REPORT AND ANALYSIS

How BMI Relates to Systolic Blood Pressure

A Weak but Visible Trend

Dataset Odd

Prepared By

Brylle Matthew A. Lupac

Naomi Christienne Tiama

1. Introduction

This dataset comprises detailed health and lifestyle profiles of 1000 individual patients, focusing on various physiological, behavioral, and clinical indicators relevant to chronic disease risk assessment. Each entry includes demographic data (Patient ID, Age, Sex), anthropometric measurements (Weight in kg, Height in cm, BMI), and cardiovascular readings (Systolic and Diastolic Blood Pressure, Heart Rate), alongside diagnostic markers such as glucose and cholesterol levels. Key lifestyle factors are also recorded, including smoking status, weekly physical activity hours, self-reported stress levels, and daily sleeping hours. Categorical indicators such as Hypertension stages, Smoking behavior, and Sex are clearly labeled both in numerical and descriptive formats, facilitating easier grouping and stratified analysis. The dataset also provides a final assessment column, "Elevated Risk," which classifies patients into categories such as "Normal/No Major Risk," "Pre-Diabetes," "Type 2 Diabetes," "Insulin Resistance," or specific conditions like "Pheochromocytoma." With this rich collection of variables, the dataset supports multifactorial health risk analysis and is especially suited for studies on the relationship between lifestyle factors and metabolic or cardiovascular conditions. Its structure enables both exploratory and predictive modeling applications for researchers aiming to understand risk contributors to chronic diseases, especially those influenced by stress, inactivity, poor sleep, and unhealthy biomarkers.

2. Methods

2.1. Data Cleaning and Preparation

- The dataset was first imported from a .csv file using the read.csv() function. Initial checks such as str(), colnames(), and glimpse() were used to examine the structure, column names, and a preview of the dataset.
- Data types were then corrected using the mutate() function from the dplyr package. Several categorical variables were converted into factors:
    - **Sex** was transformed into a factor variable.
    - **Smoking_Status** was relabeled with descriptive labels: "Non-Smoker," "Occasional," and "Chainsmoke."
    - Other variables such as **Hypertension**, **Elevated.Risk**, **Medication**, and **Stress_Level** were also converted into factor form.

2.2. Descriptive Statistical Analysis

- The group computed summary statistics for the key numeric variables using the summarise_all() function:
    - **Mean** – the average value
    - **Median** – the middle value when data are ordered
    - **Standard Deviation** – the spread of the values from the mean
- The analysis was applied to selected variables, including:
    - Age
    - Weight (kg), Height (cm), BMI
    - Systolic and Diastolic Blood Pressure
    - Heart Rate
    - Physical Activity Hours per Week

- Daily Sleeping Hours
- Glucose and Cholesterol Levels

2.3. Data Visualization

- **Scatter Plot:** A scatter plot of BMI vs. Systolic Blood Pressure was created using geom_point() and geom_smooth(method = "lm").
- **Bar Graph:** A bar graph was plotted using geom_bar() to show the number of individuals in each Smoking Status category, using labeled levels for clarity.
- **Histogram**: A histogram of Daily Sleeping Hours was generated with 12 bins to show the distribution of sleep hours among individuals.
- **Boxplot:** A boxplot compares Cholesterol levels across different Stress Levels, helping to identify trends or differences between groups.

2.4. Advanced Statistical Test

- To examine whether there was a significant relationship between Hypertension status and Smoking Status, the group performed a Chi-Square Test of Independence using the chisq.test() function.
- This test assessed whether the distribution of hypertensive and non-hypertensive individuals differed across smoking categories.
- The result provided a p-value, which was interpreted to determine whether any observed association was statistically significant.

3. Results and Discussion

3.1. Summary Statistics

Table 1. Descriptive Statistics for Selected Health Variables

Variable	Mean	Median	Standard Deviation (SD)
Age (years)	54.601	55	21.309
Weight (kg)	62.459	62	9.161
Height (cm)	154.040	155	11.622
Body Mass Index (BMI)	26.738	25.961	5.546

Systolic Blood Pressure (mm Hg)	133.402	133	10.884
Diastolic Blood Pressure (mm Hg)	81.669	82	8.420
Heart Rate (beats per minute)	99.927	99	18.486
Physical Activity (hrs/week)	8.125	8	4.826
Daily Sleeping Hours (hrs/day)	5.378	5	1.233
Glucose (mg/dL)	128.242	125	23.280
Cholesterol (mg/dL)	189.179	187	34.515

- Most participants were middle-aged, with a mean age of 54.6 years.
- The average weight was 62.5 kg, and the average height was 154 cm.
- The mean BMI was 26.7, indicating that the group was, on average, slightly overweight.
- Systolic blood pressure averaged 133.4 mm Hg, and diastolic blood pressure averaged 81.7 mm Hg, both near hypertension thresholds.
- The average heart rate was 99.9 beats per minute, slightly above the typical resting range.
- Participants engaged in about 8.1 hours of physical activity per week.
- The average sleeping duration was 5.4 hours per day, which is below recommended levels.
- The group had mean glucose levels of 128.2 mg/dL and cholesterol levels of 189.2 mg/dL, both slightly elevated.
- Small differences between mean and median values suggest that the distributions of most variables were approximately symmetric.

3.1. Data Visualizations

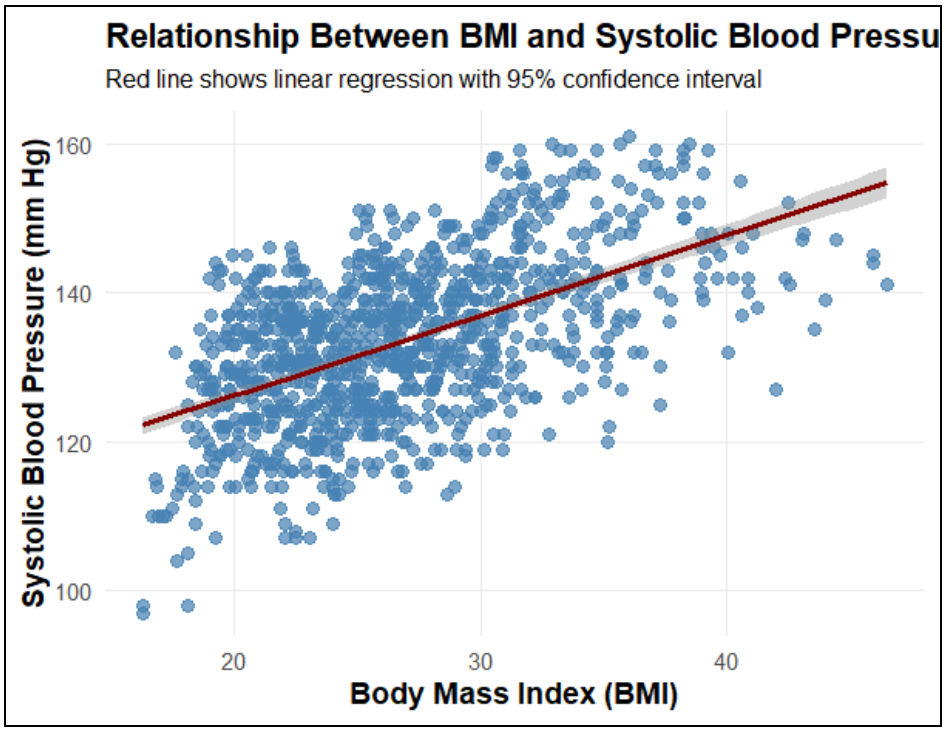

Figure 1. Scatter Plot: BMI vs. Systolic Blood Pressure

- Figure 1 shows a positive linear relationship between Body Mass Index (BMI) and Systolic Blood Pressure.
- As BMI increases, systolic blood pressure tends to rise, indicating that individuals with higher BMI generally have higher blood pressure.
- The regression line confirms this trend, suggesting a potential link between overweight/obesity and elevated blood pressure.
- The spread of points shows moderate variability, but the upward trend is consistent.

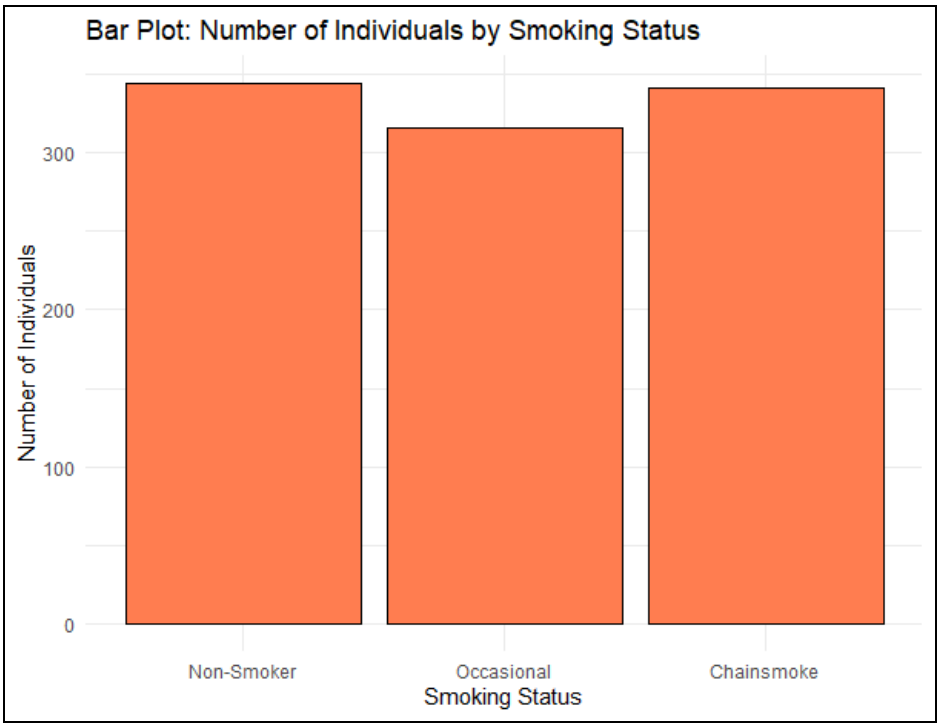

Figure 2. Bar Plot: Number of Individuals by Smoking Status

- Figure 2 displays the distribution of individuals by Smoking Status.
- The largest group was Non-Smokers (344 individuals), closely followed by Chainsmokers (341), and then Occasional Smokers (315).
- The number of individuals is fairly balanced across all three smoking categories.
- This distribution indicates a high prevalence of smoking behavior in the population, with nearly equal representation across smoking types.

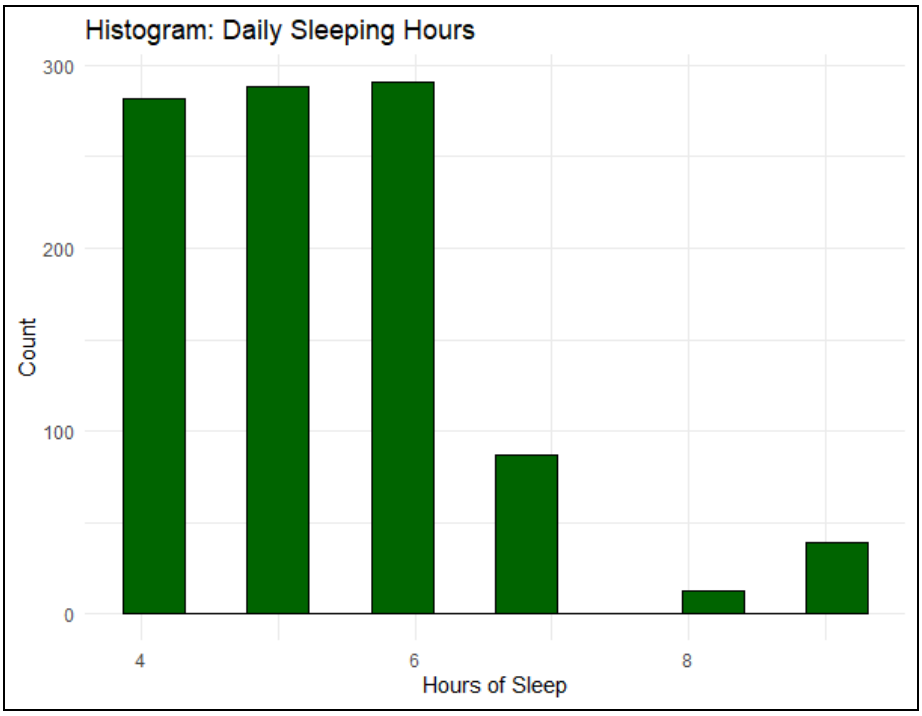

Figure 3. Histogram: Distribution of Daily Sleeping Hours

- Figure 3 shows the distribution of Daily Sleeping Hours among participants.
- The majority of individuals reported sleeping between 4 to 6 hours per day, with the highest count at 6 hours (291 individuals).
- Very few participants slept for 7 hours (87) or 8 hours (13), and only 39 individuals reported sleeping 9 hours.
- This suggests that most individuals in the dataset get less than the recommended 7–8 hours of sleep, indicating potential sleep deprivation in the group.

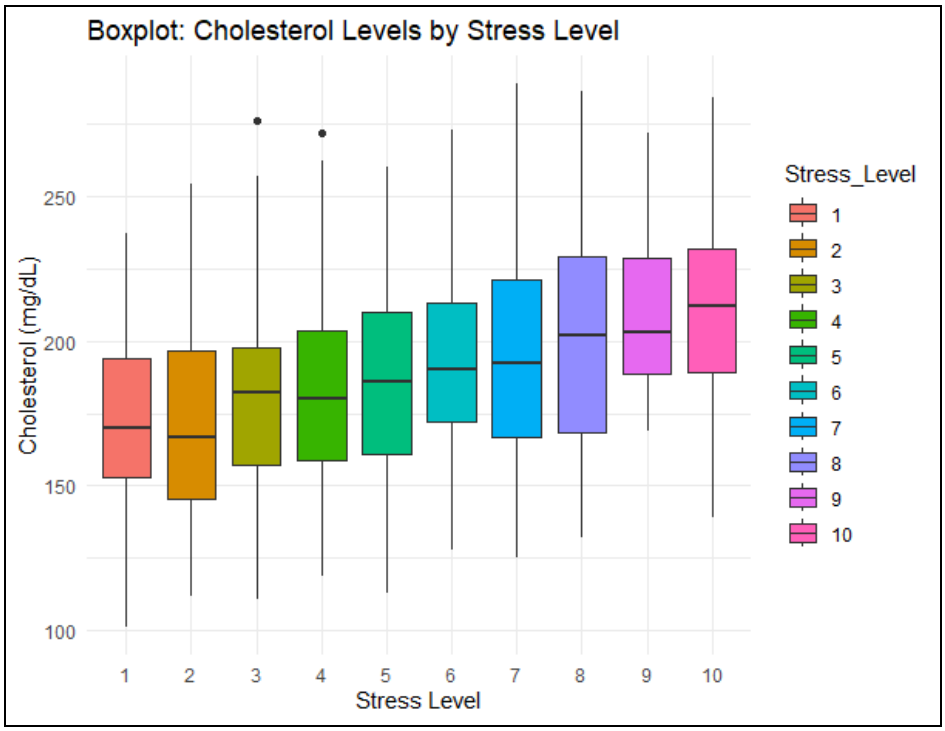

Figure 4. Boxplot: Cholesterol Levels by Stress Level

- Figure 4 shows the distribution of cholesterol levels across different stress levels using boxplots.
- Individuals with higher stress levels generally exhibited greater variation in cholesterol levels, as indicated by wider interquartile ranges and more outliers.
- Some stress levels showed higher median cholesterol values, suggesting a potential association between elevated stress and increased cholesterol.
- The presence of multiple outliers at both low and high stress levels indicates that cholesterol levels vary considerably within each group.

2.3 Advanced Statistical Insight

Table 2. Chi-Square Test of Independence Between Hypertension and Smoking Status

Test	Chi-Square Statistic (X^2)	Degrees of Freedom (df)	p-value	Conclusion
Pearson's Chi-Square Test	134.01	8	$< 2.2 \times 10^{-16}$	Significant association between variables

- The p-value is less than 0.001, indicating a statistically significant relationship between Hypertension status and Smoking Status. This suggests that individuals' likelihood of having hypertension may vary depending on their smoking habits.

4. Conclusion

This analysis highlights important relationships between health indicators and lifestyle factors in a population of 1,000 patients. On average, participants were middle-aged, slightly overweight, and had borderline high blood pressure, cholesterol, and glucose levels. Most individuals reported insufficient sleep and moderate physical activity. The data revealed clear patterns: higher BMI was linked to increased systolic blood pressure, and higher stress levels were associated with more variability in cholesterol levels. The Chi-square test showed a significant link between smoking status and hypertension, suggesting that smoking behavior affects blood pressure health. Overall, this dataset shows how lifestyle habits like smoking, physical activity, stress, and sleep are closely tied to chronic disease risk and reinforces the importance of preventive health measures.