

# Método bayesiano

Tiago M. Magalhães

Departamento de Estatística - ICE-UFJF

Juiz de Fora, 05 de junho de 2024



# Roteiro

- 1 Introdução
- 2 Método bayesiano
- 3 Aplicação
- 4 Agradecimentos
- 5 Referências bibliográficas



# Roteiro

- 1 Introdução
- 2 Método bayesiano
- 3 Aplicação
- 4 Agradecimentos
- 5 Referências bibliográficas



# Modelo de regressão linear

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

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ell = 1, 2, \dots, n, \quad (1)$$



# Modelo de regressão linear

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

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \ell = 1, 2, \dots, n, \quad (1)$$

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,  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  são variáveis aleatórias independentes e com a mesma variância  $\sigma^2$ , também desconhecida, a ser estimada.



# Modelo de regressão linear

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

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad \ell = 1, 2, \dots, n, \quad (1)$$

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,  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  são variáveis aleatórias independentes e com a mesma variância  $\sigma^2$ , também desconhecida, a ser estimada.



# Forma matricial

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

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



# Forma matricial

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

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

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

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





# Forma matricial

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

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

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

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



# 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  $\beta$  é dado por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3)$$





# Estimação

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

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3)$$

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

# Estimação

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

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3)$$

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

$$\text{QMRes} = \frac{1}{n - p} (\mathbf{Y} - \mathbf{X}\hat{\beta})^T (\mathbf{Y} - \mathbf{X}\hat{\beta}). \quad (4)$$



# Estimação

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

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3)$$

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

$$\text{QMRes} = \frac{1}{n - p} (\mathbf{Y} - \mathbf{X}\hat{\beta})^T (\mathbf{Y} - \mathbf{X}\hat{\beta}). \quad (4)$$



# Normalidade dos erros

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

$$Y \sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n),$$

# Normalidade dos erros

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

$$\mathbf{Y} \sim N_n(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n),$$

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



# Normalidade dos erros

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

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

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

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



# Normalidade dos erros

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

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

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

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



# Estimação por máxima verossimilhança

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





# Estimação por máxima verossimilhança

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



# Estimação por máxima verossimilhança

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



# Roteiro

- 1 Introdução
- 2 Método bayesiano**
- 3 Aplicação
- 4 Agradecimentos
- 5 Referências bibliográficas



# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória.



# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória.

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



# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória.

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

Esta distribuição é a combinação entre



# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória. Toda a inferência é feita através da distribuição *a posteriori*,  $\pi(\beta, \sigma^2 | \mathbf{y}, \mathbf{X})$ . Esta distribuição é a combinação entre

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



# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória. Toda a inferência é feita através da distribuição *a posteriori*,  $\pi(\beta, \sigma^2 | \mathbf{y}, \mathbf{X})$ .

Esta distribuição é a combinação entre

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





# Inferência bayesiana

Na inferência bayesiana,  $\beta$  é um vetor aleatório e  $\sigma^2$  uma variável aleatória. Toda a inferência é feita através da distribuição *a posteriori*,  $\pi(\beta, \sigma^2 | \mathbf{y}, \mathbf{X})$ .

Esta distribuição é a combinação entre

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



# Distribuição *a posteriori*

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



# Distribuição *a posteriori*

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

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) &= \frac{L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{\int_{\beta_1} \cdots \int_{\beta_p} \int_{\sigma^2} L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2) d\beta_1 \cdots d\beta_p d\sigma^2} \\ &= \frac{L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{f(\mathbf{y} | \mathbf{X})}.\end{aligned}\tag{6}$$

# Distribuição *a posteriori*

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

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) &= \frac{L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{\int_{\beta_1} \cdots \int_{\beta_p} \int_{\sigma^2} L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2) d\beta_1 \cdots d\beta_p d\sigma^2} \\ &= \frac{L(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times \pi(\boldsymbol{\beta}, \sigma^2)}{f(\mathbf{y} | \mathbf{X})}.\end{aligned}\tag{6}$$

# Distribuição *a posteriori*

Lembrando que,

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

# Distribuição *a posteriori*

Lembrando que,

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

*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:

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



# 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:

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





# Distribuição *a posteriori*

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

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



# Distribuição *a posteriori*

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

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

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

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

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

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

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, nós temos as seguintes *prioris* conjugadas:



# Distribuição *a posteriori*

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

$$\beta|\sigma^2 \sim N_p(\mu_0, \sigma^2 \Lambda_0^{-1}) \text{ e } \sigma^2 \sim \text{gama inversa}(a_0, b_0),$$



# Distribuição *a posteriori*

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

$$\beta|\sigma^2 \sim N_p(\mu_0, \sigma^2 \Lambda_0^{-1}) \text{ e } \sigma^2 \sim \text{gama inversa}(a_0, b_0),$$

em que  $a_0 = v_0/2$  e  $b_0 = v_0 s_0^2/2$ ,  $a_0$ ,  $b_0$ ,  $\mu_0$ ,  $\Lambda_0^{-1}$  são hiper-parâmetros.



## Distribuição *a posteriori*

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

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

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  $U \sim \text{gama inversa}(a, b)$ , 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  $U \sim \text{gama inversa}(a, b)$ , 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  $U \sim \text{gama inversa}(a, b)$ , 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,

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



## Distribuição *a posteriori*

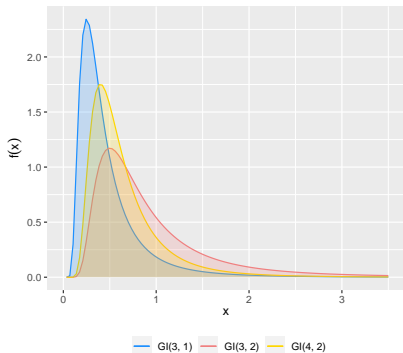
Recordando que, se  $U \sim \text{gama inversa}(a, b)$ , sua FDP é dada por:

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

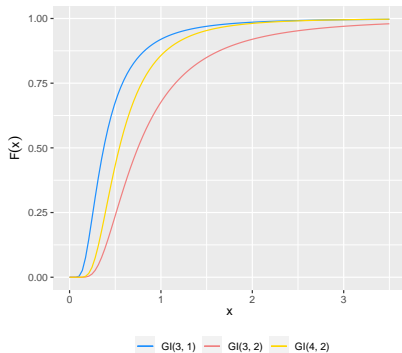
com  $a \in \mathbb{R}^+$  e  $b \in \mathbb{R}^+$ . E também,

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





(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:

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



## Distribuição *a posteriori*

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

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

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



## Distribuição *a posteriori*

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

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

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





# Distribuição *a posteriori*

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

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

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

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



## Distribuição *a posteriori*

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

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

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

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



# Algoritmo

Passo 0. Definir  $a_0, b_0, \mu_0, \Lambda_0^{-1}$  e calcular  $a_n, b_n, \mu_n, \Lambda_n^{-1}$ ;



# Algoritmo

Passo 0. Definir  $a_0$ ,  $b_0$ ,  $\mu_0$ ,  $\Lambda_0^{-1}$  e calcular  $a_n$ ,  $b_n$ ,  $\mu_n$ ,  $\Lambda_n^{-1}$ ;

Passo 1. Simular  $i$  valores de  $\beta$  e  $\sigma^2$ , 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  $\sigma^2$ , 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  $\sigma^2$ , da seguinte forma:

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



# 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  $\boldsymbol{\beta}$  e  $\sigma^2$ , da seguinte forma:

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

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  $\sigma^2$ , da seguinte forma:

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

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





# 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  $\sigma^2$ , da seguinte forma:

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

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



# Roteiro

- 1 Introdução
- 2 Método bayesiano
- 3 Aplicação**
- 4 Agradecimentos
- 5 Referências bibliográficas



# Aplicação

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

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



# 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  $Y_\ell$ : horas trabalhadas,  $x_{\ell 2}$ : tamanho do lote,  $\ell = 1, 2, \dots, 10$  e  $QMR_{\text{Res}} = 7,5$ .



# 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  $Y_\ell$ : horas trabalhadas,  $x_{\ell 2}$ : tamanho do lote,  $\ell = 1, 2, \dots, 10$  e  $QMR_{\text{Res}} = 7,5$ .



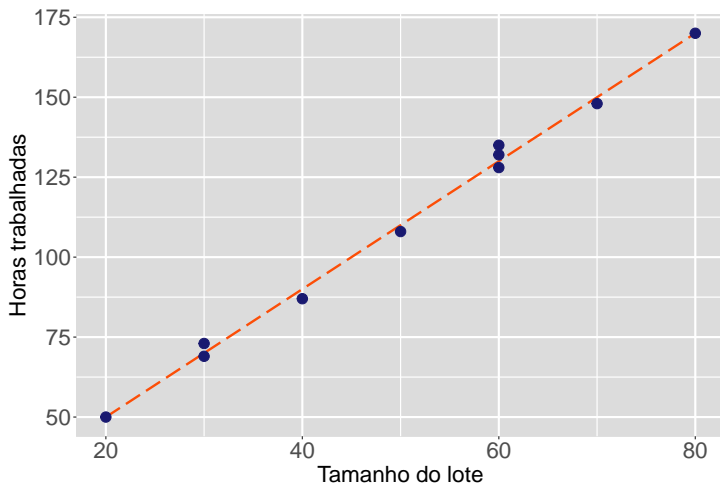


Figura 2: GD com a reta ajustada para os dados *Westwood Company*.

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;
- $\hat{\beta}_2 = 2$ : 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.

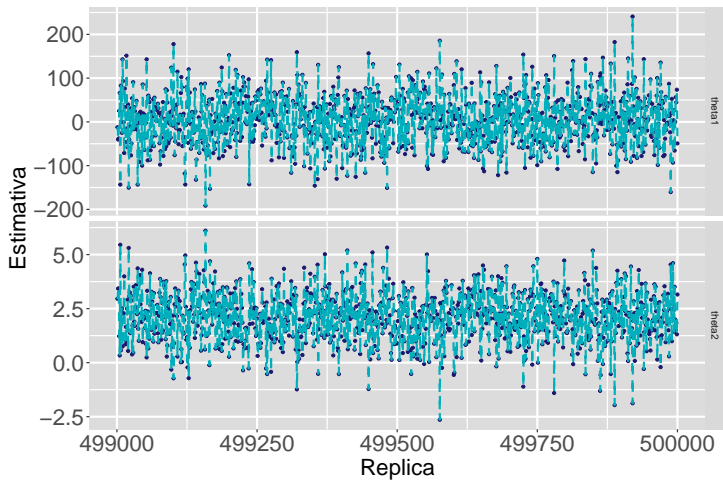


Figura 3: Estimativas observadas.

# Aplicação

Nós temos os seguintes modelos estimados,

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

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

# Roteiro

- 1 Introdução
- 2 Método bayesiano
- 3 Aplicação
- 4 Agradecimentos**
- 5 Referências bibliográficas



# Agradecimentos

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



# Roteiro

- 1 Introdução
- 2 Método bayesiano
- 3 Aplicação
- 4 Agradecimentos
- 5 Referências bibliográficas**



# 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](mailto:tiago.magalhaes@ufjf.br)

📄 [ufjf.br/tiago\\_magalhaes](https://ufjf.br/tiago_magalhaes)

🌐 Departamento de Estatística, Sala 319

