ASSIGNMENT 10.2.1

Kiran Komati

2021-05-22

```
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_knit$set(root.dir = 'C:/Users/kiran/dsc520')
```

Load the necessary libraries and load the Thoraric Surgery Data

```
library('foreign')
ThS<-read.arff('data/ThoraricSurgery.arff')
```

i) Fit a binary logistic regression model to the data set that predicts whether or not the patient survived for one year (the Risk1Y variable) after the surgery. Use the glm() function to perform the logistic regression. See Generalized Linear Models for an example. Include a summary using the summary() function in your results.

```
Ths_glm <- glm(formula = Risk1Yr ~ DGN + PRE14, family = binomial(), data = ThS)
summary(Ths_glm)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ DGN + PRE14, family = binomial(), data = ThS)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2783  -0.5265  -0.5265  -0.4220   2.2195
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -16.0341  1455.3976  -0.011  0.99121
## DGNDGN2       14.1160  1455.3976   0.010  0.99226
## DGNDGN3       13.6601  1455.3976   0.009  0.99251
## DGNDGN4       13.9480  1455.3976   0.010  0.99235
## DGNDGN5       15.5613  1455.3976   0.011  0.99147
## DGNDGN6        0.2166  1625.6133   0.000  0.99989
## DGNDGN8       15.8001  1455.3983   0.011  0.99134
## PRE14OC12      0.4680     0.3110   1.505  0.13243
## PRE14OC13      1.3476     0.5723   2.355  0.01853 *
## PRE14OC14      1.8254     0.5752   3.174  0.00151 **
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 395.61  on 469  degrees of freedom
## Residual deviance: 367.60  on 460  degrees of freedom
## AIC: 387.6
##
## Number of Fisher Scoring iterations: 14
```

ii) According to the summary, which variables had the greatest effect on the survival rate?

Residual Deviance shows that using variables DGN,PRE14 improved the ovreall model than when only the other variables are used.

iii) To compute the accuracy of your model, use the dataset to predict the outcome variable. The percent of correct predictions is the accuracy of your model. What is the accuracy of your model?

```
ThS$predicted.probabilities<-fitted(Ths_glm)
res <- predict(Ths_glm,type="response")
confMatrix <- table(Actual_Value=ThS$Risk1Yr,Predicted_Value=res>0.5)
confMatrix
```

```
##              Predicted_Value
## Actual_Value FALSE TRUE
##            F   399    1
##            T    69    1
```

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

```
## [1] 0.8510638
```