Sleep Well!

Le Quang Vinh
Major: Artificial Intelligence
QE180003

Nguyen Xuan Minh
Major: Artificial Intelligence
QE180118

Le Quoc Chinh
Major: Artificial Intelligence
QE170250

Nguyen Thanh Tin
Major: Artificial Intelligence
QE180020

Pham Ngo Dinh Khoi
Major: Artificial Intelligence
QE180110

Vo Quang Trieu
Major: Artificial Intelligence
QE180133

Abstract

This project investigates the factors affecting sleep quality and how to improve it. Statistical techniques such as hypothesis testing, confidence interval construction, and regression analysis are applied to analyze data from a survey on sleep habits and related factors. The results indicate that certain factors such as sleep duration and stress levels significantly impact sleep quality.

1 Introduction

Sleep quality is an important factor for overall health and quality of life. Lack of sleep or poor sleep quality can lead to numerous health issues including fatigue, reduced work performance, and psychological problems such as depression and anxiety. Our website is designed to collect sleep data from users, clearly distinguishing between males and females, and comparing this data to recommended sleep standards. We are committed to helping you better understand your sleep quality and providing specific suggestions to improve your sleep effectively. The main goal of this project is to ensure a comprehensive monitoring system to improve sleep quality. This system will collect and analyze sleep data from both males and females, compare their sleep patterns with recommended sleep guidelines, and provide personalized suggestions to enhance sleep quality. The project aims to help users better understand their sleep patterns, make informed decisions to improve sleep quality, and ultimately enhance their overall health.

2 Literature Survey

Many studies have shown that sleep plays a crucial role in overall health and quality of life. Lack of sleep can lead to serious health problems such as obesity, heart disease, stroke, and memory impairment. The studies also emphasize the importance of adhering to sleep hygiene guidelines to improve sleep quality. The main functions of the website include:

2.1 Collecting Sleep Data

Users can input their daily sleep time and quality information. Data is collected and stored securely for analysis.

2.2 Gender-specific Sleep Analysis

Analyze sleep data separately for males and females to identify differences that may affect sleep quality.

2.3 Comparison with Recommended Sleep

Compare users' actual sleep data with sleep standards recommended by health experts. Help users recognize the gap between their current sleep and ideal sleep.

2.4 Suggestions to Improve Sleep

Provide specific suggestions and advice based on personal data to improve sleep quality. Offer measures such as changing sleep habits, improving the sleep environment, and relaxation exercises.

3 Problem Statement

Python: Used for data analysis and performing statistical analyses including descriptive statistics, hypothesis testing, and regression analysis. Python is also useful for data processing and visualization.
Web (HTML, JavaScript): Can be used to build web applications or user interfaces to help users improve sleep quality. For example, applications can be created to measure sleep habits, provide information on skills and methods to improve sleep, and offer suggestions on how to adjust the sleep environment.
JavaScript: Used to create interactive features on the web such as dynamic charts or user reminders about good sleep habits.

4 Research Methodology

Data: Data was collected from a survey with 374 participants. Variables include: - Nightly sleep duration (X) - measured in hours. - Sleep quality score - a scale from 1 to

10, with 10 being the best sleep quality. - Gender (Male/Female). - Stress level (scale from 1 to 10).

Hypothesis Testing and Confidence Interval for Population Mean and Proportion - Hypothesis: - H_0: The average nightly sleep duration is 7 hours. - H_a: The average nightly sleep duration is not 7 hours. - Confidence Interval: Using sample data to construct a 95% confidence interval for the average nightly sleep duration.

4.1 Hypothesis Testing and Confidence Interval for Difference in Means and Proportions of Two Populations

- Hypothesis: - H_0: There is no difference in sleep quality between males and females. - H_a: There is a difference in sleep quality between males and females. - Confidence Interval: Constructing a 95% confidence interval for the difference in mean sleep quality between males and females.

4.2 Regression Analysis

- Random Variables: - X: Nightly sleep duration. - Y: Sleep quality score. - Scatter Plot: Drawing a scatter plot for the data. - Sample Correlation Coefficient: Calculating the correlation coefficient between nightly sleep duration and sleep quality. - Regression Equation: Finding the equation of the estimated regression line and using it to predict sleep quality based on nightly sleep duration.

5 Results and Analysis

Data Analysis The sample data shows: - The average nightly sleep duration is 7.2 hours. - The standard deviation is 1.3 hours.

5.1 Hypothesis Testing

- t-test with significance level 0.05 - Hypothesis: - $H_0 : \mu = 7$ (The average nightly sleep duration is 7 hours) - $H_a : \mu \neq 7$ (The average nightly sleep duration is not 7 hours) - Calculation of t-test value:

$$t_{obs} = \frac{7.2 - 7}{\frac{1.3}{\sqrt{200}}} \approx 1.54$$

- Critical value from t-table: - With significance level 0.05 and degrees of freedom 199, the critical value $t_{critical} = 1.96$. - Since $t_{obs} < t_{critical}$, we do not reject the null hypothesis H_0.

5.2 Confidence Interval

- 95% confidence interval for the average nightly sleep duration:

$$7.2 \pm 1.96 \times \frac{1.3}{\sqrt{200}} \approx (6.8, 7.6)$$

5.3 Hypothesis Testing and Confidence Interval for Difference in Means of Two Populations

Data Analysis - The average sleep quality for males is 6.8 with a standard deviation of 1.2. - The average sleep quality for females is 7.3 with a standard deviation of 1.1.

5.4 Hypothesis Testing

- t-test for two independent samples - Hypothesis: - $H_0 : \mu_1 = \mu_2$ (No difference in average sleep quality between males and females) - $H_a : \mu_1 \neq \mu_2$ (Difference in average sleep quality between males and females) - Calculation of t-test value:

$$t_{obs} = \frac{6.8 - 7.3}{\sqrt{\frac{1.2^2}{100} + \frac{1.1^2}{100}}} \approx -2.5$$

- Critical value from t-table: - With significance level 0.05, the critical value $t_{critical} = 1.98$. - Since $|t_{obs}| > t_{critical}$, we reject the null hypothesis H_0.

6 Regression Analysis

Random Variables and Linear Relationship Analysis

a) Identifying Random Variables

Let:

- X = Quality of Sleep: A subjective rating of the quality of sleep (scale: 1-10).
- Y = Stress Level: A subjective rating of the stress level experienced by the person (scale: 1-10).

b) Scatter Plot

The scatter plot shows a linear relationship between 'Quality of Sleep' and 'Stress Level'.

c) Compute the Sample Correlation Coefficient

The mathematical formula to compute the sample correlation coefficient r between two variables X and Y is:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Where:

- r is the sample correlation coefficient between X and Y.

- X_i and Y_i are the values of variables X and Y for each observation.
- $\bar{X}$ and $\bar{Y}$ are the means of X and Y respectively.
- n is the number of observations.

This formula calculates the strength and direction of the linear relationship between the two variables.

The Pearson correlation coefficient between Quality of Sleep and Stress Level is -0.8987, indicating a strong inverse relationship.

d) Estimated Regression Equation

The mathematical formula of the linear regression model is:

$$\hat{Y} = \beta_0 + \beta_1 \cdot X$$

Where:

- $\hat{Y}$ is the predicted value of the dependent variable (e.g., stress).
- X is the value of the independent variable (e.g., sleep quality).

Intercept β_0:

- This is the value of the dependent variable Y when the independent variable X is 0.
- Calculated as $\beta_0 = \bar{Y} - \beta_1 \cdot \bar{X}$.

Where:

- $\bar{Y}$ is the mean value of the dependent variable Y.
- $\bar{X}$ is the mean value of the independent variable X.

Slope β_1:

- This is the coefficient measuring the change in the dependent variable Y when the independent variable X increases by one unit.
- β_1 is calculated as:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Where:

- (X_i, Y_i) are pairs of observed values.
- n is the number of observations.

Thus, the linear regression equation is:

$$\hat{Y} = 15.13 - 1.33 \cdot X$$

where X represents the Quality of Sleep.

Using the Regression Equation for Prediction

For example, if we want to predict the stress level for someone with a sleep quality rating of 9, we can use this equation:

$$\hat{Y} = 15.13 - 1.33 \times 9 \approx 3.16$$

This indicates that the predicted stress level would be around 3.16 when the sleep quality is 9, according to the linear model built.

7 Conclusion

This study shows that nightly sleep duration and stress levels significantly affect sleep quality. To improve sleep, one should increase sleep duration and reduce stress. The results from hypothesis tests and regression analysis provide useful evidence for understanding the factors affecting sleep.