Analysis report for the Dataset Even

Almazan, Naomi Heart C., Tajon, Jaiven Kamylle

Introduction

This dataset explores the nutritional and physiological profiles of individuals through 1,000 patient entries. It includes various health indicators such as body fat percentage, muscle mass, BMI, caloric intake, and micronutrient consumption (e.g., protein, fat, carbohydrates, vitamin C, iron, and water intake). It also provides behavioral and lifestyle variables like physical activity hours and categorized BMI and activity levels.

The goal of this analysis is to understand patterns in body composition and nutrition, assess the relationships between physical activity and dietary habits, and uncover any significant statistical associations.

Materials and Methods

The dataset, containing 1,000 patient records and various health indicators, was imported into R and cleaned by removing unused columns and rows with missing values. Column names were renamed for clarity, and categorical variables such as sex, smoking status, and hypertension were converted into labeled factors.

Descriptive statistics (mean, median, standard deviation) were computed using the psych package. A new categorical variable, BMI_Category, was created based on WHO BMI classifications: Underweight, Normal, Overweight, and Obese.

Data visualizations were created using ggplot2, including histograms (e.g., BMI distribution), bar plots (e.g., BMI categories), scatter plots (e.g., BMI vs Blood Pressure), and boxplots (e.g., Glucose by Smoking Status).

Lastly, correlation analysis was used to assess relationships between continuous variables (e.g., glucose and cholesterol), and a chi-square test examined associations between categorical variables (e.g., BMI category and smoking status).

Key results and Figures

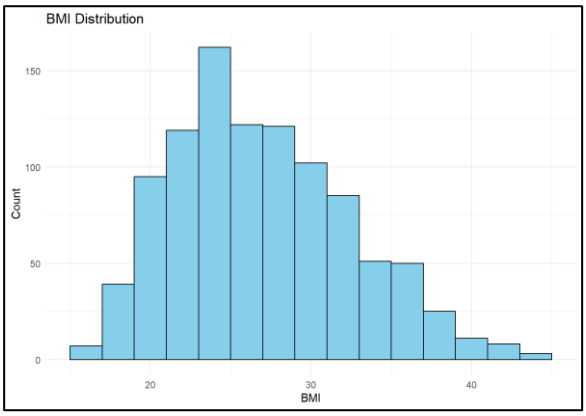

Figure 1. Histogram of BMI

The histogram shows that the majority of individuals fall within the overweight category, with a peak around BMI 27. The distribution is slightly right-skewed, suggesting that more people are above the healthy range than below it. This indicates a possible concern for obesity-related health risks in the dataset.

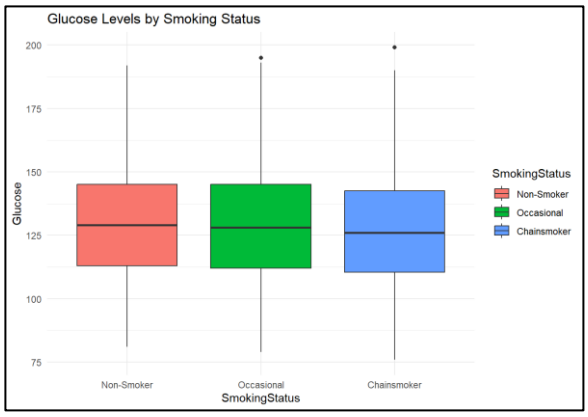

Figure 2. Scatterplot of BMI vs Systolic Blood Pressure

The boxplot shows a clear positive trend: individuals with higher BMI generally have higher systolic blood pressure. This suggests a potential risk for hypertension as BMI increases.

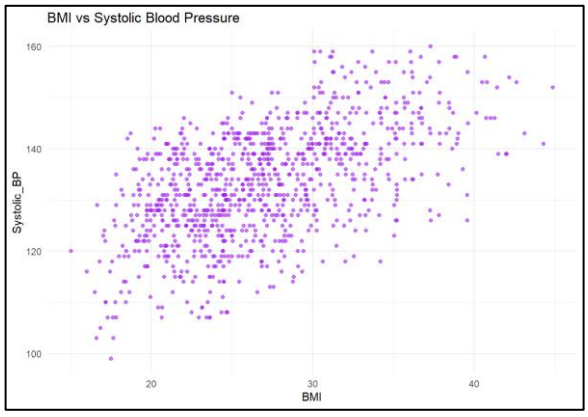

Figure 3: Scatterplot of BMI vs Systolic Blood Pressure

Figure 3 shows a clear positive trend: individuals with higher BMI generally have higher systolic blood pressure. This suggests a potential risk for hypertension as BMI increases.

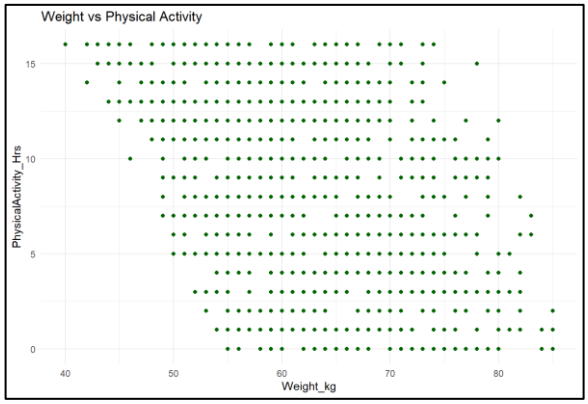

Figure 4. Weight vs Physical Activity

People with different weights are spread across all activity levels, but those who exercise less often tend to weigh less. This might mean that being more active helps maintain more muscle mass.

Test Type	Variables Compared	Result (r / χ^2)	p-value	Interpretation
Pearson Correlation	Glucose vs. Cholesterol	r = 0.693	< 2.2e-16	**Strong positive correlation**: Higher glucose is associated with higher cholesterol.
Chi-Square Test	BMI Category vs. Smoking Status	χ^2 = 6.94	0.3262	**No significant association**: Smoking status is not related to BMI category.

Table 1. Statistical Resuls

Interpretation

Most individuals fall into the *overweight* and *obese* BMI categories, as shown in the BMI histogram, signaling a possible population-level concern regarding weight management. A positive trend between BMI and systolic blood pressure suggests that higher BMI is associated with elevated cardiovascular risk.

Additionally, glucose and cholesterol levels show a **strong positive correlation (r = 0.693, p < 0.001)**, indicating that individuals with high blood. Also, the **chi-square test (p = 0.3262)** found no significant association between BMI category and smoking status, suggesting that smoking behavior in this sample does not vary meaningfully across different BMI groups.

Lifestyle indicators like physical activity hours appear scattered across all weight ranges, but those with higher weights are somewhat more active, possibly due to muscle mass maintenance or exercise prescriptions. Glucose levels by smoking status vary slightly, but without strong outliers, hinting at complex, non-linear relationships.

This analysis reveals that higher BMI is linked with increased blood pressure, and glucose strongly correlates with cholesterol, reinforcing known risks of cardiovascular and metabolic disorders. While smoking status does not significantly differ across BMI groups, the prevalence of overweight individuals and elevated

glucose-cholesterol levels indicate an urgent need for integrated interventions focusing on weight control, diet, and metabolic health.

Conclusion

This analysis reveals that higher BMI is linked with increased blood pressure, and glucose strongly correlates with cholesterol, reinforcing known risks of cardiovascular and metabolic disorders. While smoking status does not significantly differ across BMI groups, the prevalence of overweight individuals and elevated glucose-cholesterol levels indicate an urgent need for integrated interventions focusing on weight control, diet, and metabolic health.