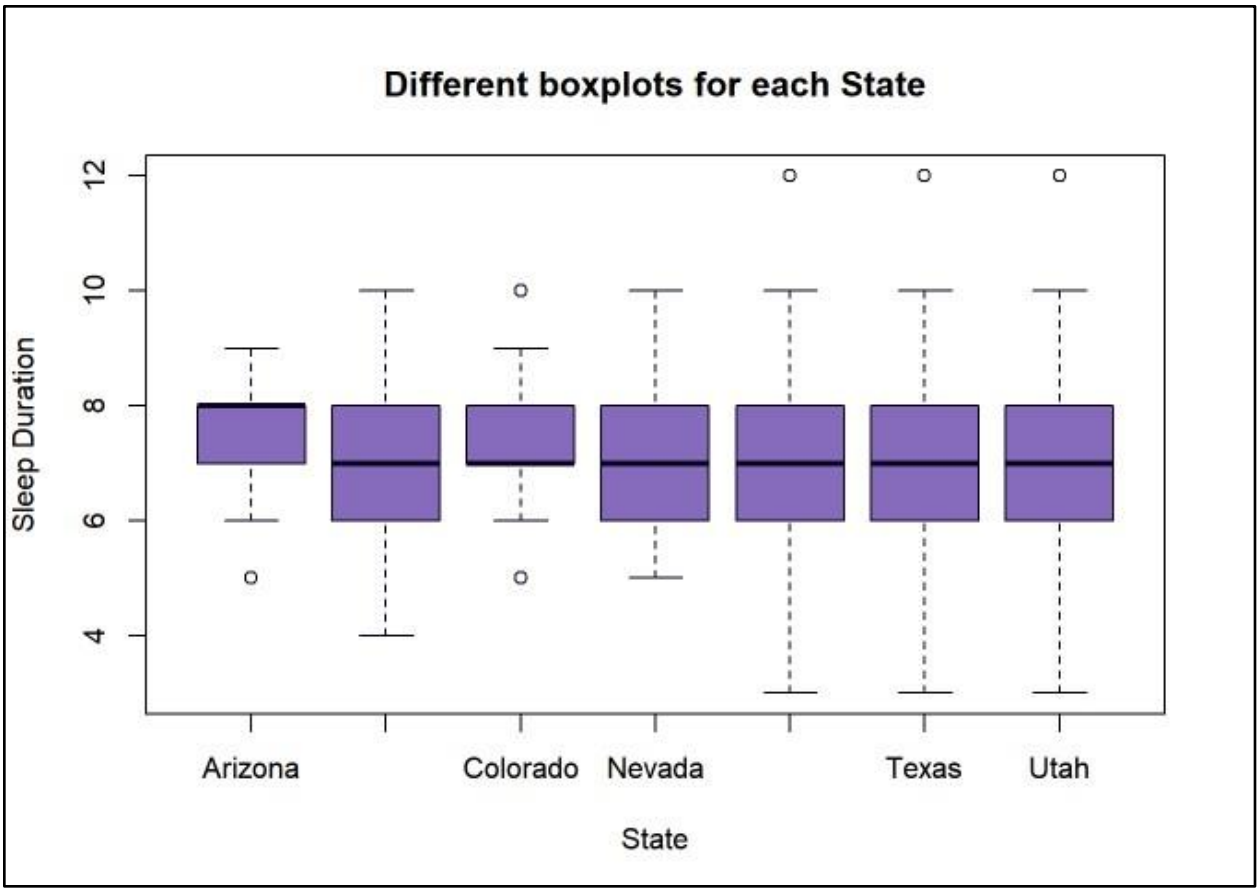

Figure 7: Boxplots

Limitations & Conclusions

There are three major limitations in this study that could be addressed in future research. First, the study focused on depression, and most of the individuals in the dataset were not suffering from depression, as shown in figure 8 (0 represents no depression, 1 represents having depression diagnosis). This further compounds the issue with our smaller dataset size, as the "depression" group is very small. Second, the data was mostly collected from telephone interviews, meaning that a lot of participants may have given false answers due to the stigma surrounding depression diagnosis, and potential pressure to lie about smoking and exercise habits. Finally, the independent variable of the study was depression, and as a categorical

variable this made a multiple linear regression model impossible, forcing us to rely on the logistic regression model.

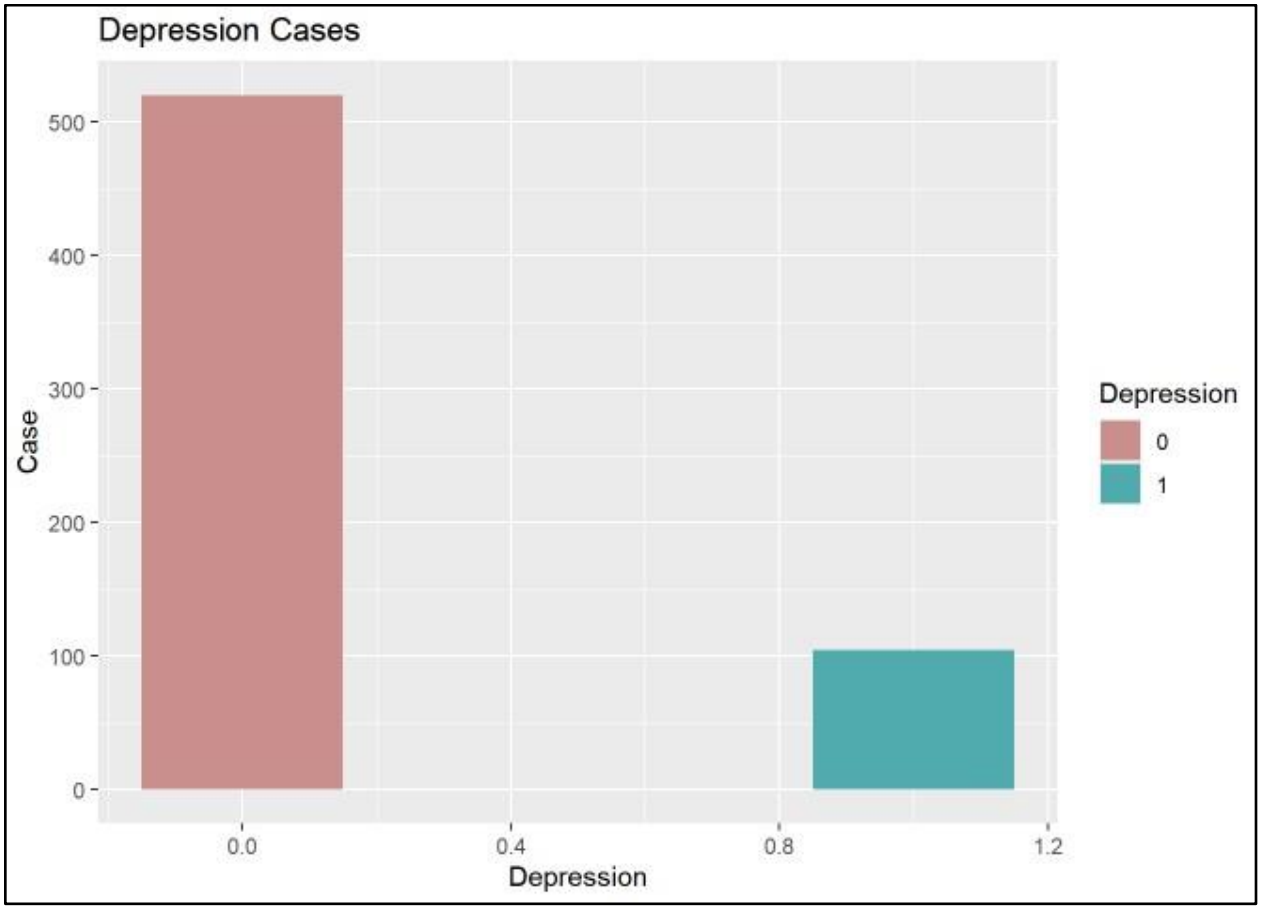

Figure 8: Depression Cases

In conclusion, our findings suggest that handling sleep quality and quitting smoking may help in handling depression. In particular, we discover a quadratic relationship between the sleep duration and depression, with both hypo and hypersomnia being predictors. However, it must be noted that these are only predictors in our models. Our model does not assign any causation. To assume a causal relationship will require more controlled clinical trials outside the scope of this paper. Moreover, this study found that although doing exercises has a great impact on

physical health, our statistical analysis showed that there is no relationship between exercise habits and depression prevalence.

References

1. Depression (major depressive disorder). (2018, February 03). Retrieved January 06, 2021, from https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007

2. Richards, E., & Pao, M. (n.d.). Sleep Disruption: Johns Hopkins Psychiatry Guide. Retrieved January 06, 2021, from https://www.hopkinsguides.com/hopkins/view/Johns_Hopkins_Psychiatry_Guide/787026/all/Sleep_Disruption#:~:text=DEFINITION,May%20be%20volitional%20or%20involuntary

3. CLEAR, J. (2020, February 04). How to Get Motivated to Workout: 3 Ways to Create an Exercise Habit. Retrieved January 6, 2021, from https://jamesclear.com/exercise-habit#:~:text=Habits%20are%20behaviors%20that%20you,getting%20started%20time%20after%20time.

4. Exercise. (n.d.). Retrieved January 6, 2021, from http://www.healthofchildren.com/E-F/Exercise.html

5. Stickley, A., Leinsalu, M., DeVylder, J., Inoue, Y., & Koyanagi, A. (2019, August 19). Sleep problems and depression among 237 023 community-dwelling adults in 46 low- and middle-income countries. Retrieved January 06, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6700183/

6. Supartini, A., Honda, T., Basri, N., Haeuchi, Y., Chen, S., Ichimiya, A., & Kumagai, S. (2016, March 03). The Impact of Sleep Timing, Sleep Duration, and Sleep Quality on Depressive Symptoms and Suicidal Ideation amongst Japanese Freshmen: The EQUSITE Study. Retrieved January 06, 2021, from https://www.hindawi.com/journals/sd/2016/8737654/#conclusions

7. Dinis, J., & Bragança, M. (2018). Quality of Sleep and Depression in College Students: A Systematic Review. *Sleep Science, 11*(4), 290-301. doi:10.5935/1984-0063.20180045

8. Gupta, R., Dahiya, S., & Bhatia, M. (2009, April). Effect of depression on sleep: Qualitative or quantitative? Retrieved January 07, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755175/

9. Fluharty, M., Munafò, M. R., Grabski, M., & Taylor, A. E. (2019, May 19). Home - PMC - NCBI. Retrieved January 7, 2021, from https://www.ncbi.nlm.nih.gov/pmc/

10. Cicek, G., Atan, T., Kamuk, Y., & İmamoğlu, O. (2015, June 1). Effects of Exercise on Levels of Depression. Retrieved January 7, 2021, from https://www.researchgate.net/publication/280979480_Effects_of_Exercise_on_Levels_of_Depression

11. Mcleod, S. (2019, July 19). Box plots (also known as box and whister plots). Retrieved January 17, 2021, from https://www.simplypsychology.org/boxplots.html#:~:text=In%20descriptive%20statistics%2C%20a%20box,(or%20percentiles)%20and%20averages.