EXERCISE 10.2

Ramsey King

2021-05-22

Fit a Logistic Regression Model to Thoracic Surgery Binary Dataset

1. 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 ~ PRE11 + PRE30 + PRE9 + PRE14, family = binomial(),
##     data = surgery_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2573  -0.5465  -0.4431  -0.3847   2.4741
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.0125     0.4656  -6.470 9.79e-11 ***
## PRE11T        0.6139     0.3263   1.882  0.05990 .
## PRE30T        0.7408     0.4315   1.717  0.08602 .
## PRE9T         1.0333     0.4422   2.337  0.01945 *
## PRE14OC12     0.4458     0.3100   1.438  0.15040
## PRE14OC13     1.2674     0.5708   2.220  0.02640 *
## PRE14OC14     1.8437     0.5616   3.283  0.00103 **
## ---
## 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: 370.81  on 463  degrees of freedom
## AIC: 384.81
##
## Number of Fisher Scoring iterations: 5
```

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

It appears that the variable PRE9T, which is Dyspnoea before surgery had the greatest impact. The reason is that the b estimate is 1.102, and the Pr(>|z|) value is 0.01, which is less than 0.05. Also, PRE14OC14, which is the largest original tumor size, had a b estimate of 1.84, and a Pr(>|z|) value of 0.00103, which is less than 0.05.

3. 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] 0.8516129
```

Fit a Logistic Regression Model

1. Fit a logistic regression model to the binary-classifier-data.csv dataset

```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(), data = binary_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3728  -1.1697  -0.9575   1.1646   1.3989
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.424809   0.117224   3.624  0.00029 ***
## x           -0.002571   0.001823  -1.411  0.15836
## y           -0.007956   0.001869  -4.257 2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2075.8  on 1497  degrees of freedom
## Residual deviance: 2052.1  on 1495  degrees of freedom
## AIC: 2058.1
##
## Number of Fisher Scoring iterations: 4
```

2. What is the accuracy of the logistic regression classifier?

```
## [1] 0.6012146
```