Mathematical Foundations I (Probability & Statistics)

Project Report

Course Code: AI&DS-ST106
Group Members: Sneha Shinde, Rishita Merchant, Sanskriti Dhar
Dataset: Heart Disease - Cleveland UCI Dataset

Introduction

The objective of this project is to explore and analyze clinical data to understand patterns and indicators related to heart disease. Using statistical methods, we investigated **relationships between physiological measurements and heart health risk factors**. The dataset we worked with contains information from 279 patients, with variables such as age, blood pressure, cholesterol, chest pain types, and maximum heart rate recorded. Our aim was to draw meaningful conclusions from this dataset using a combination of hypothesis testing, analysis of variance, and correlation studies.

Data Overview and Initial Observations

Our dataset consisted of patient profiles that included age, gender, chest pain types, resting blood pressure, cholesterol levels, and exercise-induced heart responses. One of the first things we noticed was that most individuals were middle-aged or elderly. Blood pressure readings suggested that many **patients were already within the pre-hypertensive or hypertensive range. Cholesterol levels were typically high, hinting at a common cardiovascular risk** shared by this group.

Resting Blood Pressure Analysis

We initially explored whether the **average resting blood pressure across the sample differed significantly from a commonly accepted standard of 130 mmHg**. Upon conducting a one-sample t-test, we found that the average blood pressure in the sample could reasonably be assumed to be 130 mmHg. This suggests that while the blood pressure values vary, the central tendency does not significantly differ from the threshold used in many medical contexts.

Gender-Based Differences in Cholesterol

When comparing cholesterol levels between male and female patients, an interesting pattern emerged. **Women in the dataset tended to have higher cholesterol levels than men**. This difference was not just observed in averages but was statistically significant, indicating that gender may play a role in cholesterol profiles within this patient population. Such insights are crucial for gender-specific preventive care and screening protocols.

Heart Rate vs Blood Pressure

We also examined how maximum heart rate during exertion compares to resting blood pressure. The paired t-test revealed that, for most patients, **the heart rate under stress was significantly higher than their resting blood pressure**. This confirms what we would **physiologically expect: during exercise or exertion, heart activity increases** considerably. But the degree of this increase could also signal cardiovascular strain or fitness, depending on the patient.

Chest Pain and Blood Pressure

Chest pain types were classified into categories like typical angina, atypical angina, non-anginal pain, and asymptomatic. We wanted to know whether patients experiencing different types of chest pain also showed differing blood pressure levels. The results of our ANOVA test indicated that **resting blood pressure does vary significantly across the types of chest pain**. This means the way **a person experiences chest discomfort could be linked to deeper circulatory or cardiac characteristics.**

To further explore this, we used a post-hoc Tukey test. We discovered that **patients with atypical angina and non-anginal pain had significantly lower blood pressure than those with typical angina**. Interestingly, patients who were asymptomatic did not show notable differences compared to others. These findings point to typical angina being more strongly associated with elevated blood pressure, reinforcing its clinical significance as a warning sign.

Understanding Correlations in the Data

Finally, we conducted a correlation analysis to uncover relationships between key variables. **Age showed a clear positive relationship with blood pressure, cholesterol, and the number of blocked arteries observed in angiographic results**. This aligns with the well-known reality that cardiovascular risks increase with age.

In contrast, we observed a **negative relationship between age and maximum heart rate - older individuals tended to reach lower peak heart rates.** Exercise-induced angina and ST depression (measured by ECG) also showed an inverse relationship with heart rate,

suggesting that patients experiencing these symptoms tend to have poorer cardiac response during exertion.

Some interesting clinical patterns also emerged in gender-based analyses. **Male patients, for instance, were more likely to present with abnormalities during stress tests and had a greater number of blocked vessels.** Additionally, features like ST depression and ECG slope were strongly linked, indicating that they might reflect a shared underlying cardiac stress or dysfunction.

Conclusion

Through this project, we were able to leverage statistical techniques to uncover meaningful health patterns in a real-world dataset. Our findings reinforce many established clinical truths, such as **the association between aging and cardiovascular risk, and highlight some nuances, like the elevated cholesterol levels seen in female patients**. **The relationship between symptoms like chest pain and measurable health parameters like blood pressure adds further depth to the understanding of heart disease presentation.**

In essence, this project highlights the power of statistics in uncovering hidden patterns, supporting diagnosis, and guiding personalized healthcare decisions. As data becomes more central to medicine, the ability to interpret it statistically becomes an indispensable skill for modern clinicians and data scientists alike.