

Resolução Lista II  
Modelos Lineares Generalizados

Nathalia Gabriella Ferreira dos Santos

2023-10-28

# Índice

<b>1)</b>	<b>2</b>
b) . . . . .	2
c) . . . . .	3
d) . . . . .	4
<b>2)</b>	<b>5</b>
a) . . . . .	5
b) . . . . .	6
c) . . . . .	6
d) . . . . .	6
e) . . . . .	7
f) . . . . .	8
<b>3)</b>	<b>9</b>
a) . . . . .	9
b) . . . . .	10
c) . . . . .	10
d) . . . . .	12

1)

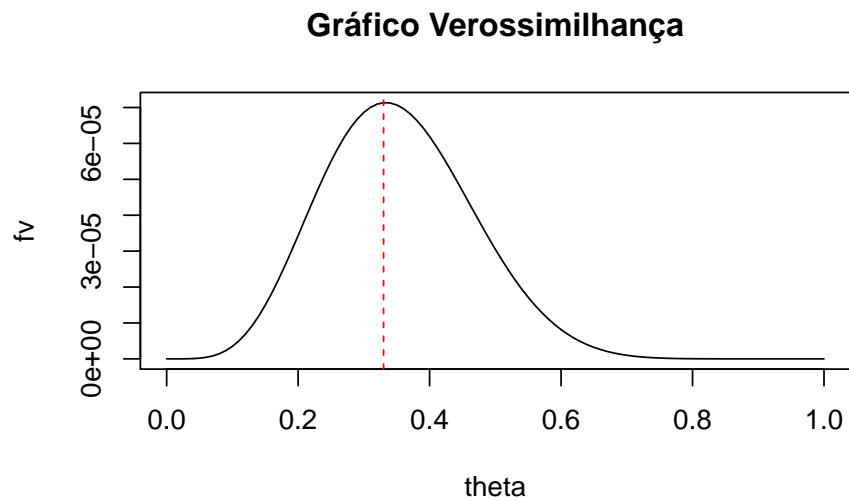
```
library(tidyverse)
rm(list = ls())
```

b)

```
y <- c(1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1)

theta <- seq(0,1,0.01)
n <- length(y)
sy <- sum(y) # soma de y

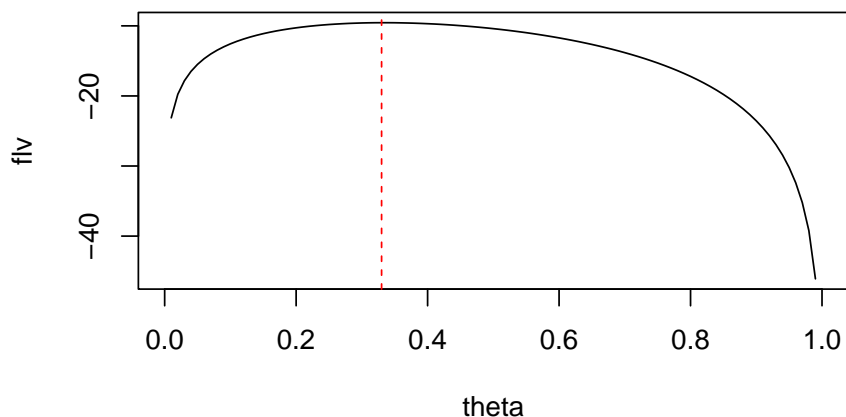
# Verossimilhança
fv <- theta^sy * (1-theta)^(n - sy)
maximo <- theta[fv == max(fv)]
plot(theta, fv, type = "line", main = "Gráfico Verossimilhança")
abline(v = maximo, col = 'red', lty = 2)
```



c)

```
# Log-Verossimilhança
flv <- sy*log(theta) + (n - sy)*log(1-theta)
maximo <- theta[flv == max(flv)]
plot(theta, flv, type = "line", main = "Gráfico Log-Verossimilhança")
abline(v = maximo, col = 'red', lty = 2)
```

### Gráfico Log-Verossimilhança



```
paste0("Valor do máximo: ", maximo)
```

```
[1] "Valor do máximo: 0.33"
```

d)

```
loglik <- function(theta, y){  
  n <- length(y)  
  sy <- sum(y)  
  log_ver <- sy*log(theta) + (n - sy)*log(1-theta)  
  return(log_ver)  
}  
  
theta0 <- 0.5  
ajuste <- optim(par = theta0, # chute inicial  
               fn = loglik, # funcao a ser otimizada  
               y = y, # parametro da funcao  
               lower = 0.0001, upper = 0.9999, # limites para o parametro  
               method = "L-BFGS-B", control = list(fnscale = -1), hessian = F) # metodo  
ajuste$par %>% round(3)
```

```
[1] 0.333
```

2)

$$Y_i \sim \text{Poisson}(\theta_i), \theta_i > 0$$

```
rm(list = ls())

y <- c(7, 2, 24, 33, 2, 5, 2, 5, 8, 8, 16, 20, 11, 4, 37, 1, 8, 8, 5, 7, 5, 4, 4, 2, 26)
x1 <- c(-0.27, 0.57, -0.74, -0.94, 0.64, 0.86, 0.5, 0.12, -0.36, 0.12, -0.55, -0.81, -0.27, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12, 0.12)

n = length(y)
```

a)

💡 Dica:

Olhar página 63 do slide 3

$$\eta_i = \beta_0 + \beta_1 X_{1i}$$

$$\theta_i = e^{\eta_i}$$

```
bet <- c(1,1) #chute inicial
w <- numeric(0)
z <- numeric(0)
for (i in 1:n) {
  eta <- bet[1] + bet[2]*x1[i]
  w[i] <- exp(eta)
  z[i] <- eta + y[i]*exp(-eta)-1
}

w = diag(w) # w eh uma matriz diagonal
w[1:5,1:5]
```

[,1] [,2] [,3] [,4] [,5]

```
[1,] 2.075081 0.000000 0.000000 0.000000 0.000000
[2,] 0.000000 4.806648 0.000000 0.000000 0.000000
[3,] 0.000000 0.000000 1.29693 0.000000 0.000000
[4,] 0.000000 0.000000 0.000000 1.061837 0.000000
[5,] 0.000000 0.000000 0.000000 0.000000 5.15517
```

b)

```
z
```

```
[1] 3.1033629 0.9860904 17.7652381 30.1382296 1.0279601 1.6383632
[7] 0.9462603 1.7513990 3.8583394 2.7302384 9.6520504 15.7291827
[13] 4.7822927 1.3961427 36.0000000 0.6465970 3.0864297 2.6830942
[19] 1.9647577 2.2866112 1.6642921 1.3882812 1.5025113 1.2489137
[25] 18.7448528
```

c)

```
X <- cbind(rep(1,n) , x1)
beta_r_1 <- solve( t(X) %*% w %*% X) %*% t(X) %*% w %*% z
beta_r_1
```

```
      [,1]
      6.153008
x1 -8.109299
```

d)

```
mod <- glm(formula = y ~ x1, family = poisson(link = "log"), start = bet)
resumo <- mod %>% summary()
resumo
```

Call:

```
glm(formula = y ~ x1, family = poisson(link = "log"), start = bet)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7780	-0.8236	0.1708	0.7132	2.0495

Coefficients:

Estimate	Std. Error	z value	Pr(> z )
----------	------------	---------	----------

```

(Intercept)    1.9798      0.0841    23.54    <2e-16 ***
x1             -1.5115      0.1240   -12.19    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 199.783  on 24  degrees of freedom
Residual deviance:  25.098  on 23  degrees of freedom
AIC: 123.27

```

Number of Fisher Scoring iterations: 15

Nota-se que:

- *Intercepto* tem coeficiente estimado em 1.98 e é significativo pois possui p-valor de  $1.535253 \times 10^{-122} < \alpha = 0,05$ .
- *x1* tem coeficiente estimado em -1.51 e é significativo pois possui p-valor de  $3.5958165 \times 10^{-34} < \alpha = 0,05$ .

e)

 Dica:

Página 17 do slide 4

```

beta0_chapeu = coef[1]
beta1_chapeu = coef[2]
mu = exp(beta0_chapeu + beta1_chapeu*x1)
de = 2*sum(y*log(y/mu)-(y-mu))
de # mesmo da saída do summary

```

```
[1] 25.09786
```

```

p <- 2 # numero de betas estimados
cbind(
  list(
    "Ponto crítico a 95%= " = qchisq(0.95, n-p) %>% round(2),
    "Ponto crítico a 99%= " = qchisq(0.99, n-p) %>% round(2)))

```

```

      [,1]
Ponto crítico a 95%= 35.17
Ponto crítico a 99%= 41.64

```



Como a Deviance Residual do modelo ( $D_{\text{modelo}} = 25.1$ ) foi menor que ambos limiares de comparação de deviance achados acima, o ajuste pode ser considerado ok. Ou seja, nem é saturado nem não significativo.

f)

```
x2p.1 = sum( ((y-mu)^2)/mu )
rp = residuals(mod, type = "pearson") # residuo de pearson
x2p.2 = sum(rp^2)
cbind(x2p.1,x2p.2)

      x2p.1    x2p.2
[1,] 26.66753 26.68479

      1-pchisq(x2p.1, n-p)

[1] 0.2703772
```

Nesse caso temos um bom ajuste pois o  $p\text{valor} > 0,05$  indica que o valor dos resíduos de pearson está abaixo do limiar. O limiar é o mesmo obtido anteriormente.

### 3)

Trata-se de um **Modelo Binomial** pois a variável resposta é binária (morrer ou não). E como há quantidade limitada na caixa, não é poisson pois há um limite máximo de mortes para cada caixa, o que viola o pressuposto da poisson ( $\theta_i > 0$ ).

```
rm(list = ls())
dados <- read.table("dados_Q3_L2_MLG.txt", col.names = c("y", "x1", "x2")) %>% as_tibble()
```

#### a)

```
m <- 10 #numero de larvas por recipiente
m1 <- glm(cbind(y,m-y)~x1+x2, family=binomial(link="logit"), data = dados)
resumo <- summary(m1)
resumo
```

Call:

```
glm(formula = cbind(y, m - y) ~ x1 + x2, family = binomial(link = "logit"),
    data = dados)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.14710	-0.79749	0.02523	0.71168	1.71217

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.0911	0.2191	-4.980	6.37e-07 ***
x1	1.5420	0.3487	4.422	9.77e-06 ***
x2	0.1722	0.2990	0.576	0.565

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 72.906 on 51 degrees of freedom  
Residual deviance: 50.796 on 49 degrees of freedom  
AIC: 193.97

Number of Fisher Scoring iterations: 4

Nota-se que:

- *Intercepto* tem coeficiente estimado em -1.09 e é significativo pois possui p-valor de  $0 < \alpha = 0,05$ .
- *x1* tem coeficiente estimado em 1.54 e é significativo pois possui p-valor de  $0 < \alpha = 0,05$ .
- *x2* tem coeficiente estimado em 0.17 e não é significativo pois possui p-valor de  $0.5646 > \alpha = 0,05$ .

b)

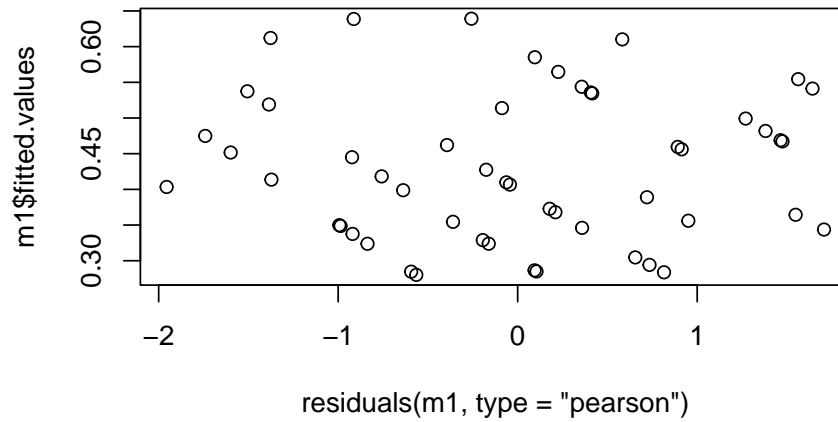
- Geral:  $ods = 100 \cdot (\exp^{\beta_1 X_{1i}} - 1)$

Quando *x1* é aumentado em 0.1 unidade, a razão de chance é variada em  $ods = 100 \cdot (\exp^{1,54 \cdot 0,1} - 1)$ . Logo  $ods = 16.65\%$ .

c)

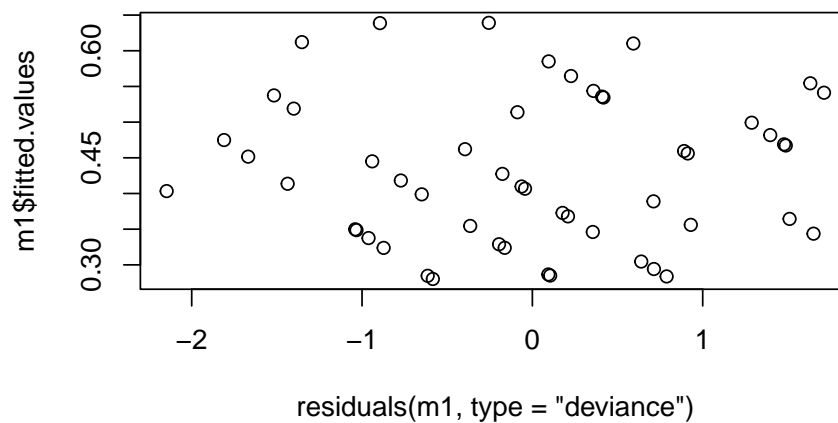
```
plot(  
  residuals(m1, type = "pearson"),  
  m1$fitted.values,  
  main = "Resíduo de Pearson vs Valores Ajustados")
```

### Resíduo de Pearson vs Valores Ajustados



```
plot(  
  residuals(m1, type = "deviance"),  
  m1$fitted.values,  
  main = "'Resíduos componente do desvio vs Valores Ajustados'")
```

### “Resíduos componente do desvio vs Valores Ajustados



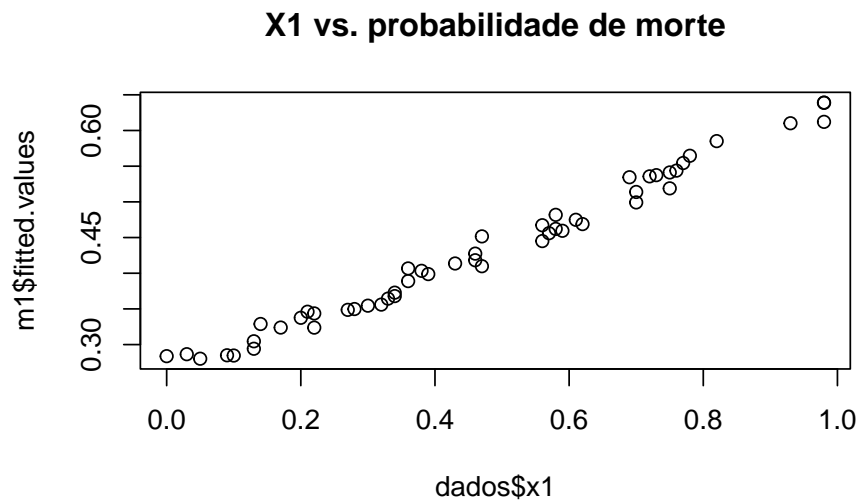
d)

```
data.frame(Recipiente = dados$y,  
           ProbMorte = m1$fitted.values %>% round(2))
```

	Recipiente	ProbMorte
1	5	0.39
2	6	0.56
3	4	0.29
4	3	0.29
5	6	0.59
6	7	0.61
7	3	0.52
8	7	0.47
9	5	0.36
10	3	0.45
11	2	0.32
12	2	0.28
13	3	0.40
14	8	0.55
15	4	0.28
16	7	0.50
17	3	0.42
18	4	0.46
19	2	0.41
20	2	0.47
21	6	0.64
22	6	0.54
23	2	0.45
24	5	0.64
25	3	0.33
26	3	0.32
27	3	0.29
28	4	0.37
29	4	0.35
30	6	0.53
31	4	0.41
32	5	0.51
33	6	0.46
34	4	0.41
35	4	0.61
36	3	0.35
37	2	0.35
38	4	0.30
39	6	0.34

40	2	0.35
41	8	0.54
42	4	0.43
43	1	0.40
44	2	0.34
45	7	0.48
46	4	0.37
47	7	0.47
48	6	0.46
49	3	0.54
50	6	0.54
51	6	0.36
52	2	0.28

```
plot(dados$x1, m1$fitted.values, main = "X1 vs. probabilidade de morte")
```



```
plot(dados$x2, m1$fitted.values, main = "X2 vs. probabilidade de morte")
```

### X2 vs. probabilidade de morte

