

Week 10 Assignment

Mithil Patel

2022-05-16

Binary Logistic Regression

Call:

```
glm(formula = Risk1Yr ~ DGN + PRE4 + PRE5 + PRE6 + PRE7 + PRE8 +  
    PRE9 + PRE10 + PRE11 + PRE14 + PRE17 + PRE19 + PRE25 + PRE30 +  
    PRE32, family = binomial(), data = df)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5863	-0.5466	-0.4222	-0.2772	2.4431

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-17.23246	2399.54488	-0.007	0.99427
DGNDGN2	14.76385	2399.54477	0.006	0.99509
DGNDGN3	14.20405	2399.54475	0.006	0.99528
DGNDGN4	14.59889	2399.54478	0.006	0.99515
DGNDGN5	16.37763	2399.54481	0.007	0.99455
DGNDGN6	0.34014	2671.68445	0.000	0.99990
DGNDGN8	18.09038	2399.54520	0.008	0.99398
PRE4	-0.20245	0.17871	-1.133	0.25730
PRE5	-0.02977	0.01784	-1.669	0.09512 .
PRE6PRZ1	-0.42701	0.51881	-0.823	0.41047
PRE6PRZ2	-0.33256	0.78708	-0.423	0.67264
PRE7T	0.70554	0.55439	1.273	0.20314
PRE8T	0.17496	0.38895	0.450	0.65284
PRE9T	1.34960	0.48723	2.770	0.00561 **
PRE10T	0.55381	0.48037	1.153	0.24895
PRE11T	0.49030	0.39313	1.247	0.21233
PRE140C12	0.42876	0.32934	1.302	0.19296
PRE140C13	1.17277	0.61741	1.899	0.05750 .
PRE140C14	1.67574	0.60931	2.750	0.00596 **
PRE17T	0.90773	0.44343	2.047	0.04065 *
PRE19T	-14.60971	1655.73244	-0.009	0.99296
PRE25T	-0.12620	1.01218	-0.125	0.90077
PRE30T	1.09028	0.50090	2.177	0.02951 *
PRE32T	-13.95862	1643.39665	-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: 341.46 on 446 degrees of freedom
AIC: 389.46

Number of Fisher Scoring iterations: 15

Based on the $\Pr(>|z|)$ value in model summary, the variable that has the greatest effect on survival rate is *PRE9T*.

Model accuracy percentage:

[1] 0.8305556

Logistic Regression Classifier

Confusion matrix:

	0	1
0	0.2863818	0.1909212
1	0.2256342	0.2970628

Model accuracy percentage:

[1] 58.34446