



escola  
britânica de  
artes criativas  
& tecnologia

## **Profissão: Cientista de Dados**

Análise de Regressão

# O que é um modelo?

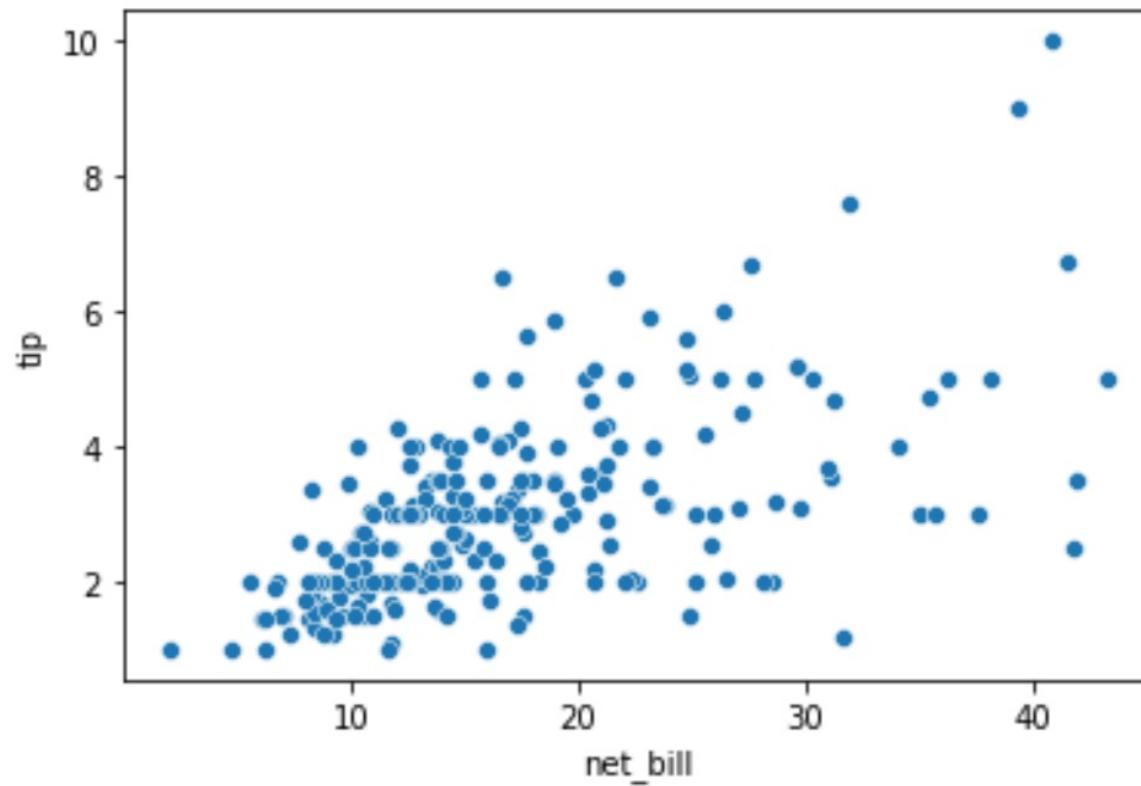
# O que é um modelo?



# Base e premissas

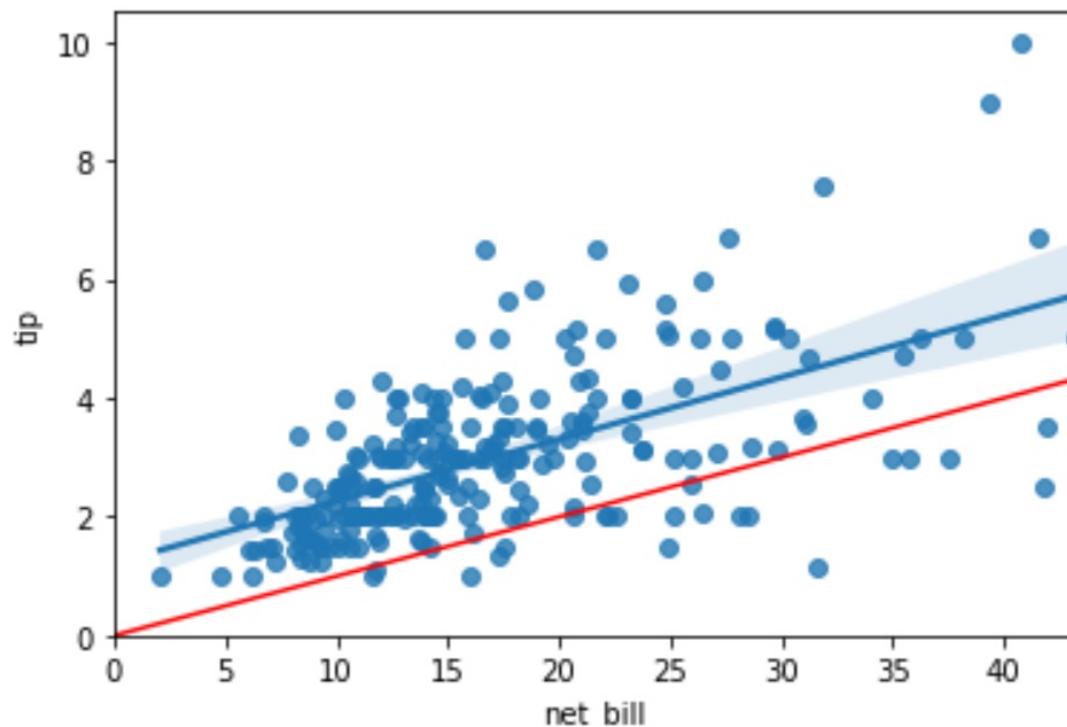
	total_bill	tip	sex	smoker	day	time	size	tip_pct
0	16.99	1.01	Female	No	Sun	Dinner	2	0.063204
1	10.34	1.66	Male	No	Sun	Dinner	3	0.191244
2	21.01	3.50	Male	No	Sun	Dinner	3	0.199886
3	23.68	3.31	Male	No	Sun	Dinner	2	0.162494
4	24.59	3.61	Female	No	Sun	Dinner	4	0.172069

# O que é um modelo?



# Introdução

# Introdução



Y = tip\_pct (% de gorjeta)  
X = net\_bill (valor da conta)

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

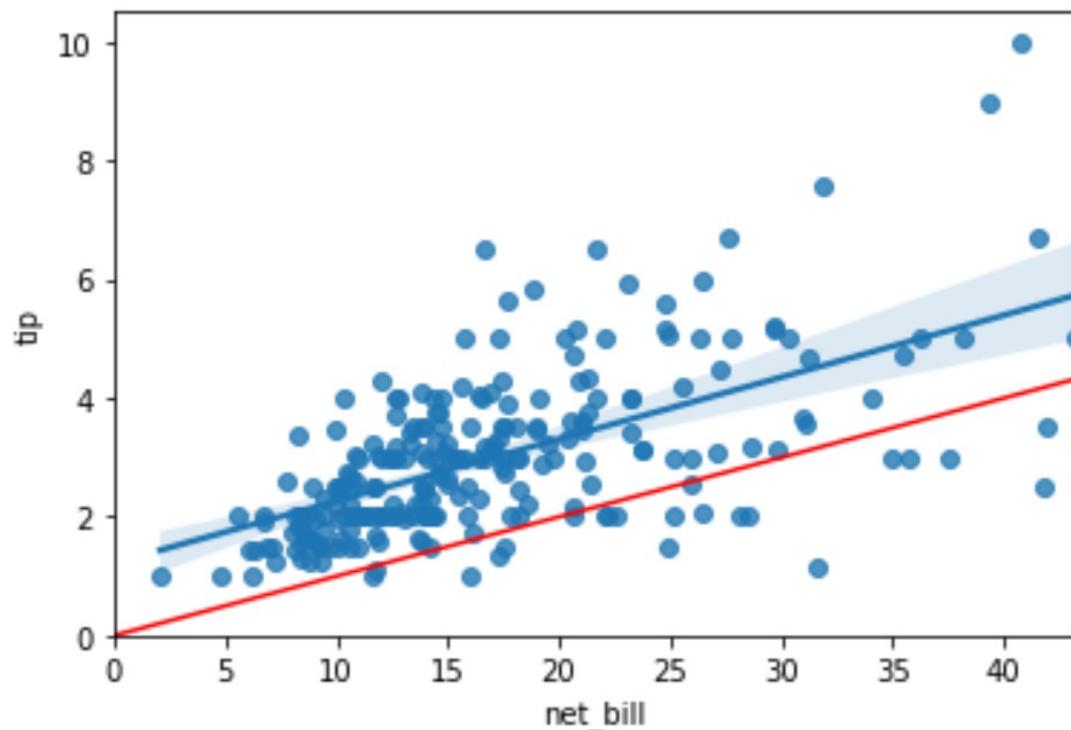
Com  $i = 1, \dots, N$   
 $\varepsilon_i \sim N(0, \sigma^2)$  i.i.d.

$\alpha$  e  $\beta$  são constantes desconhecidas.

Usando estimativas pontuais temos:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x + \hat{\varepsilon}$$

# Entendendo a equação



$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

$$\hat{\alpha} = 1,33$$

$$\hat{\beta} = 0,10$$

Se  $x = 10$ :

$$\hat{y} = 1,33 + 0,10 * 10 = 2,33$$

Se  $x = 20$ :

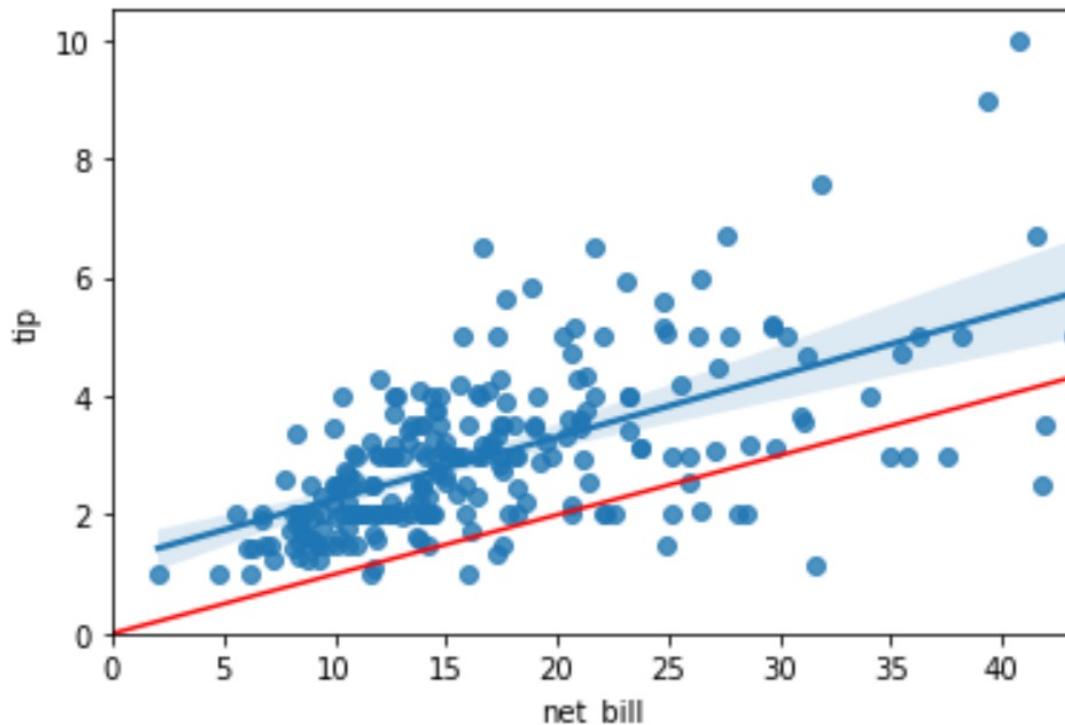
$$\hat{y} = 1,33 + 0,10 * 20 = 3,33$$

Se  $x = 30$ :

$$\hat{y} = 1,33 + 0,10 * 30 = 4,33$$

# Interpretação dos parâmetros

# Interpretação dos parâmetros



$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Com  $i = 1, \dots, N$   
 $\varepsilon_i \sim N(0, \sigma^2)$  i.i.d.

- $\alpha$  é o valor esperado de  $y$  quando  $x$  é zero.
- $\beta$  é o aumento esperado em  $y$  para cada unidade que se incrementa em  $x$ .
- $\varepsilon$  é um erro aleatório do valor de  $y$ , quando predito pelo modelo.
- $\sigma^2$  é a variância dos erros  $\varepsilon_i$

---

Isto é um  
modelo!

---

**"All models are wrong,  
but some are useful"**

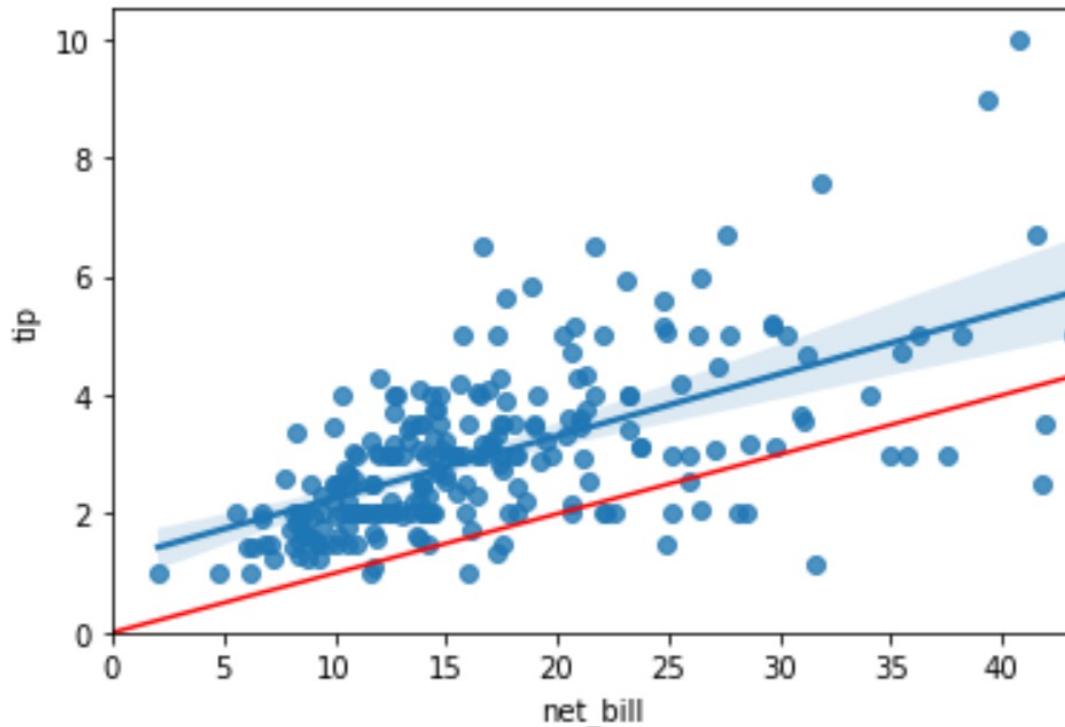
**"Todos os modelos estão errados,  
mas alguns são úteis"**

George E. P. Box



# Obtendo as estimativas

# Os erros



Y = tip\_pct (% de gorjeta)  
X = net\_bill (valor da conta)

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Com  $i = 1, \dots, N$   
 $\varepsilon_i \sim N(0, \sigma^2)$  i.i.d.

$\alpha$  e  $\beta$  são constantes desconhecidas.

Usando estimativas pontuais temos:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x + \hat{\varepsilon}$$

# Os erros

$$\begin{aligned}\hat{y}_i &= \hat{\alpha} + \hat{\beta}x_i \\ y_i &= \hat{\alpha} + \hat{\beta}x_i + \hat{\varepsilon}_i\end{aligned}\right\} y_i = \hat{y}_i + \hat{\varepsilon}_i$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad \Rightarrow \quad \hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta}x_i$$

$$QME = \frac{1}{N} \sum_i^N \hat{\varepsilon}_i^2 \quad SQE = \sum_i^N \hat{\varepsilon}_i^2$$

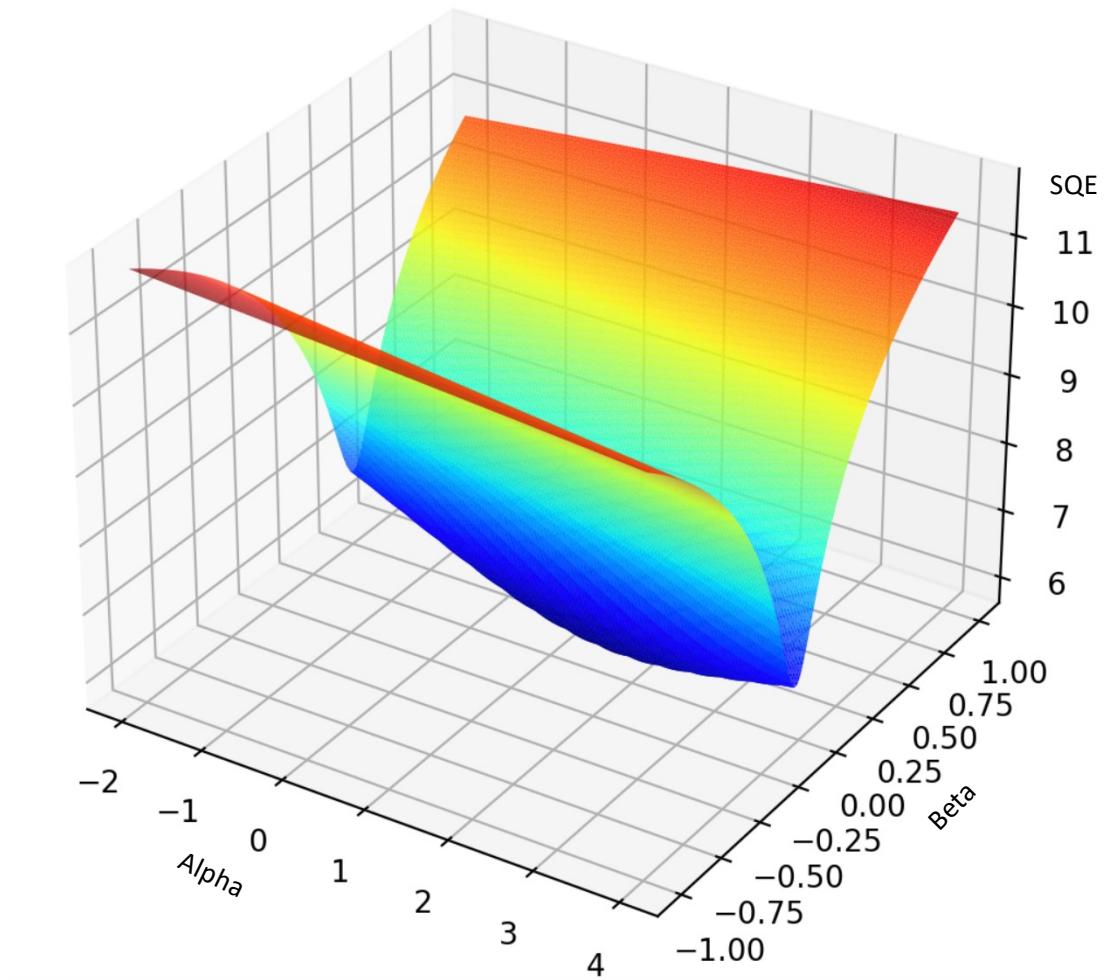
# Minimizando os erros

$$SQE = \sum_i^N (y_i - (\hat{\alpha} + \hat{\beta}x_i))^2$$

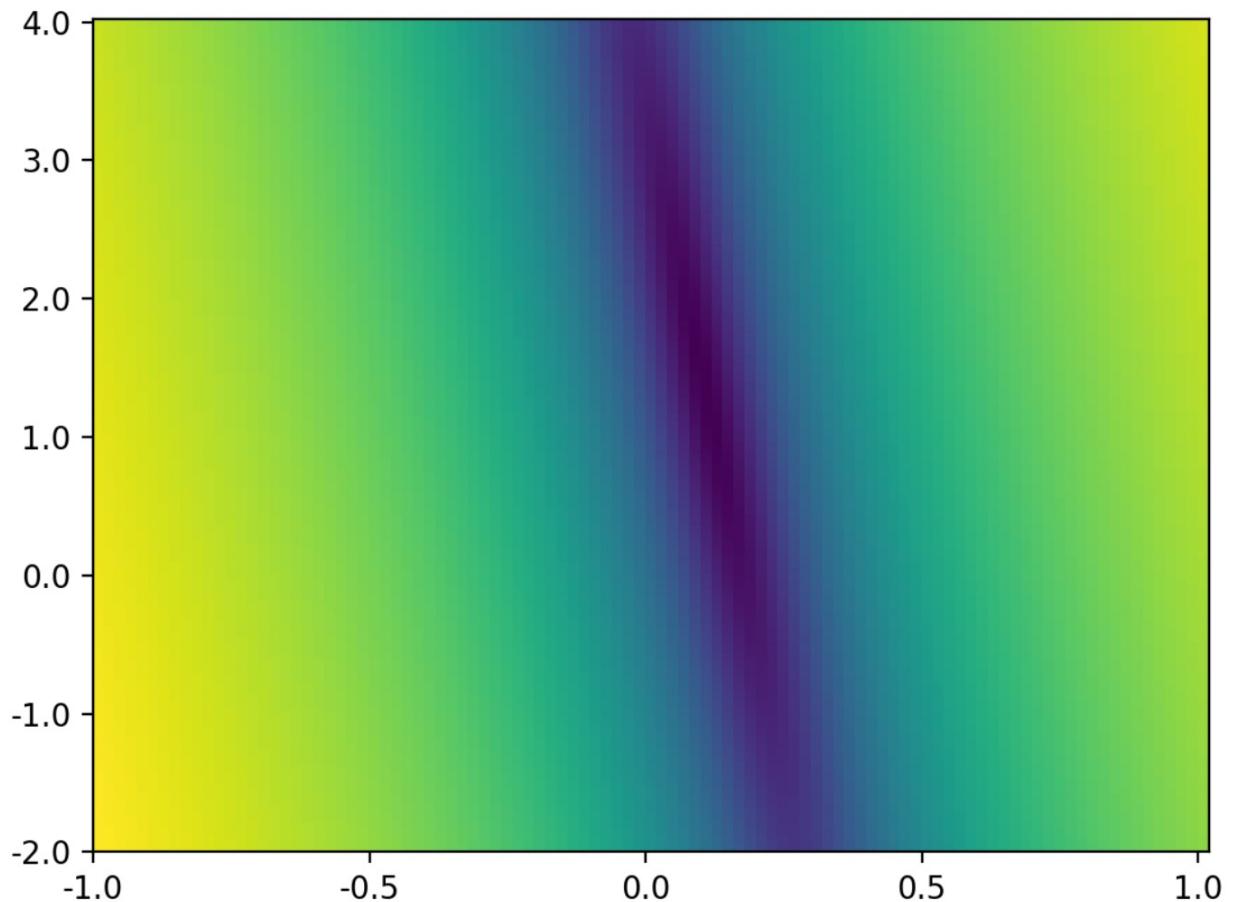
$\hat{\alpha}$  e  $\hat{\beta}$  correspondem a  $a$  e  $b$  que minimizam:

$$SQE = \sum_i^N (y_i - (a + bx_i))^2$$

# Erro em função dos parâmetros



# Erro em função dos parâmetros



# Estimador de mínimos quadrados

$$\frac{\partial \text{SQE}}{\partial \alpha} = 0 \quad \Rightarrow \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\frac{\partial \text{SQE}}{\partial \beta} = 0 \quad \Rightarrow \quad \hat{\beta} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})}$$

# Propriedades dos estimadores de mínimos quadrados

- A soma dos resíduos é zero  $\sum_{i=1}^N \hat{\varepsilon}_i = 0$
- Os resíduos não têm correlação linear com os preditores
- O ponto  $(\bar{x}, \bar{y})$  está sempre na reta
- A distribuição dos estimadores é conhecida

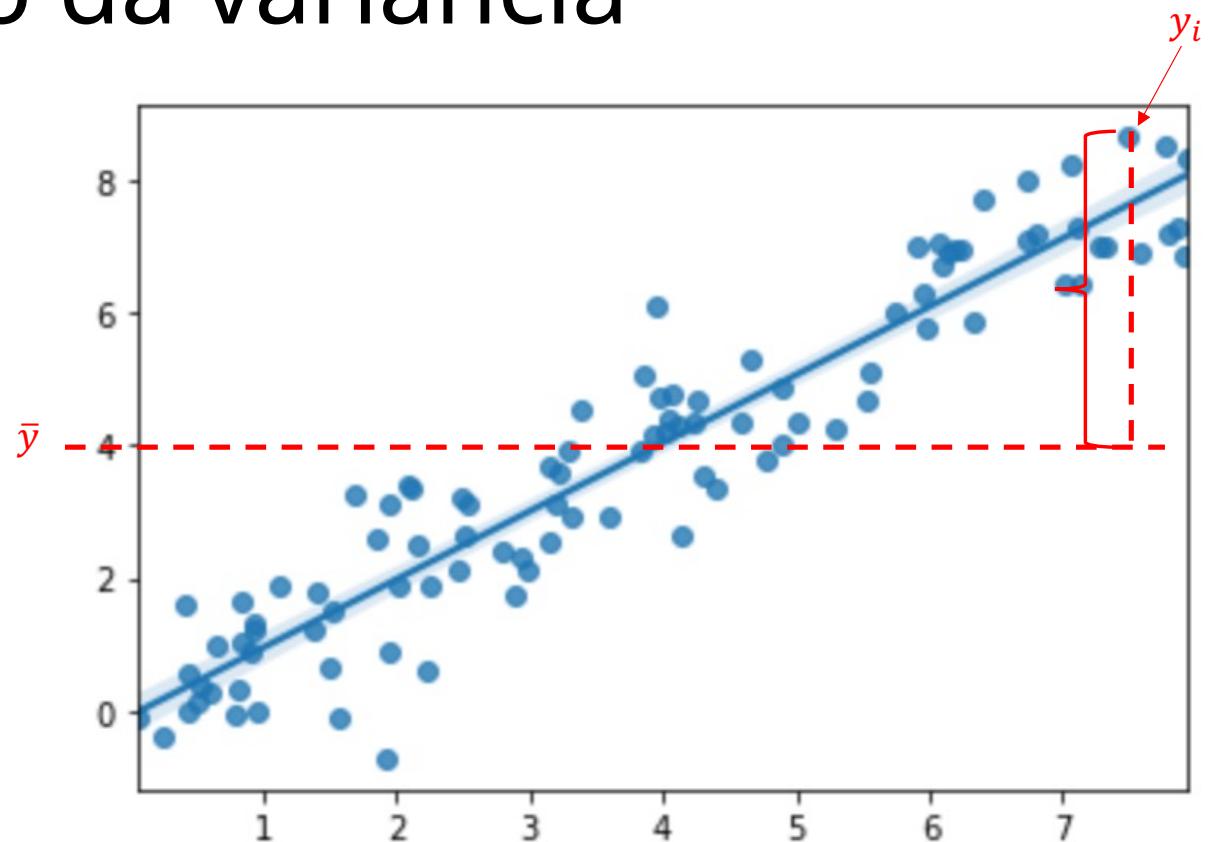
# Qualidade do modelo

# Decomposição da variância

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

$$SQM = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

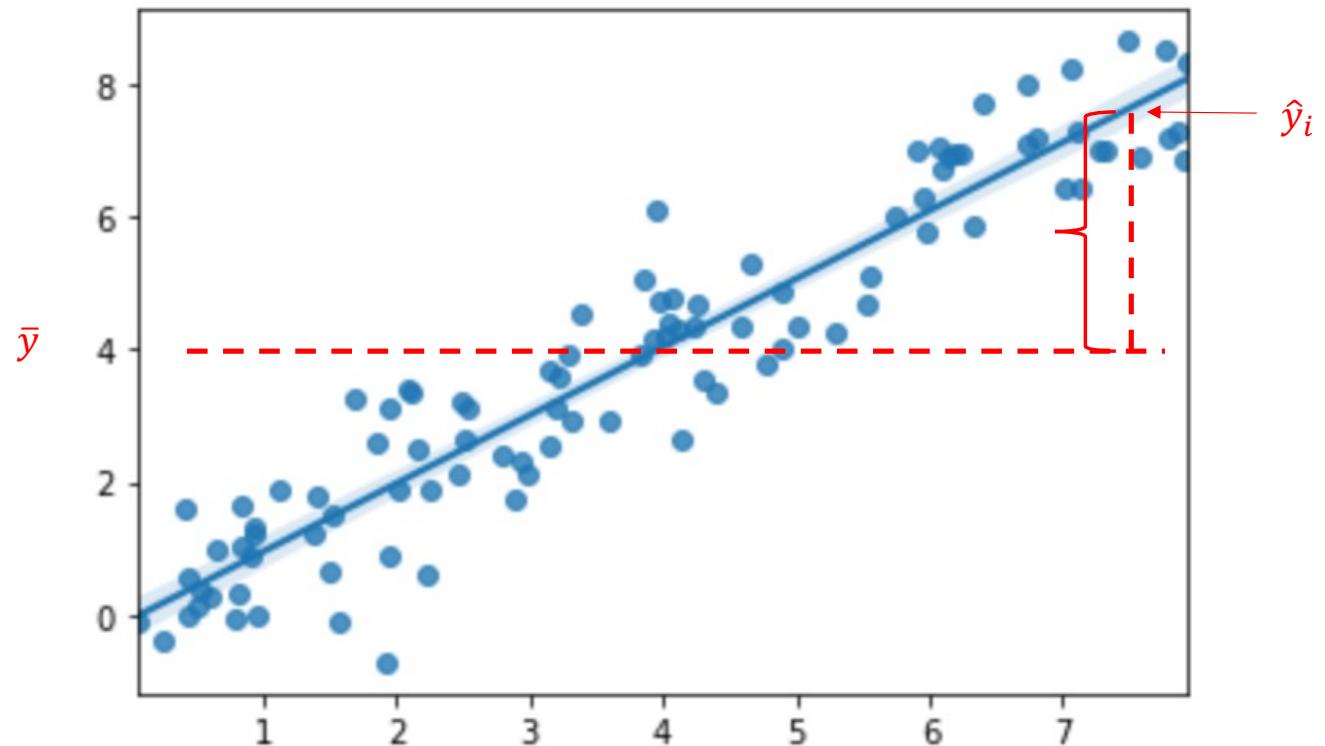


# Decomposição da variância

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

$$SQM = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



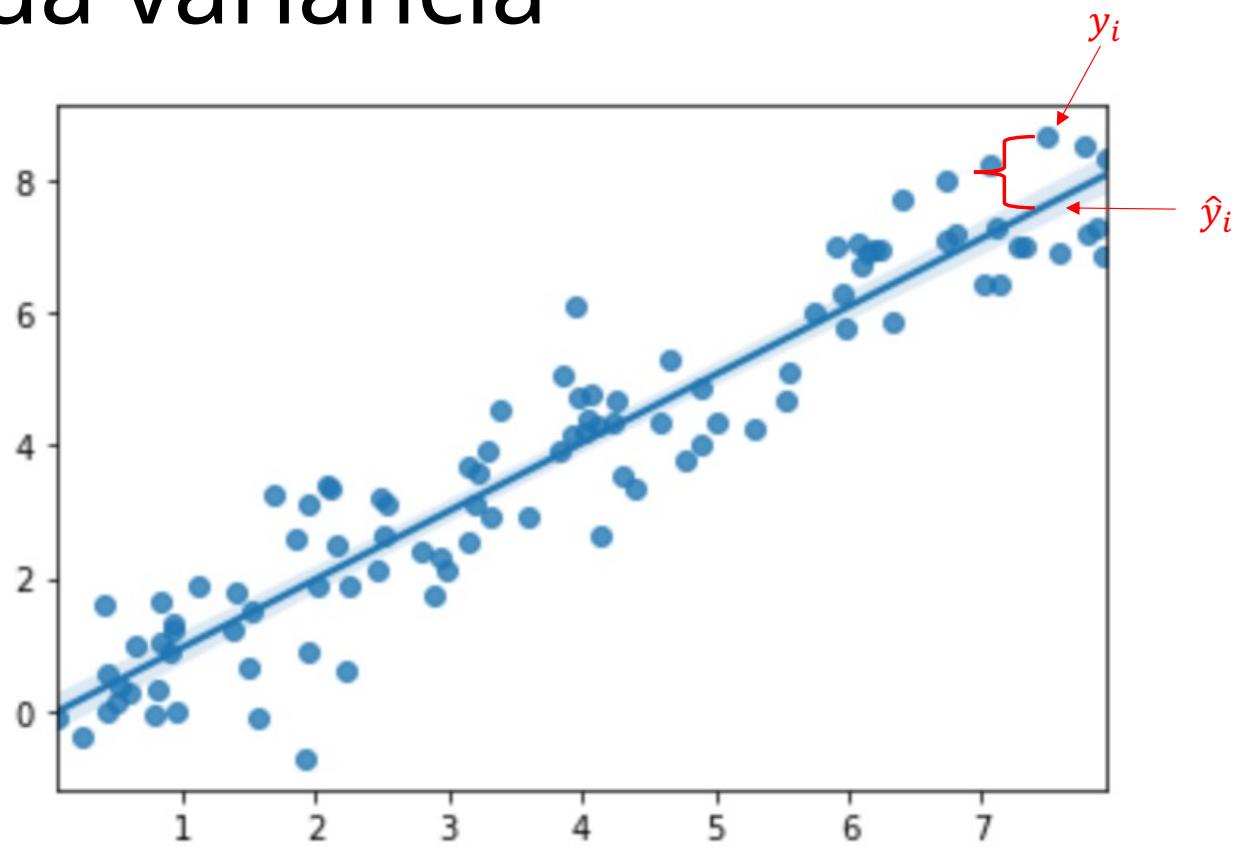
# Decomposição da variância

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

$$SQM = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SQE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

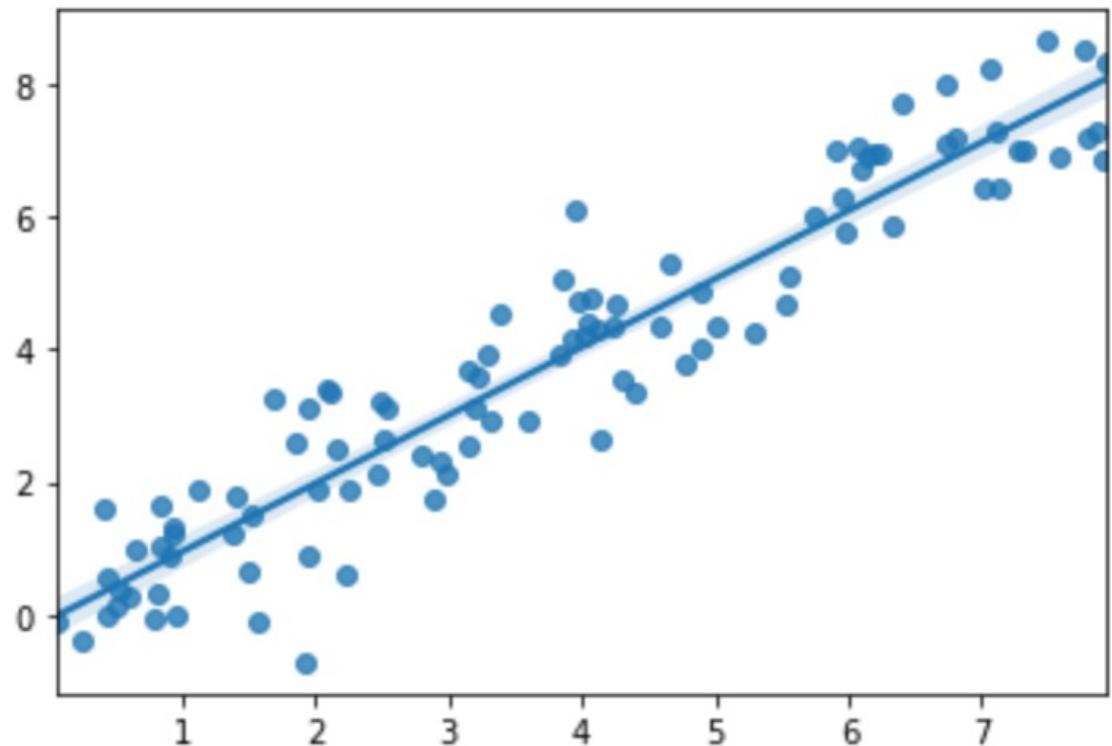
$$SQT = SQM + SQE$$



# Coeficiente de determinação

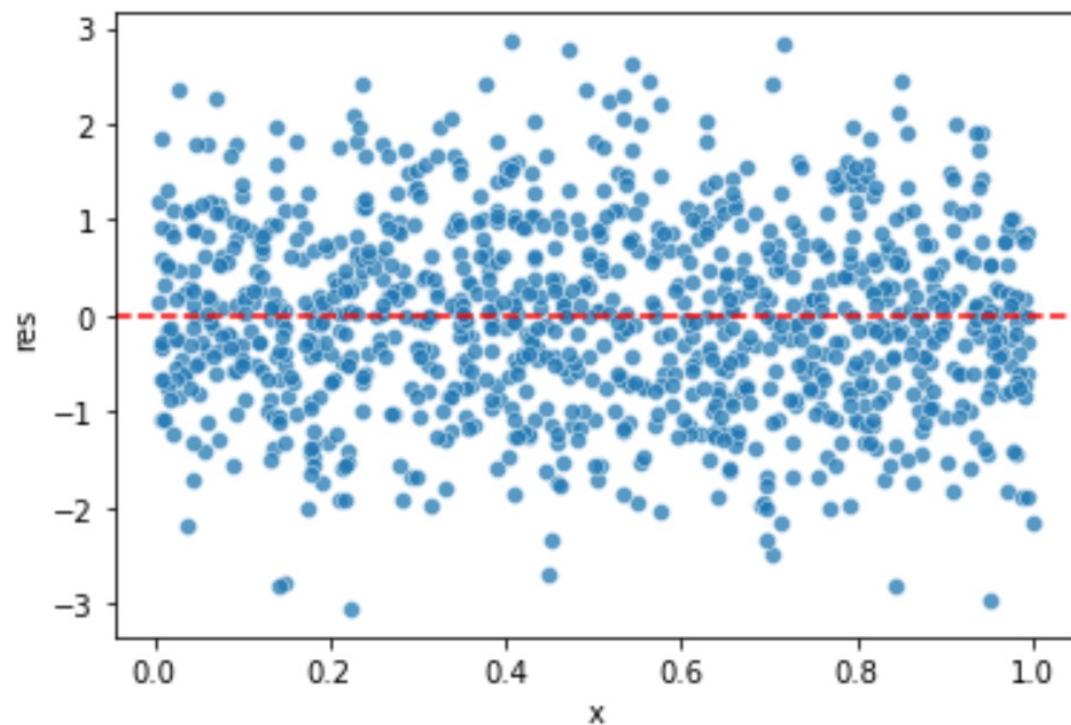
$R^2$  – coeficiente de determinação

$$R^2 = \frac{SQM}{SQT} = 1 - \frac{SQE}{SQT}$$



# Análise de Resíduos

# Análise de resíduos

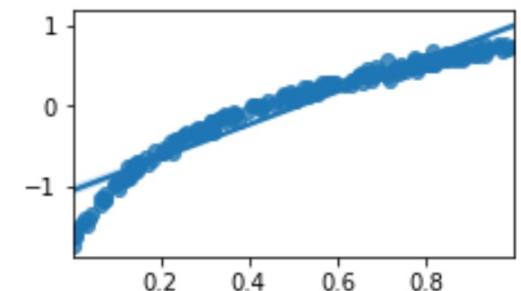
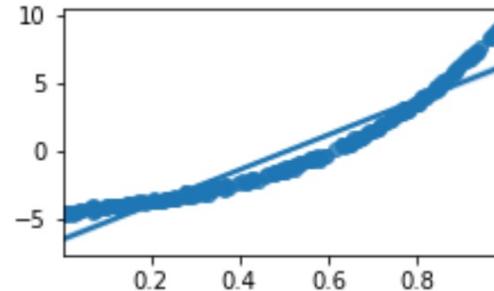
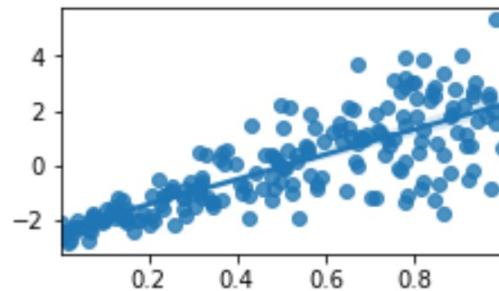


Desejado:

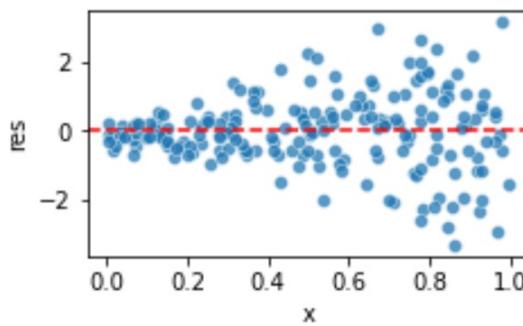
- Nenhum padrão evidente
- Aspecto de independência
- Variância uniforme

# Análise de resíduos

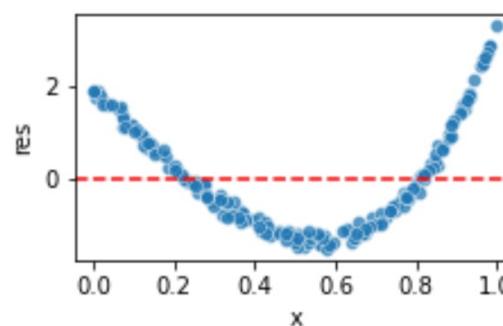
Y \* X



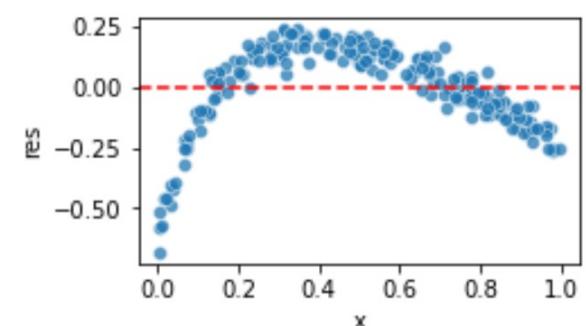
Res \* X



Resíduo aumenta



Relação convexa



Relação côncava

# Multiplicação de matrizes

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

**3 x 2**

**2 x 2**

**3 x 2**

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} (1 & 2) \\ (1 & 3) \\ (2 & 4) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

Diagram illustrating the multiplication of matrices A and B. Matrix A is a 3x2 matrix with columns (1, 2), (1, 3), and (2, 4). Matrix B is a 2x1 matrix with rows (1) and (2). The product AB is a 3x1 matrix with elements 5, 7, and 10. Red boxes highlight the first column of A and the first row of B, with arrows indicating their multiplication to produce the first element of the result matrix.

$$1 \times 1 + 2 \times 2 = 5$$

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 10 \end{pmatrix} \quad \begin{pmatrix} 11 \\ 15 \\ 22 \end{pmatrix}$$

Diagram illustrating the multiplication of matrices A and B. Matrix A is a 3x2 matrix and matrix B is a 2x2 matrix. The first element of the first row of A (1) is highlighted in green and multiplied by the first column of B (1, 2). The second element of the first row of A (2) is highlighted in green and multiplied by the second column of B (1, 2). The third element of the first row of A (3) is highlighted in blue and multiplied by the first column of B (1, 2). The fourth element of the first row of A (4) is highlighted in blue and multiplied by the second column of B (1, 2). The resulting matrix is a 3x2 matrix with elements 5, 7, 10 in the first row and 11, 15, 22 in the second row.

$$1 \times 1 + 3 \times 2 = 7$$

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

Diagram illustrating the multiplication of matrices A and B. Matrix A is a 3x2 matrix and matrix B is a 2x2 matrix. The multiplication A · B results in a 3x2 matrix. The calculation for the bottom-right element of the result matrix is shown:  $2 \times 1 + 4 \times 2 = 10$ . The diagram highlights the row from A and the column from B used for this calculation with red boxes.

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

1 x 3 + 2 x 4 = 11

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

1 x 3 + 3 x 4 = 15

# Multiplicação de matrizes

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A \cdot B = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 7 & 15 \\ 10 & 22 \end{pmatrix}$$

Diagram illustrating the multiplication of matrices A and B. Matrix A is a 3x2 matrix and matrix B is a 2x2 matrix. The multiplication A · B results in a 3x2 matrix. Red boxes highlight specific elements during the calculation:

- A red box surrounds the first row of A (1, 2) and the first column of B (1, 2), with a red arrow pointing from the bottom right corner of the box to the result 5.
- A red box surrounds the second row of A (1, 3) and the first column of B (1, 2), with a red arrow pointing from the bottom right corner of the box to the result 7.
- A red box surrounds the third row of A (2, 4) and the first column of B (1, 2), with a red arrow pointing from the bottom right corner of the box to the result 10.
- A red box surrounds the first row of A (1, 2) and the second column of B (2, 4), with a red arrow pointing from the bottom right corner of the box to the result 11.
- A red box surrounds the second row of A (1, 3) and the second column of B (2, 4), with a red arrow pointing from the bottom right corner of the box to the result 15.
- A red box surrounds the third row of A (2, 4) and the second column of B (2, 4), with a red arrow pointing from the bottom right corner of the box to the result 22.

Calculation detail:  $2 \times 3 + 4 \times 4 = 22$

# Multiplicação de matrizes e a regressão

$$X = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 4 \end{pmatrix}$$

$$X \cdot \hat{B} = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{\alpha} + 3\hat{\beta} \\ \hat{\alpha} + 2\hat{\beta} \\ \hat{\alpha} + 4\hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

# Matrizes e regressão múltipla

$$X = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ 1 & 4 & 0 \end{pmatrix}$$

$$X \cdot \hat{B} = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\alpha} + 3\hat{\beta}_1 + 1\hat{\beta}_2 \\ \hat{\alpha} + 2\hat{\beta}_1 + 0\hat{\beta}_2 \\ \hat{\alpha} + 4\hat{\beta}_1 + 0\hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{pmatrix}$$