# ica7_shuangyu_zhao

shuangyu_zhao

2023-01-31

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
heart <- read.csv("/Users/apple/Desktop/STT811 appl_stat_model/data/Heart.csv")
```

```
head(heart)
```

```
##   X Age Sex    ChestPain RestBP Chol Fbs RestECG MaxHR ExAng Oldpeak Slope Ca
## 1 1  63   1      typical    145  233   1       2   150     0     2.3     3  0
## 2 2  67   1 asymptomatic    160  286   0       2   108     1     1.5     2  3
## 3 3  67   1 asymptomatic    120  229   0       2   129     1     2.6     2  2
## 4 4  37   1   nonanginal    130  250   0       0   187     0     3.5     3  0
## 5 5  41   0   nontypical    130  204   0       2   172     0     1.4     1  0
## 6 6  56   1   nontypical    120  236   0       0   178     0     0.8     1  0
##         Thal AHD
## 1      fixed  No
## 2     normal Yes
## 3 reversable Yes
## 4     normal  No
## 5     normal  No
## 6     normal  No
```

1.

```
heart$target <- ifelse(heart$AHD=="Yes",1,0)
head(heart)
```

```
##   X Age Sex    ChestPain RestBP Chol Fbs RestECG MaxHR ExAng Oldpeak Slope Ca
## 1 1  63   1      typical    145  233   1       2   150     0     2.3     3  0
## 2 2  67   1 asymptomatic    160  286   0       2   108     1     1.5     2  3
## 3 3  67   1 asymptomatic    120  229   0       2   129     1     2.6     2  2
## 4 4  37   1   nonanginal    130  250   0       0   187     0     3.5     3  0
## 5 5  41   0   nontypical    130  204   0       2   172     0     1.4     1  0
## 6 6  56   1   nontypical    120  236   0       0   178     0     0.8     1  0
##         Thal AHD target
## 1      fixed  No      0
```

```
## 2     normal Yes      1
## 3 reversable Yes      1
## 4     normal  No      0
## 5     normal  No      0
## 6     normal  No      0
```

2.

```
mod <- glm(data = heart, target ~ MaxHR + RestBP + ChestPain, family = binomial)
summary(mod)
```

```
##
## Call:
## glm(formula = target ~ MaxHR + RestBP + ChestPain, family = binomial,
##     data = heart)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2063  -0.6994  -0.4206   0.7701   2.1752
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           2.915699   1.484988   1.963 0.049594 *
## MaxHR                -0.032991   0.007105  -4.643 3.43e-06 ***
## RestBP                0.021750   0.008534   2.549 0.010818 *
## ChestPainnonanginal -2.094570   0.344278  -6.084 1.17e-09 ***
## ChestPainnontypical -2.010243   0.437980  -4.590 4.44e-06 ***
## ChestPaintypical    -1.788426   0.540894  -3.306 0.000945 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 417.98  on 302  degrees of freedom
## Residual deviance: 300.91  on 297  degrees of freedom
## AIC: 312.91
##
## Number of Fisher Scoring iterations: 4
```

MaxHR, ChestPain are the most important coefficients.

3.

```
p <- 1/(1+exp(- c(1, 170, 145, 0, 1, 0) %*% mod$coefficients ))
p
```

```
##           [,1]
## [1,] 0.1751871
```

The probabibility of having heart disease is 17.52%.

```r
ratio <- p/(1-p)
ratio
```

```
##           [,1]
## [1,] 0.2123961
```

4.

```r
confusionMatrix(data = as.factor(as.integer(2*mod$fitted.values)), reference = as.factor(heart$target))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 127  33
##          1  37 106
##
##                Accuracy : 0.769
##                  95% CI : (0.7174, 0.8152)
##     No Information Rate : 0.5413
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.5358
##
##  Mcnemar's Test P-Value : 0.7199
##
##             Sensitivity : 0.7744
##             Specificity : 0.7626
##          Pos Pred Value : 0.7938
##          Neg Pred Value : 0.7413
##              Prevalence : 0.5413
##          Detection Rate : 0.4191
##    Detection Prevalence : 0.5281
##       Balanced Accuracy : 0.7685
##
##        'Positive' Class : 0
##
```