SCHOOL OF ENGINEERING AND TECHNOLOGY

ASSIGNMENT / PROJECT SUBMISSION FORM

PROGRAMME: BSc (Hons) Information Systems (Data Analytics)

SEMESTER: Jan / Apr / Aug 2024

SUBJECT: STAT1114/STAT2044 Statistics for Data Analytics

DEADLINE: 02 August 2024, 11.59 pm

INSTRUCTIONS TO CANDIDATES

- This is a group project.

IMPORTANT
The University requires students to adhere to submission deadlines for any form of assessment. Penalties are applied in relation to unauthorized late submission of work.

- Coursework submitted after the deadline but within 1 week will be accepted for a maximum mark of 40%.
- Work handed in following the extension of 1 week after the original deadline will be regarded as a non-submission and marked zero.

Lecturer's Remark (Use additional sheet if required)
List down the name of the group members and the student IDs here.

I **AYU WEN LI** (Student's Name) **22017867** (Student ID) received the assignment and read the comments.

***wenli* (02/8/2024)** (Signature/Date)

I **KEERTANA A/P SUBRAMANIAM** (Student's Name) **23109614** (Student ID) received the assignment and read the comments.

***keertana* (02/8/2024)** (Signature/Date)

I **SIOW QI YUNG** (Student's Name) **22053037** (Student ID) received the assignment and read the comments.

qiyung (02/8/2024) (Signature/Date)

I **WONG HUI SAN** (Student's Name) **22034540** (Student ID) received the assignment and read the comments.

huisan (02/8/2024) (Signature/Date)

Academic Honesty Acknowledgement

"I **AYU WEN LI** (Student's Name) verify that this paper contains entirely my own work. I have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, I have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. I realize the penalties *(refer to Section 5.3 (page 32) and Appendix B (page 62-70) of the Student Handbook)* for any kind of copying or collaboration on any assignment."

***wenli* (02/8/2024)** (Student's signature / Date)

"I **KEERTANA A/P SUBRAMANIAM** (Student's Name) verify that this paper contains entirely my own work. I have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, I have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. I realize the penalties *(refer to Section 5.3 (page 32) and Appendix B (page 62-70) of the Student Handbook)* for any kind of copying or collaboration on any assignment."

***keertana* (02/8/2024)** (Student's signature / Date)

"I **SIOW QI YUNG** (Student's Name) verify that this paper contains entirely my own work. I have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, I have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. I realize the penalties *(refer to Section 5.3 (page 32) and Appendix B (page 62-70) of the Student Handbook)* for any kind of copying or collaboration on any assignment."

qiyung (02/8/2024) (Student's signature / Date)

"I **WONG HUI SAN** (Student's Name) verify that this paper contains entirely my own work. I have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, I have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. I realize the penalties *(refer to Section 5.3 (page 32) and Appendix B (page 62-70) of the Student Handbook)* for any kind of copying or collaboration on any assignment."

huisan (02/8/2024) (Student's signature / Date)

Data Protection

The protection of personal data is an important concern to Sunway University and any personal data collected on this form will be treated in accordance with the Personal Data Protection Notice of the institution.
http://sunway.edu.my/pdpa/notice_english (English version)
http://sunway.edu.my/pdpa/notice_bm (Malay version)

Final Report

1. Introduction

This dataset which is titled the "Sleep Health and Lifestyle Dataset," is sourced from Kaggle. It is a synthetic dataset and is assumed to be representing a global population. The dataset includes information on individuals' sleep health and various factors related to their sleep patterns.

This data can contribute to the understanding and promotion of Sustainable Development Goal 3: Good Health and Well-being. By investigating the relationship between lifestyle factors and sleep patterns, this study aims to enhance our understanding of how various aspects of daily life and health metrics may affect sleep quality and duration. The insights gained could help shape public health strategies to promote better sleep through lifestyle changes, recognizing that good sleep is essential for overall health and well-being.

Link: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

This dataset contains information on 374 adults, ranging in age from 27 to 59 years old. It includes 13 variables that cover various aspects of their lives, including demographics, sleep patterns, physical health indicators, and lifestyle factors.

The demographic information includes a unique **Person ID** for each individual, their **Gender** (Male or Female), **Age** in years, and **Occupation**.

Sleep-related data is captured through two variables: **Sleep Duration**, measured in hours per day, and **Quality of Sleep**, rated on a scale from 1 to 10. The dataset also notes the presence or absence of **Sleep Disorders**, specifically categorizing them as None, Insomnia, or Sleep Apnea.

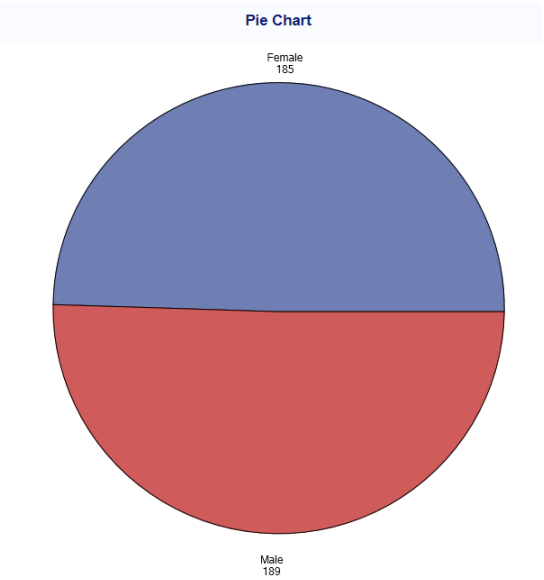
Physical health is represented by several indicators. The **BMI Category** classifies individuals as Underweight, Normal, or Overweight. **Blood Pressure** is recorded as systolic over diastolic pressure. **Heart Rate** is measured in beats per minute (bpm).

Lifestyle factors are reflected in the **Physical Activity Level**, recorded as minutes per day of engagement, and **Daily Steps** taken. The dataset also includes a **Stress Level** measurement, rated on a scale from 1 to 10.

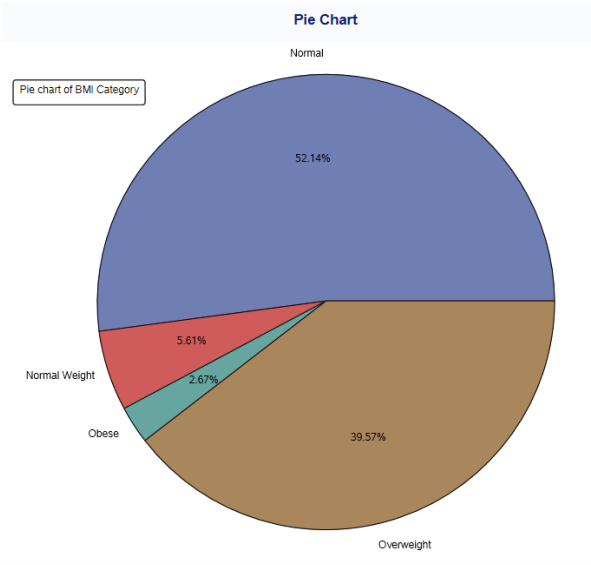
Objectives:

1. To evaluate whether the average sleep duration of respondents in our dataset aligns with the recommended sleep duration reported in a relevant article.
2. To examine potential differences in sleep quality between females and males, considering the role of gender in sleep health.
3. To analyse the impact of BMI category on sleep duration while accounting for other influential factor, such as sleep quality, to explore how body composition may influence sleep patterns.
4. To explore the factors influencing sleep duration and develop a predictive model based on various lifestyle and health indicators, aiming to identify key determinants of sleep duration.

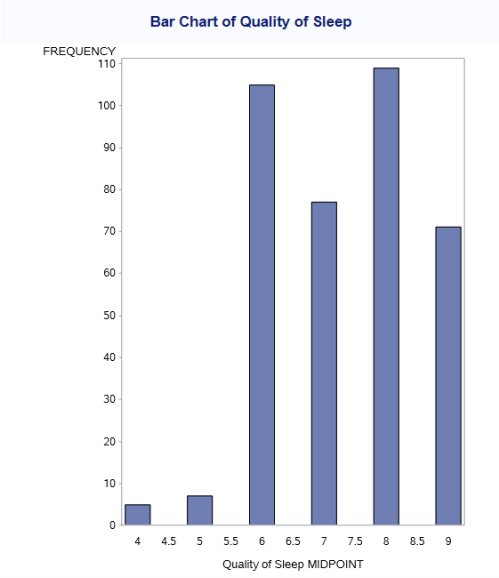
2. Descriptive Analysis



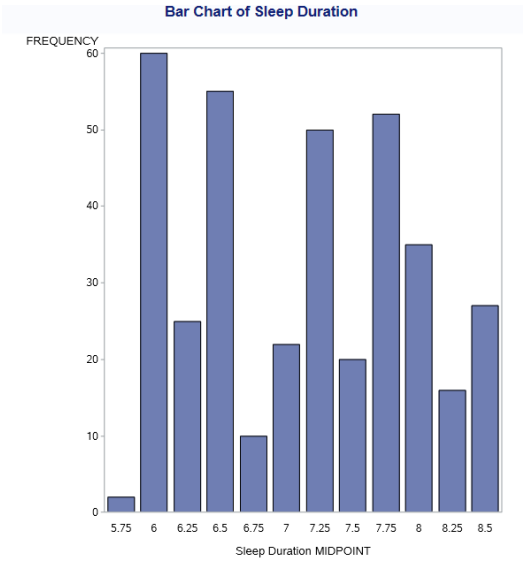
From the dataset, there are a total of 374 adults, with 185 females and 189 males.



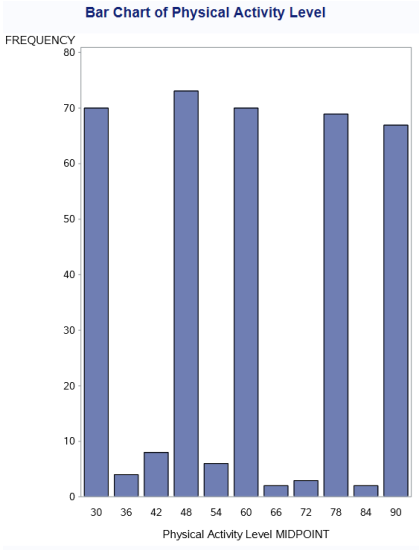
The pie chart shows BMI categories (normal, normal weight, obese, overweight) where 52.14% of adults have a normal BMI, 39.57% are overweight, and the fewest are obese. Most adults have a healthy weight due to a normal BMI, but many adults are overweight that may show health concerns.



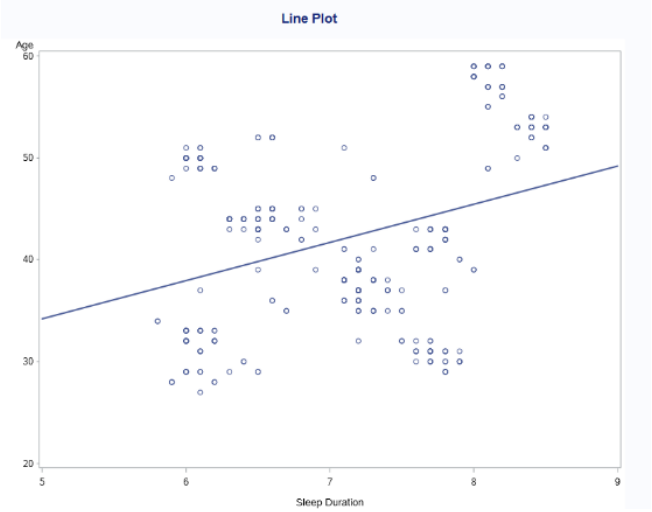
The bar chart shows the sleep quality of adults in the scale of between 1 and 10. Most of the adults have a higher value of sleep quality which is 8 and the second most is 6. There are very few adults who have a sleep quality of 4.



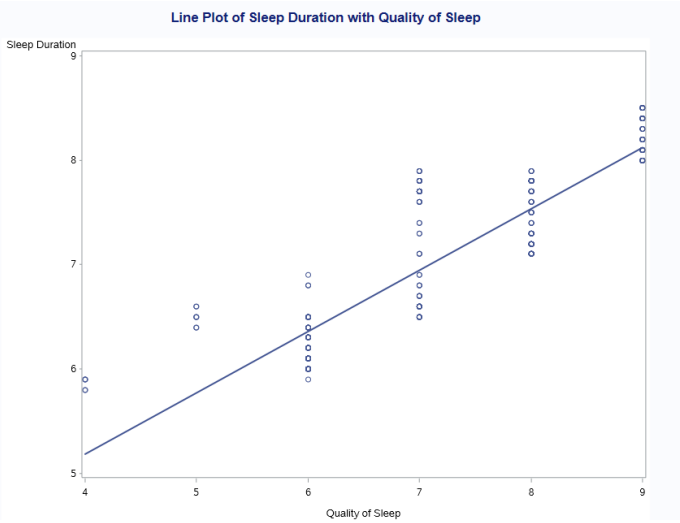
The bar chart shows the sleep duration in hours of adults. Most of the adults' sleep around 6 hours to 6.25 hours a day, followed by duration around 7.5 hours, and a few of them sleep around less than 6 hours or more than 8 hours a day.



The bar chart shows the number of minutes the adults engaged in physical activity daily. Most adults have engaged in physical activity around 30, 48, 54, 60, and 84 minutes per day. There are a few of them engaged in physical activity below 30 or between 66 and 90 minutes a day.



The line plot shows the best fit line between age and sleep duration, where the horizontal axis represents the sleep duration, and the vertical axis represents the age. The best fit line shows a slightly positive correlation between age and sleep duration. When the age of adults increases, the sleep duration will also increase.



The line plot shows the best fit line between quality of sleep and sleep duration, where the horizontal axis represents the quality of sleep, and the vertical axis represents the sleep duration. The best fit line shows a positive correlation between quality of sleep and sleep duration. Therefore, the higher the sleep duration, and higher the quality of sleep.

Summary Statistics
Results

The MEANS Procedure

Variable	Mean	Std Dev	Minimum	Maximum	N	Median
Sleep Duration	7.1320856	0.7956567	5.8000000	8.5000000	374	7.2000000
Quality of Sleep	7.3128342	1.1969559	4.0000000	9.0000000	374	7.0000000
Physical Activity Level	59.1711230	20.8308037	30.0000000	90.0000000	374	60.0000000

The summary statistics shows the minimum, maximum, mean, median and standard deviation of sleep duration, quality of sleep and physical activity level.

Mean of sleep duration shows that 374 of the adults slept an average of about 7.13 hours a day in this dataset.

Mean of quality of sleep shows that 374 of the adults have an average sleep quality of about 7.31 based on the scale of between 1 and 10.

Mean of physical activity level shows that the adults engaged in physical activity in average of about 59.17 minutes a day.

Sleep duration and quality of sleep have low standard deviation. It shows the data points are closer to the mean that shows a low variability. Low variability shows there is little spread among the values, and they do not differ much from each other.

Physical activity level has the highest standard deviation. It shows that the data points are spread out over a wide range of values and are far from the mean. It also has a high variability showing that there is a wider spread among the values.

3. Statistical Inference and Modelling

One – Sample Hypothesis (To achieve objective no.1)

Reference: (Link: Reported sleep duration reveals segmentation of the adult life-course into three phases | Nature Communications)

This article indicates that the average sleep duration recommended for adults is 7 hours.

From the dataset:

t Test (One-Sample) where the mean average is 7

The TTEST Procedure

Variable: Sleep Duration

N	Mean	Std Dev	Std Err	Minimum	Maximum
374	7.1321	0.7957	0.0411	5.8000	8.5000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
7.1321	7.0512	7.2130	0.7957	0.7424	0.8572

DF	t Value	Pr > |t|
373	3.21	0.0014

H_0 = Mean average of sleep duration is less or equal to 7 hours

H_1 = Mean average of sleep duration is more than 7 hours

$$\bar{x} = \frac{\sum x_i}{n}$$

$$= \frac{2667.4}{374}$$

$$= 7.1321$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

$$= \sqrt{\frac{236.135}{373}}$$

$$= 0.7957$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$= \frac{7.1321 - 7}{0.7957/\sqrt{374}}$$

$$= 3.2106$$

$$\alpha = 0.05$$

$$p - \text{value} = 1 - 0.99934 = 0.00066$$

∴ p-value is less than 0.05, hence null hypothesis is rejected. There is sufficient evidence to support that the mean average of sleep duration is more than 7 hours.

∴ The dataset analysis shows that the mean sleep duration is 7.1321 hours, slightly more than the recommended 7 hours, indicating that average sleep duration exceeds the recommendation.

Two – Sample Hypothesis (To achieve objective no. 2)

H_0 = Mean average of quality of sleep is the same for both gender

H_1 = Mean average of quality of sleep is different for both gender

t Test

The TTEST Procedure

Variable: Quality of Sleep

Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		185	7.6649	1.2795	0.0941	4.0000	9.0000
Male		189	6.9683	0.9995	0.0727	4.0000	9.0000
Diff (1-2)	Pooled		0.6966	1.1466	0.1186		
Diff (1-2)	Satterthwaite		0.6966		0.1189		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		7.6649	7.4793	7.8505	1.2795	1.1610	1.4250
Male		6.9683	6.8248	7.1117	0.9995	0.9079	1.1119
Diff (1-2)	Pooled	0.6966	0.4634	0.9298	1.1466	1.0698	1.2353
Diff (1-2)	Satterthwaite	0.6966	0.4628	0.9304			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	372	5.87	<.0001
Satterthwaite	Unequal	347.96	5.86	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	184	188	1.64	0.0008

For Female:

$$\bar{x}_1 = \frac{\sum x_i}{n}$$

$$= \frac{1418}{185}$$

$$= 7.6649$$

$$s_1 = \sqrt{\frac{\sum(x_i - \bar{x}_1)^2}{n - 1}}$$

$$= \sqrt{\frac{301.2216}{184}}$$

$$= 1.2795$$

For Male:

$$\bar{x}_2 = \frac{\sum x_i}{n}$$

$$= \frac{1317}{189}$$

$$= 6.9683$$

$$s_2 = \sqrt{\frac{\sum(x_i - \bar{x}_2)^2}{n - 1}}$$

$$= \sqrt{\frac{187.8095}{188}}$$

$$= 0.9995$$

Using non-pooled t-test:

$$t_0 = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$= \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{({s_1}^2/n_1) + ({s_2}^2/n_2)}}$$

$$= \frac{(7.6649 - 6.9683)}{\sqrt{(1.2795/185) + (0.9995/189)}}$$

$$= \frac{0.6966}{0.1105}$$

$$= 6.3041$$

$$p - \text{value} = < 0.0001$$

∴ Assuming the significance level is 0.05, then the p-value is < 0.05, hence the null hypothesis is rejected. There is sufficient evidence to conclude that the mean average of the quality of sleep is different for both genders.

One-Way ANOVA (To achieve objective no.3)

To verify the assumption of equal variances, first conduct Levene's Test for homogeneity of variance.

Linear Models

The GLM Procedure

Levene's Test for Homogeneity of Sleep Duration Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
BMI Category	2	0.5527	0.2763	0.64	0.5289
Error	371	160.7	0.4330		

H_0: $\sigma^2_1 = \sigma^2_2 = \sigma^2_3$

H_1: Not all population variances are equal.

∴ At the 5% significance level, the p-value is 0.5289 which is greater than 0.05. Thus, H_0 is not rejected. It indicates that the variances of the variable under consideration are the same for each treatment. The assumption of homogeneity of variance is met.

First Model (Without blocking factor)

Linear Models

The GLM Procedure

Dependent Variable: Sleep Duration

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	33.8130804	16.9065402	31.00	<.0001
Error	371	202.3218929	0.5453420		
Corrected Total	373	236.1349733			

R-Square	Coeff Var	Root MSE	Sleep Duration Mean
0.143194	10.35423	0.738473	7.132086

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BMI Category	2	33.81308037	16.90654018	31.00	<.0001

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: At least one BMI category has a mean sleep duration that is different from the others.

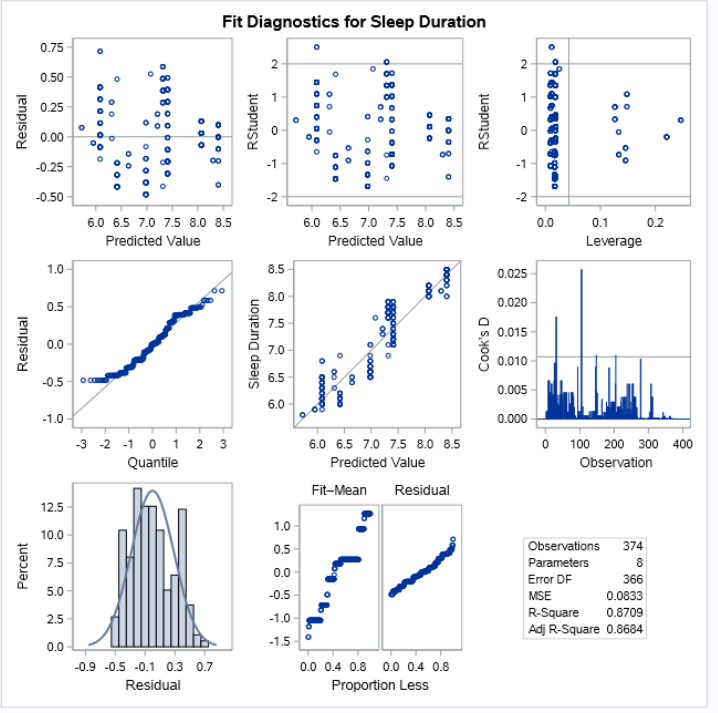
The p-value is less than 0.0001. Thus, H_0 is rejected. This indicates that there are significant differences between the means of sleep duration across BMI categories.

However, the R-square value is low (R-square = 0.143194), with high Root MSE value (Root MSE=0.738473) indicating this model does not fit the dataset well. Therefore, the model is improved by adding a block design.

Randomised Block Design

The purpose of using randomised block design is to compare the means of Sleep Duration across different BMI Categories while controlling for variability due to Quality of Sleep as a block factor.

To evaluate the assumption of normality, we examine the distribution of Q-Q Plot.



The points on the Q-Q plot follow the 45-degree line reasonably well, indicating that the residuals are approximately normally distributed. Thus, the assumption of normality is met.

Output:

Linear Models

The GLM Procedure

Class Level Information		
Class	Levels	Values
Quality of Sleep	6	4 5 6 7 8 9
BMI Category	3	Normal Obese Overweight

Number of Observations Read	374
Number of Observations Used	374

Linear Models

The GLM Procedure

Dependent Variable: Sleep Duration

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	205.6442739	29.3777534	352.64	<.0001
Error	366	30.4906994	0.0833079		
Corrected Total	373	236.1349733			

R-Square	Coeff Var	Root MSE	Sleep Duration Mean
0.870876	4.046939	0.288631	7.132086

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BMI Category	2	6.8272944	3.4136472	40.98	<.0001
Quality of Sleep	5	171.8311935	34.3662387	412.52	<.0001

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: At least one BMI category has a mean sleep duration that is different from the others.

Including Quality of Sleep in the model increases the R-Square and lowers the Root MSE, indicating improved variance explanation and prediction accuracy.

With a p-value less than 0.0001, H_0 is rejected, showing significant differences in sleep duration across BMI categories and blocks (Quality of Sleep). Both factors significantly affect sleep duration (p-values < 0.001), necessitating a post-hoc test to identify specific differences between means.

Post-Hoc: Tukey's Multiple Comparison

Linear Models

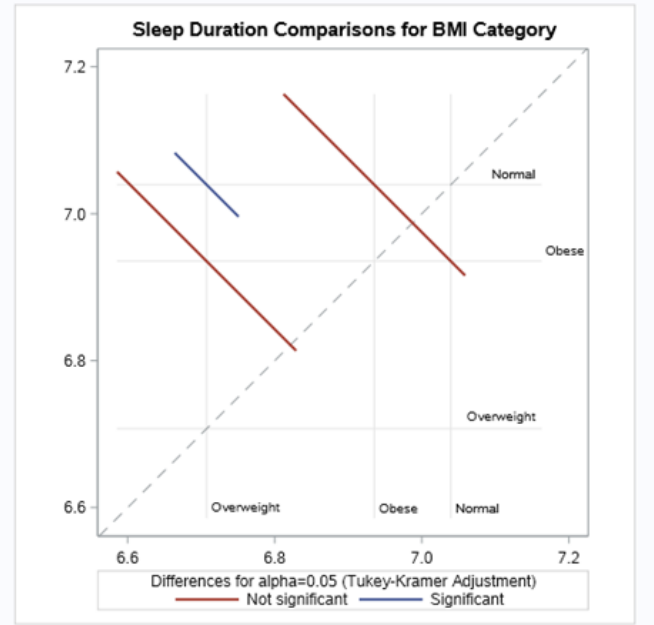
The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

BMI Category	Sleep Duration LSMEAN	LSMEAN Number
Normal	7.03946538	1
Obese	6.93537474	2
Overweight	6.70736563	3

Least Squares Means for effect BMI Category
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Sleep Duration

i/j	1	2	3
1		0.5825	<.0001
2	0.5825		0.0721
3	<.0001	0.0721	



The first part of the output shows the mean sleep duration for each BMI category. The second part of the output shows p-values from pairwise comparisons of all possible combinations of means.

The output result and the diffogram shows that the only significant pairwise difference is between Normal and Overweight with p-value less than 0.0001. There is a statistically significant difference in sleep duration between normal weight and overweight individuals, with normal weight individuals sleeping longer on average.

Multiple Linear Regression (To achieve objective no.4)

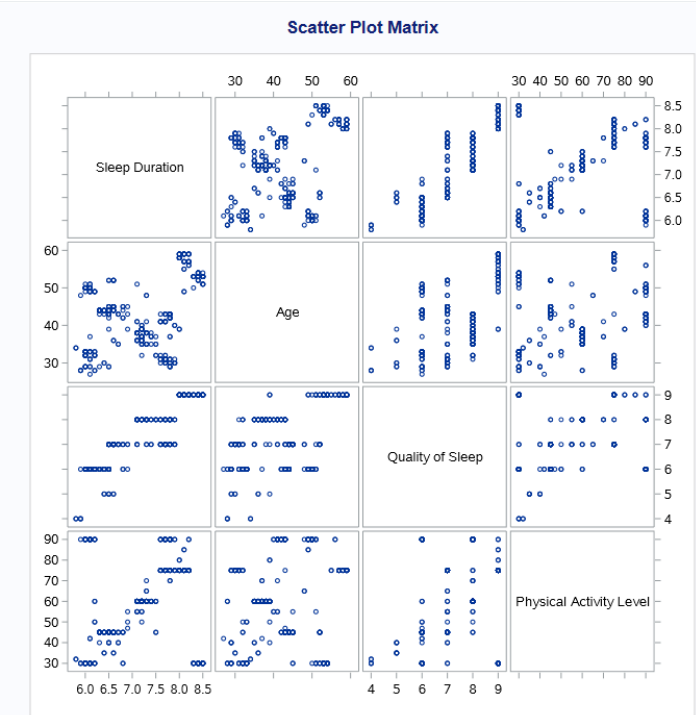
We would like to explore the factors influencing sleep duration and develop a predictive model based on various lifestyle and health indicators, aiming to identify key determinants of sleep duration.

Dependent Variable: Sleep Duration (hours)

Independent Variables: Quality of Sleep (X_1), Age (X_2), Physical activity level (X_3)

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Before conducting regression analysis, it is important to examine the relationships between the variables using scatter plot matrix and Pearson correlation coefficient to ensure they exhibit a linear relationship.



3 With Variables:	Physical Activity Level Age Quality of Sleep
1 Variables:	Sleep Duration

Pearson Correlation Coefficients, N = 374	
	Sleep Duration
Physical Activity Level	0.21236
Age	0.34471
Quality of Sleep	0.88321

The scatter plots reveal different levels of correlation between Sleep Duration and the independent variables. Sleep Duration shows a slight linear trend with Age, with a moderate positive correlation (r = 0.34471). A more obvious positive linear trend is seen with Quality of Sleep, reflected in a strong positive correlation (r = 0.88321). In contrast, the relationship with Physical Activity Level is weaker, showing a mild linear trend and a weak positive correlation (r = 0.21236).

Output:

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: Sleep Duration

Number of Observations Read	374
Number of Observations Used	374

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	186.48176	62.16059	463.20	<.0001
Error	370	49.65321	0.13420		
Corrected Total	373	236.13497			

Root MSE	0.36633	R-Square	0.7897
Dependent Mean	7.13209	Adj R-Sq	0.7880
Coeff Var	5.13637		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation	95% Confidence Limits	
Intercept	1	2.92714	0.12795	22.88	<.0001	0	2.67555	3.17873
Age	1	-0.00926	0.00250	-3.71	0.0002	1.30276	-0.01417	-0.00436
Physical Activity Level	1	0.00202	0.00093280	2.16	0.0312	1.04943	0.00018266	0.00385
Quality of Sleep	1	0.61213	0.01814	33.75	<.0001	1.30975	0.57647	0.64779

Linear Regression Equation:

$\hat{y} = 2.92714 + 0.61213X_1 - 0.00926X_2 + 0.00202X_3$

R-square (0.7897) indicates that 78.97% of the variation in Sleep Duration can be explained by the variability in Age, Physical Activity Level and Quality of Sleep, while 21.03% of the sample variability in sleep duration can't be explained by those variables.

Adj R-square (0.7880) indicates that 78.8% of the variation in sleep duration is explained by the multiple regression model while being adjusted for the number of independent variables and sample size.

Inference on Collective Influence:

(β_1 = Quality of Sleep; β_2 = Age; β_3 = Physical Activity Level)

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_1: At least one $\beta_j \neq 0$

F-value = 463.20

p-value = <0.0001

Since the p-value is less than 0.05 at the 0.05 significance level, the H_0 is rejected. Therefore, there is a linear relationship between the dependent variable and at least one of the independent variables. It means that at least one of the explanatory variables has a significant effect on the response variable.

Inference for Individual Regression Coefficients:

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

t-value = 33.75

p-value = <0.0001

Since the p-value is less than 0.05 at the 0.05 significance level, the H_0 is rejected. Therefore, there is strong evidence showing that quality of sleep is related to sleep duration, controlling for age and physical activity level. The 95% confidence interval indicates that for every one-scale increase in quality of sleep, the predicted sleep duration is estimated to increase between 0.58 hours (34.8 minutes) and 0.65 hours (39 minutes).

H_0: $\beta_2 = 0$

H_1: $\beta_2 \neq 0$

t-value = -3.72

p-value = 0.0002

Since the p-value is less than 0.05 at the 0.05 significance level, the H_0 is rejected. Therefore, there is strong evidence showing that age is related to sleep duration, controlling for quality of sleep and physical activity level. The 95% confidence interval indicates that for every one-year increase in Age, the predicted sleep duration is estimated to decrease between 0.004 hours (14.4 seconds) and 0.01 hours (36 seconds).

H_0: $\beta_3 = 0$

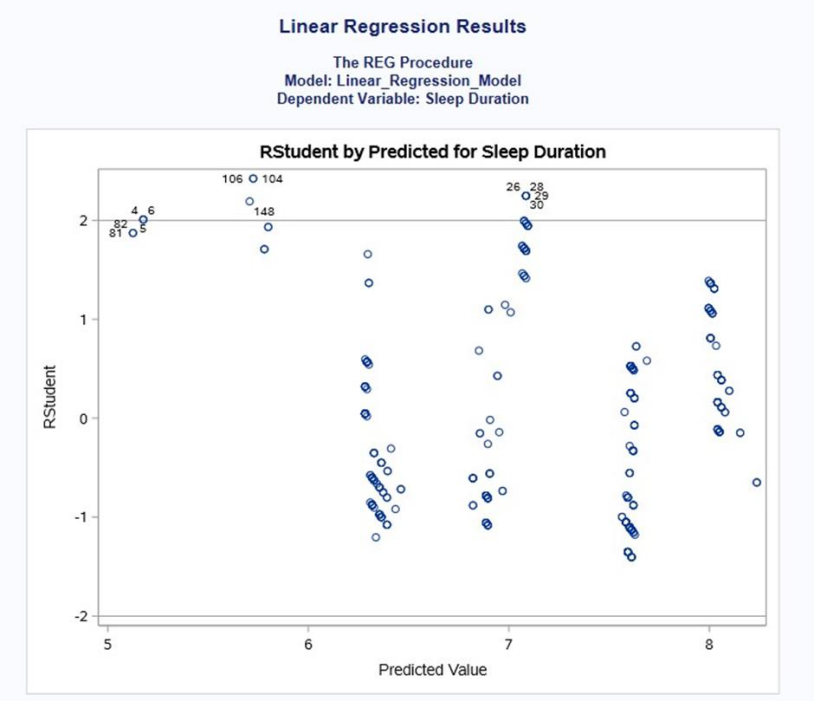
H_1: $\beta_3 \neq 0$

t-value = 2.16

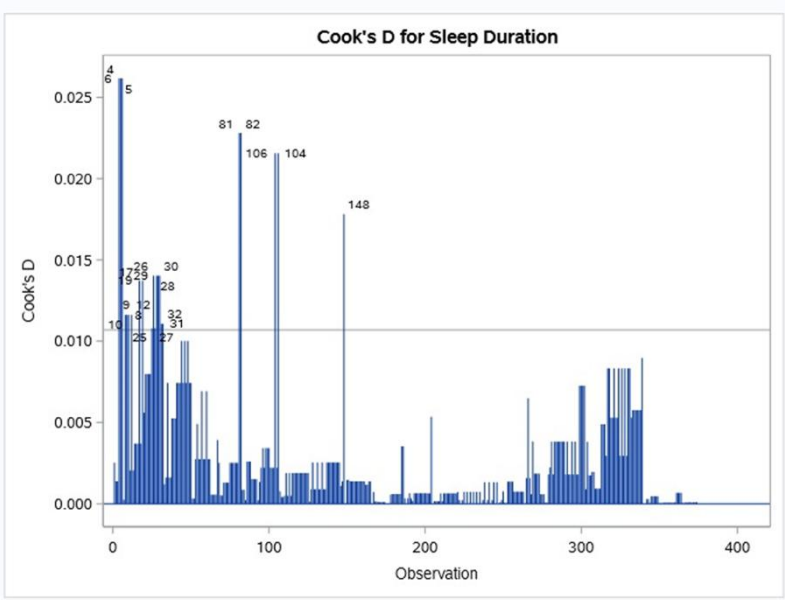
p-value = 0.0312

Since the p-value is less than 0.05 at the 0.05 significance level, the H_0 is rejected. Therefore, there is strong evidence that physical activity level is related to sleep duration, controlling for age and quality of sleep. The 95% confidence interval indicates that for every one-minute increase in physical activity level, the predicted sleep duration is estimated to increase between 0 and 0.004 hours (14.4 seconds).

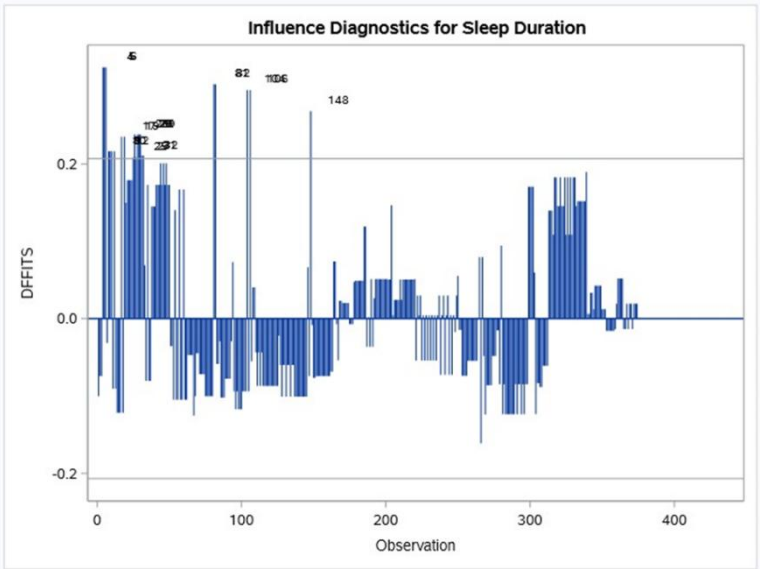
Regression Diagnostic



The RStudent by Predicted Value plot shows a number of observations beyond 2 standard errors from the mean of 0.



The Cook's D plot shows observations 4, 5, and 6 have values above 0.025, indicating they may be highly influential.



The DFFITS plot highlights observations 4, 5, 6, 81, 82, 106, 104, and 148 as potentially influential, with values significantly outside the typical range.

4. Conclusion

Our statistical analysis uncovered several key insights aligned with the study's objectives. The one-sample hypothesis test showed that the average sleep duration in our dataset exceeds 7 hours, differing from recommended guidelines and suggesting variability in sleep needs or behaviours. Gender differences in sleep quality were confirmed through the two-sample test, emphasizing the necessity for gender-specific interventions. The one-way ANOVA and subsequent post-hoc analysis indicated that BMI significantly affects sleep duration, particularly between normal weight and overweight individuals, pointing to a connection between body composition and sleep patterns. The multiple regression analysis highlighted Quality of Sleep as a crucial factor influencing Sleep Duration, with Age and Physical Activity also playing roles, albeit to a lesser extent.

Future research should explore gender-specific sleep interventions, the impact of BMI on sleep, and the development of comprehensive strategies that enhance both sleep quality and duration, potentially through long-term studies to assess the effectiveness of these approaches over time.

(2467 words)

(Excluding front page and references)

References

Coutrot, A., Lazar, A. S., Richards, M., Manley, E., Wiener, J. M., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2022). Reported sleep duration reveals segmentation of the adult life-course into three phases. Nature News. https://www.nature.com/articles/s41467-022-34624-8

Tharmalingam, L. (2023). Sleep health and lifestyle dataset. Kaggle. https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset