

Untitled

Guilherme Pazian

4 de junho de 2018

O código abaixo simula a variável resposta “ysim” com base nas variáveis explicativas “x1” e “x2”.

- As variáveis x1 e x2 são geradas a partir da distribuição normal com média 1 e variância 1 (caso queira, pode mudar estes parâmetros).
- Define-se os parâmetros beta0, beta1 e beta2. O vetor “preditorlinear” é gerado pela combinação linear das v.a. x1 e x2: $\text{beta0} + \text{beta1} * x1 + \text{beta2} * x2$
- O vetor “prob” é gerado aplicando a função inversa da função logito (que é a função de ligação usada na regressão logística) no vetor “preditorlinear”.
- São geradas variáveis aleatórias baseadas na distribuição uniforme e estas são comparadas com as probabilidades geradas no vetor “prob”, definindo assim o vetor com observações binárias simuladas.

```
set.seed(123)
source("diag_Bern.R")
source("envel_Bern.R")
# fonte: https://www.r-bloggers.com/example-7-2-simulate-data-from-a-logistic-regression/
beta0 <- 0
beta1 <- 0.5
x1 <- rnorm(1000,1,1)
beta2 <- 0.9
x2 <- rnorm(1000,1,1)
preditorlinear <- beta0 + beta1*x1 + beta2*x2
prob <- exp(preditorlinear)/(1 + exp(preditorlinear))
vaunif <- runif(1000,0,1)
ysim <- ifelse(vaunif < prob,1,0)
```

Abaixo é ajustado um modelo de regressão logística com os dados simulados anteriormente.

```
fit2 = glm(ysim ~ x1 + x2,family = "binomial")
```

Abaixo as estimativas dos parâmetros do modelo. É notável que as estimativas dos parâmetros são muito próximas dos valores dos parâmetros definidos anteriormente (beta0, beta1 e beta2).

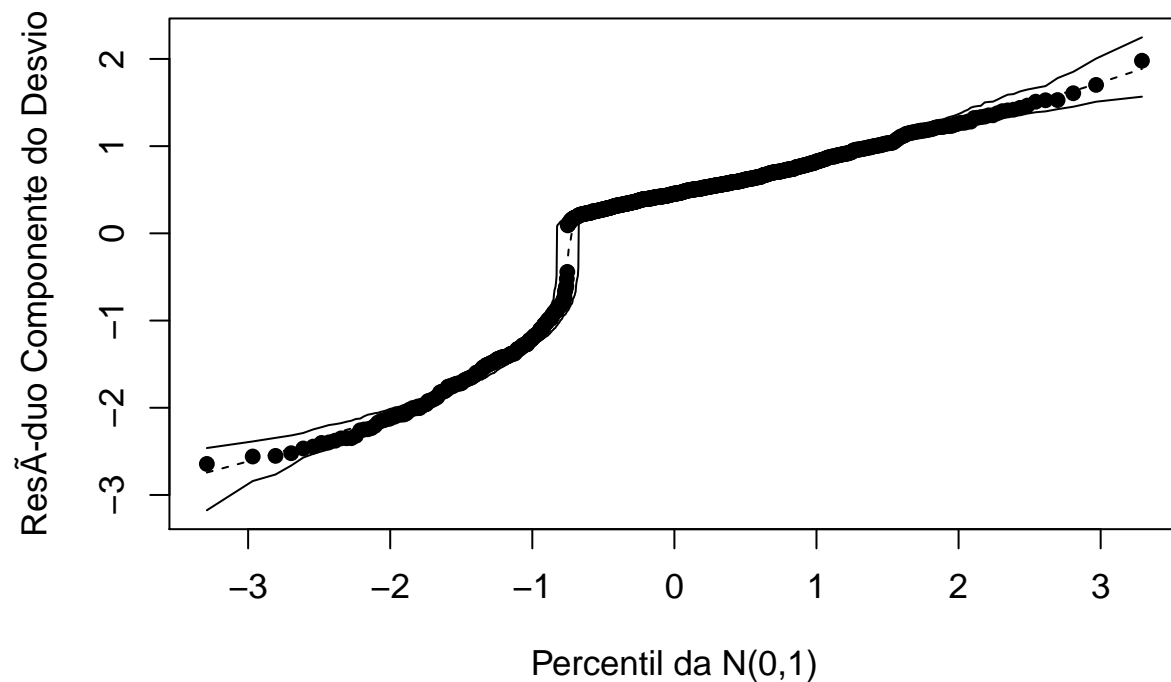
```
summary(fit2)

##
## Call:
## glm(formula = ysim ~ x1 + x2, family = "binomial")
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6418   0.2044   0.4545   0.6894   1.9715
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.07435    0.12648  -0.588   0.557
## x1           0.59157    0.09079   6.516 7.24e-11 ***
## x2           0.99715    0.09596  10.391 < 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: 1068.80  on 999  degrees of freedom
## Residual deviance:  881.74  on 997  degrees of freedom
## AIC: 887.74
##
## Number of Fisher Scoring iterations: 5
```

Análise de resíduos. Note que os resíduos se comportam da maneira prevista, evidenciando um bom ajuste do modelo:

```
source("envel_Bern.R")
envelBern(fit2,"logit")
```



Outro indicador para verificar o ajuste do modelo é a análise do Resíduo componente do desvio. Observe que o p-valor associado ao teste não é menor que os 0.05, o que indica um bom ajuste do modelo:

```
desvio <- deviance(fit2)
p <- ncol(model.matrix(fit2))
n <- length(ysim)
pdesvio <- 1-pchisq(desvio,df=n-p)

print(pdesvio)
```

```
## [1] 0.9962393
```