

Disciplina: Modelos Lineares Generalizados

Curso: Graduação em Estatística

Código: EST0011

Semestre: 2023.2

Prof. Frederico Machado Almeida

Lista de Exercícios #4

Observações:

- Questões para entregar: 1, 2(b,c), 3 e 4.
- Demais questões são apenas para estudar.
- Prazo de entrega: **17/11/2023**

Q01. Suponha que $Y_i \sim \exp(\phi)$, para $i = 1, \dots, 5$. Isto é,

$$f(y_i|\phi) = \frac{1}{\phi} \exp\left(-\frac{y_i}{\phi}\right), \text{ com } y_i > 0 \text{ e } \phi > 0.$$

Suponha que o vetor observado foi $\mathbf{y} = (2.8, 3.5, 2.4, 1.9, 3.0)^\top$ com média $\bar{y} = 2.72$.

- Implemente manualmente o algoritmo de Newton-Raphson e itere até obter a convergência. Feito isso, apresente uma tabela contendo as iterações, as estimativas (para cada iteração), e a tolerância.
- Como o valor obtido se compara com o estimador de máxima verossimilhança? Isto é, $\hat{\phi} = \bar{y}$?

Q02. Assuma que y_1, y_2, \dots, y_n é uma amostra observada de Y . Para cada distribuição abaixo, apresente a função desvio escalonado total.

- Distribuição Weibull: $f(y_i|\alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{y_i}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{y_i}{\beta}\right)^\alpha\right]$, com $y_i > 0$; $\alpha, \beta > 0$.
- Distribuição Binomial: $Y_i \sim \text{Binom}(n_i, p_i)$ $P(Y_i = y_i|\alpha, \beta) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$, com $y_i = 0, 1, 2, \dots, n_i$ e $0 < p_i < 1$.
- Distribuição Beta: $f(y_i|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1} (1 - y_i)^{\beta-1}$, com $0 < y_i < 1$, e $\Gamma(\cdot)$ denota a função gama.

Q03. Considere a função gama apresentada na questão anterior. Conforme mostrado na **Lista 3**, a distribuição gama é membro da família exponencial com ligação canônica. Usando a função de ligação logarítmica (frequentemente usada como função de ligação canônica para a distribuição Poisson), obtenha:

- A função score $U(\beta_j)$ e as entradas genéricas (j, s) da matriz de informação de Fisher, digamos $\mathcal{I}_{js}(\beta)$.

- (b) A matriz dos pesos $W(\beta)$ e a matriz $G(\beta) = \text{diagonal} \{g'(\mu_i), \text{ com } i = 1, \dots, n\}$.
- (c) A variável dependente ajustada na t -ésima iteração do Fisher Scoring, i.e., $z_i^{(t)}$.

Q04. Os dados apresentados na Tabela são referentes ao tempo até a morte y_i (em semanas) de dezassete pacientes que sofrem de leucemia, e x_i denota a contagem inicial dos glóbulos brancos, em uma escala de \log_{10} .

i	y_i	x_i	i	y_i	x_i
1	65	3,36	10	143	3,85
2	156	2,88	11	56	3,97
3	100	3,63	12	26	4,51
4	134	3,41	13	22	4,54
5	16	3,78	14	01	5,00
6	108	4,02	15	01	5,00
7	121	4,00	16	05	4,72
8	04	4,23	17	65	5,00
9	39	3,73			

- (a) Obtenha o gráfico de dispersão de y_i contra x_i . O diagrama apresenta alguma tendência?
- (b) Uma possível especificação para $\mu_i = \mathbb{E}(Y_i)$ é,

$$\mathbb{E}(Y_i) = \exp(\beta_0 + \beta_1 x_i). \quad (1)$$

Como pode ser visto, esta especificação permite que $\mathbb{E}(Y_i)$ seja sempre positivo para todo β_j , com $j = 0, 1$ e todo x_i . Qual função de ligação seria apropriada para este caso? Justifique a sua resposta.

- (c) A distribuição Exponencial tem sido frequentemente usada para modelar tempos de sobrevivência. Sua fdp é dada por: $f(y_i; \mu) = \frac{1}{\mu} \exp\left(-\frac{y_i}{\mu}\right)$. Observe que, esta distribuição é um caso particular da distribuição Gama (veja Q01-(e), Lista 3), com $\nu = 1$. Mostre que $\mathbb{E}(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \mu_i^2$.
- (d) Usando algum software estatístico, ajuste o modelo de regressão, assumindo que os dados seguem uma distribuição Exponencial, com $\mathbb{E}(Y_i)$ dada segundo a equação (1).
- (e) Para o modelo obtido em (d): (i) compare os valores observados: y_i e os estimados $\hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)$; (ii) use os resíduos padronizados $r_i = \frac{y_i - \hat{y}_i}{\hat{y}_i}$ para avaliar a qualidade de ajuste do modelo.
- (f) Interprete os valores dos coeficientes de regressão, $\hat{\beta}_0$ e $\hat{\beta}_1$.

Q05. Sabe-se que a concentração de colesterol no soro sanguíneo aumenta com a idade, mas é menos claro se o nível de colesterol também está associado ao peso corporal. As variáveis de interesse são: o nível de colesterol sérico de trinta e dois pacientes (milhões de moles por litro), idade (anos) e o índice de massa corporal (peso dividido pela altura ao quadrado, onde o peso foi medido em quilogramas e a altura em metros). Com base no banco de dados descrito, foi ajustado o modelo de regressão Gama, com função de ligação inversa (por default). Toda informação sobre o ajuste do modelo está apresentada no Apêndice.

APÊNDICE

```
##-----##
## I. Caracterizando o Banco de Dados ##
##-----##
> str(Q05.LISTA04)
'data.frame': 32 obs. of 4 variables:
 $ chol: num 5.94 6.48 4.71 8.83 5.86 5.1 5.86 5.1 6.8 4.65 ...
 $ Age : int 52 65 46 76 51 47 51 47 70 30 ...
 $ bmi : num 20.7 26.3 21.3 22.7 25.4 21.5 25.4 21.5 23.9 18.9 ...
 $ sex : int 0 0 0 1 1 1 0 0 0 0 ...
```

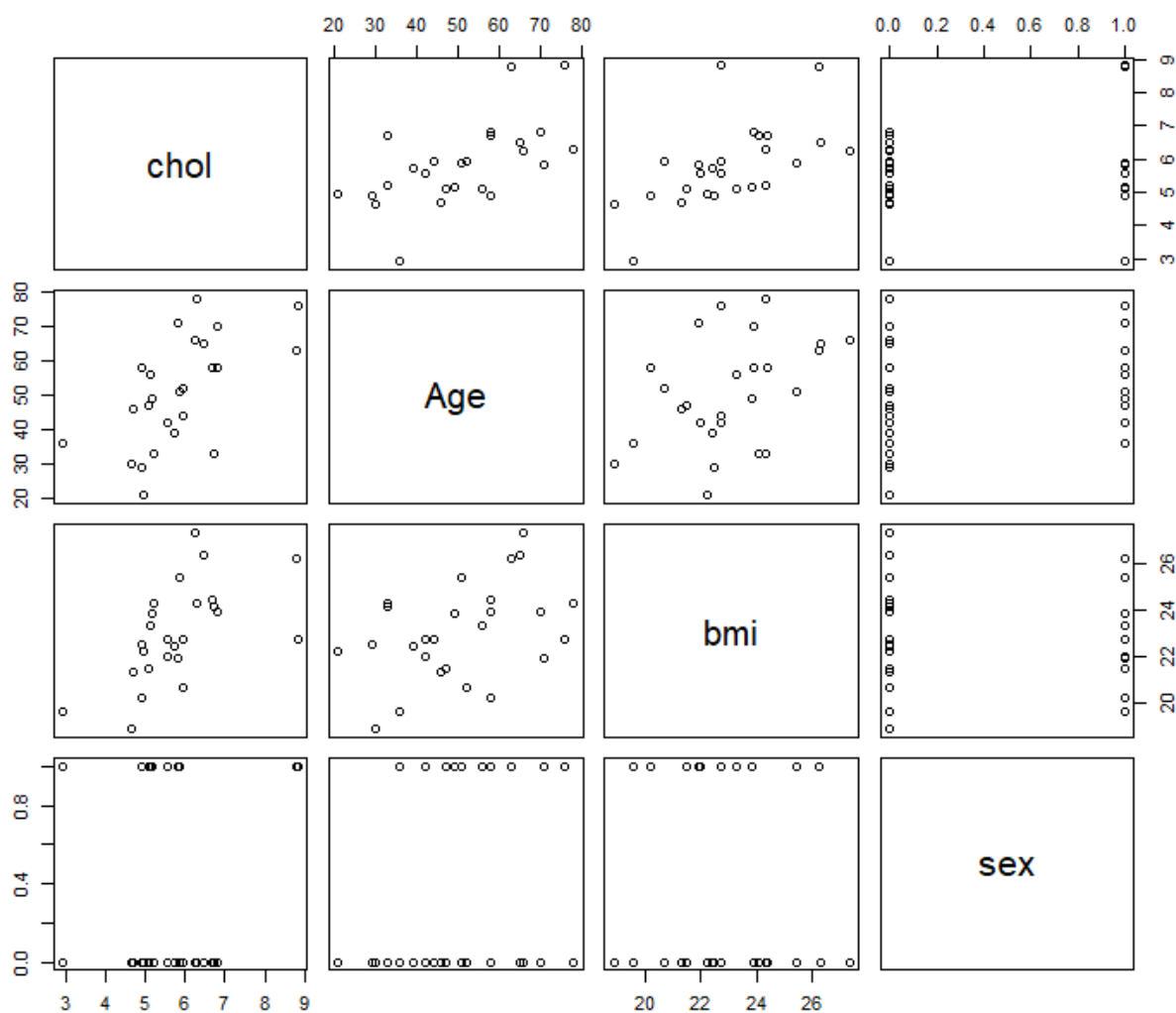


Figure 1: Scatterplot para as variáveis do banco de dados.

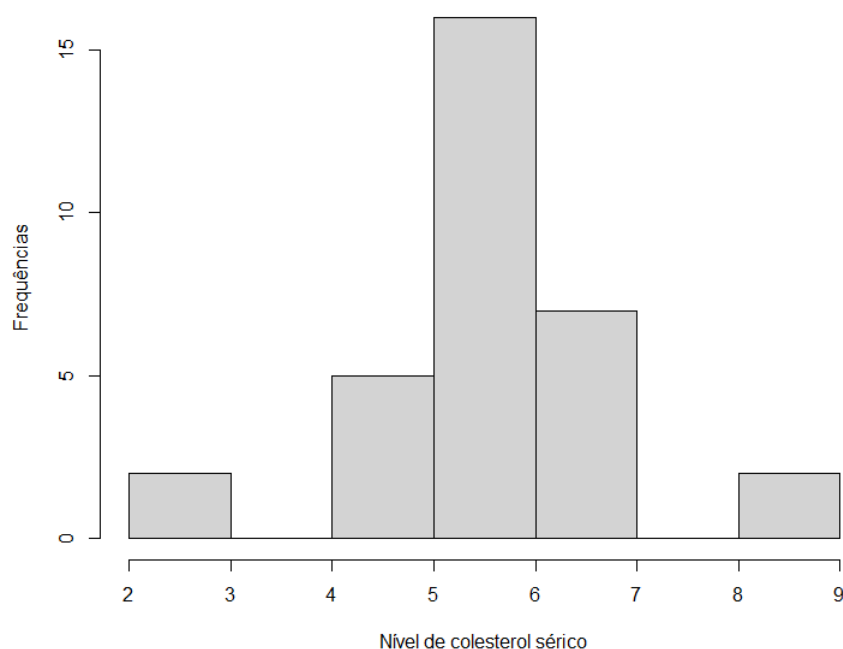


Figure 2: Histograma da variável resposta

```
##-----##
## II. Ajustando o modelo de regressão Gama ##
##-----##
> fit.Q05 <- glm(chol~Age+bmi+sex+Age:sex, data = Q05.LISTA04,
family=Gamma(link = "inverse"))
> summary(fit.Q05)
Call:
glm(formula = chol ~ Age + bmi + sex + Age:sex, family = Gamma(link =
"inverse"), data = Q04.LISTA04)

Deviance Residuals:
Min       1Q   Median       3Q      Max
-0.42723  -0.08677  -0.01425   0.09297   0.20490

Coefficients:
Estimate Std. Error t-value Pr(>|t|)
(Intercept)  0.4151666  0.0541811   7.663  3.05e-08 ***
Age          -0.0003999  0.0004282  -0.934  0.35861
bmi          -0.0094165  0.0025765  -3.655  0.00109 **
sex           0.1172457  0.0423257   2.770  0.01001 *
Age:sex      -0.0021329  0.0007419  -2.875  0.00779 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.02134601)

Null deviance: 1.52658 on 31 degrees of freedom
Residual deviance: 0.62225 on 27 degrees of freedom
AIC: 86.673
```

Number of Fisher Scoring iterations: 4

```
##-----##
## III. Qualidade de Ajuste do Modelo ##
##-----##
> ## (1) Naglekerke (1991) R-square ##

> Num<- 1-exp((fit.Q05$dev-fit.Q05$null)/nrow(Q05.LISTA05))
> Den<- 1-exp(-fit.Q05$null/nrow(Q05.LISTA04))
> R2_Nag<- Num/Den; R2_Nag
[1] 0.5981385

> ## (2) An alternative R^2 ##

> Null.dev<- fit.Q05$null.deviance
> Resi.dev<- deviance(fit.Q05)
> ##
> R2.dev<- (Null.dev-Resi.dev)/Null.dev; R2.dev
[1] 0.5923876

> ##(3) Goodness-of-fit analysis
> pchisq(deviance(fit.Q05), df.residual(fit.Q05), lower=FALSE)
[1] 1

##-----##
## IV ANÁLISE DE RESÍDUOS E MEDIDAS DE INFLUÊNCIA ##
##-----##
##
##-----##
## COVRATIO ##
##-----##
1      2      3      4      5      6      7      8      9      10
1.024 1.378 1.237 2.056 1.475 1.306 1.248 1.298 1.332 1.378
11     12     13     14     15     16     17     18     19     20
1.311 1.114 1.484 1.729 1.614 1.346 1.144 0.663 1.144 0.182
21     22     23     24     25     26     27     28     29     30
1.189 0.943 1.189 0.943 1.280 1.337 1.223 1.083 1.375 1.273
31     32
1.228 1.119

##-----##
## DFFITS ##
##-----##
> dffits(fit.Q05)
1      2      3      4      5      6      7      8      9
0.436 -0.293 -0.148 1.100 -0.138 0.189 -0.210 -0.020 0.331
10     11     12     13     14     15     16     17     18
0.132 -0.303 0.249 -0.060 -0.134 0.611 -0.137 -0.269 -1.156
19     20     21     22     23     24     25     26     27
-0.269 -1.298 0.145 0.646 0.145 0.646 -0.764 -0.151 0.145
28     29     30     31     32
0.501 -0.109 0.046 0.158 -0.965
```

V Análise de Resíduos

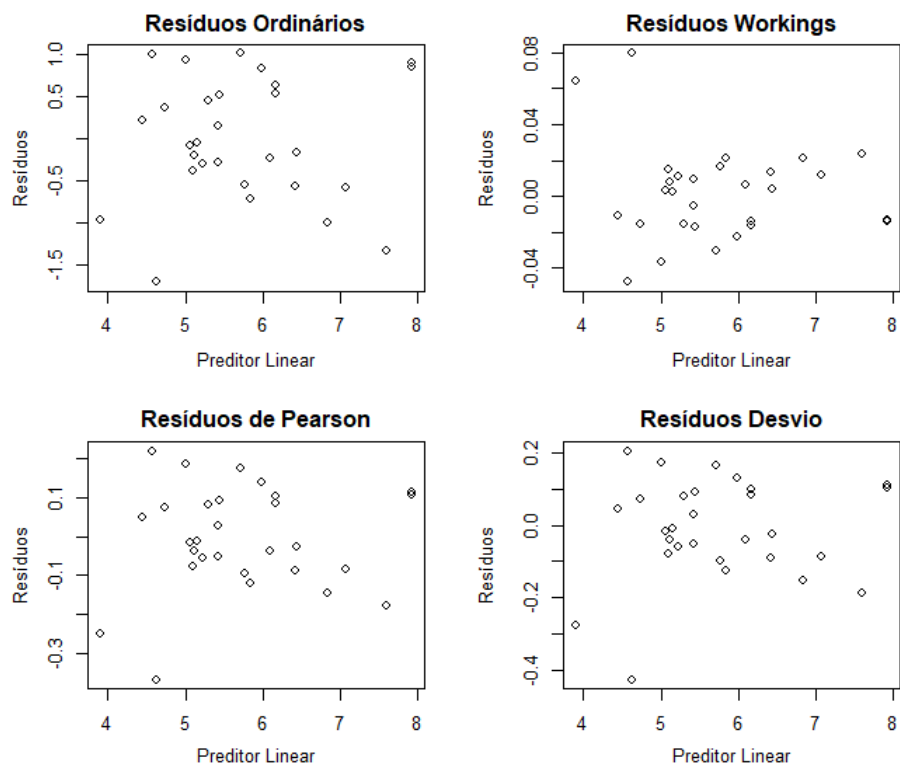


Figure 3: Comparação do grau de dispersão de diferentes tipos de resíduos em função $\hat{\eta}_i$

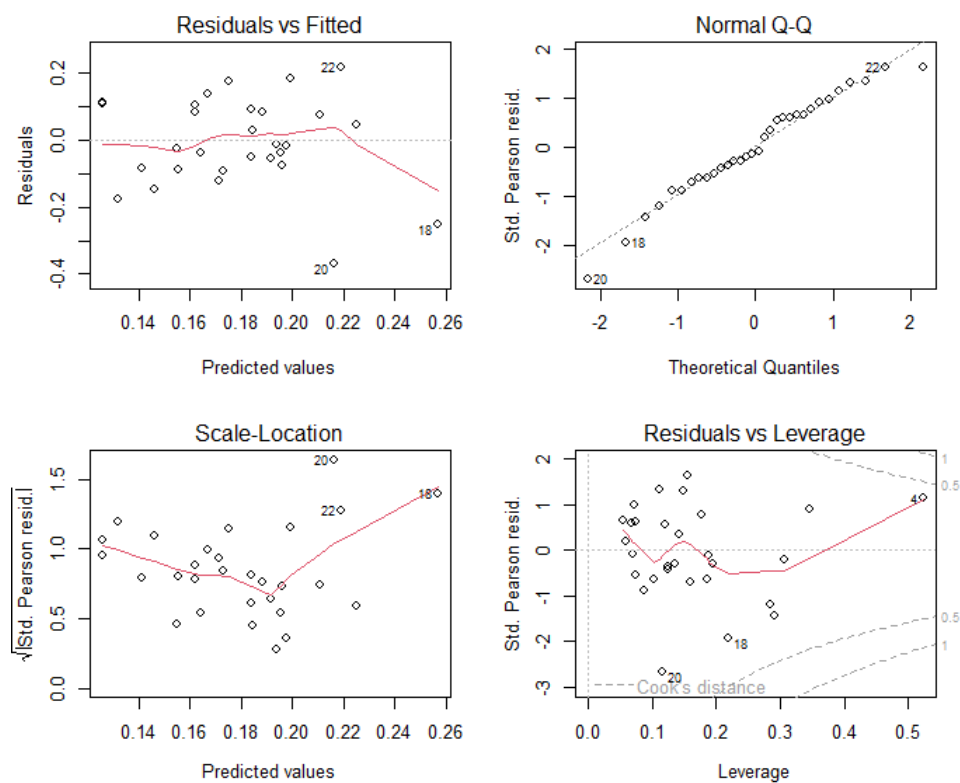


Figure 4: Análise de resíduos

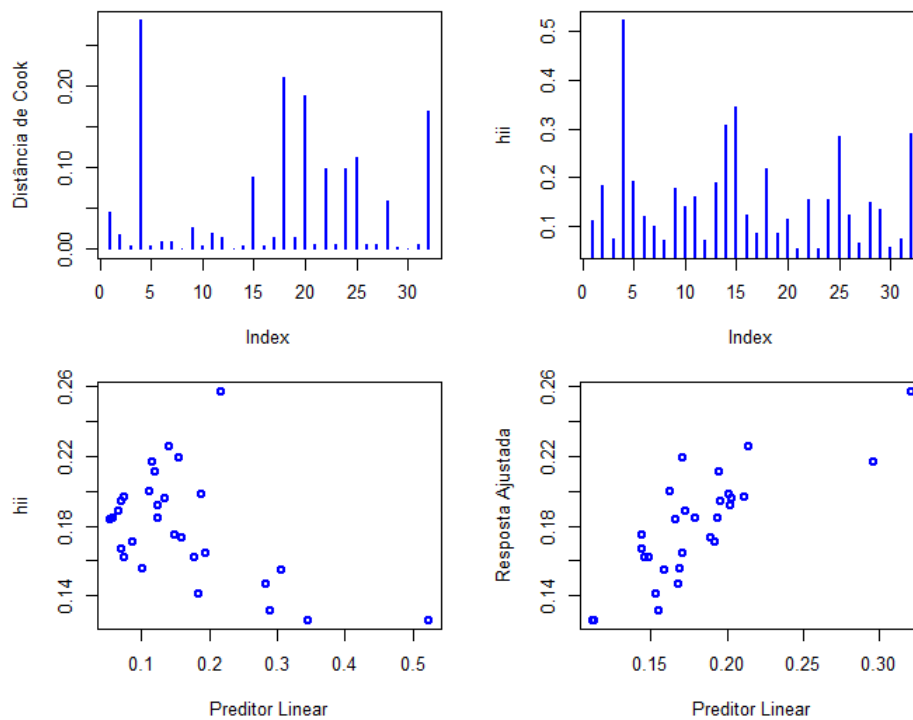


Figure 5: Medidas de influência qualidade de ajuste.

```
> ##-----##
> ## VI Análise do Desvio ##
> ##-----##

> anova(fit.Q05, test="Chisq")
Analysis of Deviance Table

Model: Gamma, link: inverse

Response: chol

Terms added sequentially (first to last)

Df Deviance      Resid.Df Resid.Dev   Pr(>Chi)
NULL                                31    1.52658
Age      1    0.50052         30    1.02606  1.283e-06 ***
bmi      1    0.22901         29    0.79705  0.001055 **
sex      1    0.00000         28    0.79705  0.996669
Age:sex  1    0.17480         27    0.62225  0.004215 **
-----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```