Karthikeyan Chellamuthu_10.2.1

Karthikeyan Chellamuthu

2/19/2022

10.2.a 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.

```
glm1 <- glm(Risk1Yr ~ .,family = binomial(),data = training_df)
```

10.2.b According to the summary, which variables had the greatest effect on the survival rate?

```
##
## Call:
## glm(formula = Risk1Yr ~ ., family = binomial(), data = training_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5213  -0.5315  -0.3852  -0.2381   2.5303
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -17.68858 2399.54556  -0.007  0.99412
## DGNDGN2       15.18836 2399.54481   0.006  0.99495
## DGNDGN3       14.36609 2399.54478   0.006  0.99522
## DGNDGN4       15.00153 2399.54482   0.006  0.99501
## DGNDGN5       16.66721 2399.54487   0.007  0.99446
## DGNDGN6        0.47857 2758.97758   0.000  0.99986
## DGNDGN8       34.95374 3393.46878   0.010  0.99178
## PRE4           0.11409    0.40253   0.283  0.77684
## PRE5          -0.23366    0.44821  -0.521  0.60215
## PRE6PRZ1      -0.72129    0.61167  -1.179  0.23832
## PRE6PRZ2      -0.86086    0.96892  -0.888  0.37429
## PRE7T          1.35541    0.64575   2.099  0.03582 *
## PRE8T         -1.09804    0.64418  -1.705  0.08828 .
## PRE9T          1.33931    0.65734   2.037  0.04160 *
## PRE10T         0.91146    0.58740   1.552  0.12073
## PRE11T         0.51341    0.48542   1.058  0.29021
## PRE14OC12      0.70084    0.42655   1.643  0.10037
## PRE14OC13      0.58796    0.82954   0.709  0.47847
## PRE14OC14      2.04690    0.71908   2.847  0.00442 **
## PRE17T         0.86074    0.52998   1.624  0.10435
## PRE19T       -14.72256 1644.14235  -0.009  0.99286
## PRE25T         0.09668    1.18193   0.082  0.93481
## PRE30T         1.11727    0.62181   1.797  0.07237 .
## PRE32T       -14.52657 2399.54478  -0.006  0.99517
## AGE           -0.00766    0.02293  -0.334  0.73838
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 276.93  on 328  degrees of freedom
## Residual deviance: 229.17  on 304  degrees of freedom
## AIC: 279.17
##
## Number of Fisher Scoring iterations: 15
```

From above, PRE7, PRE9 and PRE14 have the greatest effect on survival rate as P < 0.05.

```
glm1 <- glm(Risk1Yr ~ PRE7+PRE9+PRE14,family = binomial(), data = training_df)
```

10.2.c 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?

Prediction matrix

```
##          actual
## predicted   F   T
##        No 120  21
```

Accuracy of the model is 0.8510638