Bioinformática

Introdução a Modelagem

Barretos, 2025

Análise Estatística

Descritiva

Descrever uma coletividade dada, sem objetivo de generalizar as conclusões.

Inferencial

Baseando-se no estudo das amostras, chegar a **conclusões** que dizem respeito a população

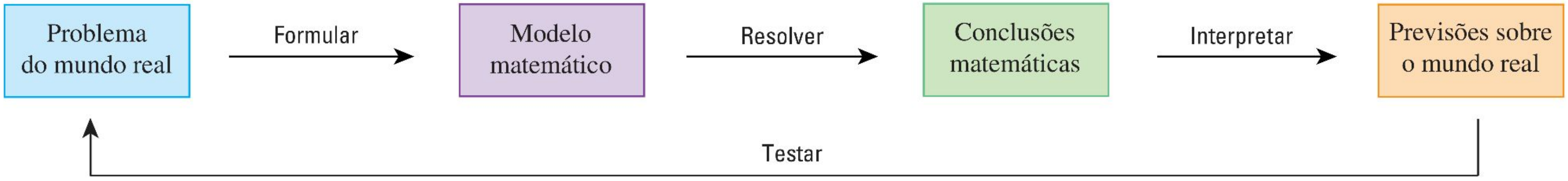
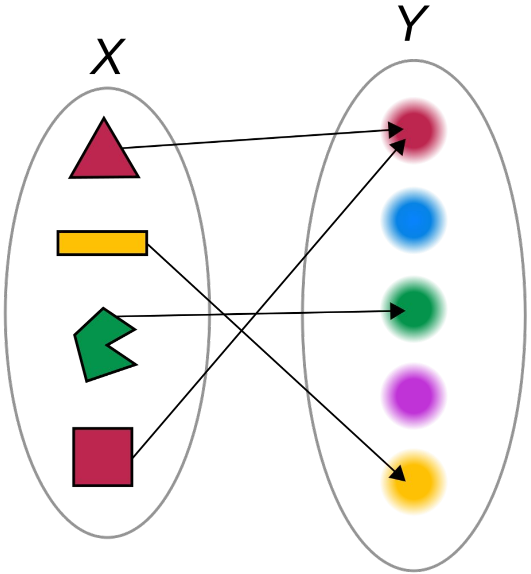


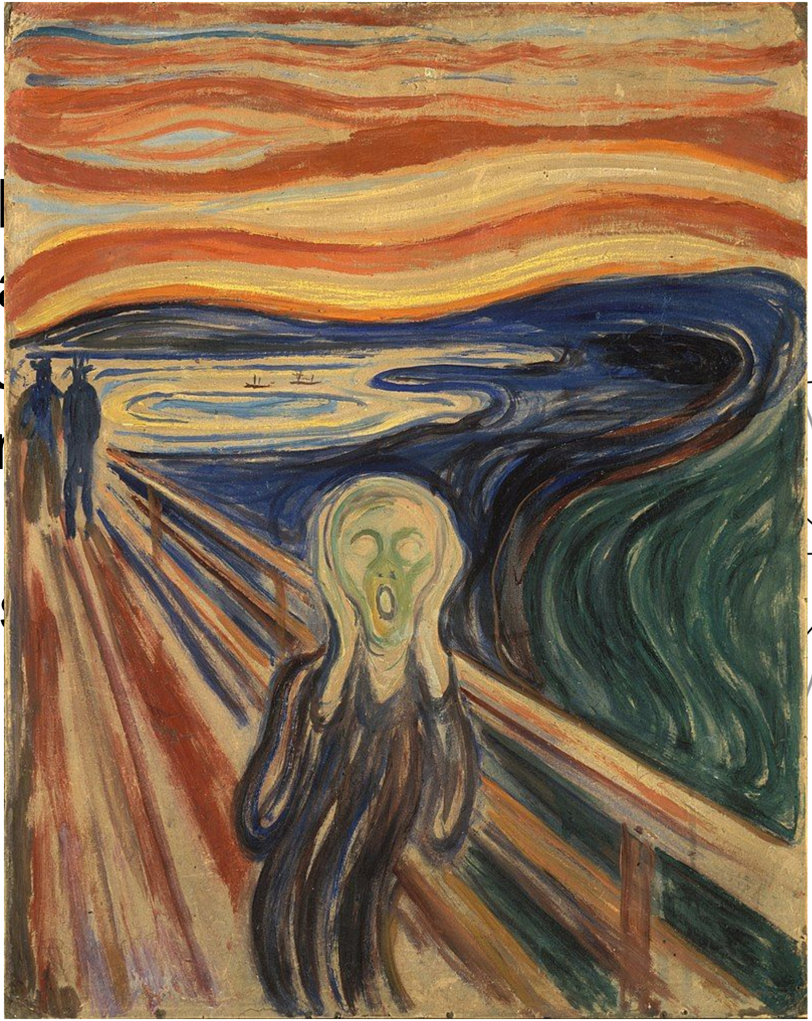
FIGURA 1 Processo de modelagem

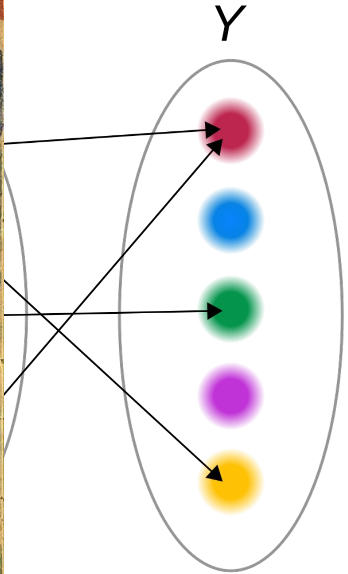
O que é um modelo ?

Função:
Se uma variável y depende de uma variável x de tal modo que cada valor de x determina exatamente um valor de y, então dizemos que y é função de x



Função:
Se uma variável
depende de uma
x de tal modo qu
valor de x deter
exatamente um
y, então dizemos
função de x

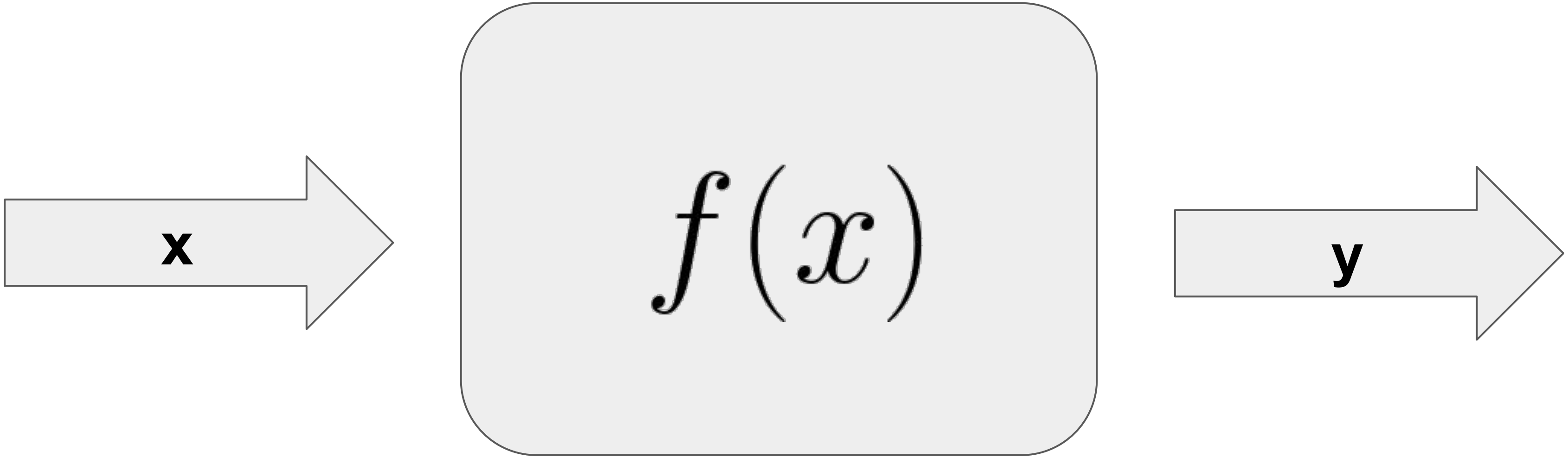




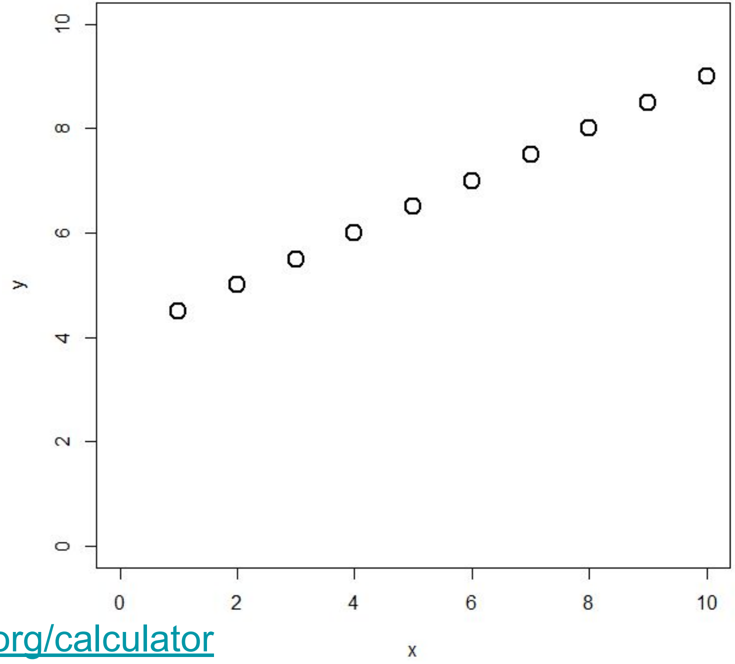

Variável Independente
Função (Modelo)
Variável Dependente


Variável Preditora
Função (Modelo)
Variável Desfecho


CAUSA
Variável Preditora
Função (Modelo)
EFEITO
Variável Desfecho


x
$f(x)$
y

x → $f(x) = a + b \cdot x$ → y

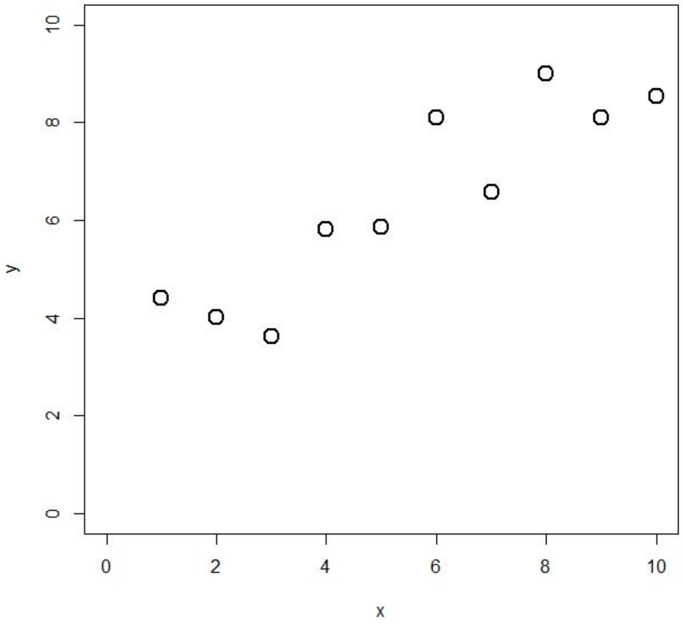


Gráficos: https://www.geogebra.org/calculator

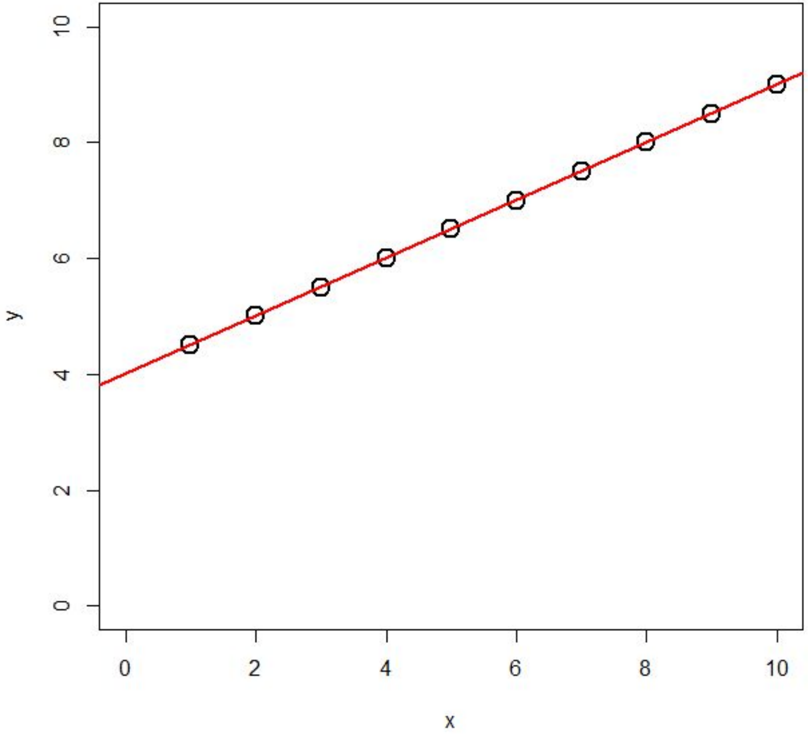
x

$$f(x) = a + b \cdot x$$

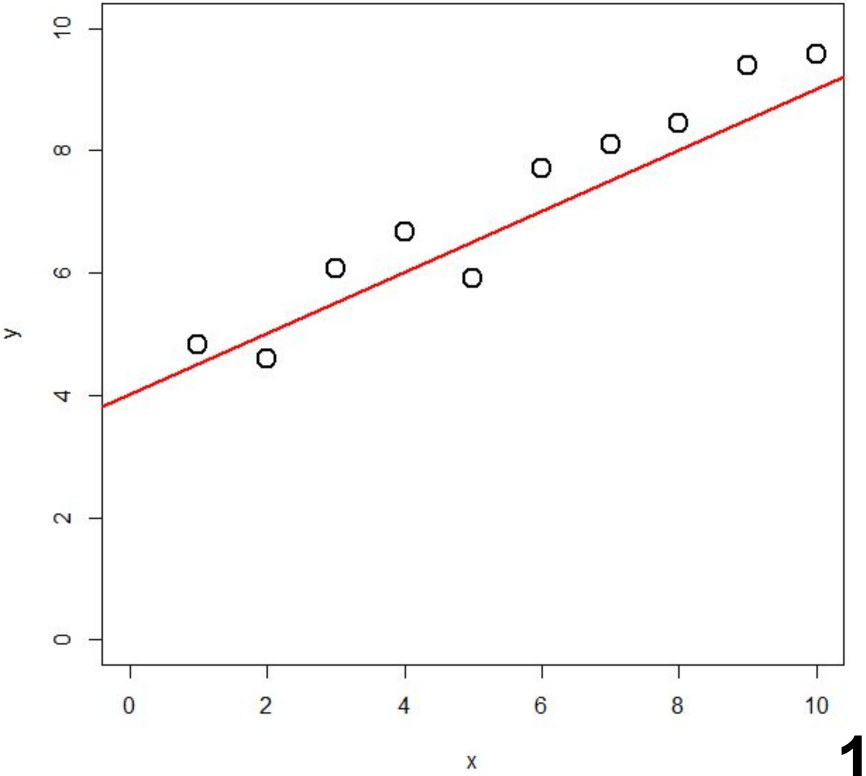
y

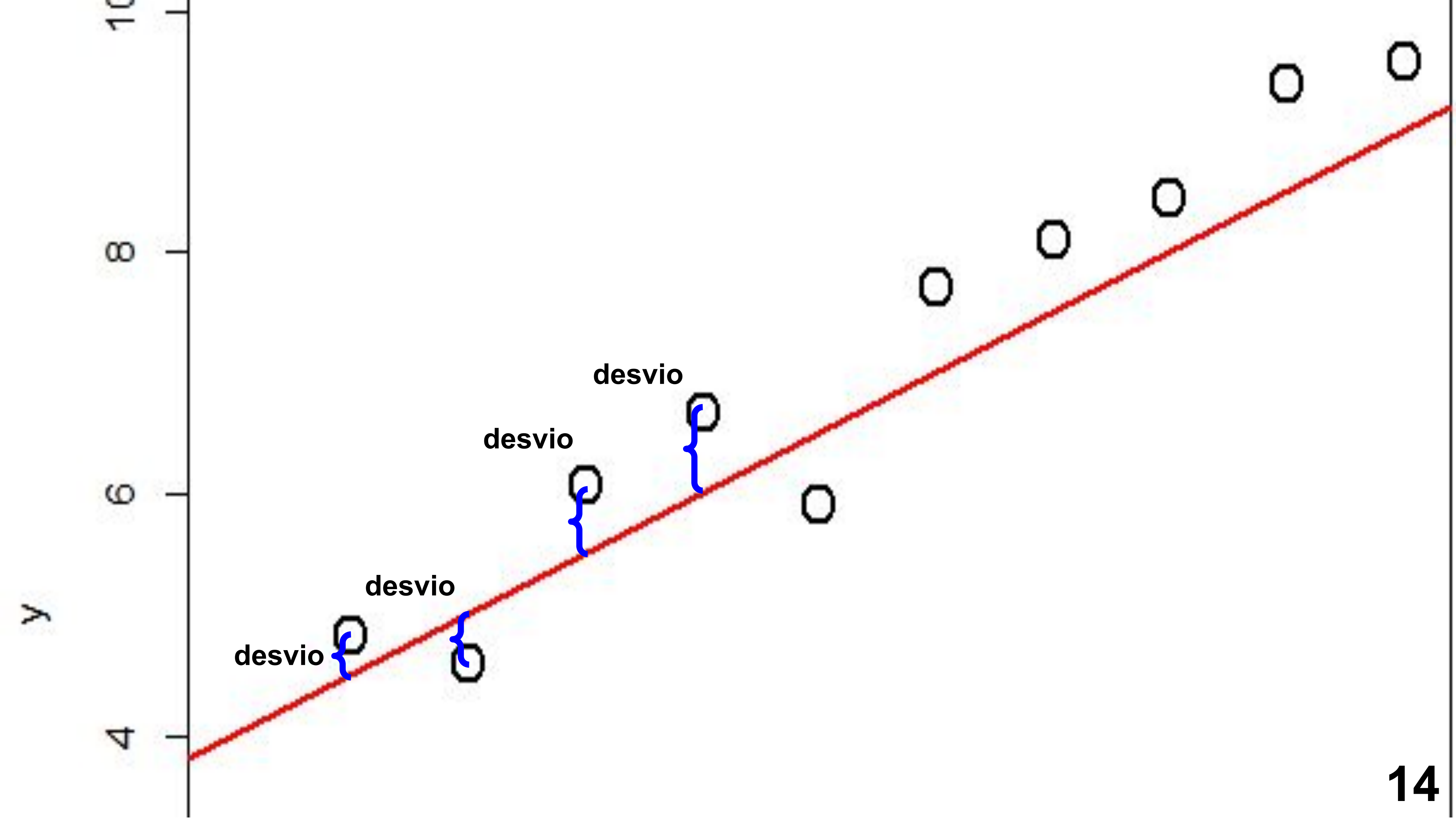


Determinístico



Aleatório (probabilístico)




desvio
desvio
desvio
desvio
8
6
4
y

Desvios

$$y - f(x) =$$

Desvios

$$y - f(x) =$$

$$= y - [a + b \cdot x]$$

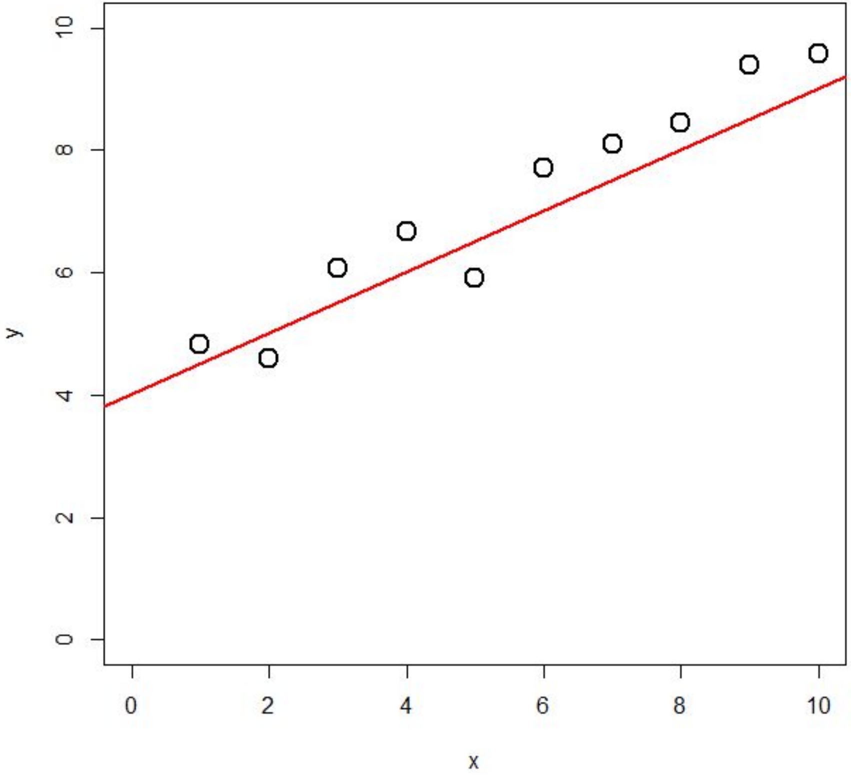
Desvios

$$y - f(x) =$$

$$= y - [a + b \cdot x]$$

$$\epsilon = (y - a - b \cdot x)$$

$$y = a + b \cdot x + \epsilon$$



Regressão Linear Simples

Modelo de Regressão Linear Simples

$$y = a + b \cdot x + \epsilon$$

$$y = \beta_0 + \beta_1 \cdot x + \epsilon$$

$$y = \alpha + \beta \cdot x + \epsilon$$

Modelo de Regressão Linear Simples

$$y = \boxed{a} + \boxed{b} \cdot x + \epsilon$$

$$y = \boxed{\beta_0} + \boxed{\beta_1} \cdot x + \epsilon$$

$$y = \boxed{\alpha} + \boxed{\beta} \cdot x + \epsilon$$

Parâmetros

		Regressor
Observation, i	Response, y	x_1
1	y_1	x_{11}
2	y_2	x_{21}
$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}

Estimação dos parâmetros

Modelo de regressão populacional

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modelo de regressão populacional

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modelo de regressão amostral

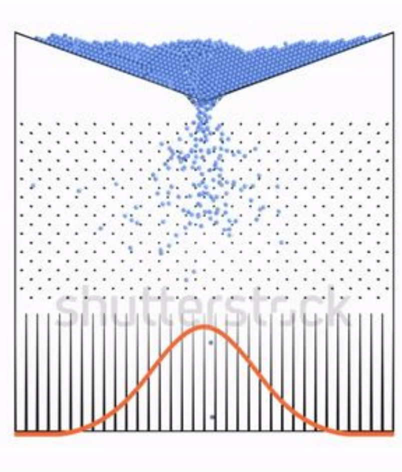
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Modelo de Regressão Linear Simples

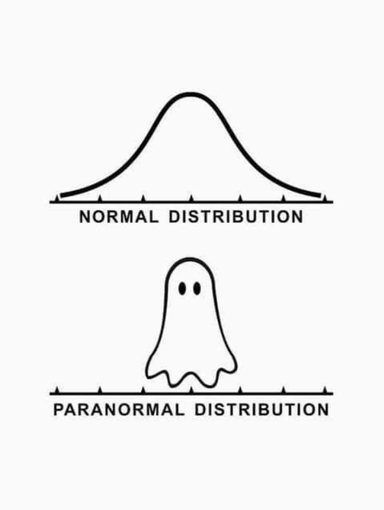
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Modelo de Regressão Linear Simples

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$



$$\epsilon_i \sim N(0, \sigma^2)$$



O modelo que mais se aproxima dos dados (verossímil)

Estimar os parâmetros

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Métodos de estimação

1. Mínimos quadrados (*least-squares*);

2. Máxima verossimilhança (*maximum likelihood*);

Métodos de estimação

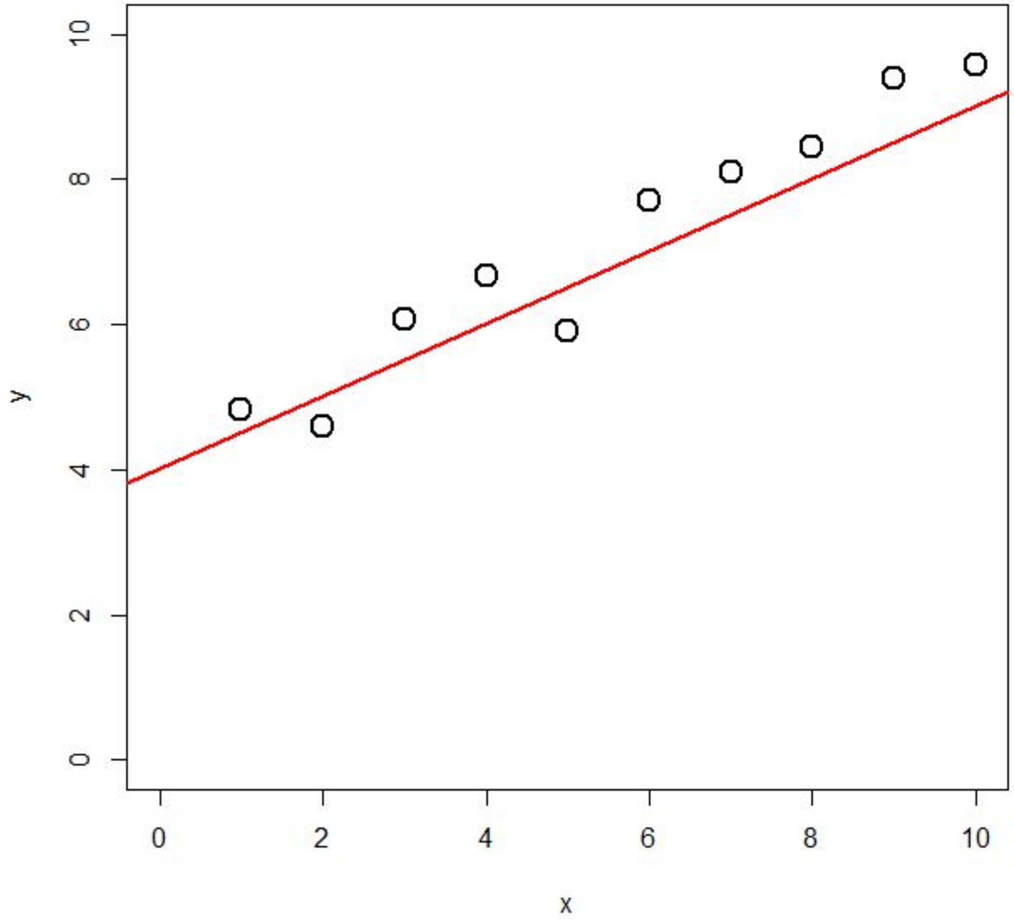
1. Mínimos quadrados (*least-squares*);

2. Máxima verossimilhança (*maximum likelihood*);

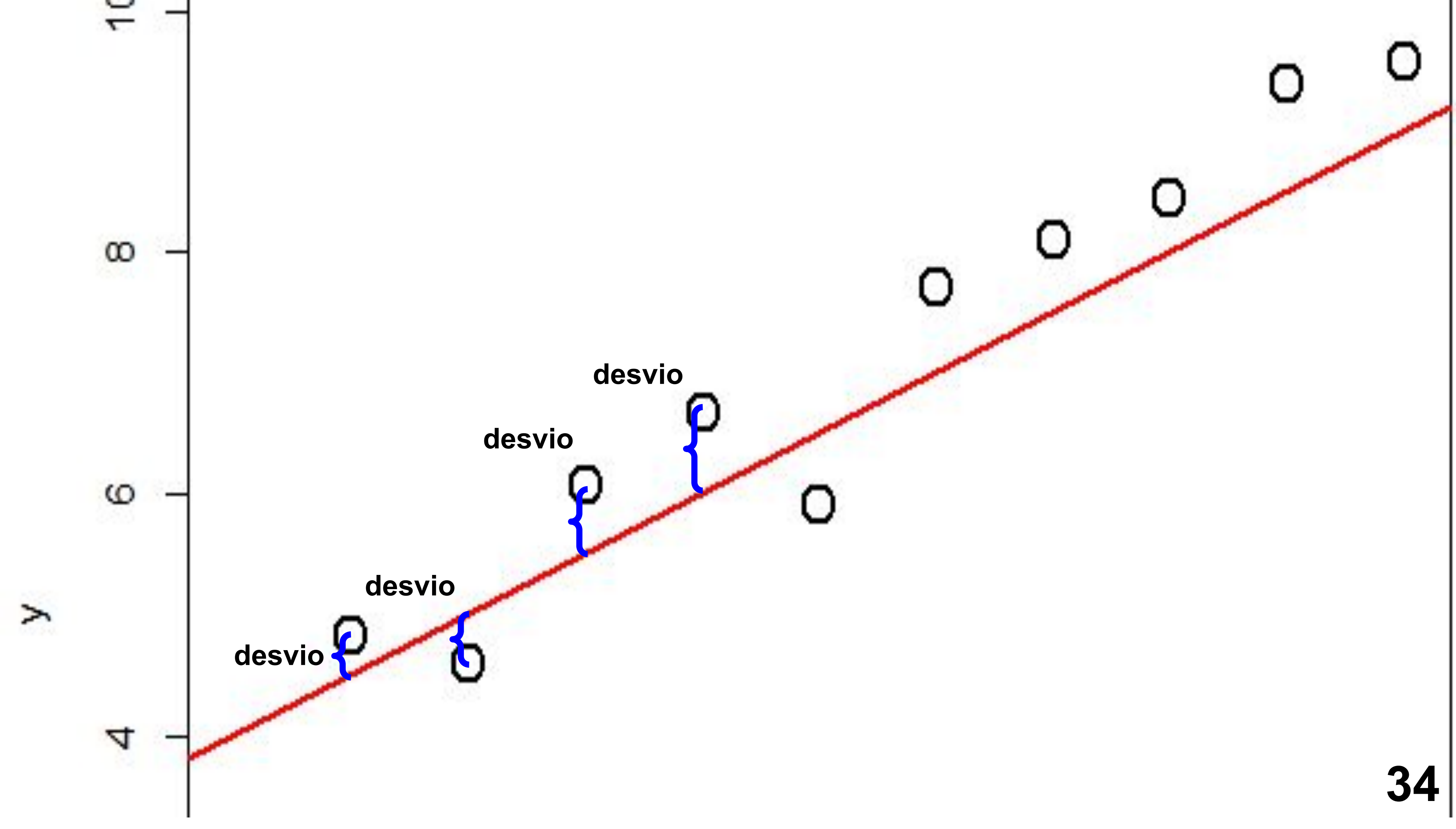
Conhecimentos a priori:

Matemática algébrica (somatória)
Cálculo diferencial (derivada)
Probabilidade (função de densidade de probabilidade)

Contextualização didática



y
0
2
4
6
8
10
x


10
8
6
4
y
desvio
desvio
desvio
desvio

Desvios

Modelo de regressão populacional

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modelo de regressão amostral

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$e_i = y_i - \hat{y}_i$$

Desvios

Modelo de regressão populacional

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modelo de regressão amostral

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \end{aligned}$$

Desvios

Modelo de regressão populacional

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Modelo de regressão amostral

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

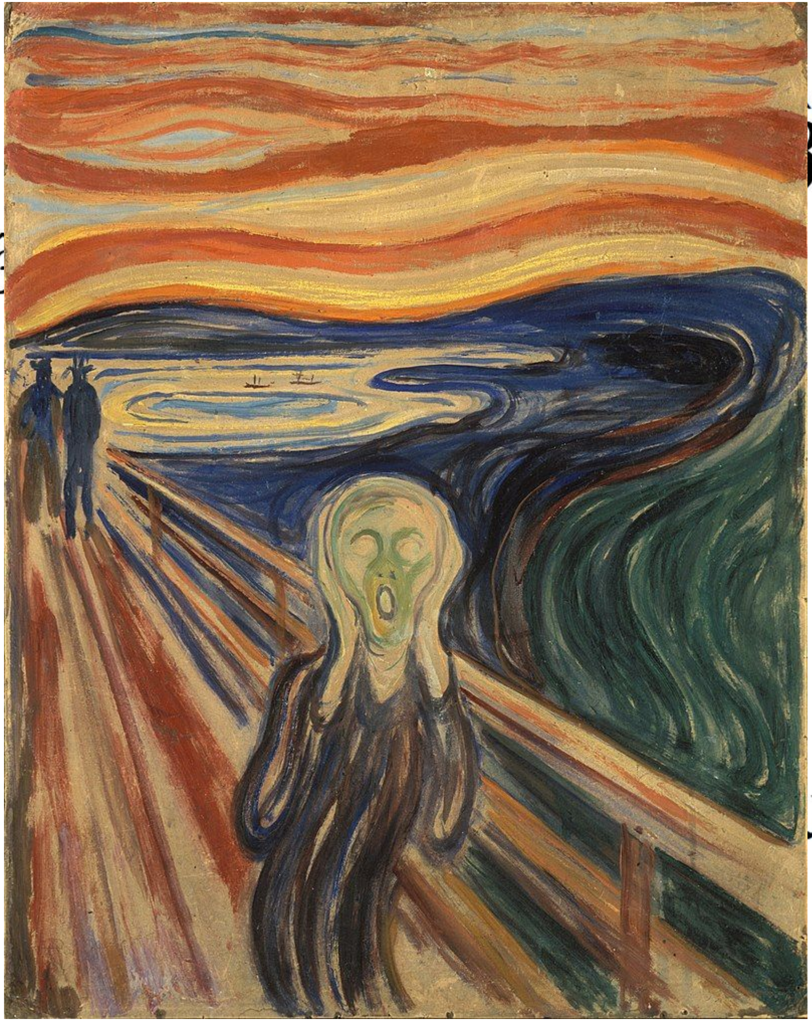
$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \end{aligned}$$

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0$$

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0,}$$



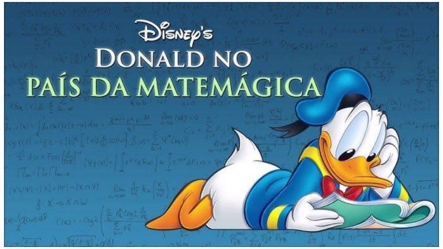
$$_1 x_i \Big) = 0$$

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1}$$

$$x_i \Big) x_i = 0$$

Equação Matemática

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$



Equação Matemática

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

Equação Matemática

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

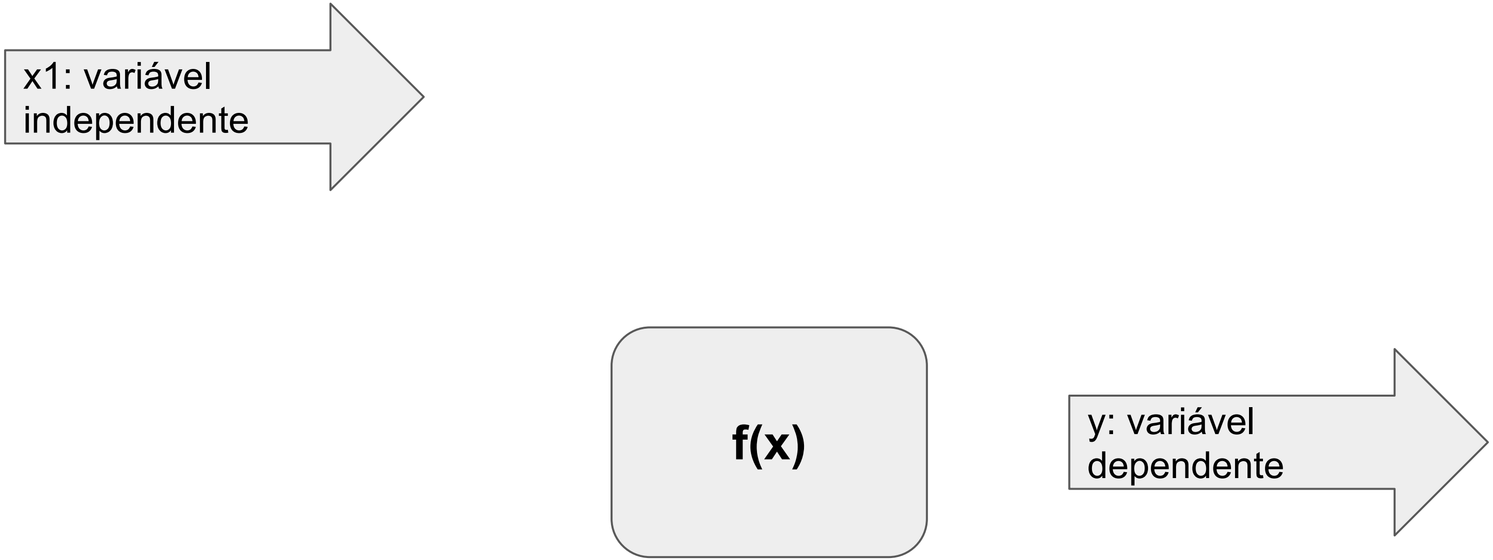
Equação Matemática

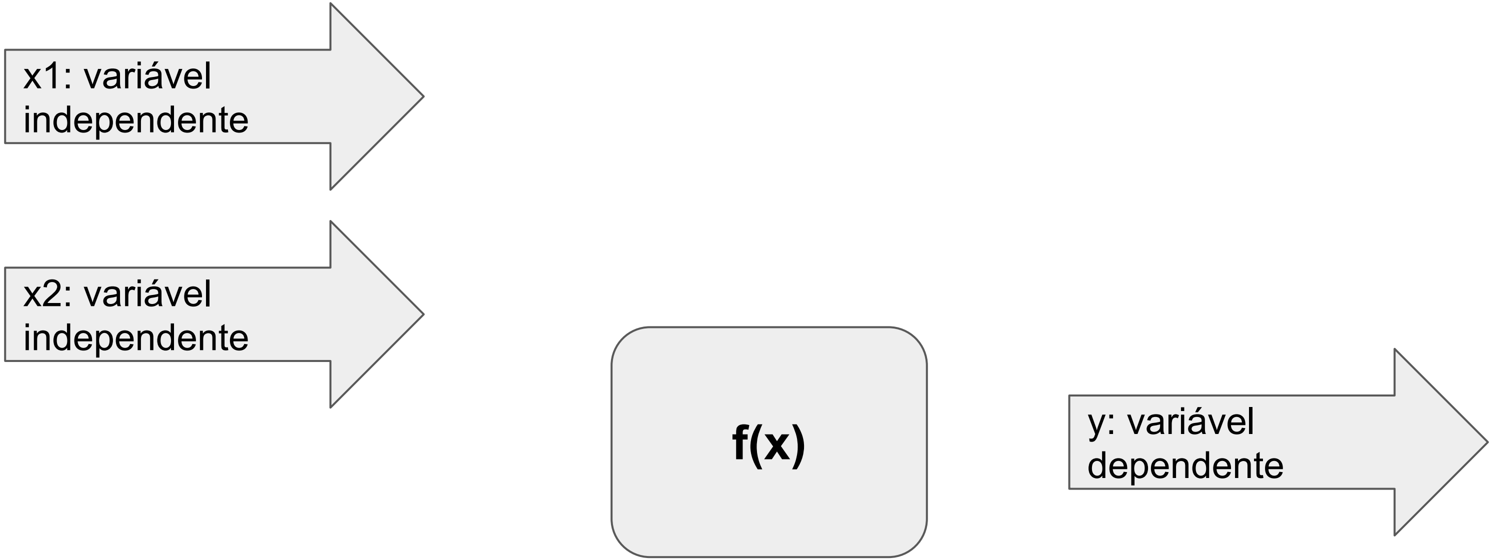
$$SS_{\text{Res}} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

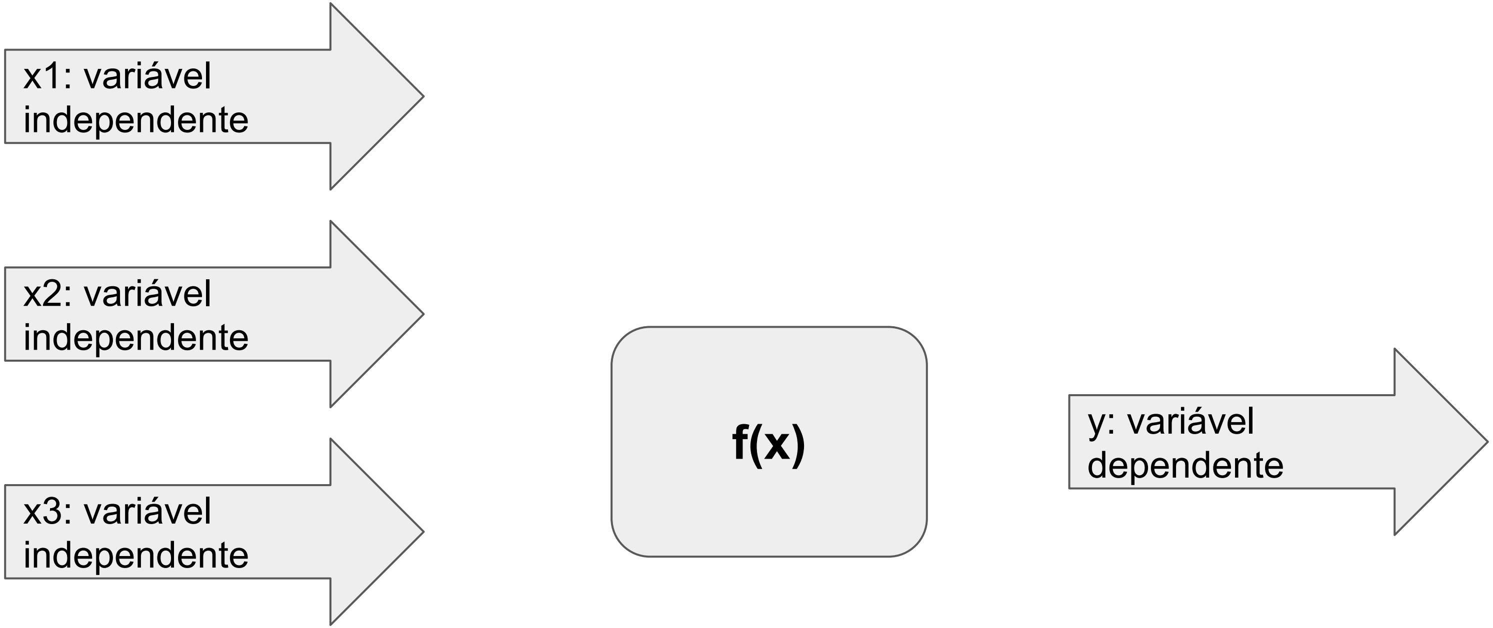
$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = MS_{\text{Res}}$$

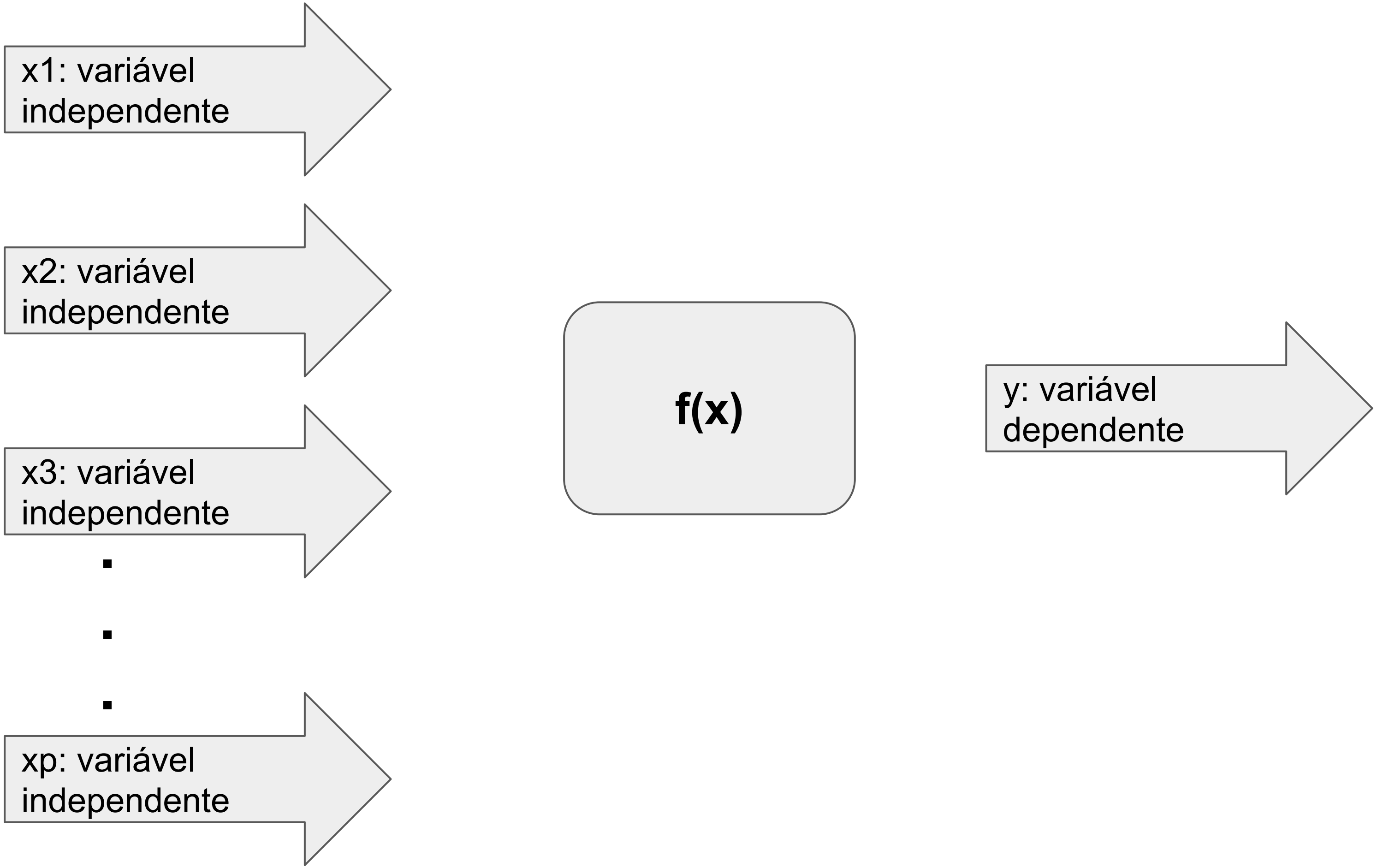
Prática 1

Regressão Linear Múltipla


x1: variável independente
f(x)
y: variável dependente


x1: variável independente
x2: variável independente
f(x)
y: variável dependente


x1: variável independente
x2: variável independente
x3: variável independente
f(x)
y: variável dependente


x1: variável independente
x2: variável independente
x3: variável independente
.
.
.
xp: variável independente
f(x)
y: variável dependente

Modelo de Regressão Múltipla

$$y = \beta_0 + \beta_1 \cdot x + \epsilon$$

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon$$

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \epsilon$$

.

.

.

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \ldots + \beta_p \cdot x_p + \epsilon$$

Modelo de Regressão Múltipla

$$y = \beta_0 + \beta_1 \cdot x + \epsilon$$

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon$$

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \epsilon$$

$$\vdots$$

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \ldots + \beta_p \cdot x_p + \epsilon$$

O modelo ajustado é mais complicado, porém os princípios básicos são os mesmos da **regressão simples**.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon_i$$

		Regressors			
Observation, i	Response, y	x_1	x_2	$\ldots$	x_k
1	y_1	x_{11}	x_{12}	$\ldots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\ldots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\ldots$	x_{nk}

Revisão Matrizes

Matriz

$$\mathbf{A} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix}.$$

Matriz

- Variável que pode armazenar n informações;
- Estrutura n dimensional;
- Homogênea.

$$\mathbf{A} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix}$$

Notação matemática

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \qquad \mathbf{A} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kp} \end{pmatrix}$$

Matriz - Identidade

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

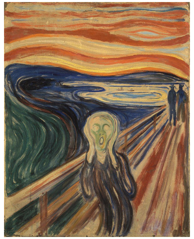
Operações Básicas - Transposição de Matriz

$$\mathbf{A} = \begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}, \qquad \mathbf{A}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}.$$

Operações Básicas - Soma

$$\begin{pmatrix} 7 & -3 & 4 \\ 2 & 8 & -5 \end{pmatrix} + \begin{pmatrix} 11 & 5 & -6 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 18 & 2 & -2 \\ 5 & 12 & -3 \end{pmatrix}$$

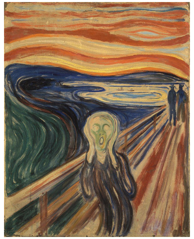
Operações Básicas - Multiplicação



$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

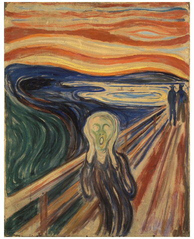
Operações Básicas - Multiplicação



$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

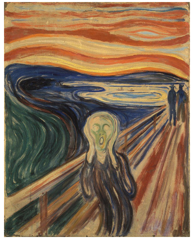
Operações Básicas - Multiplicação



$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

Operações Básicas - Multiplicação



$$\mathbf{AB} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

Operações Básicas - Multiplicação

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 6 & 5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{pmatrix}.$$

$$\mathbf{AB} = \begin{pmatrix} 2 \cdot 1 + 1 \cdot 2 + 3 \cdot 3 & 2 \cdot 4 + 1 \cdot 6 + 3 \cdot 8 \\ 4 \cdot 1 + 6 \cdot 2 + 5 \cdot 3 & 4 \cdot 4 + 6 \cdot 6 + 5 \cdot 8 \end{pmatrix}$$

$$= \begin{pmatrix} 13 & 38 \\ 31 & 92 \end{pmatrix}$$

Operações Básicas - Multiplicação

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A} \cdot \mathbf{I} = \begin{bmatrix} 2 \cdot 1 + 3 \cdot 0 & 2 \cdot 0 + 3 \cdot 1 \\ 4 \cdot 1 + 5 \cdot 0 & 4 \cdot 0 + 5 \cdot 1 \end{bmatrix}$$

Operações Básicas - Multiplicação

$$A = \begin{bmatrix} \boxed{2 \quad 3} \\ 4 \quad 5 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A \cdot I = \begin{bmatrix} \boxed{2 \cdot 1 + 3 \cdot 0} & 2 \cdot 0 + 3 \cdot 1 \\ 4 \cdot 1 + 5 \cdot 0 & 4 \cdot 0 + 5 \cdot 1 \end{bmatrix}$$

Operações Básicas - Multiplicação

$$A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A \cdot I = \begin{bmatrix} 2 \cdot 1 + 3 \cdot 0 & \boxed{2 \cdot 0 + 3 \cdot 1} \\ 4 \cdot 1 + 5 \cdot 0 & 4 \cdot 0 + 5 \cdot 1 \end{bmatrix}$$

Operações Básicas - Multiplicação

Ela não é comutativa.

A frase “a ordem do fator não altera o produto”

$$AB \neq BA.$$

Operações Básicas - Inversa de Matriz

Pra que serve uma operação inversa ???

Escalar (“números”)

Inversa da soma

$$1 + ??? = 0 \text{ (neutro da soma)}$$

Inverso da multiplicação

$$2 * ??? = 1 \text{ (neutro da multiplicação)}$$

Operações Básicas - Inversa de Matriz

Pra que serve uma operação inversa ???

Escalar (“números”)

Inversa da soma

$$1 + (-1) = 0 \text{ (neutro da soma)}$$

Inverso da multiplicação

$$2 * \tfrac{1}{2} = 2/2 = 1 \text{ (neutro da multiplicação)}$$

Operações Básicas - Inversa de Matriz

E pra isso, temos a inversa da multiplicação de matriz

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Voltando …

Observation, i	Response, y	Regressors			
		x_1	x_2	$\dots$	x_k
1	y_1	x_{11}	x_{12}	$\dots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\dots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\dots$	x_{nk}

Cálculo Matemática

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + \varepsilon_n$$

Cálculo Matemática

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Cálculo Matemática

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Cálculo Matemática

Modelo de regressão populacional

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Modelo de regressão amostral

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

Cálculo Matemática - Matricial

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Cálculo Matemática

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$SS_{\text{Res}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}$$

$$MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n - p}$$

Prática 2

Considerações finais

1. Tudo é feito com 1 linha de código;

Considerações finais

1. Tudo é feito com 1 linha de código;
2. Operações matriciais são muito importantes para a estatística;
    a. Interpretações biológicas

Cálculo Matemática

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Observações independentes;
- Sem correlação
- Variância constante (homogeneidade)

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \qquad \varepsilon_i \sim N(0, \sigma^2) \qquad \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

- Observações são correlacionadas por um grupo (simetria composta);
- Variância constante (homogeneidade)

$$
\begin{bmatrix}
\sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2
\end{bmatrix}
$$

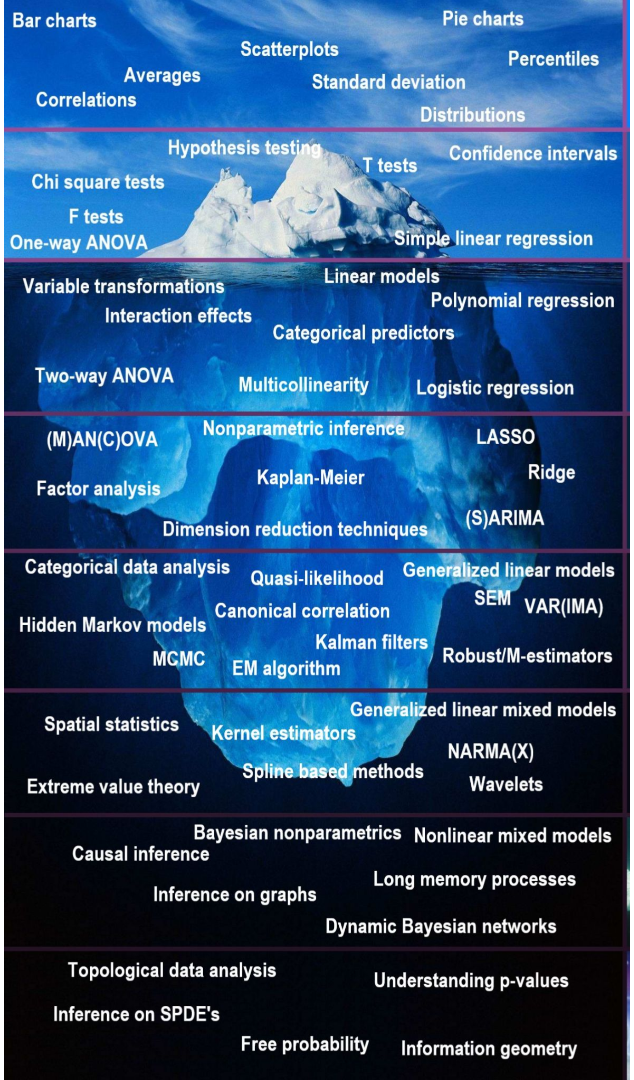
- Observações são correlacionadas por um grupo (simetria composta);
- Variância constante (homogeneidade)

Modelos de efeitos mistos

$$
\begin{bmatrix}
\sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2 & \sigma_s^2 \\
\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma^2+\sigma_s^2
\end{bmatrix}
$$

Considerações finais

1. Tudo é feito com 1 linha de código;
2. Operações matriciais são muito importantes para a estatística;
   a. Interpretações biológicas
3. Apenas um *nutshell*


Bar charts
Pie charts
Scatterplots
Percentiles
Averages
Standard deviation
Correlations
Distributions
Hypothesis testing
T tests
Confidence intervals
Chi square tests
F tests
One-way ANOVA
Simple linear regression
Variable transformations
Linear models
Polynomial regression
Interaction effects
Categorical predictors
Two-way ANOVA
Multicollinearity
Logistic regression
(M)AN(C)OVA
Nonparametric inference
LASSO
Ridge
Factor analysis
Kaplan-Meier
(S)ARIMA
Dimension reduction techniques
Categorical data analysis
Quasi-likelihood
Generalized linear models
SEM
VAR(IMA)
Canonical correlation
Hidden Markov models
Kalman filters
Robust/M-estimators
MCMC
EM algorithm
Generalized linear mixed models
Spatial statistics
Kernel estimators
NARMA(X)
Spline based methods
Wavelets
Extreme value theory
Bayesian nonparametrics
Nonlinear mixed models
Causal inference
Long memory processes
Inference on graphs
Dynamic Bayesian networks
Topological data analysis
Understanding p-values
Inference on SPDE's
Free probability
Information geometry


THIS IS FINE.

Estatística - Regressão linear múltiplo

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Ciência de dados - Regressão Ridge

$$\hat{\beta}_{\mathrm{R}} = (\mathbf{X}_{\mathrm{A}}'\mathbf{X}_{\mathrm{A}})^{-1}\mathbf{X}_{\mathrm{A}}'\mathbf{y}_{\mathrm{A}} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

Estatística - Regressão linear múltiplo

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Ciência de dados - Regressão Ridge

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = (\mathbf{X}_{\mathrm{A}}'\mathbf{X}_{\mathrm{A}})^{-1}\mathbf{X}_{\mathrm{A}}'\mathbf{y}_{\mathrm{A}} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

Ciência de dados - Suporte vetor máquina

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{H}\hat{\beta} = (\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I})^{-1}\mathbf{H}\mathbf{H}^T\mathbf{y}.$$

$$\text{kernel } \{\mathbf{H}\mathbf{H}^T\}_{i,i'} = K(x_i, x_{i'})$$

Considerações finais

1. Tudo é feito com 1 linha de código;
2. Operações matriciais são muito importantes para a estatística;
    a. Interpretações biológicas
3. Apenas um *nutshell;*
4. **Não tenham medo da matemática, atualmente podemos testar muita coisa !!!**

Referência (matemática)