Forma matricial do MRLS

Prof. Juvêncio Santos Nobre

Departamento de Estatística e Matemática Aplicada

Universidade Federal do Ceará-Brasil

http://www.dema.ufc.br/∼juvencio

DEMA-UFC

Capital do **Ceará**, setembro de 2022

Conteúdo

1 Forma matricial do MRLS

2 Método de Mínimos Quadrados

MRLS

Considere o MRLS com forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ i = 1, \ldots, n. \tag{1}$$

organizando lexicograficamente, temos que

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + e_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + e_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_n + e_n. \end{aligned}$$

MRLS

Perceba que podemos reescrever a estrutura do modelo da seguinte forma

$$\begin{aligned}
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \\
&= \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}. \qquad (2)
\end{aligned}$$

MRLS

Denotando

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz $(n \times 2)$ de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

- $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$: vetor (2×1) de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ da fonte de variação/erro de medida.
- Então, podemos reescrever o MRLS (1) da seguinte forma

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}.$$

MRLS

Denotando

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz $(n \times 2)$ de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

- $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$: vetor (2×1) de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ da fonte de variação/erro de medida.
- Então, podemos reescrever o MRLS (1) da seguinte forma

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}.$$

MRLS

Denotando

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz $(n \times 2)$ de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

- $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$: vetor (2×1) de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ da fonte de variação/erro de medida.
- Então, podemos reescrever o MRLS (1) da seguinte forma

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}.$$

MRLS

Denotando

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz $(n \times 2)$ de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

- $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$: vetor (2×1) de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ da fonte de variação/erro de medida.
- Então, podemos reescrever o MRLS (1) da seguinte forma

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}.$$

MRLS

Denotando

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz $(n \times 2)$ de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

- $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$: vetor (2×1) de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ da fonte de variação/erro de medida.
- Então, podemos reescrever o MRLS (1) da seguinte forma

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}.$$

MRLS

- Em geral, para ajustar o MRLS assume-se que

$$\boldsymbol{e} \sim (\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n).$$

- Tal suposição implica que $e_1, \ldots, e_n$ representa uma sequência de variáveis aleatórias de média zero, homoscedásticos com variância σ^2 e não-correlacionados.
- Para se realizar inferências de segunda ordem (construção de intervalos de confiança e testes de hipóteses, por exemplo) é comum assumir normalidade, i.e.,

$$\boldsymbol{e} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n).$$

MRLS

- Em geral, para ajustar o MRLS assume-se que

$$\boldsymbol{e} \sim (\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

- Tal suposição implica que $e_1, \ldots, e_n$ representa uma sequência de variáveis aleatórias de média zero, homoscedásticos com variância σ^2 e não-correlacionados.

- Para se realizar inferências de segunda ordem (construção de intervalos de confiança e testes de hipóteses, por exemplo) é comum assumir normalidade, i.e.,

$$\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

MRLS

- Em geral, para ajustar o MRLS assume-se que

$$\boldsymbol{e} \sim (\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

- Tal suposição implica que $e_1, \ldots, e_n$ representa uma sequência de variáveis aleatórias de média zero, homoscedásticos com variância σ^2 e não-correlacionados.
- Para se realizar inferências de segunda ordem (construção de intervalos de confiança e testes de hipóteses, por exemplo) é comum assumir normalidade, i.e.,

$$\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

Exercícios (entregar próxima aula)

Exercício 1: Faça um breve ensaio sobre distribuição normal multivariada e suas propriedades apresentadas de forma matricial. Em especial, apresente as 3 definições equivalentes da distribuição normal multivariada (algumas sem a necessidade da fdp).

Exercício 2: Usando um software de sua preferência, plote a função densidade de probabilidade e a respectiva curva de nível (mapa de contorno) da $\mathcal{N}_2(0, 0, 1, 1, \rho)$, para $\rho = 0, \pm 0.1, \pm 0.5, \pm 0.9$. Interprete os gráficos.

Exercícios (entregar próxima aula)

Exercício 3: Faça um breve ensaio sobre distribuição t-Student multivariada e suas propriedades apresentadas de forma matricial.

Exercício 4: Usando um software de sua preferência, plote a função densidade de probabilidade e a respectiva curva de nível (mapa de contorno) da $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, para $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \rho \boldsymbol{I}_2$, $\rho = 0, \pm 0.1, \pm 0.5, \pm 0.9$ e $\nu = 1, 2, 10, 30$ e 100. Interprete os gráficos.

Exercício 5: Apresente um resumo sobre:

i) Distribuições χ^2 e F não centrais.

ii) Distribuições de formas lineares e quadráticas (sob suposição de normalidade).

Método de Mínimos Quadrados

- Sob as suposições usuais associadas ao MRLS é possível obter o Estimador de Mínimos Quadrados (EMQ) de $\boldsymbol{\beta}$ através da minimização da seguinte função objetivo

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

- Lembrando que qualquer soma de quadrados pode ser reescrita como (Lista 0) uma forma quadrática

$$\sum_{i=1}^{n} a_i^2 = \boldsymbol{a}^\top \boldsymbol{a} = \|\boldsymbol{a}\|^2$$

com $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$.

- Então podemos reescrever (3) como

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = (\boldsymbol{y} - \boldsymbol{X}\beta)^\top (\boldsymbol{y} - \boldsymbol{X}\beta) = \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2,$$

i.e., $Q(\beta)$ é uma forma quadrática.

Método de Mínimos Quadrados

- Sob as suposições usuais associadas ao MRLS é possível obter o Estimador de Mínimos Quadrados (EMQ) de $\boldsymbol{\beta}$ através da minimização da seguinte função objetivo

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

- Lembrando que qualquer soma de quadrados pode ser reescrita como (Lista 0) uma forma quadrática

$$\sum_{i=1}^{n} a_i^2 = \boldsymbol{a}^\top \boldsymbol{a} = \|\boldsymbol{a}\|^2$$

com $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$.

- Então podemos reescrever (3) como

$$Q(\beta) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = (\boldsymbol{y} - \boldsymbol{X}\beta)^\top(\boldsymbol{y} - \boldsymbol{X}\beta) = \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2,$$

i.e., $Q(\beta)$ é uma forma quadrática.

Método de Mínimos Quadrados

- Sob as suposições usuais associadas ao MRLS é possível obter o Estimador de Mínimos Quadrados (EMQ) de $\boldsymbol{\beta}$ através da minimização da seguinte função objetivo

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

- Lembrando que qualquer soma de quadrados pode ser reescrita como (Lista 0) uma forma quadrática

$$\sum_{i=1}^{n} a_i^2 = \boldsymbol{a}^\top \boldsymbol{a} = \|\boldsymbol{a}\|^2$$

com $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$.

- Então podemos reescrever (3) como

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2,$$

i.e., $Q(\boldsymbol{\beta})$ é uma forma quadrática.

Método de Mínimos Quadrados

Note que

$$\begin{aligned} Q(\beta) &= (\boldsymbol{y} - \boldsymbol{X}\beta)^\top(\boldsymbol{y} - \boldsymbol{X}\beta) \\ &= \boldsymbol{y}^\top\boldsymbol{y} - \boldsymbol{y}^\top\boldsymbol{X}\beta - \beta^\top\boldsymbol{X}^\top\boldsymbol{y} + \beta^\top\boldsymbol{X}^\top\boldsymbol{X}\beta. \end{aligned}$$

Dado que $\boldsymbol{y}^\top\boldsymbol{X}\beta = \beta^\top\boldsymbol{X}^\top\boldsymbol{y}$ (Por que isso é verdade?), então a função objetivo se simplifica a

$$Q(\beta) = \boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top\boldsymbol{X}\beta + \beta^\top(\boldsymbol{X}^\top\boldsymbol{X})\beta.$$

Resultados - Q15 Lista 0

Lembrete: Considere $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ e $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$ vetores reais de dimensão $n \times 1$ e $\boldsymbol{A}$ uma matriz quadrada de ordem n. Então,

i) $\dfrac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}} = \dfrac{\partial \boldsymbol{x}^\top \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}$.

ii) $\dfrac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^\top$.

iii) $\dfrac{\partial \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}^\top)\boldsymbol{x}$.

Equações normais

Logo, dado que a matriz $\boldsymbol{X}^\top\boldsymbol{X}$ é simétrica, tem-se que

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^\top\boldsymbol{y} + 2(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{\beta},$$

de forma que o sistema de equações normais (equação de estimação) é dado por

$$\left.\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0},$$

implicando que

$$-2\boldsymbol{X}^\top\boldsymbol{y} + 2(\boldsymbol{X}^\top\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \mathbf{0} \Leftrightarrow (\boldsymbol{X}^\top\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{y}.$$

Considerando que $\boldsymbol{X}$ é de **posto completo**, implicando que $\boldsymbol{X}^\top\boldsymbol{X}$ é não singular, tem-se que a solução do sistema de equações, e respectivo candidato a EMQ de $\boldsymbol{\beta}$, é

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}. \tag{4}$$

Matriz de segundas derivadas

A matriz de segundas derivadas, com relação ao vetor de parâmetros, da função objetivo é dada por

$$\frac{\partial^2 Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = 2(\boldsymbol{X}^\top\boldsymbol{X})$$

que é uma matriz positiva definida (exercício - fazer quadro). Portanto,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

corresponde ao estimador de Mínimos Quadrados de $\boldsymbol{\beta}$.

Exercícios (entregar próxima aula)

Exercício 6: Discuta sob que condição a matriz de especificação $\boldsymbol{X}$ associada ao MRLS (2) é de posto completo.

Exercício 7: Mostre que $\boldsymbol{A}^\top\boldsymbol{A}$ é não-singular, **se e somente se**, $\boldsymbol{A}$ é de posto completo, sendo $\boldsymbol{A} \in \mathbb{R}^{n\times p}$, i.e., uma matriz de dimensão $n \times p$.

Exercício 8: Se $\boldsymbol{A} \in \mathbb{R}^{n\times p}$, mostre que $\boldsymbol{A}^\top\boldsymbol{A}$ é positiva definida, i.e., $\forall \boldsymbol{x} \in \mathbb{R}^p$

$$\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} \geq 0.$$

Exercício 9: Mostre que

$$Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \left\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 + \left\|\boldsymbol{X}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right)\right\|^2.$$

Observações

Obs1. Note que

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{A_x y},$$

i.e., $\widehat{\boldsymbol{\beta}}$ é uma transformação linear do vetor de variáveis respostas, implicando que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são combinações lineares de $\boldsymbol{y}$ como já provamos.

Obs2. Perceba que

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} n & n\bar{x}_n \\ n\bar{x}_n & \sum_{i=1}^{n} x_i^2 \end{pmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} n\bar{y}_n \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$\text{e} \quad (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \frac{1}{S_{xx}} \begin{pmatrix} \sum_{i=1}^{n} x_i^2/n & -\bar{x}_n \\ -\bar{x}_n & 1 \end{pmatrix}.$$

Observações

Portanto,

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} = \frac{1}{S_{xx}} \begin{pmatrix} \sum_{i=1}^{n} x_i^2/n & -\bar{x}_n \\ -\bar{x}_n & 1 \end{pmatrix} \begin{pmatrix} n\bar{y}_n \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}.$$

Lembrando que

$$\sum_{i=1}^{n} x_i^2 \quad = \quad S_{xx} + n\bar{x}_n^2 \ \text{ e } \ \sum_{i=1}^{n} x_i y_i = S_{xy} + n\bar{x}_n\bar{y}_n,$$

então

$$(\sum_{i=1}^{n} x_i^2/n)n\bar{y}_n - \bar{x}_n \sum_{i=1}^{n} x_i y_i = S_{xx}\bar{y}_n - \bar{x}_n S_{xy}$$

e

$$-\bar{x}_n n\bar{y}_n + \sum_{i=1}^{n} x_i y_i = S_{xy}.$$

Observações

Logo, pode-se concluir que

$$\therefore (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} = \frac{1}{S_{xx}} \begin{pmatrix} S_{xx}\bar{y}_n - \bar{x}_n S_{xy} \\ S_{xy} \end{pmatrix} = \begin{pmatrix} \bar{y}_n - \bar{x}_n S_{xy}/S_{xx} \\ S_{xy}/S_{xx} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix}$$

Obs3. Sob o MRLS tem-se que

$$\boldsymbol{y} \sim (\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n),$$

em que $\boldsymbol{I}_n$ denota a matriz identidade de ordem, n. Adicionalmente, sob a suposição de normalidade

$$\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n).$$

Observações

Obs4. Tem-se que

$$\begin{aligned}\mathbb{E}[\widehat{\boldsymbol{\beta}}] &= \mathbb{E}[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}] = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbb{E}[\boldsymbol{y}] \\ &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta},\end{aligned}$$

i.e., $\widehat{\boldsymbol{\beta}}$ é um estimador não viesado de $\boldsymbol{\beta}$, como já havíamos provado de forma marginal. Por outro lado, a matriz de variâncias-covariâncias do EMQ de $\boldsymbol{\beta}$ é dada por

$$\begin{aligned}\mathrm{Var}[\widehat{\boldsymbol{\beta}}] &= \mathrm{Var}[\boldsymbol{A_x}\boldsymbol{y}] = \boldsymbol{A_x}\mathrm{Var}[\boldsymbol{y}]\boldsymbol{A_x}^\top = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\sigma^2\boldsymbol{I}_n[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top]^\top \\ &= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\ &= \frac{\sigma^2}{S_{xx}}\begin{pmatrix}\sum_{i=1}^n x_i^2/n & -\bar{x}_n \\ -\bar{x}_n & 1\end{pmatrix} = \sigma^2\begin{pmatrix}\frac{1}{n}+\frac{\bar{x}_n^2}{S_{xx}} & -\frac{\bar{x}_n}{S_{xx}} \\ -\frac{\bar{x}_n}{S_{xx}} & \frac{1}{S_{xx}}\end{pmatrix},\end{aligned}$$

dado que $\sum_{i=1}^n x_i^2 = S_{xx} + n\bar{x}_n^2$.

Observações

Obs5. Sob a suposição de normalidade, tem-se que

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_2 \left[\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right].$$

Obs6. (Teorema de Gauss-Markov) Sob as suposições do MRLS, $\widehat{\boldsymbol{\beta}}$ é o BLUE de $\boldsymbol{\beta}$.

A versão **multivariada** do teorema de Gauss-Markov será demonstrada nas próximas aulas.

Obs7. O vetor de valores preditos é dado por

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{H}\boldsymbol{y},$$

que é uma transformação linear das observações. A matriz $\boldsymbol{H} := \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ é uma matriz simétrica e idempotente, que representa a matriz de projeção ortogonal do subespaço gerado pelas colunas de $\boldsymbol{X}$ ($\mathcal{C}(\boldsymbol{X})$).

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial\widehat{\boldsymbol{y}}}{\partial\boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial\widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial \widehat{\boldsymbol{y}}}{\partial \boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial \widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial\widehat{\boldsymbol{y}}}{\partial\boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial\widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial\widehat{\boldsymbol{y}}}{\partial\boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial\widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial\widehat{\boldsymbol{y}}}{\partial\boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial\widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial \widehat{\boldsymbol{y}}}{\partial \boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial \widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial\widehat{\boldsymbol{y}}}{\partial\boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial\widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Obs7. (cont.) A matriz $\boldsymbol{H}$ também é comumente denominada de matriz Hat (chapéu em Inglês), denominação dada pelo J.W. Tukey, se deve ao fato de $\boldsymbol{H}$ ser base da transformação linear do vetor dos valores observados no vetor de valores ajustados, $\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, ou seja, $\boldsymbol{H}$ "coloca" o chapéu no vetor $\boldsymbol{y}$. Ela também é comumente chamada de matriz de predição. Abaixo, seguem algumas propriedades e características da matriz $\boldsymbol{H}$:

- $\boldsymbol{H}^2 = \boldsymbol{H}$ e $\boldsymbol{H}^\top = \boldsymbol{H}$.
- $\boldsymbol{H}\boldsymbol{X} = \boldsymbol{X}$ e $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$.
- $\mathrm{Var}[\widehat{\boldsymbol{y}}] = \mathrm{Var}[\boldsymbol{H}\boldsymbol{y}] = \sigma^2\boldsymbol{H}\boldsymbol{H}^\top = \sigma^2\boldsymbol{H}^2 = \sigma^2\boldsymbol{H}$.
- $\mathrm{Var}[\widehat{y}_i] = \sigma^2 h_{ii}$, ou seja quanto maior for h_{ii} maior é a imprecisão na predição de y_i.
- $\dfrac{\partial \widehat{\boldsymbol{y}}}{\partial \boldsymbol{y}} = \boldsymbol{H}$.
- $\dfrac{\partial \widehat{y}_i}{\partial y_i} = h_{ii}$ que reflete a taxa de variação de $\widehat{y}_i$ quando y_i varia de forma infinitesimal, ou seja, h_{ii} reflete a influência de y_i no seu respectivo valor predito.
- h_{ii} é chamada de alavancagem (*high leverage*) da i-ésima observação.

Observações

Dado que $\boldsymbol{H}$ é idempotente, obtemos que $\forall i = 1, \ldots, n$

$$h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$$
$$\Rightarrow h_{ii}(1 - h_{ii}) = \sum_{j \neq i} h_{ij}^2 \geq 0,$$

implicando que $h_{ii} \in [0, 1]$, $\forall i = 1, \ldots, n$.

No caso do MRLS é possível mostrar que (exercício Lista I)

$$h_{ij} \quad = \quad \left\{ \frac{1}{n} + \frac{(x_i - \bar{x}_n)(x_j - \bar{x}_n)}{S_{xx}} \right\}, \forall i, j = 1, \ldots, n.$$

Outras propriedades e extensões serão apresentadas em formato de exercício.

Observações

Obs8. O vetor de resíduos ordinários é dado por

$$\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{y} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y},$$

que é uma transformação linear de $\boldsymbol{y}$. Dado que $(\boldsymbol{I}_n - \boldsymbol{H})$ é simétrica e idempotente, então sob a suposição de normalidade tem-se

$$\widehat{\boldsymbol{e}} \sim \mathcal{N}_n[\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})],$$

implicando que os resíduos podem não ser homoscedásticos. Ademais, como $(\boldsymbol{I}_n - \boldsymbol{H})$ é idempotente, então $\mathrm{posto}(\boldsymbol{I}_n - \boldsymbol{H}) = \mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{H}) = n - 2$, implicando que a mesma é singular. Logo,

$$\widehat{\boldsymbol{e}} \sim \mathcal{N}_n[\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})],$$

é normal multivariada singular.

Observações

Obs9. Sob normalidade, tem-se que

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_2 \left[\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right],$$

implicando que $\forall \boldsymbol{c} \in \mathbb{R}^2 \neq \boldsymbol{0}$

$$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} \sim \mathcal{N} \left[\boldsymbol{c}^\top \boldsymbol{\beta}, \sigma^2 \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}\right].$$

Observações

Casos particulares importantes:

i) Se $\boldsymbol{c} = (1, 0)^\top$, então

$$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 \sim \mathcal{N}\left[\beta_0, \sigma^2 \left\{\frac{1}{n} + \frac{\bar{x}_n^2}{S_{xx}}\right\}\right].$$

ii) Se $\boldsymbol{c} = (0, 1)^\top$, então

$$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} = \widehat{\beta}_1 \sim \mathcal{N}\left[\beta_1, \frac{\sigma^2}{S_{xx}}\right].$$

iii) Se $\boldsymbol{c} = (1, x_i)^\top = \boldsymbol{x}_i$, então

$$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 + \widehat{\beta}_1 x_i = \widehat{\mu}_{x_i} \sim \mathcal{N}\left[\mu_{x_i}, \sigma^2 \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i\right],$$

em que $\mu_{x_i} = \beta_0 + \beta_1 x_i$ e $\boldsymbol{x}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_i = \left\{\frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}$.

Observações

Obs10. Lembrando que $(\boldsymbol{I}_n - \boldsymbol{H})$ é uma matriz simétrica e idempotente, então podemos reescrever SQRes como

$$\begin{aligned}
\mathrm{SQRes} &= \sum_{i=1}^{n} \widehat{e}_i^2 = \|\widehat{\boldsymbol{e}}\|^2 = \widehat{\boldsymbol{e}}^\top \widehat{\boldsymbol{e}} \\
&= ((\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y})^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} = \boldsymbol{y}^\top (\boldsymbol{I}_n - \boldsymbol{H})^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} \\
&= \boldsymbol{y}^\top (\boldsymbol{I}_n - \boldsymbol{H})(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} \\
&= \boldsymbol{y}^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y},
\end{aligned}$$

que é uma forma quadrática.

Dado que para $\mathrm{posto}(\boldsymbol{I} - \boldsymbol{H}) = \mathrm{tr}(\boldsymbol{I} - \boldsymbol{H}) = n - 2$ que corresponde aos g.l. da respectiva forma quadrática. Logo, um ENV (na verdade o MINQUE) de σ^2 é dado por

$$\widehat{\sigma}^2 = \mathrm{QMRes} := \frac{\mathrm{SQRes}}{n-2} = \frac{\boldsymbol{y}^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}}{n-2}.$$

Observações

Obs11. Note que

$$\mathrm{SQT} = \sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2.$$

Mas, basta relembrar a Questão 21 da Lista 0

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\mathbf{1}_n^\top \boldsymbol{y} = \frac{1}{n}\boldsymbol{y}^\top \mathbf{1}_n,$$

em que $\mathbf{1}_n$ representa um vetor $(n \times 1)$ com todos elementos iguais a 1. Por outro lado, dado que $\sum_{i=1}^{n} y_i^2 = \boldsymbol{y}^\top \boldsymbol{y}$, então podemos reescrever SQT como

$$\begin{aligned} \mathrm{SQT} &= \boldsymbol{y}^\top\boldsymbol{y} - n\bar{y}_n\bar{y}_n = \boldsymbol{y}^\top\boldsymbol{y} - n\left(\frac{1}{n}\boldsymbol{y}^\top\mathbf{1}_n\right)\left(\frac{1}{n}\mathbf{1}_n^\top\boldsymbol{y}\right) \\ &= \boldsymbol{y}^\top\boldsymbol{y} - \boldsymbol{y}^\top\frac{\mathbf{1}_n\mathbf{1}_n^\top}{n}\boldsymbol{y} = \boldsymbol{y}^\top\left(\boldsymbol{I}_n - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y}, \end{aligned}$$

em que $\boldsymbol{J}_n = \mathbf{1}_n\mathbf{1}_n^\top$ representa uma matriz quadrada de dimensão n com todos os elementos iguais a 1.

Observações

Portanto, pelas observações 10 e 11, temos que

$$\begin{aligned}\text{SQReg} &= \text{SQT} - \text{SQRes} \\ &= \boldsymbol{y}^\top \left(\boldsymbol{I}_n - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y} - \boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} \\ &= \boldsymbol{y}^\top \left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y}.\end{aligned}$$

Exercícios (entregar próxima aula)

Exercício 10: Considere o MRLS e as respectivas somas de quadrados reescritas como formas quadráticas. Mostre que as matrizes núcleo são idempotentes, i.e., que $(\boldsymbol{I}_n - \boldsymbol{H})$, $(\boldsymbol{I}_n - \boldsymbol{J}_n/n)$ e $(\boldsymbol{H} - \boldsymbol{J}_n/n)$ são idempotentes. Adicionalmente, mostre que

$$
\begin{aligned}
\mathrm{posto}(\boldsymbol{I}_n - \boldsymbol{H}) &= n - 2 \\
\mathrm{posto}(\boldsymbol{I}_n - \boldsymbol{J}_n/n) &= n - 1 \\
\mathrm{posto}(\boldsymbol{H} - \boldsymbol{J}_n/n) &= 1.
\end{aligned}
$$

Exercício 11: Mostre que se $\boldsymbol{X} = \mathbf{1}_n$, i.e., se não existe regressão, então $\boldsymbol{H} = \boldsymbol{J}_n/n$, implicando que $\mathrm{SQReg} = 0$.

Exercício 12: Expresse a tabela de ANOVA do MRLS em termos matriciais.

Exercício 13: Considere o MRLS de **intercepto nulo** e todas as suas pressuposições. Expresse todas as quantidades de interesse usando a notação matricial.

Observações

Obs12. Sob a suposição de normalidade, temos que $\widehat{\boldsymbol{\beta}} \perp \mathrm{SQRes}$ e que

$$\frac{\mathrm{SQRes}}{\sigma^2} \sim \chi^2_{(n-2)}.$$

Iremos demonstrar isto posteriormente usando resultados associados a distribuições de transformações lineares e de formas quadráticas associadas a normal multivariada. 😎

Obs13. Todos os testes de hipóteses considerados anteriormente são casos particulares da classe abaixo

$$\mathcal{H}_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = a \text{ vs } \mathcal{H}_1 : \boldsymbol{c}^\top \boldsymbol{\beta} \neq a, \tag{5}$$

sendo $a \in \mathbb{R}$ especificado.

Observações

Para este fim, podemos utilizar como base para construção da estatística de teste a seguinte quantidade pivotal (válida sob a suposição de normalidade):

$$\frac{\dfrac{\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}-\boldsymbol{c}^\top\boldsymbol{\beta}}{\sqrt{\sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}}=\frac{\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}-\boldsymbol{c}^\top\boldsymbol{\beta}}{\sqrt{\mathrm{QMRes}\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}}}\sim t_{(n-2)}. \tag{6}$$

Para testar (5) podemos usar a seguinte estatística de teste

$$t_0=\frac{\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}-a}{\sqrt{\mathrm{QMRes}\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}}}.$$

Sob $\mathcal{H}_0:\boldsymbol{c}^\top\boldsymbol{\beta}=a$, tem-se que $t\sim t_{(n-2)}$. De forma que rejeitamos $\mathcal{H}_0$ ao nível α se $|t_0|>t_{(1-\alpha/2,n-2)}$. O valor-p associado é dado por

$$\text{Valor-p}=2\min\{\mathbb{P}[t_{(n-2)}>t_0],\mathbb{P}[t_{(n-2)}<t_0]\}.$$

Exercício - Entregar próxima aula

Exercício 14: Baseado na estatística de teste apresentada anteriormente, especifique a região crítica e como obter o valor-p nos seguintes casos:

i) $\mathcal{H}_0 : \boldsymbol{c}^\top\boldsymbol{\beta} = a \text{ vs } \mathcal{H}_1 : \boldsymbol{c}^\top\boldsymbol{\beta} > a$, a especificado.

ii) $\mathcal{H}_0 : \boldsymbol{c}^\top\boldsymbol{\beta} = a \text{ vs } \mathcal{H}_1 : \boldsymbol{c}^\top\boldsymbol{\beta} < a$, a especificado.

Observações

Casos particulares importantes:

i) Se $\boldsymbol{c} = (1,0)^\top$, então $\mathcal{H}_0 : \beta_0 = a$ vs $\mathcal{H}_1 : \beta_0 \neq a$ e

$$t_0 = \frac{\widehat{\beta}_0 - a}{\sqrt{\mathrm{QMRes}\left\{\frac{1}{n} + \frac{\bar{x}_n^2}{S_{xx}}\right\}}}$$

ii) Se $\boldsymbol{c} = (0,1)^\top$, então $\mathcal{H}_0 : \beta_1 = a$ vs $\mathcal{H}_1 : \beta_1 \neq a$ e

$$t_0 = \frac{\widehat{\beta}_1 - a}{\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}}$$

iii) Se $\boldsymbol{c} = (1,x_i)^\top = \boldsymbol{x}_i$, então $\mathcal{H}_0 : \mu_{x_i} = a$ vs $\mathcal{H}_1 : \mu_{x_i} \neq a$ e

$$t_0 = \frac{\widehat{\mu}_{x_i} - a}{\sqrt{\mathrm{QMRes}\left\{\frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}}}.$$

Observações

Para construção do intervalo de confiança para $\boldsymbol{c}^\top\boldsymbol{\beta}$ podemos usar (6). Logo, pelo método da quantidade pivotal, temos que o melhor intervalo de confiança para $\boldsymbol{c}^\top\boldsymbol{\beta}$ com nível $(1-\alpha)$ é dado por

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{c}^\top\boldsymbol{\beta}) = \left[\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}} \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\,\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}}\right]$$

Casos particulares importantes:

i) Se $\boldsymbol{c} = (1,0)^\top$, então

$$\mathrm{IC}_{1-\alpha}(\beta_0) = \left[\widehat{\beta}_0 \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

ii) Se $\boldsymbol{c} = (0,1)^\top$, então

$$\mathrm{IC}_{1-\alpha}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-2)}\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}\right],$$

Observações

iii) Se $\boldsymbol{c} = (1, x_i)^\top = \boldsymbol{x}_i$, então

$$\begin{aligned} \mathrm{IC}_{1-\alpha}(\mu_{x_i}) &= \left[\widehat{y}_i \pm t_{(1-\alpha/2,n-2)}\sqrt{h_{ii}\mathrm{QMRes}}\right] \\ &= \left[\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}(\boldsymbol{x}_i^\top\{\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i\}}\right], \end{aligned}$$

em que $h_{ii} = \left\{\frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}$.

iv) Temos também que

$$\begin{aligned} \mathrm{IP}_{1-\alpha}(Y_{x_i}) &= \left[\widehat{y}_i \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\{1 + h_{ii}\}}\right] \\ &= \left[\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left\{1 + \boldsymbol{x}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_i\right\}}\right]. \end{aligned}$$