

Exploração de dados - Banco Czech

Bruno Santos Wance de Souza

Lucas de Jesus Matias
Luiz Cesar Costa Raymundo

21 de novembro de 2018

Contents

Pagamento de Empréstimo

3

Pagamento de Empréstimo

```
read.csv2("./dados/pagamento_emprestimo.csv", stringsAsFactors = FALSE) -> pagamentoEmprestimo
glm(data = pagamentoEmprestimo, formula = pagamento ~ estadocivil + idade + sexo, family = binomial) ->
summary(glmPagamento)

##
## Call:
## glm(formula = pagamento ~ estadocivil + idade + sexo, family = binomial,
##      data = pagamentoEmprestimo)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4892  -0.4015   0.4166   0.5905   2.1662
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.96591    1.12267  -1.751  0.07993 .
## estadocivil -2.95095    0.58293  -5.062 4.14e-07 ***
## idade       0.11614    0.04432   2.621  0.00877 **
## sexo        1.30123    0.43861   2.967  0.00301 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 212.70  on 179  degrees of freedom
## Residual deviance: 146.65  on 176  degrees of freedom
## AIC: 154.65
##
## Number of Fisher Scoring iterations: 5

#Os valores Ps das variáveis rejeitam a hipótese inicial de que são irrelevantes para o modelo, portan
#Para testar o modelo, fazemos a predição
glmprobs <- predict(glmPagamento, type="response")
nLinhas <- nrow(pagamentoEmprestimo)
glmpred <- rep(0, nLinhas)
#Utilizaremos maior que 0,5 como sucesso no pagamento
glmpred[ glmprobs > 0.5 ] <- 1

#Comparando a predição com os dados que já conhecíamos, obtivemos 24 True Negatives, 125 True Positives
table(glmpred, pagamentoEmprestimo$pagamento)

##
## glmpred    0    1
##      0    24    5
##      1    26 125
```