What Factors Affects Mortality from Heart Disease?
140XP Final Project

Lior Levy Meruk, Mary McSweeney, Henry Johnson, Alex Clare, Maya Hirsch

Abstract

This study attempts to answer one main question: what inherited and acquired traits and general factors significantly influence mortality from heart disease? We chose 4 variables to explore: age, high cholesterol, diabetes, and gender. Using 2-sample t-tests, Chi-square goodness-of-fit tests, and a logistic regression model, we explored which factors are significant to mortality in heart disease patients, which we found all are significant. This is consistent with outside research our team has found.

1. Introduction

Heart disease is an increasing concern and a leading cause of mortality around the world. This raises a need for deeper research and effective predictive tools in order to reduce the risks and improve patient health. Our research question is, what are the significant factors impacting mortality in heart disease patients? This main research question guides the analysis, seeking to uncover important variables such as age, cholesterol levels, diabetes, and gender that could affect patient outcomes. In Pakistan, where healthcare resources are more limited, understanding the key factors influencing heart disease outcomes is especially important. This report aims to analyze a dataset which documents heart disease mortality in Pakistan, leveraging data science methods to gather insights that could inform clinical practices and improve patient health.

2. Background

Cardiovascular diseases (CVDs) are one of the world's most prominent public health concerns, contributing to considerable morbidity and mortality. It is known that factors such as age, gender, lifestyle, and comorbidities like diabetes and hypertension substantially influence cardiovascular disease outcomes. Furthermore, additional obstacles such as late diagnosis, limited access to advanced healthcare, and disparities in patient management enhance these risks.

*Corresponding author

This dataset, collected from hospital records in Pakistan, encompasses several clinical and demographic information for heart patients. It includes variables including age, gender, blood pressure, cholesterol levels, smoking status, and additional health markers. The dataset offers a valuable opportunity to explore trends and correlations that could influence early intervention approaches, improve risk assessment, and better treatment plans. Through using machine learning techniques and statistical analyses, this paper seeks to reveal actionable insights that can contribute to both academic understanding of heart disease and its management in places with limited resources.

3. Research Question

Our main research question is:

- What are the significant factors impacting mortality in heart disease patients?

More specifically, we wish to gain insight on:

- Does diabetes have an effect on mortality in heart disease patients?
- Does high cholesterol have an effect on mortality in heart disease patients?
- Does a person's gender have an effect on mortality in heart disease patients?
- Does a person's age have an effect on mortality in heart disease patients?

4. EDA - Exploratory Data Analysis

We conducted Exploratory Data Analysis (EDA) on Age, Cholesterol, Diabetes, and Gender, which were several key variables we were interested in. This exploratory data analysis was performed focusing on the primary variable of interest, mortality, to investigate potential patterns and relationships.

To improve clarity during analysis, we transformed the Mortality variable (0 - Alive, 1 - Deceased) into a new variable, Deceased (0 - No, 1 - Yes). Our exploratory data analysis focused on Age, Cholesterol, Diabetes, and Gender, examining their relationships with mortality. Summary statistics revealed that deceased patients tended to have a higher mean age (54.8 vs. 52.6 years) and lower average cholesterol levels (244 vs. 266 mg/dL) compared to survivors.

Visualizations, such as histograms, boxplots, and bar charts, highlighted trends, including higher mortality proportions among individuals with diabetes and males. We will later perform statistical tests to test these trends. This analysis provided insights into key factors linked to mortality, informing subsequent multivariate modeling.

Gender, Diabetes, Cholesterol, Age EDA: Figure 1.

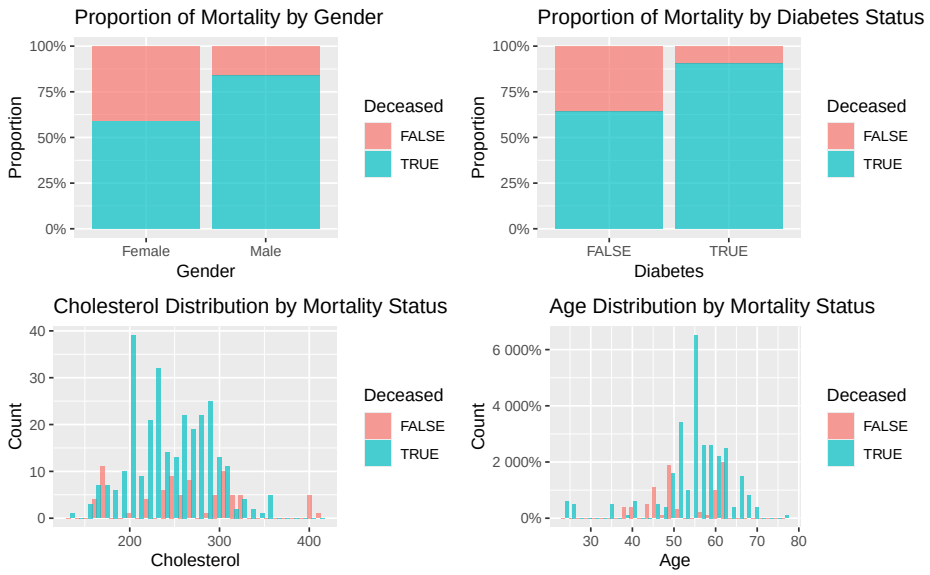

Figure 1: Variables by Mortality

5. Hypothesis Testing

Table 1: Hypothesis Test Summary

Hypothesis	Test	P-value
Significant association between gender and mortality	Chi-square test	2.955e-06
Significant association between diabetes and mortality	Chi-square test	2.955e-06
Significant association between age and mortality	T-test	0.04704
Significant association between cholesterol and mortality	T-test	0.00693
Age, cholesterol, diabetes, and gender are predictors of mortality	Logistic regression	All predictors p < 0.01

6. Equations

Here is an equation:

$$f_X(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; \alpha, \beta, x > 0.$$

Inline equations work as well: $\sum_{i=2}^{\infty}\{\alpha_i^{\beta}\}$

7. Figures and tables

Figure 2 is generated using an R chunk.

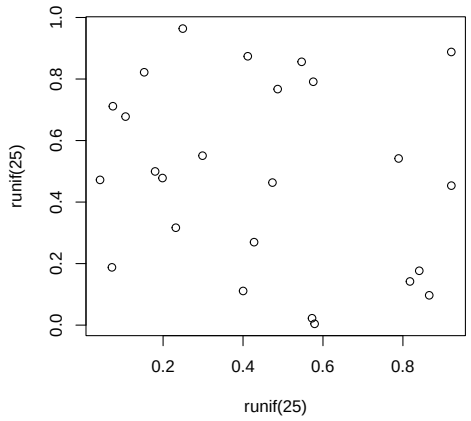

Figure 2: A meaningless scatterplot

8. Tables coming from R

Tables can also be generated using R chunks, as shown in Table 2 example.

```
knitr::kable(head(mtcars)[,1:4])
```

Table 2: Caption centered above table

	mpg	cyl	disp	hp
Mazda RX4	21.0	6	160	110
Mazda RX4 Wag	21.0	6	160	110
Datsun 710	22.8	4	108	93
Hornet 4 Drive	21.4	6	258	110
Hornet Sportabout	18.7	8	360	175
Valiant	18.1	6	225	105

References