Método bayesiano

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 15 de abril de 2026

Roteiro

Roteiro

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell,\ \ell = 1, 2, \ldots, n, \tag{1}$$

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell,\ \ell = 1, 2, \ldots, n, \tag{1}$$

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

Modelo de regressão linear

Suponham que $Y_1, Y_2, \ldots, Y_n$ tais que

$$Y_\ell = \boldsymbol{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell,\ \ell = 1, 2, \ldots, n, \tag{1}$$

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

Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \tag{2}$$

Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

em que $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$ é a matriz de planejamento e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, com

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ e } \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n.$$

Forma matricial

A Equação (1) pode ser escrita de forma matricial, da seguinte forma:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \tag{2}$$

em que $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)^\top$ é a matriz de planejamento e $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, com

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0} \text{ e } \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n.$$

Suposições

Resumindo,

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

Suposições

Resumindo,

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

Suposições

Resumindo,

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

Suposições

Resumindo,

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

Suposições

Resumindo,

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

Suposições

Resumindo,

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

Estimação

O estimador de mínimos quadrados (EMQ) do vetor $\boldsymbol{\beta}$ é dado por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}. \tag{3}$$

Estimação

O estimador de mínimos quadrados (EMQ) do vetor $\boldsymbol{\beta}$ é dado por:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}. \tag{3}$$

Enquanto, um estimador para σ^2 é dado pelo quadrado médio do resíduo (QMRes), i.e.,

Estimação

O estimador de mínimos quadrados (EMQ) do vetor β é dado por:

$$\widehat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}. \tag{3}$$

Enquanto, um estimador para σ^2 é dado pelo quadrado médio do resíduo (QMRes), i.e.,

$$\text{QMRes} = \frac{1}{n-p} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right)^\top \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right). \tag{4}$$

Estimação

O estimador de mínimos quadrados (EMQ) do vetor β é dado por:

$$\widehat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}. \tag{3}$$

Enquanto, um estimador para σ^2 é dado pelo quadrado médio do resíduo (QMRes), i.e.,

$$\text{QMRes} = \frac{1}{n-p} \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right)^\top \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta} \right). \tag{4}$$

Normalidade dos erros

Se, adicionalmente, nós supomos que, $\boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, nós temos o seguinte:

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

Normalidade dos erros

Se, adicionalmente, nós supomos que, $\varepsilon \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, nós temos o seguinte:

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

i.e., a função densidade de probabilidade (FDP) de $\boldsymbol{Y}$ pode ser escrita como:

Normalidade dos erros

Se, adicionalmente, nós supomos que, $\boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, nós temos o seguinte:

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

i.e., a função densidade de probabilidade (FDP) de $\boldsymbol{Y}$ pode ser escrita como:

$$f(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)^\top(\boldsymbol{y} - \boldsymbol{X}\beta)\right\} \mathbb{I}_{\mathbb{R}^n}(\boldsymbol{y}). \qquad (5)$$

Normalidade dos erros

Se, adicionalmente, nós supomos que, $\boldsymbol{\varepsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, nós temos o seguinte:

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

i.e., a função densidade de probabilidade (FDP) de $\boldsymbol{Y}$ pode ser escrita como:

$$f(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)^\top(\boldsymbol{y} - \boldsymbol{X}\beta)\right\} \mathbb{I}_{\mathbb{R}^n}(\boldsymbol{y}). \tag{5}$$

Estimação por máxima verossimilhança

A FDP de $\boldsymbol{Y}$, em (5), também é a função de verossimilhança de β e σ^2. O estimador de máxima verossimilhança (EMV) de β coincide com o EMQ, dado em (3),

Estimação por máxima verossimilhança

A FDP de $\boldsymbol{Y}$, em (5), também é a função de verossimilhança de β e σ^2. O estimador de máxima verossimilhança (EMV) de β coincide com o EMQ, dado em (3), enquanto o EMV de σ^2 pode ser escrito em função do QMRes, dado em (4): $\hat{\sigma}^2 = [n/(n-p)]$ QMRes.

Estimação por máxima verossimilhança

A FDP de $\boldsymbol{Y}$, em (5), também é a função de verossimilhança de $\boldsymbol{\beta}$ e σ^2. O estimador de máxima verossimilhança (EMV) de $\boldsymbol{\beta}$ coincide com o EMQ, dado em (3), enquanto o EMV de σ^2 pode ser escrito em função do QMRes, dado em (4): $\hat{\sigma}^2 = [n/(n-p)]$ QMRes.

Roteiro

Inferência bayesiana

Na inferência bayesiana, β é um vetor aleatório e σ^2 uma variável aleatória.

Inferência bayesiana

Na inferência bayesiana, $\boldsymbol{\beta}$ é um vetor aleatório e σ^2 uma variável aleatória.

Toda a inferência é feita através da distribuição *a posteriori*, $\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})$.

Inferência bayesiana

Na inferência bayesiana, $\boldsymbol{\beta}$ é um vetor aleatório e σ^2 uma variável aleatória.

Toda a inferência é feita através da distribuição *a posteriori*, $\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})$.

Esta distribuição é a combinação entre

Inferência bayesiana

Na inferência bayesiana, $\boldsymbol{\beta}$ é um vetor aleatório e σ^2 uma variável aleatória.

Toda a inferência é feita através da distribuição *a posteriori*, $\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})$.

Esta distribuição é a combinação entre

- a função de verossimilhança, $L(\boldsymbol{y} | \boldsymbol{X}, \beta, \sigma^2)$, a informação obtida dos dados;

Inferência bayesiana

Na inferência bayesiana, $\boldsymbol{\beta}$ é um vetor aleatório e σ^2 uma variável aleatória.

Toda a inferência é feita através da distribuição *a posteriori*, $\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})$.

Esta distribuição é a combinação entre

- a função de verossimilhança, $L(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2)$, a informação obtida dos dados;
- e a distribuição *a priori*, $\pi(\boldsymbol{\beta}, \sigma^2)$, uma informação anterior aos dados.

Inferência bayesiana

Na inferência bayesiana, $\boldsymbol{\beta}$ é um vetor aleatório e σ^2 uma variável aleatória.

Toda a inferência é feita através da distribuição *a posteriori*, $\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X})$.

Esta distribuição é a combinação entre

- a função de verossimilhança, $L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2)$, a informação obtida dos dados;
- e a distribuição *a priori*, $\pi(\boldsymbol{\beta}, \sigma^2)$, uma informação anterior aos dados.

Distribuição *a posteriori*

Pelo Teorema de Bayes, a distribuição *a posteriori* é dada por

$$\pi(\beta, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \frac{L(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta, \sigma^2)}{\int_{\beta_1} \cdots \int_{\beta_p} \int_{\sigma^2} L(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta, \sigma^2)\, d\beta_1 \cdots d\beta_p d\sigma^2}$$

$$= \frac{L(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta, \sigma^2)}{f(\boldsymbol{y}|\boldsymbol{X})}. \qquad (6)$$

Distribuição *a posteriori*

Pelo Teorema de Bayes, a distribuição *a posteriori* é dada por

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \frac{L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{\int_{\beta_1} \cdots \int_{\beta_p} \int_{\sigma^2} L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2) \, d\beta_1 \cdots d\beta_p d\sigma^2}$$

$$= \frac{L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{f(\boldsymbol{y}|\boldsymbol{X})}. \tag{6}$$

Distribuição *a posteriori*

Lembrando que,

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}|\sigma^2) \times \pi(\sigma^2),$$

Distribuição *a posteriori*

Lembrando que,

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}|\sigma^2) \times \pi(\sigma^2),$$

a posteriori em (6) pode ser expressa da seguinte forma:

Distribuição *a posteriori*

Lembrando que,

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}|\sigma^2) \times \pi(\sigma^2),$$

a posteriori em (6) pode ser expressa da seguinte forma:

$$\pi(\beta, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) = \frac{L(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta|\sigma^2) \times \pi(\sigma^2)}{f(\boldsymbol{y}|\boldsymbol{X})}. \quad (7)$$

Distribuição *a posteriori*

Lembrando que,

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}|\sigma^2) \times \pi(\sigma^2),$$

a posteriori em (6) pode ser expressa da seguinte forma:

$$\pi(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) = \frac{L(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}|\sigma^2) \times \pi(\sigma^2)}{f(\boldsymbol{y}|\boldsymbol{X})}. \tag{7}$$

Distribuição *a posteriori*

A distribuição *a posteriori* em (7) ainda pode ser representada por

$$\pi(\beta, \sigma^2|\boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{y}|\boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta|\sigma^2) \times \pi(\sigma^2). \tag{8}$$

Distribuição *a posteriori*

A distribuição *a posteriori* em (7) ainda pode ser representada por

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{y} | \boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta | \sigma^2) \times \pi(\sigma^2). \tag{8}$$

A verossimilhança em (8) está dada em (5). Enquanto, *a priori* depende de alguma crença.

Distribuição *a posteriori*

A distribuição *a posteriori* em (7) ainda pode ser representada por

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{y} | \boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta | \sigma^2) \times \pi(\sigma^2). \tag{8}$$

A verossimilhança em (8) está dada em (5). Enquanto, *a priori* depende de alguma crença. Porém, não há garantia que *a posteriori* tenha forma fechada.

Distribuição *a posteriori*

A distribuição *a posteriori* em (7) ainda pode ser representada por

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{y} | \boldsymbol{X}, \beta, \sigma^2) \times \pi(\beta | \sigma^2) \times \pi(\sigma^2). \tag{8}$$

A verossimilhança em (8) está dada em (5). Enquanto, *a priori* depende de alguma crença. Porém, não há garantia que *a posteriori* tenha forma fechada.

Distribuição *a posteriori*

Para formas fechadas, se propõe *priori* conjugadas. Para os modelos de regressão,

Distribuição *a posteriori*

Para formas fechadas, se propõe *priori* conjugadas. Para os modelos de regressão, nós temos as seguintes *prioris* conjugadas:

$$\beta|\sigma^2 \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Lambda}_0^{-1}) \text{ e } \sigma^2 \sim \text{gama inversa}(a_0, b_0),$$

Distribuição *a posteriori*

Para formas fechadas, se propõe *priori* conjugadas. Para os modelos de regressão, nós temos as seguintes *prioris* conjugadas:

$$\boldsymbol{\beta}|\sigma^2 \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Lambda}_0^{-1}) \text{ e } \sigma^2 \sim \text{gama inversa}(a_0, b_0),$$

em que $a_0 = v_0/2$ e $b_0 = v_0 s_0^2/2$, a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ são hiper-parâmetros.

Distribuição *a posteriori*

Para formas fechadas, se propõe *priori* conjugadas. Para os modelos de regressão, nós temos as seguintes *prioris* conjugadas:

$$\boldsymbol{\beta}|\sigma^2 \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \sigma^2\boldsymbol{\Lambda}_0^{-1}) \text{ e } \sigma^2 \sim \text{gama inversa}(a_0, b_0),$$

em que $a_0 = v_0/2$ e $b_0 = v_0 s_0^2/2$, a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ são hiper-parâmetros.

Distribuição *a posteriori*

Recordando que, se $U \sim$ gama inversa(a, b), sua FDP é dada por:

$$f(u) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{u}\right)^{a+1} \exp\left\{-\frac{b}{u}\right\} \mathbb{I}_{\mathbb{R}^+}(u),$$

Distribuição *a posteriori*

Recordando que, se $U \sim$ gama inversa(a, b), sua FDP é dada por:

$$f(u) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{u}\right)^{a+1} \exp\left\{-\frac{b}{u}\right\} \mathbb{I}_{\mathbb{R}^+}(u),$$

com $a \in \mathbb{R}^+$ e $b \in \mathbb{R}^+$.

Distribuição *a posteriori*

Recordando que, se $U \sim$ gama inversa(a, b), sua FDP é dada por:

$$f(u) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{u}\right)^{a+1} \exp\left\{-\frac{b}{u}\right\} \mathbb{I}_{\mathbb{R}^+}(u),$$

com $a \in \mathbb{R}^+$ e $b \in \mathbb{R}^+$. E também,

$$\mathbb{E}(U) = \frac{b}{a-1}, \text{ para } a > 1 \text{ e } \mathrm{Var}(U) = \frac{b^2}{(a-1)^2(a-2)}, \text{ para } a > 2.$$

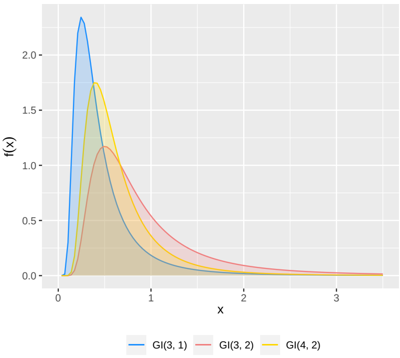
Distribuição *a posteriori*

Recordando que, se $U \sim$ gama inversa(a, b), sua FDP é dada por:

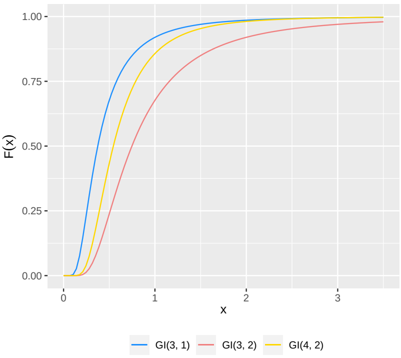
$$f(u) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{u}\right)^{a+1} \exp\left\{-\frac{b}{u}\right\} \mathbb{I}_{\mathbb{R}^+}(u),$$

com $a \in \mathbb{R}^+$ e $b \in \mathbb{R}^+$. E também,

$$\mathbb{E}(U) = \frac{b}{a-1}, \text{ para } a > 1 \text{ e } \mathrm{Var}(U) = \frac{b^2}{(a-1)^2(a-2)}, \text{ para } a > 2.$$



(a) FDP.



(b) FDA.

Figura 1: Gama inversa para diferentes valores de a e b.

Distribuição *a posteriori*

Após algumas manipulações com as *prioris*, (8) pode ser escrita como:

$$\pi(\beta, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \pi(\beta | \sigma^2, \boldsymbol{y}, \boldsymbol{X}) \times \pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X}), \qquad (9)$$

Distribuição *a posteriori*

Após algumas manipulações com as *prioris*, (8) pode ser escrita como:

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X}) \times \pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X}), \tag{9}$$

em que $\pi(\beta | \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ e $\pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X})$ são as funções densidade de probabilidades $\mathrm{N}_p(\boldsymbol{\mu}_n, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$

Distribuição *a posteriori*

Após algumas manipulações com as *prioris*, (8) pode ser escrita como:

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X}) \times \pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X}), \tag{9}$$

em que $\pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ e $\pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X})$ são as funções densidade de probabilidades $\mathrm{N}_p(\boldsymbol{\mu}_n, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$ e gama inversa(a_n, b_n), respectivamente,

Distribuição *a posteriori*

Após algumas manipulações com as *prioris*, (8) pode ser escrita como:

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X}) \times \pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X}), \tag{9}$$

em que $\pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ e $\pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X})$ são as funções densidade de probabilidades $\mathrm{N}_p(\boldsymbol{\mu}_n, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$ e gama inversa(a_n, b_n), respectivamente, com

$$\Lambda_n = \boldsymbol{X}^\top \boldsymbol{X} + \Lambda_0, \quad \boldsymbol{\mu}_n = \Lambda_n^{-1} \left(\boldsymbol{X}^\top \boldsymbol{X} \widehat{\beta} + \Lambda_0 \boldsymbol{\mu}_0 \right),$$

$$a_n = a_0 + \frac{n}{2}, \quad b_n = b_0 + \frac{1}{2} \left(\boldsymbol{y}^\top \boldsymbol{y} + \boldsymbol{\mu}_0^\top \Lambda_0 \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top \Lambda_n \boldsymbol{\mu}_n \right).$$

Distribuição *a posteriori*

Após algumas manipulações com as *prioris*, (8) pode ser escrita como:

$$\pi(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}, \boldsymbol{X}) = \pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X}) \times \pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X}), \tag{9}$$

em que $\pi(\boldsymbol{\beta} | \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ e $\pi(\sigma^2 | \boldsymbol{y}, \boldsymbol{X})$ são as funções densidade de probabilidades $\mathrm{N}_p(\boldsymbol{\mu}_n, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$ e gama inversa(a_n, b_n), respectivamente, com

$$\boldsymbol{\Lambda}_n = \boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{\Lambda}_0, \quad \boldsymbol{\mu}_n = \boldsymbol{\Lambda}_n^{-1} \left(\boldsymbol{X}^\top \boldsymbol{X} \widehat{\boldsymbol{\beta}} + \boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0 \right),$$
$$a_n = a_0 + \frac{n}{2}, \quad b_n = b_0 + \frac{1}{2} \left(\boldsymbol{y}^\top \boldsymbol{y} + \boldsymbol{\mu}_0^\top \boldsymbol{\Lambda}_0 \boldsymbol{\mu}_0 - \boldsymbol{\mu}_n^\top \boldsymbol{\Lambda}_n \boldsymbol{\mu}_n \right).$$

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de β e σ^2, da seguinte forma:

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de $\boldsymbol{\beta}$ e σ^2, da seguinte forma:

- $\sigma^{2(i)} \sim \text{gama inversa}\,(a_n,\, b_n)$;

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de $\boldsymbol{\beta}$ e σ^2, da seguinte forma:

- $\sigma^{2(i)} \sim$ gama inversa $(a_n,\, b_n)$;
- $\beta^{(i)}|\sigma^{2(i)} \sim \mathrm{N}_p\left(\boldsymbol{\mu}_n,\, \sigma^2\boldsymbol{\Lambda}_n^{-1}\right)$.

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de $\boldsymbol{\beta}$ e σ^2, da seguinte forma:

- $\sigma^{2(i)} \sim \text{gama inversa}\,(a_n,\, b_n)$;
- $\boldsymbol{\beta}^{(i)} | \sigma^{2(i)} \sim \text{N}_p\,(\boldsymbol{\mu}_n,\, \sigma^2 \boldsymbol{\Lambda}_n^{-1})$.

`Passo 2.` Com os i valores, calcular média, mediana ou moda

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de $\boldsymbol{\beta}$ e σ^2, da seguinte forma:

- $\sigma^{2(i)} \sim \text{gama inversa}\,(a_n,\, b_n)$;
- $\boldsymbol{\beta}^{(i)}|\sigma^{2(i)} \sim \mathrm{N}_p\left(\boldsymbol{\mu}_n,\, \sigma^2\boldsymbol{\Lambda}_n^{-1}\right)$.

`Passo 2.` Com os i valores, calcular média, mediana ou moda e esta quantidade calculada será a estimativa bayesiana de β e σ^2.

Algoritmo

`Passo 0.` Definir a_0, b_0, $\boldsymbol{\mu}_0$, $\boldsymbol{\Lambda}_0^{-1}$ e calcular a_n, b_n, $\boldsymbol{\mu}_n$, $\boldsymbol{\Lambda}_n^{-1}$;

`Passo 1.` Simular i valores de $\boldsymbol{\beta}$ e σ^2, da seguinte forma:

- $\sigma^{2(i)} \sim$ gama inversa $(a_n,\, b_n)$;
- $\boldsymbol{\beta}^{(i)}|\sigma^{2(i)} \sim \mathrm{N}_p\left(\boldsymbol{\mu}_n,\, \sigma^2\boldsymbol{\Lambda}_n^{-1}\right)$.

`Passo 2.` Com os i valores, calcular média, mediana ou moda e esta quantidade calculada será a estimativa bayesiana de $\boldsymbol{\beta}$ e σ^2.

Roteiro

Aplicação

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

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

Aplicação

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, $\ell = 1, 2, \ldots, 10$ e QMRes $= 7{,}5$.

Aplicação

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, $\ell = 1, 2, \ldots, 10$ e QMRes $= 7{,}5$.

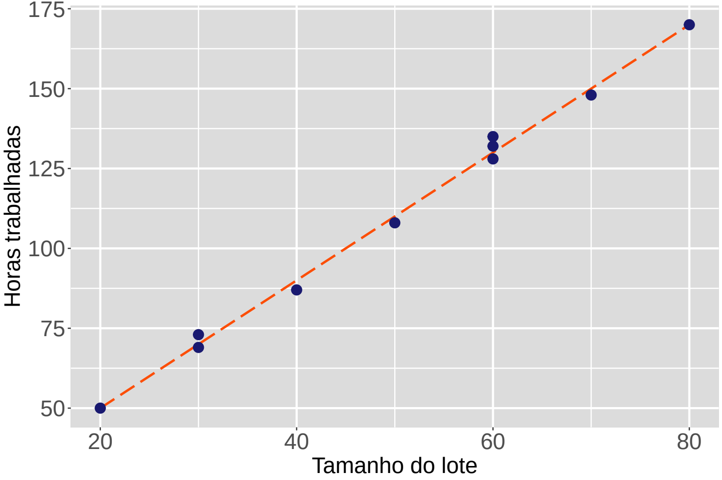


Figura 2: GD com a reta ajustada para os dados *Westwood Company*.

Aplicações

Os parâmetros podem ser interpretados da seguinte forma

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

Aplicações

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.

Aplicações

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.

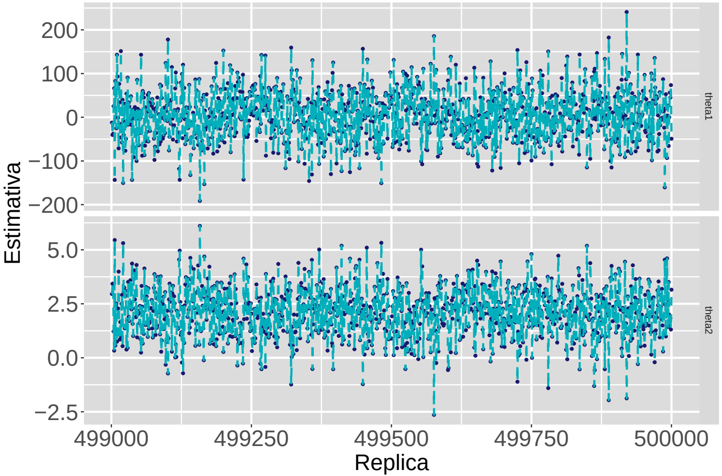


Figura 3: Estimativas observadas.

Aplicação

Nós temos os seguintes modelos estimados,

$$\hat{Y}_\ell = 5{,}96 + 2{,}07 x_{\ell 2}, \text{ (Média)}$$

$$\hat{Y}_\ell = 6{,}05 + 2{,}07 x_{\ell 2}. \text{ (Mediana)}$$

Roteiro

Agradecimentos

Ao Prof. Márcio A. Diniz (Icahn School of Medicine at Mount Sinai, USA).

Roteiro

Referências bibliográficas I

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

Obrigado!

✉ tiago.magalhaes@ufjf.br

ufjf.br/tiago_magalhaes

Departamento de Estatística, Sala 319