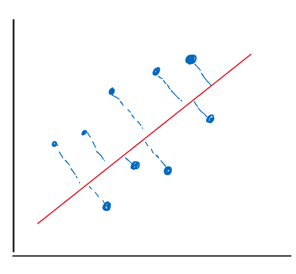
Princípios de Álgebra Linear

Regressão linear simples $\Rightarrow$ univariada



$\mathbb{E}[x]$ = média populacional

↳ esperança estatística

$\bar{x}$ = média amostral

Esperança condicional: $\mathbb{E}[Y|X]$ $\Rightarrow$ média de Y quando fixamos X em uma condição

$\Downarrow$

$\mathbb{E}[y \mid 7\% \le x \le 10\%]$

Regressão é uma aproximação linear das médias condicionais

$\bar{Y}|X = \hat{\beta}_0 + \hat{\beta}_1 x$

↳ em média, dadas determinadas condições, chegamos a determinados valores de Y.

↳ x

No caso da regressão múltipla,

$y_i = \beta_0 + \beta_{i1} x_1 + \beta_{i2} \cdot x_2$ $\Rightarrow$ plano, e não uma reta de regressão.

A regressão em notação matricial

$y = X\beta + \epsilon$

$$\underbrace{y_i}_{\text{escalar}} = \underbrace{\vec{x}_i}_{\text{vetor}} \overbrace{\beta}^{\text{vetor}} + \underbrace{\epsilon_i}_{\text{escalar}}$$

Para representar a regressão na forma vetorial, suponha:

$$\begin{array}{cc} \text{cte} & \text{edu} \\ \end{array} \qquad \text{Yrenda}$$

$$\underbrace{\begin{bmatrix} 1 & 10 \\ 1 & 5 \\ 1 & 15 \\ 1 & 12 \\ \vdots & \vdots \\ 1 & 8 \end{bmatrix}}_{x_1 \quad x_2} \qquad \underbrace{\begin{bmatrix} 1200 \\ 1300 \\ 500 \\ 4100 \\ \vdots \\ 800 \end{bmatrix}}_{y}$$

existe uma combinação linear de x_1 e x_2 que me permita aproximar y?

A regressão linear assume β_0 e β_1 fixos para todos os indivíduos.

Suponha $\beta_0 = 50$ e $\beta_1 = 100$

$$\underbrace{\begin{bmatrix} 1 & 10 \\ 1 & 5 \\ 1 & 15 \\ 1 & 12 \\ \vdots & \vdots \\ 1 & 8 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} 50 \\ 100 \end{bmatrix}}_{\vec{\beta}} = \underbrace{\begin{bmatrix} 1200 \\ 1300 \\ 500 \\ 4100 \\ \vdots \\ 800 \end{bmatrix}}_{\vec{y}} + \underbrace{\overset{\text{erros}}{\begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix}}}_{\epsilon}$$

$\vec{y} = X\vec{\beta} + \vec{e}$ $\Rightarrow$ isso, no entanto, não existe, pq X não é quadrada

$\rightarrow X^{-1}\vec{y} = \cancel{X^{-1}X}\vec{\beta} + \vec{e}$

$\Rightarrow \vec{b} = X^{-1}\vec{y} - \vec{e}$

Para isso funcionar, precisamos de um truque:

$$X\vec{\beta} = \vec{y} \Rightarrow \underbrace{X^T X}_{\text{quadrada}} \vec{\beta} = X^T \vec{y}$$

$$\Rightarrow \cancel{(X^T X)^{-1}(X^T X)}\vec{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$$\Rightarrow \mathbb{I}\vec{\beta} = (X^T X)^{-1} X^T \vec{y}$$

Note que, na ocasião anterior, fizemos a conta sem o termo referente ao erro. De fato, $y = X\beta$ apenas aproxima y. Ainda temos o erro.

$$\vec{y} = X\vec{\beta} + \vec{\epsilon} \Rightarrow X^T y = X^T X \vec{\beta} + \boxed{X^T \vec{\epsilon}}$$

↳ segundo o ML4, isso é 0.

$$\Rightarrow (X^T X)^{-1} X^T y = \cancel{(X^T X)^{-1} X^T X}\vec{\beta} + \cancel{(X^T X)^{-1} X^T \vec{\epsilon}}$$

$$\Rightarrow (X^T X)^{-1} X^T y = \vec{\beta} + \cancel{(X^T X)^{-1} X^T \vec{\epsilon}}$$

$$\Rightarrow \vec{\beta} = (X^T X)^{-1} X^T (\vec{y} \cancel{- \vec{\epsilon}})$$

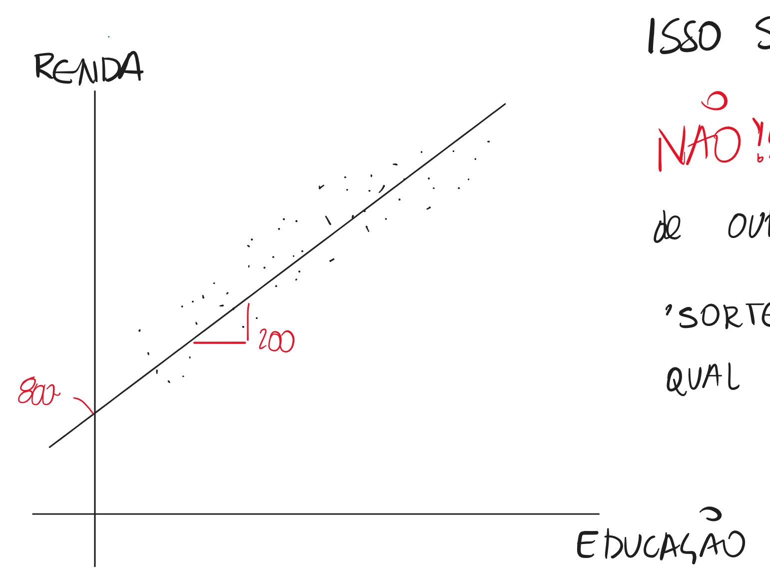
* Pressupostos do modelo linear

ML1: é possível obter y por meio de uma combinação linear das colunas de X

ML2: cada observação de X gera o y correspondente usando os mesmos coeficientes beta

ML3: $X^T X$ é inversível

ML4: $X^T \epsilon = 0$. Pressuposto da exogeneidade.



ISSO SIGNIFICA QUE EDUCAÇÃO CAUSA RENDA?

NÃO!! à medida que x varia, uma série de outras coisas variam também.

"SORTEIA UMA PESSOA NO BRASIL E ELA TEM DOUTORADO. QUAL COR VOCÊ DIRIA QUE ESSA PESSOA TEM?"

$y = X\vec{\beta} + \epsilon$ $\longrightarrow$ resíduo de todo o resto que determina a renda.

Quando dizemos que $X^T e = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, implicitamente estamos dizendo que o erro se distribui aleatoriamente, ou ceteris paribus, ou seja, todo o resto se mantém constante.