Surgery

Stephen Smitshoek

19/05/2022

```
## Warning: package 'caTools' was built under R version 4.1.3
```

#Generalized Linear Model

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##
## Call:
## glm(formula = Risk1Yr ~ DGN + PRE4 + PRE5 + PRE7 + PRE8 + PRE9 +
##      PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 + PRE32 +
##      AGE, family = binomial(), data = surgery_training)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5492  -0.4729  -0.4013  -0.1333   2.4375
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -18.92498 3956.18101  -0.005  0.99618
## DGNDGN2       15.31467 3956.18043   0.004  0.99691
## DGNDGN3       15.45367 3956.18038   0.004  0.99688
## DGNDGN4       14.94343 3956.18045   0.004  0.99699
## DGNDGN5       18.37253 3956.18045   0.005  0.99629
## DGNDGN6        1.03555 4516.80991   0.000  0.99982
## DGNDGN8        2.50717 5594.88400   0.000  0.99964
## PRE4           0.10914    0.51190   0.213  0.83117
## PRE5          -0.34224    0.59243  -0.578  0.56347
## PRE7T          0.69537    0.82621   0.842  0.39999
## PRE8T          0.15447    0.61162   0.253  0.80061
## PRE9T          1.89147    0.70342   2.689  0.00717 **
## PRE10T        -0.16896    0.48719  -0.347  0.72873
## PRE11T         0.23354    0.51511   0.453  0.65028
## PRE14OC12      0.37230    0.46260   0.805  0.42093
## PRE14OC13      0.41772    0.96271   0.434  0.66436
## PRE14OC14      1.93607    0.89510   2.163  0.03054 *
## PRE17T         0.95645    0.61299   1.560  0.11869
## PRE19T       -15.49146 2769.73682  -0.006  0.99554
## PRE25T       -16.31404 1497.11660  -0.011  0.99131
## PRE30T         2.39677    1.01779   2.355  0.01853 *
## PRE32T       -14.92813 2543.82797  -0.006  0.99532
## AGE           -0.01398    0.02483  -0.563  0.57342
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 217.37  on 304  degrees of freedom
## Residual deviance: 179.04  on 282  degrees of freedom
## AIC: 225.04
##
## Number of Fisher Scoring iterations: 16
```

The factors that are the most significant in the model are PRE4, PRE9, PRE14 and PRE17. These are the only factors with a P value less than 0.05

Confidence Matrix

```
##              Predicted_Value
## Actual_Value FALSE TRUE
##           F   126    4
##           T    32    3
```

Accuracy

```
## [1] 0.7818182
```

```
library(foreign)
library(caTools)

setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")
surgery_data <- read.arff('data\\ThoraricSurgery.arff')

surgery_data_split <- sample.split(surgery_data, SplitRatio = 0.7)

surgery_training <- subset(surgery_data, surgery_data_split==TRUE)
surgery_testing <- subset(surgery_data, surgery_data_split==FALSE)

surgery_glm <- glm(Risk1Yr ~ DGN + PRE4 + PRE5 + PRE7 + PRE8 + PRE9 + PRE10 +
                   PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 + PRE32 +
AGE,
                   data=surgery_training, family=binomial())


summary(surgery_glm)

response <- predict(surgery_glm, surgery_testing, type="response")

confmatrix <- table(Actual_Value=surgery_testing$Risk1Yr, Predicted_Value =
response > 0.5)
confmatrix

(confmatrix[[1,1]] + confmatrix[[2,2]]) / sum(confmatrix)
```