

Indications of Heart Failure

Proud Jiao, Xiaocong Xuan, Yijie Lu, Xiaoyan Wei, Lukas Lam, Felix Zhou

Statistics Department, UCLA

Abstract

The main question we wanted to answer is what are significant indicators that could predict heart failures. The methods we used include Exploratory Data Analysis (EDA) to find patterns and especially correlations between predictors and heart failure and ANOVA to determine statistical evidence for correlation. The results indicate that common indicators such as Age, Sex, Resting Blood Pressure, Fasting Blood Sugar Level, and Max Heart Rate, all have significant correlations with heart disease.

1. Introduction

Heart Disease is the leading cause of death within the US, with one person dying every 34 seconds due to a heart related issue. Up to 90% of these deaths are preventable, but a mere 8% of Americans recognize this (CDC). We wanted to take the initiative and find out the leading symptoms that indicate heart disease.

Our research question is, for Heart Failure, what factors increase the chances of being diagnosed for heart failure? We will take a look at various predictors, within the Heart Disease data set, and closely examine the correlation between predictors and whether heart failure is present. Additionally, we validate our statistical findings with credible medical sources, such as Mayo Clinic and Center for Disease Control (CDC).

2. Background

We preformed our statistical analysis on the Heart Disease data set, which includes 368 different observations and 60 different predictors. However, we narrowed our predictors to more common indicators such as, Age, Sex, Resting Blood Pressure, Fasting Blood Sugar Level, and Max Heart Rate.

In our statistical analysis, we encoded those that are positive for heart disease as 1, or commonly displayed as a blue color, and 0 for those that are negative for heart disease, or commonly displayed as a red color. Following the encoding for positive or negative for heart disease, we also encoded a Fasting Blood Sugar Level < 120 mg/dl as red, and a Fasting Blood Sugar Level > 120 mg/dl as blue.

3. Methods | Exploratory Data Analysis

3.1. Correlation Matrix

We first investigated the general trend of correlation between predictors using the correlation matrix¹. The last column in this graph specifically shows correlations between heart failures and other predictors. For example, MaxHR has a negative correlation with heart failure, and fasting blood sugar level has a positive correlation with heart failure, etc.

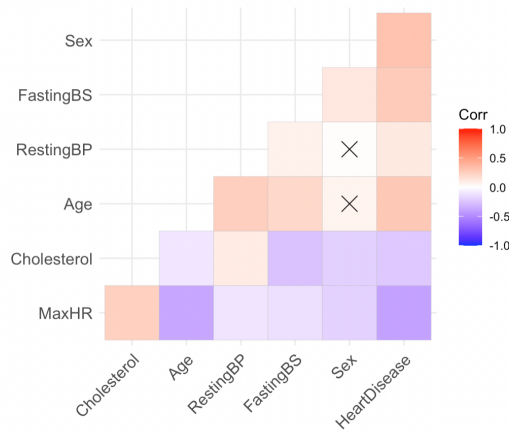


Figure 1. Correlation Matrix with color encoding

3.2. Distribution of Age

We then investigate individual predictors by looking at their distributions grouped by heart disease². The way to look at the graph is that given an age (x), look at the proportion of people with heart disease to that without heart disease. We clearly see that before 50 years old, more people in the sample did not have heart disease, whereas people of 50+ years tend to have more heart disease.

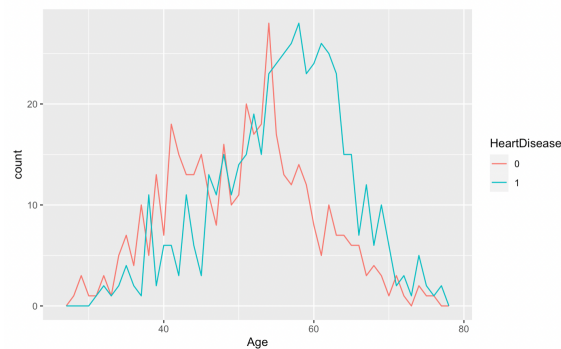


Figure 2. Distribution of Age separated by heart disease

3.3. Distribution of Resting Blood Pressure

For resting blood pressure, the distributions look overlapped a lot². However, in later significance testing, we will find that resting blood pressure is still significant predictor.

3.4. Distribution of Cholesterol

When investigating the cholesterol distribution, we found that there were abnormally large number of 0s recorded in the data, as the huge spikes on the left of the graph indicates⁴. Since it is biologically impossible to have a cholesterol level of 0, a healthy cholesterol level is less than 200 mg/dl, we decide that they were missing data and replaced them with NAs. After fixing the data, the new distribution shows a strong, positive relationship between cholesterol and heart disease⁵. As a side note, the correlation matrix above included the false 0s, thus showing a negative relationship between cholesterol and heart disease, when in fact it should be positive¹.

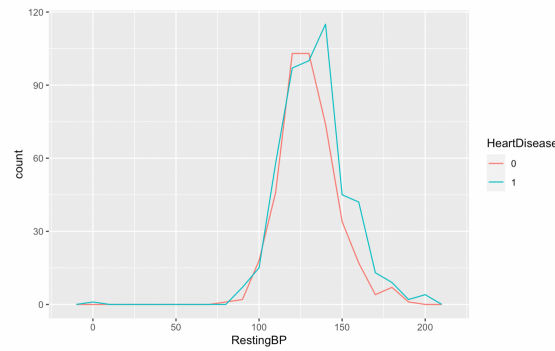


Figure 3. Distribution of Resting Blood Pressure separated by heart disease

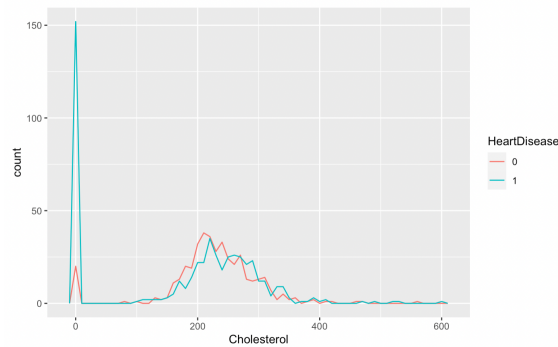


Figure 4. Distribution of Cholesterol Before data fixing

3.5. Distribution of Max Heart Rate, Sex, and Fasting Blood Sugar Level

A healthy max heart rate is calculated with the following formula: $220 - \text{Age}$. For example, a healthy max heart rate for a 70 year old patient would be $220 - 70 = 150$ bpm. From the graphs, it seems that max heart rate have a negative correlation with heart disease. For the two categorical variables Sex and FastingBS, we used a stacked dot plot with the dot size representing sample size in that certain group. While males have a 50-50 heart disease ratio, female patients have proportionally significantly fewer heart diseases. For the Fasting Blood sugar level, we can see that when blood sugar level goes beyond 120 mg/dl, the risks of heart failure increase significantly.

4. ANOVA

We use ANOVA to test if the correlation we visualized were significant. We ran ANOVA for each individual predictor and Heart Disease because multi-collinearity might affect the scores if all predictors are ran at once. The ANOVA table shows all predictors are significant⁹.

5. Result and Conclusion

After looking the various predictors within the heart disease data set, we initially created a correlation matrix to find the relation between common symptoms of heart disease and heart disease. Following the correlation matrix, we took a closer at each predictor through an analysis of positive cases and negative cases of heart disease in relation with given predictor.

Through our findings we can conclude that all six given predictors, Sex, Age, Fasting BS, Resting BP, Cholesterol, and Max Heart Rate, play a large role in indicating heart disease.

Sex: Males are 250% more likely to get heart disease. Age: 60+ are 66% more likely to get heart disease. Fasting BS: Above 120 mg/dl are 65% more likely to get heart disease. Resting BP: Above 145 bpm are 27% more likely to get heart disease. Cholesterol: Above 280 mg/dl are 30%

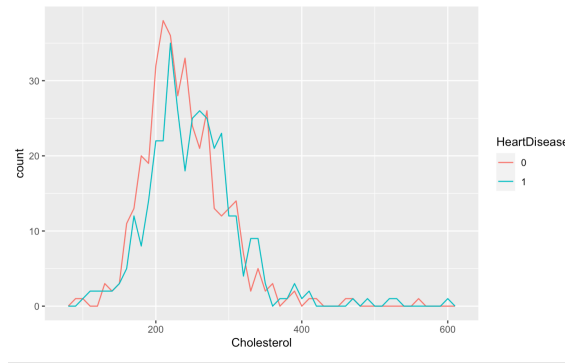


Figure 5. Distribution of Cholesterol After data fixing

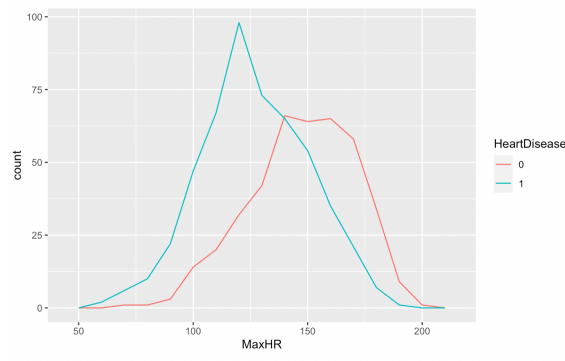


Figure 6. Distribution of Max Heart rate

more likely to get heart disease. Max Heart Rate: Below 150 bpm are 132% more likely to get heart disease.

References

- [1] Kaggle[Online]. Available: <https://www.kaggle.com/datasets/asgharalikhan/mortality-rate-heart-patient-pakistan-hospital>
- [2] R. R. Septiawan, S. Viridi, and Suprijadi, "The Effect of Particle Size Ratio on Porosity of a Particles Deposition Process," *Key Engineering Materials*, vol. 675–676, pp. 647–650, 2016. [Online]. Available: <https://www.scientific.net/KEM.675-676.647>.

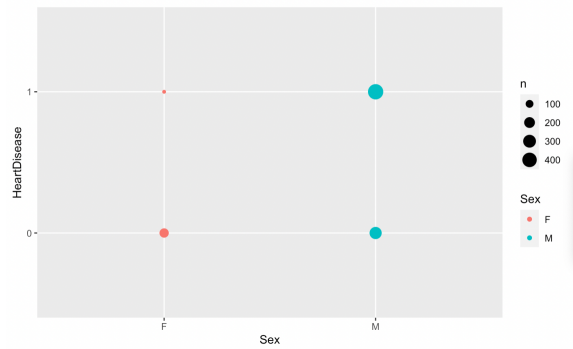


Figure 7. Distribution of Heart Failure by Sex

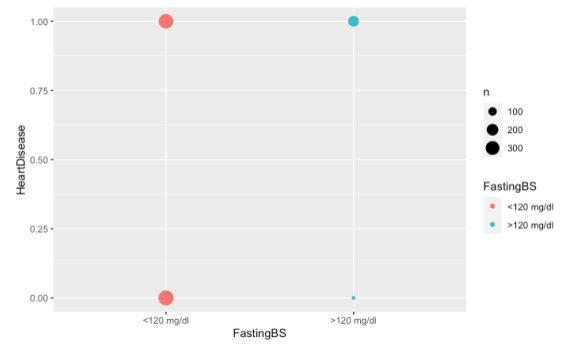


Figure 8. Distribution of Heart Failure by Fasting Blood Sugar Level

Predictor	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significant?
Age	1	18.05	18.048	79.16	<2e-16	yes
Sex	1	21.17	21.168	94.25	<2e-16	yes
Cholesterol	1	2.01	2.0078	8.114	0.00451	yes
MaxHR	1	36.38	36.38	174.9	<2e-16	yes
RestingBP	1	2.63	2.6263	10.73	0.0011	yes
FastingBS	1	16.21	16.21	70.48	<2e-16	yes

Figure 9. ANOVA Table