Week10_Thoracic_Surgery

Akila Selvaraj

2/19/2022

Logistic regression

```
library('mlogit')
```

```
## Loading required package: dfidx
```

```
##
## Attaching package: 'dfidx'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
mlChat <- mlogit.data(Thoracic_df, choice = "Risk1Yr", shape = "wide")

Thoracic_glm <- glm(Risk1Yr ~ AGE + DGN + PRE4 + PRE5  + PRE6 + PRE7 + PRE8 + PRE9 +
                        PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 +
                        PRE30 + PRE32 , data = Thoracic_df, family = binomial(link='logit'))
```

Summary of logistic model

```
summary(Thoracic_glm)
```

```
##
## Call:
## glm(formula = Risk1Yr ~ AGE + DGN + PRE4 + PRE5 + PRE6 + PRE7 +
##     PRE8 + PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 +
##     PRE30 + PRE32, family = binomial(link = "logit"), data = Thoracic_df)
##
## 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) -1.655e+01  2.400e+03  -0.007  0.99450
## AGE         -9.506e-03  1.810e-02  -0.525  0.59944
## 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
## ---
## 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
```

Out of all the variables, few of the variables like PRE9T,PRE14OC14,PRE17T,PRE30T turned out to be significant.

PRE9T - The beta coefficient against this variable is 1.368. Let us convert this value into odds by taking the $\exp(1.368) = 3.927$. The value indicates that the odds of an individual being in Risk1Yr increases by 3.9% than the one with PRE9 F value.

Out of 2 levels, F level became the reference, and thus all PRE variables are inferred in comparison to the referenced variable. That is how we interpret the categorical variables.

Age - The beta coefficient of the age variable is 0.009506, which is in the logit of odds terms. When we convert this to odds by taking $\exp(0.023362)$ we get 1.0095. The value indicates that as age increase by one more unit, then the odds of an individual being in the Risk1Yr F group will increase by 2%.

Predicting outcome variable

```
predicted_df <- predict(Thoracic_glm, Thoracic_df, type = "response")
summary(predicted_df)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.07105 0.10992 0.14894 0.17779 0.74465
```

Percentage of correct predictions - Accuracy of the model

```
library("performance")
performance_pcp(Thoracic_glm, ci = 0.95, method = "Herron", verbose = TRUE)
```

```
## # Percentage of Correct Predictions from Logistic Regression Model
##
##    Full model: 77.77% [74.01% - 81.53%]
##    Null model: 74.65% [70.72% - 78.58%]
##
## # Likelihood-Ratio-Test
##
##    Chi-squared: 54.421
##    df: 24.000
##    p-value:  0.000
```

AS PCP is above 50%, we can say that the PCP of the full model is considerably above the null model's PCP.

The likelihood-ratio test indicates that the model has a significantly better fit than the null-model as p is less than 0.05.

The accuracy of the model from percentage of correct predictions is 77%.