“SLEEP EFFICIENCY AMONG VARIOUS CRITERIA: ANOVA ”

A course based project report submitted in partial fulfilment of the requirement for the award of the certificate of

COURSE BASED PROJECT

Under the Guidance Of
Dr. N. POTHANNA
(Associate Professor, VNRVJIET)

Department of Humanities &Sciences

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTE
(Approved by AICTE, New Delhi, Govt of T.S and Affiliated to JNTU, Hyderabad)
Accredited by NBA and NAAC with ‘A++’ Grade
Bachupally, Hyderabad – 500090, Telangana, India.
January 2024

VALLURUPALLI NAGESWARA RAO VIGNANA JYOTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY
AN AUTONOMOUS INSTITUTE
(Approved by AICTE, New Delhi, Govt of T.S and Affiliated to JNTU, Hyderabad)
Accredited by NBA and NAAC with 'A++' Grade
Bachupally, Hyderabad – 500090
Telangana, India.

Estd.1995

This is to certify that the summer online course-based project entitled **"SLEEP EFFICIENCY AMONG VARIOUS CRITERIA: ANOVA"** has been carried out and submitted by the following students during 11-09-2023 to 06-01-2024.

S.No	Name of the Student	Hallticket number	S.No	Name of the Student	Hallticket number
1	B. Nived Krishna	22071A0570	7	S. Navrin	22071A05B0
2	K. Devi Sri Charan	22071A0585	8	T. Adithya	22071A05B2
3	R. Prashanth	22071A05A5	9	T. Navya	22071A05B4
4	R. Sudhakar	22071A05A6	10	T. Aniketh	23075A05C2
5	S. Adarsh	22071A05A8	11	U. Sandeep	22075A05C3
6	S. Sai Teja Akhil	22071A05A9	12	V. Indupriya	22071A05C8

Dr.N. POTHANNA
Project Guide
Associate Professor
Humanities and Sciences
VNRVJIET

Dr. S. Nagini
Head of the Department
Professor
Computer Science and Engineering
VNRVJIET

DECLARATION

We, the undersigned declare that the course-based project report entitled **"SLEEP EFFICINECY AMONG VARIOUS CRITERIA: ANOVA"** has been carried out and submitted. It is an authentic work and has not been submitted to any other Institute/University.

PLACE: HYDERABAD

DATE:

B. Nived Krishna
(22071A0570)

K. Devi Sri Charan
(22071A0585)

R. Prashanth
(22071A05A5)

R. Sudhakar
(22071A05A6)

S. Adarsh
(22071A05A8)

S. Sai Teja Akhil
(22071A05A9)

S. Navrin
(22071A05B0)

T. Adithya
(22071A05B2)

T. Navya
(22071A05C3)

T. Aniketh
(22071A05C2)

U. Sandeep
(22075A05C3)

V. Indupriya
(22075A05C&)

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without mentioning the people who have made it possible, because success is the epitome of hard work. So, with gratitude, we acknowledge all those whose guidance and encouragement have made our efforts successful.

We wish to express our deep sense of gratitude to our Principal **Dr. C.D. NAIDU,** Director of Advancement Dr. **B. Chennakesava Rao** and Head of the Department of Computer Science and Engineering **Dr. S. NAGINI** for their encouragement, which went a long way in the successful completion of this course-based project.

We express our sincere gratitude to **Dr. N. POTHANNA**, Associate Professor, Department of Humanities and Sciences who throughout the course of internship guided us without his consent we would not have been able to complete this project.

ABSTRACT

This study dives deep into understanding sleep efficiency by using ANOVA analysis as a strong statistical method. We want to explore lots of different things about sleep, like how long people sleep, disturbances, how quickly they fall asleep, and more. We're doing a super detailed analysis to figure out all the little connections that affect sleep quality.

Our main goal is to find out the big differences in sleep among different groups of people, like based on their demographics and behaviours. ANOVA analysis helps us see these subtle patterns and how they affect sleep in different situations and for different groups. It's like trying to untangle a really tangled-up knot to understand how all the parts fit together.

This study isn't just for school. It could be super helpful for things like healthcare and how we live our lives. The stuff we learn from our analysis could help make personalized plans for better sleep. Plus, it might help make new rules and plans to help everyone have healthier sleep habits and be happier overall.

Basically, our goal is to use ANOVA analysis to understand sleep better and use that knowledge to help people sleep well. It's not just about individuals; it's about making things better for everyone by understanding sleep in a really detailed way.

CONTENTS

CHAPTER 1

INTRODUCTION

Sleep efficiency significantly impacts overall health and well-being. Adequate and restorative sleep is essential for physical health, mental clarity, emotional balance, and a robust immune system. Conversely, poor-quality or insufficient sleep is associated with increased stress, weakened immunity, cognitive impairment, and a higher risk of chronic conditions like hypertension, diabetes, and mental health disorders.

Several factors influence sleep efficiency. These include sleep duration, disturbances, sleep onset latency, depth of sleep stages, environmental factors (like bedroom conditions, noise, and temperature), lifestyle habits (such as exercise routines, diet, and stress levels), and demographic factors (like age, gender, and individual health conditions). Understanding these multifaceted influences is crucial in addressing sleep health.

The duration of sleep plays a pivotal role in sleep efficiency. Both prolonged sleep deprivation and excessive sleep have been linked to adverse health outcomes. Disturbances during sleep, such as frequent awakenings or irregular sleep schedules, also impact sleep quality significantly. Additionally, sleep disorders like insomnia, sleep apnea, and circadian rhythm disorders can profoundly affect sleep efficiency and overall health.

Behavioural and lifestyle choices significantly affect sleep efficiency. Regular exercise, a balanced diet, and stress management practices positively impact sleep quality. Conversely, habits like caffeine and alcohol consumption, irregular sleep schedules, excessive screen time before bed, and high stress levels can detrimentally affect sleep quality. Environmental factors like light, noise, and bedroom comfort also play crucial roles. Demographic factors like age, gender, and underlying health conditions influence sleep efficiency. Age-related changes in

sleep patterns, gender-specific influences on sleep quality, and the presence of health issues significantly impact sleep efficiency. Understanding these demographic and health-related impacts is vital for tailoring interventions to optimize sleep health across different populations. In summary, sleep efficiency is intricately connected to various aspects of health. Recognizing and addressing the multifaceted factors influencing sleep quality are crucial steps in improving overall health outcomes.

CHAPTER 2

LITERATURE REVIEW

The statistical technique known as Analysis of Variance (ANOVA) is frequently used in a variety of academic fields to investigate variations in means between several groups. This overview of the literature examines the background, core ideas, and various uses of ANOVA, emphasising its importance in improving research methods.

1. **Historical Overview:**

   Sir Ronald A. Fisher's groundbreaking work in the early 20th century is where ANOVA first emerged. Fisher created the idea of variance analysis to solve problems with agricultural experimentation. ANOVA developed into a potent statistical tool throughout time, finding use in a variety of disciplines outside of agriculture, such as psychology, biology, economics, and more.

2. **Fundamental Concepts:**

   ANOVA is predicated on basic ideas like dividing variance into its component parts, contrasting group averages, and determining the importance of observed differences. The underlying presumptions of homogeneity of variances and normal distribution within groups are among them. It is essential to comprehend these ideas in order to apply ANOVA accurately in a variety of research settings.

3. **ANOVA in the Behavior Sciences and Psychology:**

   ANOVA is frequently used in psychology to examine variations in mean scores between experimental groups, which aids in the comprehension of behavioral and cognitive phenomena. This use case also includes the behavioral sciences, where researchers can examine how different populations are affected by treatments with the help of ANOVA.

4. **ANOVA in Clinical and Biomedical Research:**

   By comparing mean outcomes across many patient groups, biomedical research use ANOVA to assess the effectiveness of treatment. This is especially important for clinical trials, where ANOVA is used to determine the best course of action for treating patients with different features.

5. Modern Applications and Technological Developments:

As computing and technology have progressed, so too have ANOVA's uses. Big data and advanced statistical software have made it possible for academics to perform intricate ANOVA studies, which enables more in-depth evaluations of group differences across a range of fields.

In conclusion, the significance of ANOVA as a foundational statistical tool across disciplines is highlighted by its historical evolution, fundamental concepts, and wide applications. ANOVA is still a reliable and flexible way to examine mean differences across several groups as research approaches change, which advances knowledge and supports evidence-based decision-making.

CHAPTER 3

METHODOLOGY

3.1 DATA DESCRIPTION

1. Data Collection:

This dataset gathers age, gender, bedtime, wakeup time, sleep duration, efficiency, REM, deep and light sleep percentages, awakenings, caffeine and alcohol intake, smoking habits, and exercise frequency. It compiles diverse demographic and sleep-related metrics, allowing exploration of potential correlations among these factors within this sample.

2. Data Preprocessing:

Eliminate any anomalies or discrepancies from the dataset. Make sure every variable is prepared

for analysis in the right format. Give categorical variables, such "Male" and "Female," numerical values.

3. Grouping by Categories:

Divide the dataset into distinct groups based on the provided Criteria:

- Exercise efficiency
- Gender
- Smoking Status
- Caffeine Consumption
- Alcohol Consumption

4. Descriptive Statistics:

Determine the mean and standard deviation of the Sleep Efficiency readings for every group. This gives an early summary of the information and draws attention to any possible variations between age groups.

5. ANOVA Analysis:

Perform a one-way Analysis of Variance (ANOVA) to assess the statistical significance of Sleep efficiency values differences among the identified groups based on the individual category. The null hypothesis's (H0) assumes that there is no significant difference in mean Sleep Efficiency levels across groups of all categories.

ANOVA FORMULA:

$$F=\left(\frac{MSBetween}{MSWithin}\right)$$

- $MS_{Between}$ is the mean square between groups, representing the variance due to differences between age groups.
- MS_{Within} is the mean square within groups, representing the variance within each age group.

6.Post hoc Tests (if necessary):

Conduct post hoc tests, such as Tukey's HSD, to identify specific groups that exhibit significant differences in Sleep Efficiency levels. This helps to pinpoint where the differences lie if the ANOVA indicates a significant overall effect.

7. Interpretation of Results:

Interpret the results, considering both the overall ANOVA significance and post hoc test outcomes. Determine whether there are statistically significant differences in mean Sleep efficiency levels among the different age groups.

8. Implications and Recommendations:

Discuss the implications of the findings for diabetes management and highlight any specific recommendations for tailoring interventions based on a specific patterns observed in Sleep Efficiency levels.

This methodology provides a structured approach to analyse Sleep Efficiency variability among groups: Exercising, Gender, Smoking, Alcohol consumption, Caffeine consumption, employing ANOVA as a robust statistical tool to uncover potential differences and contribute to a more nuanced understanding of diabetes management across various age spectrums.

CHAPTER 4
PROGRAM

4. PYTHON CODE

4.1 IMPORTED MODULES

```
import numpy as np
import pandas as pd
from scipy.stats import f_oneway
import matplotlib.pyplot as plt
import seaborn as sns
```

4.2 LOADING DATASET

```
data=pd.read_csv("dataforanalysis.csv")
```

4.3 ANOVA: Sleep Efficiency vs. Exercise Efficiency

```
# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Exercise frequency', y='Sleep efficiency', data=data)
plt.xlabel('Exercise frequency')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Exercise frequency')
plt.show()
```

Performing the hypothesis test:

```
# Separate data for each Exercise Frequency
# Filter out rows with missing Exercise Frequency values
data_filtered = data.dropna(subset=['Exercise frequency', 'Sleep duration'])
# Get unique exercise categories after removing missing values
exercise_categories = sorted(data_filtered['Exercise frequency'].unique())
sleep_duration_by_exercise = [data[data['Exercise frequency'] ==
category]['Sleep efficiency'] for category in exercise_categories]
# Perform the ANOVA test
f_statistic_1, p_value_1 = f_oneway(*sleep_duration_by_exercise)
print(p_value_1)
# Define the significance level (alpha)
alpha = 0.05
# Interpret the results
if p_value_1 < alpha:
```

```
    result = "There is significant difference in Sleep efficiency based on
Exercise Frequency"
else:
    result = "There is no significant difference in Sleep efficiency based on
Exercise Frequency"
result
```

4.4 ANOVA: Sleep vs. Gender

```
# Create a box plot for Gender Vs. Sleep Efficiency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Gender', y='Sleep efficiency', data=data)
plt.xlabel('Gender')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Gender')
plt.show()
```

Performing the hypothesis test:

```
# Filter out rows with missing Gender and Sleep efficiency values
data_filtered = data.dropna(subset=['Gender', 'Sleep efficiency'])

# Get unique gender categories after removing missing values
gender_categories = sorted(data_filtered['Gender'].unique())

# Generate sleep efficiency data for each gender category
sleep_efficiency_by_gender = [
    data_filtered[data_filtered['Gender'] == category]['Sleep efficiency']
    for category in gender_categories
]

# Perform the ANOVA test
f_statistic_2, p_value_2 = f_oneway(*sleep_efficiency_by_gender)
print("P-value:", p_value_2)

# Define the significance level (alpha)
alpha = 0.05

# Interpret the results
if p_value_2 < alpha:
    result = "There is a significant difference in Sleep efficiency based on
Gender."
else:
```

```
    result = "There is no significant difference in Sleep efficiency based on
Gender."

print("Result:", result)
```

4.5 ANOVA: Sleep vs. Smoking

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Smoking status', y='Sleep efficiency', data=data)
plt.xlabel('Smoking Status')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Smoking Habit')
plt.show()
```

Performing the hypothesis test:

```
# Separate data for each group Based on smoking habits
Somking_Yes = data[data['Smoking status'] == 'Yes']['Sleep efficiency']
Smoking_No = data[data['Smoking status'] == 'No']['Sleep efficiency']

# Perform the ANOVA test
f_statistic_3, p_value_3 = f_oneway(Somking_Yes,Smoking_No)
print(p_value_3)
# Define the significance level (alpha)
alpha = 0.05

# Interpret the results
if p_value_3 < alpha:
    result = "significant"
else:
    result = "not significant"
result
```

4.6 ANOVA: Sleep vs. Alcohol Consumption

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Alcohol consumption', y='Sleep efficiency', data=data)
plt.xlabel('Alcohol consumption')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Alcohol consumption')
plt.show()
```

Performing the hypothesis test:

```
# Filter out rows with missing Alcohol consumption and Sleep efficiency values
data_filtered = data.dropna(subset=['Alcohol consumption', 'Sleep efficiency'])

# Get unique Alcohol consumption categories after removing missing values
Alcohol_consumption_categories = sorted(data_filtered['Alcohol
consumption'].unique())

# Generate sleep efficiency data for each Alcohol consumption category
sleep_efficiency_by_Alcohol_consumption = [
    data_filtered[data_filtered['Alcohol consumption'] == category]['Sleep
efficiency']
    for category in Alcohol_consumption_categories
]

# Perform the ANOVA test
f_statistic_4, p_value_4 = f_oneway(*sleep_efficiency_by_Alcohol_consumption)
print("P-value:", p_value_4)

# Define the significance level (alpha)
alpha = 0.05

# Interpret the results
if p_value_4 < alpha:
    result = "There is a significant difference in Sleep efficiency based on
Alcohol consumption."
else:
    result = "There is no significant difference in Sleep efficiency based on
Alcohol consumption."

print("Result:", result)
```

4.7 ANOVA: Sleep Efficiency vs. Caffeine Consumption

```
# Create a box plot for Sleep efficiency vs. Caffeine consumption
plt.figure(figsize=(8, 6))
sns.boxplot(x='Caffeine consumption', y='Sleep efficiency', data=data)
plt.xlabel('Caffeine consumption')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Caffeine consumption')
plt.show()
```

Performing the hypothesis test:

```
# Filter out rows with missing Caffeine consumption and Sleep efficiency values
data_filtered = data.dropna(subset=['Caffeine consumption', 'Sleep
efficiency'])

# Get unique Caffeine consumption categories after removing missing values
Caffeine_consumption_categories = sorted(data_filtered['Caffeine
consumption'].unique())

# Generate sleep efficiency data for each Caffeine consumption category
sleep_efficiency_by_Caffeine_consumption = [
    data_filtered[data_filtered['Caffeine consumption'] == category]['Sleep
efficiency']
    for category in Caffeine_consumption_categories
]

# Perform the ANOVA test
f_statistic_5, p_value_5 = f_oneway(*sleep_efficiency_by_Caffeine_consumption)
print("P-value:", p_value_5)

# Define the significance level (alpha)
alpha = 0.05

# Interpret the results
if p_value_5 < alpha:
    result = "There is a significant difference in Sleep efficiency based on
Caffeine consumption'."
else:
    result = "There is no significant difference in Sleep efficiency based on
Caffeine consumption'."

print("Result:", result)
```

Report:

```
#print("Categories: Exercise Frequecny, Gender, Smoking Status, Alchol
Consumption, Caffeine Consumption")
print("Project Report: Testing Sleep Efficiency Differences among different
Categories")
print("Categories: Exercise Frequecny, Gender, Smoking Status, Alchol
Consumption, Caffeine Consumption")
print("-------------------------------------------------------------------------
-------------------------------------------------------------------------\n"*3)
print("Category 1 :Testing Sleep Efficiency Differences among
different  Exercise Frequecies")
```

```
print("Hypotheses:")
print("Null Hypothesis (H0): There is no statistical difference in Sleep
Efficiency among the differing Exercise Frequencies.")
print("Alternative Hypothesis (Ha): There is a statistical difference in Sleep
Efficiency the differing Exercise Frequencies.")
print("")

print("Dataset Information:")
print(data.groupby('Exercise frequency')['Sleep efficiency'].describe())
print("")

# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Exercise frequency', y='Sleep efficiency', data=data)
plt.xlabel('Exercise frequency')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Exercise frequency')
plt.show()
print("")

print("Statistical Analysis:")
print("F-statistic:", f_statistic_1)
print("p-value:", p_value_1)
print("")

print("Conclusion:")
print(f"The p-value is {p_value_1:.4f}.")
print(f"Based on the significance level of {alpha}, the test is {result}.")
if p_value_1 < alpha:
    print("We reject the null hypothesis, and there is statistical difference
in Sleep Efficiency among the differing Exercise Frequencies")
else:
    print("We fail to reject the null hypothesis, and there is no statistical
difference in Sleep Efficiency among the differing Exercise Frequencies")
print("-------------------------------------------------------------------------
-------------------------------------------------------------------------\n"*3)
print("Category 2 :Testing Sleep Efficiency Differences among different
Genders")
print("Hypotheses:")
print("Null Hypothesis (H0): There is no statistical difference in Sleep
Efficiency among the Genders.")
print("Alternative Hypothesis (Ha): There is a statistical difference in Sleep
Efficiency the Genders.")
print("")
```

```
print("Dataset Information:")
print(data.groupby('Gender')['Sleep efficiency'].describe())
print("")

# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Gender', y='Sleep efficiency', data=data)
plt.xlabel('Gender')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Gender')
plt.show()
print("")

print("Statistical Analysis:")
print("F-statistic:", f_statistic_2)
print("p-value:", p_value_2)
print("")
print("Conclusion:")
print(f"The p-value is {p_value_2:.4f}.")
print(f"Based on the significance level of {alpha}, the test is {result}.")
if p_value_2 < alpha:
    print("We reject the null hypothesis, and there is statistical difference
in Sleep Efficiency among the differing Exercise Frequencies")
else:
    print("We fail to reject the null hypothesis, and there is no statistical
difference in Sleep Efficiency among the differing Exercise Frequencies")
print("----------------------------------------------------------------------
----------------------------------------------------------------------\n"*3)
print("Category 3 :Testing Sleep Efficiency Differences among different Smoking
Status")
print("Hypotheses:")
print("Null Hypothesis (H0): There is no statistical difference in Sleep
Efficiency among the differing Smoking Habits.")
print("Alternative Hypothesis (Ha): There is a statistical difference in Sleep
Efficiency the differing Smoking Habits.")
print("")

print("Dataset Information:")
print(data.groupby('Smoking status')['Sleep efficiency'].describe())
print("")

# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Smoking status', y='Sleep efficiency', data=data)
plt.xlabel('Smoking Status')
```

```
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Smoking Status')
plt.show()
print("")

print("Statistical Analysis:")
print("F-statistic:", f_statistic_3)
print("p-value:", p_value_3)
print("")
print("Conclusion:")
print(f"The p-value is {p_value_3:.4f}.")
print(f"Based on the significance level of {alpha}, the test is {result}.")
if p_value_3 < alpha:
    print("We reject the null hypothesis, and there is statistical difference
in Sleep Efficiency among  differing Smoking Habits.")
else:
    print("We fail to reject the null hypothesis, and there is no statistical
difference in Sleep Efficiency among  differing Smoking Habits.")
print("-------------------------------------------------------------------------
-------------------------------------------------------------------------\n"*3)
print("Category 4 :Testing Sleep Efficiency Differences among different Alcohol
Consumption Habits")
print("Hypotheses:")
print("Null Hypothesis (H0): There is no statistical difference in Sleep
Efficiency among the differing Alcohol Consumption Habits.")
print("Alternative Hypothesis (Ha): There is a statistical difference in Sleep
Efficiency the differing Alcohol Consumption Habits.")
print("")

print("Dataset Information:")
print(data.groupby('Alcohol consumption')['Sleep efficiency'].describe())
print("")

# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Alcohol consumption', y='Sleep efficiency', data=data)
plt.xlabel('Alcohol consumption')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Alcohol consumption')
plt.show()
print("")

print("Statistical Analysis:")
print("F-statistic:", f_statistic_4)
print("p-value:", p_value_4)
```

```
print("")
print("Conclusion:")
print(f"The p-value is {p_value_4:.4f}.")
print(f"Based on the significance level of {alpha}, the test is {result}.")
if p_value_4 < alpha:
    print("We reject the null hypothesis, and there is statistical difference
in Sleep Efficiency among the differing Alcohol Consumption Habits.")
else:
    print("We fail to reject the null hypothesis, and there is no statistical
difference in Sleep Efficiency among the differing Alcohol Consumption
Habits.")
print("-------------------------------------------------------------------------
-------------------------------------------------------------------------\n"*3)

print("Category 5 :Testing Sleep Efficiency Differences among different
Caffeine consumption")
print("Hypotheses:")
print("Null Hypothesis (H0): There is no statistical difference in Sleep
Efficiency among the differing Caffeine consumption Habits.")
print("Alternative Hypothesis (Ha): There is a statistical difference in Sleep
Efficiency the differing Caffeine consumption Habits.")
print("")

print("Dataset Information:")
print(data.groupby('Caffeine consumption')['Sleep efficiency'].describe())
print("")

# Create a box plot for Sleep Efficiency vs. Exercise frequency
plt.figure(figsize=(8, 6))
sns.boxplot(x='Caffeine consumption', y='Sleep efficiency', data=data)
plt.xlabel('Caffeine consumption')
plt.ylabel('Sleep efficiency')
plt.title('Sleep Quality By Caffeine consumption')
plt.show()
print("")

print("Statistical Analysis:")
print("F-statistic:", f_statistic_5)
print("p-value:", p_value_5)
print("")
print("Conclusion:")
print(f"The p-value is {p_value_5:.4f}.")
print(f"Based on the significance level of {alpha}, the test is {result}.")
if p_value_5 < alpha:
```

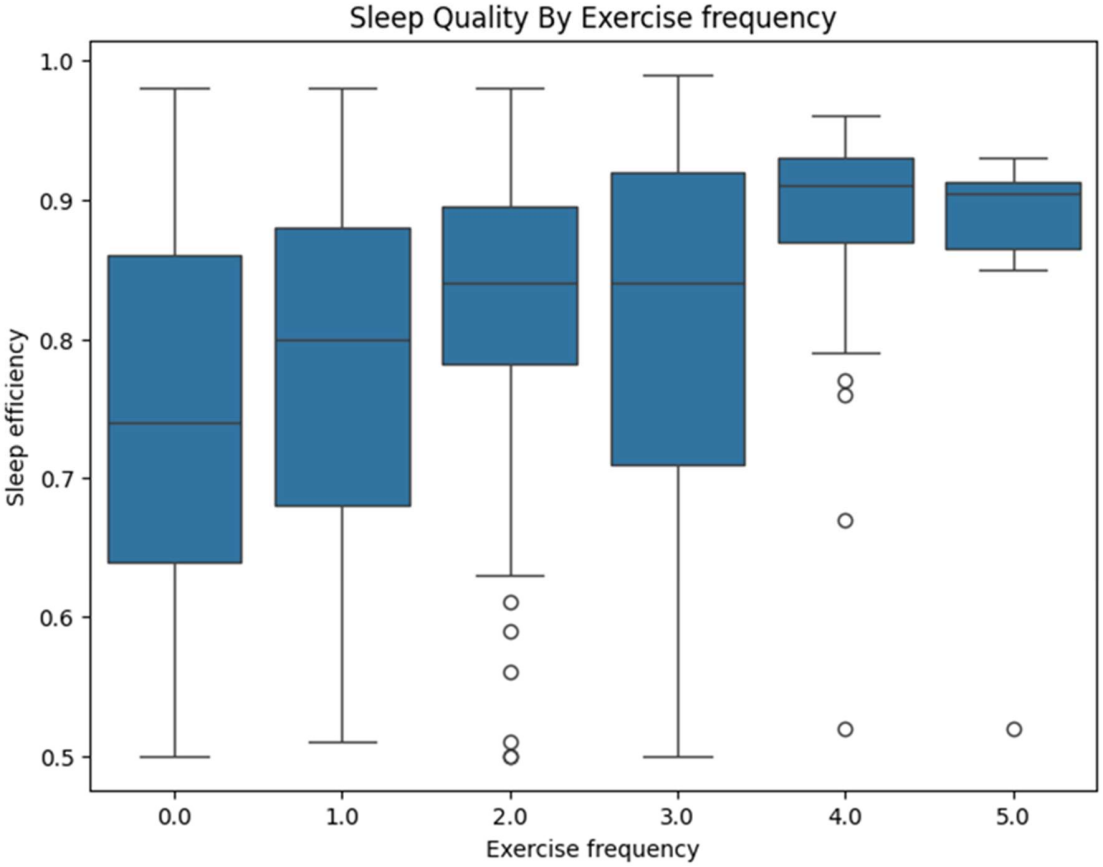
```
    print("We reject the null hypothesis, and there is statistical difference
in Sleep Efficiency among the differing Caffeine consumption Habits.")
else:
    print("We fail to reject the null hypothesis, and there is no statistical
difference in Sleep Efficiency among the differing Caffeine consumption
Habits.")
print("--------------------------------------------------------------------------
------------------------------------------------------------------------\n"*3)
```

CHAPTER 5

RESULTS AND DISCUSSIONS WITH FIGURES

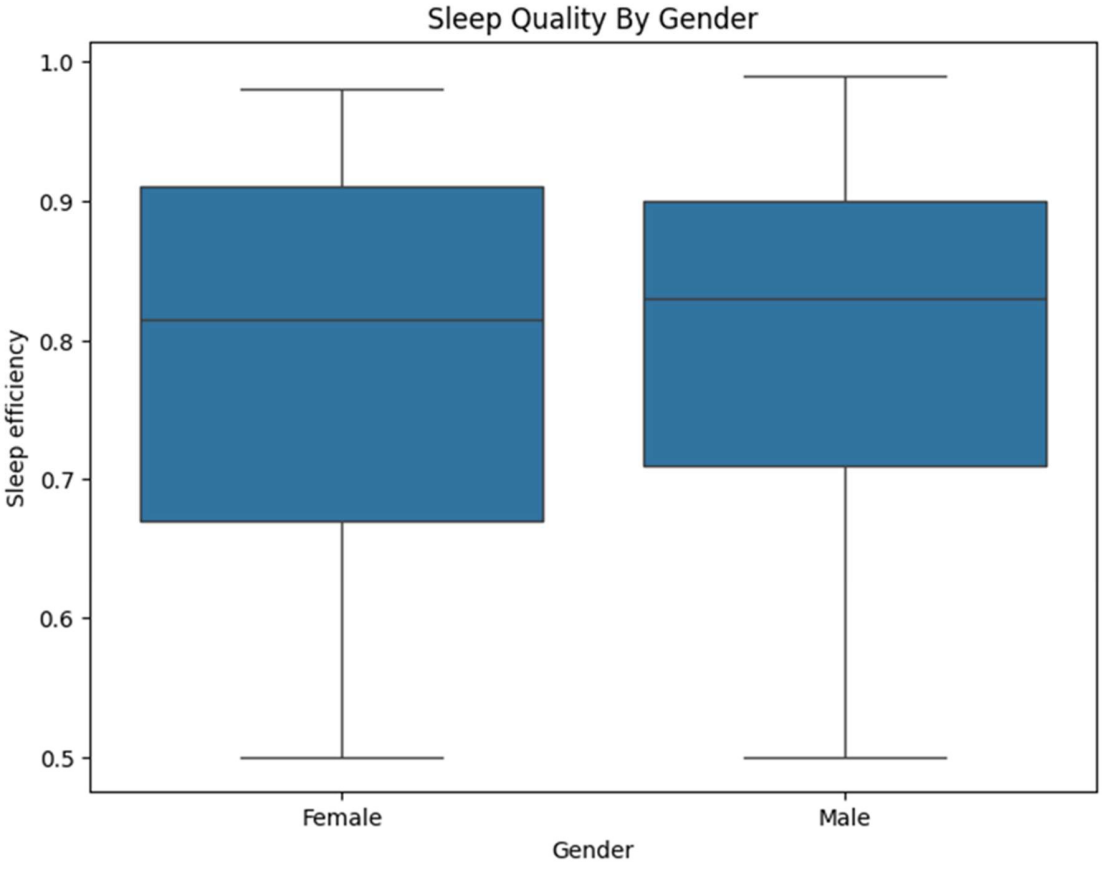
5.1 ANOVA: Sleep Efficiency vs. Exercise Efficiency

Output:



Hypothesis Test Result:
3.345181657726702e-07
'There is significant difference in Sleep efficiency based on Exercise Frequency'

5.2 ANOVA: Sleep Efficiency vs. Gender

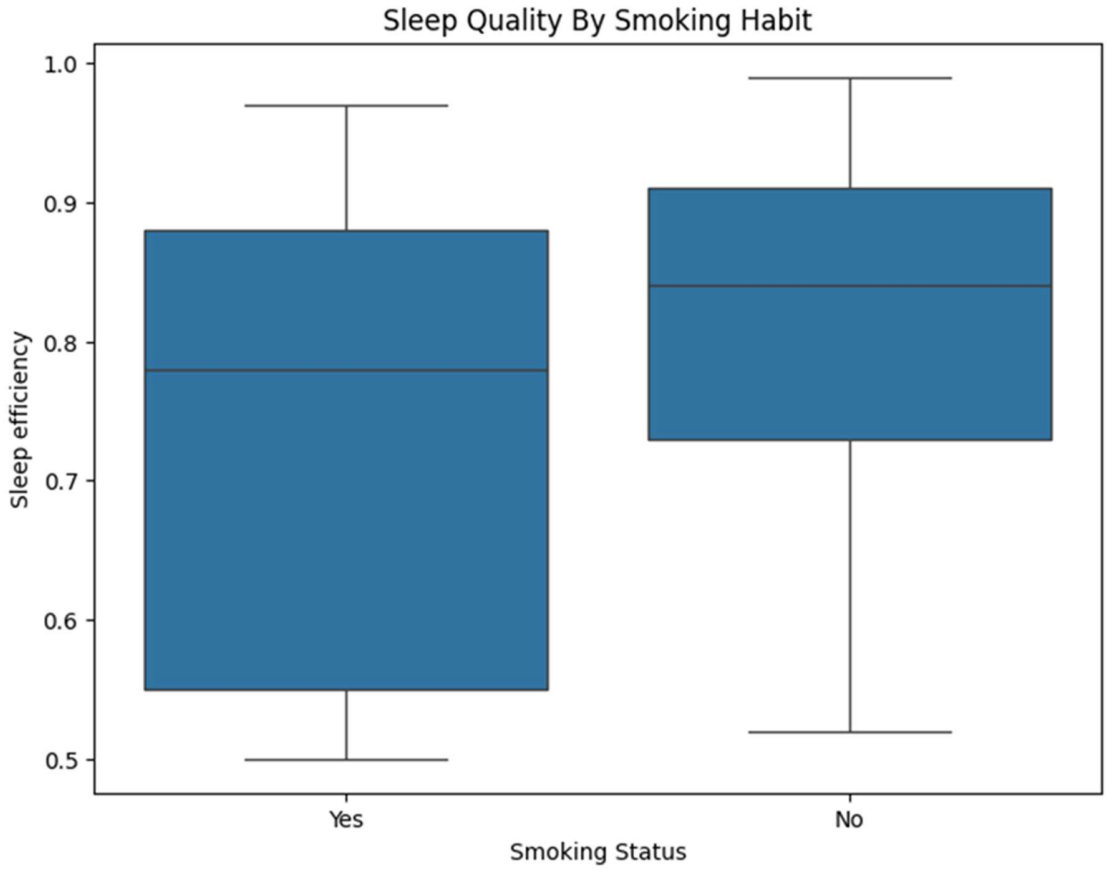


Hypothesis testing Result:

P-value: 0.831071888458071
Result: There is no significant difference in Sleep efficiency based on Gender.

5.3 ANOVA: Sleep Efficiency vs. Smoking

Output:

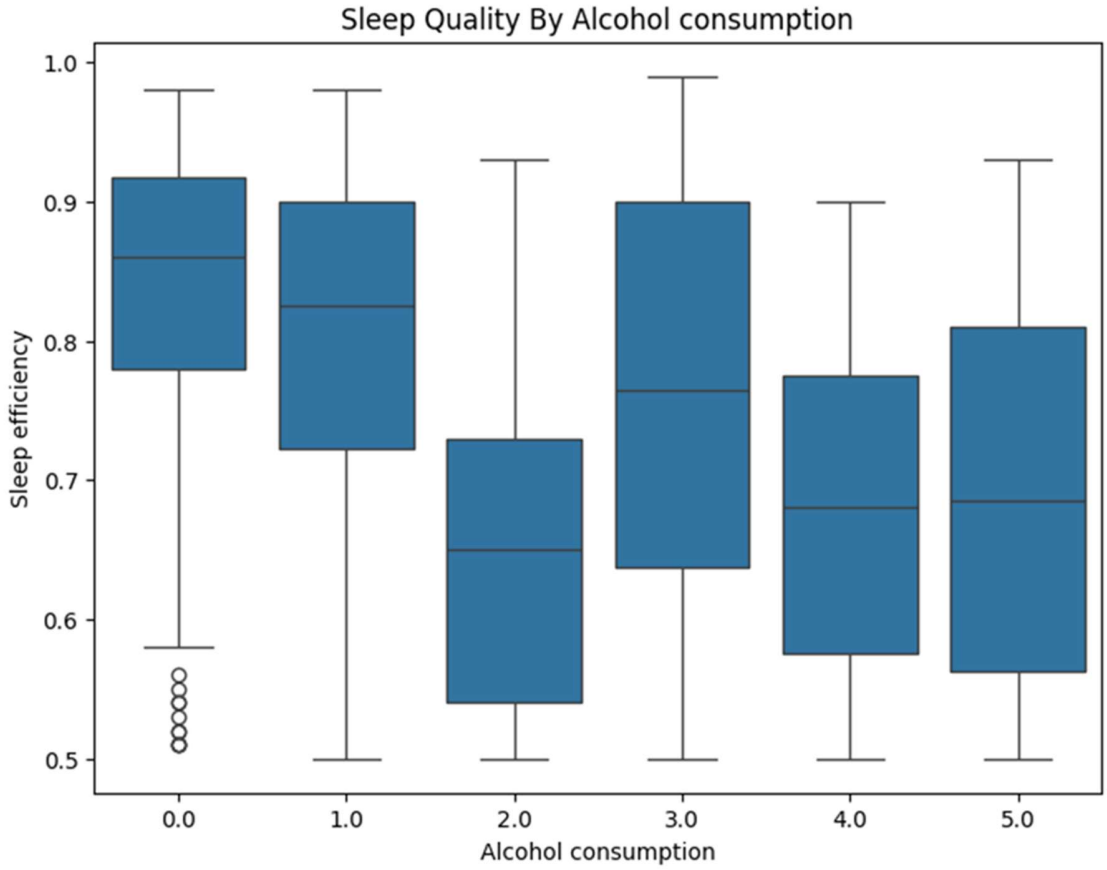


Hypothesis Testing Result:

3.288310567776604e-10
'significant'

5.4 ANOVA: Sleep Efficiency vs. Alcohol Consumption

Output:



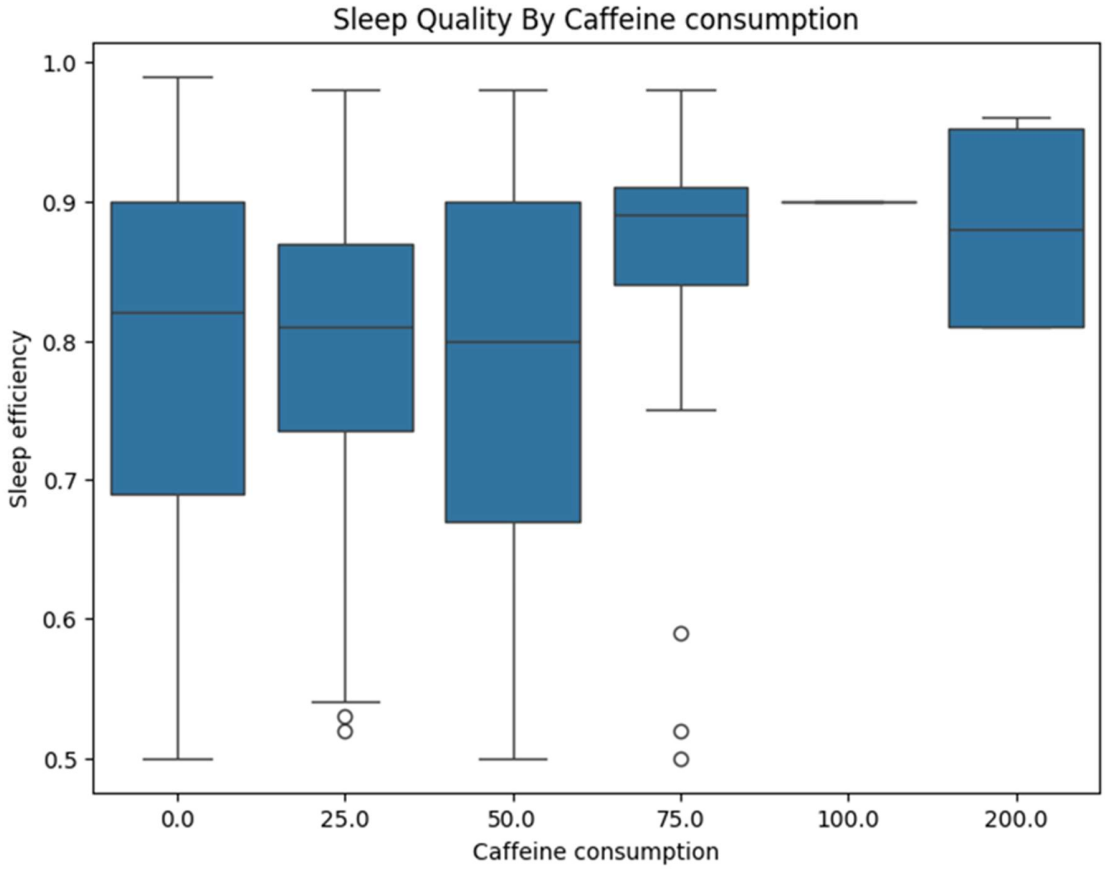
Hypothesis Testing Results:

P-value: 5.397864814571545e-19

Result: There is a significant difference in Sleep efficiency based on Alcohol consumption.

5.5 ANOVA: Sleep Efficiency vs. Caffeine Consumption

Output:



Hypothesis Testing Result:

P-value: 0.09672345209472184

Result: There is no significant difference in Sleep efficiency based on Caffeine consumption'.

5.6 Project Report:

Project Report: Testing Sleep Efficiency Differences among different Categories
Categories: Exercise Frequency, Gender, Smoking Status, Alcohol Consumption, Caffeine Consumption

---

Category 1 :Testing Sleep Efficiency Differences among different Exercise Frequencies
Hypotheses:
Null Hypothesis (H0): There is no statistical difference in Sleep Efficiency among the differing Exercise Frequencies.
Alternative Hypothesis (Ha): There is a statistical difference in Sleep Efficiency the differing Exercise Frequencies.

Dataset Information:

	count	mean	std	min	25%	50%	75%	\
Exercise frequency								
0.0	116.0	0.743879	0.136883	0.50	0.6400	0.740	0.8600	
1.0	97.0	0.774124	0.130948	0.51	0.6800	0.800	0.8800	
2.0	54.0	0.810000	0.121810	0.50	0.7825	0.840	0.8950	
3.0	130.0	0.799154	0.138195	0.50	0.7100	0.840	0.9200	
4.0	41.0	0.881463	0.085163	0.52	0.8700	0.910	0.9300	
5.0	8.0	0.851250	0.136428	0.52	0.8650	0.905	0.9125	

	max
Exercise frequency	
0.0	0.98
1.0	0.98
...	
3.0	0.99
4.0	0.96
5.0	0.93

Statistical Analysis:
F-statistic: 7.966486458712289
p-value: 3.345181657726702e-07

Conclusion:
The p-value is 0.0000.
Based on the significance level of 0.05, the test is There is no significant difference in Sleep efficiency based on Caffeine consumption'..
We reject the null hypothesis, and there is statistical difference in Sleep Efficiency among the differing Exercise Frequencies

---

Category 2: Testing Sleep Efficiency Differences among different Genders
Hypotheses:
Null Hypothesis (H0): There is no statistical difference in Sleep Efficiency among the Genders.
Alternative Hypothesis (Ha): There is a statistical difference in Sleep Efficiency the Genders.

Dataset Information:

	count	mean	std	min	25%	50%	75%	max
Gender								
Female	224.0	0.787545	0.141965	0.5	0.67	0.815	0.91	0.98
Male	228.0	0.790263	0.128584	0.5	0.71	0.830	0.90	0.99

Statistical Analysis:
F-statistic: 0.045561322468459745
p-value: 0.831071888458071

Conclusion:
The p-value is 0.8311.
Based on the significance level of 0.05, the test is There is no significant difference in Sleep efficiency based on Caffeine consumption'.
We fail to reject the null hypothesis, and there is no statistical difference in Sleep Efficiency among the differing Exercise Frequencies

---

Category 3 :Testing Sleep Efficiency Differences among different Smoking Status
Hypotheses:

Null Hypothesis (H0): There is no statistical difference in Sleep Efficiency among the differing Smoking Habits.
Alternative Hypothesis (Ha): There is a statistical difference in Sleep Efficiency the differing Smoking Habits.

Dataset Information:

	count	mean	std	min	25%	50%	75%	max
Smoking status								
No	298.0	0.817081	0.109384	0.52	0.73	0.84	0.91	0.99
Yes	154.0	0.734416	0.161711	0.50	0.55	0.78	0.88	0.97

Statistical Analysis:
F-statistic: 41.32804631786465
p-value: 3.288310567776604e-10

Conclusion:
The p-value is 0.0000.
Based on the significance level of 0.05, the test is There is no significant difference in Sleep efficiency based on Caffeine consumption'..
We reject the null hypothesis, and there is statistical difference in Sleep Efficiency among differing Smoking Habits.

Category 4 :Testing Sleep Efficiency Differences among different Alcohol Consumption Habits
Hypotheses:
Null Hypothesis (H0): There is no statistical difference in Sleep Efficiency among the differing Alcohol Consumption Habits.
Alternative Hypothesis (Ha): There is a statistical difference in Sleep Efficiency the differing Alcohol Consumption Habits.

Dataset Information:

	count	mean	std	min	25%	50%	75% \
Alcohol consumption							
0.0	246.0	0.832033	0.109693	0.51	0.7800	0.860	0.9175
1.0	54.0	0.803889	0.122296	0.50	0.7225	0.825	0.9000
2.0	37.0	0.671351	0.134602	0.50	0.5400	0.650	0.7300
3.0	48.0	0.750417	0.152803	0.50	0.6375	0.765	0.9000

...
3.0 0.99
4.0 0.90
5.0 0.93

Statistical Analysis:
F-statistic: 21.22248558048939
p-value: 5.397864814571545e-19

Conclusion:
The p-value is 0.0000.
Based on the significance level of 0.05, the test is There is no significant difference in Sleep efficiency based on Caffeine consumption'..
We reject the null hypothesis, and there is statistical difference in Sleep Efficiency among the differing Alcohol Consumption Habits.

Category 5 :Testing Sleep Efficiency Differences among different Caffeine consumption
Hypotheses:
Null Hypothesis (H0): There is no statistical difference in Sleep Efficiency among the differing Caffeine consumption Habits.
Alternative Hypothesis (Ha): There is a statistical difference in Sleep Efficiency the differing Caffeine consumption Habits.

Dataset Information:

Caffeine consumption	count	mean	std	min	25%	50%	75%	\
0.0	211.0	0.786209	0.137449	0.50	0.690	0.82	0.9000	
25.0	79.0	0.799114	0.114815	0.52	0.735	0.81	0.8700	
50.0	107.0	0.771589	0.143508	0.50	0.670	0.80	0.9000	
75.0	25.0	0.845600	0.127413	0.50	0.840	0.89	0.9100	

...

75.0 0.98
100.0 0.90
200.0 0.96

Statistical Analysis:
F-statistic: 1.879395768792106
p-value: 0.09672345209472184

Conclusion:
The p-value is 0.0967.
Based on the significance level of 0.05, the test is There is no significant difference in Sleep efficiency based on Caffeine consumption'..
We fail to reject the null hypothesis, and there is no statistical difference in Sleep Efficiency among the differing Caffeine consumption Habits.

CHAPTER 6
CONCLUSION

Exercise Frequency: Higher exercise frequency showed a significant link with improved sleep efficiency.

2. Gender: Interestingly, gender didn't appear to significantly influence sleep efficiency based on this dataset.

3. Smoking Status: Non-smokers exhibited notably higher sleep efficiency compared to smokers, indicating a significant impact of smoking habits on sleep quality.

4. Alcohol Consumption: Sleep efficiency showed variations based on alcohol consumption, with lower consumption associated with better sleep efficiency.

5. Caffeine Consumption: While no significant difference was observed in sleep efficiency concerning caffeine intake, a trend hinted that higher intake might slightly impact sleep efficiency.

Conclusion:
The study highlights the substantial impact of lifestyle choices like exercise, smoking, and alcohol consumption on sleep quality. Although gender and caffeine intake didn't emerge as major influencers in this dataset, exercise frequency, smoking status, and alcohol consumption notably affected sleep efficiency.

These findings stress the importance of lifestyle modifications in potentially enhancing sleep quality. They underscore the potential benefits of regular exercise, reducing or quitting smoking, and moderating alcohol intake for improved sleep patterns. This study lays a foundation for further research and suggests tailored interventions to enhance sleep quality by addressing specific lifestyle factors.

REFERENCES

1. https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/types-of-correlation.html

2. https://projecteuclid.org/journals/statistical-science/volume-4/issue-2/Francis-Galtons-Account-of-the-Invention-of-Correlation/10.1214/ss/1177012580.full?tab=ArticleFirstPage

3. https://seaborn.pydata.org/examples/index.html

4. https://www.geeksforgeeks.org/how-to-perform-a-one-way-anova-in-python/

5. https://www.javatpoint.com/anova-test-in-python

6. https://towardsdatascience.com/anova-test-with-python-cfbf4013328b

7. https://analyticsindiamag.com/a-complete-python-guide-to-anova/

8. https://rowannicholls.github.io/python/statistics/hypothesis_testing/anova.html

9. https://www.statology.org/one-way-anova-python/