

Disciplina: Modelos Lineares Generalizados

Curso: Mestrado em Estatística

Código: PPGEST2217

Semestre: 2025.2

Prof. Frederico Machado Almeida

GABARITO LISTA #02

Observações:

- Questões para entregar: 2, 4 (a, b), 6 (c), 9 e 10
- Demais questões são apenas para estudar.
- Prazo de entrega: **22/10/2025**

Q1. Mostre que:

- (a) GA é um g-inverso de GA se $G = A^{-1}$.
- (b) Seja c um escalar. Mostre que, um g-inverso de cA é G/c .
- (c) Mostre que \mathbf{I}_n/n é um g-inverso da matriz \mathbf{J}_n .

Q2. Seja G um g-inverso da matriz $A^\top A$. Mostre que,

- (a) $AG^\top A^\top A = A$

Resp: A prova dos itens desta questão, segue das propriedades matriciais apresentadas no SLIDE-GLM03, pag. 03, da definição de g -inverso, e usando o fato de que a matriz inversa generalizada G é simétrica, isto é, $G^\top = G$. Desta forma, de acordo com a definição, obtemos:

$$A^\top AGA^\top A \stackrel{\text{def}}{=} A^\top A \iff (A^\top AGA^\top A)^\top = (A^\top A)^\top \iff A^\top AG^\top A^\top A = A^\top A.$$

Assim, segue da igualdade do produto matricial apresentado no SLIDE-GLM03, pag. 03 que, $AG^\top A^\top A = A$.

- (b) $A^\top AGA^\top = A^\top$

Resp: Segue a mesma ideia do item (a), e da definição do g -inverso que:

$$A^\top AGA^\top A \stackrel{\text{def}}{=} A^\top A \iff A^\top AGA^\top A = A^\top A \iff A^\top AGA^\top = A^\top.$$

- (c) $A^\top AG^\top A^\top = A^\top$

Resp: Segue de forma similar ao resultado do item (a) que,

$$A^\top AGA^\top A \stackrel{\text{def}}{=} A^\top A \iff (A^\top AGA^\top A)^\top = (A^\top A)^\top \iff A^\top AG^\top A^\top A = A^\top A.$$

Consequentemente, $AG^\top A^\top = A^\top$. Esse resultado pode ser obtido imediatamente usando o fato de G ser uma matriz simétrica.

- (d) $AG^\top A^\top = \textcolor{red}{AG}A^\top$

Resp: Em primeiro lugar, a pergunta estava mal formulada. Após a correção da mesma, e usando o fato de G ser uma matriz simétrica, segue que:

$$AG^\top A^\top \stackrel{\text{simetria}}{=} AGA^\top.$$

(e) $AG^T A^T$ é simétrica.

Resp: Para mostrar que $AG^T A^T$ é simétrica, segue que:

$$(AG^T A^T)^T = (A^T)^T (G^T)^T A^T = AG^T A^T.$$

Segue da simetria de G que, $(G^T)^T = G^T = G$.

Q3. Seja \mathbf{X} uma matriz $n \times p$, mostre que,

(a) $\mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{X}^T \mathbf{X}$, onde $\mathbf{W} = G \mathbf{X}^T \mathbf{X}$ e $G = (\mathbf{X}^T \mathbf{X})^{-1}$.

(b) $\mathbf{I}_n - \mathbf{X} G \mathbf{X}^T$ é simétrica e idempotente.

(c) Se $H = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ então, $H \mathbf{X} = \mathbf{X}$.

(d) Supondo que a matriz \mathbf{X} seja particionada em $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, onde \mathbf{X}_1 é uma matriz $n \times k$ e \mathbf{X}_2 é uma matriz de dimensão $n \times (p - k)$. Mostre que:

(i) As matrizes $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ e $\mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$ são idempotentes.

(ii) $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T$ é idempotente.

(iii) Obtenha o traço de $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T$.

(iv) Obtenha o traço de $\mathbf{X}_2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1} \mathbf{X}_2^T$.

Q4. Seja $H = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, a matriz de projeção no plano gerado pelas colunas de \mathbf{X} . Pode-se para mostrar as seguintes propriedades:

(a) $0 \leq h_{ii} \leq 1$

Resp: Temos a seguinte matriz de projeção $H = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, tal que, h_{ij} denotam as entradas genéricas (i, j) de \mathbf{H} . Observe também que, da equação (1) temos que, tendo em conta a definição e propriedades dos elementos de \mathbf{H} , segue que,

$$h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2. \quad (1)$$

Da equação (1), é fácil observar que, como h_{ii} pode ser escrito como a soma dos quadrados de h_{ij} , então, segue que $h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 \geq 0^{(*)}$, para todo i, j . Assim, concluímos que

$h_{ii} \geq 0$. Desta forma, podemos provar a preposição anterior de duas maneiras: (i) Se assumirmos que $h_{ii} > 1$, a expressão em $(*)$ resultaria em um absurdo. Logo, como $h_{ii} \geq 0$, concluímos que $0 \leq h_{ii} \leq 1$.

(ii) A segunda maneira de provar a propriedade anterior, seria considerar o resultado da expressão em $(*)$, tal que, $h_{ii} \geq h_{ii}^2 \iff h_{ii} - h_{ii}^2 \geq 0^{(**)}$, que é uma inequação quadrática com $a = -1$, $b = 1$ e $c = 0$. Assim, $h_{ii} - h_{ii}^2 = h_{ii} (1 - h_{ii}) = 0$, resultado nas seguintes raízes, $h_{ii} = 0$ ou $h_{ii} = 1$. Como $a = -1 < 0$ então, o conjunto solução da inequação será $0 \leq h_{ii} \leq 1$.

(b) $-0.5 \leq h_{ij} \leq 0.5$

Resp: Para mostrar essa propriedade, vamos usar o resultado do item anterior. Tal que,

$$h_{ii} = h_{ii}^2 + h_{ij}^2 + \sum_{i, j \neq \ell} h_{i\ell}^2 \geq 0.$$

Consequentemente, $h_{ii} \geq h_{ii}^2 + h_{ij}^2 \iff h_{ii} - h_{ii}^2 \geq h_{ij}^2 \iff h_{ij}^2 \leq \max \{h_{ii}(1 - h_{ii})\}$. Como $h_{ii} = 0$ ou $h_{ii} = 1$ são as possíveis soluções da equação quadrática $h_{ii} - h_{ii}^2 = 0$, o ponto máximo da função será alcançado no ponto x_v (x -vértice), isto é, $h_{ii}^v = \frac{1+0}{2} = 1/2$. Por fim, $h_{ij}^2 \leq \max \{h_{ii}^v(1 - h_{ii}^v)\} = 1/4 \iff h_{ij}^2 - 1/4 \leq 0$. Resolvendo essa inequação quadrática obtemos $-\frac{1}{2} \leq h_{ij} \leq \frac{1}{2}$.

(c) $\sum_{i=1}^n \sum_{j=1}^n h_{ij}^2 = p$.

Q5. Para um modelo de regressão linear múltipla dado por $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, com $\boldsymbol{\epsilon} \sim \mathcal{NM}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Denote $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$, o estimador dos mínimos quadrados de $\boldsymbol{\beta}$, $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, os valores ajustados e H a matriz de projeção no plano gerado pelas colunas de \mathbf{X} . Nos itens (a) e (b), pede-se para mostrar que:

(a) $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - H) \boldsymbol{\epsilon}$. Sendo \mathbf{e} um estimador para o $\boldsymbol{\epsilon}$.

(b) $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}^\top \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y}$.

(c) Nos itens (i)-(iv), obtenha:

(i) $Cov[(\mathbf{I} - H) \mathbf{y}, H \mathbf{y}]$.

(ii) $Cov(\mathbf{y}, \hat{\mathbf{y}})$.

(iii) $Cov(\hat{\mathbf{y}}, \hat{\boldsymbol{\beta}})$.

(iv) $Cov(\mathbf{e}, \hat{\boldsymbol{\beta}})$.

Q6. Considere o enunciado da Questão 05, e denote por SQT , $SQReg$ e SQE as somas do quadrado dos totais, da regressão e dos erros. Mostre que:

(a) $\mathbb{E}(SQT) = (n - 1)\sigma^2 + (\mathbf{X}\boldsymbol{\beta})^\top \left(\mathbf{I} - \frac{1}{n} \mathbf{J}\right) (\mathbf{X}\boldsymbol{\beta})$.

(b) $\mathbb{E}(SQE) = (n - p)\sigma^2$.

(c) $\mathbb{E}(SQReg) = (p - 1)\sigma^2 + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}$.

Resp: A resolução desse item segue a mesma ideia do item (a). Isto é, com base nas notas de aulas SLIDE-GLM03, pag.32 temos o seguinte resultado: $SQReg = \mathbf{y}^\top \left(\mathbf{H} - \frac{1}{n} \mathbf{J}_n\right) \mathbf{y}$. Por simplicidade, assumamos que $W = \left(\mathbf{H} - \frac{1}{n} \mathbf{J}_n\right)$, tal que, $tr(W) = tr\left(\mathbf{H} - \frac{1}{n} \mathbf{J}_n\right) = tr(\mathbf{H}) - \frac{1}{n} tr(\mathbf{J}_n) = p - 1$. Consequentemente,

$$\begin{aligned} \mathbb{E}(SQReg) &= \mathbb{E}(\mathbf{y}^\top W \mathbf{y}) = tr\{\mathbb{E}(\mathbf{y}^\top W \mathbf{y})\} = \mathbb{E}\{tr(\mathbf{y}^\top W \mathbf{y})\} \\ &= \mathbb{E}\{tr(W \mathbf{y}^\top \mathbf{y})\} = tr\{\mathbb{E}(W \mathbf{y}^\top \mathbf{y})\} = tr\{W \mathbb{E}(\mathbf{y}^\top \mathbf{y})\} \\ &= tr\{W (Var(\mathbf{y}) + \mathbb{E}(\mathbf{y})^\top \mathbb{E}(\mathbf{y}))\} \\ &= tr\{W (\sigma^2 \mathbf{I}_n + (\mathbf{X}\boldsymbol{\beta})^\top (\mathbf{X}\boldsymbol{\beta}))\} \\ &= \sigma^2 tr\{W\} + tr\{W (\mathbf{X}\boldsymbol{\beta})^\top (\mathbf{X}\boldsymbol{\beta})\} \\ &= \sigma^2 (p - 1) + (\mathbf{X}\boldsymbol{\beta})^\top W (\mathbf{X}\boldsymbol{\beta}) \\ &= \sigma^2 (p - 1) + (\mathbf{X}\boldsymbol{\beta})^\top \left(\mathbf{H} - \frac{1}{n} \mathbf{J}_n\right) (\mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

No entanto, como $\mathbf{X} = [\mathbf{1}, X_1]$, e a soma dos quadrados da regressão não leva em conta o intercepto, defina $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \dots, \beta_p)^\top$. Consequentemente, $\mathbb{E}(SQReg) = \sigma^2 (p - 1) + (\mathbf{X}_1 \boldsymbol{\beta}_1)^\top \left(\mathbf{H} - \frac{1}{n} \mathbf{J}_n\right) (\mathbf{X}_1 \boldsymbol{\beta}_1)$.

(d) A estatística F da tabela ANOVA pode ser dada por, $F_{cal} = \frac{R^2}{1-R^2} \frac{n-p}{p-1} \sim F_{(p-1, n-p)}$, sob uma H_0 verdadeira.

Q7. Considere o modelo de regressão linear múltipla, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, com $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

(a) Mostre que a função de verossimilhança do modelo é dada por:

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= \frac{1}{(2\pi)^{n/2} |\sigma^2 \mathbf{I}|^{1/2}} e^{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) / 2\sigma^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) / 2\sigma^2}, \text{ com } \sigma^2 > 0. \end{aligned}$$

(b) Obtenha o estimador dos mínimos quadrados do vetor $\boldsymbol{\beta}$.

(c) Obtenha o estimador para a variância dos erros $\hat{\sigma}^2$.

(d) Mostre que $\hat{\sigma}^2$ é um estimador tendencioso para σ^2 .

Q8. Considere a regressão linear simples dada por $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, com $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, tal que, $\sigma_i^2 = \text{Var}(\epsilon_i | x_i) = x_i^2 \sigma^2$, $i = 1, \dots, n$.

(a) Obtenha os estimadores dos mínimos quadrados ponderados para β_0 e β_1 e as respectivas variâncias.

(b) Obtenha as respectivas somas dos quadrados dos totais, da regressão e dos erros.

(c) Considere agora que o modelo de regressão passa pela origem (isto é, assumindo que $\beta_0 = 0$), tal que, $y_i = \beta x_i + \epsilon_i$, com $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. Obtenha o estimador dos mínimos quadrados para β .

Q9. Uma consultora pretende avaliar a influência de uma variável independente x na resposta y . No entanto, a análise feita a partir do conjunto de dados apresentado no ficheiro [Q09-data.txt](#) mostrou que algumas das principais suposições do modelo clássico de regressão não foram satisfeitas. Usando o método de estimação por mínimos quadrados ponderados, proponha uma solução para o problema.

Resp: Note que o conjunto de dados é composto por uma única variável explicativa, e os resultados do modelo ajustado são apresentados abaixo. Como podemos observar, os resultados do ajuste indicam que a variável independente é importante para modelar as variações da resposta. No entanto, a validade do modelo está atrelado a verificação das suposições básicas de um modelo clássico de regressão.

Abaixo, é apresentado o síntese do ajuste dos modelos onde diferentes configurações para os pesos, w_i 's são consideradas. Como pode ser observado, o modelo usual (com $w_i = 1$ para todo i), ou sem ponderação apresentou resíduos (ordinários e internos estudentizados) claramente heterocedásticos, com variância decrescendo para valores pequenos de x_i , estabiliza levemente para algum intervalo de x_i e depois volta a crescer para valores maiores de x_i . Como consequência, os resíduos ordinários e os estudentizado não estão normalmente distribuídos (vide no teste de Shapiro-Wilks para o modelo sem ponderação, e o QQ plot apresentado na Figura 1, painel (a)).

Várias ponderações, w_i 's foram especificadas para os modelos ajustados nos Métodos (1) à (3). Porém, como podemos observar nos resultados reportados abaixo, e nos gráficos apresentados nas Figuras 1 e 2. Portanto, entre várias ponderações, a única que mostrou-se ser a mais eficiente é aquela baseada nos pesos, $w_i = 1/x_i^2$ (modelo apresentado no Método

3). Para esse modelo, os resíduos internos estudentizados estão normalmente distribuídos (ou estudentizados), conforme apresentado no painel (d) da figura abaixo.

De igual forma, o teste de Shapiro-Wilks mostra que a hipótese de normalidade dos resíduos não deve ser rejeitada, pois, o valor-p associado a estatística de Shapiro-Wilks (valor-p=0,5227) é maior do que qualquer nível de significância, α . Conforme apresentado em aula, a variância dos resíduos ordinários, $e_i = y_i - \hat{y}_i$ depende dos valores de alavancagem, isto é, $e_i \sim \mathcal{N}(0, (1 - h_{ii})\sigma^2)$, por esse motivo, os resíduos internos estudentizados, $r_i = \frac{e_i}{\hat{\sigma}\sqrt{(1-h_{ii})}}$ são mais informativos que os resíduos ordinários. Como conclusão, a versão do modelo final com variância aproximadamente constante, e resíduos internos aproximadamente normais é:

```
Call:lm(formula = y ~ ., data = Q09.data)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.5932	-0.7230	-0.1380	0.8016	5.9595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.85196	0.13258	-6.426	9.55e-10 ***
x	0.39307	0.06601	5.954	1.17e-08 ***

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

Residual standard error: 1.866 on 198 degrees of freedom

Multiple R-squared: 0.1519, Adjusted R-squared: 0.1476

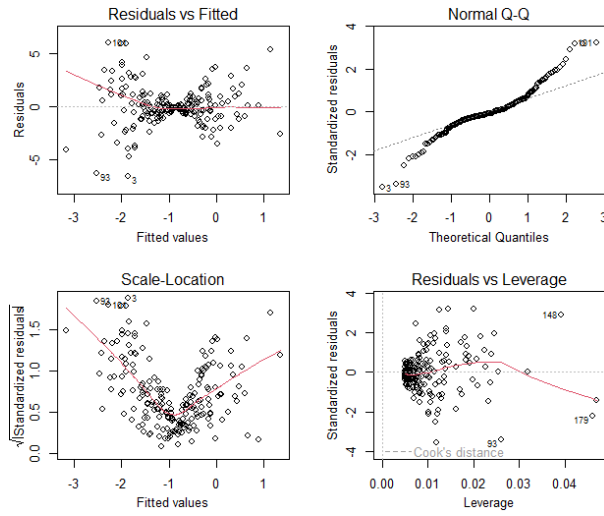
F-statistic: 35.45 on 1 and 198 DF, p-value: 1.173e-08

```
##-----##
## Modelo sem ponderação (wi=1) ##
##-----##
> ## Teste de Shapiro-Wilks
> ## H0: Os resíduos são normais
> ## H1: Os resísdos não são normais
> shapiro.test(studResFit)
##
Shapiro-Wilk normality test
data: studResFit
W = 0.93846, p-value = 1.688e-07
##
##
##-----##
## Método 1: Usando wi=1/|ei.hat| ##
##-----##
> ## Teste de Shapiro-Wilks
> ## H0: Os resíduos são normais
> ## H1: Os resísdos não são normais
> shapiro.test(studRes1)
##
Shapiro-Wilk normality test
data: studRes1
```

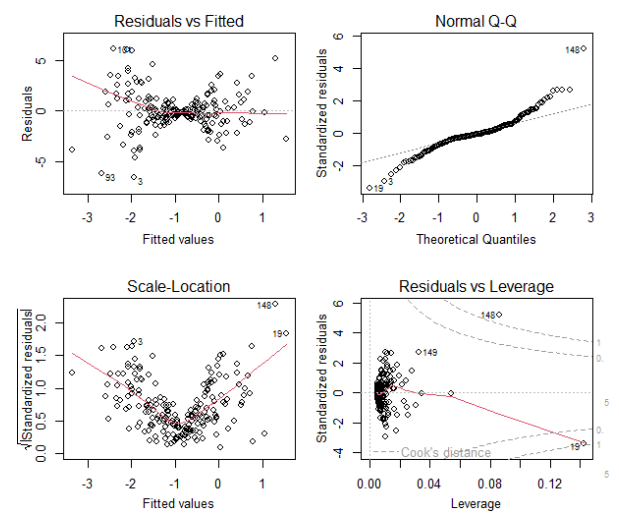
```
W = 0.92019, p-value = 6.008e-09
##
##
> ##-----##
> ## Método 2: Usando wi=1/|x| ##
> ##-----##
> ## Teste de Shapiro-Wilks
> ## H0: Os resíduos são normais
> ## H1: Os resísudos não são normais
> shapiro.test(studRes2)
##
Shapiro-Wilk normality test
data: studRes2
W = 0.98366, p-value = 0.02019
##
##
> ##-----##
> ## Método 3: Usando wi=1/x^2 ##
> ##-----##
> ##
> ## Teste de Shapiro-Wilks
> ## H0: Os resíduos são normais
> ## H1: Os resísudos não são normais
> shapiro.test(studRes3)
##
Shapiro-Wilk normality test
data: studRes3
W = 0.99346, p-value = 0.5227

> summary(fit.weig3)
Call:lm(formula = y ~ x, data = Q09.data, weights = 1/wght3^2)
##
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.998995  0.001076 -928.547  < 2e-16 ***
x            0.402871  0.067923   5.931 1.32e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
Residual standard error: 0.9568 on 198 degrees of freedom
Multiple R-squared:  0.1509, Adjusted R-squared:  0.1466
F-statistic: 35.18 on 1 and 198 DF, p-value: 1.32e-08
```

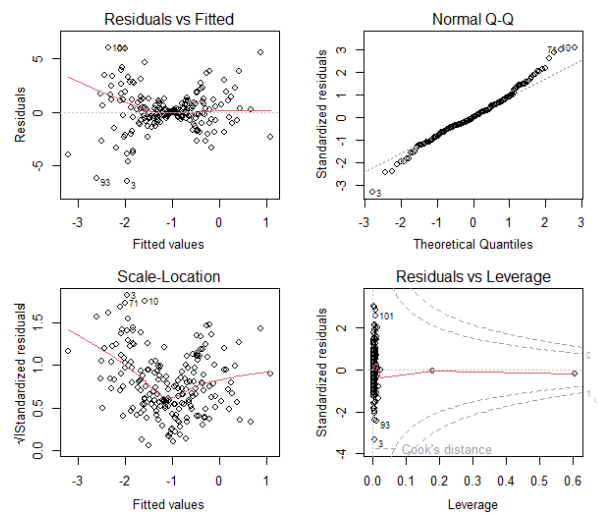
Painel (a): Modelo sem ponderação



Painel (b): Modelo com ponderação $W_i = 1/|e_i|$



Painel (c): Modelo com ponderação $W_i = 1/|x_i|$



Painel (d): Modelo com ponderação $W_i = 1/x_i^2$

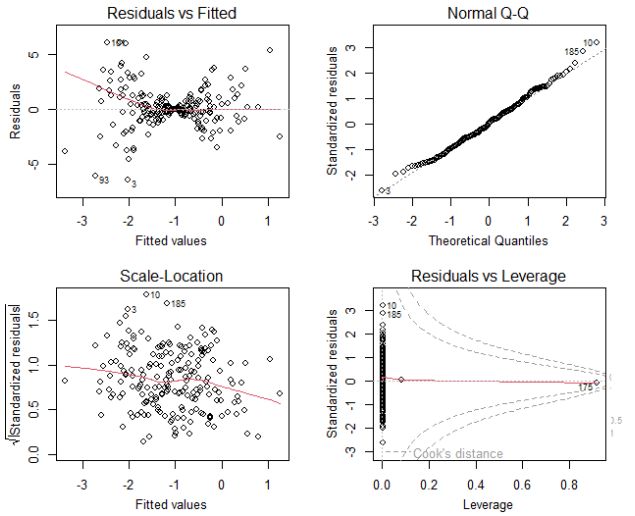


Figure 1: Gráfico dos resíduos de diferentes modelos. Diferentes pesos w_i foram considerados.

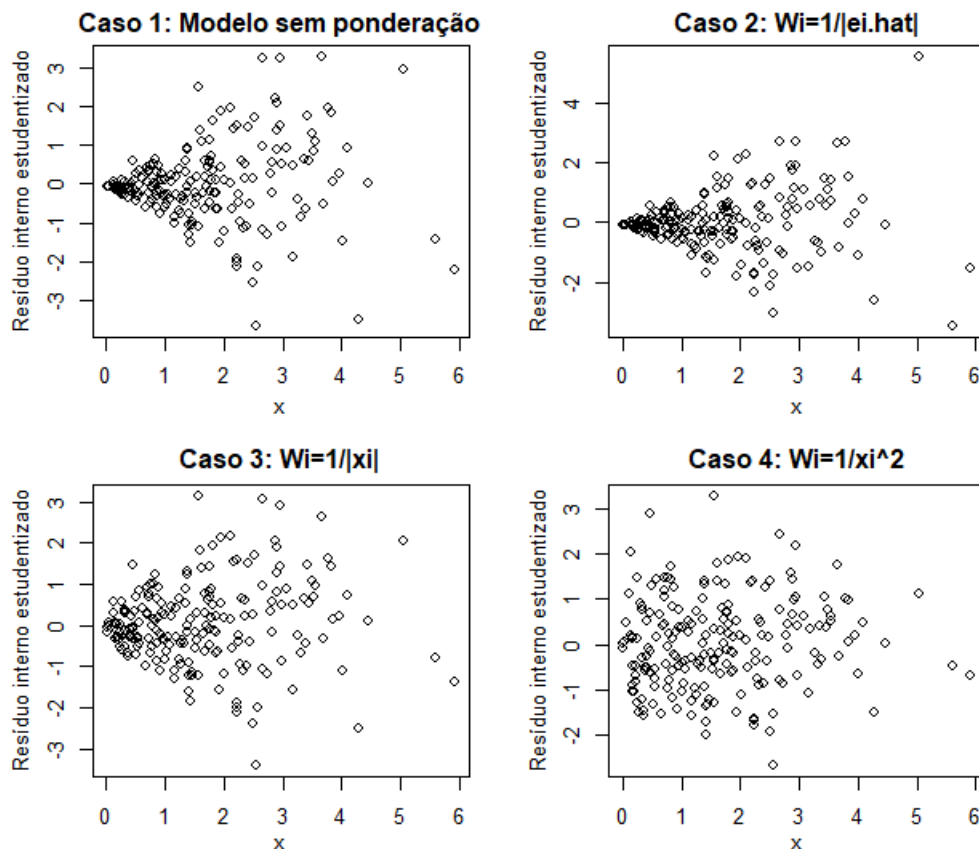


Figure 2: Gráfico dos resíduos internos estudentizados para os 4 modelos considerados.

Q10. Uma equipe de pesquisadores de saúde mental deseja comparar três métodos de tratamento da depressão grave (A, B e C=referência). Eles também gostariam de estudar a relação entre idade e eficácia do tratamento, bem como a interação (se houver) entre idade e tratamento. Cada elemento da amostra aleatória simples de 36 pacientes, foi selecionado aleatoriamente para receber o tratamento A, B ou C. Os dados obtidos podem ser encontrados no ficheiro [Q10-data.txt](#). A variável dependente y é a eficácia do tratamento; as variáveis independentes são: a idade do paciente no aniversário mais próximo, e o tipo de tratamento administrado (use 3% de significância durante as análises).

```
> summary(Q10)
eficacia      idade      tratamento
Min.   :25.00   Min.   :19.00   Length:36
1st Qu.:46.75   1st Qu.:32.25   Class :character
Median :58.00   Median :44.00   Mode  :character
Mean   :55.17   Mean   :44.11
3rd Qu.:63.25   3rd Qu.:56.50
Max.   :73.00   Max.   :67.00
> str(Q02)
'data.frame': 36 obs. of 3 variables:
 $ eficacia  : int  56 41 40 28 55 25 46 71 48 63 ...
 $ idade     : int  21 23 30 19 28 23 33 67 42 33 ...
 $ tratamento: chr  "A" "B" "B" "C" ...
```

(a) Ajuste o modelo de regressão linear e interprete os resultados obtidos.

Resp: Como é do vosso conhecimento, atendendo o fato de termos três níveis do fator **tratamento**, é possível construir duas variáveis dummies. E conforme o enunciado, vamos considerar o nível **C** como sendo de referência. Assim, obtemos:

$$x_{Ai} = \begin{cases} 1, & \text{para o tto A} \\ 0, & \text{Caso contrário.} \end{cases} \quad \text{e} \quad x_{Bi} = \begin{cases} 1, & \text{para o tto B} \\ 0, & \text{Caso contrário.} \end{cases}$$

Assim, o modelo ajustado no item (a) é apresentado abaixo:

Call:

```
lm(formula = y ~ x1 + xA + xB, data = Q10.data)
```

####

Residuals:

Min	1Q	Median	3Q	Max
-12.5732	-3.3922	0.9829	3.9613	9.5062

####

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	22.29059	3.50510	6.359	3.85e-07 ***
x1	0.66446	0.06978	9.522	7.42e-11 ***
xA	10.25276	2.46542	4.159	0.000224 ***
xB	0.44518	2.46399	0.181	0.857763

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

####

Residual standard error: 6.035 on 32 degrees of freedom

Multiple R-squared: 0.784, Adjusted R-squared: 0.7637

F-statistic: 38.71 on 3 and 32 DF, p-value: 9.287e-11

(i) O modelo ajustado para as três variáveis independentes sem levar em conta a interação entre as variáveis mostrou que, o tratamento *B* para a depressão grave não é importante para explicar a eficácia (*y*). Os demais coeficientes, incluindo o intercepto, são todos significativos ao nível de significância de 5% (porque apresentam valores-p inferiores a 0,05). Lembrando que, estes resultados foram obtidos testando a significância individual dos β_j 's, por meio das seguintes hipóteses, $H_0 : \beta_j = 0$ contra $H_1 : \beta_j \neq 0$, com $j = 0, 1, 2, 3$.

(ii) Interpretação dos coeficientes de regressão: observe que, para o estudo em particular, o intercepto $\hat{\beta}_0 = 22,29$ não tem uma interpretação prática. Pois, como o interesse é de estudar o impacto das três variáveis independentes na eficácia do tratamento da depressão grave, não faz sentido assumir um valor zero, para a variável idade. Ademais, $\hat{\beta}_1 = 0,66$ indica a magnitude do aumento na eficácia do tratamento a cada ano que passa, mantendo constante os demais fatores de risco. Tal resultado, não parece se um pouco conflitante com a realidade. Por outro lado, o valor $\hat{\beta}_2 = 10,25$ indica o aumento na eficácia para pacientes administrados com tratamento do tipo A, mantendo constante os demais fatores. Ou seja, a contribuição do tratamento A nas variações de *y* (eficiência) é maior que o do tratamento B.

(iii) O coeficientes de determinação indica que, 78,4% das variações da resposta (*y*) são explicadas pelo modelo. Desta forma, a quantidade remanescente, digamos, 21,6% das variações do *y* são atribuídas à causas aleatórias (efeito do erro aleatório). O teste de

significância global (testa F), indica a existência de ao menos um coeficiente de regressão diferente de 0 (vide a tabela ANOVA apresentada no item *b*).

- (b) Obtenha a tabela ANOVA para o modelo obtido no item (a) e interprete os resultados.

Resp: Usando a função `anova(object)` do R, obtemos a tabela com as respectivas somas dos quadrados parciais. Ou seja, esses resultados tem como propósito de avaliar a contribuição da cada covariável (ao incluí-los separadamente no modelo) em reduzir a soma dos quadrados dos erros (ou aumentar a soma dos quadrados da regressão)

```
##-----##
## Anova for Extra Sum Squares ##
##-----##
> anova(fit.q2)
Analysis of Variance Table

Response: y
Df Sum Sq Mean Sq F value    Pr(>F)
x1      1 3424.4   3424.4  94.0153 4.797e-11 ***
xA      1  803.8    803.8  22.0679 4.775e-05 ***
xB      1    1.2     1.2   0.0326  0.8578
Residuals 32 1165.6    36.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
##
##-----##
## Anova for Global Test ##
##-----##
as.table(Anova.table)
```

Sources	SumSquare	df	MeanSquare	Fcal	p-value
Regression	4229.425	3	1409.808	38.705	9.28808293396552e-11
Error	1165.575	32	36.424		
Total	5395	35			

O teste de significância global objetiva testar as seguintes hipóteses de interesse, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ vs $H_1 : \exists \beta_j \neq 0$, com $j = 1, 2, 3$. Os resultados da tabela ANOVA, são os mesmos reportados no item anterior. Ou seja, o teste de significância global (testa F), indica a existência de ao menos um coeficiente de regressão diferente de 0.

- (c) Considere a possibilidade de incluir a interação entre as variáveis independentes, i.e., assumo que o modelo a ser ajustado tem a seguinte formulação: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{Ai} + \beta_3 x_{Bi} + \beta_4 x_{1i} x_{Ai} + \beta_5 x_{1i} x_{Bi} + \epsilon_i$, com a suposição de que $\epsilon_i \sim N(0, \sigma^2)$. Com base no modelo anterior,

- (c₁) Liste todos os submodelos que podem ser obtidos considerando os tratamentos, *A*, *B* e *C*.

Resp: Com base nas variáveis dummies apresentadas no item anterior, segue que, os possíveis submodelos que podem surgir, tendo em conta cada nível do tratamento administrado são:

Modelo 1 (tto A, $x_{Ai} = 1; x_{Bi} = 0$) : $y_i = \beta_0 + \beta_2 + (\beta_1 + \beta_4) x_{1i} + \epsilon_i$

Modelo 2 (tto B; $x_{Ai} = 0; x_{Bi} = 1$) : $y_i = \beta_0 + \beta_3 + (\beta_1 + \beta_5) x_{1i} + \epsilon_i$

Modelo 3 (tto C; $x_{Ai} = 0; x_{Bi} = 0$) : $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$

(c₂) Interprete os coeficientes de regressão associados aos fatores de interação.

Resp: Ajuste do modelo com termos de interação.

Call:

```
lm(formula = y ~ x1 + xA + xB + x1:xA + x1:xB, data = Q10.data)
```

###

Residuals:

Min	1Q	Median	3Q	Max
-6.4366	-2.7637	0.1887	2.9075	6.5634

###

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.21138	3.34964	1.854	0.073545 .
x1	1.03339	0.07233	14.288	6.34e-15 ***
xA	41.30421	5.08453	8.124	4.56e-09 ***
xB	22.70682	5.09097	4.460	0.000106 ***
x1:xA	-0.70288	0.10896	-6.451	3.98e-07 ***
x1:xB	-0.50971	0.11039	-4.617	6.85e-05 ***

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

###

Residual standard error: 3.925 on 30 degrees of freedom

Multiple R-squared: 0.9143, Adjusted R-squared: 0.9001

F-statistic: 64.04 on 5 and 30 DF, p-value: 4.264e-15

(i) Os resultados apresentados acima, mostram claramente que, a inclusão dos termos de interação entre a variável idade (x_1) com os dois níveis de tratamento administrado, melhorou significativamente a importância dos fatores de risco. Observe que, o tratamento B (x_{Bi}) que não foi significativo no modelo sem interação, passou a ser, depois que os termos de interação foram adicionados ao modelo. Os resultados mostram ainda que os fatores de interação devem permanecer no modelo.

(ii) Podemos afirmar também que, depois que os fatores $x_1 * x_A$ e $x_1 * x_B$ foram incluídos no modelo, o poder do modelo melhorou, pois, a proporção das variações da resposta y que são explicadas pelo modelo, passou de 78,4 para 91,43%. Ademais, o teste de significância global dos coeficientes de regressão, leva-nos a concluir que existe pelo menos um β_j diferente de 0.

(iii) Interpretação do modelo: usando os submodelos apresentados no item (a – i), temos que,

Modelo 1 (tto A, $x_{Ai} = 1; x_{Bi} = 0$) : $\hat{y}_{Ai} = \hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_4) x_{1i} = 47,52 + 0,33x_{1i}$

Modelo 2 (tto B; $x_{Ai} = 0; x_{Bi} = 1$) : $\hat{y}_{Bi} = \hat{\beta}_0 + \hat{\beta}_3 + (\hat{\beta}_1 + \hat{\beta}_5) x_{1i} = 28,92 + 0,52x_{1i}$

Modelo 3 (tto C; $x_{Ai} = 0; x_{Bi} = 0$) : $\hat{y}_{Ai} = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} = 6,21 + 1,03x_{1i}$.

Assim, o valor 0,33 representa o aumento na eficácia do tratamento em pacientes que receberam o tto A, para cada variação unitária da variável idade. De igual forma, para cada variação unitária do fator idade, a eficácia do tratamento aumenta em 0,52 unidades, em pacientes que receberam o tto B; e 1,03 é o aumento na eficácia do tratamento, para cada variação unitária do fator idade, em pacientes que não receberam o tto A, nem o B.

- (c₃) Apresente a tabela anova para testar as seguintes hipóteses, $H_0 : \beta_4 = \beta_5 = 0$ contra $H_1 : \exists \beta_j \neq 0$, com $j = 1, 4, 5$. Há evidências suficientes para mudar o modelo apresentado em (c)?

H0: b4=b5=0 (Modelo reduzido) vs H1: Ambos os coeficientes são diferente de zero (Modelo mais completo).

##

##-----##

Modelo Sob H0 (modelo reduzido)

##-----##

> anova(reduzido)

Analysis of Variance Table

###

Response: y

Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	3424.4	3424.4	94.0153 4.797e-11 ***
xA	1	803.8	803.8	22.0679 4.775e-05 ***
xB	1	1.2	1.2	0.0326 0.8578
Residuals	32	1165.6	36.4	

##

##

##-----##

Modelo Sob H1 (modelo completo)

##-----##

> anova(completo)

Analysis of Variance Table

###

Response: y

Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	3424.4	3424.4	222.2946 2.059e-15 ***
xA	1	803.8	803.8	52.1784 4.857e-08 ***
xB	1	1.2	1.2	0.0772 0.7831
x1:xA	1	375.0	375.0	24.3430 2.808e-05 ***
x1:xB	1	328.4	328.4	21.3194 6.850e-05 ***
Residuals	30	462.1	15.4	

A tabela ANOVA parcial, para testar as hipóteses formuladas anteriormente, requer a obtenção da soma dos quadrados extra $SQReg(x_3, x_4|x_1, x_A, x_B)$, com $x_3 = x_1 * x_A$ e $x_4 = x_1 * x_B$, respectivamente.

Observe que, sob H_0 verdadeira, o modelo reduzido tem o seguinte formato: $y_i^R = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{Ai} + \beta_3 x_{Bi} + \epsilon_i^R$, cujos resultados do modelo reduzidos estão apresentados acima. Desta forma, em decorrência do modelo reduzido obtemos:

$$\begin{aligned}SQE(x_1, x_A, x_B) &= 1165,6 \\SQReg(x_1, x_A, x_B) &= SQReg(x_1) + SQReg(x_A|x_1) + SQReg(x_B|x_1, x_A) \\&= 3424,40 + 803,80 + 1,20 = 4229,40.\end{aligned}$$

De igual forma, supondo H_0 falsa, ou seja, sob H_1 obtemos o modelo completo. Isto é, $y_i^C = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{Ai} + \beta_3 x_{Bi} + \beta_4 x_{1i} x_{Ai} + \beta_5 x_{1i} x_{Bi} + \epsilon_i^C$, cujos resultados da tabela ANOVA parcial são apresentados acima. Onde, x_2 e x_3 denotam os termos de interação (segue da representação anterior). Portanto,

$$\begin{aligned}SQE(x_1, x_A, x_B, x_2, x_3) &= 462,10 \\SQReg(x_1, x_A, x_B) &= SQReg(x_1) + SQReg(x_A x_1) + SQReg(x_B|x_1, x_A) \\&= 3424,40 + 803,80 + 1,20 + 375,0 + 328,4 = 4932,80.\end{aligned}$$

Desta forma, a soma dos quadrados da regressão extra, que resulta da inclusão dos termos de interação no modelo que tinha apenas x_1 , x_A e x_B é dada por:

$$\begin{aligned}SQReg(x_2, x_3|x_1, x_A, x_B) &= SQE(x_1, x_A, x_B) - SQE(x_1, x_A, x_B, x_2, x_3) \\&= 1165,60 - 462,10 \approx 703,40. \quad \text{ou}\end{aligned}$$

$$\begin{aligned}SQReg(x_2, x_3|x_1, x_A, x_B) &= SQReg(x_1, x_A, x_B, x_2, x_3) - SQReg(x_1, x_A, x_B)^* \\&= 4932,80 - 4229,40 \approx 703,40 \quad (\text{mesmo resultado que o anterior}).\end{aligned}$$

Tal como foi visto em sala de aula, a inclusão (ou exclusão) de covariáveis no modelo altera apenas as somas dos quadrados da regressão e do erro. Ou seja, a soma dos quadrados dos totais continua inalterável. Assim, segue de (*) que, a soma dos quadrados do erro pode ser obtida da seguinte forma:

$$\begin{aligned}SQE.Extra &= SQT - SQReg(x_2, x_3|x_1, x_A, x_B) - SQReg(x_1, x_A, x_B) \\&= 5395,0 - 703,40 - 4229,40 = 462,10 \quad \text{ou},\end{aligned}$$

podemos obter a soma dos quadrados dos erros extra, fazendo a diferença entre a soma dos quadrados dos erros para o modelo reduzido, e a soma dos quadrados da regressão extra, resultado em: $SQE.Extra = 1165,57 - 703,43 \approx 462,10$ (que é a soma dos quadrados dos erros do modelo completo!!!)

Fonte de variação	SQ	GL	QM	F_{cal}	valor - p
Regressão	703,40	2**	351,7	22,84	9.377822e - 07
Erro	462,10	30***	15,40		

Table 1: Legenda: FV-fonte de variação; SQ-soma dos quadrados; GL-graus de liberdade e QM-quadrado médio.

Obs.: (**) o número de graus de liberdade é dado conforme apresentado em sala de aula. E, este valor equivale ao número de parâmetros que estão sendo testados na H_0 . Já, em (***) denota o número de graus de liberdades do erro no modelo completo. Por fim, os resultados acima corroboram com os resultados da comparação de modelos apresentados abaixo.

Analysis of Variance Table

###

```
> anova(res.model, augment.model)
```

Analysis of Variance Table

Model 1: $y \sim x1 + xA + xB$

Model 2: $y \sim x1 + xA + xB + x1:xA + x1:xB$

	Res.Df	RSS	Df	Sum-of-Sq	F	Pr(>F)
1	32	1165.57				
2	30	462.15	2	703.43	22.831	9.41e-07 ***

Com base nos resultados acima, concluímos que há evidências suficientes para rejeitar a hipótese nula. Ou seja, o modelo completo (que inclui os termos de interação), é melhor que o reduzido (sem termos de interação).

(c₄) Faça uma análise completa dos resíduos do modelo.

Resp: Os gráficos da Figura c₄, mostram claramente que, embora algumas observações sugiram levemente um padrão de contorno, é correto afirmar que, tais resíduos estão espalhados aleatoriamente em torno do zero. Ademais, a suposição de normalidade dos resíduos pode ser observada a partir do gráfico Q-Q plot. Nota-se que, apesar de existirem observações fora reta, os pontos desviantes não tem força o suficiente para afetar a normalidade dos resíduos. Tais conclusões podem ser confirmadas por meio do teste de Shapiro-Wilks apresentado abaixo.

De igual forma, tanto a análise gráfica apresentada na Figura c₄, como o teste de Breusch-Pagan, ambos levam-nos a concluir que os resíduos são homoscedásticos. Com base nos gráficos da análise dos resíduos, é correto afirmar que, não há evidências suficientes para rejeitar a hipótese de independência entre os resíduos (Teste de Durbin-Watson).

No que diz respeito à existência de pontos de influência, os gráficos sugerem que, as observações 1, 2, 30 e 35 são potenciais pontos de influência, segundo os critérios COVRATIO, e a observação 32 é um potencial ponto de influência nos valores ajustados segundo o DFFITS (veja a Figura c₄).

```
##-----##
##      Shapiro-Wilks Normality test      ##
##-----##
Shapiro-Wilk normality test
data:  resid(fit.q2_interaction)
W = 0.96286, p-value = 0.263
##
##-----##
## Test for heterocedascity (Breuch-Pagan) ##
##-----##
studentized Breusch-Pagan test
data:  fit.q2_interaction
BP = 2.9958, df = 5, p-value = 0.7006
##
##-----##
##      Test for errors independence (DW)  ##
##-----##
Durbin-Watson test
```

data: fit.q2_interaction
DW = 1.6326, p-value = 0.1214
alternative hypothesis: true autocorrelation is greater than 0

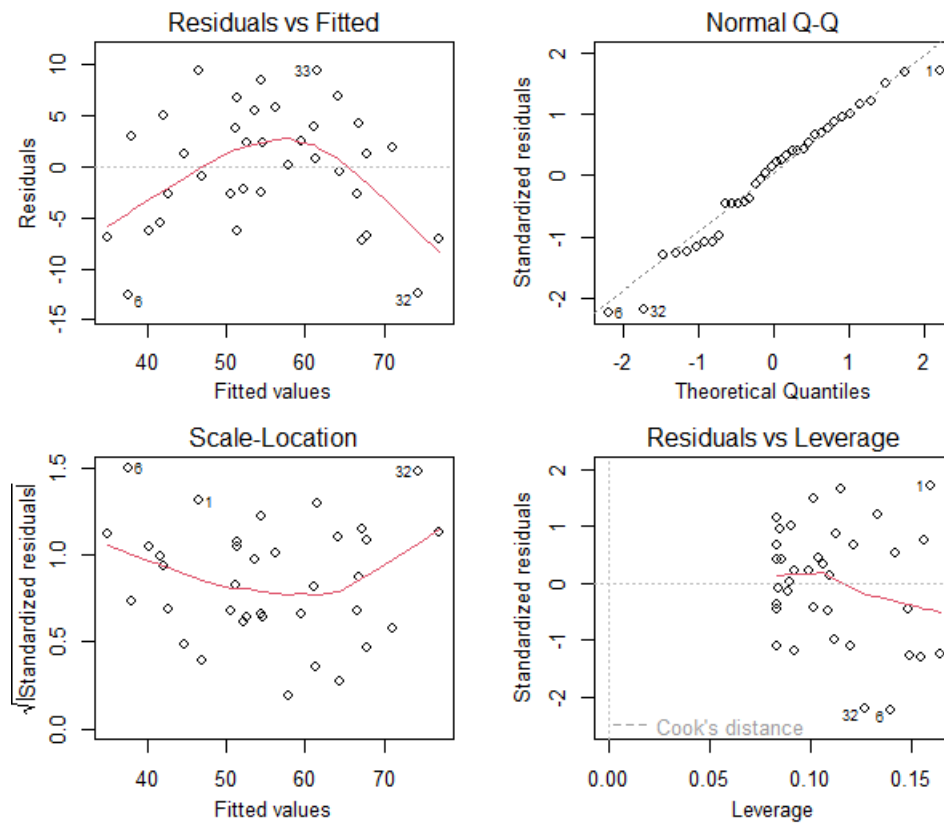


Figure 3: Análise de resíduos

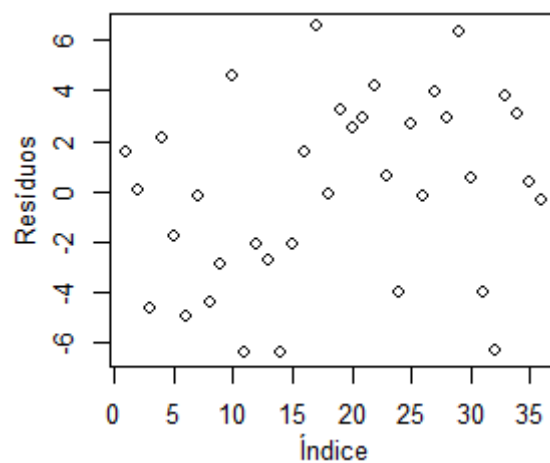


Figure 4: Gráfico de autocorrelação serial.

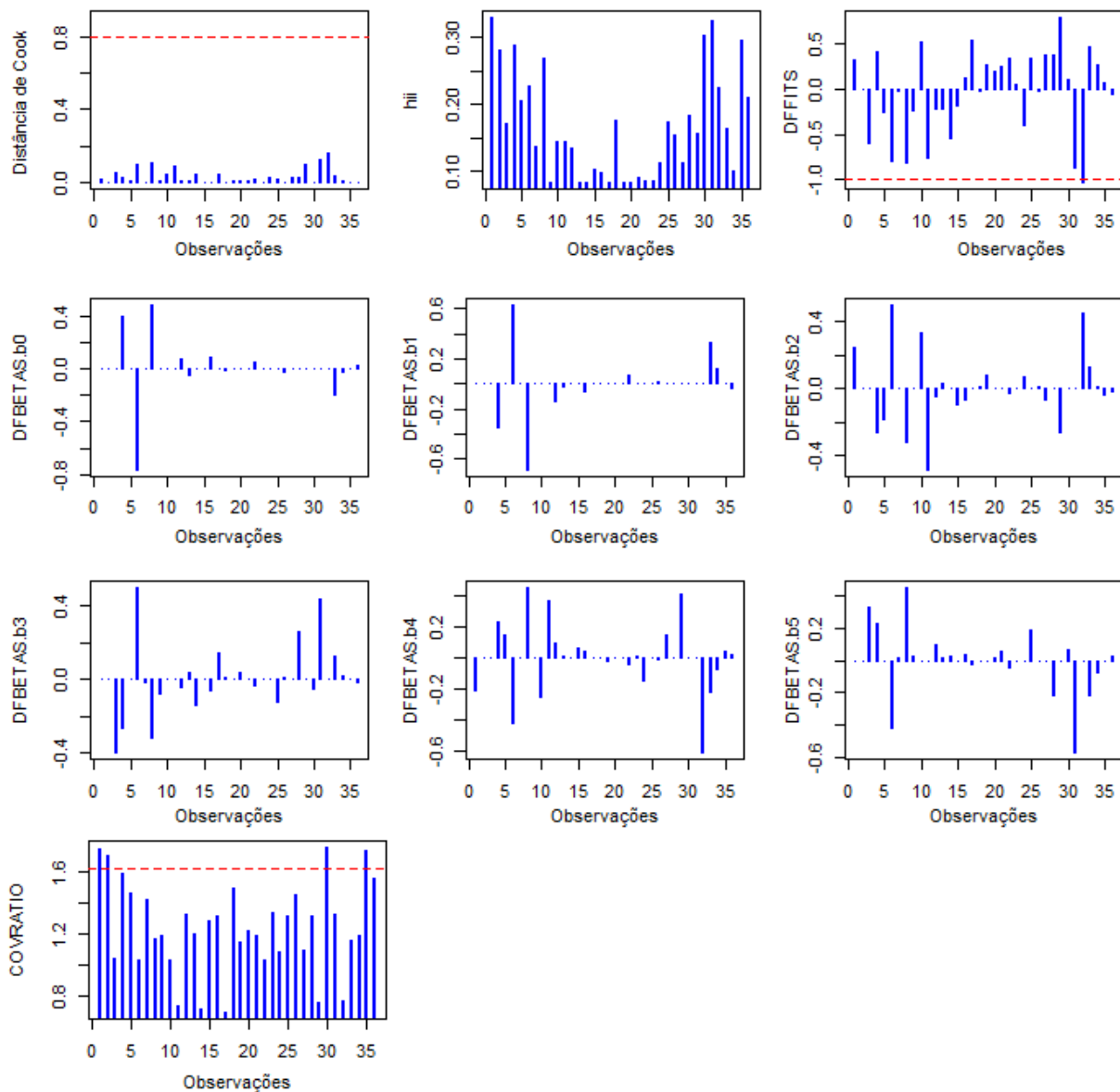


Figure 5: Medidas de influência.

(d) Considere os seguintes modelos, **modelo I**: $y_i^I = \beta_0 + \beta_1 x_{1i} + \epsilon_i^I$, e **Modelo II**: $y_i^{II} = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{Bi} + \beta_4 x_{1i} x_{Bi} + \epsilon_i^{II}$.

(d₁) Considere a possibilidade de incluir no **modelo I**, os fatores de risco x_{Ai} , x_{Bi} e a interação dos dois com x_{1i} . Formule as hipóteses de interesse e conclua ao nível de 5% de significância.

Resp: Para esse item temos as seguintes hipóteses:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 \text{ Modelo reduzido (modelo I)}$$

$$H_1 : \exists \beta_j \neq 0 \text{ com } j = 2, 3, 4, 5 \text{ Modelo completo}$$

```
##-----##
## Comparação de Modelos Aninhados ##
##-----##
> anova(res.model,completo)
```


Analysis of Variance Table

###

Model 1: $y \sim x_1$ (Sob H_0)

Model 2: $y \sim x_1 + x_A + x_B + x_1:x_A + x_1:x_B$ (Sob H_1)

Res.Df	RSS	Df	Sum-of-Sq	F	Pr(>F)
1	34	1970.57			
2	30	462.15	4	1508.4	24.48 4.458e-09 ***

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

Resp: Os resultados apresentados acima, mostram que há evidências suficientes para rejeitar a hipótese nula. Ou seja, o modelo completo (sob H_1), é melhor que o reduzido.

- (d_2) Ainda sobre o item (d_1), qual é a variação na soma dos quadrados da regressão, e dos erros, resultante em decorrência da inclusão de tais variáveis no modelo?

Resp: Observem que, no modelo com apenas uma variável independente, digamos, x_1 , a soma dos quadrados da regressão foi de $SQReg(x_1) = 3424,40$ e a soma dos quadrados dos erros foi de $SQE(x_1) = 1970,6$. Entretanto, depois que foram adicionados ao modelo os fatores de mencionados no item (d_1), as quantidades anteriores passaram a ser iguais a $SQReg(x_1, x_A, x_B, x_2, x_3) = 4932,80$ e $SQE(x_1, x_A, x_B, x_2, x_3) = 462,10$. Ou seja, a $SQReg$ aumentou em $\Delta_{Reg} = 4932,80 - 1970,6 = 2962,20$ unidades, e a SQE diminuiu em $\Delta_{Erro} = 1970,6 - 462,10 = 1508,5$ unidades.

##-----##

Modelo reduzido (sob H_0)

##-----##

> anova(res.model)

Analysis of Variance Table

###

Response: y

Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	3424.4	3424.4	59.085 6.155e-09 ***
Residuals	34	1970.6	58.0	

##

##-----##

Modelo completo (sob H_1)

##-----##

> anova(completo)

Analysis of Variance Table

Response: y

Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	3424.4	3424.4	222.2946 2.059e-15 ***
xA	1	803.8	803.8	52.1784 4.857e-08 ***
xB	1	1.2	1.2	0.0772 0.7831
x1:xA	1	375.0	375.0	24.3430 2.808e-05 ***
x1:xB	1	328.4	328.4	21.3194 6.850e-05 ***
Residuals	30	462.1	15.4	

- (d_3) Suponha que estejamos interessados em incluir o fator x_{Ai} no modelo II. Formule as hipóteses de interesse, e conclua quanto a viabilidade ou não da inclusão de tal fator de risco.

Resp: Se o interesse for de incluir o fator x_{Ai} no modelo II, nossas hipóteses de interesse são:

H_0 : $\beta_2 = 0$ modelo reduzido (ou modelo II)

H_1 : $\beta_2 \neq 0$ modelo aumentado $y_i^{II} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{Ai} + \beta_3 x_{Bi} + \beta_4 x_{1i} x_{Bi} + \epsilon_i^{II}$

Os resultados apresentados abaixo mostram que há evidências suficientes para rejeitar a hipótese nula. Ou seja, o modelo aumentado, com x_{Ai} deve ser escolhido.

```
##-----##
## Comparação de Modelos Aninhados ##
##-----##
>anova(res.model, augment.model)
Analysis of Variance Table

Model 1: y ~ x1 + xB + x1:xB
Model 2: y ~ x1 + xA + xB + x1:xA + x1:xB
Res.Df    RSS      Df Sum-of-Sq    F      Pr(>F)
1      32    1723.75
2      30     462.15   2     1261.6   40.948 2.658e-09 ***
```