Assignment 7.1: Fit a Logistic Regression Model to a Thoracic Surgery Binary Dataset

Kurt Stoneburner

July 13th 2020

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.

```
##
## Call:
## glm(formula = Risk1Yr ~ DGN + PRE7 + PRE8 + PRE10 + PRE11 + PRE14 +
##     PRE30, family = binomial(), data = dataset_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5035  -0.5373  -0.4556  -0.3467   2.3586
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -16.9400  1455.3976  -0.012  0.99071
## DGNDGN2       13.8053  1455.3976   0.009  0.99243
## DGNDGN3       13.3529  1455.3976   0.009  0.99268
## DGNDGN4       13.7082  1455.3976   0.009  0.99248
## DGNDGN5       15.2847  1455.3977   0.011  0.99162
## DGNDGN6        0.1518  1623.6378   0.000  0.99993
## DGNDGN8       16.7114  1455.3983   0.011  0.99084
## PRE7T          0.4124     0.5193   0.794  0.42715
## PRE8T          0.2609     0.3743   0.697  0.48581
## PRE10T         0.3484     0.3445   1.011  0.31194
## PRE11T         0.4593     0.3381   1.359  0.17426
## PRE14OC12      0.4571     0.3204   1.427  0.15363
## PRE14OC13      1.2538     0.6008   2.087  0.03690 *
## PRE14OC14      1.7179     0.5924   2.900  0.00373 **
## PRE30T         0.9168     0.4782   1.917  0.05524 .
## ---
## 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: 357.20  on 455  degrees of freedom
## AIC: 387.2
##
```

```
## Number of Fisher Scoring iterations: 14
```

b. According to the summary, which variables had the greatest effect on the survival rate? The DGN variable had the greatest effect on survival rate in the model. DGN values of DGN2, DGN3,DGN5,DG5, and DGN8 greatly influenced the survival rate. However the DGN value DGN6 had very little effect on the survival rate. These values represent standardized diagnosis. It stands to reason, that the primary predictor of one year survivability is the condition that is to be cured.

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?

```
## [1] "The model successfully predicted 398 outcomes. Out of 470 elements."
```

```
## [1] "Leaving 72 elements predicted incorrectly. For an accuracy score of 85%"
```