10.2.1 Exercise

Felipe Rodriguez

2023-02-19

```
options(scipen=999)
```

```
library(foreign)
setwd('/Users/feliperodriguez/Library/CloudStorage/OneDrive-BellevueUniversity/Github/dsc520//data/')
surgery_data <- read.arff('ThoraricSurgery.arff')
```

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.

```
risk1yr_glm <- glm(Risk1Yr~., data=surgery_data, family=binomial)
summary(risk1yr_glm)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ ., family = binomial, data = surgery_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6084  -0.5439  -0.4199  -0.2762   2.4929
##
## Coefficients:
##                Estimate  Std. Error z value Pr(>|z|)
## (Intercept)  -16.551698 2399.545216  -0.007  0.99450
## DGNDGN2       14.736276 2399.544756   0.006  0.99510
## DGNDGN3       14.180552 2399.544735   0.006  0.99528
## DGNDGN4       14.608329 2399.544765   0.006  0.99514
## DGNDGN5       16.381321 2399.544797   0.007  0.99455
## DGNDGN6        0.408853 2673.049070   0.000  0.99988
## DGNDGN8       18.032862 2399.545187   0.008  0.99400
## PRE4          -0.227245    0.184911  -1.229  0.21909
## PRE5          -0.030304    0.017858  -1.697  0.08971 .
## PRE6PRZ1      -0.442715    0.519908  -0.852  0.39448
## PRE6PRZ2      -0.293701    0.790690  -0.371  0.71030
## PRE7T          0.715341    0.555560   1.288  0.19788
## PRE8T          0.174337    0.389186   0.448  0.65419
```

```
## PRE9T          1.368216    0.486768   2.811  0.00494 **
## PRE10T         0.576958    0.482570   1.196  0.23185
## PRE11T         0.516181    0.396480   1.302  0.19295
## PRE14OC12      0.439364    0.330092   1.331  0.18318
## PRE14OC13      1.179207    0.616546   1.913  0.05580 .
## PRE14OC14      1.652973    0.609362   2.713  0.00668 **
## PRE17T         0.926593    0.444462   2.085  0.03709 *
## PRE19T       -14.655378 1653.541054  -0.009  0.99293
## PRE25T        -0.097894    1.003314  -0.098  0.92227
## PRE30T         1.083997    0.499030   2.172  0.02984 *
## PRE32T       -13.983295 1645.313892  -0.008  0.99322
## AGE           -0.009506    0.018099  -0.525  0.59944
## ---
## 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: 341.19  on 445  degrees of freedom
## AIC: 391.19
##
## Number of Fisher Scoring iterations: 15
```

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

In order to identify which variables have the greatest effect on survival rate we look at the Significance codes at the bottom of the summary. Any value that has an asterisk can be considered a variable that has an affect on survival rate since their P-Value is less than .05. The values identified, from greatest to least are: PRE14OC14, PRE9T, PRE17T, and PRE30T.

Predict the Accuracy of your moodel

```
response_glm <- predict(risk1yr_glm, type = "response")
prediction <- table(Actual_Value = surgery_data$Risk1Yr, Predicted_Value = response_glm > .5)
prediction
```

```
##             Predicted_Value
## Actual_Value FALSE TRUE
##            F   390   10
##            T    67    3
```

```
accuracy_risk1yr <- ((prediction[[1,1]] + prediction[[2,2]]) / sum(prediction))
accuracy_risk1yr
```

```
## [1] 0.8361702
```