**National University of Computer and Emerging Sciences, Lahore Campus**



**MG2009 – Data Analysis for Business-II**

**Final Project Report**

**Submitted To:**

Ms. Aroosa Safdar

**Submitted By:**

Amna Azfar (23L-5047)

**BSBA:**

3A

### **Abstract**

This project investigates patterns in gym attendance and exercise behavior using a dataset comprising gym members' demographic and activity records. The study aims to identify significant trends and relationships among variables such as age, gender, membership type, and exercise frequency. Through the application of descriptive and inferential statistical techniques in IBM SPSS, this analysis seeks to uncover key factors that influence gym utilization and member engagement.

The data analysis encompasses various methods, including descriptive statistics to summarize demographic profiles and visual representations to illustrate exercise patterns. Inferential techniques such as t-tests, ANOVA, and correlation analysis are employed to test hypotheses related to demographic impacts and exercise trends. The goal is to derive actionable insights that can inform gym management strategies and improve member retention and satisfaction.

The results are expected to contribute to a deeper understanding of how different factors interact with gym attendance and exercise routines, aiding in the development of targeted programs and marketing approaches. This project emphasizes data-driven decision-making, underscoring the importance of comprehensive data analysis in the health and fitness industry.

### **Acknowledgement**

I would like to express my sincere gratitude to everyone who supported me during the completion of this data analysis project. First and foremost, I extend my deepest appreciation to my course instructor, Miss Aroosa Safdar, for her invaluable guidance, encouragement, and insights throughout the project. Her expertise and mentorship have been instrumental in shaping my understanding of statistical analysis and ensuring the quality of my work.

I am also thankful to my group members for their dedication, teamwork, and contributions at every stage of the project. Their collaborative efforts, thoughtful discussions, and unwavering support played a crucial role in completing this analysis successfully. The shared commitment and exchange of ideas among our group made the project experience both productive and enriching.

Finally, I would like to express my gratitude to FAST NUCES for fostering an environment that encourages academic exploration and excellence. The facilities, resources, and academic community at the university have greatly contributed to the successful completion of this project.

Thank you to everyone who played a role in making this project possible.

**Table of Contents**

**1. Introduction ---------------------------------------------------------------------------------------------------------------7**

1.1 Objectives of the Study -------------------------------------------------------------------------------------------8

2. Data Collection and Methodology ------------------------------------------------------------------------------------9

**3. Analysis and Results ---------------------------------------------------------------------------------------------------11**

3.1. Descriptive Statistics ---------------------------------------------------------------------------------------------12

3.2. Inferential Statistics ----------------------------------------------------------------------------------------------27

**4. Conclusion ---------------------------------------------------------------------------------------------------------------48**

**5. Limitations ---------------------------------------------------------------------------------------------------------------49**

**6. References ---------------------------------------------------------------------------------------------------------------50**

**7. Appendix -----------------------------------------------------------------------------------------------------------------51**

**List of Figures**

[Figure 1 11](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423086)

[Figure 2 12](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423087)

[Figure 3 13](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423088)

[Figure 4 14](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423089)

[Figure 5 15](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423090)

[Figure 6 16](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423091)

[Figure 7 17](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423092)

[Figure 8 19](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423093)

[Figure 9 21](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423094)

[Figure 10 23](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423095)

[Figure 11 25](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423096)

[Figure 12 25](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423097)

[Figure 13 27](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423098)

[Figure 14 27](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423099)

[Figure 15 29](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423100)

[Figure 16 29](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423101)

[Figure 17 31](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423102)

[Figure 18 31](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423103)

[Figure 19 33](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423104)

[Figure 20 33](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423105)

[Figure 21 36](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423106)

[Figure 22 38](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423107)

[Figure 23 41](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423108)

[Figure 24 41](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423109)

[Figure 25 42](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423110)

[Figure 26 42](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423111)

[Figure 27 46](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423112)

**List of Tables**

[Table 1 12](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423248)

[Table 2 14](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423249)

[Table 3 16](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423250)

[Table 4 17](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423251)

[Table 5 18](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423252)

[Table 6 19](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423253)

[Table 7 19](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423254)

[Table 8 21](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423255)

[Table 9 21](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423256)

[Table 10 23](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423257)

[Table 11 23](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423258)

[Table 12 25](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423259)

[Table 13 25](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423260)

[Table 14 27](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423261)

[Table 15 29](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423262)

[Table 16 31](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423263)

[Table 17 33](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423264)

[Table 18 35](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423265)

[Table 19 36](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423266)

[Table 20 38](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423267)

[Table 21 40](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423268)

[Table 22 40](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423269)

[Table 23 40](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423270)

[Table 24 44](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423271)

[Table 25 45](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423272)

[Table 26 46](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423273)

[Table 27 47](file:///C:\Users\user1\Downloads\DA%20Final%20Project.docx#_Toc184423274)

#### **Introduction**

This project focuses on analyzing gym attendance and exercise behavior among members. As the health and fitness industry continues to grow, data analytics play a pivotal role in understanding member behavior and optimizing services. The primary aim of this study is to explore how demographic factors such as age, gender, and membership type influence gym attendance and exercise patterns. This information is essential for gyms seeking to enhance member satisfaction, improve retention rates, and develop targeted fitness programs.

A critical issue faced by fitness centers is ensuring consistent member engagement. Despite various membership plans and exercise options, many gyms struggle with low attendance and high dropout rates, which can affect operational efficiency and financial performance. This project addresses these challenges by investigating the relationships between key variables and exercise behavior, aiming to identify patterns that can guide strategic decisions..

The study is structured around research questions designed to probe the association between demographic characteristics and exercise habits. Questions include whether age impacts the frequency of gym visits, if gender differences exist in exercise patterns, and whether membership type influences gym attendance. These questions lead to hypotheses such as: "There is a significant difference in exercise frequency between genders" and "Membership type significantly impacts gym attendance."

The dataset for this analysis is derived from gym member records and includes variables such as age, gender, membership type, exercise frequency, and session duration. By employing IBM SPSS for data analysis, the project utilizes descriptive statistics to summarize data trends, as well as inferential tests like t-tests, ANOVA, and correlation analysis to validate the hypotheses. These methods will not only reveal key insights but also provide evidence-based recommendations for improving member engagement and optimizing gym services.

Overall, the project combines research, statistical analysis, and practical implications to offer comprehensive insights into gym member behavior, contributing valuable information that can support data-driven decision-making in the fitness industry.

### **Objectives of the Study**

The primary objectives of this study are as follows:

* **Analyze patterns in gym attendance and exercise frequency** among members to identify trends.
* **To apply statistical inference and data analysis techniques** to investigate relationships or trends in gym member data
* **Examine the impact of demographic factors** (e.g., age, gender) on exercise behavior and participation.
* **Test hypotheses regarding demographic differences in exercise patterns** to validate significant relationships.
* **Provide actionable insights for gym management** to enhance member engagement and develop targeted programs.
* **To evaluate the fitness levels of members** based on variables like heart rate, workout type, and experience level.
* **To determine factors influencing key outcomes** such as calories burned, workout frequency, and BMI.

These objectives will inform strategic decisions to improve member retention and optimize services.

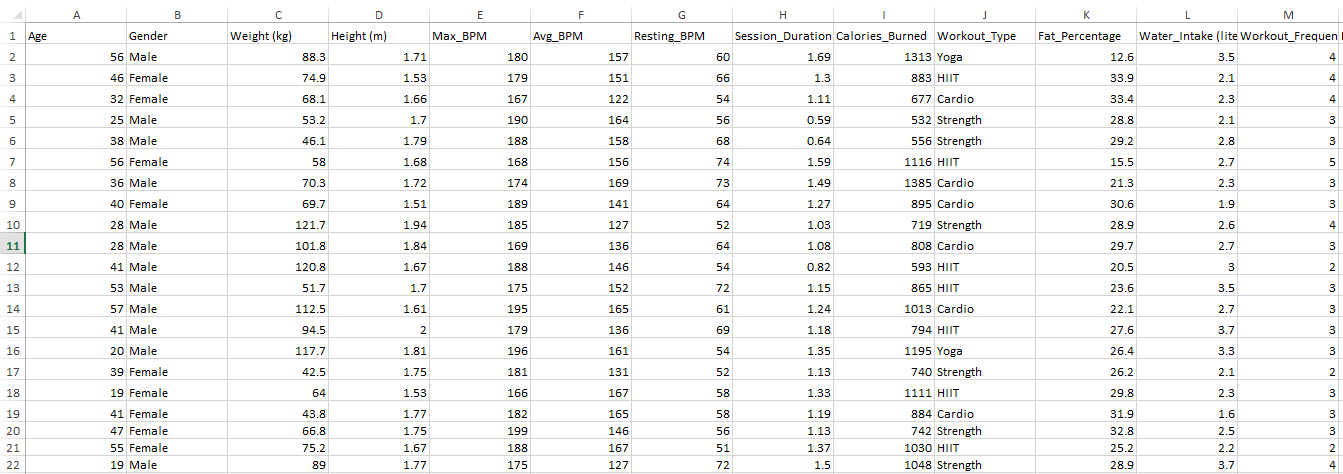
1. **Data Collection & Methodology**

The dataset used for this analysis was sourced from Kaggle, a platform known for hosting a wide range of high-quality datasets. It contains detailed records of 973 gym members, capturing various aspects of their fitness journeys. The dataset includes demographic information such as age and gender, alongside physical attributes like weight, height, and BMI. Additionally, it provides data on cardiovascular metrics, including maximum, average, and resting BPM (beats per minute), as well as workout-related details such as session duration, calories burned, workout type, and frequency of exercise sessions per week. Health metrics like fat percentage and water intake are also included, along with members' experience levels in fitness activities.

This dataset was chosen due to its comprehensive nature and the absence of missing values, making it suitable for direct analysis without extensive preprocessing. The analysis focuses on identifying trends and correlations between variables such as workout frequency, caloric expenditure, and BMI to understand members' fitness behaviors. Using this data, we aim to generate actionable insights into gym members' health and exercise patterns.

**LIST OF ABBREVIATIONS**

* + **Age** - Age of the gym member (in years).
  + **Gender** - Gender of the gym member (Male/Female).
  + **Weight (kg)** - Body weight of the member (in kilograms).
  + **Height (m)** - Height of the member (in meters).
  + **Max\_BPM** - Maximum Beats per Minute recorded during a workout session.
  + **Avg\_BPM** - Average Beats per Minute recorded during a workout session.
  + **Resting\_BPM** - Resting Beats per Minute, measured while the member is at rest.
  + **Session\_Duration (hours)** - Duration of the workout session (in hours).
  + **Calories\_Burned** - Total calories burned during a workout session.
  + **Workout\_Type** - Type of workout performed (e.g., cardio, strength training).
  + **Fat\_Percentage** - Body fat percentage of the member.
  + **Water\_Intake (liters)** - Amount of water consumed by the member (in liters).
  + **Workout\_Frequency (days/week)** - Number of days the member works out per week.
  + **Experience\_Level** - Fitness experience level, typically on a scale (e.g., beginner, intermediate, advanced).
  + **BMI** - Body Mass Index, calculated as weight (kg) divided by height (m²).

****

1. **ANALYSIS & RESULTS**

### The **Analysis & Results** section provides a comprehensive examination of the gym member dataset to uncover trends and relationships. This part of the report details the statistical methods used and presents findings in a structured manner.

### **3.1. Descriptive Statistics**

This section summarizes key demographic and exercise variables, showing distributions through frequencies, percentages, and cumulative frequencies. Measures of central tendency (mean, median) and dispersion (standard deviation) provide insight into typical values and variability. Cross-tabulation compares groups (e.g., gender differences in exercise frequency), while visual tools like bar charts and histograms depict data trends. Normality tests, including the Shapiro-Wilk test, assess distribution shapes, with skewness and kurtosis commented on. Outliers are identified using boxplots and z-scores, and data normalization is considered if necessary.

### **3.2. Inferential Statistics**

Inferential analysis includes hypothesis testing to validate relationships. Independent t-tests compare exercise means between groups (e.g., gender), and ANOVA assesses differences across membership types. Correlation analysis explores links between continuous variables like age and exercise duration. SPSS output tables are used to interpret p-values and confidence intervals, guiding decisions to reject or accept hypotheses. Results are discussed in the context of the study's objectives, providing key insights into gym member behavior.

### **Descriptive Statistics**

### For Categorical Variables:

* **Frequencies and percentages**

1. **Gender :**

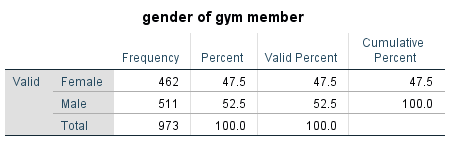


Table 1

* **Plots:**
* **Bar Chart**

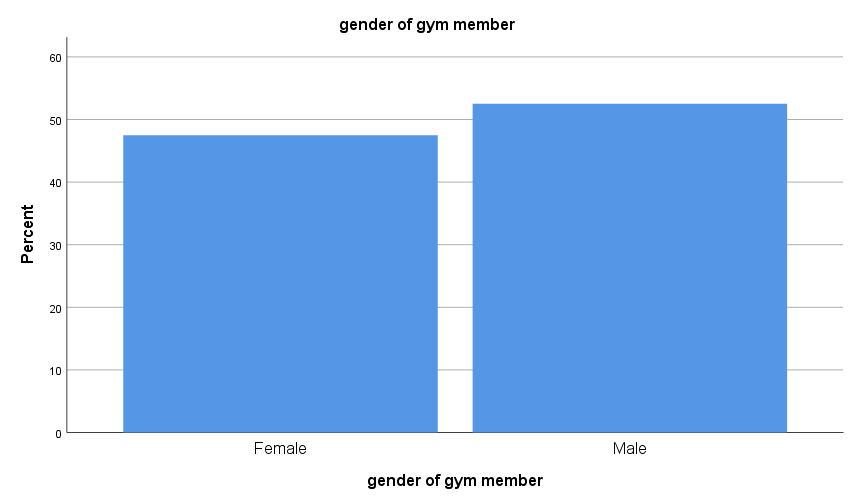
****

Figure 1

* **Pie Chart**

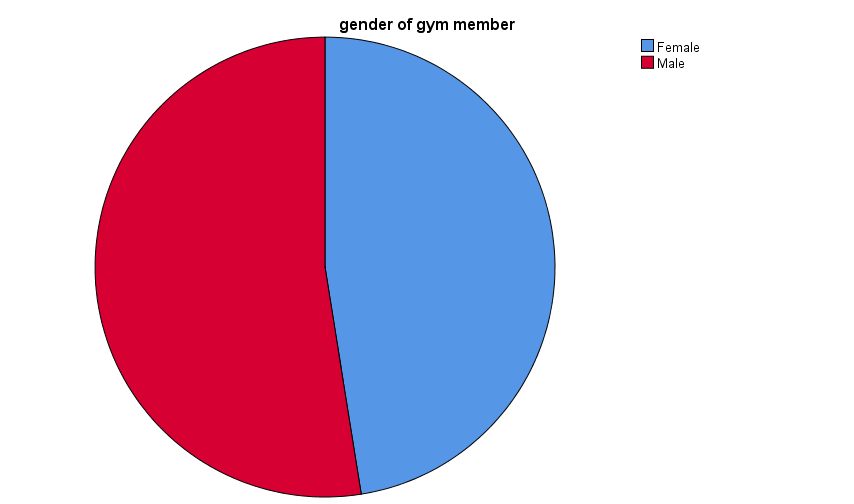


Figure 2

**Interpretation:**

The Gym membership is nearly evenly split between males and females, with each gender comprising about 50% of the total which means that the data shows a balanced distribution among both genders.

1. **Workout Type :**

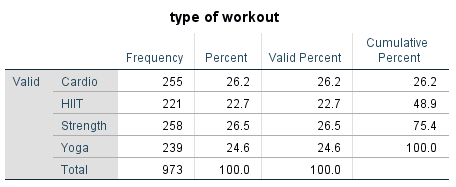


Table 2

* **Plots:**
* **Bar Chart**

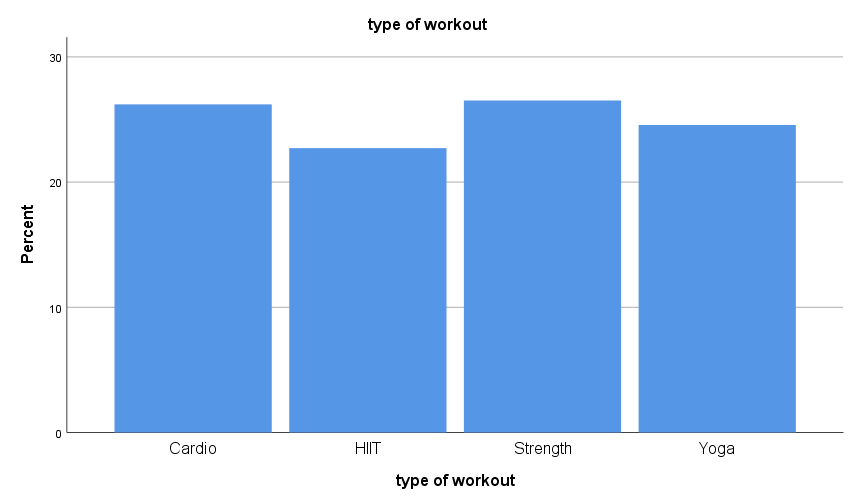
****

Figure 3

* **Pie Chart**

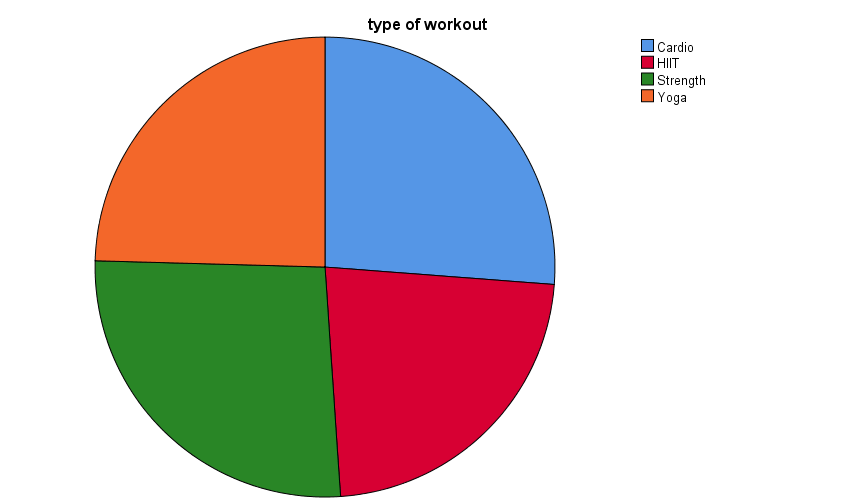
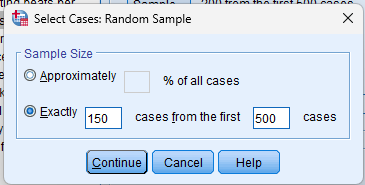
****

Figure 4

**Interpretation:**

The data shows that Strength (26.5%) is the most popular workout type, followed closely by Cardio (26.2%) and Yoga (24.6%), while HIIT (22.7%) is the least popular. Together, over 75% of participants prefer Strength, Cardio, or HIIT, with Yoga completing the total. The distribution indicates fairly balanced preferences, with a slight preference for Strength.

1. **Level of Experience through Random Sampling :**

****

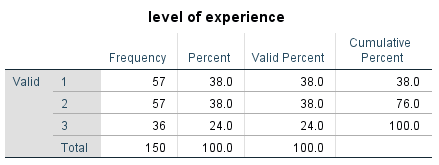
****

Table 3

* **Plots:**

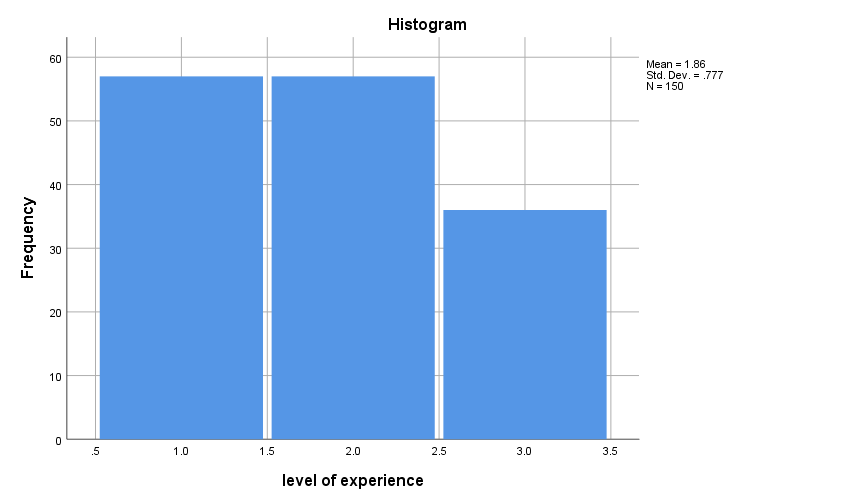
****

Figure 5

**Interpretation:**

The Data shows that most respondents have either 1 or 2 years of experience which is 38% of all, while 24% of Gym members have 3 years of experience.

* **Cross-tabulation**

1. **Gender vs. Type of Workout :**

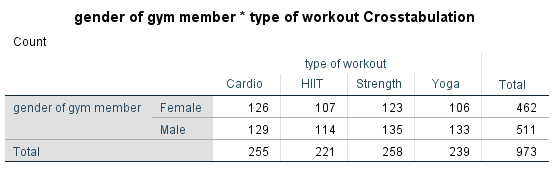
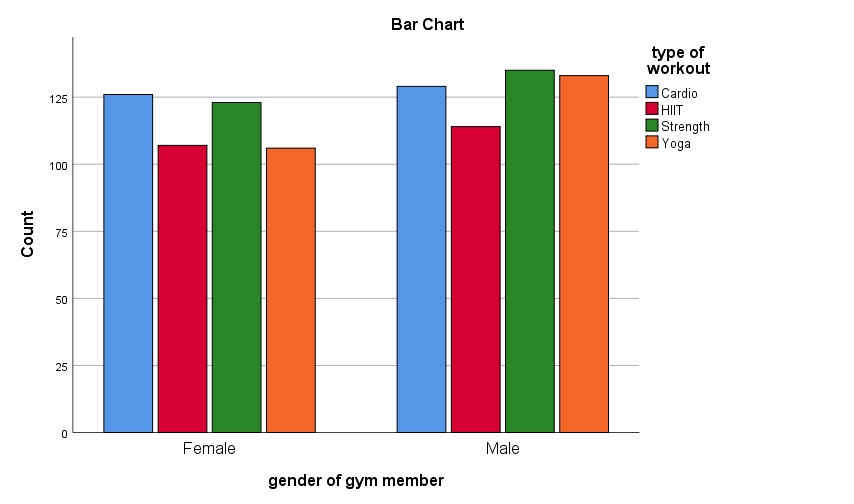
****

Table 4

* **Plots:**



‘

Figure 6

**Interpretation:**

The data shows the relationship between gym members' gender and their preferred workout types. Overall, there are slightly more male members (511) than female members (462). Among all workout types, Strength is the most popular, followed by Cardio, Yoga, and HIIT. For females, Cardio and Strength are almost equally preferred, with HIIT and Yoga being slightly less popular. Males favor Strength the most, followed closely by Cardio and Yoga, while HIIT is the least preferred. This suggests that while both genders show interest in a variety of workouts, males lean more towards Strength, and females have a balanced preference for Cardio and Strength.

1. **Gender vs. Experience Level**

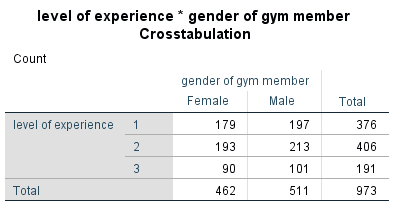


Table 5

* **Plots:**

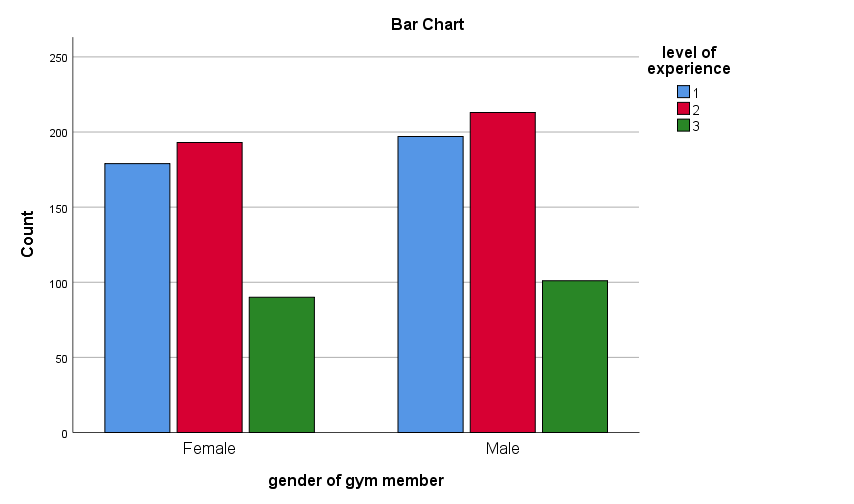


Figure 7

**Interpretation:**

The comparison of data shows that there are more Male members than female members. There are 462 females and 511 Males and the most common level of experience among both is 2 years.

For Continuous Variables:

* **Age**

1. **Central tendency and Dispersion :**

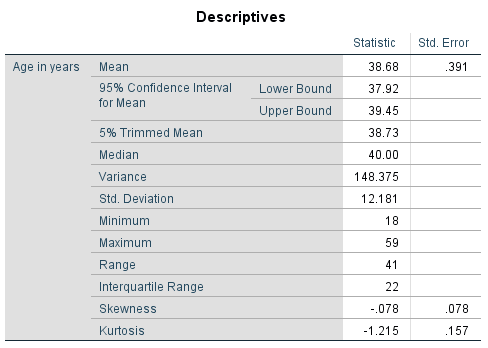
****

Table 6

1. **Boxplot and Outliers :**

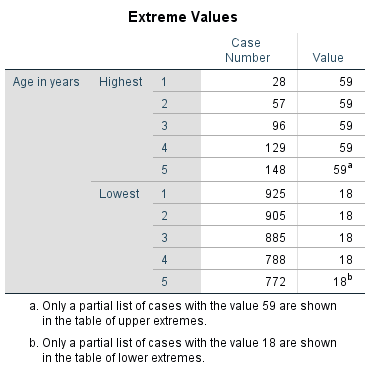
****

Table 7

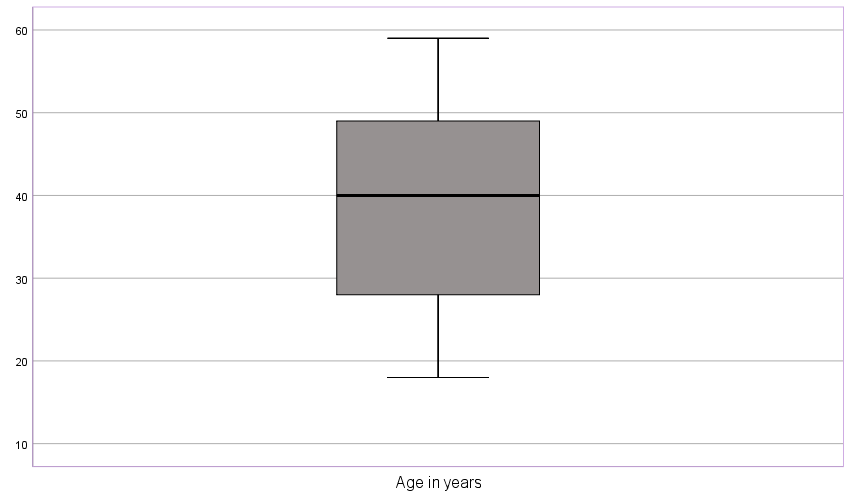
****

Figure 8

**Interpretation:**

The table provides descriptive statistics for the ages of gym members, revealing key insights into the distribution. The average age of members is 38.68 years, with a 95% confidence interval indicating that the true mean likely falls between 37.92 and 39.45 years. The 5% trimmed mean, slightly higher at 38.73 years, suggests minimal influence from outliers. The median age is 40 years, indicating a nearly symmetrical distribution, as the mean is close to the median. The variance of 148.375 and a standard deviation of 12.181 years reflect moderate variability in the ages of gym members.

The age range spans from 18 to 59 years, covering a difference of 41 years, while the interquartile range (IQR) of 22 years shows that the middle 50% of ages are relatively concentrated. A slight negative skewness of -0.078 suggests the data is nearly symmetrical, with a small tendency toward younger ages. Additionally, the negative kurtosis value of -1.215 indicates a relatively flat distribution compared to a normal curve. Overall, the statistics highlight a diverse age group among gym members, with most ages centered around the average.

* **Workout Frequency**

1. **Central tendency and Dispersion :**

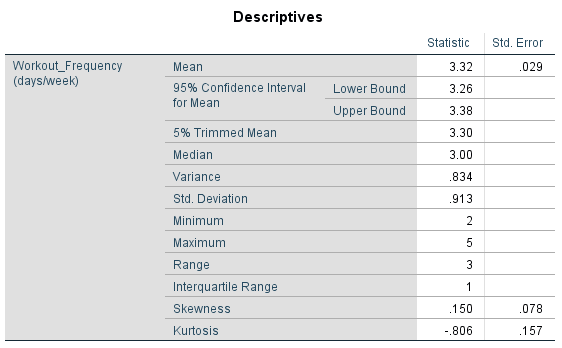
****

Table 8

1. **Boxplot and Outliers :**

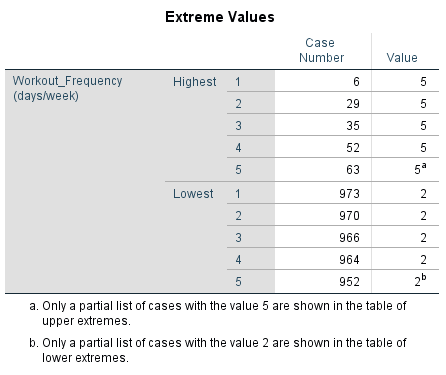
****

Table 9

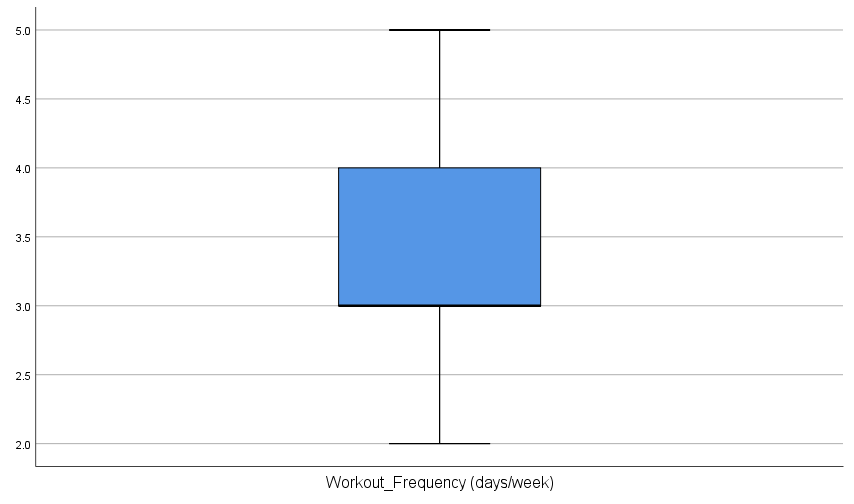
****

Figure 9

**Interpretation:**

The table provides descriptive statistics for "Workout Frequency" in days per week. The mean frequency is 3.32 days, with a 95% confidence interval suggesting the true mean lies between 3.26 and 3.38 days. The median frequency is 3 days, closely aligning with the mean, while the 5% trimmed mean of 3.30 days indicates minimal impact from extreme values.

The variability in workout frequency is low, with a standard deviation of 0.913 and a variance of 0.834. The minimum reported frequency is 2 days per week, and the maximum is 5 days, resulting in a range of 3 days. The interquartile range (IQR) is 1 day, showing that the middle 50% of members have workout frequencies concentrated within a narrow range.

The skewness value of 0.150 suggests a slight positive skew, indicating a minor tilt toward higher workout frequencies. The kurtosis value of -0.806 reflects a flatter distribution compared to a normal curve, indicating fewer extreme values.

The boxplots highlight few outliers with exceptionally high or low workout frequencies. This reflects that most members maintain moderate engagement, with extreme cases being rare.

In summary, the data indicates that most gym members follow a consistent workout routine, typically exercising 3 to 4 days per week.

* **Session Duration**

1. **Central tendency and Dispersion :**

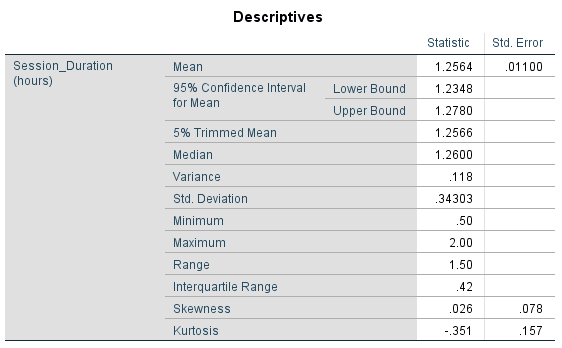
****

Table 10

1. **Boxplot and Outliers :**

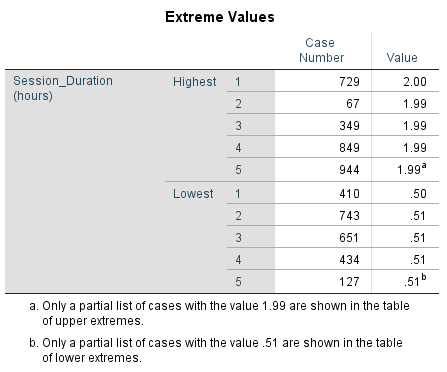


Table 11

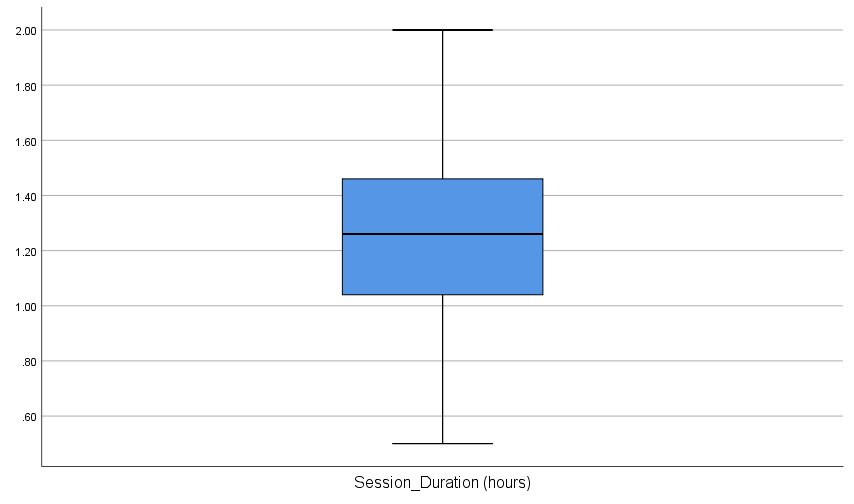
****

Figure 10

**Interpretation:**

The table summarizes the descriptive statistics for "Session Duration" in hours. The mean session duration is 1.2564 hours, with a 95% confidence interval indicating that the true mean likely falls between 1.2348 and 1.2780 hours. The 5% trimmed mean is 1.2566 hours, reflecting minimal influence from outliers, and the median duration is 1.2600 hours, closely aligned with the mean.

The standard deviation of 0.34303 hours and variance of 0.118 indicate moderate variability in session durations. The minimum reported session duration is 0.5 hours, and the maximum is 2 hours, resulting in a range of 1.5 hours. The interquartile range (IQR) is 0.42 hours, suggesting that the middle 50% of sessions are relatively consistent in duration.

A slight positive skewness of 0.026 indicates that the distribution is nearly symmetrical, with a very minor tilt toward longer sessions. The kurtosis value of -0.351 suggests a relatively flat distribution compared to a normal curve, indicating fewer extreme session durations. There are several outliers both above and below the whiskers. These outliers indicate session duration that are significantly longer or shorter than the majority of the data.

Overall, gym members tend to have consistent session durations, averaging slightly over an hour per session.

* **Calories Burned**

1. **Central tendency and Dispersion :**

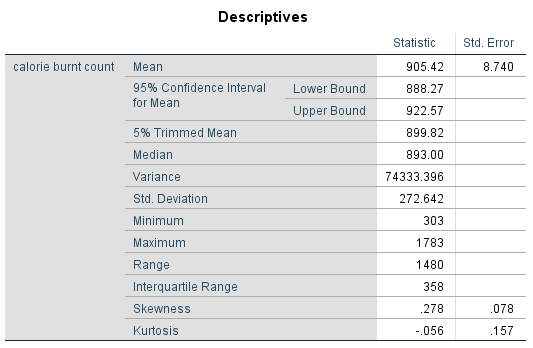
****

Table 12

1. **Boxplot and Outliers :**

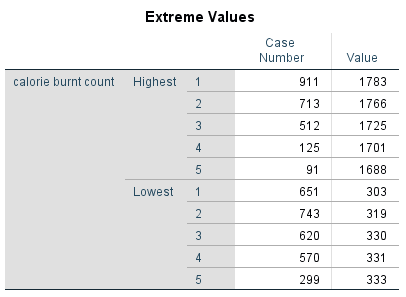
****

Table 13

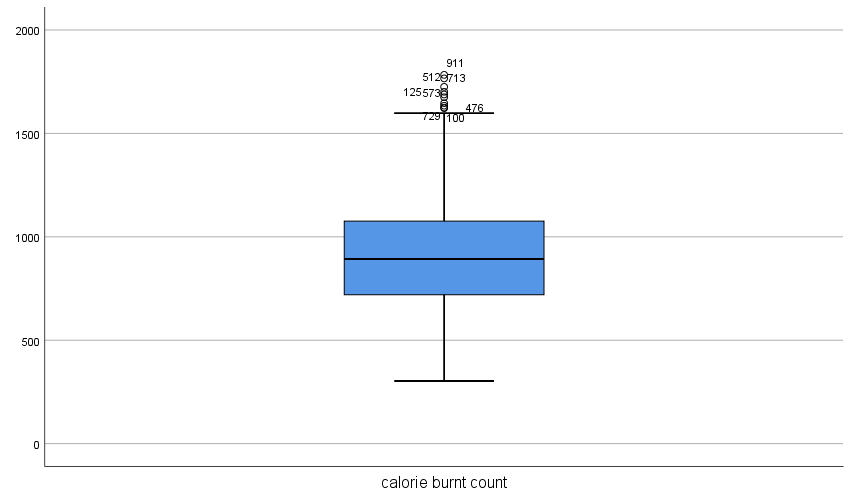
****

Figure 11

1. **Plots :**

* **Stem and Leaf**

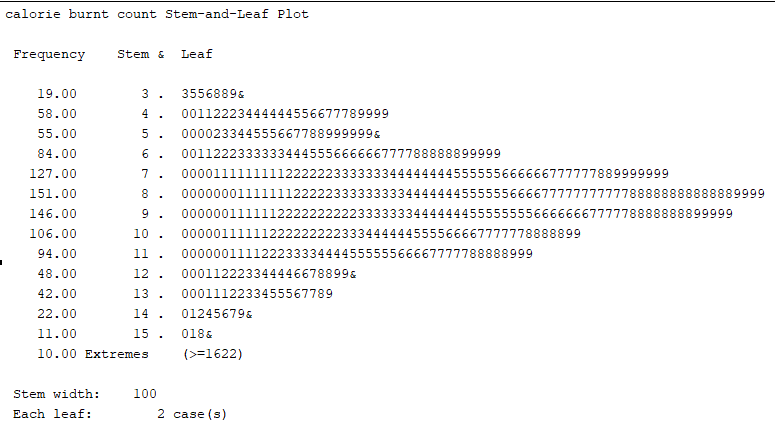
****

Figure 12

**Interpretation:**

The average calorie count burn is 905.42. Half of the sessions burned 893 calories or less, while the other half burned more. A variance of 74333.396 indicates significant variability. The standard deviation shows that the calorie counts vary by about 272.642 calories on average. A value of 278 indicates strong positive skewness, suggesting that most members burn a typical range of calories, while a few burn significantly higher amounts. The presence of outliers also suggests that a few sessions resulted in significantly higher or lower calorie burn compared to typical range. Kurtosis of -0.56 suggests a flatter distribution.

### **Inferential Statistics**

* **Test of Normality:**

1. **Workout\_Frequency**

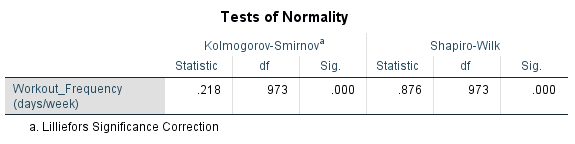


Table 14

* **Hypotheses Testing**:

**Step 1:**

**H₀**: The workout frequency of gym members follows a normal distribution.

**H₁**: The workout frequency of gym members does not follow a normal distribution.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

The workout frequency of gym members does not follow a normal distribution.

* **Plots:**

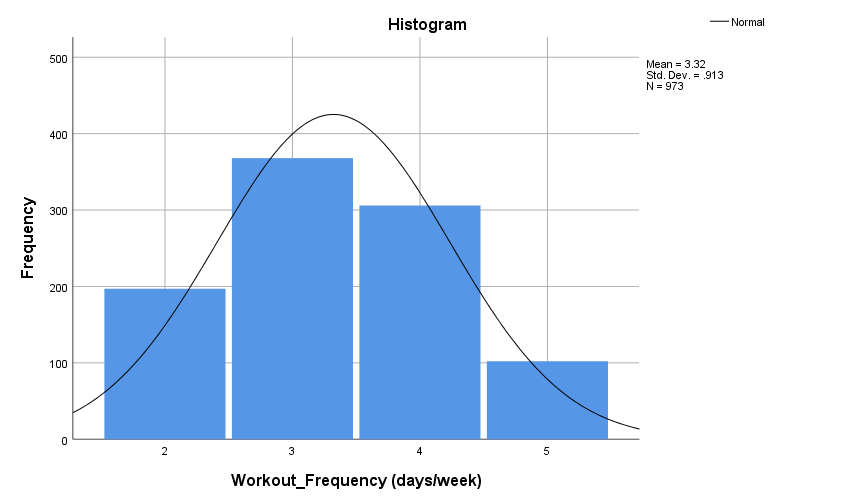
****

Figure 13

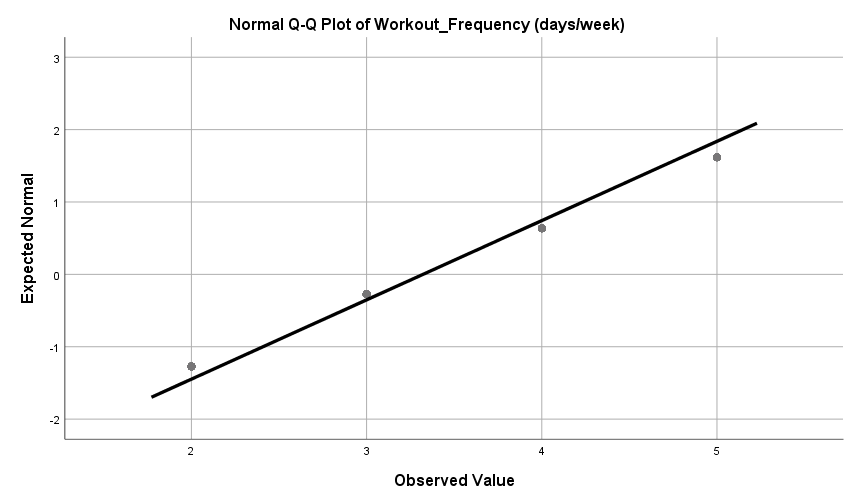


Figure 14

1. **Age**

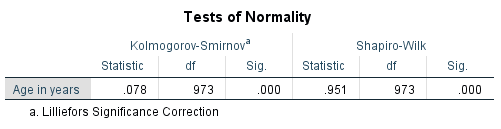
****

Table 15

* **Hypotheses Testing**:

**Step 1:**

**H₀**: The age of gym members follows a normal distribution.

**H₁**: The age of gym members does not follow a normal distribution.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

The age of gym members does not follow a normal distribution.

* **Plots:**

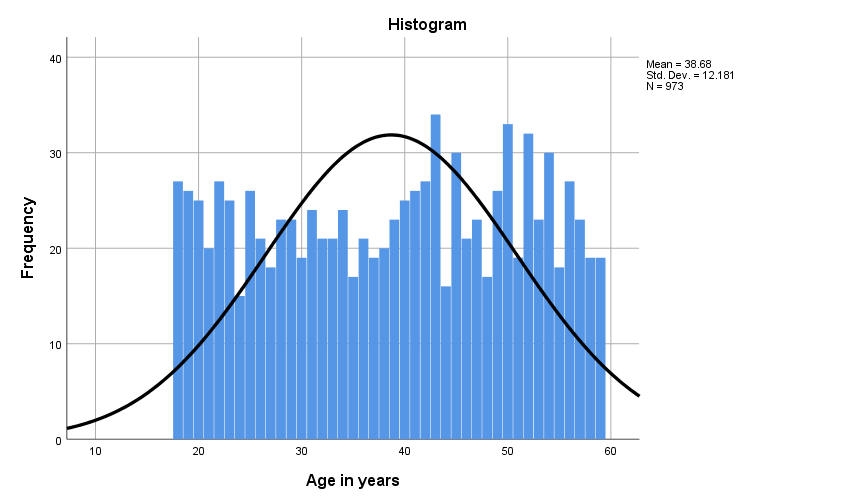
****

Figure 15

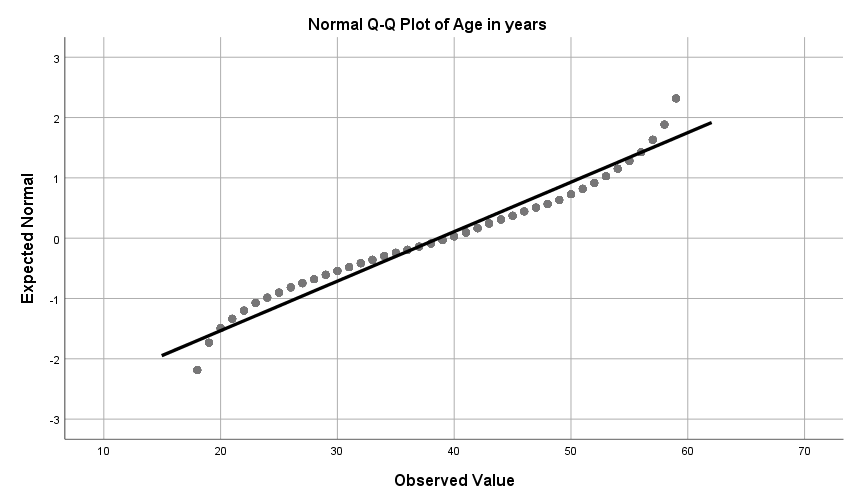


Figure 16

1. **Session\_duration**

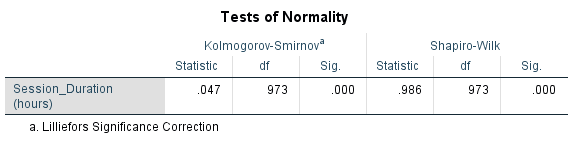
****

Table 16

* **Hypotheses Testing**:

**Step 1:**

**H₀**: The session duration of gym members follows a normal distribution.

**H₁**: The session duration of gym members does not follow a normal distribution.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

The session duration of gym members does not follow a normal distribution.

* **Plots:**

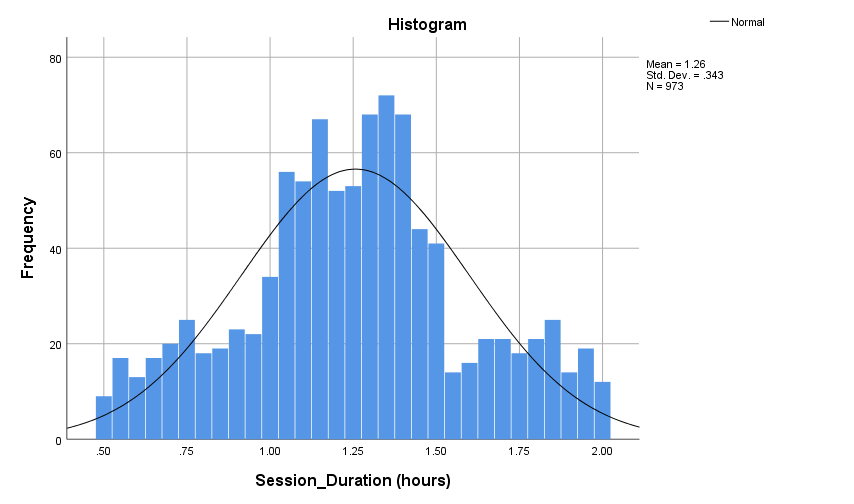
****

Figure 17

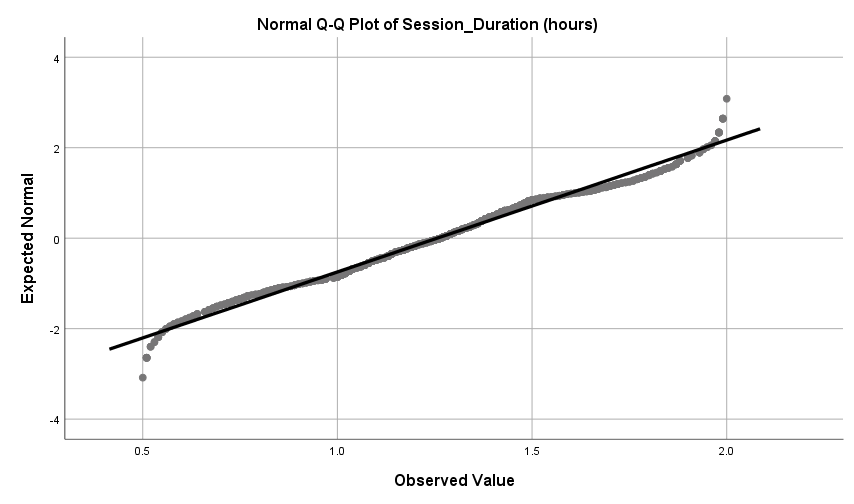


Figure 18

1. **Calories\_Burned**

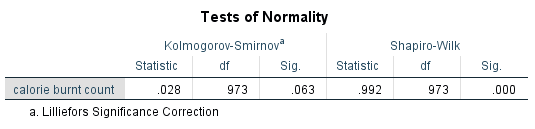


Table 17

* **Hypotheses Testing**:

**Step 1:**

**H₀**: The distribution of calories burned is normal.

**H₁**: The distribution of calories burned is not normal.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Fail to Reject H0

**Step 6:**

The distribution of calories burned is normal.

* **Plots:**

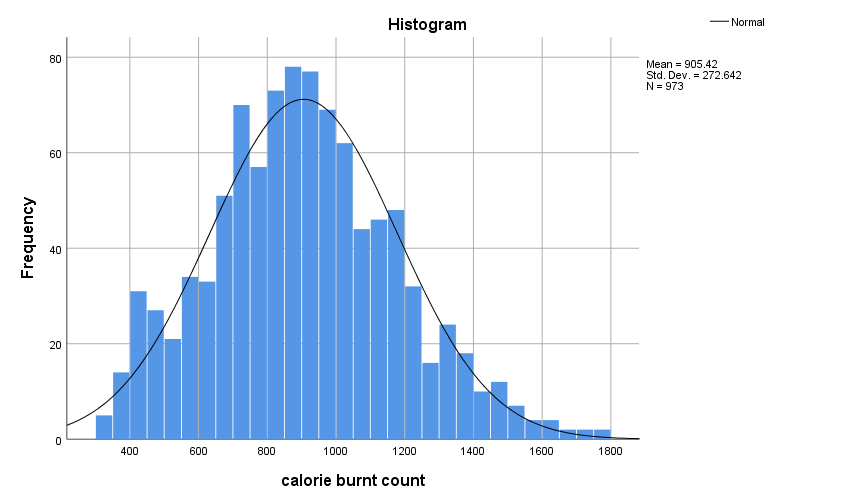


Figure 19

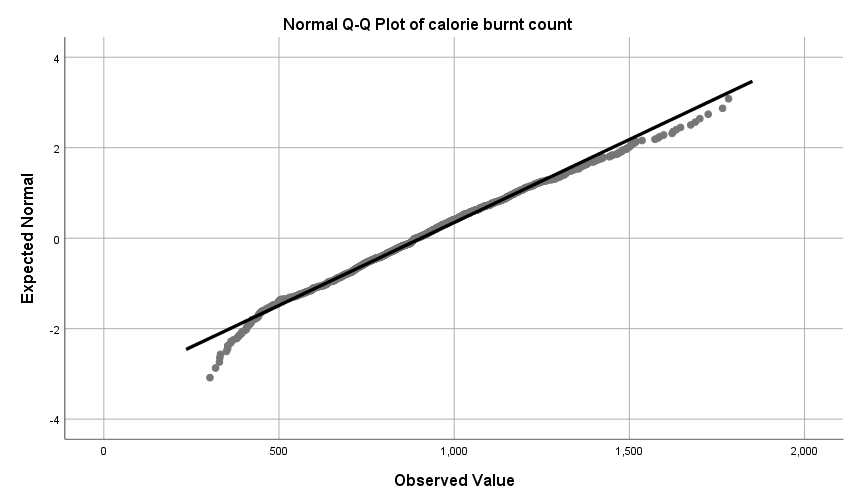


Figure 20

* **Parametric Tests:**

1. **T- test**

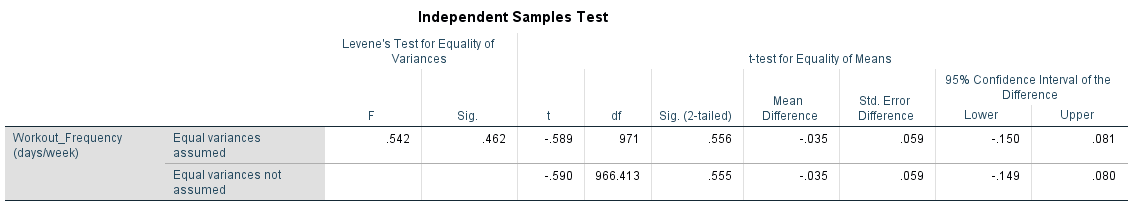
Compare **Workout Frequency** between **Gender** groups.

Table 18

* **Hypotheses Testing**:

**Step 1:**

**H₀**: There is no significant difference in means of ***workout frequency*** between males and females.

**H₁**: There is a significant difference in means of ***workout frequency*** between males and females.

**Step 2:**

α = 0.05

**Step 3:**

F = 0.542 and P-value = 0.462.

Since the **p-value > 0.05**, we fail to reject the null hypothesis of equal variances. Therefore, we proceed under the assumption of **equal variances**.

**Step 4:**

Using the row for "Equal variances assumed"

t = -0.589, df = 971, and P-value = 0.556

**Step 5:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 6:**

Since **p-value (0.556) > 0.05**, we fail to reject the null hypothesis. So, there is no significant difference in means of ***workout frequency*** between males and females.

1. **Pearson Correlation**

Examine the relationship between **Age** and **Calories Burned.**

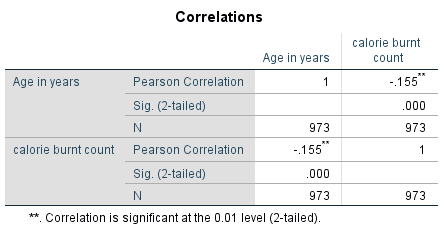
****

Table 19

* **Hypotheses Testing**:

**Step 1:**

**H₀**: There is no significant correlation between **Age** and **Calories Burned.**

**H₁**: There is a significant correlation between **Age** and **Calories Burned.**

**Step 2:**

α = 0.01

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

There is a significant correlation between Age and Calories Burned. The value of correlation coefficient is -0.155 which indicates a **weak negative relationship** between age and calorie burn count.

* **Plots:**



Figure 21

**Interpretation:**

The weak correlation is visually evident in the nearly flat trend line and the wide spread of points across all ages. The low R² value (0.024) confirms that only 2.4% of the variation in calorie burn count is explained by age.

While the relationship between age and calorie burn count is statistically significant, its practical significance is minimal due to the weak correlation.

Other factors likely have a more substantial influence on calorie burn count than age.

1. **One-Way ANOVA**

Compare ***Experience Level*** and **Workout\_frequency.**

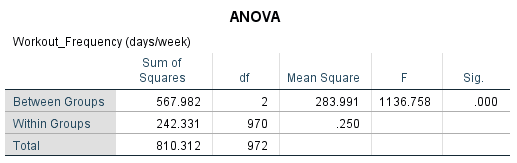
****

Table 20

* **Hypotheses Testing**:

**Step 1:**

**H₀**: There is no significant difference in mean workout frequency across different experience levels.

**H₁**: There is a significant difference in mean workout frequency across at least 2 experience levels.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

There is no significant difference in mean workout frequency across different experience levels.

* **Plots:**

****

Figure 22

**Interpretation:**

The plot shows a clear increase in **mean workout frequency** (days/week) as the **level of experience** rises. The ANOVA test revealed a statistically significant difference (p=0.000) in mean workout frequency across the three experience levels. This means experience level has a meaningful impact on workout frequency. The results support the alternative hypothesis

**(H1):** **Experience level significantly influences workout frequency.** The positive linear trend confirms that individuals with greater experience tend to work out more consistently.

1. **Multiple Linear Regression**

Predicting ***Calories Burned*** based on ***Session Duration*** and ***Avg\_BPM.***

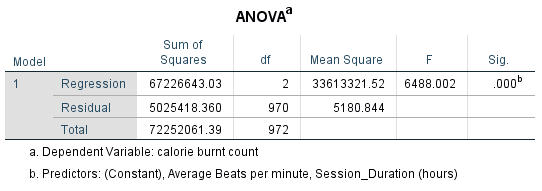


Table 21

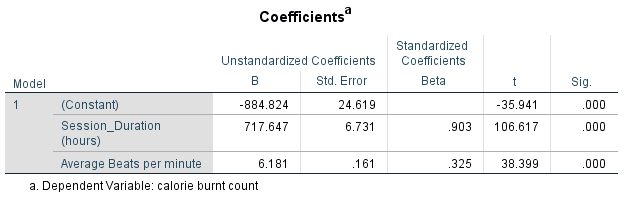


Table 22

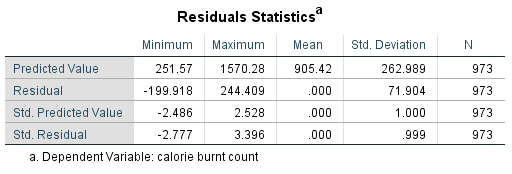
******

Table 23

* **Hypotheses Testing**:

**Step 1:**

**H₀**: Session duration and Avg\_BPM do not significantly predict calories burned.

**H₁**: Session duration and Avg\_BPM significantly predict calories burned.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

Session duration and Avg\_BPM significantly predict calories burned.

**Conclusion:**

The hypothesis test indicates that **session duration** and **average BPM (beats per minute)** are significant predictors of **calories burned** (p=0.000p = 0.000p=0.000). This means that both variables have a meaningful impact on calorie expenditure, and changes in these factors are likely to influence the number of calories burned during a session.

* **Plots:**



Figure 23

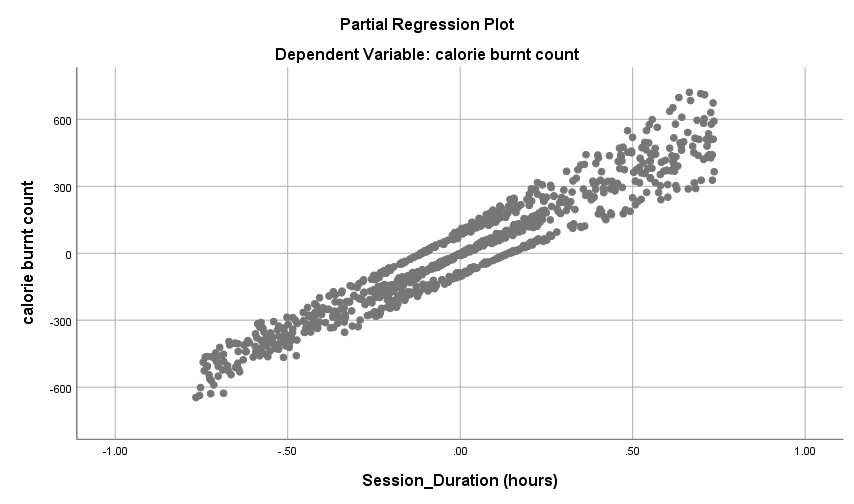
****

Figure 24

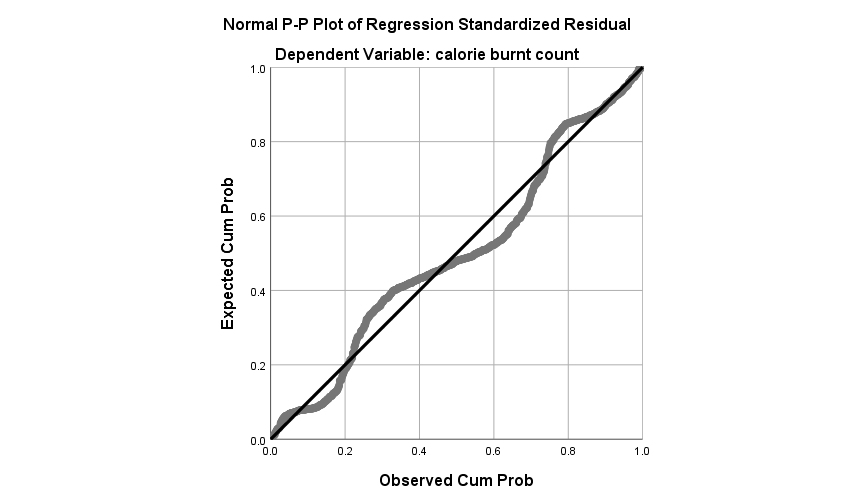
******

Figure 25

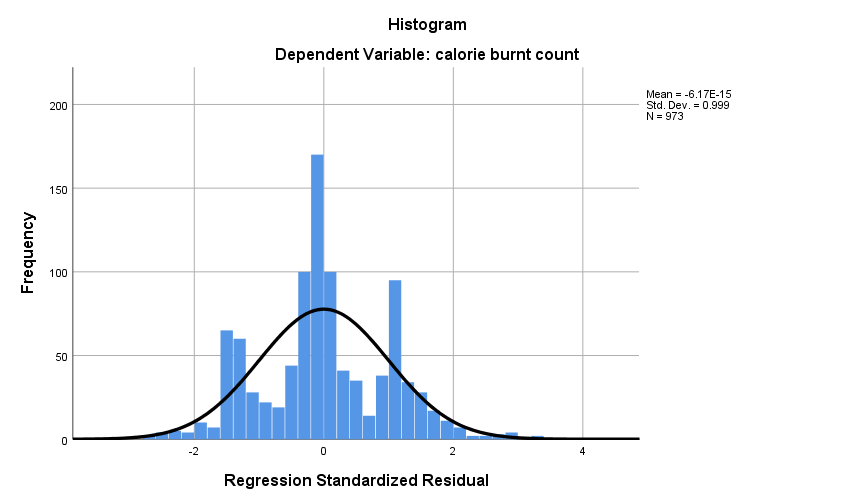
******

Figure 26

* **Non-Parametric Tests:**

1. **Mann-Whitney U Test**

Compare ***Calories Burned*** between ***genders.***

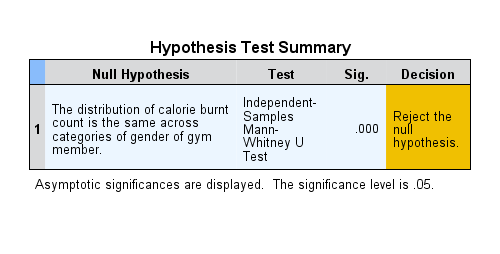


Table 24

* **Hypotheses Testing**:

**Step 1:**

**H₀**: There is not a significant difference in distribution of ***calories burned*** between male and female.

**H₁**: There is a significant difference in distribution of ***calories burned*** between males and females.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

There is a significant difference in distribution of calories burned between males and females.

1. **Spearman Correlation**

Relationship between ***Fat Percentage*** and ***BMI****.*

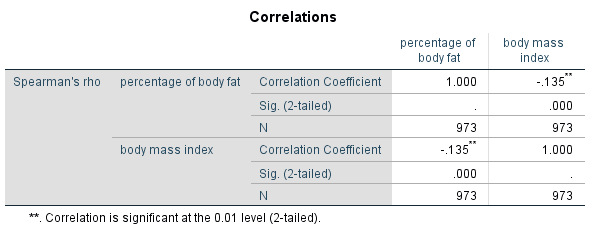


Table 25

* **Hypothesis Testing:**

**Step 1:**

**H₀**: There is no significant correlation between the percentage of body fat and BMI.

**H₁**: There is a significant correlation between the percentage of body fat and BMI.

**Step 2:**

α = 0.01

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.000

**Step 5:**

Reject H0

**Step 6:**

There is a significant correlation between the percentage of body fat and BMI.

1. **Kruskal-Wallis Test**

Compare ***Avg\_BPM*** and **BMI** across ***experience levels.***

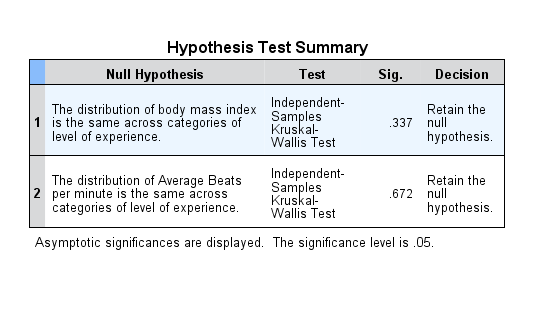
****

Table 26

* **Hypotheses Testing**:

**Step 1:**

**H₀**: The distribution of ***BMI*** and ***Avg\_BPM*** is the same across categories of ***level of experience.***

**H₁**: The distribution of ***BMI*** and ***Avg\_BPM*** is not same across categories of ***level of experience.***

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

**P-value for** BMI **= 0.337**

**P-value for** Avg\_BPM **= 0.672**

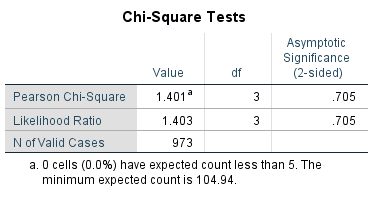
**Step 5:**

Do not Reject H0.

**Step 6:**

The distribution of ***BMI*** and ***Avg\_BPM*** is the same across categories of ***level of experience***.

1. **Chi-square Test of Independence**

****Relationship between ***Gender*** and ***Workout Type***.

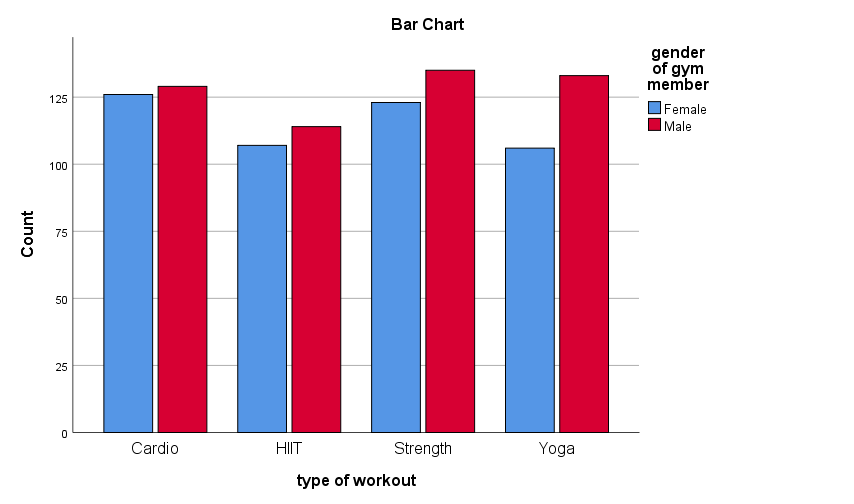
****

Table 27

Figure 27

* **Hypothesis Testing:**

**Step 1:**

**H₀**: There is no association between ***gender*** and ***workout type.***

**H₁**: There is an association between ***gender*** and ***workout type***.

**Step 2:**

α = 0.05

**Step 3:**

If p ≤ α, reject H0 otherwise fail to reject H0

**Step 4:**

P-value = 0.705

df = 3

**Step 5:**

Fail to Reject H0

**Step 6:**

There is no association between ***gender*** and ***workout type*** which means that both variables are independent.

### **Conclusion**

The analysis of gym attendance and exercise behavior provided valuable insights using data from 973 gym members. Through IBM SPSS, demographic variables, workout preferences, and fitness metrics were analyzed using descriptive and inferential statistical methods.

Key findings highlighted balanced gender representation and a preference for strength training, cardio, and yoga. The average member age was 38.68 years, with age influencing participation levels. Most members followed moderate workout routines, averaging 3.32 sessions per week, with session durations typically around one hour. Outlier analyses identified individuals with exceptionally high or low workout activity.

Inferential analyses offered deeper insights. T-tests revealed no significant difference in workout frequency between genders, while one-way ANOVA showed differences in workout frequency based on experience levels. Correlation analyses identified meaningful relationships, such as between age and calories burned, and between BMI and fat percentage. Regression analysis demonstrated session duration and average BPM as strong predictors of calories burned. Non-parametric tests like the Mann-Whitney U test confirmed differences in calories burned by gender, while chi-square tests found no association between gender and workout type.

This study emphasizes the importance of demographic and behavioral factors in gym attendance and exercise preferences. It provides actionable insights for gym management to develop tailored fitness programs and engagement strategies, enhancing member retention and satisfaction. However, limitations such as dataset scope, static data, and unexamined external factors suggest the need for further research to refine data-driven strategies in the fitness industry.

### **Limitations**

Despite its strengths, this study faced certain limitations that may affect the generalizability of the results:

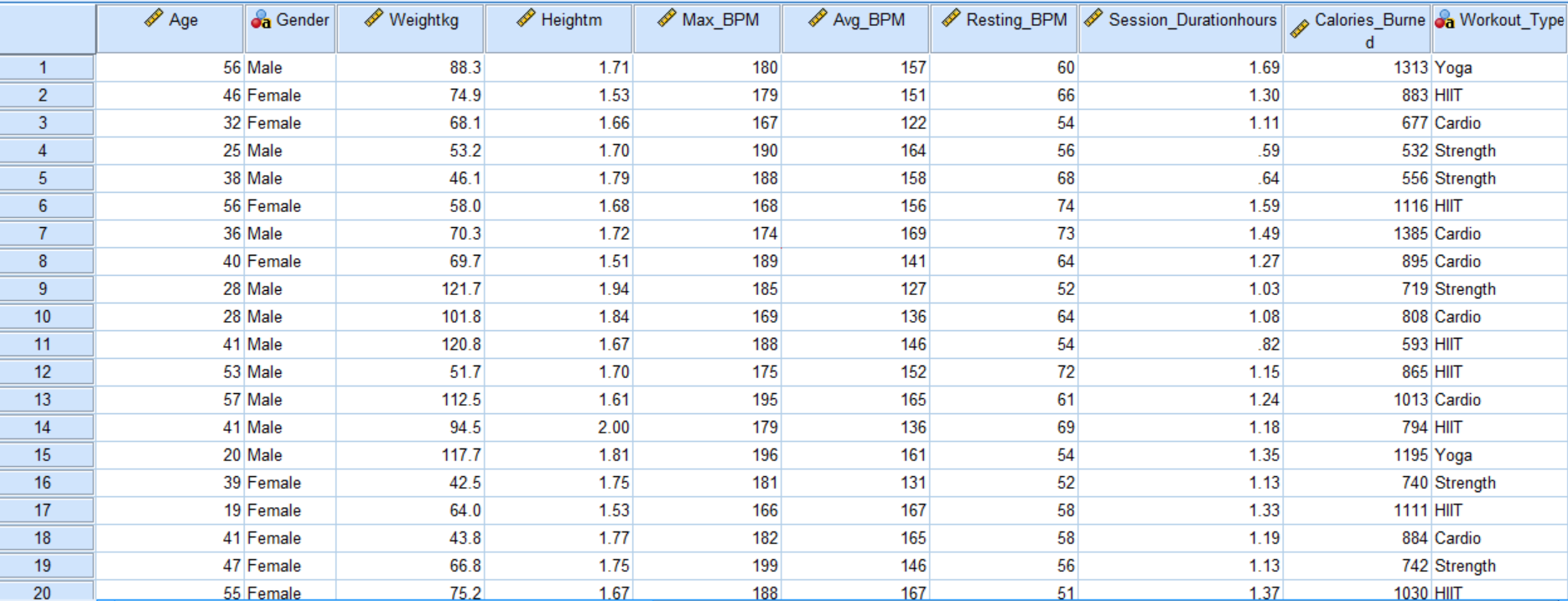
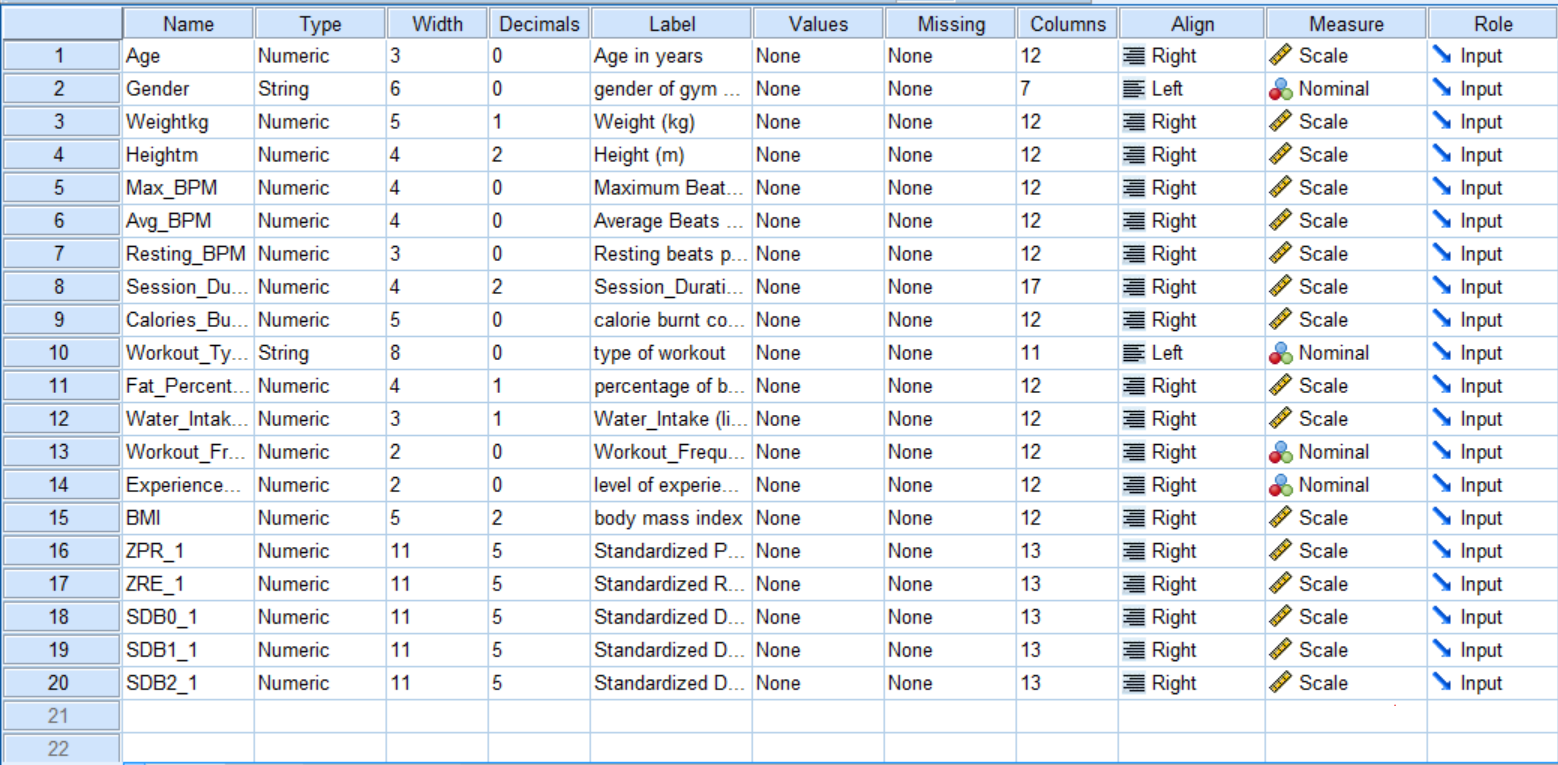
1. **Dataset Scope:** The dataset was limited to one source and may not represent the diverse population of gym members globally. Variations in gym cultures, geographic locations, and facilities were not accounted for.
2. **Static Data:** The dataset captured members' attributes at a fixed point, failing to reflect changes over time or long-term trends in fitness behaviors.
3. **Causality vs. Correlation:** While significant relationships were observed between variables, the analysis cannot establish causation, only correlation.
4. **Experience Classification:** The experience level variable lacked detailed differentiation, which may have masked nuanced insights across intermediate levels.
5. **Unexamined Factors:** External influences such as diet, socioeconomic status, or motivation levels, which likely impact exercise behavior, were not included in the dataset.
6. **Sampling Bias:** There is a possibility of bias in the dataset’s composition, such as an overrepresentation of specific demographics or workout preferences.

Future research could address these limitations by incorporating longitudinal data, expanding the dataset to include more diverse populations, and exploring external factors to provide a more holistic understanding of gym attendance and exercise behavior.

1. **References**

* Gjestvang, C., Abrahamsen, F., & Haakstad, L. A. H. (2021). What Makes Individuals Stick to Their Exercise Regime? A One-Year Follow-Up Study Among Novice Exercisers in a Fitness Club Setting. Frontiers in Psychology, 12, Article 638928. <https://doi.org/10.3389/fpsyg.2021.638928>.
* Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. Health Psychology, 28(6), 690–701.
* Middelkamp, J., Van Rooijen, M., & Steenbergen, B. (2016). Attendance behavior of ex-members in fitness clubs: A retrospective study applying the stages of change. Perceptual and Motor Skills, 122(2), 350–359.
* Research Square. (2018). A systematic review of interventions to increase attendance at health and fitness venues: Identifying key behavior change techniques. Retrieved from <https://www.researchsquare.com>.
* BMC Sports Science, Medicine and Rehabilitation. (2021). Long-term members’ use of fitness centers: A qualitative study. Retrieved from <https://bmcsportsscimedrehabil.biomedcentral.com>.
* Kaggle. (2023). Gym Members Exercise Dataset. Retrieved from <https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset>.

1. **Appendix**

* **Data View:**
* **Variable View:**