Assignment 4 - Logistic Regression

Son Nguyen

2024-02-13

## Question 4

### Part a

d = read.csv('https://bryantstats.github.io/math461/assignments/HospitalCosts.csv')  
  
# Calculate median charge  
median\_cost = median(d$TOTCHG)  
  
# Create a column to compare the cost with median charge  
# this column takes value 0 of the charge is less than the median and 1 otherwise  
  
d$charge = ifelse(d$TOTCHG < median\_cost, 0, 1)  
  
  
# Train  
model = glm(charge ~ AGE + LOS,   
 data = d,  
 family = binomial(link = "logit"))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

# Report model summary  
summary(model)

##   
## Call:  
## glm(formula = charge ~ AGE + LOS, family = binomial(link = "logit"),   
## data = d)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.32670 0.44248 -9.778 < 2e-16 \*\*\*  
## AGE 0.06929 0.01858 3.729 0.000192 \*\*\*  
## LOS 1.76979 0.18351 9.644 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 693.15 on 499 degrees of freedom  
## Residual deviance: 474.28 on 497 degrees of freedom  
## AIC: 480.28  
##   
## Number of Fisher Scoring iterations: 7

predicted\_prob = predict(model, list(AGE = 20, LOS = 1),   
 type = 'response')  
predicted\_prob

## 1   
## 0.2366538

predicted\_class = ifelse(predicted\_prob>=.5, 1, 0)  
predicted\_class

## 1   
## 0

d$predicted\_prob\_charge = predict(model, d, type = "response")  
  
d$predicted\_charge = ifelse(d$predicted\_prob\_charge <.5, 0, 1)  
  
mean(d$predicted\_charge == d$charge)

## [1] 0.814