Statistical Analysis of Sleep Health and Lifestyle

Student Name: *Namitha Bhaskar*

RIT Email ID: nb6425@rit.edu

Course: STAT 614 : Applied Statistics
Instructor: Carly Metcalfe
Date: 18 December 2023

1. INTRODUCTION

The objective of this study is to assess if sleep is related to various factors like physical activity, daily steps, BMI, gender, and stress level, and recommend actions based on these results.

The research questions of the study are as follows :

Aim 1 : Is the mean Quality of sleep different for males and females?

Null hypothesis $H_0: \mu_{Male} = \mu_{Female}$

Alternative hypothesis $H_1: \mu_{Male} \neq \mu_{Female}$

Aim 2 : Is there a significant difference in the mean sleep duration for the people in the 3 BMI categories (underweight, normal, and overweight)?

Null hypothesis $H_0: \mu_{underweight} = \mu_{normal} = \mu_{overweight}$

Alternative hypothesis $H_1: \mu_{underweight} \neq \mu_{normal} \neq \mu_{overweight}$

Aim 3 : Is there a relationship between the Duration of Sleep with Physical Activity and Daily Step Count?

Null hypothesis $H_0: \beta_{Physical\ Activity\ level} = \beta_{Daily\ steps} = 0$

Alternative hypothesis $H_1: \beta_{Physical\ activity\ level} \neq \beta_{Daily\ steps} \neq 0$

Aim 4 : Does the type of sleep disorder vary by BMI categories?

Null hypothesis H_0: There is no relationship between sleep disorder and BMI category

Alternative hypothesis H_1: There is a relationship between sleep disorder and BMI category

2. METHODS

The dataset chosen is a sleep and lifestyle questionnaire which analyzes the sleep pattern, health statistics of 374 men and women. This dataset focuses on Comprehensive sleep metrics, lifestyle factors, cardiovascular health and sleep disorder analysis. It has a total of 13 variables ranging from gender, age, occupation to health parameters like heart rate and blood pressure. The variables used in this analysis are categorical and continuous variables. They include gender (binomial - male / female), sleep duration (continuous - hours), quality of sleep (continuous - scale of 1-10), physical activity level (continuous - minutes), BMI category (categorical - underweight / normal / overweight), daily steps (continuous - steps), and sleep disorder (categorical - normal, insomnia, sleep apnea)

Statistical analysis for Aim 1 is done using the Two sample t-test for means. Additionally, To determine equal or unequal variance, the F-Test for variance is performed. The response variable is the quality of sleep and the predictor variable is gender. For Aim 2, One Way Analysis of Variance is done with the response variable as sleep duration and the predictor variable as the BMI category. Aim 3 is analyzed by Multiple Linear Regression with the response variable as Duration of sleep and predictor variables as physical activity level and daily step count of an individual. A test of significance for the regression slope will also be performed. Aim 4 will be carried out using the Chi-Square

statistical test with categorical variables sleep disorder and BMI category. All tests of significance for this study will be conducted using an alpha level of 0.05 ($\alpha = 0.05$)

3. RESULTS

In this dataset of 374 people, there are 185 females and 189 males

Test	F Ratio	DFNum	DFDen	p-value
F-Test 2 sided	1.6387	184	188	0.0008

Table 1 : Test of Variance for quality of sleep by gender

Table 1 represents the result of F -test for two sample variance with Null Hypothesis $\sigma^2_{Male} = \sigma^2_{Female}$ and Alternate Hypothesis $\sigma^2_{Male} \neq \sigma^2_{Female}$. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that the variance is unequal. Therefore the two sample t-test for means is done assuming unequal variances.

Difference	-0.69661
Std Err Dif	0.11889
Upper CL Dif	-0.46278
Lower CL Dif	-0.93044
Confidence	0.95
t Ratio	-5.85932
Prob > |t|	< 0.0001

Table 2: two sample t-test for means for quality of sleep by gender

Table 2 represents the result of two sample t-test for means assuming unequal variance. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there is sufficient evidence to claim that the quality of sleep is different between males and females.

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob>F
BMI Category	2	33.81308	16.9065	31.0017	< 0.0001
Error	371	202.32189	0.5453		
C.Total	373	236.13497			

Table 3: ANOVA F-Test for sleep duration by BMI Category

Table 3 gives the result of One Way ANOVA test for sleep duration based on BMI categories. Since the p-value is less than 0.05, the null hypothesis is rejected. Hence we conclude that there is a significant difference in the mean duration of sleep for people in the 3 BMI categories. People who fall under the normal BMI category, have the maximum sleep duration followed by overweight people and lastly obese BMI category people. The statistical software output in the appendix shows plot and the Tukey analysis for all pairs of data.

Term	Estimate	t-Ratio	Prob > |t|
Intercept	7.4637	44.70	< 0.0001
Physical Activity Level	0.0230	7.99	< 0.0001
Daily steps	-0.0002	-6.69	< 0.0001

Table 4: Parameter Estimates for physical activity and daily steps

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob>F
Model	2	34.95085	17.4754	32.2261	< 0.0001
Error	371	201.18413	0.5423		
C.Total	373	236.13497			

Table 5: Response for sleep duration by physical activity level and daily steps

Table 4 gives the result of the Multiple Linear Regression for sleep duration based on the daily steps and physical activity level. Based on the results, equation of the regression model is:

$$\hat{y} = 7.4367 + 0.0230 * physicalActivitylevel - 0.0002 * dailysteps$$

Table 5, reports the significance test for the regression model where the p-value of the F-Test is less than 0.05, therefore we reject the null hypothesis and conclude that there is a relationship between sleep duration and physical activity level and daily steps. From the model we can see that for every one minute increase in the physical activity level, there is 0.0230 increase in sleep duration and for every one step increase of daily step count, there is 0.0002 decrease in the sleep duration. The highest sleep duration is for people who have a physical activity time of 90 minutes and who have a daily step count of 3000 to 4000 steps. This can be seen from the cube plot in Fig.a8 and the prediction profiler plot in Fig.a9(in appendix). The residual VS predicted plot in Fig.a6 does not show much of a scatter, which might suggest some discrepancy with the data. The quantile plot for the residuals shown in Fig.a7 seem to be normal despite few outlier points.

Test	ChiSquare	Prob > ChiSq
Pearson	245.665	< 0.0001

Table 5: Response for Sleep disorder based on BMI

Table 5 provides the result of the Chi square test on sleep disorder based on the BMI category. Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there is a relationship between sleep disorder and the BMI.

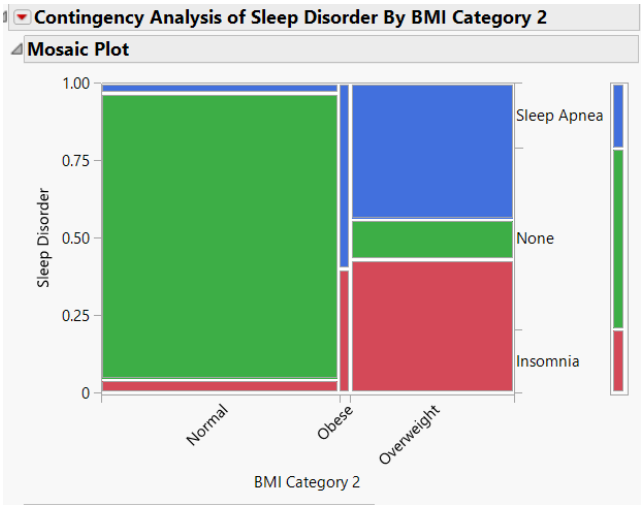


Fig. 1: Mosaic plot for chi square test on sleep disorder

Based on the mosaic plot in Fig 1, we can see that the majority of people under normal BMI category have no sleep disorder. Most of the people under the overweight BMI category suffer from either sleep apnea or insomnia. A very small portion of people in this category have no sleep disorders. Majority of people in the obese BMI category suffer from sleep apnea. Based on the plot, people who are obese have sleep disorder.

4. CONCLUSION

Based on the study results, it is clear that sleep quality and sleep duration is associated with factors like gender, physical activity, daily steps, and BMI. To increase the duration of sleep for people in the overweight and obese BMI category, they have to start reducing their weight. This will also help in the reduction of sleep disorders that arise based on the BMI category. Physical activity plays a major role in the duration of sleep. People with more minutes of physical activity per day tend to sleep more hours. These recommendations will help achieve better quality of sleep and higher duration of uninterrupted sleep. It is recommended that future studies should collect more data on obese people to get more accurate results for the study.

APPENDIX

Oneway Analysis of Quality of Sleep By Gender

Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Female	185	1.279482	1.126370	1.081081
Male	189	0.999493	0.853448	0.846561

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	16.3700	1	372	<.0001*
Brown-Forsythe	12.0807	1	372	0.0006*
Levene	22.2059	1	372	<.0001*
Bartlett	11.2189	1	.	0.0008*
F Test 2-sided	1.6387	184	188	0.0008*

Fig. a1: Result of F-Test for 2 sample variance

Oneway Analysis of Quality of Sleep By Gender

t Test

Male-Female

Assuming unequal variances

Difference	-0.69661	t Ratio	-5.85932
Std Err Dif	0.11889	DF	347.9558
Upper CL Dif	-0.46278	Prob > |t|	<.0001*
Lower CL Dif	-0.93044	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*

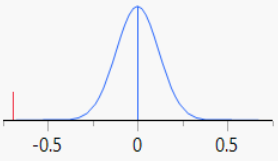
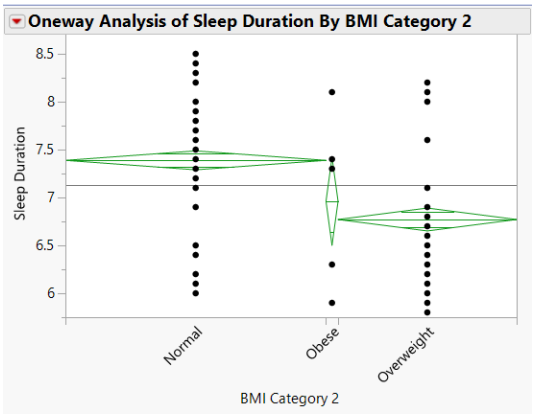


Fig. a2: Result of t-Test for 2 means



Oneway Anova

Summary of Fit

Rsquare	0.143194
Adj Rsquare	0.138575
Root Mean Square Error	0.738473
Mean of Response	7.132086
Observations (or Sum Wgts)	374

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
BMI Category 2	2	33.81308	16.9065	31.0017	<.0001*
Error	371	202.32189	0.5453		
C. Total	373	236.13497			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Normal	216	7.38796	0.05025	7.2892	7.4868
Obese	10	6.96000	0.23353	6.5008	7.4192
Overweight	148	6.77027	0.06070	6.6509	6.8896

Std Error uses a pooled estimate of error variance

Fig. a3: Result if One Way ANOVA

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

HSD Threshold Matrix

Connecting Letters Report

Level		Mean
Normal	A	7.3879630
Obese	A B	6.9600000
Overweight	B	6.7702703

Levels not connected by same letter are significantly different.

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Normal	Overweight	0.6176927	0.0788002	0.432263	0.8031221	<.0001*
Normal	Obese	0.4279630	0.2388701	-0.134136	0.9900622	0.1738
Obese	Overweight	0.1897297	0.2412860	-0.378055	0.7575140	0.7117

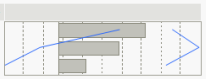


Fig. a4: Comparison of means by Tukey test

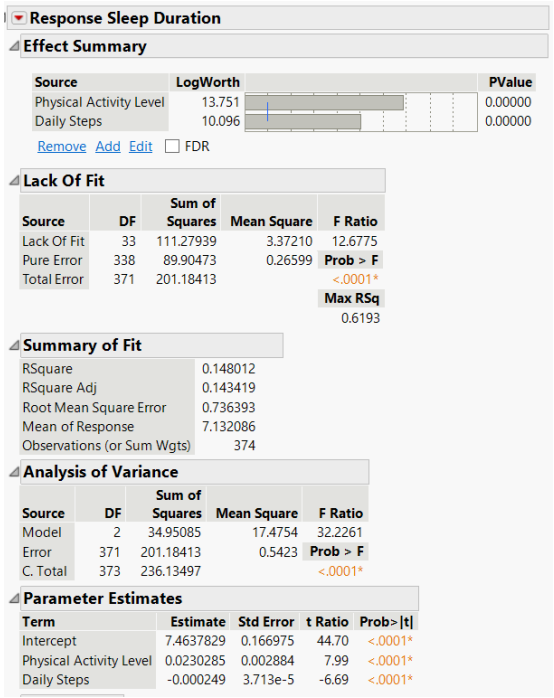


Fig. a5: Result of multiple linear regression of sleep duration

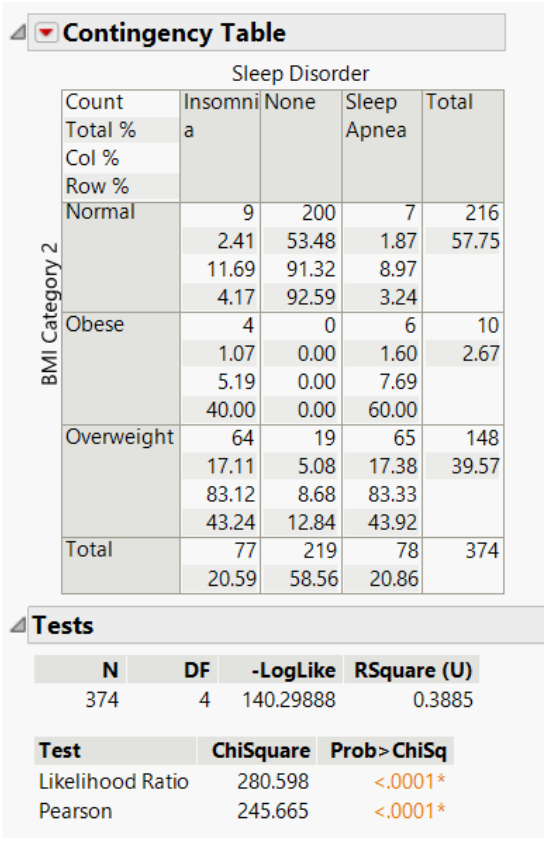


Fig. a6: Chi square test for sleep disorder by BMI category

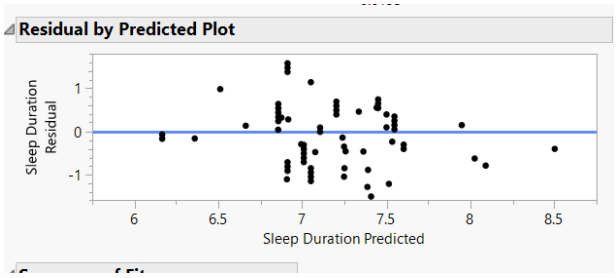


Fig. a6: residual by predicted plot for regression

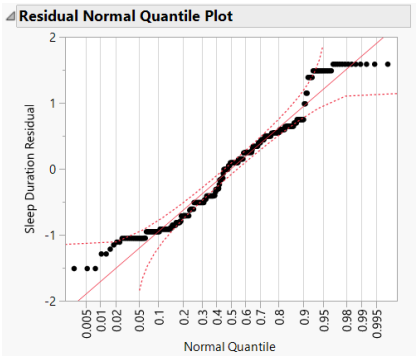


Fig. a7: Normal quantile plot for residuals

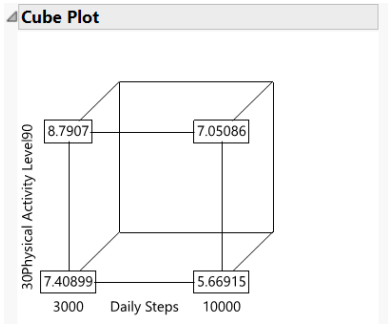


Fig. a8: Cube Plot for sleep duration

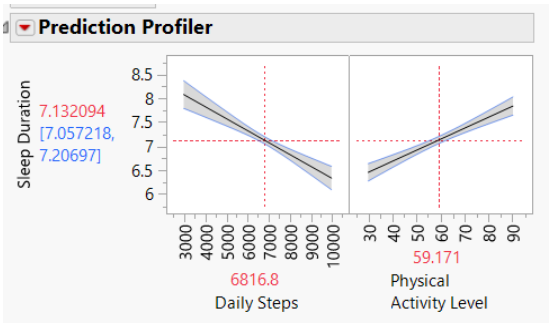


Fig. a9: prediction profiler for sleep duration