University of Calgary

Course Number: DATA 603

A Multiple Regression Analysis of Lifestyle, Mental, and Environmental Factors on Sleep Quality

Group 16

Name	UCID
Ray Al-Saidi	(00271729)
Graeme Ko	(30235563)
Soyebal Saad Adnan	(30216885)
Annie Shamirian	(30223202)

Table of Contents

Table of Figures

Introduction

Sleep, a fundamental aspect of human existence, exceeds being a mere period of rest. It is a complex physiological process crucial for cognitive functions, emotional equilibrium, and physical vitality. The complicated interplay between body and mind during the nocturnal hours unfolds through distinct stages in the sleep cycle, each characterized by unique features. Given today's fast-paced lifestyles contributing to issues like insomnia and insufficient sleep, our quantitative investigation seeks to uncover the intricate relationships between various variables. Sleep affects us all, and understanding these connections may be important to improving sleep quality.

Objectives

This project aims to investigate the complex interplay between lifestyle factors and sleep metrics, emphasizing linear relationships for predictive modeling. By analyzing associations between sleep duration, efficiency, and stages with age and gender, the goal is to facilitate the creation of personalized sleep strategies tailored to individual characteristics. Additionally, the project aims to contribute to the field by investigating the linear connections among bedtime, wakeup time, and sleep metrics, providing a basis for predictive models that optimize sleep schedules. Lastly, the exploration of linear correlations between various sleep metrics seeks to offer comprehensive insights for the development of models geared towards enhancing overall sleep health. Below is a summary of the objectives.

- Timing aspect of sleep
- Associations between sleep metrics
- Age and gender's impact on sleep
- Sleep patterns and lifestyle factors
- Create a predictive model for sleep efficiency

Dataset and Cleaning

The dataset is accessible on Kaggle, encompassing 452 records without duplicates and including 65 entries with missing values. The table below provides details on each variable's name, type, and unit of measurement. The variables Wakeup time and Bedtime were initially in string format, prompting us to convert them into datetime format. Subsequently, we excluded the year, month, and day, assigning a score based on their proximity to midnight. Additionally, any missing values (Na) were removed from the dataset.

Our response variable is sleep efficiency, quantified as the total time spent asleep divided by the total time spent in bed. Consequently, individuals experiencing frequent tossing, turning, and prolonged wakefulness during the night would exhibit a lower sleep efficiency percentage.

Variable	**Type**	**Units**
Age	Quantitative	Years
Gender	Qualitative	Male/Female
Bedtime	Quantitative	Time
Wakeup time	Quantitative	Time
Sleep Duration	Quantitative	Hours
Sleep Efficiency	Quantitative	Percentage
REM sleep Percentage	Quantitative	Percentage
Deep sleep Percentage	Quantitative	Percentage
Light Sleep Percentage	Quantitative	Percentage
Awakenings	Quantitative	Numeric
Caffeine consumption	Quantitative	mg in the last 24 hours
Alcohol Consumption	Quantitative	Oz in the last 24 hours
Smoking Status	Qualitative	Yes/No
Exercise Frequency	Quantitative	Number of times per week

Methodology:

Task Distribution:

Ray Al-Saidi	Data Exploration and Cleaning, Visualizations, Analysis
Graeme Ko	Model Building and Prediction, Analysis
Soyebal Saad Adnan	Interpretations, Assumption tests, Analysis
Annie Shamirian	Interpretations, Assumption tests, Analysis

Multicollinearity

Light sleep exhibited extremely high collinearity with deep sleep, preventing the successful execution of the regression R function. Consequently, we opted to eliminate light sleep from the analysis, prioritizing deep sleep due to its perceived greater logical significance as it is a commonly used sleep phase metric for determining sleep quality. The variables Wakeup time and Bedtime displayed some collinearity as well, leading us to exclude Wakeup time from the analysis, as it exhibited a higher Variance Inflation Factor (VIF) score. We checked the correlation (0.757) between bedtime and wakeup time in the figure below.

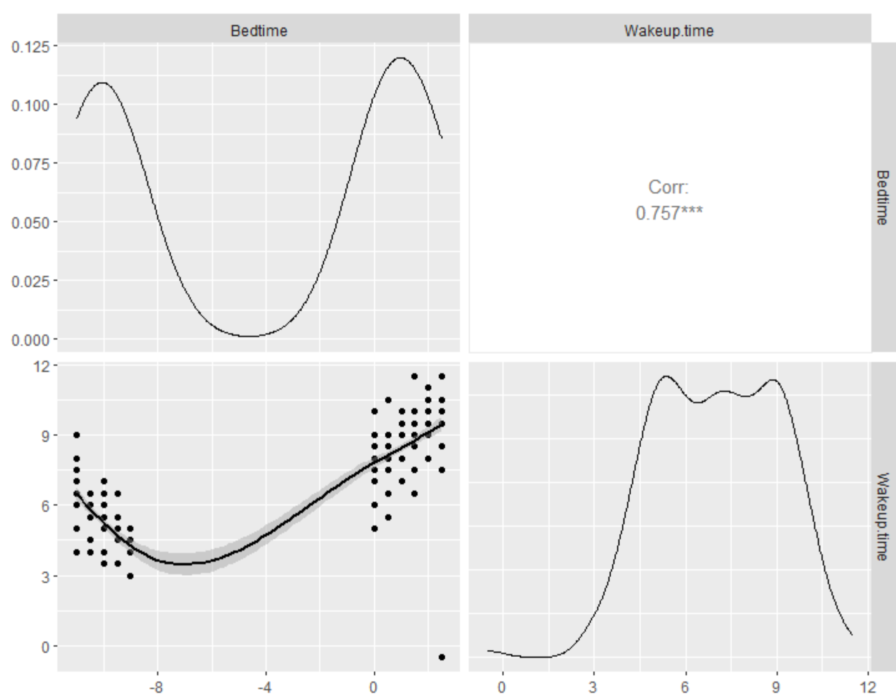

Figure 1 Correlation between Bedtime and Wakeup time

Following the removal of colinear variables, the variables of interest to select the model from are the following:

Age, Gender, Bedtime, Sleep.duration, REM.sleep.percentage, Deep.sleep.percentage, Awakenings, Caffeine.consumption, Alcohol.consumption, Smoking.status, and Exercise.frequency. Below is the output of the regression function in R using these variables.

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              0.3435592  0.0420132   8.177 4.51e-15 ***
Age                      0.0009755  0.0002450   3.982 8.22e-05 ***
factor(Gender)Male       0.0017635  0.0069683   0.253  0.80034
Bedtime                 -0.0004761  0.0006268  -0.760  0.44803
Sleep.duration           0.0020734  0.0035846   0.578  0.56334
REM.sleep.percentage     0.0066984  0.0009398   7.128 5.25e-12 ***
Deep.sleep.percentage    0.0055716  0.0002378  23.429  < 2e-16 ***
Awakenings              -0.0317907  0.0025290 -12.570  < 2e-16 ***
Caffeine.consumption     0.0002450  0.0001135   2.159  0.03150 *
Alcohol.consumption     -0.0061233  0.0021163  -2.893  0.00403 **
factor(Smoking.status)Yes -0.0449020  0.0069311  -6.478 2.91e-10 ***
Exercise.frequency       0.0056560  0.0024895   2.272  0.02366 *
```

Model Selection

We explored the selection of our base model using both the regsubsets and ols best subset methods. In particular, below are the key metrics derived from the ols best subset method output. Despite the closely matched high adjusted R-squared values, our focus shifted to the 8-variable model, driven by its low cp-value indicating heightened model precision. Additionally, attention was drawn to the 11-variable model, where the cp value exactly equaled p+1, or 11 + 1 in this context, implying a potential slight bias within that model. Notably, the 11-variable model boasted the lowest AIC and BIC values among the higher variable models considered for our dataset.

```
      rsquare         cp       BIC     RMSE AdjustedR
 [1,] 0.6226586 344.914740 -366.2247 2.689344 0.6216810
 [2,] 0.7300297 139.504685 -490.1810 1.924101 0.7286273
 [3,] 0.7642764  75.349899 -536.8530 1.680021 0.7624349
 [4,] 0.7877469  32.011795 -571.5856 1.512746 0.7855301
 [5,] 0.7946762  20.626359 -578.5028 1.463360 0.7919887
 [6,] 0.7988358  14.591269 -580.4828 1.433714 0.7956678
 [7,] 0.8026031   9.313934 -581.8570 1.406864 0.7989668
 [8,] 0.8049135   6.850856 -580.4641 1.390398 0.8007956
 [9,] 0.8051562   8.381974 -574.9861 1.388668 0.8005171
[10,] 0.8053208  10.064050 -569.3530 1.387495 0.8001569
[11,] 0.8053540  12.000000 -563.4581 1.387259 0.7996595
```

Below is the summary for the eleven-variable model, yielding an adjusted R^2 of 0.7991 and RMSE of 0.06082. The 8-variable model would be the same as if we were to manually select the model using t-tests with a significance level of 0.05. However, despite initial considerations, we opted not to proceed with the 8-variable model. This decision was influenced by the failure of the 8-variable model to meet linear regression assumptions. Consequently, we have chosen to retain the eleven-variable model and will concentrate on it for the remainder of the report. The RMD file includes all the code and tests used to vet the 8-variable model. Please refer to the file if needed.

```
Coefficients:
                        Estimate Std. Error t value
(Intercept)            0.3436989  0.0434220   7.915
Age                    0.0009757  0.0002459   3.968
GenderMale             0.0017585  0.0069883   0.252
Bedtime               -0.0004669  0.0009456  -0.494
Wakeup.time           -0.0000356  0.0027415  -0.013
Sleep.duration         0.0021003  0.0041473   0.506
REM.sleep.percentage   0.0066962  0.0009558   7.006
Deep.sleep.percentage  0.0055714  0.0002387  23.343
Awakenings            -0.0317903  0.0025326 -12.553
Caffeine.consumption   0.0002451  0.0001142   2.147
Alcohol.consumption   -0.0061242  0.0021201  -2.889
Smoking.statusYes     -0.0448981  0.0069467  -6.463
Exercise.frequency     0.0056564  0.0024931   2.269
                      Pr(>|t|)
(Intercept)           2.81e-14 ***
```

```
Age                 8.68e-05 ***
GenderMale           0.80146
Bedtime             0.62177
Wakeup.time          0.98965
Sleep.duration       0.61285
REM.sleep.percentage  1.14e-11 ***
Deep.sleep.percentage  < 2e-16 ***
Awakenings           < 2e-16 ***
Caffeine.consumption   0.03246 *
Alcohol.consumption    0.00409 **
Smoking.statusYes     3.19e-10 ***
Exercise.frequency     0.02384 *
```

Shown below are the plots for the optimal number of variables for when different criteria such as BIC, CP, RMSE, R^2 and adjust R^2 are considered.

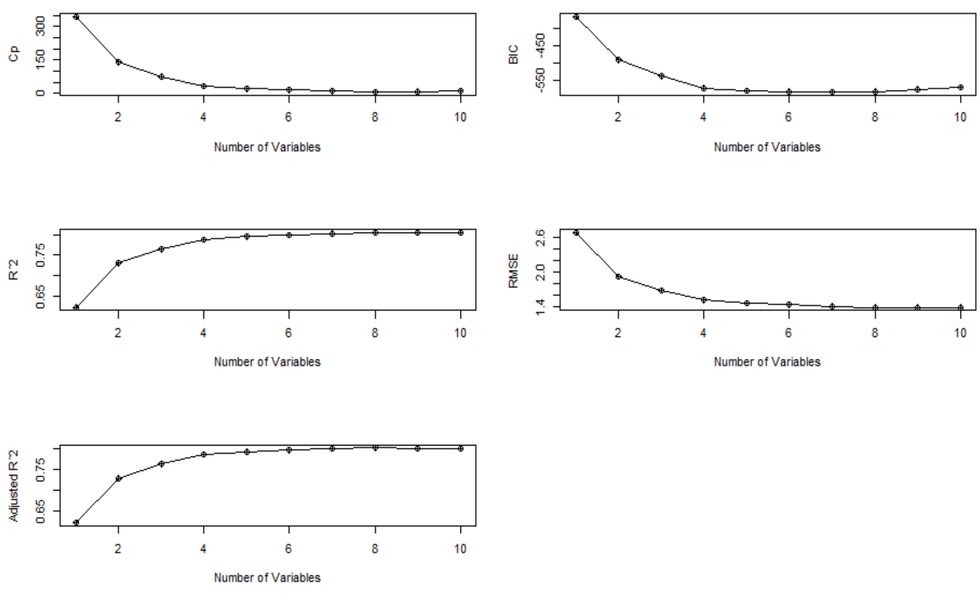

Figure 2 Subset Method Number of Variables Selection

Eight Variable Model

The 8-variable model included the following factors and interactions:

- Age
- REM sleep percentage
- Deep sleep percentage
- Awakenings
- Caffeine consumption
- Alcohol consumption
- Smoking status (categorical)
- Exercise frequency
- Interaction: Age and Smoking status
- Interaction: Deep sleep percentage and Awakenings
- Interaction: Deep sleep percentage and Smoking status
- Interaction: Awakenings and Alcohol consumption
- Interaction: Alcohol consumption and Exercise frequency
- Age squared (quadratic term)

This model passed the Shapiro-Wilk normality test (p-value = 0.2121) but failed the Breusch-Pagan test (p-value = 0.00799). The residual vs. fitted plot did not reveal any discernible pattern, meeting the linearity test. However, even after a Box-Cox transformation with a lambda of 0.959596, the homoscedasticity condition remained unmet. Although no points exceeded Cook's distance of 1, there were leverage points at 82, 258, 303, 397, and 425. Despite removing these points, homoskedasticity still failed. Consequently, we opted for the 11-variable model, as it satisfied the conditions that were not met in the 8-variable model.

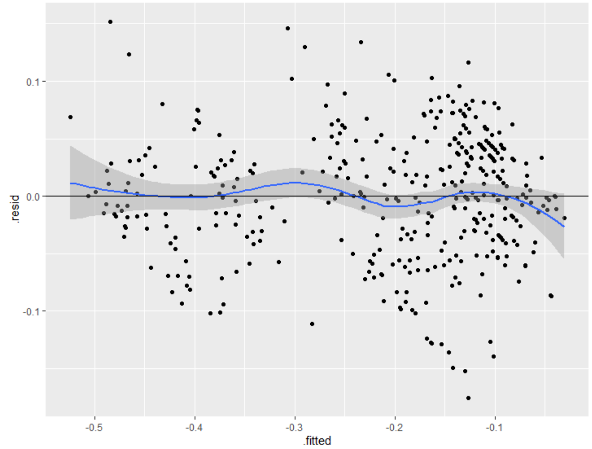

Figure 3 Eight Variable Residual Plot

Therefore, all further sections will be considering the eleven-variable model we have chosen.

Model Building - Interactions

Using the eleven-variable model, we then built an interaction model for it seen in Figure 4. Following the removal of interaction terms with insignificant t-test values, the model has been refined to encompass 12 meaningful interaction terms between variables. This iterative process led to our final model achieving an adjusted R^2 value of 0.8503 and RMSE of 0.0525, indicating some explanatory power for the relationships captured within the interaction terms.

```
Coefficients:
                                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                                   7.057e-02  5.912e-02   1.194 0.233442
Age                                           5.554e-03  8.772e-04   6.332 7.15e-10 ***
factor(Gender)Male                            2.029e-02  7.660e-03   2.649 0.008429 **
Bedtime                                      -1.463e-04  5.618e-04  -0.260 0.794771
Sleep.duration                                5.764e-03  3.792e-03   1.520 0.129338
REM.sleep.percentage                          8.768e-03  1.046e-03   8.384 1.14e-15 ***
Deep.sleep.percentage                         8.702e-03  7.065e-04  12.318  < 2e-16 ***
Awakenings                                    2.063e-02  1.137e-02   1.815 0.070317 .
Caffeine.consumption                          1.000e-04  1.028e-04   0.973 0.331172
Alcohol.consumption                           3.431e-02  1.980e-02   1.733 0.083953 .
factor(Smoking.status)Yes                    -1.398e-01  3.011e-02  -4.643 4.80e-06 ***
Exercise.frequency                            6.123e-03  3.221e-03   1.901 0.058107 .
Age:factor(Smoking.status)Yes                -1.479e-03  4.559e-04  -3.245 0.001284 **
Age:Deep.sleep.percentage                    -6.903e-05  1.365e-05  -5.057 6.77e-07 ***
Age:Awakenings                               -5.248e-04  1.628e-04  -3.224 0.001378 **
factor(Gender)Male:Alcohol.consumption       -1.018e-02  3.626e-03  -2.808 0.005249 **
Sleep.duration:Alcohol.consumption           -4.084e-03  1.999e-03  -2.043 0.041779 *
REM.sleep.percentage:Alcohol.consumption     -1.040e-03  5.029e-04  -2.067 0.039418 *
Deep.sleep.percentage:Awakenings             -6.121e-04  1.620e-04  -3.778 0.000185 ***
Deep.sleep.percentage:factor(Smoking.status)Yes  2.580e-03  4.041e-04   6.385 5.22e-10 ***
Awakenings:Alcohol.consumption                4.096e-03  1.443e-03   2.839 0.004780 **
Awakenings:factor(Smoking.status)Yes          1.294e-02  4.934e-03   2.622 0.009096 **
Awakenings:Exercise.frequency                -3.031e-03  1.531e-03  -1.979 0.048548 *
Alcohol.consumption:Exercise.frequency        5.481e-03  1.338e-03   4.095 5.20e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0525 on 364 degrees of freedom
Multiple R-squared:  0.8592,    Adjusted R-squared:  0.8503
F-statistic: 96.61 on 23 and 364 DF,  p-value: < 2.2e-16
```

Figure 4 Model's interactions Output

Model Building – Higher Order Terms

For the determination of including higher-order terms within the model, we looked at scatterplots of each variable to investigate any non-linear trends between the variables and sleep efficiency. In Figure 5, included the scatterplot for the Age variable, as this was the main predictor we tested for having higher order terms within our model. However, the power terms in the resulting model were not found to be significant and did not significantly improve the fit of the model. Therefore, we decided to keep our interaction model previously discussed with only linear terms.

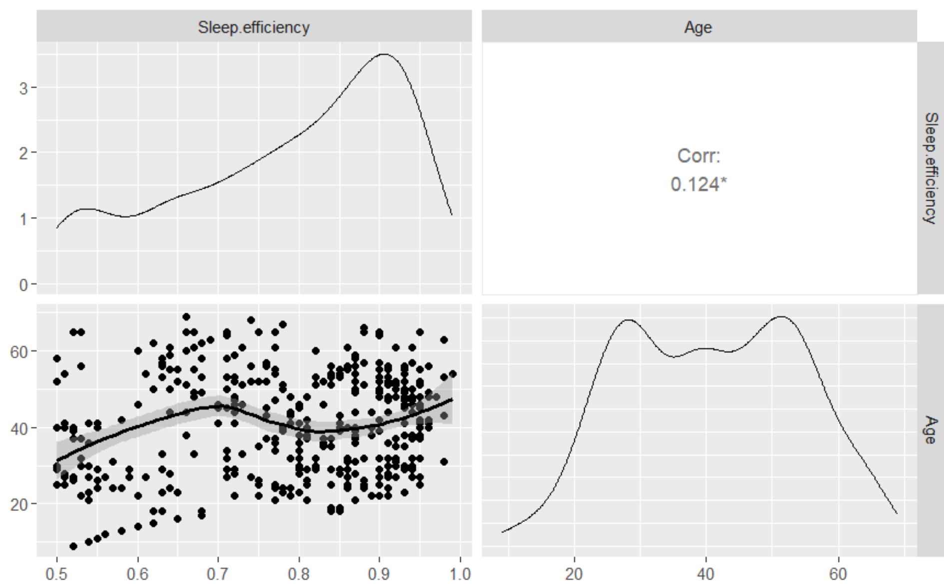

Figure 5 Age Vs. Efficiency Scatterplot3

Outliers

In the analysis of our final 11-variable model, we employed Cook's distance to identify influential points. Notably, no points with a Cook's distance greater than 1 were identified. However, during the examination of high leverage points, we identified the data points 395, 217, and 78 as high leverage influential points. Subsequently, to ensure the accuracy of our model, we made the decision to remove these high-leverage points from the analysis.

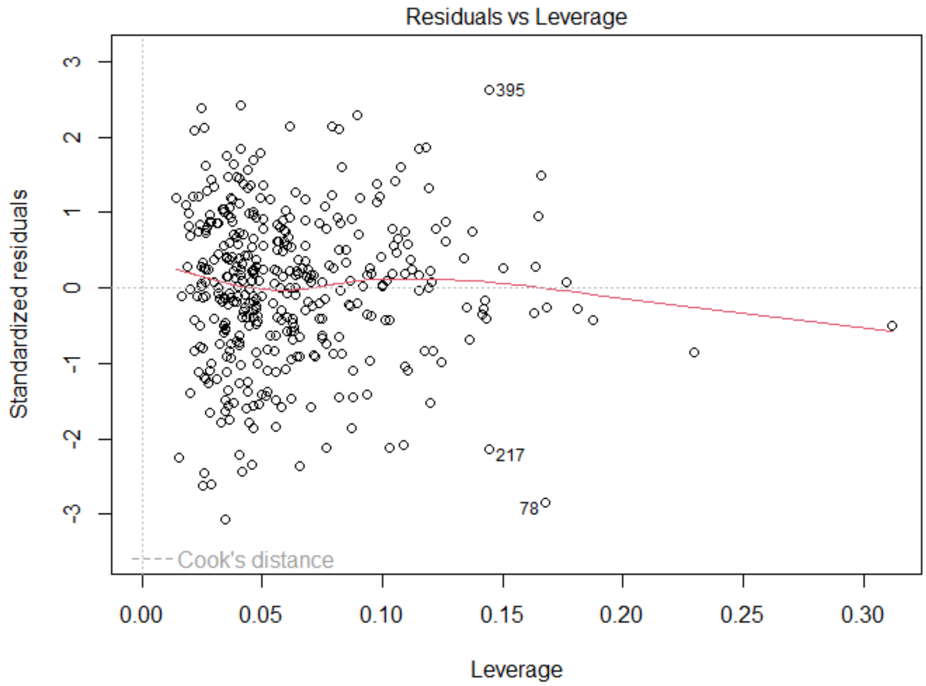

Figure 6 Residuals vs Leverage

In Figure 7, points 78, 217, and 395 clearly stand out in the Cook's distance plot, although they do not exceed a value of one.

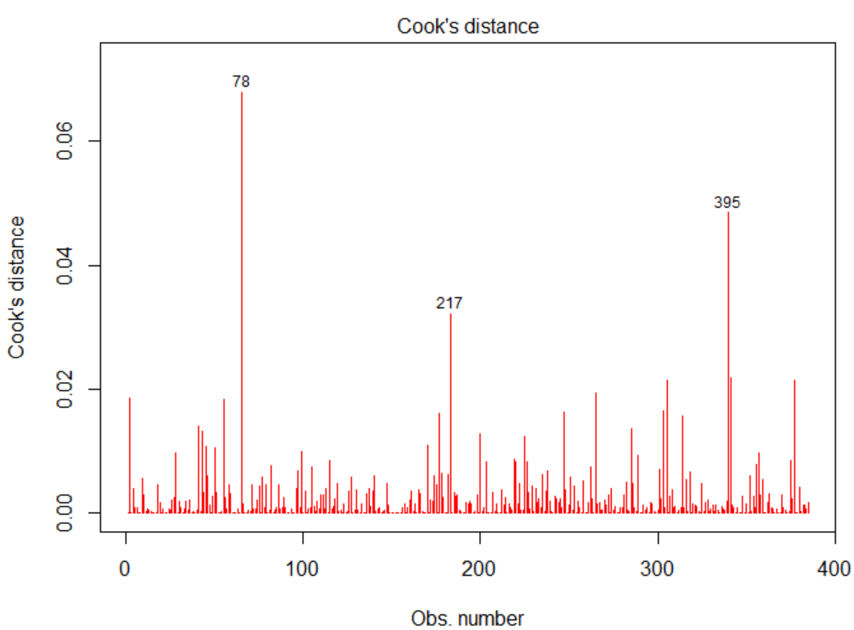

Figure 7 Cook's Distance

Checking the Regression Assumptions

1. Linearity Assumption

The lack of any noticeable pattern in the residual plot depicted in Figure 8 suggests that our assumption of linearity remains valid.

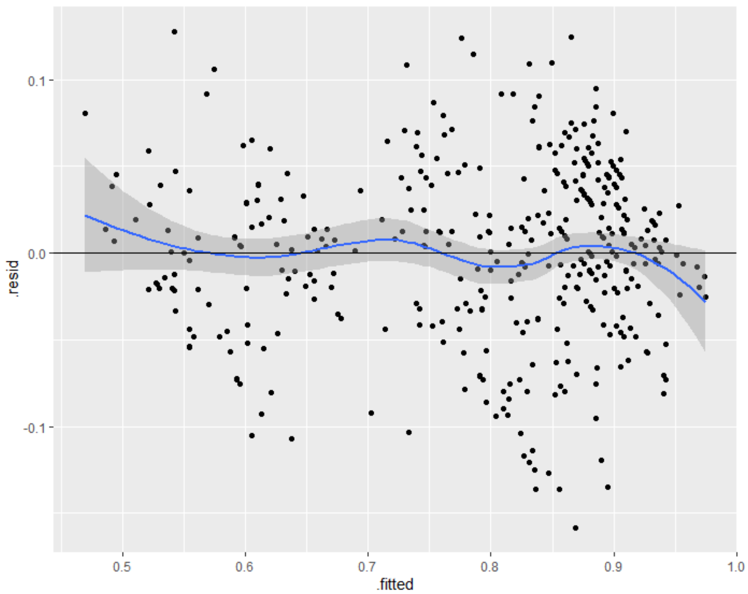

Figure 8 Residual vs Fitted Plot

2. Normality Assumption

H_0 : The sample data residuals are significantly normally distributed

H_A : The sample data residuals are not significantly normally distributed

````
```{r}
shapiro.test(residuals(finalelevenmod))
```


        Shapiro-Wilk normality test

data:  residuals(finalelevenmod)
W = 0.99315, p-value = 0.07705
````

Figure 9 Shapiro's Test Output

With a p-value of 0.07, we do not reject the null hypothesis, indicating that the normality condition is satisfied. Also, we checked the Q-Q plot and points seem to fall close to the diagonal reference line

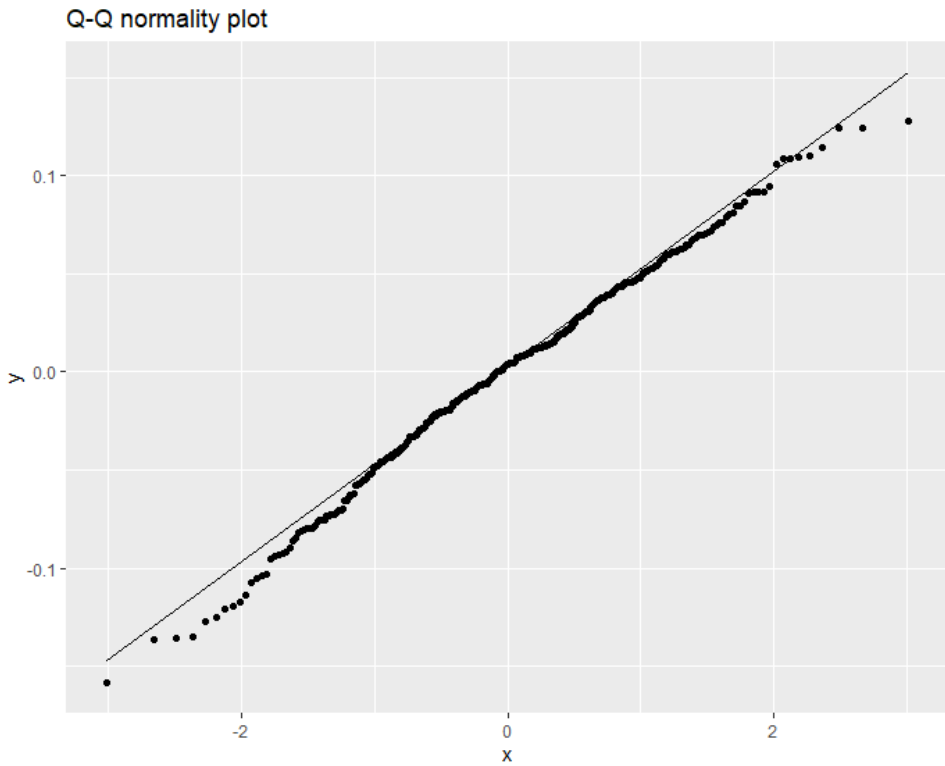

Figure 10 Q-Q Normality Plot

3. Equal Variance Assumption

H_0: Heteroscedasticity is not present (homoscedasticity)

H_A: Heteroscedasticity is present

```
```{r}
bptest(finalelevenmod)
```


        studentized Breusch-Pagan test

 data: finalelevenmod
 BP = 23.071, df = 23, p-value = 0.4566
```

Figure 11 Breusch-Pagan Test Output

With a p-value of 0.4566, we do not reject the null hypothesis, indicating that the homoscedasticity condition is satisfied.

4. **Independence Assumption**

Given that each observation recorded in our dataset belongs to a different person, we assume that it does not have time-series related error and that the data satisfies the independence assumption.

Interpretations

Sleep Pattern and Lifestyle Factors

Extracting individual coefficients from our final model for Caffiene, Alcohol, Smoking gives us -

Alcohol = (B9 + B15 Gender + B16 SleepDuration +B17 REMSleepDuration)

Alcohol = (0.0362 - 0.01082 Gender - 0.00418 SleepDuration – 0.00110 REMSleepDuration)

Caffeine = B8 (no interaction terms)

Caffeine = 0.000101

Smoking = (B10 + B12 Age + B19 DeepSleepPercentage)

Smoking = (-0.1404 - 0.00148 Age + 0.00259 DeepSleepPercentage)

Alcohol Consumption Coefficient Interpretation:

For Every 1 unit increase in Alcohol Consumption leads to a change in sleep efficiency by

(0.0362 - 0.01082 Gender - 0.00418 SleepDuration – 0.00110 REMSleepDuration)

Caffeine Consumption Coefficient Interpretation:

For each additional unit increase in caffeine consumption, there is an estimated increase of 0.000101 units in sleep efficiency, holding other variables constant.

Smoking Coefficient Interpretation:

For Every 1 unit increase in smoking leads to a change in sleep efficiency by

(-0.1404 - 0.00148 Age + 0.00259 DeepSleepPercentage)

In our exploration of factors influencing sleep quality, compared to the standalone impact of caffeine and alcohol, when assessing their effects collectively with other variables, their role may become more apparent. On a different note, our findings highlight a negative correlation between smoking and sleep quality, signifying that smoking tends to diminish overall sleep quality. Conversely, the positive influence of higher exercise frequency on sleep quality stands out as a significant contributor. These insights underscore the importance of considering the interactive effects of various lifestyle factors when evaluating their impact on sleep quality, providing a more nuanced understanding of the complex relationship between habits and the quality of one's sleep.

Age and Gender's Impact on Sleep

Coefficients:

Extracting individual coefficients from our final model for Gender and Age gives us -

Gender = (B2 + B15 Alcohol Consumption)

Gender = (0.0206 -0.01082 Alcohol Consumption)

Age = (B1 + B12 Smoking + B13 Deep Sleep Percentage + B14 Awakening)

Age = (0.00553 - 0.00148 Smoking – 6.89e-5 Deep Sleep Percentage – 0.000504 Awakenings

Age Coefficient Interpretation:

For Every 1 unit increase in age leads to a change in sleep efficiency by

(0.00553 - 0.00148 Smoking – 6.89e-5 Deep Sleep Percentage – 0.000504 Awakenings

Gender Coefficient Interpretation:

For Males, the change in sleep efficiency is caused by:

(0.0206 -0.01082 Alcohol Consumption)

The examination of sleep efficiency average reveals comparable values for both genders. Nevertheless, the data pertaining to females exhibit a greater degree of spread, indicating increased variance in female sleep patterns. On the other hand, the data for males display less variance, suggesting a more consistent sleep efficiency among this demographic.

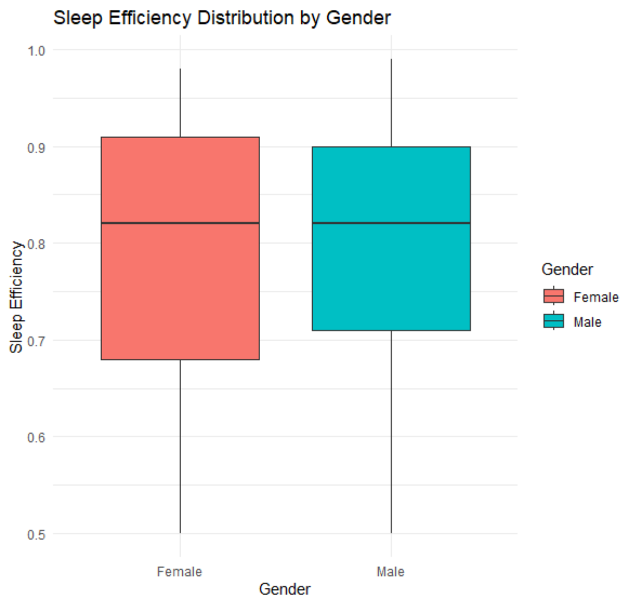

Figure 12 Sleep Efficiency by Gender

The analysis of the chart data reveals variations in sleep efficiency between genders across different age groups. Specifically, females exhibit a lower average sleep efficiency in the earlier years of life. However, this efficiency gradually increases until reaching its peak in the 40s to early 50s, after which a slight decline is observed during the mid-50s to 60s. In contrast, the sleep efficiency for men appears to peak later in life, specifically during their 60s, with notable dips occurring in their late 20s.

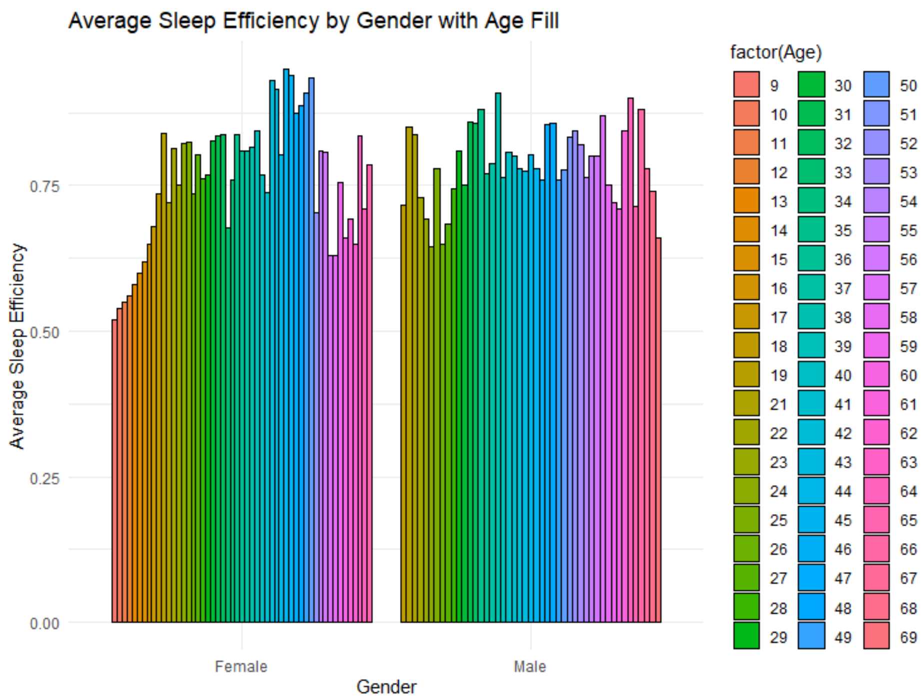

Figure 13 Sleep Efficiency by Gender and Age

Time Aspect of Sleep

Coefficients:

Extracting individual coefficients from our final model for Bedtime gives us -

Bedtime = B3 (no interaction terms)

Bedtime = -0.000123

Bedtime Coefficient Interpretation:

For each additional unit increase in bedtime, there is an estimated decrease of 0.000123 units in sleep efficiency, assuming other variables remain constant.

In our regression analysis of sleep patterns, individual factors such as bedtime and sleep duration exhibit limited significance. However, their relevance becomes apparent when considering their interactions with specific variables. Notably, the frequency of awakenings gains significance in relation to age, revealing an age-dependent influence. Additionally, the depth of deep sleep shows statistical importance when linked to the frequency of awakenings, while sleep duration takes on significance in conjunction with alcohol consumption. This nuanced approach sheds light on the intricate dynamics of sleep patterns, emphasizing the interconnectedness of various factors in our regression framework.

Associations Between Sleep Metrics

Coefficients:

Extracting individual coefficients from our final model for RemSleep, DeepSleep and Sleep Duration -

RemSleepPercentage = (B5 + B17 Alcohol Consumption)

RemSleepPercentage = (0.00883 - 0.00110 Alcohol Consumption)

DeepSleepPercentage = (B6 + B13 Age + B18 Awakening + B19 Smoking)

DeepSleepPercentage = (0.0087 - 6.89e-5 Age – 0.000617 Awakening + 0.00259 Smoking)

Sleep Duration = (B4 + B16 Alcohol)

Sleep Duration = (0.00579 - 0.00418 Alcohol)

Awakening = (B7 + B14 Age + B18 DeepSleepPercentage + B20 Alcohol + B21 Exercise)

Awakening = (0.0201 - 0.000504 Age - 0.000617 DeepSleepPercentage + 0.00440 Alcohol + 0.0124 Exercise)

REMSleepPercentage Coefficient Interpretation:

For Every 1 unit increase in REMSleepPercentage leads to a change in sleep efficiency by (0.00883 - 0.00110 Alcohol Consumption)

DeepSleepPercentage Coefficient Interpretation:

For Every 1 unit increase in DeepSleepPercentage leads to a change in sleep efficiency by (0.0087 - 6.89e-5 Age – 0.000617 Awakening + 0.00259 Smoking)

Sleep Duration Coefficient Interpretation:

For Every 1 unit increase in Sleep Duration leads to a change in sleep efficiency by (0.00579 - 0.00418 Alcohol)

Awakening Coefficient Interpretation:

For Every 1 unit increase in Awakening leads to a change in sleep efficiency by (0.0201 - 0.000504 Age - 0.000617 DeepSleepPercentage + 0.00440 Alcohol + 0.0124 Exercise)

Optimizing different stages of sleep plays a crucial role in enhancing overall sleep efficiency. The data shows that a higher deep sleep percentage contributes to improved sleep efficiency. Deep sleep is characterized by slow-wave sleep, during which the body undergoes essential restorative processes.

Similarly, a higher percentage of REM (Rapid Eye Movement) sleep is also linked to enhanced sleep efficiency. REM sleep is associated with vivid dreaming and cognitive restoration. Interestingly, the relationship between the number of awakenings and sleep efficiency suggests that having more brief awakenings throughout the night can positively impact overall sleep quality. These insights underscore the multifaceted nature of sleep and highlight the importance of considering various sleep stages and patterns for achieving optimal sleep efficiency.

Model Predictions

We found it exciting to assess our model using sleep data from one of our group members. Fortunately, we had access to Graeme's historical sleep data recorded from the "Sleep as Android" app, encompassing all the variables employed in our model. Graeme's actual sleep efficiency for one night was measured at 72.88%, whereas the model predicted a sleep efficiency of 63.53%. The actual sleep efficiency value still falls within the 95% prediction interval of 52.83% to 74.23% estimated by the model, suggesting a reasonable alignment between the predicted and observed values for this case. However, it's important to note that additional data would be necessary for a more comprehensive and robust validation of the model. Figure 14 below shows the data used for the prediction.

Graeme's Values	
Sleep.efficiency	0.729
Age	21
Gender	"Male"
Bedtime	1.4
Sleep.duration	8.25
REM.sleep.percentage	17.9
Deep.sleep.percentage	40
Awakenings	4
Caffeine.consumption	0
Alcohol.consumption	0
Smoking.status	"No"
Exercise.frequency	1

Figure 14 Graeme's Sleep Data 1

Furthermore, we experimented with a more recent night of sleep characterized by a somewhat irregular pattern due to higher alcohol content, frequent awakenings, and a shorter duration. Our model's efficiency prediction differed by less than 2% from the measured value, although it's worth noting that the 95%

confidence interval for this test was relatively large. The actual sleep efficiency was 68.00%. The model predicted Graeme's sleep efficiency to be 69.69%, with a 95% prediction interval ranging from 50.69% to 88.69%.

Graeme's Values 2	
Sleep.efficiency	0.68
Age	23
Gender	"Male"
Bedtime	1.8
Sleep.duration	4.417
REM.sleep.percentage	14.34
Deep.sleep.percentage	25
Awakenings	7
Caffeine.consumption	0
Alcohol.consumption	9
Smoking.status	"No"
Exercise.frequency	0

Figure 15 Graeme's Sleep Data 2

Conclusion

Our comprehensive analysis of various factors influencing sleep quality has yielded several key findings. First, individual factors such as caffeine and alcohol consumption were found to be insignificant on their own, while smoking was identified as a detriment to sleep quality. Conversely, a higher frequency of exercise demonstrated a positive impact on sleep quality. Second, examining the influence of age and gender on sleep revealed that older individuals generally experience better sleep quality, and men tend to achieve better sleep than women. Third, in exploring the temporal aspects of sleep, bedtime and sleep duration were found to be insignificant on their own, with significance emerging when considering interactions such as Awakenings:Age, DeepSleep:Awakenings, and SleepDuration:AlcoholConsumption. Lastly, investigating associations between various sleep metrics uncovered that a higher percentage of deep sleep and REM sleep, as well as an increased number of awakenings, were associated with improved sleep efficiency. These nuanced insights contribute to a more comprehensive understanding of the intricate interplay of factors influencing sleep quality, providing valuable information for both research and practical applications in promoting better sleep health.

References

Sleep Efficiency Dataset. https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency. Accessed 24 Nov. 2023.