Week 9 Logical Regression

Xin TANG

2023-08-1

Logical Regression: surgery and binary data

Surgery fit model

working with the thoracic surgery data set

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. include a summary()

Inline Code

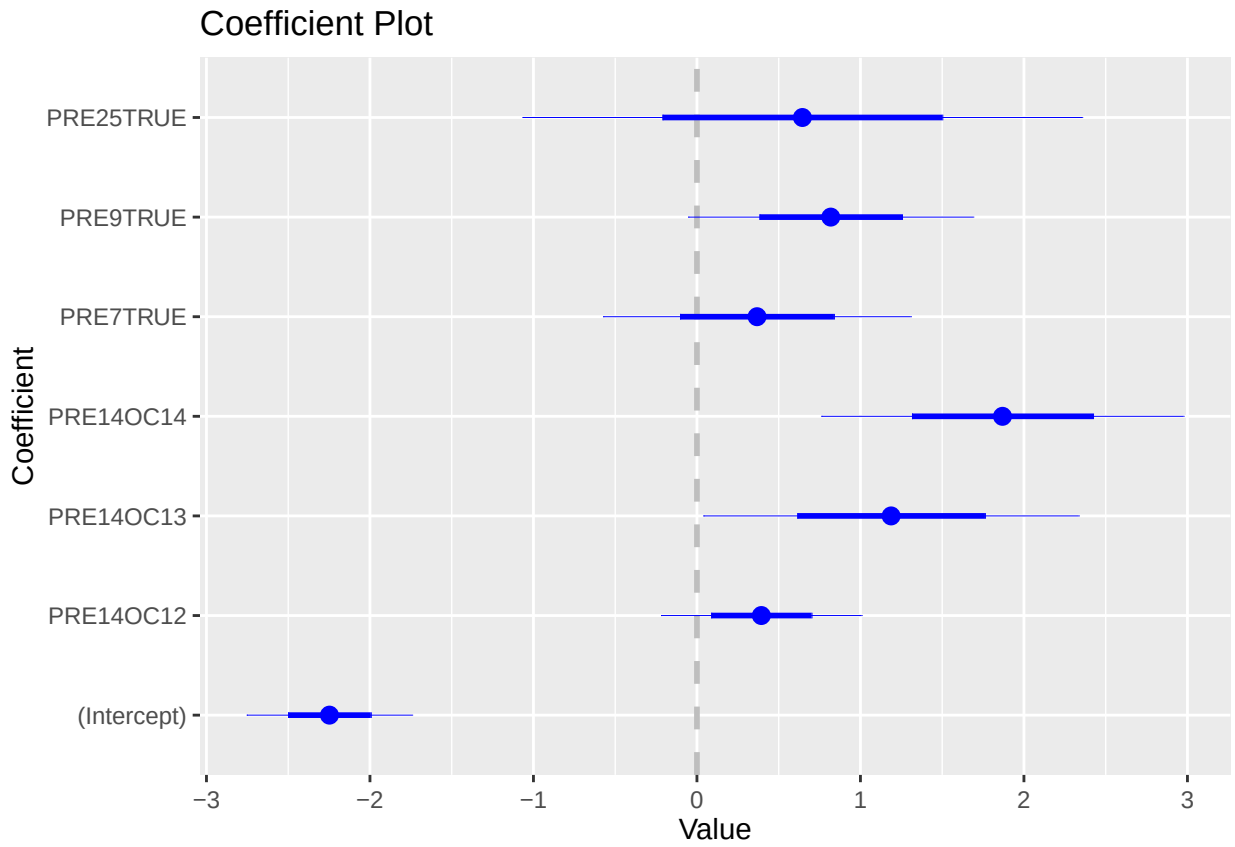
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Attaching package: 'scales'

## The following object is masked from 'package:readr':
##
##     col_factor
```




```
##
## Call:
## glm(formula = Risk1Yr ~ PRE14 + PRE7 + PRE9 + PRE25, family = binomial(link = "logit"),
##     data = surgery9)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.2476     0.2528  -8.892  < 2e-16 ***
## PRE14OC12     0.3939     0.3067   1.284 0.198989
## PRE14OC13     1.1876     0.5738   2.070 0.038481 *
## PRE14OC14     1.8689     0.5537   3.375 0.000737 ***
## PRE7TRUE      0.3675     0.4704   0.781 0.434661
## PRE9TRUE      0.8184     0.4358   1.878 0.060366 .
## PRE25TRUE     0.6450     0.8556   0.754 0.450930
## ---
## 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: 377.28  on 463  degrees of freedom
## AIC: 391.28
##
## Number of Fisher Scoring iterations: 4
```

2. From model and chart, the **PRE14** has greatest effect on the survival rate

3. Compute the accuracy of your model, use the dataset to predict the outcome variable.

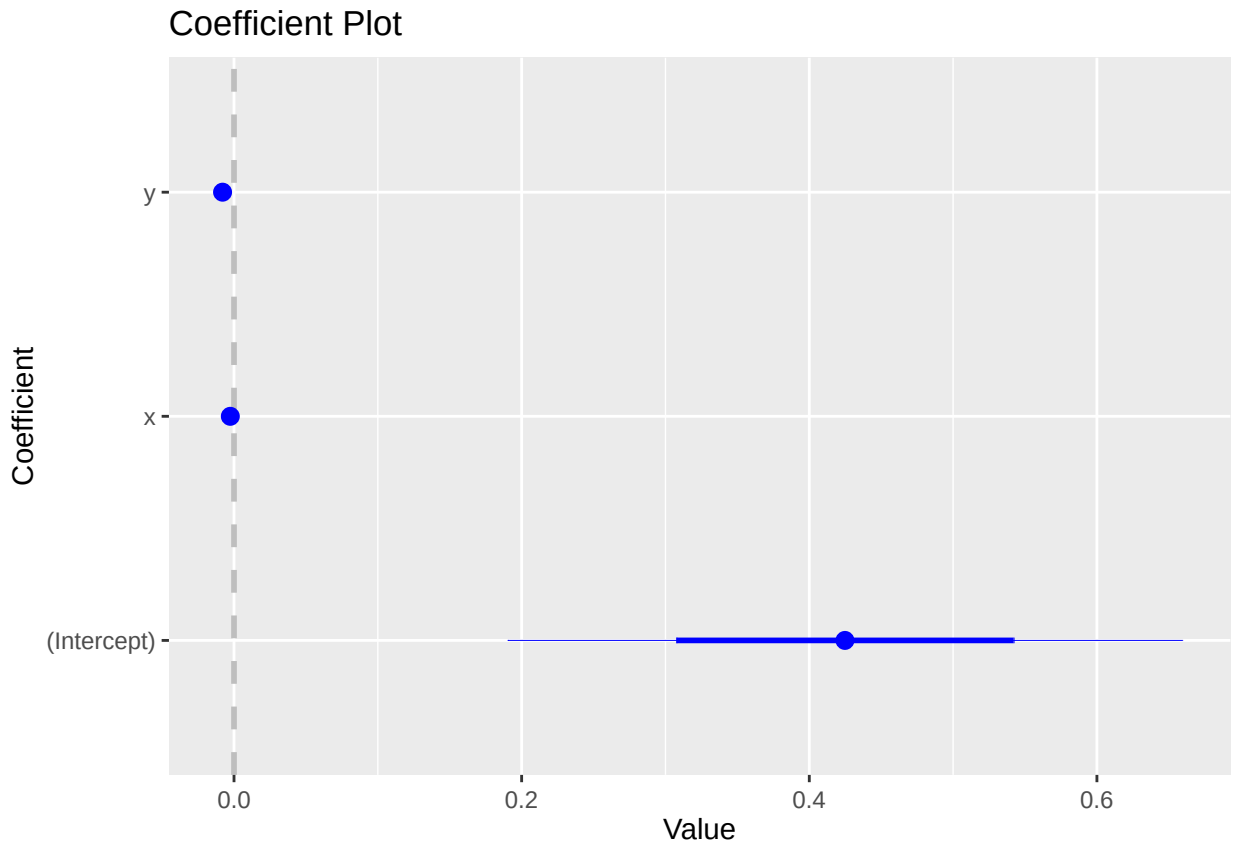
```
##        predict
## actual  FALSE
##   FALSE   400
##   TRUE     70
```

```
## [1] 0.8510638
```

The **accuracy** of the model is **85%**

Logical Regression model on binary data.

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




```
##
## Call:
## glm(formula = label ~ x + y, family = binomial(link = "logit"),
##     data = binary9)
##
## 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
```

What is the accuracy of the logistic regression classifier?

```
##       predict
## actual FALSE TRUE
##      0   429  338
##      1   286  445
```

```
## [1] 0.5834446
```

The **accuracy** of the model is **58%**

End of Markdown