

CS 210
Course Project

Dilay Durak
28141

18.01.2024

. Sabancı .
Universitesi

Motivation

I am working on this project because I have always wanted to exercise regularly throughout my life, and I have often used the excuse of weather conditions, whether it's too hot or too cold, to avoid it. I chose to undertake this project to find out if these excuses have had any impact on my workouts and daily calorie expenditure.

Data Source - My Own Data

- **Kcal per week:** This column quantifies the average calories that I burned per week, serving as a measure of physical activity. I took that data from my fitness application, and this application measure kcal by my apple watch.
- **Weather:** Reflecting the prevailing weather conditions during each recorded period and location (centigrade format).
- **Date:** The start date of the week I am calculating. I prepared it in chronological order, with 7 days (1 week) between each data.
- **City:** Specifies the geographical location associated with each data entry. I used that column to find correct weather.

	A	B	C	D	E
1	kcal per week	weather	date	city	
2	92	10.14	2.01.2023	istanbul	
3	111	10.07	9.01.2023	istanbul	
4	150	13.43	16.01.2023	istanbul	
5	48	8.93	23.01.2023	istanbul	
6	89	4.58	30.01.2023	istanbul	
7	72	3.14	6.02.2023	istanbul	
8	53	7.93	13.02.2023	istanbul	
9	84	12.14	20.02.2023	istanbul	
10	67	11.00	27.02.2023	istanbul	
11	77	13.29	6.03.2023	istanbul	

Hypothesis Testing

Hypothesis testing is a statistical method used to make inferences about a population based on a sample of data. It involves formulating a hypothesis about the population parameter, collecting and analyzing data, and then deciding whether to accept or reject the null hypothesis.

Hypotheses:

- **H0 (Null Hypothesis):** “There is no correlation between weather conditions and the calories burned per week.”

- **H1 (Alternative Hypothesis):** “There is a correlation between weather conditions and the calories burned per week.”

Data Analysis

`“correlation_coefficient, p_value = pearsonr(df['weather'], df['kcal per week'])”`

The code calculates the Pearson correlation coefficient and p-value between the 'weather' and 'kcal per week' columns in the DataFrame (`df`). The correlation coefficient, assigned to `correlation_coefficient`, indicates the strength and direction of the linear relationship between the two variables. A value close to 1 or -1 implies a stronger correlation, while 0 suggests no linear correlation. The p-value, assigned to `p_value`, represents the probability of observing the calculated correlation if there were no true correlation. A smaller p-value indicates stronger evidence against the null hypothesis. These statistical measures provide insights into the association between 'weather' and 'kcal per week,' informing subsequent hypothesis testing and decision-making in the code. This line of code calculates the Pearson correlation coefficient and p-value between the 'weather' and 'kcal per week' columns, providing insights into the linear relationship (if any) between these two variables. The subsequent analysis in the code involves interpreting these results and making decisions based on the calculated values and a predefined significance level (`alpha`).

1. Pearson Correlation Coefficient

The calculated Pearson correlation coefficient is approximately 0.375. This positive value suggests a moderate positive correlation between weather conditions and the calories burned per week. The coefficient's magnitude indicates a discernible relationship, while the positive sign implies that as weather conditions improve, there tends to be an increase in the calories burned.

2. P-value

The p-value associated with the correlation coefficient is 0.0062, which is less than the common significance level of 0.05. As a result, we reject the null hypothesis (H_0) that posits there is no correlation between weather and calories burned per week. The low p-value indicates that the observed correlation is statistically significant.

Findings

P-value (0.0062) being less than the significance level of 0.05, you would reject the null hypothesis, indicating that there is no evidence to suggest a correlation between weather and kcal per week.

“Reject H_0 : There is no correlation between weather and kcal per week.”

Based on the results, I realized that my excuses have not had any impact on my life and calorie expenditure.

Limitations and Future Work

Despite the valuable insights gained, the project has certain limitations. The dataset's scope may not encompass all relevant factors influencing physical activity, and external variables may play a role. Future work could involve expanding the dataset to include additional variables, considering seasonal variations, or incorporating geographical factors. Additionally, a more extensive analysis of outliers and a deeper exploration of specific weather conditions could enhance the precision of the findings. The project lays the groundwork for future endeavors in personalized health analytics and contributes to ongoing discussions about the interplay between environmental factors and individual well-being.

References

<https://www.accuweather.com/>