Is your sleep good enough?

Sleep Efficiency Analysis

Ali Afkhami	30271805
Daniela Mañozca Cruz	30262558
Evan Losier	30022571
Luisa Alejandra Sierra Guerra	30261956
Ruby Nouri Kermani	30261323

Faculty of Graduate Studies - Statistical Modelling with Data (Data 603)

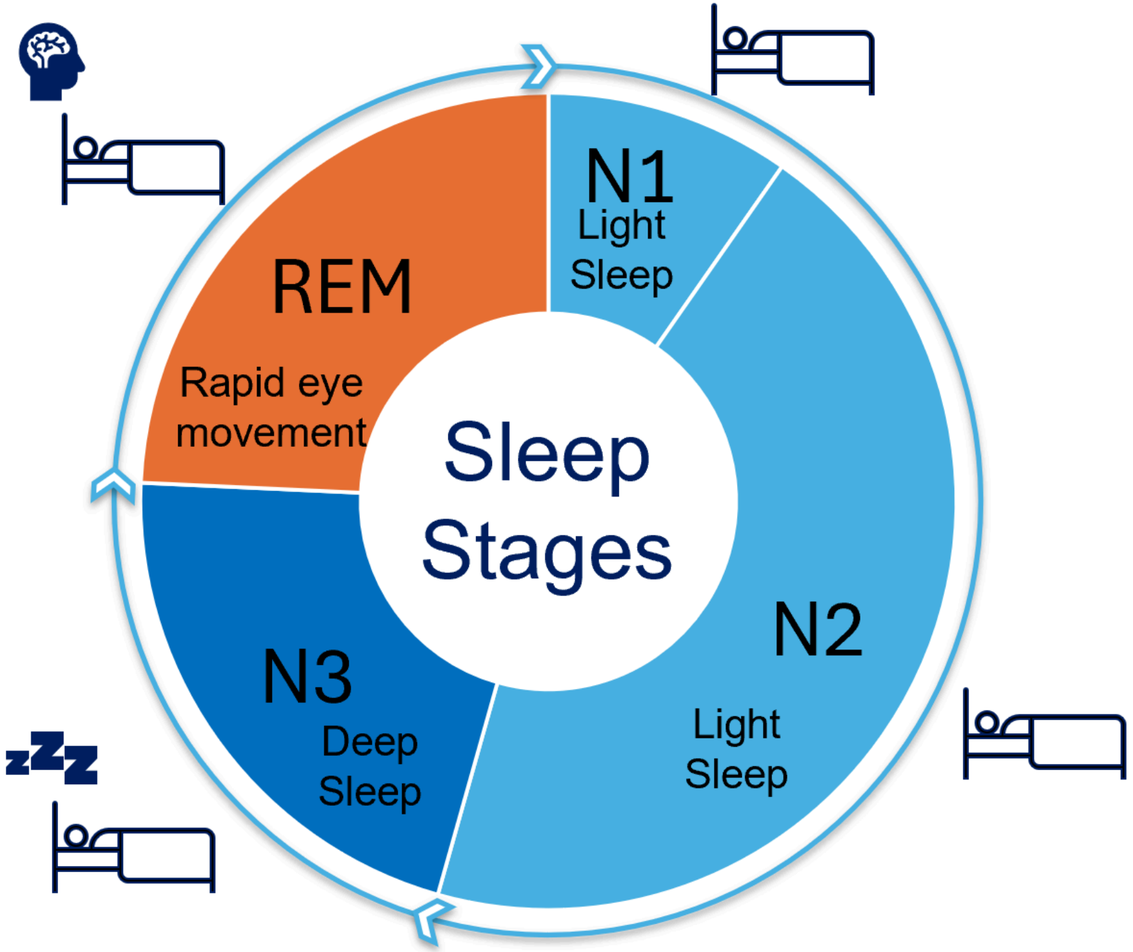
March 1st , 2025

UNIVERSITY OF CALGARY

Agenda

- Introduction
- Data
- Objective
- Methodology – Workflow
- Analysis & Results
- Conclusions
- References

Introduction



Sleep is not uniform. Instead, over the course of the night, your total sleep is made up of several rounds of the sleep cycle, which is composed of four individual stages. **In a typical night, a person goes through four to six sleep cycles** . Not all sleep cycles are the same length, but on average they last about 90 minutes each.

UNIVERSITY OF CALGARY

Data

Data collected from a study conducted in Morocco by a group of artificial intelligence engineering students from ENSIAS.

Records from local community collected over several months. The data was collected using a combination of self-reported surveys, actigraphy, and polysomnography which is a sleep monitoring technique.

452 records

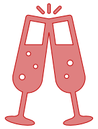
ID	Age	Gender	Bedtime	Wakeup time	Sleep duration	REM sleep %	Deep sleep %	Light sleep %	Awakenings	Caffeine consumption	Alcohol consumption	Smoking status	Exercise frequency	Sleep efficiency
1	65	Female	3/6/2021 1:00	3/6/2021 7:00	6	18	70	12	0	0	0	Yes	3	0.88
2	69	Male	12/5/2021 2:00	12/5/2021 9:00	7	19	28	53	3	0	3	Yes	3	0.66
3	40	Female	5/25/2021 21:30	5/25/2021 5:30	8	20	70	10	1	0	0	No	3	0.89
4	40	Female	11/3/2021 2:30	11/3/2021 8:30	6	23	25	52	3	50	5	Yes	1	0.51
5	57	Male	3/13/2021 1:00	3/13/2021 9:00	8	27	55	18	3	0	3	No	3	0.76
6	36	Female	7/1/2021 21:00	7/1/2021 4:30	7.5	23	60	17	0		0	No	1	0.9

UNIVERSITY OF CALGARY

Objectives

With this project, we aim to develop a model to predict sleep efficiency. To achieve this, we will analyze 11 available variables to determine which ones have an impact on sleep efficiency. Since we are using Multiple Regression Modeling (MRM) techniques, we will validate the assumptions to ensure the model we generate is reliable. Through this process, we aim to answer the following questions:

Is the belief that going to bed early improves sleep efficiency actually valid?



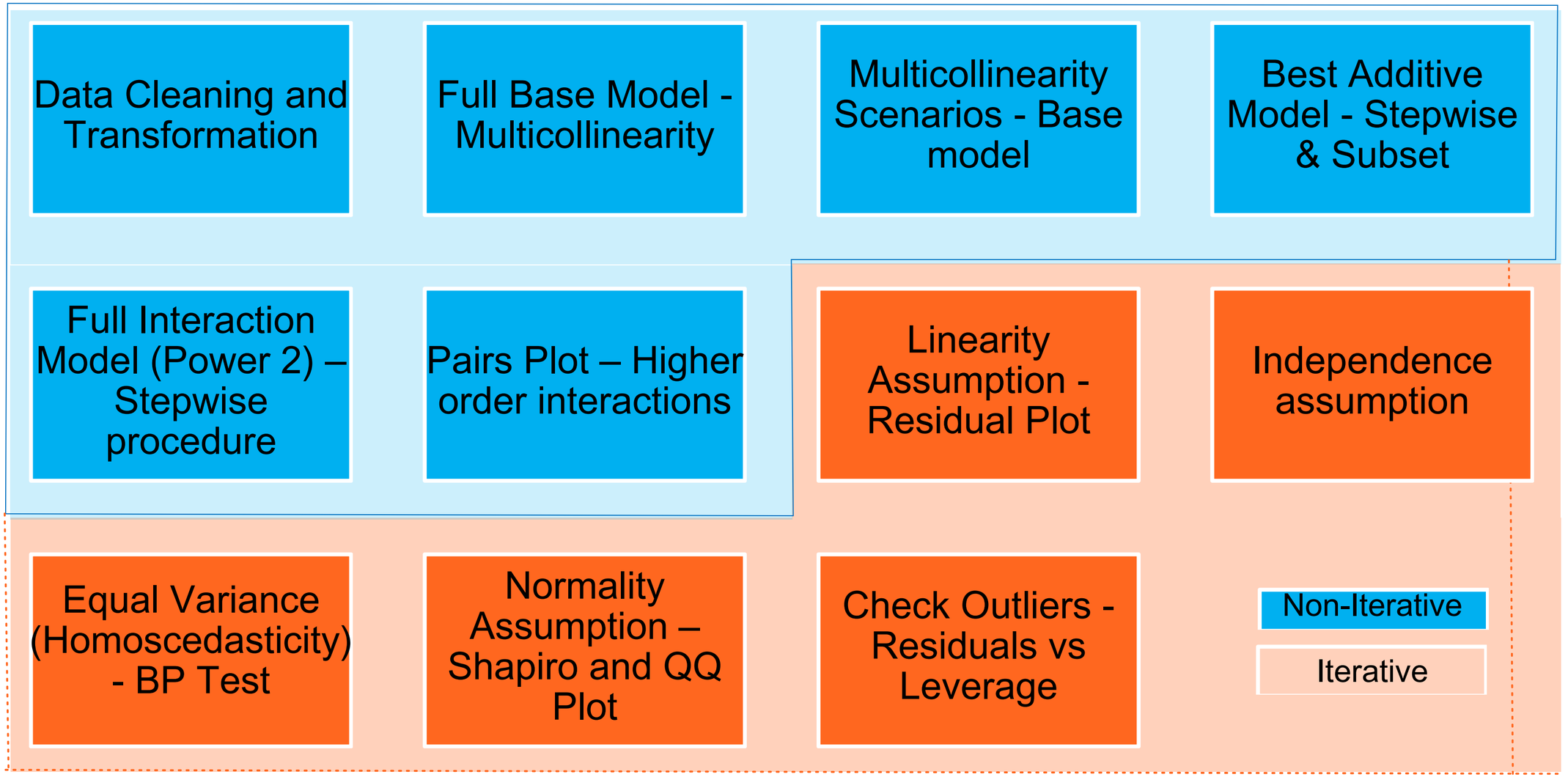
Do unhealthy habits, such as caffeine consumption, alcohol intake, and smoking, have an impact on sleep efficiency?

Can healthy habits, such as working out, influence sleep efficiency?

Should we consider different variables to improve the prediction of sleep efficiency?

UNIVERSITY OF CALGARY

Methodolgy



UNIVERSITY OF CALGARY

6

Analysis and Result

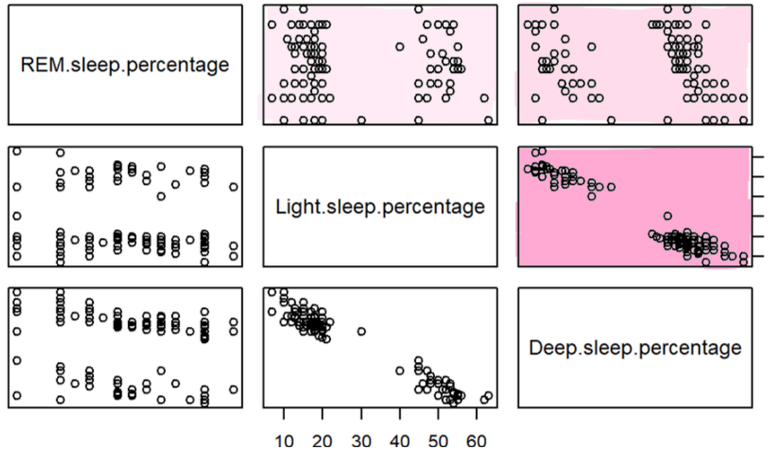
Full Additive Model

$$\begin{aligned}\text{Sleep Efficiency} = {} & \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Factor(Gender)} + \beta_3 \cdot \text{Sleep Duration} + \beta_4 \cdot \text{REM Sleep Percentage} \\ & + \beta_5 \cdot \text{Deep Sleep Percentage} + \beta_6 \cdot \text{Light Sleep Percentage} + \beta_7 \cdot \text{Awakenings} + \beta_{10} \cdot \text{Factor(Smoking Status)} \\ & + \beta_8 \cdot \text{Caffeine Consumption} + \beta_9 \cdot \text{Alcohol Consumption} + \beta_{11} \cdot \text{Exercise Frequency} \\ & + \beta_{12} \cdot \text{Bedtime Shifted} + \beta_{13} \cdot \text{Wake-up Time Shifted} + \epsilon\end{aligned}$$

Individual T - test

(Intercept)	0.5075240	0.2279653	2.226	0.026499	*
Age	0.0007893	0.0002312	3.414	0.000698	***
factor(Gender)Male	0.0032557	0.0064663	0.503	0.614877	
Sleep.duration	0.0025007	0.0045606	0.548	0.583743	
REM.sleep.percentage	0.0071102	0.0008797	8.082	6.23e-15	***
Deep.sleep.percentage	0.0057485	0.0002190	26.245	< 2e-16	***
Light.sleep.percentage	NA	NA	NA	NA	
Awakenings	-0.0317817	0.0022906	-13.875	< 2e-16	***
Caffeine.consumption	0.0001186	0.0001011	1.173	0.241281	
Alcohol.consumption	-0.0058723	0.0019610	-2.994	0.002904	**
factor(Smoking.status)Yes	-0.0425256	0.0064834	-6.559	1.52e-10	***
Exercise.frequency	0.0053861	0.0023593	2.283	0.022912	*
Bedtime_shifted	-0.0287926	0.0316728	-0.909	0.363815	
Wakeuptime_shifted	-0.1277314	0.1907116	-0.670	0.503362	

Multicollinearity



VIF Multicollinearity Diagnostics

	VIF	detection
Age	1.1226	0
factor(Gender)Male	1.2682	0
Sleep.duration	1.8912	0
REM.sleep.percentage	Inf	1
Deep.sleep.percentage	Inf	1
Light.sleep.percentage	Inf	1
Awakenings	1.1900	0
Caffeine.consumption	1.1028	0
Alcohol.consumption	1.2052	0
factor(Smoking.status)Yes	1.1456	0
Exercise.frequency	1.3845	0
Bedtime_shifted	3.7704	0
Wakeuptime_shifted	4.7279	0

UNIVERSITY OF CALGARY

Analysis and Result

Best Additive Model

$$\widehat{Sleep.efficiency} = 0.3526447 + 0.0057675X_{Deep.sleep.percentage} + 0.0072077X_{REM.sleep.percentage} - 0.0413003X_{Smoking.status(YES)} + 0.0007721X_{Age} - 0.0319148X_{Awakenings} - 0.0059629X_{Alcohol.consumption} + 0.0047631X_{Exercise.frequency}$$

Multicollinearity

Model 1.'Deep.sleep.percentage' & 'REM.sleep.percentage'
Model 2. 'REM.sleep.percentage' & 'Light.sleep.percentage'
Model 3. 'Deep.sleep.percentage' & 'Light.sleep.percentage'

VIF Multicollinearity Diagnostics

	VIF	detection
Age	1.1226	0
factor(Gender)Male	1.2682	0
Sleep.duration	1.8912	0
REM.sleep.percentage	1.1649	0
Deep.sleep.percentage	1.4233	0
Awakenings	1.1900	0
Caffeine.consumption	1.1028	0
Alcohol.consumption	1.2052	0
factor(Smoking.status)Yes	1.1456	0
Exercise.frequency	1.3845	0
Bedtime_shifted	3.7704	0
Wakeuptime_shifted	4.7279	0

Stepwise Selection: Both & Subset Best

Model	Adj. R-Square	AIC
1	0.6191	-958.1619
2	0.7282	-1109.6911
3	0.7685	-1181.2158
4	0.7872	-1218.4071
5	0.7923	-1228.2978
6	0.7956	-1234.7116
7	0.7976	-1238.0046
8	0.7976	-1237.0627
9	0.7973	-1235.5429
10	0.7970	-1233.8503
11	0.7967	-1232.0683
12	0.7963	-1230.3293

UNIVERSITY OF CALGARY

Analysis and Result

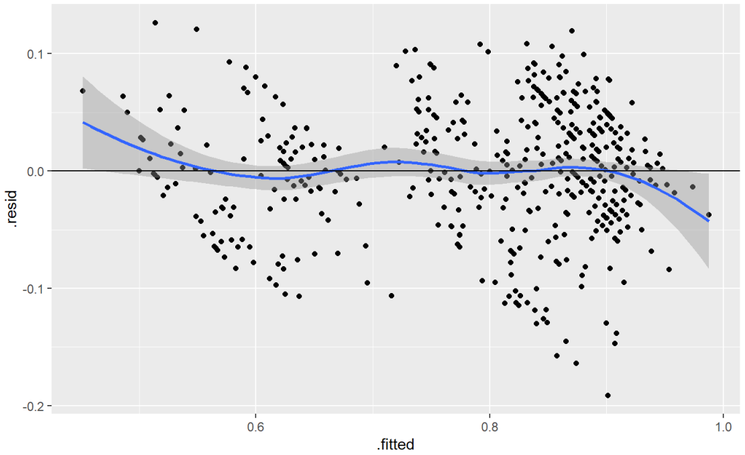
Best Interactive Model

$$\widehat{Sleep.efficiency} = 0.1513 + 0.00274X_{Age} + 0.05622X_{Awakenings} + 0.00653X_{Exercise.frequency} - 0.1448X_{Smoking.statusYes}$$
$$- 0.005624X_{Alcohol.consumption} + 0.008218X_{Deep.sleep.percentage}$$
$$+ 0.002023X_{Smoking.statusYes} \times X_{Deep.sleep.percentage}$$
$$- 0.0008188X_{Awakenings} \times X_{Deep.sleep.percentage}$$
$$- 0.00004226X_{Age} \times X_{Deep.sleep.percentage}$$
$$- 0.001915X_{REM.sleep.percentage} \times X_{Awakenings}$$

OLS Stepwise (Both):
p_enter = 0.05,
p_remove = 0.3

Smoking.status × Deep.sleep.percentage
Awakening × Deep.sleep.percentage
Age × Deep.sleep.percentage
REM.sleep.percentage × Awakenings

Linearity Assumption

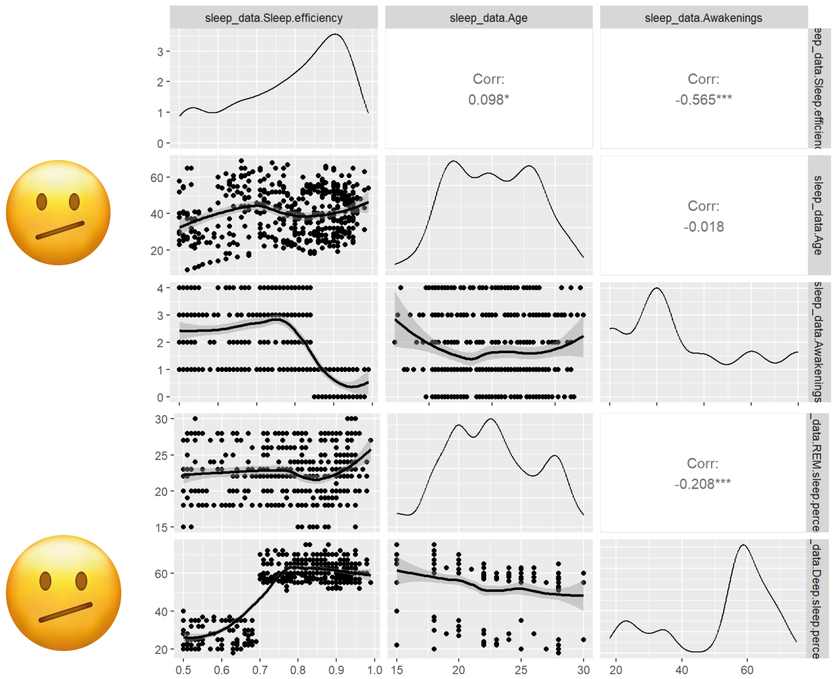


UNIVERSITY OF CALGARY

Analysis and Result

High-order model

To satisfy the **linearity assumption**, we examined which variables could be high-order terms.



Evanmodel = lm(Sleep.efficiency ~ REM.sleep.percentage + Age + Awakenings + Exercise.frequency + Smoking.status + Alcohol.consumption + Deep.sleep.percentage + Smoking.status × Deep.sleep.percentage + Awakenings × Deep.sleep.percentage + Age × Deep.sleep.percentage + REM.sleep.percentage × Awakenings + I(Deep.sleep.percentage^2) + I(Deep.sleep.percentage^3) + I(Awakenings^2) + I(Awakenings^3) + I(Awakenings^4), data=sleep_data)

Danielamodel = lm(Sleep.efficiency ~ REM.sleep.percentage + Age + Awakenings + Exercise.frequency + Smoking.status + Alcohol.consumption + Deep.sleep.percentage + Smoking.status × Deep.sleep.percentage + Awakenings × Deep.sleep.percentage + Age × Deep.sleep.percentage + REM.sleep.percentage × Awakenings + I(Age^2)+I(Age^3)+I(Age^4)+I(Age^5)+ I(Deep.sleep.percentage^2)+I(Deep.sleep.percentage^3)+I(Deep.sleep.percentage^4)+I(Deep.sleep.percentage^5), data=sleep_data)

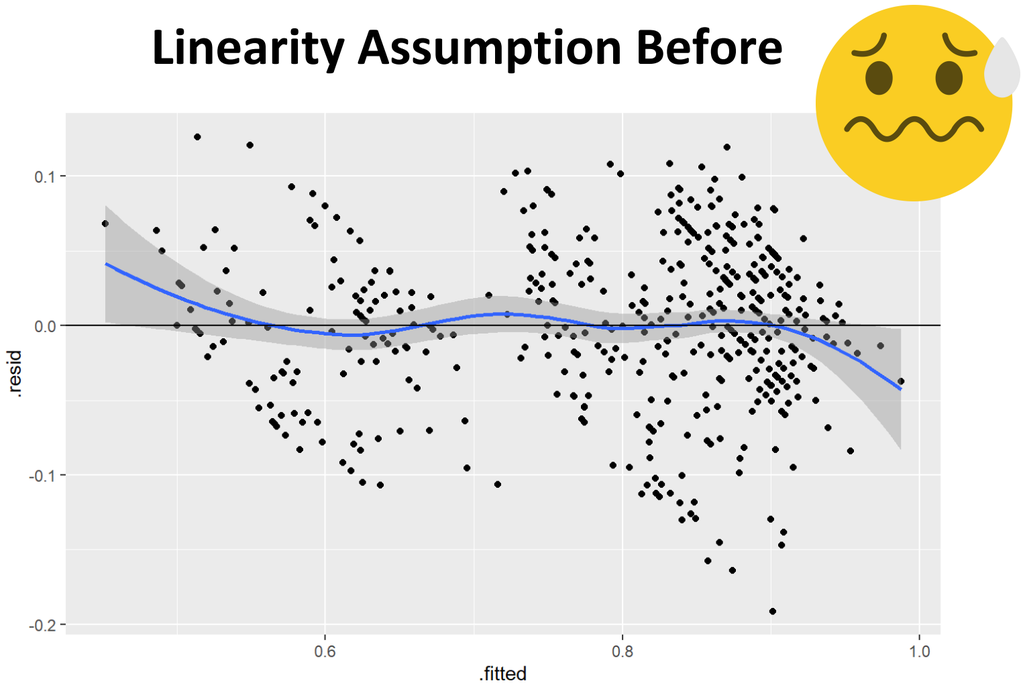
Rubymodel = lm(Sleep.efficiency ~ REM.sleep.percentage + Age + Awakenings + Exercise.frequency + Smoking.status + Alcohol.consumption + Deep.sleep.percentage + Smoking.status × Deep.sleep.percentage + Awakenings × Deep.sleep.percentage + Age × Deep.sleep.percentage + REM.sleep.percentage × Awakenings + I(Awakenings^2) + I(Awakenings^3) + I(Awakenings$^{4)}$ + I(Deep.sleep.percentage^2) + I(Deep.sleep.percentage^3) + I(Age^2), data = sleep_data)

UNIVERSITY OF CALGARY

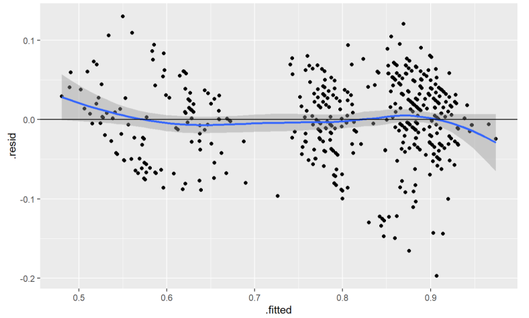
Analysis and Result

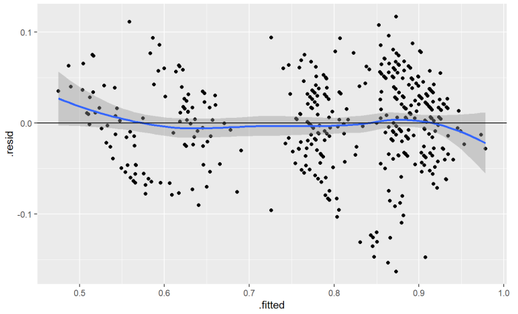
High-order model

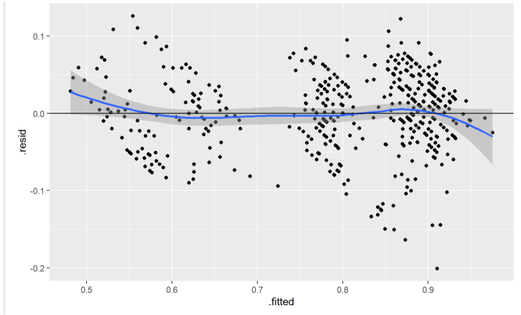
Linearity Assumption Before



Linearity Assumption After







Evan's model

Daniela's model

Ruby's model

UNIVERSITY OF CALGARY

Analysis and Result

Equal Variance Assumption

H_0 : heteroscedasticity is not present (homoscedasticity) ✓

H_a : heteroscedasticity is present

Normality Assumption

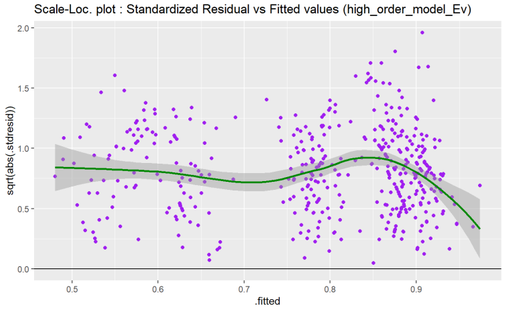
H_0: the sample data are significantly normally distributed ✓

H_a: the sample data are not significantly normally distributed

Analysis and Result

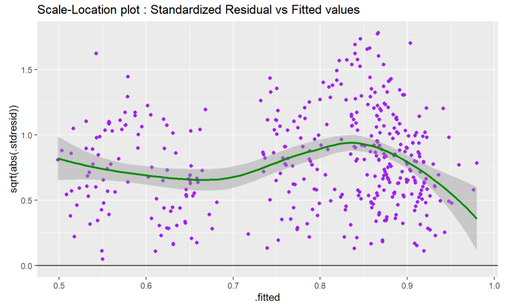
Equal Variance and Normality

Evan's model



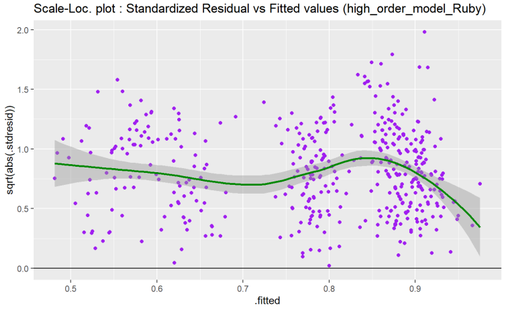

```
        studentized Breusch-Pagan test

data:  evanmodel
BP = 27.33, df = 16, p-value = 0.03796


        Shapiro-Wilk normality test

data:  residuals(evanmodel)
W = 0.98279, p-value = 3.379e-05
```

Daniela's model




```
        studentized Breusch-Pagan test

data:  Danielamodel
BP = 29.795, df = 19, p-value = 0.05447


        Shapiro-Wilk normality test

data:  residuals(Danielamodel)
W = 0.98203, p-value = 5.387e-05
```

Ruby's model



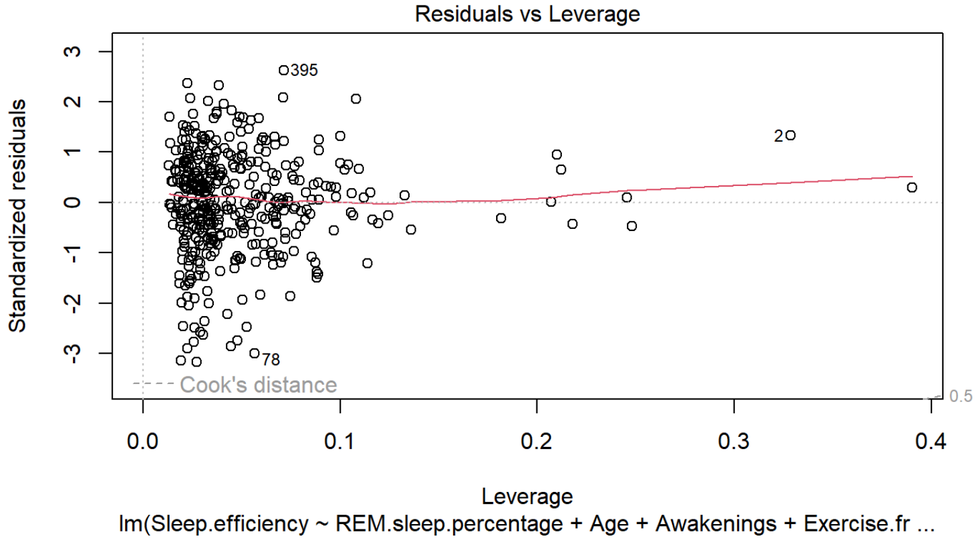

```
        studentized Breusch-Pagan test

data:  rubymodel
BP = 54.522, df = 17, p-value = 8.175e-06


        Shapiro-Wilk normality test

data:  residuals(rubymodel)
W = 0.98478, p-value = 0.0002534
```

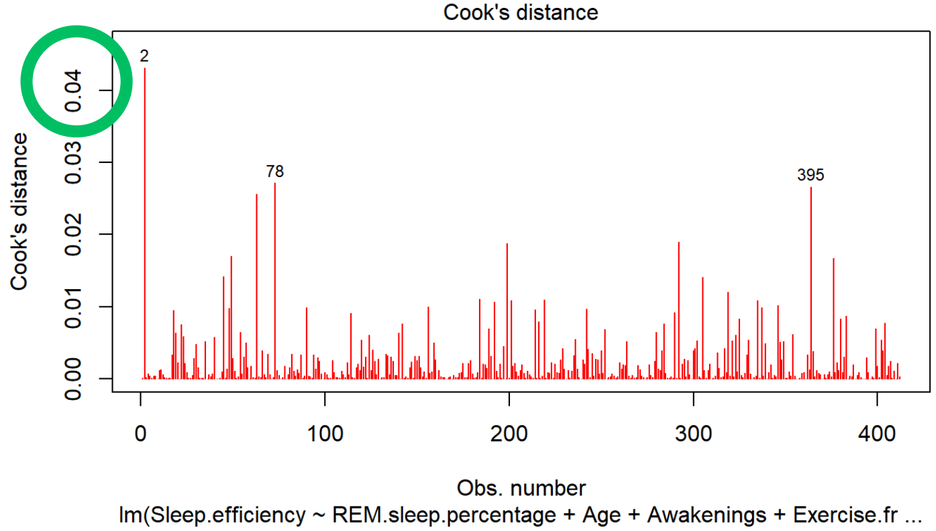
UNIVERSITY OF CALGARY

Analysis and Result

Outliers

Daniela's model





UNIVERSITY OF CALGARY

Analysis and Result

Best Model

Additive Model

Adj. R^2: 0.7976
RSE: 0.0609

Interactive Model

Adj. R^2: 0.8315
RSE: 0.05551

High-order model

Adj. R^2: 0.865
RSE: 0.04994

Source of Variation	df	Sum of Squares	Mean Square	F Statistic
Regression	19	6.609	0.348	138.3
Residual	392	0.986	0.00252	-
Total	411	7.595	-	-

1. Multicollinearity ✓
2. Linearity assumption ✓
3. Independence assumption ✓
4. Equal Variance assuption ✓
5. Outlier ✓
6. Normality assumption ✗

Analysis and Result

Best Model

$$
\begin{aligned}
\widehat{Sleep.efficiency} = & -2.308 + 0.0058X_{REM.sleep.percentage} + 0.1067X_{Age} + 0.0779X_{Awakenings} + 0.0071X_{Exercise.frequency} \\
& - 0.1514X_{Smoking.statusYes} - 0.0062X_{Alcohol.consumption} + 0.2858X_{Deep.sleep.percentage} \\
& - 0.0061X_{Age}^2 + 0.00017X_{Age}^3 - 0.00000216X_{Age}^4 + 0.0000000105X_{Age}^5 \\
& - 0.0148X_{Deep.sleep.percentage}^2 + 0.00036X_{Deep.sleep.percentage}^3 - 0.00000412X_{Deep.sleep.percentage}^4 + 0.0000000177X_{Deep.sleep.percentage}^5 \\
& + 0.0022X_{Smoking.statusYes} * X_{Deep.sleep.percentage} \\
& - 0.00099X_{Awakenings} * X_{Deep.sleep.percentage} \\
& - 0.0000261X_{Age} * X_{Deep.sleep.percentage} \\
& - 0.0024X_{REM.sleep.percentage} * X_{Awakenings}
\end{aligned}
$$

Analysis and Result

Best Model

$$\widehat{Sleep.efficiency} = -2.308 + 0.0058X_{REM.sleep.percentage} + 0.1067X_{Age} + 0.0779X_{Awakenings} + 0.0071X_{Exercise.frequency} - 0.1514X_{Smoking.statusYes} - 0.0062X_{Alcohol.consumption} + 0.2858X_{Deep.sleep.percentage} - 0.0061X^2_{Age} + 0.00017X^3_{Age} - 0.00000216X^4_{Age} + 0.0000000105X^5_{Age} - 0.0148X^2_{Deep.sleep.percentage} + 0.00036X^3_{Deep.sleep.percentage} - 0.00000412X^4_{Deep.sleep.percentage} + 0.0000000177X^5_{Deep.sleep.percentage} + 0.0022X_{Smoking.statusYes} * X_{Deep.sleep.percentage} - 0.00099X_{Awakenings} * X_{Deep.sleep.percentage} - 0.0000261X_{Age} * X_{Deep.sleep.percentage} - 0.0024X_{REM.sleep.percentage} * X_{Awakenings}$$

Submodel for Smokers

$$\widehat{Sleep.efficiency} = -2.4594 + 0.0058X_{REM.sleep.percentage} + 0.1067X_{Age} + 0.0779X_{Awakenings} + 0.0071X_{Exercise.frequency} - 0.0062X_{Alcohol.consumption} + 0.2880X_{Deep.sleep.percentage} - 0.0061X^2_{Age} + 0.00017X^3_{Age} - 0.00000216X^4_{Age} + 0.0000000105X^5_{Age} - 0.0148X^2_{Deep.sleep.percentage} + 0.00036X^3_{Deep.sleep.percentage} - 0.00000412X^4_{Deep.sleep.percentage} + 0.0000000177X^5_{Deep.sleep.percentage} - 0.00099X_{Awakenings} * X_{Deep.sleep.percentage} - 0.0000261X_{Age} * X_{Deep.sleep.percentage} - 0.0024X_{REM.sleep.percentage} * X_{Awakenings}$$

Submodel for Non - Smokers

$$\widehat{Sleep.efficiency} = -2.308 + 0.0058X_{REM.sleep.percentage} + 0.1067X_{Age} + 0.0779X_{Awakenings} + 0.0071X_{Exercise.frequency} - 0.0062X_{Alcohol.consumption} + 0.2858X_{Deep.sleep.percentage} - 0.0061X^2_{Age} + 0.00017X^3_{Age} - 0.00000216X^4_{Age} + 0.0000000105X^5_{Age} - 0.0148X^2_{Deep.sleep.percentage} + 0.00036X^3_{Deep.sleep.percentage} - 0.00000412X^4_{Deep.sleep.percentage} + 0.0000000177X^5_{Deep.sleep.percentage} - 0.00099X_{Awakenings} * X_{Deep.sleep.percentage} - 0.0000261X_{Age} * X_{Deep.sleep.percentage} - 0.0024X_{REM.sleep.percentage} * X_{Awakenings}$$

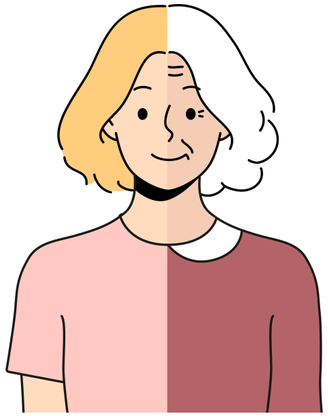
- **Exercise frequency:** Has a small positive effect (+0.0071), supporting that physical activity improves sleep.
- **Smoke status:** Smoking reduces sleep efficiency by -0.1514 points (holding other factors constant).
- **Alcohol consumption:** Reduces sleep efficiency (-0.0062), consistent with research on alcohol's disruptive effects.
- **Age:** Shows a non-linear relationship (polynomial terms up to 5), suggesting sleep efficiency changes differently at various life stages.
- **Deep sleep percentage:** Also follows a complex curve (polynomial terms up to 5), meaning both too little and too much deep sleep could be suboptimal.
- **REM sleep percentage:** Positively affects sleep (+0.0058), but its benefit is reduced with more awakenings (interaction term -0.0024).
- **Awakenings:** More awakenings increase sleep efficiency (+0.0779 per awakening), but this effect weakens with higher deep sleep (interaction term -0.00099)

UNIVERSITY OF CALGARY

Conclusions

Lifestyle factors

- Exercise frequency (mild positive effect).
- Alcohol consumption (small negative effect).



Demographics

- Age (a complex nonlinear relationship, captured by polynomial terms up to the 5th degree, suggesting sleep efficiency changes non-monotonically with age).

Sleep-related variables

- REM sleep percentage (positively associated with sleep efficiency).
- Deep sleep percentage (strong positive effect, with nonlinear polynomial terms indicating diminishing returns at very high levels).
- Awakenings (more awakenings reduce sleep efficiency).

Conclusions

The model indicates that REM sleep percentage (**+0.0058**) and deep sleep percentage (**+0.2858**, with adjustments for smokers) contribute to better sleep efficiency, whereas frequent awakenings (**+0.0779** per episode) decrease it. Regular exercise (**+0.0071**) has a small positive effect, while alcohol consumption (**-0.0062**) and smoking (**-0.1514**) negatively impact sleep. Age influences sleep efficiency in a nonlinear way, meaning its effect varies at different life stages. Additionally, the interaction between deep sleep and awakenings (**-0.00099**) suggests that the benefits of deep sleep diminish when sleep is frequently interrupted.

Recomendations

To improve sleep efficiency, it is beneficial to increase REM and deep sleep by maintaining a consistent sleep schedule and creating a restful environment. Reducing nighttime awakenings—by minimizing noise, stress, and caffeine intake—can further enhance sleep quality. Regular exercise provides a slight boost to sleep efficiency, but avoiding alcohol and smoking is crucial, as they significantly impair sleep. Since age affects sleep differently over time, adapting sleep habits to individual needs is essential. Lastly, prioritizing uninterrupted deep sleep can maximize its restorative benefits.

References

- The City of Calgary, 2025. *The City of Calgary´s Open Data Portal.* [Online] Available at: https://data.calgary.ca/stories/s/About/mgd4-4snb [Accessed 18 01 2025].

- Pandas, n.d. Pandas. [Online] Available at: https://pandas.pydata.org/docs/reference/api/pandas.MultiIndex.from_product.html[Accessed 2025].

- Ebrahim, M., 2024. Like geeks. [Online] Available at: https://likegeeks.com/seaborn-line-plot-secondary-y-axis/[Accessed 2025].

- Plotting Choropleth Maps using Python (Plotly). 2021. [Film] s.l.: Indian Pythonisa Youtube Channel.

UNIVERSITY OF CALGARY

Thank you

Questions?