

Estimação

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 25 de março de 2026



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação
- 6 Referências bibliográficas



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação
- 6 Referências bibliográficas



Modelo de regressão linear

Suponham que Y_1, Y_2, \dots, Y_n tais que

$$Y_l = \mathbf{x}_l^T \boldsymbol{\beta} + \varepsilon_l, \quad (1)$$



Modelo de regressão linear

Suponham que Y_1, Y_2, \dots, Y_n tais que

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad (1)$$

em que $\mathbf{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \dots, x_{\ell p})^\top$ é conhecido, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ são variáveis aleatórias independentes e com a mesma variância σ^2 , também desconhecida, a ser estimada.



Modelo de regressão linear

Suponham que Y_1, Y_2, \dots, Y_n tais que

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad (1)$$

em que $\mathbf{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \dots, x_{\ell p})^\top$ é conhecido, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos a serem estimados, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ são variáveis aleatórias independentes e com a mesma variância σ^2 , também desconhecida, a ser estimada.



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros,



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros,
 - têm média zero;



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros,
 - têm média zero;
 - variância constante;



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros,
 - têm média zero;
 - variância constante;
 - e são não correlacionados.



Suposições

Em resumo, as **suposições** de um modelo de regressão linear são de que:

- A relação entre as variáveis resposta e as preditoras é linear;
- Os erros,
 - têm média zero;
 - variância constante;
 - e são não correlacionados.



Forma matricial

A Equação (1), que nós definimos como **modelo de regressão linear** (MRL), podendo ser escrita de forma matricial, da seguinte forma:

$$Y = X\beta + \varepsilon, \quad (2)$$



Forma matricial

A Equação (1), que nós definimos como **modelo de regressão linear** (MRL), podendo ser escrita de forma matricial, da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, sendo $\mathbf{0}$ o vetor nulo de dimensão n , \mathbf{I}_n a matriz identidade de ordem n e $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$



Forma matricial

A Equação (1), que nós definimos como **modelo de regressão linear** (MRL), podendo ser escrita de forma matricial, da seguinte forma:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, sendo $\mathbf{0}$ o vetor nulo de dimensão n , \mathbf{I}_n a matriz identidade de ordem n e $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$



Forma matricial

ou

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

A matriz \mathbf{X} tem n linhas (tamanho da amostra) e p colunas (número de variáveis preditoras), de posto p ($n \geq p$), \mathbf{X} é chamada de matriz de planejamento.



Forma matricial

ou

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}.$$

A matriz \mathbf{X} tem n linhas (tamanho da amostra) e p colunas (número de variáveis preditoras), de posto p ($n \geq p$), \mathbf{X} é chamada de matriz de planejamento.



Modelo de regressão normal linear

Para que procedimentos inferenciais possam ser aplicados, nós precisamos assumir que, $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$ e $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, em (1) e (2), respectivamente. O MRL passa a ser denominado de **modelo de regressão normal linear** (MNL).



Modelo de regressão normal linear

Para que procedimentos inferenciais possam ser aplicados, nós precisamos assumir que, $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$ e $\varepsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, em (1) e (2), respectivamente. O MRL passa a ser denominado de **modelo de regressão normal linear** (MNL).



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados**
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação
- 6 Referências bibliográficas



Método dos mínimos quadrados

Definição

Os estimadores de mínimos quadrados (EMQ) obtém estimativas para os parâmetros desconhecidos **minimizando** a soma de quadrado dos erros.



Método dos mínimos quadrados

Definição

Os estimadores de mínimos quadrados (EMQ) obtém estimativas para os parâmetros desconhecidos **minimizando** a soma de quadrado dos erros.



Método dos mínimos quadrados

Se $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^\top$ é o EMQ de β , ele deve ser tal que minimize

$$Q = \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})]^2. \quad (3)$$

Método dos mínimos quadrados

Se $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^\top$ é o EMQ de β , ele deve ser tal que minimize

$$Q = \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})]^2. \quad (3)$$

Método dos mínimos quadrados

Derivando (3), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial Q}{\partial \beta_m} = -2 \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (4)$$



Método dos mínimos quadrados

Derivando (3), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial Q}{\partial \beta_m} = -2 \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (4)$$

Igualando (4) a zero,



Método dos mínimos quadrados

Derivando (3), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial Q}{\partial \beta_m} = -2 \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (4)$$

Igualando (4) a zero, nós temos que

$$\hat{\beta}_1 \sum_{\ell=1}^n x_{\ell 1} x_{\ell m} + \dots + \hat{\beta}_m \sum_{\ell=1}^n x_{\ell m}^2 + \dots + \hat{\beta}_p \sum_{\ell=1}^n x_{\ell p} x_{\ell m} = \sum_{\ell=1}^n x_{\ell m} Y_\ell. \quad (5)$$



Método dos mínimos quadrados

Derivando (3), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial Q}{\partial \beta_m} = -2 \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (4)$$

Igualando (4) a zero, nós temos que

$$\hat{\beta}_1 \sum_{\ell=1}^n x_{\ell 1} x_{\ell m} + \dots + \hat{\beta}_m \sum_{\ell=1}^n x_{\ell m}^2 + \dots + \hat{\beta}_p \sum_{\ell=1}^n x_{\ell p} x_{\ell m} = \sum_{\ell=1}^n x_{\ell m} Y_\ell. \quad (5)$$



Método dos mínimos quadrados

As p equações em (5) são chamadas de **equações normais**. E o **estimador de mínimos quadrados** é dado por:



Método dos mínimos quadrados

As p equações em (5) são chamadas de **equações normais**. E o **estimador de mínimos quadrados** é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (6)$$



Método dos mínimos quadrados

As p equações em (5) são chamadas de **equações normais**. E o **estimador de mínimos quadrados** é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (6)$$



Método dos mínimos quadrados

Uma alternativa para a obtenção de (6), é observar que (3) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta).$$



Método dos mínimos quadrados

Uma alternativa para a obtenção de (6), é observar que (3) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta).$$

e derivar Q em relação à β .



Método dos mínimos quadrados

Uma alternativa para a obtenção de (6), é observar que (3) pode ser escrito como:

$$Q = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta).$$

e derivar Q em relação à β .



Método dos mínimos quadrados

Antes, nós podemos observar:

$$\begin{aligned}Q &= (\mathbf{Y}^\top - \beta^\top \mathbf{X}^\top)(\mathbf{Y} - \mathbf{X}\beta) \\&= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\beta - \beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \\&= \mathbf{Y}^\top \mathbf{Y} - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta.\end{aligned}$$

Agora, derivando Q em relação à β , nós temos:

$$\frac{\partial Q}{\partial \beta} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\beta.$$



Método dos mínimos quadrados

Antes, nós podemos observar:

$$\begin{aligned}Q &= (\mathbf{Y}^\top - \beta^\top \mathbf{X}^\top)(\mathbf{Y} - \mathbf{X}\beta) \\&= \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\beta - \beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta \\&= \mathbf{Y}^\top \mathbf{Y} - 2\beta^\top \mathbf{X}^\top \mathbf{Y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta.\end{aligned}$$

Agora, derivando Q em relação à β , nós temos:

$$\frac{\partial Q}{\partial \beta} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\beta.$$



Método dos mínimos quadrados

Igualando a derivada ao vetor nulo, nós temos:

$$-2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{0}$$

$$\Rightarrow \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$$

$$\Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$



Método dos mínimos quadrados

Observação

O EMQ existirá se existir a inversa $(\mathbf{X}^T \mathbf{X})^{-1}$.

Método dos mínimos quadrados

Observação

O EMQ existirá se existir a inversa $(\mathbf{X}^T \mathbf{X})^{-1}$. E a matriz $(\mathbf{X}^T \mathbf{X})^{-1}$ existe se as variáveis preditoras forem linearmente independentes, isto é, se nenhuma coluna da matriz \mathbf{X} é uma combinação linear das outras.

Método dos mínimos quadrados

Observação

O EMQ existirá se existir a inversa $(\mathbf{X}^T \mathbf{X})^{-1}$. E a matriz $(\mathbf{X}^T \mathbf{X})^{-1}$ existe se as variáveis preditoras forem linearmente independentes, isto é, se nenhuma coluna da matriz \mathbf{X} é uma combinação linear das outras.



Propriedades do EMQ

A esperança de $\hat{\beta}$ é dada por

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}(\mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbb{E}(\mathbf{X}\beta + \varepsilon) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{X}\beta + \mathbb{E}(\varepsilon)] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\beta = \beta.\end{aligned}$$



Propriedades do EMQ

A esperança de $\hat{\beta}$ é dada por

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}(\mathbf{Y}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}(\mathbf{X}\beta + \varepsilon) \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top [\mathbf{X}\beta + \mathbb{E}(\varepsilon)] \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\beta = \beta.\end{aligned}$$



Propriedades do EMQ

E a variância de $\hat{\beta}$ é dada por

$$\begin{aligned}\text{Var}(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{X}\beta + \varepsilon) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\varepsilon) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 I_n \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.\end{aligned}$$



Propriedades do EMQ

E a variância de $\hat{\beta}$ é dada por

$$\begin{aligned}\text{Var}(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{X}\beta + \varepsilon) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\varepsilon) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I}_n \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.\end{aligned}$$



Propriedades do EMQ

Pelo Teorema de Gauss-Markov (Plackett, 1949), os EMQ

- são não viesados;



Propriedades do EMQ

Pelo Teorema de Gauss-Markov (Plackett, 1949), os EMQ

- são não viesados;
- e têm menor variância entre todos os estimadores lineares não viesados.



Propriedades do EMQ

Pelo Teorema de Gauss-Markov (Plackett, 1949), os EMQ

- são não viesados;
- e têm menor variância entre todos os estimadores lineares não viesados.



Valores ajustados

E assim, o valor ajustado (estimado) da ℓ -ésima observação é dado por



Valores ajustados

E assim, o valor ajustado (estimado) da l -ésima observação é dado por

$$\hat{Y}_l = \mathbf{x}_l^\top \hat{\boldsymbol{\beta}},$$

$$l = 1, 2, \dots, n.$$

Valores ajustados

E assim, o valor ajustado (estimado) da l -ésima observação é dado por

$$\hat{Y}_l = \mathbf{x}_l^\top \hat{\boldsymbol{\beta}},$$

$l = 1, 2, \dots, n$. De forma matricial,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{H}\mathbf{Y}.$$



Valores ajustados

E assim, o valor ajustado (estimado) da ℓ -ésima observação é dado por

$$\hat{Y}_\ell = \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}},$$

$\ell = 1, 2, \dots, n$. De forma matricial,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{H}\mathbf{Y}.$$

A matriz \mathbf{H} , $n \times n$, é chamada de matriz chapéu (*hat matrix*).



Valores ajustados

E assim, o valor ajustado (estimado) da ℓ -ésima observação é dado por

$$\hat{Y}_\ell = \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}},$$

$\ell = 1, 2, \dots, n$. De forma matricial,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{H}\mathbf{Y}.$$

A matriz \mathbf{H} , $n \times n$, é chamada de matriz chapéu (*hat matrix*).



Matriz chapéu

Propriedades da matriz H .

- H é simétrica e idempotente;



Matriz chapéu

Propriedades da matriz \mathbf{H} .

- \mathbf{H} é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})$ é simétrica e idempotente;



Matriz chapéu

Propriedades da matriz \mathbf{H} .

- \mathbf{H} é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})$ é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})\mathbf{H} = \mathbf{0}$.



Matriz chapéu

Propriedades da matriz \mathbf{H} .

- \mathbf{H} é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})$ é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})\mathbf{H} = \mathbf{0}$.

Observação: uma matriz \mathbf{M} é dita idempotente se e somente se $\mathbf{MM} = \mathbf{M}$.



Matriz chapéu

Propriedades da matriz \mathbf{H} .

- \mathbf{H} é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})$ é simétrica e idempotente;
- $(\mathbf{I}_n - \mathbf{H})\mathbf{H} = \mathbf{0}$.

Observação: uma matriz \mathbf{M} é dita idempotente se e somente se $\mathbf{MM} = \mathbf{M}$.



É a diferença entre o valor observado pelo correspondente valor ajustado.

$$\begin{aligned}e &= Y - \hat{Y} = Y - X\hat{\beta} \\ &= Y - HY = (I_n - H)Y,\end{aligned}$$

É a diferença entre o valor observado pelo correspondente valor ajustado.

$$\begin{aligned}\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{HY} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y},\end{aligned}$$

em que $\mathbf{e} = (e_1, e_2, \dots, e_n)^\top$ e $e_\ell = Y_\ell - \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}}$ é o ℓ -ésimo resíduo, $\ell = 1, 2, \dots, n$.

Resíduos

É a diferença entre o valor observado pelo correspondente valor ajustado.

$$\begin{aligned}\mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{Y} - \mathbf{HY} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y},\end{aligned}$$

em que $\mathbf{e} = (e_1, e_2, \dots, e_n)^\top$ e $e_\ell = Y_\ell - \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}}$ é o ℓ -ésimo resíduo, $\ell = 1, 2, \dots, n$.

A esperança de \mathbf{e} é dada por

$$\begin{aligned}\mathbb{E}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\mathbb{E}(\mathbf{Y}) = (\mathbf{I}_n - \mathbf{H})\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\mathbf{I}_n - \mathbf{H})[\mathbf{X}\boldsymbol{\beta} + \mathbb{E}(\boldsymbol{\varepsilon})] = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}.\end{aligned}$$

A esperança de \mathbf{e} é dada por

$$\begin{aligned}\mathbb{E}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\mathbb{E}(\mathbf{Y}) = (\mathbf{I}_n - \mathbf{H})\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\mathbf{I}_n - \mathbf{H})[\mathbf{X}\boldsymbol{\beta} + \mathbb{E}(\boldsymbol{\varepsilon})] = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}.\end{aligned}$$

E a variância de \mathbf{e} é dada por

$$\begin{aligned}\text{Var}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{Y})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\sigma^2 \mathbf{I}_n (\mathbf{I}_n - \mathbf{H})^\top \\ &= \sigma^2 (\mathbf{I}_n - \mathbf{H}).\end{aligned}$$

E a variância de \mathbf{e} é dada por

$$\begin{aligned}\text{Var}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{Y})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\sigma^2\mathbf{I}_n(\mathbf{I}_n - \mathbf{H})^\top \\ &= \sigma^2(\mathbf{I}_n - \mathbf{H}).\end{aligned}$$

Lembrando que $(\mathbf{I}_n - \mathbf{H})$ é idempotente e simétrica.

E a variância de \mathbf{e} é dada por

$$\begin{aligned}\text{Var}(\mathbf{e}) &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{Y})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\text{Var}(\boldsymbol{\varepsilon})(\mathbf{I}_n - \mathbf{H})^\top \\ &= (\mathbf{I}_n - \mathbf{H})\sigma^2\mathbf{I}_n(\mathbf{I}_n - \mathbf{H})^\top \\ &= \sigma^2(\mathbf{I}_n - \mathbf{H}).\end{aligned}$$

Lembrando que $(\mathbf{I}_n - \mathbf{H})$ é idempotente e simétrica.

Resíduos

Seja h_{ij} , o (i, j) -ésimo termo da matriz \mathbf{H} , $i, j = 1, 2, \dots, n$, então nós temos que:

- $\mathbb{E}(e_j) = 0$;

Resíduos

Seja h_{ij} , o (i, j) -ésimo termo da matriz \mathbf{H} , $i, j = 1, 2, \dots, n$, então nós temos que:

- $\mathbb{E}(e_j) = 0$;
- $\text{Var}(e_j) = \sigma^2(1 - h_{jj})$;

Resíduos

Seja h_{ij} , o (i, j) -ésimo termo da matriz \mathbf{H} , $i, j = 1, 2, \dots, n$, então nós temos que:

- $\mathbb{E}(e_i) = 0$;
- $\text{Var}(e_i) = \sigma^2(1 - h_{ii})$;
- $\text{Cov}(e_i, e_j) = -\sigma^2 h_{ij}$.



Resíduos

Seja h_{ij} , o (i, j) -ésimo termo da matriz \mathbf{H} , $i, j = 1, 2, \dots, n$, então nós temos que:

- $\mathbb{E}(e_i) = 0$;
- $\text{Var}(e_i) = \sigma^2(1 - h_{ii})$;
- $\text{Cov}(e_i, e_j) = -\sigma^2 h_{ij}$.



Estimação da variância do erro

O estimador da variância, $\hat{\sigma}^2$, é dado por

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2}{n-p} = \frac{1}{n-p} \mathbf{e}^{\top} \mathbf{e} \\ &= \frac{1}{n-p} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^{\top} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right).\end{aligned}\tag{7}$$

Estimação da variância do erro

O estimador da variância, σ^2 , é dado por

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2}{n-p} = \frac{1}{n-p} \mathbf{e}^{\top} \mathbf{e} \\ &= \frac{1}{n-p} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^{\top} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right).\end{aligned}\tag{7}$$

O estimador $\hat{\sigma}^2$, dado em (7), também é denominado de **quadrado médio do resíduo** (QMRes).



Estimação da variância do erro

O estimador da variância, σ^2 , é dado por

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2}{n-p} = \frac{1}{n-p} \mathbf{e}^{\top} \mathbf{e} \\ &= \frac{1}{n-p} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^{\top} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right).\end{aligned}\quad (7)$$

O estimador $\hat{\sigma}^2$, dado em (7), também é denominado de **quadrado médio do resíduo** (QMRes). E $\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2$ é denominado de **soma de quadrados do resíduo** (SQRes).



Estimação da variância do erro

O estimador da variância, σ^2 , é dado por

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2}{n-p} = \frac{1}{n-p} \mathbf{e}^{\top} \mathbf{e} \\ &= \frac{1}{n-p} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^{\top} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right).\end{aligned}\quad (7)$$

O estimador $\hat{\sigma}^2$, dado em (7), também é denominado de **quadrado médio do resíduo** (QMRes). E $\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2$ é denominado de **soma de quadrados do resíduo** (SQRes).



Estimação da variância do erro

A esperança de $\hat{\sigma}^2$ é dada por

$$\begin{aligned}\mathbb{E}(\hat{\sigma}^2) &= \frac{1}{n-p} \mathbb{E}(\mathbf{e}^\top \mathbf{e}) = \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})\mathbf{Y}]^\top (\mathbf{I}_n - \mathbf{H})\mathbf{Y} \right\} \\ &= \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})]^\top (\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \right\} \\ &\stackrel{*}{=} \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon}]^\top (\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right\} \\ &= \frac{1}{n-p} \mathbb{E} \left[\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right] = \frac{1}{n-p} \mathbb{E} \left[\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right] \\ &\stackrel{**}{=} \frac{1}{n-p} \text{tr} \left[(\mathbf{I}_n - \mathbf{H})\sigma^2 \mathbf{I}_n \right] = \frac{\sigma^2}{n-p} \text{tr}(\mathbf{I}_n - \mathbf{H}) \\ &\stackrel{***}{=} \frac{\sigma^2}{n-p} (n-p) = \sigma^2.\end{aligned}$$



Estimação da variância do erro

A esperança de $\hat{\sigma}^2$ é dada por

$$\begin{aligned}\mathbb{E}(\hat{\sigma}^2) &= \frac{1}{n-p} \mathbb{E}(\mathbf{e}^\top \mathbf{e}) = \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})\mathbf{Y}]^\top (\mathbf{I}_n - \mathbf{H})\mathbf{Y} \right\} \\ &= \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})]^\top (\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \right\} \\ &\stackrel{*}{=} \frac{1}{n-p} \mathbb{E} \left\{ [(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon}]^\top (\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right\} \\ &= \frac{1}{n-p} \mathbb{E} \left[\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right] = \frac{1}{n-p} \mathbb{E} \left[\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{H})\boldsymbol{\varepsilon} \right] \\ &\stackrel{**}{=} \frac{1}{n-p} \text{tr} \left[(\mathbf{I}_n - \mathbf{H})\sigma^2 \mathbf{I}_n \right] = \frac{\sigma^2}{n-p} \text{tr}(\mathbf{I}_n - \mathbf{H}) \\ &\stackrel{***}{=} \frac{\sigma^2}{n-p} (n-p) = \sigma^2.\end{aligned}$$



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então

$$\mathbb{E} \left[\mathbf{v}^T \mathbf{W} \mathbf{v} \right] = \text{tr} [\mathbf{W} \mathbf{C}] + \mathbf{m}^T \mathbf{W} \mathbf{m}.$$



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então

$$\mathbb{E} \left[\mathbf{v}^T \mathbf{W} \mathbf{v} \right] = \text{tr} [\mathbf{W} \mathbf{C}] + \mathbf{m}^T \mathbf{W} \mathbf{m}.$$

***: $\text{tr}(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{H})$.



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então

$$\mathbb{E} \left[\mathbf{v}^T \mathbf{W} \mathbf{v} \right] = \text{tr} [\mathbf{W} \mathbf{C}] + \mathbf{m}^T \mathbf{W} \mathbf{m}.$$

***: $\text{tr}(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{H})$. Como $\text{tr}(\mathbf{I}_n) = n$ e

$$\text{tr}(\mathbf{H}) = \text{tr} \left[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] = \text{tr} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \right] = \text{tr}(\mathbf{I}_p) = p.$$



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então

$$\mathbb{E} \left[\mathbf{v}^T \mathbf{W} \mathbf{v} \right] = \text{tr} [\mathbf{W} \mathbf{C}] + \mathbf{m}^T \mathbf{W} \mathbf{m}.$$

***: $\text{tr}(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{H})$. Como $\text{tr}(\mathbf{I}_n) = n$ e

$$\text{tr}(\mathbf{H}) = \text{tr} \left[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] = \text{tr} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \right] = \text{tr}(\mathbf{I}_p) = p.$$

Logo, $\text{tr}(\mathbf{I}_n - \mathbf{H}) = n - p$.



Estimação da variância do erro

*: Ver $\mathbb{E}(\mathbf{e})$.

** : Sejam \mathbf{v} um vetor aleatório de comprimento v , com vetor de médias \mathbf{m} e matriz de covariâncias \mathbf{C} e \mathbf{W} uma matriz de dimensão $v \times v$, então

$$\mathbb{E} \left[\mathbf{v}^T \mathbf{W} \mathbf{v} \right] = \text{tr} [\mathbf{W} \mathbf{C}] + \mathbf{m}^T \mathbf{W} \mathbf{m}.$$

***: $\text{tr}(\mathbf{I}_n - \mathbf{H}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{H})$. Como $\text{tr}(\mathbf{I}_n) = n$ e

$$\text{tr}(\mathbf{H}) = \text{tr} \left[\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] = \text{tr} \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \right] = \text{tr}(\mathbf{I}_p) = p.$$

Logo, $\text{tr}(\mathbf{I}_n - \mathbf{H}) = n - p$.



Estimação da variância do erro

E assim, nós temos que o QMRes, dado em (7), é um estimador não viciado para σ^2 .



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação
- 6 Referências bibliográficas



Método da máxima verossimilhança

Definição

Os estimadores de máxima verossimilhança (EMV) obtém estimativas para parâmetros desconhecidos que **maximizam** a função de verossimilhança.

Método da máxima verossimilhança

Definição

Os estimadores de máxima verossimilhança (EMV) obtém estimativas para parâmetros desconhecidos que **maximizam** a função de verossimilhança.

Método da máxima verossimilhança

A máxima verossimilhança (juntamente com algumas de suas variantes) é o método de estimação mais amplamente utilizado, e uma lista de suas aplicações cobriria praticamente todo o campo da estatística (Lehmann e Casella, 1998, p. 468).

O método da máxima verossimilhança é, de longe, a técnica mais popular para derivar estimadores (Casella e Berger, 2002, p. 315).



Método da máxima verossimilhança

A máxima verossimilhança (juntamente com algumas de suas variantes) é o método de estimação mais amplamente utilizado, e uma lista de suas aplicações cobriria praticamente todo o campo da estatística (Lehmann e Casella, 1998, p. 468).

O método da máxima verossimilhança é, de longe, a técnica mais popular para derivar estimadores (Casella e Berger, 2002, p. 315).



Método da máxima verossimilhança

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, com $Y_\ell \sim \mathcal{N}(\mu_\ell, \sigma^2)$,



Método da máxima verossimilhança

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, com $Y_\ell \sim \mathcal{N}(\mu_\ell, \sigma^2)$,
em que $\mu_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta}$, $\ell = 1, 2, \dots, n$,



Método da máxima verossimilhança

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, com $Y_\ell \sim \mathcal{N}(\mu_\ell, \sigma^2)$, em que $\mu_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta}$, $\ell = 1, 2, \dots, n$, a função de verossimilhança é dada por

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\ell=1}^n (Y_\ell - \mathbf{x}_\ell^\top \boldsymbol{\beta})^2 \right\}. \quad (8)$$



Método da máxima verossimilhança

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias independentes, com $Y_\ell \sim \mathcal{N}(\mu_\ell, \sigma^2)$, em que $\mu_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta}$, $\ell = 1, 2, \dots, n$, a função de verossimilhança é dada por

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\ell=1}^n (Y_\ell - \mathbf{x}_\ell^\top \boldsymbol{\beta})^2 \right\}. \quad (8)$$



Método da máxima verossimilhança

Maximizar (8) é equivalente a maximizar o seu logaritmo, dado por

$$l(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{\ell=1}^n (Y_{\ell} - \mathbf{x}_{\ell}^{\top} \beta)^2. \quad (9)$$



Método da máxima verossimilhança

Maximizar (8) é equivalente a maximizar o seu logaritmo, dado por

$$l(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{\ell=1}^n (Y_{\ell} - \mathbf{x}_{\ell}^{\top} \beta)^2. \quad (9)$$



Método da máxima verossimilhança

Derivando (9), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta_m} = \frac{1}{\sigma^2} \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (10)$$



Método da máxima verossimilhança

Derivando (9), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta_m} = \frac{1}{\sigma^2} \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (10)$$

Igualando (10) a zero,

Método da máxima verossimilhança

Derivando (9), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta_m} = \frac{1}{\sigma^2} \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (10)$$

Igualando (10) a zero, nós temos que

$$\hat{\beta}_1 \sum_{\ell=1}^n x_{\ell 1} x_{\ell m} + \dots + \hat{\beta}_m \sum_{\ell=1}^n x_{\ell m}^2 + \dots + \hat{\beta}_p \sum_{\ell=1}^n x_{\ell p} x_{\ell m} = \sum_{\ell=1}^n x_{\ell m} Y_\ell. \quad (11)$$



Método da máxima verossimilhança

Derivando (9), em relação ao parâmetro β_m , $m = 1, 2, \dots, p$, nós temos:

$$\frac{\partial l(\beta, \sigma^2)}{\partial \beta_m} = \frac{1}{\sigma^2} \sum_{\ell=1}^n [Y_\ell - (\beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \dots + \beta_p x_{\ell p})] x_{\ell m}. \quad (10)$$

Igualando (10) a zero, nós temos que

$$\hat{\beta}_1 \sum_{\ell=1}^n x_{\ell 1} x_{\ell m} + \dots + \hat{\beta}_m \sum_{\ell=1}^n x_{\ell m}^2 + \dots + \hat{\beta}_p \sum_{\ell=1}^n x_{\ell p} x_{\ell m} = \sum_{\ell=1}^n x_{\ell m} Y_\ell. \quad (11)$$



Método da máxima verossimilhança

O **estimador de máxima verossimilhança** de β_m , obtido da solução das p equações em (11), é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (12)$$



Método da máxima verossimilhança

O **estimador de máxima verossimilhança** de β_m , obtido da solução das p equações em (11), é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (12)$$

O EMV em (12) coincide com o EMQ, apresentado em (6).



Método da máxima verossimilhança

O **estimador de máxima verossimilhança** de β_m , obtido da solução das p equações em (11), é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (12)$$

O EMV em (12) coincide com o EMQ, apresentado em (6).



Método da máxima verossimilhança

Agora, derivando (9), em relação ao parâmetro σ^2 , nós temos:

$$\frac{\partial l(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \beta \right)^2. \quad (13)$$



Método da máxima verossimilhança

Agora, derivando (9), em relação ao parâmetro σ^2 , nós temos:

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \boldsymbol{\beta} \right)^2. \quad (13)$$

Igualando (13) a zero,



Método da máxima verossimilhança

Agora, derivando (9), em relação ao parâmetro σ^2 , nós temos:

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \boldsymbol{\beta} \right)^2. \quad (13)$$

Igualando (13) a zero, nós temos que

$$\frac{1}{2(\hat{\sigma}^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2 = \frac{n}{2\hat{\sigma}^2}. \quad (14)$$



Método da máxima verossimilhança

Agora, derivando (9), em relação ao parâmetro σ^2 , nós temos:

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \boldsymbol{\beta} \right)^2. \quad (13)$$

Igualando (13) a zero, nós temos que

$$\frac{1}{2(\hat{\sigma}^2)^2} \sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2 = \frac{n}{2\hat{\sigma}^2}. \quad (14)$$



Método da máxima verossimilhança

O **estimador de máxima verossimilhança** de σ^2 , obtido da solução da equação (14), é dado por:

$$\hat{\sigma}_{MV}^2 = \frac{\sum_{\ell=1}^n (Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}})^2}{n} = \frac{n-p}{n} \text{QMRes.} \quad (15)$$



Método da máxima verossimilhança

O **estimador de máxima verossimilhança** de σ^2 , obtido da solução da equação (14), é dado por:

$$\hat{\sigma}_{MV}^2 = \frac{\sum_{\ell=1}^n \left(Y_{\ell} - \mathbf{x}_{\ell}^{\top} \hat{\boldsymbol{\beta}} \right)^2}{n} = \frac{n-p}{n} \text{QMRes.} \quad (15)$$



Método da máxima verossimilhança

O estimador para β obtido pelo método da máxima verossimilhança coincide com o obtido pelo método dos mínimos quadrados, por conta disso, o EMV de β é **não viesado** e tem menor variância entre todos os estimadores lineares não viesados.



Método da máxima verossimilhança

O estimador para β obtido pelo método da máxima verossimilhança coincide com o obtido pelo método dos mínimos quadrados, por conta disso, o EMV de β é **não viesado** e tem menor variância entre todos os estimadores lineares não viesados.



Método da máxima verossimilhança

O mesmo não acontece com o EMV de σ^2 . De fato, $\hat{\sigma}_{MV}^2$ é um **estimador viciado**,



Método da máxima verossimilhança

O mesmo não acontece com o EMV de σ^2 . De fato, $\hat{\sigma}_{MV}^2$ é um **estimador viciado**, para tamanhos de amostras pequenos, esse viés é considerável.



Método da máxima verossimilhança

O mesmo não acontece com o EMV de σ^2 . De fato, $\hat{\sigma}_{MV}^2$ é um **estimador viciado**, para tamanhos de amostras pequenos, esse viés é considerável.

Adicionalmente, os EMV de β e σ^2 são estimadores consistentes e suficientes.



Método da máxima verossimilhança

O mesmo não acontece com o EMV de σ^2 . De fato, $\hat{\sigma}_{MV}^2$ é um **estimador viciado**, para tamanhos de amostras pequenos, esse viés é considerável.

Adicionalmente, os EMV de β e σ^2 são estimadores consistentes e suficientes.

A suposição de normalidade e a consequente estimação por máxima verossimilhança permitem a construção de procedimentos inferenciais, como intervalos de confiança.



Método da máxima verossimilhança

O mesmo não acontece com o EMV de σ^2 . De fato, $\hat{\sigma}_{MV}^2$ é um **estimador viciado**, para tamanhos de amostras pequenos, esse viés é considerável.

Adicionalmente, os EMV de β e σ^2 são estimadores consistentes e suficientes.

A suposição de normalidade e a consequente estimação por máxima verossimilhança permitem a construção de procedimentos inferenciais, como intervalos de confiança.



Método da máxima verossimilhança

Derivando (9), uma segunda vez, em relação ao parâmetro $\beta_{m'}$, $m' = 1, 2, \dots, p$, nós temos:

$$\frac{\partial^2 l(\beta, \sigma^2)}{\partial \beta_m \partial \beta_{m'}} = -\frac{1}{\sigma^2} \sum_{l=1}^n x_{lm} x_{lm'}. \quad (16)$$



Método da máxima verossimilhança

Derivando (9), uma segunda vez, em relação ao parâmetro $\beta_{m'}$, $m' = 1, 2, \dots, p$, nós temos:

$$\frac{\partial^2 l(\beta, \sigma^2)}{\partial \beta_m \partial \beta_{m'}} = -\frac{1}{\sigma^2} \sum_{\ell=1}^n x_{\ell m} x_{\ell m'}. \quad (16)$$

Se nós aplicamos a esperança e multiplicamos por -1 em (16), obtemos (m, m') -ésimo elemento da matriz de informação de Fisher do EMV de β .



Método da máxima verossimilhança

Derivando (9), uma segunda vez, em relação ao parâmetro $\beta_{m'}$, $m' = 1, 2, \dots, p$, nós temos:

$$\frac{\partial^2 l(\beta, \sigma^2)}{\partial \beta_m \partial \beta_{m'}} = -\frac{1}{\sigma^2} \sum_{\ell=1}^n x_{\ell m} x_{\ell m'}. \quad (16)$$

Se nós aplicamos a esperança e multiplicamos por -1 em (16), obtemos (m, m') -ésimo elemento da matriz de informação de Fisher do EMV de β .



Método da máxima verossimilhança

Lembrando que, a equação (10) é a função escore. Em notação matricial, o vetor escore e a matriz de informação de Fisher são dadas, respectivamente, por



Método da máxima verossimilhança

Lembrando que, a equação (10) é a função escore. Em notação matricial, o vetor escore e a matriz de informação de Fisher são dadas, respectivamente, por

$$\mathbf{U}_\beta = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\beta) \text{ e } \mathbf{K}_\beta = \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X}. \quad (17)$$



Método da máxima verossimilhança

Lembrando que, a equação (10) é a função escore. Em notação matricial, o vetor escore e a matriz de informação de Fisher são dadas, respectivamente, por

$$\mathbf{U}_\beta = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\beta) \text{ e } \mathbf{K}_\beta = \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X}. \quad (17)$$



Considerações

Embora seja comum utilizar $\hat{Y}_l = \mathbf{x}_l^\top \hat{\beta}$,

Considerações

Embora seja comum utilizar $\hat{Y}_\ell = \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}}$, o que, de fato, nós estamos estimando é μ_ℓ .

Considerações

Embora seja comum utilizar $\hat{Y}_\ell = \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}}$, o que, de fato, nós estamos estimando é μ_ℓ . Basta lembrar que, $\mu_\ell = \mathbb{E}(Y_\ell) = \mathbf{x}_\ell^\top \boldsymbol{\beta}$.



Considerações

Embora seja comum utilizar $\hat{Y}_\ell = \mathbf{x}_\ell^\top \hat{\boldsymbol{\beta}}$, o que, de fato, nós estamos estimando é μ_ℓ . Basta lembrar que, $\mu_\ell = \mathbb{E}(Y_\ell) = \mathbf{x}_\ell^\top \boldsymbol{\beta}$.



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações**
- 5 Estudo de simulação
- 6 Referências bibliográficas



Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste (estimação dos parâmetros), nós temos o seguinte modelo estimado,

$$\hat{Y}_l = 10 + 2x_{l2},$$

Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste (estimação dos parâmetros), nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

em que Y_ℓ : horas trabalhadas, $x_{\ell 2}$: tamanho do lote, $\ell = 1, 2, \dots, 10$ e $QMRes = 7,5$.

Aplicação 1. Conjunto de dados *Westwood Company* (Neter et al., 1983, p. 36). Após o ajuste (estimação dos parâmetros), nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

em que Y_ℓ : horas trabalhadas, $x_{\ell 2}$: tamanho do lote, $\ell = 1, 2, \dots, 10$ e $QMRes = 7,5$.

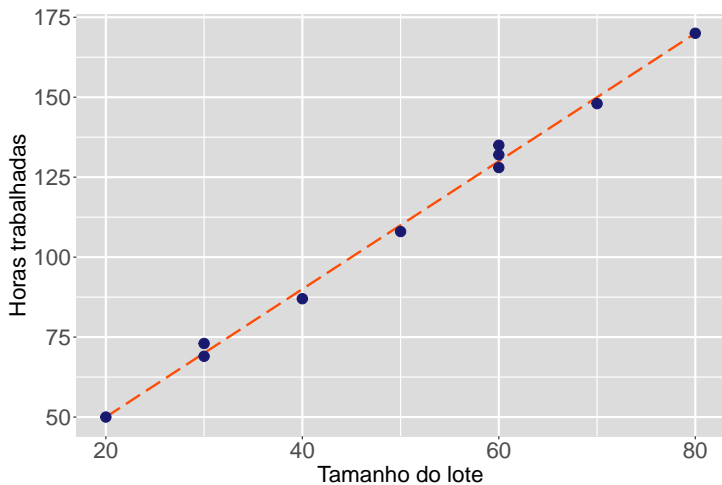


Figura 1: GD com a reta ajustada para os dados *Westwood Company*.

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = 10$:

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = 10$: se o tamanho do lote fosse zero, haveria 10 horas trabalhadas;

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = 10$: se o tamanho do lote fosse zero, haveria 10 horas trabalhadas;
- $\hat{\beta}_2 = 2$:

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = 10$: se o tamanho do lote fosse zero, haveria 10 horas trabalhadas;
- $\hat{\beta}_2 = 2$: a cada uma unidade aumentada no lote, ocorreria um aumento médio de 2 horas trabalhadas.

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = 10$: se o tamanho do lote fosse zero, haveria 10 horas trabalhadas;
- $\hat{\beta}_2 = 2$: a cada uma unidade aumentada no lote, ocorreria um aumento médio de 2 horas trabalhadas.

Aplicação 2. Conjunto de dados Walker (1962). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -7,211 + 3,603x_{\ell 2} - 10,065x_{\ell 3},$$

Aplicação 2. Conjunto de dados Walker (1962). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -7,211 + 3,603x_{\ell 2} - 10,065x_{\ell 3},$$

em que Y_ℓ : número de pulsações, $x_{\ell 2}$: temperatura, $x_{\ell 3}$ a espécie do grilo, $x_{\ell 3} \in \{Oecanthus exclamationis (ex), Oecanthus niveus (niv)\}$, $\ell = 1, 2, \dots, 31$ e $QMRes = 12,764$.

Aplicação 2. Conjunto de dados Walker (1962). Após o ajuste, nós temos o seguinte modelo estimado,

$$\hat{Y}_\ell = -7,211 + 3,603x_{\ell 2} - 10,065x_{\ell 3},$$

em que Y_ℓ : número de pulsações, $x_{\ell 2}$: temperatura, $x_{\ell 3}$ a espécie do grilo, $x_{\ell 3} \in \{Oecanthus exclamationis (ex), Oecanthus niveus (niv)\}$, $\ell = 1, 2, \dots, 31$ e $QMRes = 12,764$.

Uma forma alternativa de apresentar o modelo ajustado é a seguinte:

$$\left\{ \begin{array}{l} \textit{Oecanthus niveus} : \quad \hat{Y}_\ell = -17,276 + 3,603x_{\ell 2} \\ \textit{Oecanthus exclamationis} : \quad \hat{Y}_\ell = -7,211 + 3,603x_{\ell 2} \end{array} \right.$$

Uma forma alternativa de apresentar o modelo ajustado é a seguinte:

$$\left\{ \begin{array}{l} \textit{Oecanthus niveus} : \quad \hat{Y}_\ell = -17,276 + 3,603x_{\ell 2} \\ \textit{Oecanthus exclamationis} : \quad \hat{Y}_\ell = -7,211 + 3,603x_{\ell 2} \end{array} \right. .$$

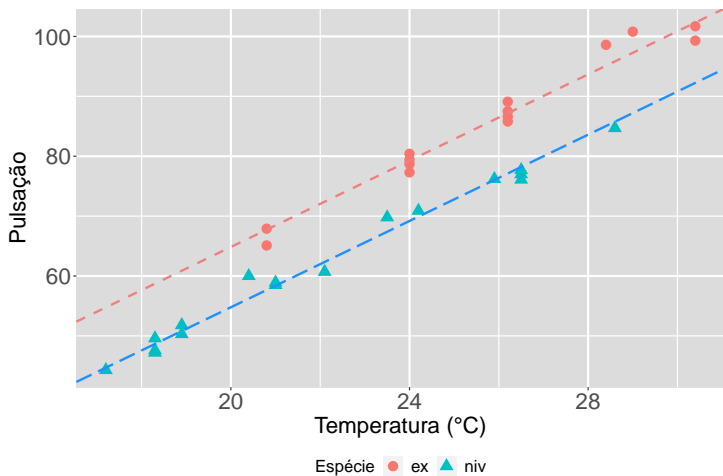


Figura 2: GD com as retas ajustadas para os dados de Walker (1962).

Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$:



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);
- $\hat{\beta}_2 = 3,603$:



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);
- $\hat{\beta}_2 = 3,603$: a cada uma unidade aumentada na temperatura, haveria um aumento médio de 3,603 nas pulsações;



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);
- $\hat{\beta}_2 = 3,603$: a cada uma unidade aumentada na temperatura, haveria um aumento médio de 3,603 nas pulsações;
- $\hat{\beta}_3 = -10,065$:



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);
- $\hat{\beta}_2 = 3,603$: a cada uma unidade aumentada na temperatura, haveria um aumento médio de 3,603 nas pulsações;
- $\hat{\beta}_3 = -10,065$: a espécie *Oecanthus niveus* tem, em média, - 10,065 pulsações a menos que a *Oecanthus exclamationis*.



Aplicações

Os parâmetros podem ser interpretados da seguinte forma

- $\hat{\beta}_1 = -7,211$: se a temperatura fosse zero e a espécie do grilo fosse a *Oecanthus exclamationis*, o número médio de pulsações seria de -7,211 (faz sentido?);
- $\hat{\beta}_2 = 3,603$: a cada uma unidade aumentada na temperatura, haveria um aumento médio de 3,603 nas pulsações;
- $\hat{\beta}_3 = -10,065$: a espécie *Oecanthus niveus* tem, em média, - 10,065 pulsações a menos que a *Oecanthus exclamationis*.



Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação**
- 6 Referências bibliográficas



Estudo de simulação

Suponham o seguinte MNL,

$$Y_l = 1 + x_{l2} + \varepsilon_l,$$

Estudo de simulação

Suponham o seguinte MNL,

$$Y_\ell = 1 + x_{\ell 2} + \varepsilon_\ell,$$

em que $x_{\ell 2}$ é quantidade conhecida e $\varepsilon_\ell \sim \mathcal{N}(0, 4)$, $\ell = 1, 2, \dots, n$.

Estudo de simulação

Suponham o seguinte MNL,

$$Y_\ell = 1 + x_{\ell 2} + \varepsilon_\ell,$$

em que $x_{\ell 2}$ é quantidade conhecida e $\varepsilon_\ell \sim \mathcal{N}(0, 4)$, $\ell = 1, 2, \dots, n$. E nós vamos assumir que $x_{\ell 2} \sim$ uniforme no intervalo $(0, 1)$.

Estudo de simulação

Suponham o seguinte MNL,

$$Y_\ell = 1 + x_{\ell 2} + \varepsilon_\ell,$$

em que $x_{\ell 2}$ é quantidade conhecida e $\varepsilon_\ell \sim \mathcal{N}(0, 4)$, $\ell = 1, 2, \dots, n$. E nós vamos assumir que $x_{\ell 2} \sim$ uniforme no intervalo $(0, 1)$.

Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 ,



Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 , neste último, utilizando o QMRes e o EMV,



Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 , neste último, utilizando o QMRes e o EMV, para $n \in \{10, 20, 40, 80, 160\}$.



Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 , neste último, utilizando o QMRes e o EMV, para $n \in \{10, 20, 40, 80, 160\}$. Sem perda de generalidade, o viés absoluto para o estimador de θ é dada por

$$\text{Viés absoluto} = \left| \hat{\theta}_{MC} - \theta \right|,$$



Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 , neste último, utilizando o QMRes e o EMV, para $n \in \{10, 20, 40, 80, 160\}$. Sem perda de generalidade, o viés absoluto para o estimador de θ é dada por

$$\text{Viés absoluto} = \left| \hat{\theta}_{MC} - \theta \right|,$$

em que $\hat{\theta}_{MC} = 5.000^{-1} \sum_{i=1}^{5.000} \hat{\theta}_i$.



Estudo de simulação

Nós faremos um estudo de Monte Carlo, com 5.000 réplicas, onde avaliaremos o viés absoluto das estimativas dos parâmetros β_1 , β_2 e σ^2 , neste último, utilizando o QMRes e o EMV, para $n \in \{10, 20, 40, 80, 160\}$. Sem perda de generalidade, o viés absoluto para o estimador de θ é dada por

$$\text{Viés absoluto} = \left| \hat{\theta}_{MC} - \theta \right|,$$

em que $\hat{\theta}_{MC} = 5.000^{-1} \sum_{i=1}^{5.000} \hat{\theta}_i$.



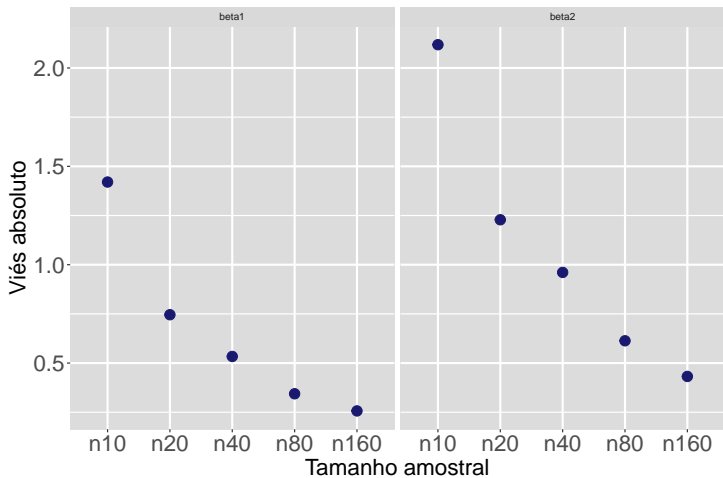


Figura 3: Viés absoluto para β .

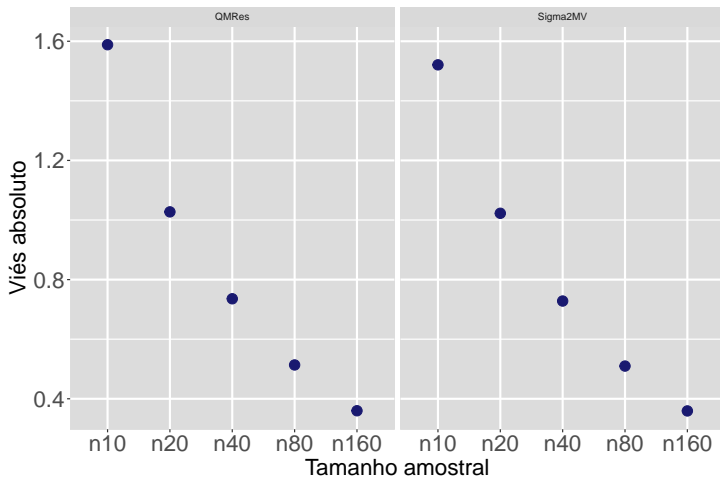


Figura 4: Viés absoluto para σ^2 , utilizando o QMRes e o EMV.

Estudo de simulação

Comentários

Pelas Figuras 3 e 4, nós podemos perceber que, a medida que o tamanho amostral cresce, o viés absoluto converge para zero.

Estudo de simulação

Comentários

Pelas Figuras 3 e 4, nós podemos perceber que, a medida que o tamanho amostral cresce, o viés absoluto converge para zero.

Roteiro

- 1 Introdução
- 2 Método dos mínimos quadrados
- 3 Método da máxima verossimilhança
- 4 Aplicações
- 5 Estudo de simulação
- 6 Referências bibliográficas**



Referências bibliográficas I

Casella, G. e Berger, R. L. (2002), *Statistical Inference*, 2nd edn, Duxbury/Thomson Learning, Pacific Grove, CA, USA.

Lehmann, E. L. e Casella, G. (1998), *Theory of Point Estimation*, 2nd edn, Springer, New York, NY, USA.

Neter, J., Wasserman, W. e Kutner, M. H. (1983), *Applied linear regression models*, Richard D. Irwin Inc, Homewood, Illinois.

Plackett, R. L. (1949), 'A historical note on the method of least squares', *Biometrika* **36**(3/4), 458–460.



Referências bibliográficas II

Walker, T. J. (1962), 'The Taxonomy and Calling Songs of United States Tree Crickets (Orthoptera: Gryllidae: Oecanthinae). I. The Genus *Neoxabea* and the *niveus* and *varicornis* Groups of the Genus *Oecanthus*', *Annals of the Entomological Society of America* **55**(3), 303–322.



Obrigado!

✉ tiago.magalhaes@ufjf.br

📄 ufjf.br/tiago_magalhaes

🌐 Departamento de Estatística, Sala 319

