exercicio4.R

gusta

2024-06-03

base <- read.csv("titanic\_data.csv", sep = ";")  
  
#a)  
set.seed(09092003)  
amostra = base[sample(nrow(base), 300),]  
  
#b)  
amostra$Survived <- as.factor(amostra$Survived)  
amostra$Pclass <- as.factor(amostra$Pclass)  
amostra$Sex <- as.factor(amostra$Sex)  
amostra$Embarked <- as.factor(amostra$Embarked)  
  
#c)  
modelo\_inicial <- glm(Survived ~ 1, family = binomial, data = amostra)  
modelo\_final <- step(modelo\_inicial, direction = "forward", scope = ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked)

## Start: AIC=394.05  
## Survived ~ 1  
##   
## Df Deviance AIC  
## + Sex 1 376.47 380.47  
## + Parch 1 382.82 386.82  
## + Pclass 2 386.20 392.20  
## + Embarked 2 387.08 393.08  
## <none> 392.05 394.05  
## + Age 1 390.21 394.21  
## + SibSp 1 391.44 395.44  
## + Fare 144 179.61 469.61  
##   
## Step: AIC=380.47  
## Survived ~ Sex  
##   
## Df Deviance AIC  
## + Parch 1 369.50 375.50  
## + Pclass 2 371.18 379.18  
## + Embarked 2 372.32 380.32  
## + Age 1 374.39 380.39  
## <none> 376.47 380.47  
## + SibSp 1 375.79 381.79  
## + Fare 144 175.78 467.78  
##   
## Step: AIC=375.5  
## Survived ~ Sex + Parch  
##   
## Df Deviance AIC  
## + Embarked 2 363.04 373.04  
## + Pclass 2 364.29 374.29  
## <none> 369.50 375.50  
## + Age 1 367.88 375.88  
## + SibSp 1 369.49 377.49  
## + Fare 144 175.53 469.53  
##   
## Step: AIC=373.04  
## Survived ~ Sex + Parch + Embarked

## Warning: glm.fit: probabilidades ajustadas numericamente 0 ou 1 ocorreu

## Df Deviance AIC  
## + Pclass 2 354.10 368.10  
## + Age 1 360.61 372.61  
## <none> 363.04 373.04  
## + SibSp 1 363.03 375.03  
## + Fare 144 173.78 471.78  
##   
## Step: AIC=368.1  
## Survived ~ Sex + Parch + Embarked + Pclass

## Warning: glm.fit: algoritmo não convergiu  
## Warning: glm.fit: probabilidades ajustadas numericamente 0 ou 1 ocorreu

## Df Deviance AIC  
## + Age 1 346.8 362.8  
## <none> 354.1 368.1  
## + SibSp 1 354.1 370.1  
## + Fare 143 3460.2 3760.2  
##   
## Step: AIC=362.82  
## Survived ~ Sex + Parch + Embarked + Pclass + Age

## Warning: glm.fit: algoritmo não convergiu  
## Warning: glm.fit: probabilidades ajustadas numericamente 0 ou 1 ocorreu

## Df Deviance AIC  
## <none> 346.8 362.8  
## + SibSp 1 346.8 364.8  
## + Fare 143 3748.5 4050.5

summary(modelo\_final)

##   
## Call:  
## glm(formula = Survived ~ Sex + Parch + Embarked + Pclass + Age,   
## family = binomial, data = amostra)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.250569 0.531160 2.354 0.018552 \*   
## Sexmale -0.899922 0.263053 -3.421 0.000624 \*\*\*  
## Parch 0.425564 0.149923 2.839 0.004532 \*\*   
## EmbarkedQ 1.304644 0.463109 2.817 0.004845 \*\*   
## EmbarkedS -0.136713 0.322958 -0.423 0.672066   
## Pclass2 -0.717134 0.404395 -1.773 0.076171 .   
## Pclass3 -1.284891 0.356806 -3.601 0.000317 \*\*\*  
## Age -0.025824 0.009859 -2.619 0.008809 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 392.05 on 299 degrees of freedom  
## Residual deviance: 346.82 on 292 degrees of freedom  
## AIC: 362.82  
##   
## Number of Fisher Scoring iterations: 4

#d)  
OR <- exp(coef(modelo\_final))  
IC <- exp(confint(modelo\_final))

## Waiting for profiling to be done...

resultado <- data.frame(OR, IC)  
print(resultado)

## OR X2.5.. X97.5..  
## (Intercept) 3.4923290 1.2547649 10.1353705  
## Sexmale 0.4066014 0.2413050 0.6781061  
## Parch 1.5304539 1.1569622 2.0909694  
## EmbarkedQ 3.6863780 1.5007888 9.2797158  
## EmbarkedS 0.8722208 0.4648942 1.6556080  
## Pclass2 0.4881494 0.2181207 1.0701097  
## Pclass3 0.2766807 0.1354713 0.5510195  
## Age 0.9745063 0.9553203 0.9930831