Data 603 Project Presentation

A Multiple Regression Analysis of Lifestyle, Demographic, and Sleep Metric Factors on Sleep Efficiency

By:
Ray Al-Saidi
Graeme Ko
Soyebal Saad Adnan
Annie Shamirian

Table of contents

Introduction

...Zzzzzz

Dataset & Cleaning

452 Records **No** Duplicates **65** NAs **X** Format

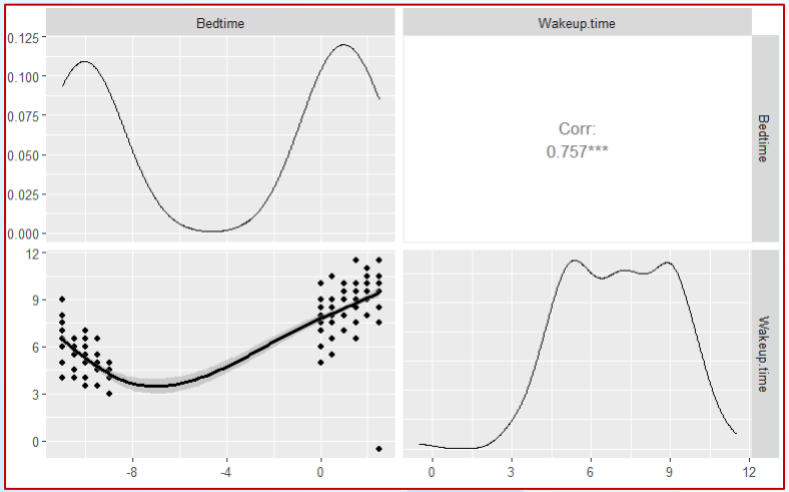
 ID	 Wakeup Time	 Deep Sleep Percentage	 Alcohol Consumption
 AGE	 Sleep Duration	 Light Sleep Percentage	 Smoking Status
 Gender	 Sleep Efficiency	 Awakening	 Exercise Frequency
 Bedtime	 REM Sleep Percentage	 Caffeine Consumption	

Multicollinearity

```r
vif(mod1)
```

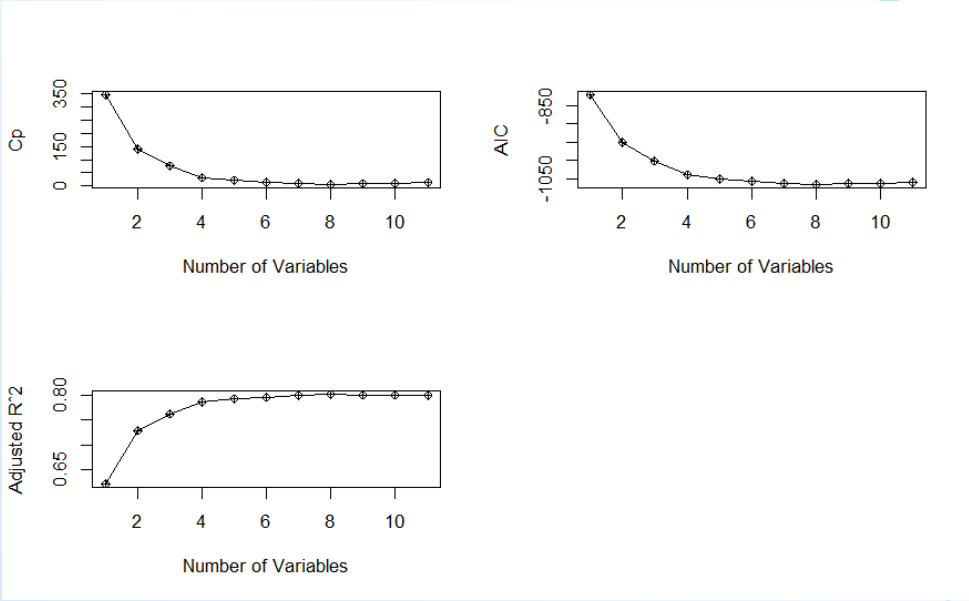
Age	Gender	Bedtime	Wakeup.time	Sleep.duration	REM.sleep.percentage	Deep.sleep.percentage
1.136319	1.280531	2.923099	3.245744	1.379663	1.124628	1.444977
Awakenings	Caffeine.consumption	Alcohol.consumption	Smoking.status	Exercise.frequency		
1.233673	1.146675	1.223044	1.140210	1.362979		

- **Bedtime and Wakeup time is moderately collinear.**
- **VIF > 1.5**
- **Therefore, remove Wakeup time.**



Model Building - Base Model

	AdjustedR	cp	AIC
[1,]	0.6216810	344.914740	-821.9265
[2,]	0.7286273	139.504685	-949.8438
[3,]	0.7624349	75.349899	-1000.4768
[4,]	0.7855301	32.011795	-1039.1704
[5,]	0.7919887	20.626359	-1050.0486
[6,]	0.7956678	14.591269	-1055.9896
[7,]	0.7989668	9.313934	-1061.3248
[8,]	0.8007956	6.850856	-1063.8929
[9,]	0.8005171	8.381974	-1062.3759
[10,]	0.8001569	10.064050	-1060.7038
[11,]	0.7996595	12.000000	-1058.7699



Model Building - Base Model

```
Sleep.efficiency ~ Age + factor(Gender) + Bedtime + Sleep.duration +
REM.sleep.percentage + Deep.sleep.percentage + Awakenings + Caffeine.consumption +
Alcohol.consumption + factor(Smoking.status) + Exercise.frequency
```

```
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 *
---
```

X

X

X

```
Residual standard error: 0.06074 on 376 degrees of freedom
Multiple R-squared:  0.8054,    Adjusted R-squared:  0.7997
F-statistic: 141.4 on 11 and 376 DF,  p-value: < 2.2e-16
```

Model Building - Interactions

```
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 ***
---
```

```
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
```

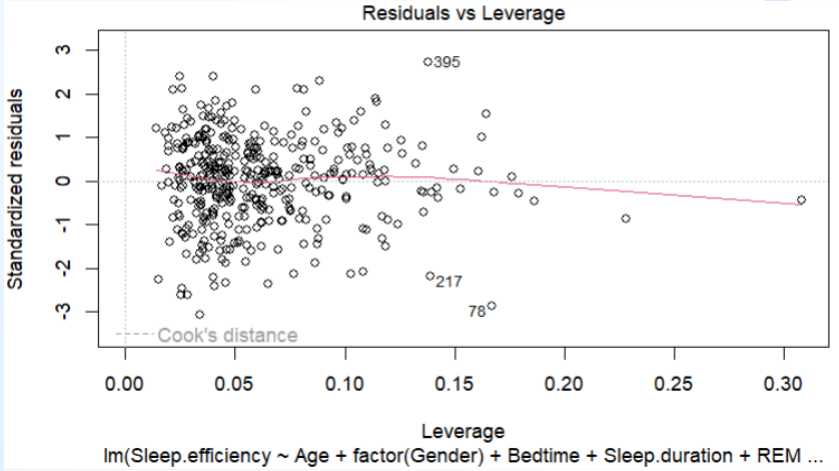
Outlier Detection

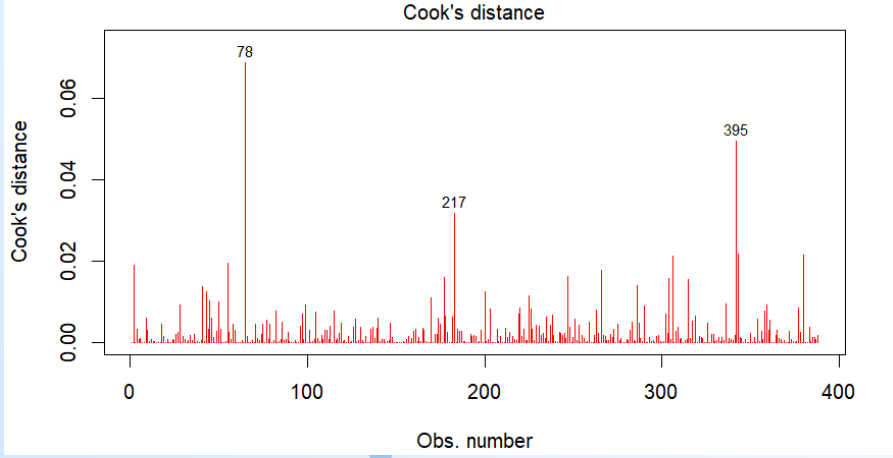
No points with Cook's Distance > 1

Three high-leverage points removed

```{r}
lev=hatvalues(finalelevenmod)
p = length(coef(finalelevenmod))
n = nrow(df)
outlier3p = lev[lev>(3*p/n)]
print(outlier3p)
```

```
      258       336       379
0.3081001 0.2280643 0.1862574
```





Assumption Test

Normality Test

Hypothesis:
H0 : the sample data are significantly normally distributed
Ha : the sample data are not significantly normally distributed

```
        Shapiro-Wilk normality test

data:  residuals(q2.model)
W = 0.98407, p-value = 0.2716
```

Cannot Reject Null Hypothesis.
Sample is normally distributed

Hypothesis:
$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$
HA: at least one sigma_i^{2} different from the others i = 1,2,...,n

Heteroscedasticity Test

```
        studentized Breusch-Pagan test

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

Cannot Reject Null Hypothesis.
The homoskedasticity condition is satisfied

Interpretations

1. *Sleep patterns and lifestyle factors*

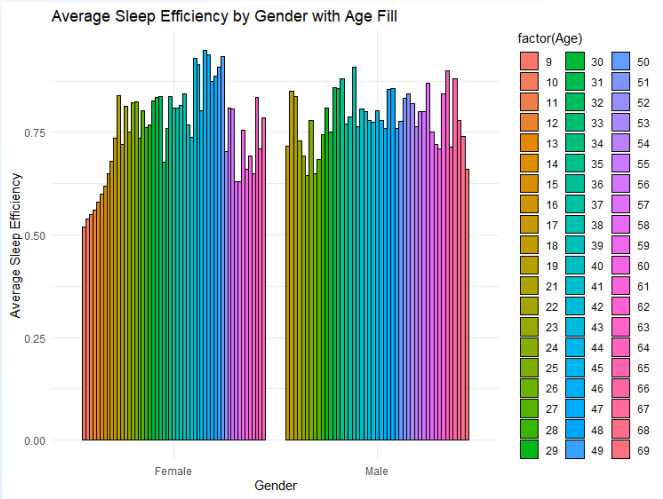
- Caffeine and alcohol consumption are insignificant on their own
- Smoking reduces sleep quality
- Higher exercise frequency improves sleep quality

$$
\begin{aligned}
\text{Sleep.efficiency} &= 6.986 \times 10^{-2} + 5.571 \times 10^{-3} \times \text{Age} + 2.048 \times 10^{-2} \times \text{factor(Gender)Male} - 9.460 \times 10^{-5} \times \text{Bedtime} \\
&+ 5.619 \times 10^{-3} \times \text{Sleep.duration} + 8.796 \times 10^{-3} \times \text{REM.sleep.percentage} + 8.746 \times 10^{-3} \times \text{Deep.sleep.percentage} + 2.078 \times 10^{-2} \times \text{Awakenings} \\
&+ 9.969 \times 10^{-5} \times \text{Caffeine.consumption} + 3.646 \times 10^{-2} \times \text{Alcohol.consumption} - 1.395 \times 10^{-1} \times \text{factor(Smoking.status)Yes} + 6.848 \times 10^{-3} \times \text{Exercise.frequency} \\
&- 1.508 \times 10^{-3} \times \text{Age:factor(Smoking.status)Yes} - 6.970 \times 10^{-5} \times \text{Age:Deep.sleep.percentage} - 5.062 \times 10^{-4} \times \text{Age:Awakenings} \\
&- 1.083 \times 10^{-2} \times \text{factor(Gender)Male:Alcohol.consumption} - 4.228 \times 10^{-3} \times \text{Sleep.duration:Alcohol.consumption} - 1.085 \times 10^{-3} \times \text{REM.sleep.percentage:Alcohol.consumption} \\
&- 6.255 \times 10^{-4} \times \text{Deep.sleep.percentage:Awakenings} + 2.586 \times 10^{-3} \times \text{Deep.sleep.percentage:factor(Smoking.status)Yes} + 4.322 \times 10^{-3} \times \text{Awakenings:Alcohol.consumption} \\
&+ 1.244 \times 10^{-2} \times \text{Awakenings:factor(Smoking.status)Yes} - 3.102 \times 10^{-3} \times \text{Awakenings:Exercise.frequency} + 5.476 \times 10^{-3} \times \text{Alcohol.consumption:Exercise.frequency}
\end{aligned}
$$

Interpretations

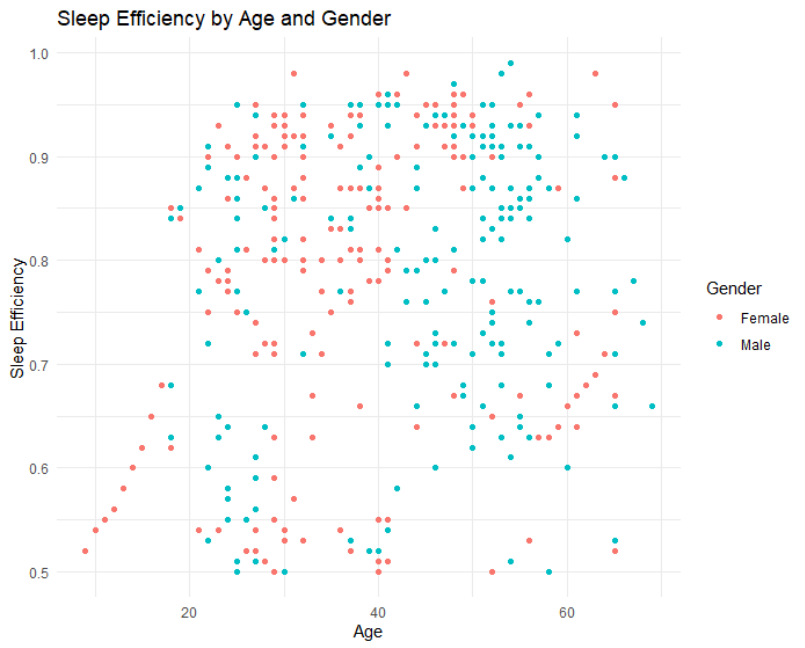
2. Age and gender's impact on sleep

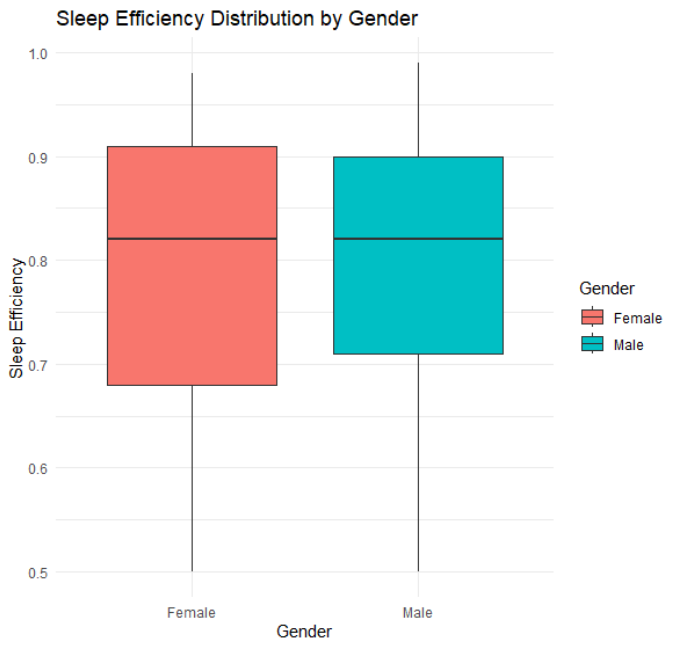
- Older test subjects experience better sleep quality
- Men generally get better sleep



$$
\begin{aligned}
\text{Sleep.efficiency} &= 6.986 \times 10^{-2} + 5.571 \times 10^{-3} \times \text{Age} + 2.048 \times 10^{-2} \times \text{factor(Gender)Male} - 9.460 \times 10^{-5} \times \text{Bedtime} \\
&+ 5.619 \times 10^{-3} \times \text{Sleep.duration} + 8.796 \times 10^{-3} \times \text{REM.sleep.percentage} + 8.746 \times 10^{-3} \times \text{Deep.sleep.percentage} + 2.078 \times 10^{-2} \times \text{Awakenings} \\
&+ 9.969 \times 10^{-5} \times \text{Caffeine.consumption} + 3.646 \times 10^{-2} \times \text{Alcohol.consumption} - 1.395 \times 10^{-1} \times \text{factor(Smoking.status)Yes} + 6.848 \times 10^{-3} \times \text{Exercise.frequency} \\
&- 1.508 \times 10^{-3} \times \text{Age:factor(Smoking.status)Yes} - 6.970 \times 10^{-5} \times \text{Age:Deep.sleep.percentage} - 5.062 \times 10^{-4} \times \text{Age:Awakenings} \\
&- 1.083 \times 10^{-2} \times \text{factor(Gender)Male:Alcohol.consumption} - 4.228 \times 10^{-3} \times \text{Sleep.duration:Alcohol.consumption} - 1.085 \times 10^{-3} \times \text{REM.sleep.percentage:Alcohol.consumption} \\
&- 6.255 \times 10^{-4} \times \text{Deep.sleep.percentage:Awakenings} + 2.586 \times 10^{-3} \times \text{Deep.sleep.percentage:factor(Smoking.status)Yes} + 4.322 \times 10^{-3} \times \text{Awakenings:Alcohol.consumption} \\
&+ 1.244 \times 10^{-2} \times \text{Awakenings:factor(Smoking.status)Yes} - 3.102 \times 10^{-3} \times \text{Awakenings:Exercise.frequency} + 5.476 \times 10^{-3} \times \text{Alcohol.consumption:Exercise.frequency}
\end{aligned}
$$

Sleep Efficiency by Gender





Interpretations

3. Time aspect of sleep

- Bedtime and sleep duration are insignificant on their own
- Significance with Awakenings*Age, DeepSleep*Awakenings, SleepDuration*AlcoholConsumption

$$
\begin{aligned}
\text{Sleep.efficiency} &= 6.986 \times 10^{-2} + 5.571 \times 10^{-3} \times \text{Age} + 2.048 \times 10^{-2} \times \text{factor(Gender)Male} - 9.460 \times 10^{-5} \times \text{Bedtime} \\
&+ 5.619 \times 10^{-3} \times \text{Sleep.duration} + 8.796 \times 10^{-3} \times \text{REM.sleep.percentage} + 8.746 \times 10^{-3} \times \text{Deep.sleep.percentage} + 2.078 \times 10^{-2} \times \text{Awakenings} \\
&+ 9.969 \times 10^{-5} \times \text{Caffeine.consumption} + 3.646 \times 10^{-2} \times \text{Alcohol.consumption} - 1.395 \times 10^{-1} \times \text{factor(Smoking.status)Yes} + 6.848 \times 10^{-3} \times \text{Exercise.frequency} \\
&- 1.508 \times 10^{-3} \times \text{Age:factor(Smoking.status)Yes} - 6.970 \times 10^{-5} \times \text{Age:Deep.sleep.percentage} - 5.062 \times 10^{-4} \times \text{Age:Awakenings} \\
&- 1.083 \times 10^{-2} \times \text{factor(Gender)Male:Alcohol.consumption} - 4.228 \times 10^{-3} \times \text{Sleep.duration:Alcohol.consumption} - 1.085 \times 10^{-3} \times \text{REM.sleep.percentage:Alcohol.consumption} \\
&- 6.255 \times 10^{-4} \times \text{Deep.sleep.percentage:Awakenings} + 2.586 \times 10^{-3} \times \text{Deep.sleep.percentage:factor(Smoking.status)Yes} + 4.322 \times 10^{-3} \times \text{Awakenings:Alcohol.consumption} \\
&+ 1.244 \times 10^{-2} \times \text{Awakenings:factor(Smoking.status)Yes} - 3.102 \times 10^{-3} \times \text{Awakenings:Exercise.frequency} + 5.476 \times 10^{-3} \times \text{Alcohol.consumption:Exercise.frequency}
\end{aligned}
$$

Interpretations

4. Associations between sleep metrics

- Higher deep sleep percentage improves sleep efficiency

- Higher REM sleep percentage improves sleep efficiency

- More awakenings improves sleep efficiency

$$
\begin{aligned}
\text{Sleep.efficiency} &= 6.986 \times 10^{-2} + 5.571 \times 10^{-3} \times \text{Age} + 2.048 \times 10^{-2} \times \text{factor(Gender)Male} - 9.460 \times 10^{-5} \times \text{Bedtime} \\
&+ 5.619 \times 10^{-3} \times \text{Sleep.duration} + 8.796 \times 10^{-3} \times \text{REM.sleep.percentage} + 8.746 \times 10^{-3} \times \text{Deep.sleep.percentage} + 2.078 \times 10^{-2} \times \text{Awakenings} \\
&+ 9.969 \times 10^{-5} \times \text{Caffeine.consumption} + 3.646 \times 10^{-2} \times \text{Alcohol.consumption} - 1.395 \times 10^{-1} \times \text{factor(Smoking.status)Yes} + 6.848 \times 10^{-3} \times \text{Exercise.frequency} \\
&- 1.508 \times 10^{-3} \times \text{Age:factor(Smoking.status)Yes} - 6.970 \times 10^{-5} \times \text{Age:Deep.sleep.percentage} - 5.062 \times 10^{-4} \times \text{Age:Awakenings} \\
&- 1.083 \times 10^{-2} \times \text{factor(Gender)Male:Alcohol.consumption} - 4.228 \times 10^{-3} \times \text{Sleep.duration:Alcohol.consumption} - 1.085 \times 10^{-3} \times \text{REM.sleep.percentage:Alcohol.consumption} \\
&- 6.255 \times 10^{-4} \times \text{Deep.sleep.percentage:Awakenings} + 2.586 \times 10^{-3} \times \text{Deep.sleep.percentage:factor(Smoking.status)Yes} + 4.322 \times 10^{-3} \times \text{Awakenings:Alcohol.consumption} \\
&+ 1.244 \times 10^{-2} \times \text{Awakenings:factor(Smoking.status)Yes} - 3.102 \times 10^{-3} \times \text{Awakenings:Exercise.frequency} + 5.476 \times 10^{-3} \times \text{Alcohol.consumption:Exercise.frequency}
\end{aligned}
$$

Personal Prediction

- Predicted sleep efficiency: 0.635
- 95% Interval prediction: (0.528, 0.742)
- Actual sleep efficiency: 0.729

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

Thank you for listening!

Any questions?

References

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