Exploring Vital Health Factors in Heart Disease Risk Assessment

Logistic Regression Model

Bhavani Priya Ganji : bganji@umass.edu

UMassAmherst

Data Analytics and Computational Social Science Program

Background

Heart disease is a global health concern, influenced by age and various health factors. Factors like high cholesterol, hypertension, diabetes, obesity, inactivity, and smoking contribute to its development.

Gender-specific differences exist in prevalence and outcomes, underscoring the importance of tailored approaches.

Recent research suggests that **maximum heart rate** during exercise also plays a role, with gender-specific associations. Understanding these factors is crucial for advancing heart disease research and enhancing preventive strategies.

Dataset: The Heart Disease UCI dataset comprises 303 observations and 14 attributes (including the target variable)..

Research Question

To what extent do age, gender and key health factors, such as cholesterol levels, blood sugar levels, and resting blood pressure, individually contribute to the risk of heart disease?

Does gender, age and the maximum heart rate achieved play a significant role in the occurrence of heart disease?

Hypothesis

Hypothesis 1:

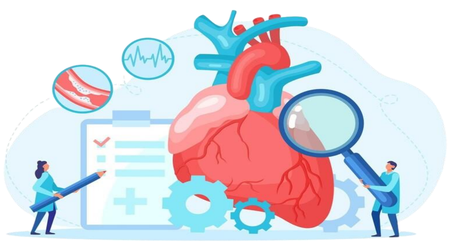
Higher blood sugar levels increase the risk of heart disease, while controlling for age and gender.

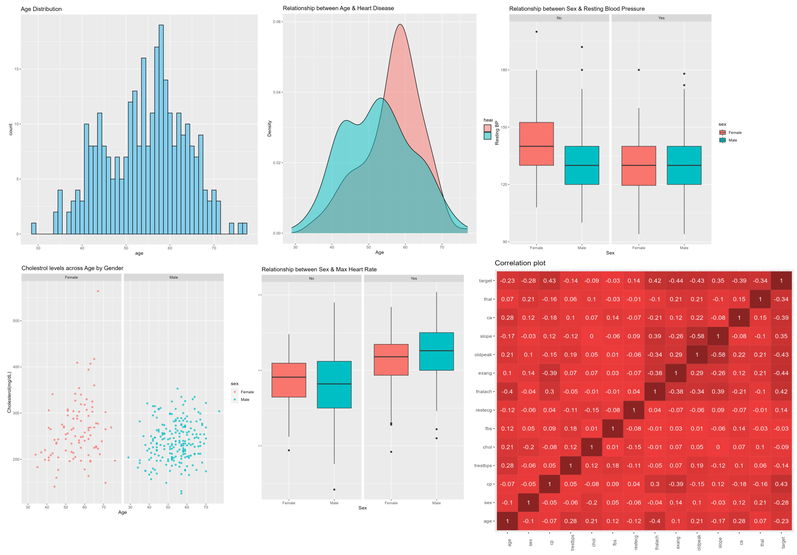
Elevated cholesterol levels significantly impact the risk of heart disease, while controlling for age and gender.

Elevated blood pressure levels have a significant effect on the risk of heart disease, while controlling for age and gender.

Hypothesis 2:

Maximum heart rate achieved during exercise have significant association with the occurrence of heart disease while controlling for age and gender

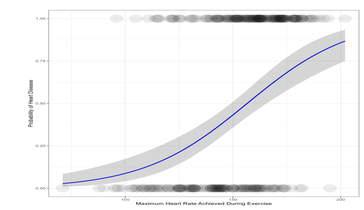




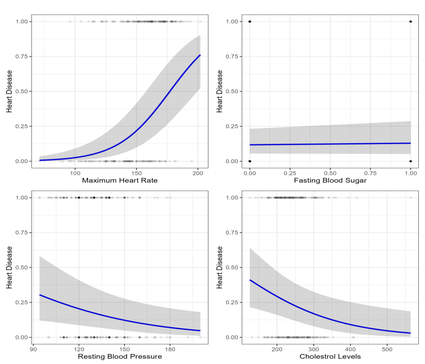
Hypothesis Testing

- Maximum heart rate during exercise shows a significant positive association with heart disease risk.
- Fasting blood sugar does not show a significant association with heart disease risk. Higher cholesterol levels are associated with a significant decrease in heart disease risk. Elevated resting blood pressure is associated with a significant decrease in heart disease.
- Age does not show a significant association with heart disease risk. Being male is significantly associated with a decreased risk of heart disease compared to being female.

```
Call:
glm(formula = heart_disease_binary ~ max_heart_rate + age + sex_binary,
    family = binomial(link = "logit"), data = heart_data_new)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.187349   1.616480  -1.972   0.0486 *
max_heart_rate  0.041393   0.007135   5.802 6.57e-09 ***
age            -0.031774   0.016472  -1.929   0.0537 .
sex_binary     -1.545892   0.311924  -4.956 7.20e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 417.64  on 302  degrees of freedom
Residual deviance: 329.91  on 299  degrees of freedom
AIC: 337.91

Number of Fisher Scoring iterations: 4

[1] "Adjusted R-Squared"

[1] 0.7065894
```




```
Call:
glm(formula = heart_disease_binary ~ max_heart_rate + fasting_bs_binary +
    cholestrol + resting_BP + age + sex_binary, family = binomial(link = "logit"),
    data = heart_data_new)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.293415   1.841907  -0.159   0.8734
max_heart_rate  0.046941   0.007705   6.092 1.12e-09 ***
fasting_bs_binary  0.107264   0.376736   0.285   0.7759
cholestrol     -0.007085   0.002764  -2.563   0.0104 *
resting_BP     -0.020430   0.008444  -2.420   0.0155 *
age            -0.014864   0.017645  -0.842   0.3996
sex_binary     -1.864455   0.343373  -5.430 5.64e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 417.64  on 302  degrees of freedom
Residual deviance: 316.61  on 296  degrees of freedom
AIC: 330.61

Number of Fisher Scoring iterations: 4

[1] "Adjusted R-Squared"

[1] 0.658686
```

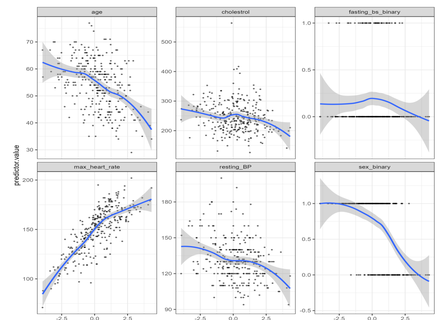


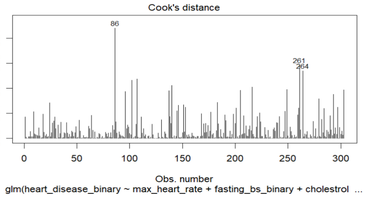
Model Discussion

Comparing models based on AIC and adjusted R-squared scores reveals that the best fit model is the full model including all independent variables, with age and sex serving as control variables. Below is the stargazer output for both full models, second with control variables.

Below are the diagnostics of full model

	Dependent variable:	
	heart_disease_binary	
	(1)	(2)
max_heart_rate	0.044***	0.047***
	(0.007)	(0.008)
fasting_bs_binary	-0.030	0.107
	(0.365)	(0.377)
cholestrol	-0.003	-0.007**
	(0.002)	(0.003)
resting_BP	-0.016**	-0.020**
	(0.008)	(0.008)
age		-0.015
		(0.018)
sex_binary		-1.864***
		(0.343)
Constant	-3.512**	-0.293
	(1.438)	(1.842)
Observations	303	303
Log Likelihood	-175.930	-158.305
Akaike Inf. Crit.	361.860	330.609
Note:	*p<0.1; **p<0.05; ***p<0.01	





Conclusion

- **Maximum heart rate during exercise** is significantly associated with increased heart disease risk, underscoring the importance of considering both physiological and demographic factors in risk assessment.
- Contrary to common assumptions, **males exhibit a lower risk of heart disease compared to females**. However, fasting blood sugar levels above 120mg/dL, cholesterol levels, and resting blood pressure do not significantly predict heart disease risk, indicating the need for further exploration or consideration of additional variables in cardiovascular health assessments.

References: