SLEEP DISORDER

Model & Analysis

Prayogi Adista
Statistics for Business

Table of Contents

1. Introduction

The development of modern society has caused common problems in sleep over the past few decades, namely sleep disorders. Sleep disorders involve problem with the duration and sleep quality (Wang et al., 2022)

Mitigating sleep disorders is important because it can improve people's physical health, mental health, cognitive function, and overall well-being. Thus, knowing variables that affect sleep problems can give us insights to improve our sleep quality so that people can be more effective in work and daily activities.

In this project, a statistical approach is used to understand parameters that influence individual's sleep disorders.

1.1 Definition of Concepts

In order to have a proper understanding of the report, we have defined some terminologies which are imperative to this cause.

Sleep disorders are a group of medical conditions that affect a person's ability to get sufficient, restful, and regular sleep. These disorders can disrupt a person's sleep patterns, leading to various physical, emotional, and cognitive problems.

BMI stands for Body Mass Index, and it is a numerical measure of a person's body weight in relation to their height.

Confirmatory analysis, also known as confirmatory data analysis or confirmatory research, is a research approach and statistical analysis method used to test specific hypotheses or theories in a structured and predetermined manner.

Logistic regression is a statistical method used for analyzing datasets where the outcome or dependent variable is categorical. It is particularly well-suited for binary classification problems, where the goal is to predict one of two possible outcomes, such as yes/no, 1/0, or true/false.

1.2 Objectives

The objectives of this project are:

1. To investigate variables that affect sleep disorders
2. To create and interpret a model that can predict sleep disorders

2. Data Sources

The Sleep Health and Lifestyle Dataset comprises 400 rows and 13 columns, covering a wide range of variables related to sleep and daily habits. It includes details such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders.

Key Features of the Dataset:

- Comprehensive Sleep Metrics: Explore sleep duration, quality, and factors influencing sleep patterns.

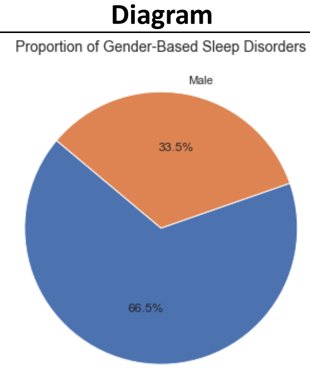
- Lifestyle Factors: Analyze physical activity levels, stress levels, and BMI categories.
- Cardiovascular Health: Examine blood pressure and heart rate measurements.
- Sleep Disorder Analysis: Identify the occurrence of sleep disorders such as Insomnia and Sleep Apnea.

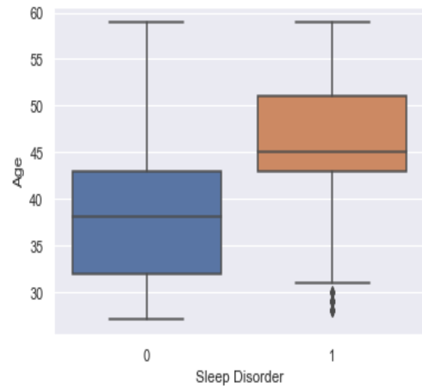
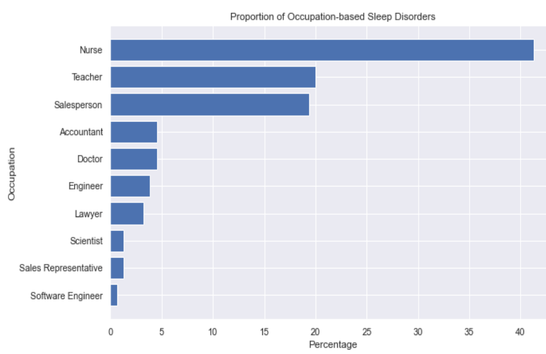
Table 2.1 The description of the dataset columns.

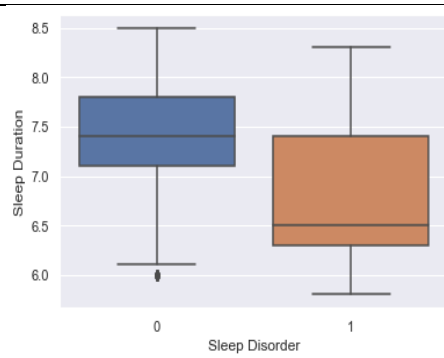
No	**Column**	**Description**
1	Person ID	An identifier for each individual.
2	Gender	The gender of the person (Male/Female).
3	Age	The age of the person in years.
4	Occupation	The occupation or profession of the person.
5	Sleep Duration	The number of hours the person sleeps per day.
6	Quality of Sleep	A subjective rating of the quality of sleep, ranging from 1 to 10.
7	Physical Activity Level	The number of minutes the person engages in physical activity daily (minutes/day).
8	Stress Level	A subjective rating of the stress level experienced by the person, ranging from 1 to 10.
9	BMI Category	The BMI category of the person (e.g., Underweight, Normal, Overweight).
10	Blood Pressure	The blood pressure measurement of the person, indicated as systolic pressure over diastolic pressure (systolic/diastolic)
11	Heart Rate	The resting heart rate of the person in beats per minute (bpm).
12	Daily Steps	The number of steps the person takes per day.
13	Sleep Disorder	The presence or absence of a sleep disorder in the person (None, Insomnia, Sleep Apnea). • None: The individual does not exhibit any specific sleep disorder. • Insomnia: The individual experiences difficulty falling asleep or staying asleep, leading to inadequate or poor-quality sleep. • Sleep Apnea: The individual suffers from pauses in breathing during sleep, resulting in disrupted sleep patterns and potential health risks.

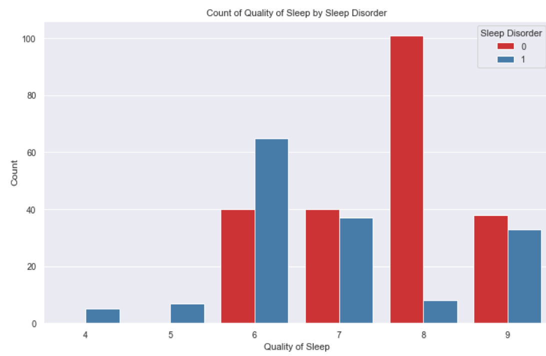
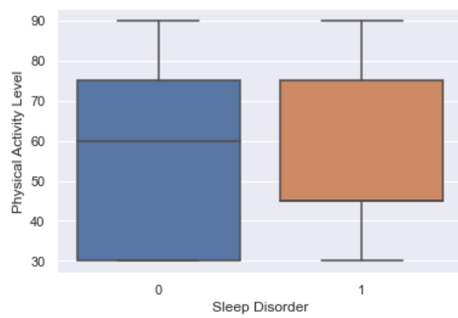
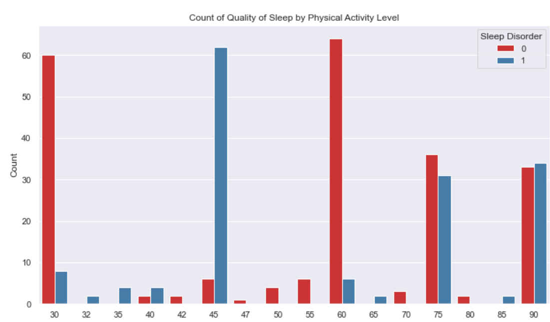
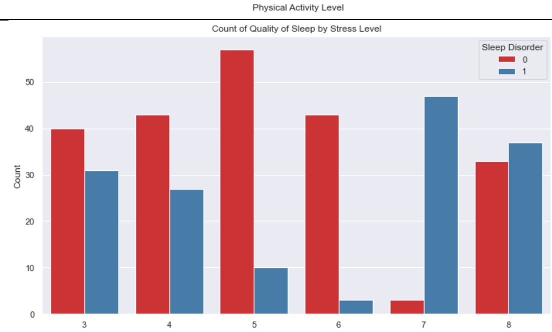
3. Exploratory Data Analysis

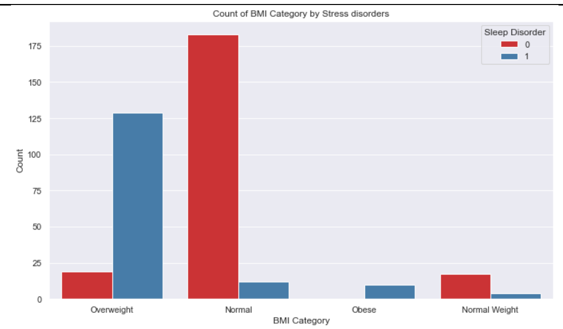
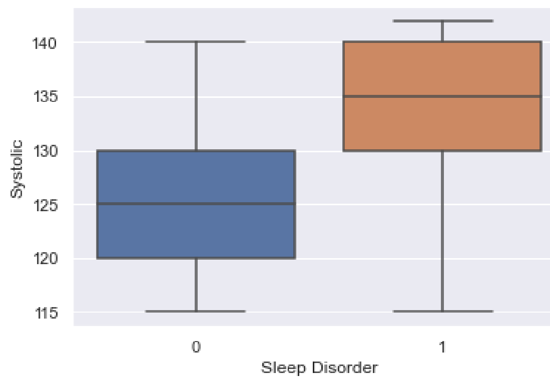
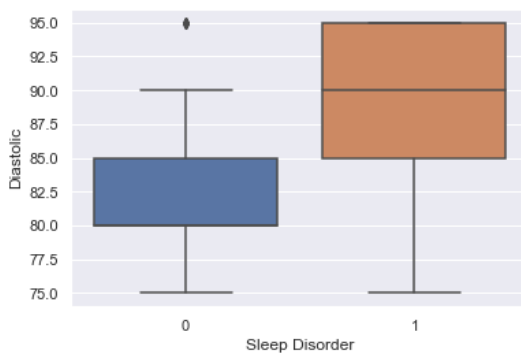
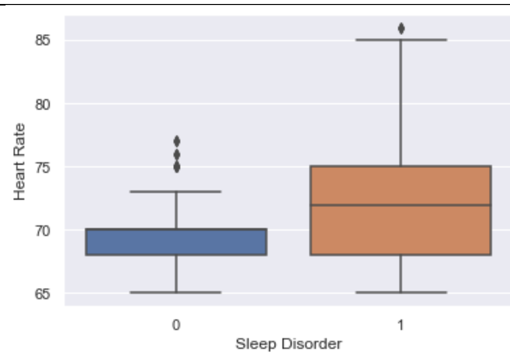
Here are the findings of each variable after exploring the dataset compared to sleep disorders.

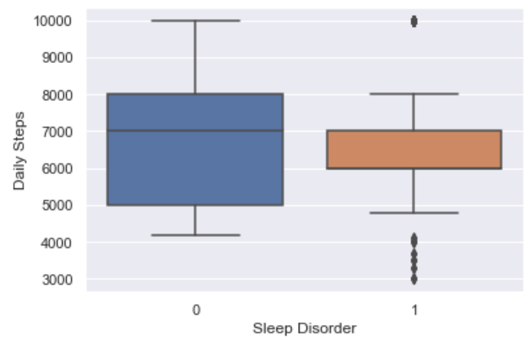
No	**Variable**	**Diagram**	**Explanation**
1	Gender		The figure represents the percentage of individuals who suffered from sleep disorders, grouped by gender. The number of women is greater than that of men, which accounts for 66.5 %. Females are twice as likely to have sleep problems than males.

<table>
<tr><td>2</td><td>Age</td><td>
</td><td>The box plot shows age distribution by the binary sleep disorders value. Older people tend to experience sleep problems than younger people.</td></tr>
<tr><td>3</td><td>Occupation</td><td><table>
<tr><th></th><th>Occupation</th><th>Percentage</th></tr>
<tr><td>0</td><td>Nurse</td><td>19.518717</td></tr>
<tr><td>1</td><td>Doctor</td><td>18.983957</td></tr>
<tr><td>2</td><td>Engineer</td><td>16.844920</td></tr>
<tr><td>3</td><td>Lawyer</td><td>12.566845</td></tr>
<tr><td>4</td><td>Teacher</td><td>10.695187</td></tr>
<tr><td>5</td><td>Accountant</td><td>9.893048</td></tr>
<tr><td>6</td><td>Salesperson</td><td>8.556150</td></tr>
<tr><td>7</td><td>Software Engineer</td><td>1.069519</td></tr>
<tr><td>8</td><td>Scientist</td><td>1.069519</td></tr>
<tr><td>9</td><td>Sales Representative</td><td>0.534759</td></tr>
<tr><td>10</td><td>Manager</td><td>0.267380</td></tr>
</table>

</td><td>The table gives information about the proportion of the type of jobs in the dataset. Nurse accounts for around one-fifth of the total population, while Manager makes up the lowest percentage, with a value of 0.2 %. There is unbalanced data in each type of job, so it can be difficult to interpret.

The figure shows the percentage of people experiencing sleep problems in each job. Nursing is the type of job that has the highest percentage, which accounts for 40 %. On the other hand, software engineer has the lowest proportion, which made up less than 1%.</td></tr>
<tr><td>4</td><td>Sleep Duration</td><td>
</td><td>The box plot shows sleep duration distribution by the binary sleep disorders value. Individuals are prone to experience sleep problems if they lack sleep.</td></tr>
</table>

5	Quality of Sleep		The multi-bar graph shows how many individuals have sleep problems or not based on their quality of sleep. People with sleep quality below six tend to have more sleep problems than those with better sleep quality. A small number of people with an eight-number sleep quality have sleep problems. Eight is the best number of sleep quality because a large number of people have no issues with sleep compared to the rest of the number of sleep quality.
6	Physical Activity Level	 	People with a forty-five-number physical activity level are more likely to have sleep disorders than individuals with other numbers.
7	Stress Level		People with stress levels below and equal to six sleeps better than those with a higher stress level.

<table>
<tr><td>8</td><td>BMI Category</td><td>
</td><td>Overweight people tend to have sleep problems</td></tr>
<tr><td>9</td><td>Systolic</td><td>
</td><td>People with a higher systolic pressure tend to have sleep disorders</td></tr>
<tr><td>10</td><td>Diastolic</td><td>
</td><td>People with a lower diastolic pressure tend to have no issues with sleep</td></tr>
<tr><td>11</td><td>Heart Rate</td><td>
</td><td>People with a higher heart rate tend to have sleep problems</td></tr>
</table>

<table>
<tr>
<td>11</td>
<td>Daily Steps</td>
<td>
</td>
<td>People with sleep problems have small variance in their daily steps than those with no sleep disorders.</td>
</tr>
</table>

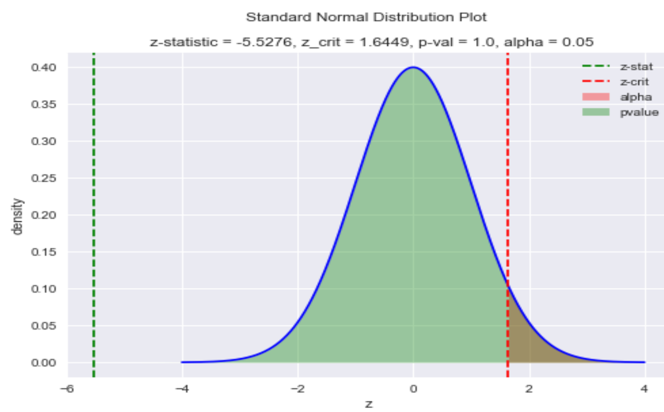
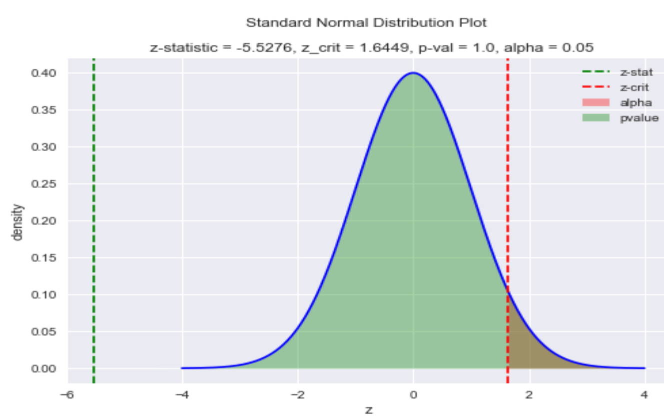
The summaries of EDA are:

1. Females are twice as likely to have sleep problems than males.
2. Older people tend to experience sleep problems than younger people.
3. Occupation is hard to interpret.
4. Individuals are prone to experience sleep problems if they lack sleep (shorter sleep length)
5. Higher sleep quality is better. The parameter is too subjective.
6. People with a forty-five-number physical activity level are more likely to have sleep disorders than individuals with other numbers. The variable is hard to interpret.
7. People with stress levels below and equal to six sleeps better than those with a higher stress level. The variable is hard to interpret.
8. Overweight people tend to have sleep problems.
9. People with higher heart rates tend to have sleep problems.
10. People with sleep problems have small variance in their daily steps than those with no sleep disorders.
11. People with higher systolic pressure tend to have sleep disorders.
12. People with lower diastolic pressure tend to have no issues with sleep.

Hence, considered variables are gender, age, sleep duration, quality of sleep, BMI category, heart rate, systolic, & diastolic.

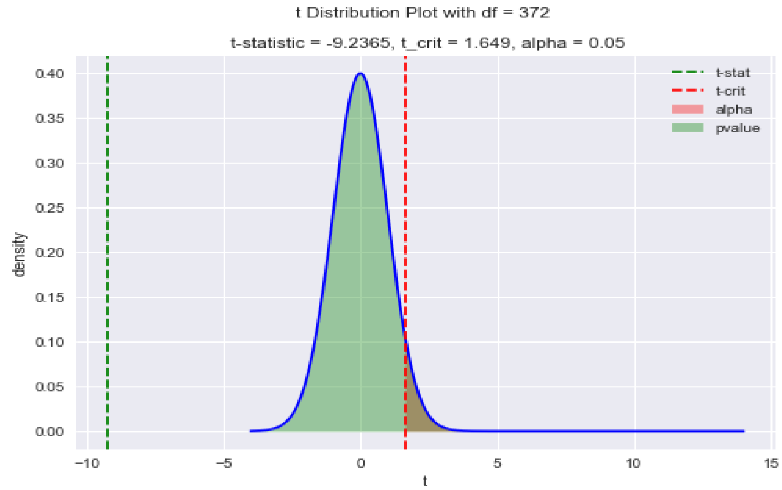
4. Confirmatory Analysis

4.1 T-test for Two Proportions

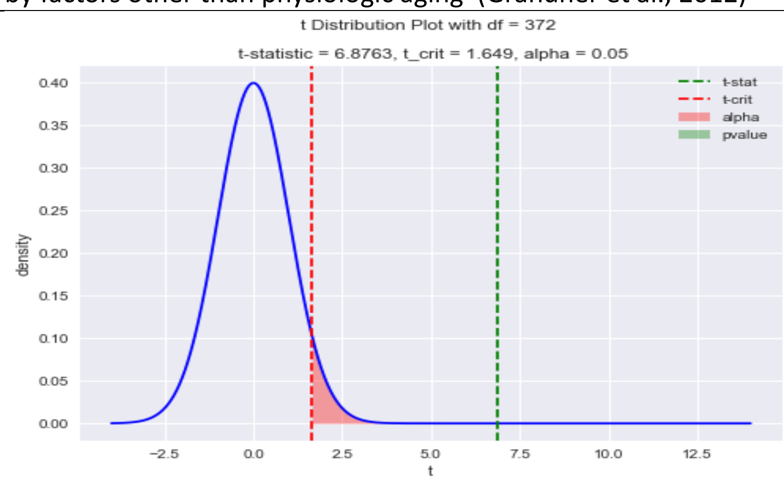
No	Variable	Hypothesis	Graph & Explanation
1	Gender	Ho = Proportion Sleep-disorders Male ≤ Proportion Sleep-disorders Female H1 = Proportion Sleep-disorders Male > Proportion Sleep-disorders Female	 From the result, with an alpha number of 0,05, we see that z-stat is far below than z-critical. We therefore do not have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the percentage of males who have sleep disorders is less than equals to the percentage of females. Gender is considered as a predictor for a model.
2	BMI Category	Ho = Proportion Sleep-disorders Normal Weight ≥ Proportion Sleep-disorders Overweight H1 = Proportion Sleep-disorders Normal Weight < Proportion Sleep-disorders Overweight	 From the result, with an alpha number of 0,05, we see that z-stat is far below than z-critical. We therefore do have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the percentage of normal-weight people who have sleep disorders is less than the percentage of overweight people. BMI Category is considered as a predictor for a model.

4.2 T-test for Two Means Population

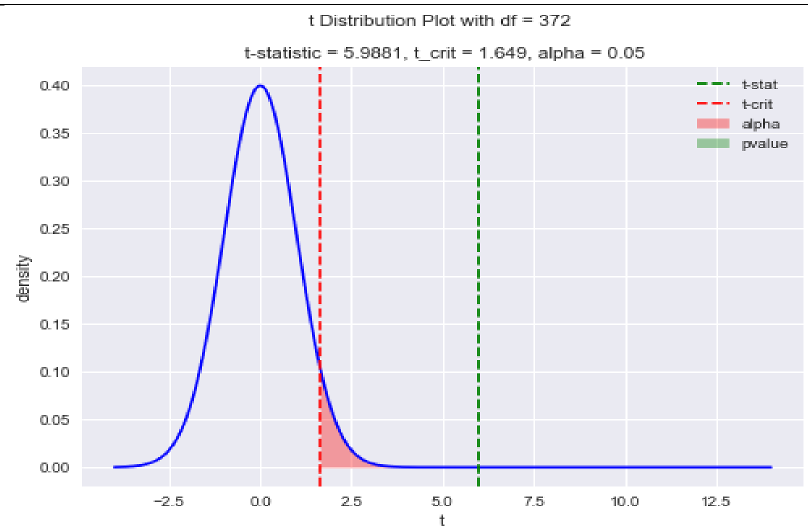
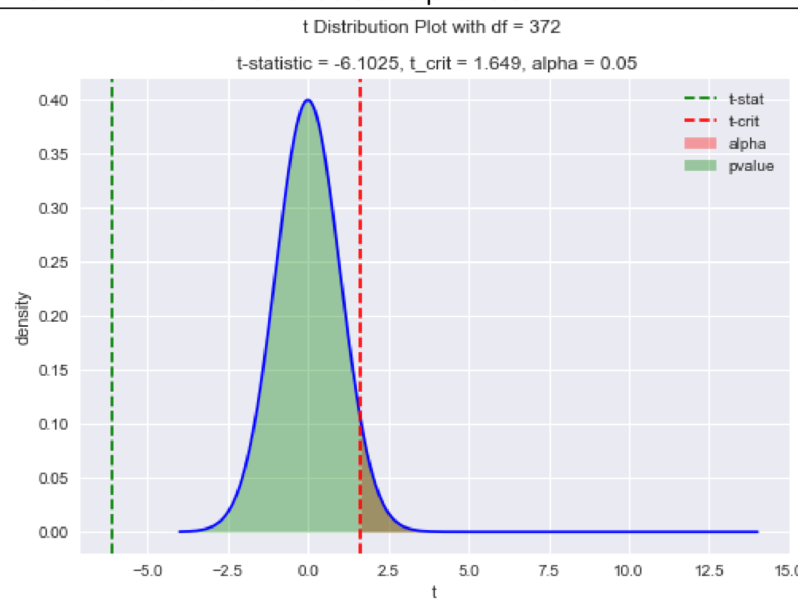
<table>
<tr><th>No</th><th>Variable</th><th>Hypothesis</th><th>Graph & Explanation</th></tr>
<tr><td>1</td><td>Age</td><td>Ho = The mean age sleep-disorders people ≤ mean age people with no sleep disorders

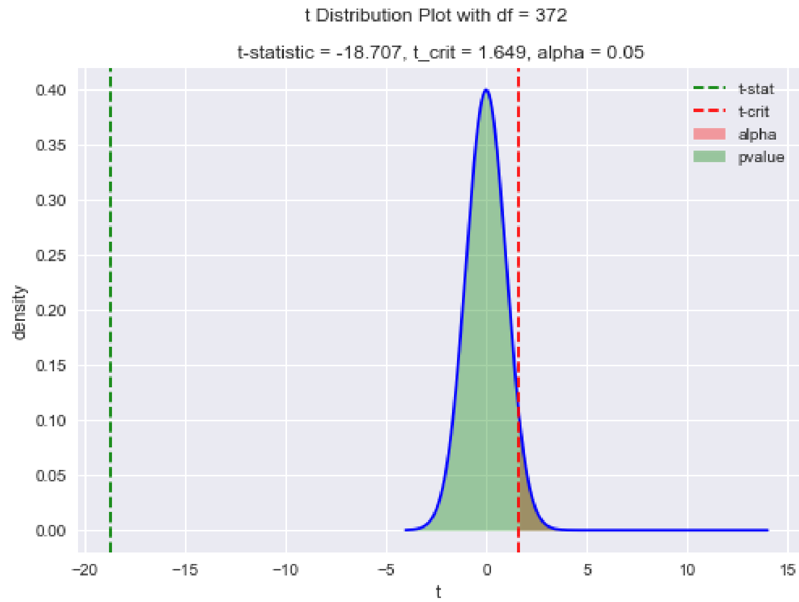
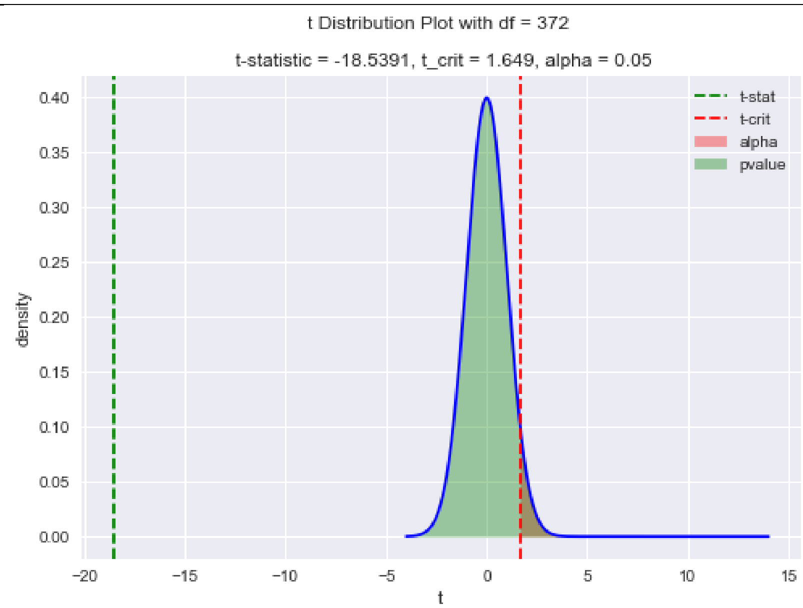
H1 = The mean age sleep-disorders people greater than mean age people with no sleep disorders</td><td>

From the result, with an alpha number of 0,05, we see that t-stat is far below than t-critical. We therefore do not have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean age of people who have sleep problems is less than equal to the mean age of those who sleep normally. The hypothetical testing result is the opposite of the exploratory data result. Based on the previous research, the often-reported increase in sleep problems with age is a nonlinear phenomenon, mediated by factors other than physiologic aging (Grandner et al., 2012)</td></tr>
<tr><td>2</td><td>Sleep Duration</td><td>Ho = mean sleep duration sleep-disorders people ≥ mean sleep duration people with no sleep disorders

H1 = mean sleep duration sleep-disorders people less than mean sleep duration people with no sleep disorders</td><td>

From the result, with an alpha number of 0,05, we see that t-stat is far above than t-critical. We therefore do not have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean sleep duration of people with sleep disorders is more than equal to people who sleep normally. The result of the hypothetical testing is contradictive with the previous research, which stated that short-sleep duration was associated with the risk of sleep disorders (Dilixiati et al, 2023). Hence, we do not include sleep-duration as a predictor variable.</td></tr>
</table>

3	Quality of Sleep	Ho = mean sleep quality sleep-disorders people ≥ mean sleep quality with no sleep disorders H1 = mean sleep quality sleep-disorders people less than mean sleep quality people with no sleep disorders	 From the result, with an alpha number of 0,05, we see that t-stat is far above than t-critical. We therefore do not have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean sleep quality of people with sleep disorders is more than equal to people who sleep normally. However, quality of sleep is a subjective rating of the quality of sleep, ranging from 1 to 10, so we do not include the variable as a predictor.
4	Heart Rate	Ho = mean heart rate sleep-disorders people ≥ mean heart rate people with no sleep disorders H1 = mean heart rate sleep-disorders people less than mean heart rate people with no sleep disorders	 From the result, with an alpha number of 0,05, we see that t-stat is far below than t-critical. We therefore do have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean heart rate of people with sleep disorders is less than people who sleep normally. Heart rate is considered as a predictor for a model.

5	Systolic	Ho = mean systolic sleep-disorders people ≥ mean systolic people with no sleep disorders H1 = mean systolic sleep-disorders people less than mean systolic people with no sleep disorders	 From the result, with an alpha number of 0,05, we see that t-stat is far below than t-critical. We therefore do have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean systolic pressure of people with sleep disorders is less than people who sleep normally. Systolic is considered as a predictor for a model.
6	Diastolic	Ho = mean Diastolic sleep-disorders people ≥ mean Diastolic people with no sleep disorders H1 = mean Diastolic sleep-disorders people less than mean Diastolic people with no sleep disorders	 From the result, with an alpha number of 0,05, we see that t-stat is far below than t-critical. We therefore do have sufficient evidence to reject the null hypothesis. We can, therefore, conclude that the mean diastolic pressure of people with sleep disorders is less than people who sleep normally. Diastolic is considered as a predictor for a model.

Based on the hypothetical testing, predictor variables are gender, BMI category, heart rate, systolic and diastolic.

4.3 Model Fitting

$$P(Sleep\ Disorder) = logit^{-1}(-33.31 - 0.478\ gender + 2.87 BMIcategory$$
$$+0.112 heartate + 0.26 systolic - 0.117 diastolic$$

Table coefficient and standard error

	coef	std err
Intercept	-33.309732	8.167365
gender	-0.478539	0.437875
bmicategory	2.875856	0.458072
heartrate	0.111738	0.051645
systolic	0.262157	0.174480
diastolic	-0.116855	0.208470

The formula is a logistic regression model that predicts the probability of having a sleep disorder based on several predictor variables, including gender, BMI category, heart rate, systolic blood pressure, and diastolic blood pressure. Here's an interpretation of the coefficients in the logistic regression equation:

- **Intercept** (-33.31): The intercept is the log-odds of having a sleep disorder when all the predictor variables are set to zero (or reference levels for categorical variables). In this case, it represents the baseline log-odds of having a sleep disorder for a reference group. A negative intercept suggests that the baseline odds of having a sleep disorder are low.
- **Gender** (-0.478): This coefficient represents the change in the log-odds of having a sleep disorder associated with a one-unit change in the gender variable. If gender is coded as binary (e.g., 0 for female, 1 for male), a negative coefficient suggests that being male is associated with a decrease in the log-odds of having a sleep disorder compared to being female.
- **BMI Category** (2.87): This coefficient represents the change in the log-odds of having a sleep disorder associated with a one-unit change in the BMI category variable. If BMI categories are ordinal (e.g., 0 for normal, 1 for overweight), a higher BMI category is associated with a greater increase in the log-odds of having a sleep disorder.
- **Heart Rate** (0.112): a difference of a one-unit of change in heart rate corresponds to a 0.112 positive difference in the logit probability of having sleep disorders.
- **Systolic Blood Pressure** (0.26): a difference of a one-unit of change in systolic blood pressure corresponds to a 0.26 positive difference in the logit probability of having sleep disorders.
- **Diastolic Blood Pressure** (-0.117): a difference of a one-unit of change in diastolic blood pressure corresponds to a 0.117 positive difference in the logit probability of having sleep disorders. A negative coefficient suggests that higher diastolic blood pressure is associated with a decrease in the log-odds of having a sleep disorder.

- All of the coefficients are greater than their standard error, except the diastolic blood pressure parameter.

5. Limitations of the Research Work

In the course of carrying out this work, the limitations we encountered are explained below:

1. Lack of relevant and correct sources that can provide data on significant predictors of probability of having sleep disorders.
2. Lack of data in age-group number and occupation.
3. The total of data is low, which consists of 374 data.

6. Conclusion

Based on the analysis, we can conclude that:

1. Predictor variables that satisfy exploratory data analysis and confirmation analysis are gender, BMI category, heart rate, systolic and diastolic.
2. The model for predicting the probability of having sleep disorders are:

$$P(Sleep\ Disorder) = logit^{-1}(-33.31 - 0.478\ gender + 2.87 BMIcategory$$

$$+0.112 heartate + 0.26 systolic - 0.117 diastolic$$

References

Dilixiati D, Kadier K, Laihaiti D, Lu JD, Rezhake R, Azhati B, Rexiati M. The relationship between sleep disorders, quality, duration and sexual dysfunction: a systematic review and meta-analysis. J Sex Med. 2023 Apr 26: qdad054. doi: 10.1093/jsxmed/qdad054. Epub ahead of print. PMID: 37186140.

Grandner, M. A., Martin, J. L., Patel, N. P., Jackson, N. J., Gehrman, P. R., Pien, G., Perlis, M. L., Xie, D., Sha, D., Weaver, T., & Gooneratne, N. S. (2012). Age and sleep disturbances among American men and women: Data from the U.S. Behavioral Risk Factor Surveillance System. *Sleep*, *35*(3), 395–406. https://doi.org/10.5665/sleep.1704

Wang, Z., Wang, Y., Huang, J., Xu, J., Chen, F., Zhu, Z., Gao, L., Qin, J., Liu, B., & Liang, C. (2022). Sleep Disturbance Affects Immune Factors in Clinical Liver Cancer Patients. *Current Oncology*, *29*(10), 7943–7952. https://doi.org/10.3390/curroncol29100628