Predicting Heart Disease:

Refer Book : Predictive Analytics : Chapter 3: Logistics Regression

Book download Link : ebook-dl.com/book/1425

Source of Data : Machine Repository's website at <https://archive.ics.uci.edu/ml/datasets/Statlog+(Heart)>

Data has input for 270 patients of which 120 had heart disease

Parameters Measured:

"AGE", "SEX", "CHESTPAIN", "RESTBP", "CHOL", "SUGAR", "ECG", "MAXHR", "ANGINA", "DEP", "EXERCISE", "FLUOR", "THAL", "OUTPUT")

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Type** | **Description** |
| Age | Numerical | Numerical Age in Years |
| Sex | Binary | Gender |
| Chestpain | Categorical | 4 valued chest pain type |
| RestBP | Numerical | Resting Blood Pressure |
| Chol | Numerical | Serum Cholesterol mg/ dl |
| Sugar | Binary | Is the fasting Sugar level > 120 mg / dl |
| ECG | Categorical | 3 valued resting ECG results |
| MAXHR | Numerical | Max heart beats achieved – beats per minute |
| ANGINA | Binary | Was ANGINA Induced by exercise |
| DEP | Numerical | ST Depression induced by exercise relative to rest |
| EXERCISE | Ordered Categorical | Slope of the Peak Exercise St Segment |
| FLUOR | Numerical | The Number of major vessels colored by fluoroscopy |
| THAL | Categorical | 3 valued thal |
| OUTPUT | Binary | Presence or absence of heart disease |

Summary of the Regression Model Output

Deviance Residuals:

Min 1Q Median 3Q Max

-2.7146 -0.4359 -0.1346 0.3475 2.7891

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -8.326299 3.547833 -2.347 0.018932 \*

AGE -0.020516 0.029579 -0.694 0.487920

SEX 1.625086 0.654564 2.483 0.013039 \*

CHESTPAIN2 1.243550 0.989153 1.257 0.208686

CHESTPAIN3 0.544792 0.860819 0.633 0.526814

CHESTPAIN4 2.330774 0.817282 2.852 0.004346 \*\*

RESTBP 0.025985 0.013270 1.958 0.050218 .

CHOL 0.008379 0.004756 1.762 0.078099 .

SUGAR -1.284948 0.730036 -1.760 0.078388 .

ECG1 1.421208 3.274582 0.434 0.664280

ECG2 0.562649 0.460964 1.221 0.222241

MAXHR -0.014524 0.012699 -1.144 0.252747

ANGINA 1.027517 0.521951 1.969 0.048998 \*

DEP 0.178891 0.280485 0.638 0.523611

EXERCISE 0.616951 0.476428 1.295 0.195337

FLUOR 1.313432 0.307707 4.268 1.97e-05 \*\*\*

THAL6 0.351865 0.991316 0.355 0.722629

THAL7 1.726956 0.509173 3.392 0.000695 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 315.90 on 229 degrees of freedom

Residual deviance: 140.55 on 212 degrees of freedom

AIC: 176.55

Number of Fisher Scoring iterations: 6

# Interpretations

FLUOR THAL7 and CHESTPAIN4 are Strongest Feature Predictors for heart disease

Note We find Age has High P value and thus appears as a weak feature Candidate: The Explanation is thus : Age in the presence of other factors is a weak feature nut if independently regressed will have a positive coeff and low p value as shown below in the <only Age> model

Deviance Residuals:

Min 1Q Median 3Q Max

-1.5027 -1.0691 -0.8435 1.2061 1.6759

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.71136 0.86348 -3.140 0.00169 \*\*

AGE 0.04539 0.01552 2.925 0.00344 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 315.90 on 229 degrees of freedom

Residual deviance: 306.89 on 228 degrees of freedom

AIC: 310.89

***Table showing Model Run on Test Data (1- Heart Disease, 0 – No Heart Disease) Accuracy – 77.5%***

