An statistical analysis of factors influencing the likelihood of a heart attack

Débora Oliveira Santana[1]
[1] Universidade Federal da Bahia

Abbreviated abstract: In Brazil, there are more than 150 thousand cases of heart attack per year, being one of the main causes of mortality in the country. This study aims to analyze the correlation coefficients in 14 risk variables that influence the probability of having a heart attack in an individual in order to identify whether there is any relationship between the variability of these variables and quantify this relationship through hypotheses.

Related publications:
- Silva, M.A. D. de Sousa, A. G., e Schargodsky, H. (1998). Fatores de risco para infarto do miocárdio no brasil: estudo fricas. *Arquivos brasileiros de cardiologia,* 71:667-675.

Previous work, challenge, and approach

This dataset contains a subset of 14 attributes from the Cleveland database. The "**target**" field corresponds to the presence of heart disease in the patient where 0 = greater chance of heart attack and 1 = less chance of heart attack.
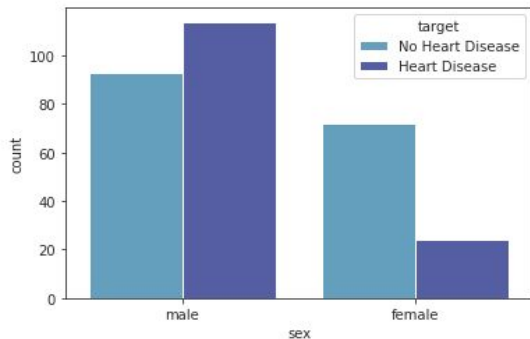
Attribute	Score
thalach	10188.320472
oldpeak	72.644253
ca	66.440765
cp	62.598098
exang	38.914377
chol	23.936394
age	23.286624
trestbps	14.823925
slope	9.804095
sex	7.576835

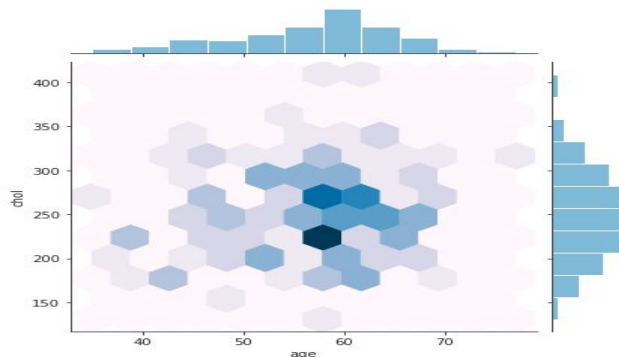**Chi-square Analysis**

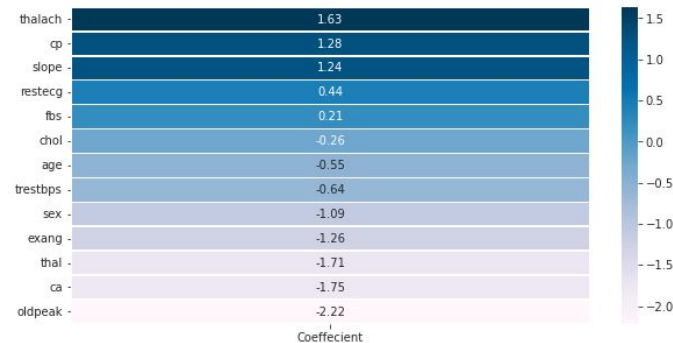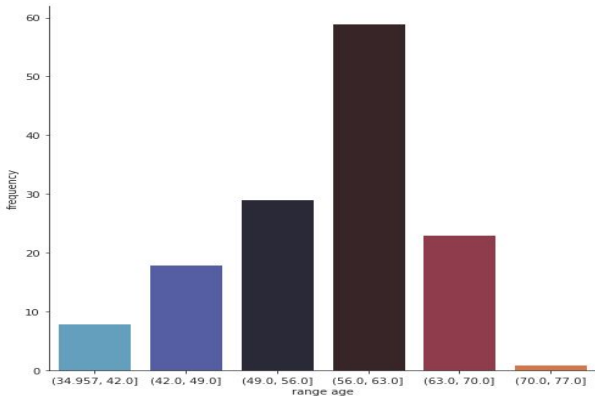**Confusion Matrix**

deboraoliveira@ufba.br
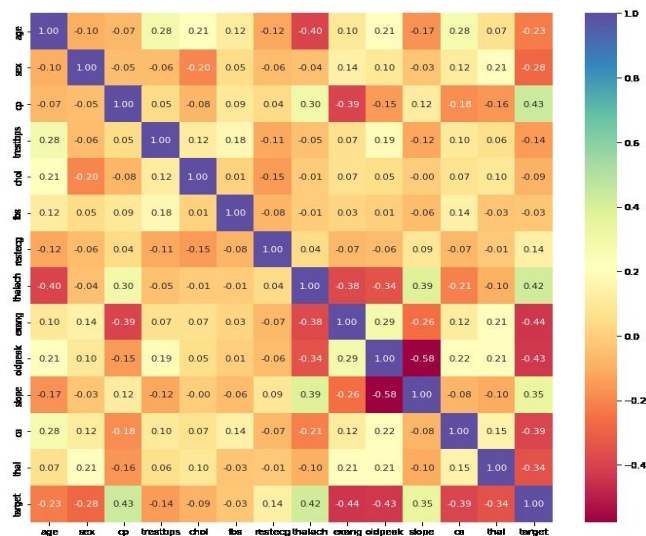
Techniques and Methods

The Cleveland dataset showed that heart attack is more prevalent among men when compared to women.

It is also observed that most individuals with heart disease in the age group above 50 years old or below 65 years old tend to have cholesterol between 200 mg/dL and 300 mg/dL.

After classifying the frequency of data referring to age, it is possible to identify that, in the Cleveland database, the probability of heart attack is more common in individuals aged between 56 and 63 years.

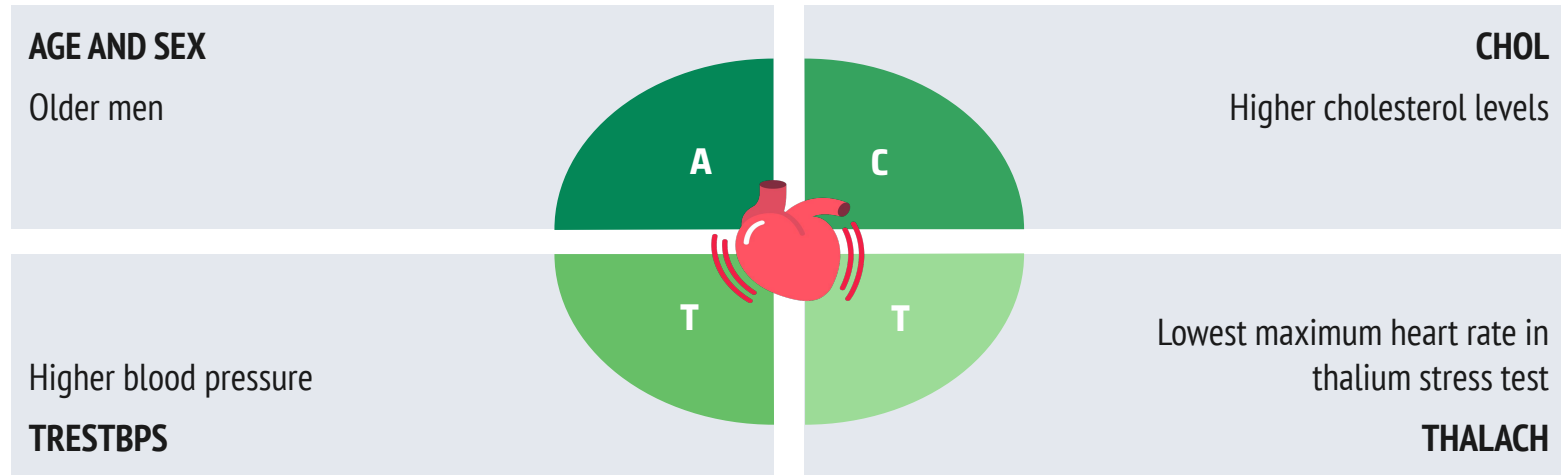|	Coeficient
thalach	1.63
cp	1.28
slope	1.24
restecg	0.44
fbs	0.21
chol	-0.26
age	-0.55
trestbps	-0.64
sex	-1.09
exang	-1.26
thal	-1.71
ca	-1.75
oldpeak	-2.22

Furthermore, after the model predictions, we also identified that the blood pressure and heart rate variables have significant coefficients in relation to the target variable.

deboraoliveira@ufba.br

CSDS 2021

Results and Conclusions

It has been identified that the people who have the greatest chance of developing a heart attack are:

AGE AND SEX	**CHOL**
Older men	Higher cholesterol levels
A	**C**
T	**T**
Higher blood pressure	Lowest maximum heart rate in thalium stress test
TRESTBPS	**THALACH**

The best way to prevent a heart attack is to prevent these risk factors.

ECD
UFBA

CSDS 2021

deboraoliveira@ufba.br