Week 10 Assingment

Surenther

2024-10-30

Load ARFF file

```
library(foreign)
myfile <- read.arff("ThoracicSurgery.arff")
```

Logistic Regression

```
# Convert 'Risk1Yr' to a factor for binary logistic regression
myfile$Risk1Yr <- factor(myfile$Risk1Yr, levels = c("F", "T"))

# Fit a binary logistic regression model
mymodel <- glm(Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + PRE8 + PRE9 +
               PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 +
               PRE32 + AGE, data = myfile, family = binomial)
summary(mymodel)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + PRE8 +
##     PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 +
##     PRE32 + AGE, family = binomial, data = myfile)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.655e+01  2.400e+03  -0.007  0.99450
## DGNDGN2      1.474e+01  2.400e+03   0.006  0.99510
## DGNDGN3      1.418e+01  2.400e+03   0.006  0.99528
## DGNDGN4      1.461e+01  2.400e+03   0.006  0.99514
## DGNDGN5      1.638e+01  2.400e+03   0.007  0.99455
## DGNDGN6      4.089e-01  2.673e+03   0.000  0.99988
## DGNDGN8      1.803e+01  2.400e+03   0.008  0.99400
## PRE4        -2.272e-01  1.849e-01  -1.229  0.21909
## PRE5        -3.030e-02  1.786e-02  -1.697  0.08971 .
## PRE6PRZ1    -4.427e-01  5.199e-01  -0.852  0.39448
## PRE6PRZ2    -2.937e-01  7.907e-01  -0.371  0.71030
## PRE7T        7.153e-01  5.556e-01   1.288  0.19788
## PRE8T        1.743e-01  3.892e-01   0.448  0.65419
## PRE9T        1.368e+00  4.868e-01   2.811  0.00494 **
```

```
## PRE10T       5.770e-01  4.826e-01   1.196  0.23185
## PRE11T       5.162e-01  3.965e-01   1.302  0.19295
## PRE14OC12    4.394e-01  3.301e-01   1.331  0.18318
## PRE14OC13    1.179e+00  6.165e-01   1.913  0.05580 .
## PRE14OC14    1.653e+00  6.094e-01   2.713  0.00668 **
## PRE17T       9.266e-01  4.445e-01   2.085  0.03709 *
## PRE19T      -1.466e+01  1.654e+03  -0.009  0.99293
## PRE25T      -9.789e-02  1.003e+00  -0.098  0.92227
## PRE30T       1.084e+00  4.990e-01   2.172  0.02984 *
## PRE32T      -1.398e+01  1.645e+03  -0.008  0.99322
## AGE         -9.506e-03  1.810e-02  -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
```

Findings

From the summary, the variables with the most significant effect on survival (Risk1Yr) are:

PRE9T (Dyspnoea before surgery True): This predictor has a coefficient of 1.368 and a p-value of 0.00494, indicating a statistically significant positive impact on survival.

PRE14OC14 (clinical TNM Large): With a coefficient of 1.653 and a p-value of 0.00668, this variable also shows a strong positive influence on survival.

PRE17T (Type 2 DM - diabetes mellitus True): This predictor has a coefficient of 0.9266 and a p-value of 0.03709, indicating a positive and statistically significant effect on survival.

PRE30T (Smoking True): This variable has a coefficient of 1.084 and a p-value of 0.02984, suggesting it positively influences survival.

These variables showed the lowest p-values, indicating they are the most influential predictors in the model for determining one-year survival post-surgery.

Predict the accuracy

```r
# Predict probabilities
predicted_probs <- predict(mymodel, type = "response")

# Convert probabilities to binary predictions with a 0.5 threshold
predicted_classes <- ifelse(predicted_probs > 0.5, "T", "F")

# Calculate accuracy
```

```r
accuracy <- mean(predicted_classes == myfile$Risk1Yr)
accuracy
```

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

Findings

An accuracy of 0.836 (or 83.6%) means that the logistic regression model correctly predicted the survival outcome (Risk1Yr) after surgery for approximately 83.6% of the patients in the dataset. This suggests that the model is reasonably effective at distinguishing between patients who survived one year post-surgery