Lab 2: Description Using Models

Describing the relationship between sleep and mental health.

Geon Kim   Chris Rezny   Jordan Andersen

April 17, 2025

Introduction

This study investigates how sleep duration is influenced by mental health, general health, physical health, and age. While prior research often examines how sleep impacts health, we reverse the focus to explore how these external factors affect sleep itself. Previous studies have found links between mental health and sleep in college populations (Milojevich and Lukowski 2016), as well as differences in sleep by age and gender (Luca et al. 2015). Understanding these dynamics can inform personal and public health strategies.

We use data from the 2022 Behavioral Risk Factor Surveillance System (BRFSS) (Centers for Disease Control and Prevention 2023), which surveys non-institutionalized adults (18+) across all U.S. states and territories via landline and mobile interviews. The sampling design selects one adult per household and includes random mobile respondents to enhance representation. Focusing on self-reported days of poor mental, general and physical health, along with age, we evaluate how these variables relate to sleep. Our aim is to quantify how much these factors contribute to variations in sleep duration across a diverse population.

Operationalization and Wrangling

The BRFSS dataset includes a wide range of health topics, many unrelated to our focus. To align with our research question, we retained only variables related to sleep, mental health, physical activity, general health and age. Among the many available indicators, we selected general measures such as the number of days in the past month a respondent experienced poor mental or general health. This allows us to broadly assess the relationship between health and sleep while leaving room for future analysis of more detailed factors.We also recoded some numeric variables into string labels to improve clarity, making the data more human-readable and flexible for analysis.

To support model development and reduce researcher degrees of freedom, we split our dataset into two parts: a 30% experimental set and a 70% confirmation set. The experimental set was used for exploration and modeling; the confirmation set validated findings. The experimental set contained a sample of 12,686 observations from the original 400,000+. We created plots to confirm that the experimental sample maintained the distributional properties of the full dataset, preserving approximate random sampling and i.i.d. conditions (Appendix A).

Model Specification

1st model

Our base model distributes the number of hours slept (SLEPTIM1) on number of days reported of poor mental health (MENTHLTH_CAT). The results of this model show that the mental health category does have a statistically

significant effect on hours slept, but it only explains a small portion of why people sleep more or less, with the adjusted R-Squared resulting in a low 0.018.

Table 1: Dependent Variable

Variable	Subgroup	Estimate (SE, Stars)
Intercept	—	7.147*** (0.003)
MENTHLTH_CAT	1−10	-0.204*** (0.006)
	11−20	-0.428*** (0.011)
	>20	-0.673*** (0.01)
	Num.Obs.	296013
	Residual Std. Error	1.447
	R²	0.019
	R² Adj.	0.019
	F	1867.476

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

The variance on prediction is about 1.45 hours. We can see in the plot below that those who reported no days of poor mental health in the past month slept on average 7.1 hours per night. On the other end of the spectrum, those who reported 20 or more days of poor mental health in the past month slept 6.5 hours per night on average, with a much wider spread of responses and with the median being closer to 6 hours. This shows that our model describes some parts of the variance of hours slept, but there is still a lot to investigate.
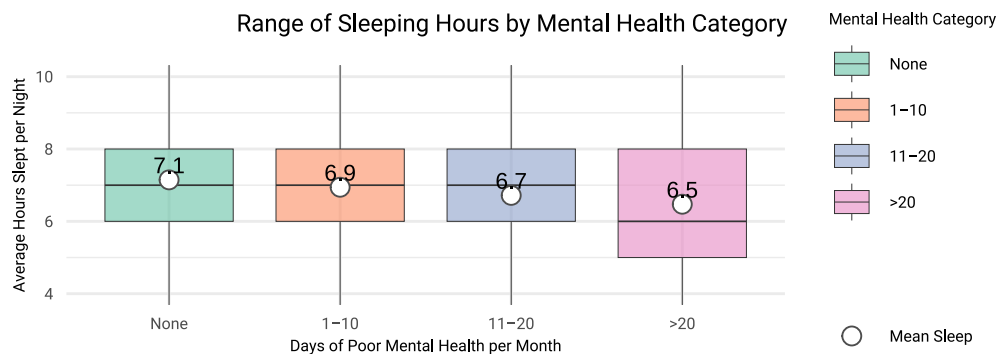

Figure 1: While the distribution of the hours of reported sleep for the first 3 catagories is almost identical, we do see a change in the means. Additionally, the expanded range of sleep hours of those with higher occurances of poor mental health aligns with the anaylsis of the above model.

2nd Model:

We continued to explore the statistical and practical significance of including other variables in the model to determine how we can build a second model that more accurately describes the trends in the data. We seperated variables into demographic and health variables, and explored how those variables affect overall sleep time. Our most successful model from this exploration was sleep time distributed on mental health, physical health, general health and age.

This model resulted in an adjusted R squared of 0.036, showing improvement from the original base model, however still leaving room for significant influence from other factors.

Table 2: Dependent Variable

Variable	Subgroup	Estimate (SE, Stars)
Intercept	—	7.294*** (0.012)
MENTHLTH_CAT	1–10	-0.114*** (0.006)
	11–20	-0.286*** (0.011)
	>20	-0.509*** (0.011)
PHYSHLTH	—	-0.004*** (0)
GENHLTH	Very good	-0.078*** (0.008)
	Good	-0.162*** (0.008)
	Fair	-0.237*** (0.011)
	Poor	-0.222*** (0.017)
AGE_CAT	25-29	-0.147*** (0.016)
	30-49	-0.283*** (0.012)
	50-69	-0.108*** (0.012)
	>70	0.256*** (0.012)
	Num.Obs.	290336
	Residual Std. Error	1.427
	R²	0.038
	R² Adj.	0.038
	F	957.528

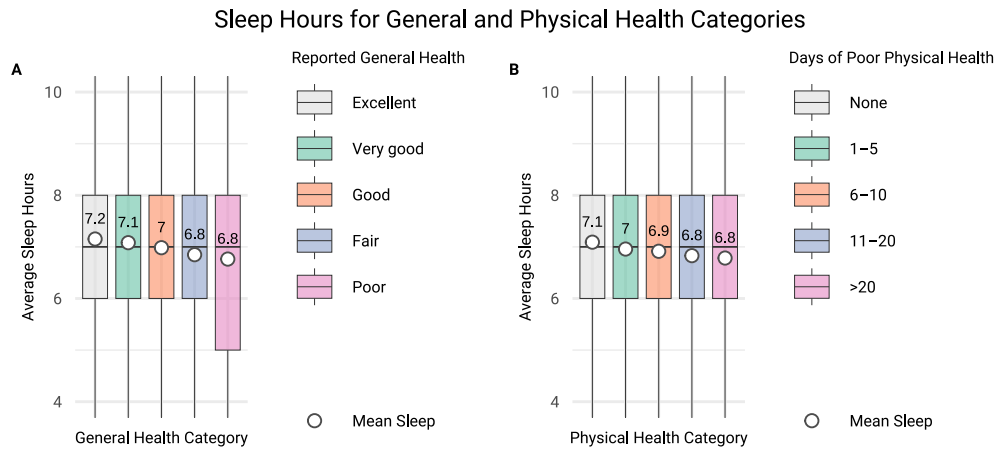+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Figure 2: (A) The distribution of sleep hours based on the reported health of participants shows a shift in the means as general health declines, although the overall distribution of sleep hours remains relatively stable. (B) The distribution of sleep hours per day according to the number of poor physical health days per month similarly indicates a decrease in the means as the number of poor physical health days increases.

Model Assumptions

Assumption 1: Data are IID

The sampling approach attempts to reduce dependence by selecting one adult per household and including random cell phone respondents, however true independence is not guaranteed due to potential clustering within geographic regions and shared household or community environments. Although all participants are US adults, they are not sampled identically. The BRFSS uses a complex sampling design involving stratification and weighting to correct for unequal probabilities of selection, non-response, and to improve representation. Therefore raw data points are not identically distributed, though appropriate application of survey weights can produce approximately unbiased population-level estimates. So, while BRFSS data are not strictly IID, the survey methodology allows for valid inferences.

Assumption 2: Linear Conditional Expectation

To fulfill the Linear Conditional Expectation assumption, no independent variables can be an exact linear combination of the others. Otherwise stated, the plot of their residuals vs fitted values do not have a slope of 0. In our models, the effect size is very small, so the slope is near zero, but not exactly 0 (Appendix B).

Assumption 3: No Perfect Collinearity

We can assume that our models do not have Perfect Collinearity, since no variables were dropped when running a regression. This is illustrated in the summary tables above. One way to confirm that there is no perfect collinearity between coefficients is by applying the variance inflation factor vif() to our model. If we get very high values in return, for example vif > 10, that could suggest that there is multicollinearity within our coefficients.

The vif for the variables MENTHLTH_CAT, GENHLTH, PHYSHLTH and AGE_CAT all fall between 1 and 1.6, representing a low factor (Appendix D). From this we can assume that there is no significant multicollinearity, and our matrix is likely not singular.

Assumption 4: Homoskedastic Conditional Variance

To assess whether the variance of the residuals is constant across levels of the predicted values, we plotted the residuals against the predicted sleep hours. As shown in Figure A of the Residuals vs Fitted Values section (Appendix B), the spread of residuals does not fan out or taper off drastically, suggesting that heteroskedasticity is not visually obvious. However, there are some faint curved patterns that could indicate mild heteroskedasticity.

To confirm this assumption, we conducted a Breusch-Pagan test (Appendix E), which statistically tests whether the residual variance is dependent on the fitted values. A significant p-value ($p < 0.05$) would indicate heteroskedasticity. The p-value result from the Breusch–Pagan test is less than 2.2e-16, which is far below the 0.05 threshold. This result provides strong evidence to reject the null hypothesis of constant variance. In other words, heteroskedasticity is likely present in our model's residuals. While this does not invalidate our model, it suggests that inference based on standard errors should be interpreted with caution.

Given the significant result from the Breusch–Pagan test and evidence of heteroskedasticity, we recalculated our model using heteroskedasticity-robust standard errors via the vcovHC() method with HC1 correction. The results confirmed our original findings, with slightly adjusted standard errors. Most key variables remained statistically significant, but this correction ensures more valid inference (Table 3).

Assumption 5: Normally Distributed Errors

A key assumption of the classical linear model is that residuals are normally distributed. While non-normal residuals do not bias regression coefficients, they may affect hypothesis testing accuracy. In large datasets like ours, the Central Limit Theorem tends to reduce this concern. To check for normality, we generated a Q-Q plot and performed a Shapiro–Wilk test on a random sample of 5,000 residuals.

From the Q-Q plot (Appendix C), we see that the central portion of the residual distribution follows the 45-degree reference line fairly closely. However, the tails deviate from normality, particularly at both extremes, where residuals are more spread out than expected under a normal distribution. This "heavy-tailed" pattern is not uncommon in real-world data.

Table 3: Table: Robust Coefficients

Variable	Subgroup	Estimate	Std. Error	t value	p-value
Intercept	—	7.294	0.012	597.573	0.000 ***
MENTHLTH_CAT	1−10	-0.114	0.006	-18.417	0.000 ***
	11−20	-0.286	0.013	-22.324	0.000 ***
	>20	-0.509	0.015	-34.525	0.000 ***
PHYSHLTH	—	-0.004	0.000	-7.410	0.000 ***
GENHLTH	Very good	-0.078	0.007	-11.266	0.000 ***
	Good	-0.162	0.007	-21.728	0.000 ***
	Fair	-0.237	0.012	-20.553	0.000 ***
	Poor	-0.222	0.024	-9.064	0.000 ***
AGE_CAT	25-29	-0.147	0.015	-9.555	0.000 ***
	30-49	-0.283	0.012	-23.869	0.000 ***
	50-69	-0.108	0.012	-9.146	0.000 ***
	>70	0.256	0.012	20.584	0.000 ***

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Shapiro-Wilk normality test

data: sample(resid(model2), 5000)
W = 0.95389, p-value < 2.2e-16

Executing a Shapiro−Wilk test returned a W statistic of 0.94 and a p-value of < 2.2e-16. This leads us to reject the null hypothesis of normally distributed residuals. Due to the very large sample size, even slight deviations from normality yield statistically significant results. The visual evidence suggests that residuals are not perfectly normal, but close enough in the center of the distribution to satisfy this assumption for regression inference given the size of our dataset.

Overall Effect

Our findings suggest that sleep duration is modestly influenced by mental health, physical health, general health, and age. Individuals who reported more days of poor mental or physical health tended to sleep slightly less, and those with poor self-reported general health also showed lower average sleep hours. While these relationships were statistically significant, the effect sizes were small — for instance, people with over 20 days of poor mental health slept only about half an hour less per night on average than those with none. The expanded model, which included all health and age variables, more than doubled the explanatory power of our base model, yet still explained only about 3.7% of the variance in sleep. This highlights that while health and demographic factors are relevant, sleep is likely shaped by a wider range of behavioral, environmental, and psychological influences not captured in our analysis.

That said, even modest effect sizes should be interpreted in context. For a variable to account for 30 minutes of sleep per night is meaningful when compared relative to the average sleep duration of 7 hours. Small changes may be very significant in real world applications.

Appendix

Appendix A: *Plots for distribution of key variables*
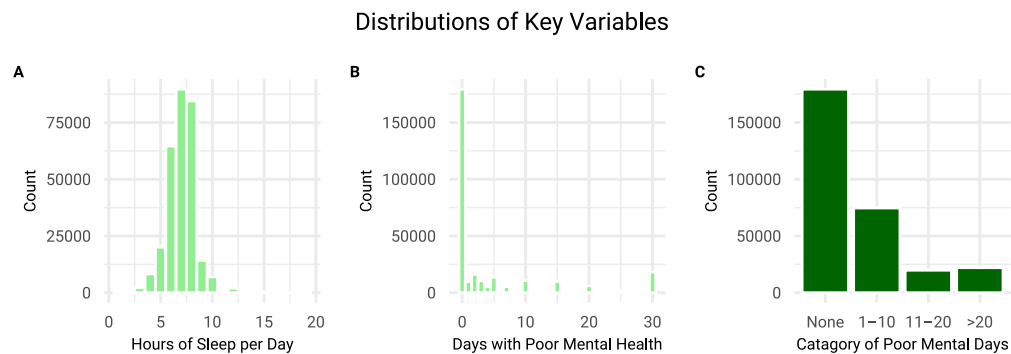
Distributions of Key Variables

Figure 3: Appendix A- The distribution of hours of sleep per day (SLEPTIM1) follows a normal distribution (A). The distribution of the number of poor mental health days during the past 30 days does not follow a normal distribution (B). This suggested that creating descrete variables would necessitate a better analysis (C).

Appendix B: *Plot for Linear Conditional Expectation Assumption*
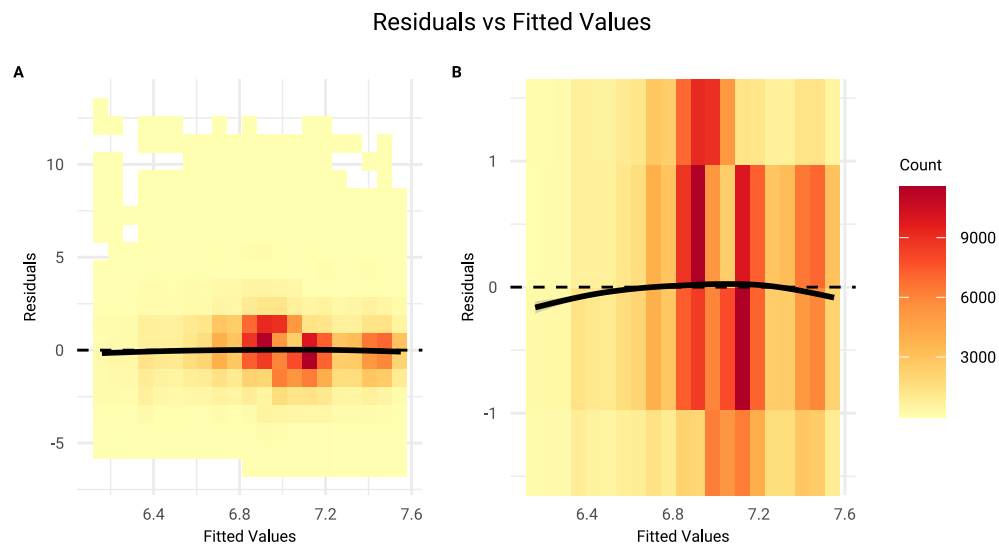
Residuals vs Fitted Values

Figure 4: Appendix B- We plot the residual vs. fitted values of our model to see if the Linear Conditional Expectation (LCE) assumption holds. On the initial plot (A) we see a very flat line, and some serious striation. The striation is expected due to the binned numerical inputs of sleep (whole hours of sleep). Zooming in (B), we see some bending of the regression line. Due to the size of the dataset we used a generalized additive model curve (GAM) instead of a LOESS smoother. While the bend is not extreme it suggests that our model might not contain only nonlinear relationships. This indicates a mild violation of the LCE assumption; however it might not be significant given the complicated nature of sleep.

Appendix C: *Q-Q Plot*
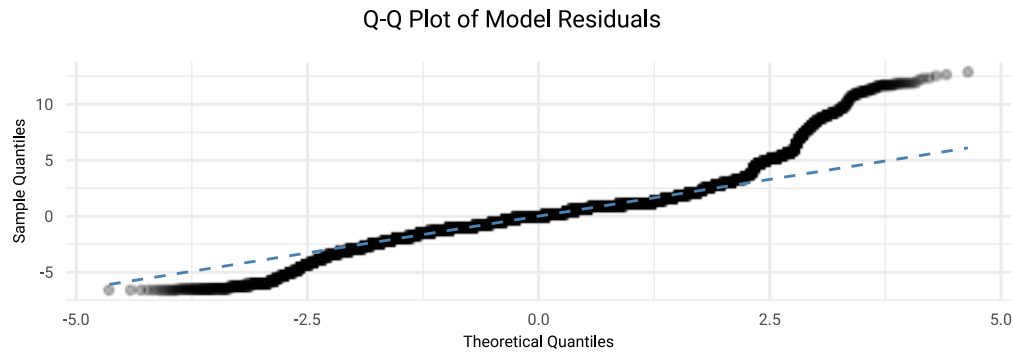
6

Q-Q Plot of Model Residuals

Figure 5: Appendix C- As shown in the Q-Q plot, the central portion of the residual distribution follows the 45-degree reference line fairly closely. However, the tails deviate from normality, particularly at both extremes, where residuals are more spread out than expected under a normal distribution. This "heavy-tailed" pattern is not uncommon in real-world data.

Appendix D: *Vif() factor summary*

Table 4: Variance Inflation Factors

Variable	GVIF	Df	GVIF Adjusted (GVIF$^{1/2Df}$)
MENTHLTH_CAT	1.255	3	1.039
PHYSHLTH	1.594	1	1.263
GENHLTH	1.607	4	1.061
AGE_CAT	1.127	4	1.015

Appendix E: *Breusch-Pagan Test Results*

studentized Breusch-Pagan test

data: model2
BP = 10643, df = 12, p-value < 2.2e-16

Github Repository

Github Project Repository

References

Centers for Disease Control and Prevention. 2023. "Behavioral Risk Factor Surveillance System (BRFSS) 2022 Survey Data." https://www.cdc.gov/brfss/annual_data/annual_2022.html.

Luca, G, J Haba Rubio, D Andries, N Tobback, P Vollenweider, G Waeber, and M Tafti. 2015. "Age and Gender Variations of Sleep in Subjects Without Sleep Disorders." *Annals of Medicine* 47 (6): 482–91. https://doi.org/10.3109/07853890.2015.1074271.

Milojevich, HM, and AF Lukowski. 2016. "Sleep and Mental Health in Undergraduate Students with Generally Healthy Sleep Habits." *PLoS ONE* 11 (6): e0156372. https://doi.org/10.1371/journal.pone.0156372.