Total number of patients: 118

**Case 1: (a)** All patients considered with y=bin(HsTNT) and x=CBC|max HsTNT .

The threshold values used to identify heart attacks are 15.4 for female patients and 34.2 for male patients.

Stepwise regression:

pval =

0.0098-WBC

0.2758-RBCi

0.3020-HGB

0.8077-MCV

0.6999-RDW

0.8021-MCH

0.7906-MCHC

0.3223-HCT

0.0472-PLT

0.1981-Age

Discriminant analysis:

Crossvalidation mean square error=0.2373 (stratified 10 fold crossvalidation used-incorrectly classified/correctly classified)

Using Logistic regression from MATLAB (glmfit)

The confusion matrix is

|  |  |
| --- | --- |
| (0,0) 11 | (0,1) 8 |
| (1,0) 13 | (1,1) 63 |

This is considering the cut-off probability as 0.7.

**(b)** y=HsTNT and x=CBC|max HsTNT . The continuous values for HsTNT are taken as the response variable

Stepwise regression:

pvalc =

0.0019

0.9435

0.9342

0.8858

0.7312

0.8748

0.8811

0.9783

0.0149

0.6284

This is ignoring all patients who have their maximum of the HsTNT measurement above 2000.

RMSE=3.461228287493799e+02



When this exclusion is not made, the predictions are pull toward a very high value (~5e4). However, even in that case, the important predictors as WBC and PLT.

However, when patients with HsTNT>500 are excluded, the only crucial parameter is WBC through the stepwise regression

pvalc =

0.0107

0.9751

0.9821

0.5046

0.3486

0.7580

0.2940

0.9320

0.0961

0.3495

“Step 1, added column 1, p=0.0106849

Final columns included: 1”

A log transformation was made for HsTNT to reduce the variation in HsTNT due to some extreme measurements (5e4)

In this case, the significant term was WBC, according to stepwise

Adding x1, FStat = 6.3192, pValue = 0.014003



**c)** y=bin(HsTNT) and x=CBC|t=0

Using stepwise regression, the significant variables are still WBC and PLT

pval =

0.0032

0.8057

0.8508

0.8617

0.5576

0.8668

0.8683

0.8793

0.0350

0.1216

Discriminant analysis:

Crossvalidation mean square error= 0.2627 (stratified 10 fold crossvalidation used-incorrectly classified/correctly classified)

Using Logistic regression from MATLAB (glmfit)

The confusion matrix is

|  |  |
| --- | --- |
| (0,0) 13 | (0,1) 6 |
| (1,0) 9 | (1,1) 67 |

This is considering the cut-off probability as 0.7.

For the test set, considering the same probability

Ctest =

0 4

6 13

d) y=HsTNT and x=CBC|t=0 . The continuous values for HsTNT are taken as the response variable

Stepwise regression:

pvalc =

0.0019

0.9435

0.9342

0.8858

0.7312

0.8748

0.8811

0.9783

0.0149

0.6284

This is ignoring all patients who have their maximum of the HsTNT measurement above 2000.

RMSE = 3.461228e+02

 

When this exclusion is not made, the predictions are pull toward a very high value (~5e4). However, even in that case, the important predictors as WBC and PLT.

However, when patients with HsTNT>500 are excluded, the only crucial parameter is WBC through the stepwise regression

pvalc =

0.0107

0.9751

0.9821

0.5046

0.3486

0.7580

0.2940

0.9320

0.0961

0.3495

“Step 1, added column 1, p=0.0106849

Final columns included: 1”

A log transformation was made for HsTNT to reduce the variation in HsTNT due to some extreme measurements (5e4)

**Case 2:** Subset analysis on patients who have an increasing HsTNT tendency over time (+ve slope)

The Predictors are the CBC measurements at max HsTNT

80 patients had +ve slope for HsTNT. Patients 2, 30 and 110 were excluded since they did not have a monotonic curve for HsTNT variation over time.

For binary outcomes, a logistic regression model was performed using Holdout crossvalidation. The confusion matrix considering the cutoff probability as 0.7

|  |  |
| --- | --- |
| (0,0) 12 | (0,1) 4 |
| (1,0) 15 | (1,1) 33 |

Confusion matrix for the test set isolated using Holdout cross validation

3 4

1 8

Using continuous HsTNT and ignoring patients with HsTNT higher than 2000, stepwise regression also suggests WBC and PLT as the significant variables

pvalc =

0.0000

0.2533

0.5138

0.5365

0.4612

0.3575

0.3770

0.4380

0.0214

0.5228

 

When the threshold for exclusion was reduced to 500, the results are similar to that of the complete patient data set. WBC turns out to be most significant. Also, with log transformation, the results are similar to before (only WBC significant).

“Adding x1, FStat = 6.3192, pValue = 0.014003”

The observations when considering the predictors as CBC|t=0 are same as that observed for the predictors=CBC measurements at maximum HsTNT.