



**INSTITUTO FEDERAL**

Rio Grande do Sul

Campus Erechim

---

# ESTATÍSTICA

## - REGRESSÃO LINEAR -

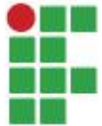
Profa. Claudia Turik de Oliveira



# Objetivo

---

- A correlação linear mostra o *quanto* duas variáveis estão relacionadas.
- A regressão linear mostra *como* elas estão relacionadas.
- Estimar valores de uma variável com base em valores conhecidos de outra variável.





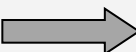

# Exemplos

---

- Deseja-se prever quanto tempo será necessário para executar uma determinada tarefa por uma pessoa, com base no tempo de treinamento.
- Um gerente deseja estimar as vendas semanais com base nas vendas das segundas e terças feiras.

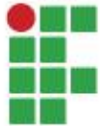
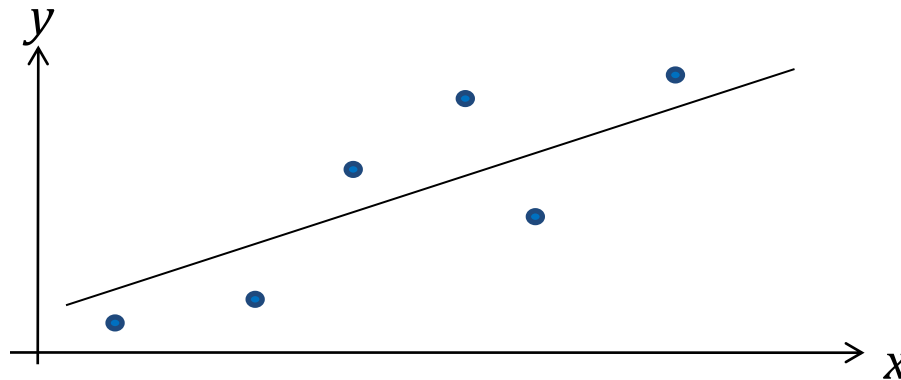


# Exemplos

Variável independente, $X$		Variável dependente, $Y$	
			
Renda (R\$)		Consumo (R\$)	
Memória RAM do computador (Gb)		Tempo de resposta do sistema (seg.)	
Área construída do imóvel (m <sup>2</sup> )		Preço do imóvel (R\$)	

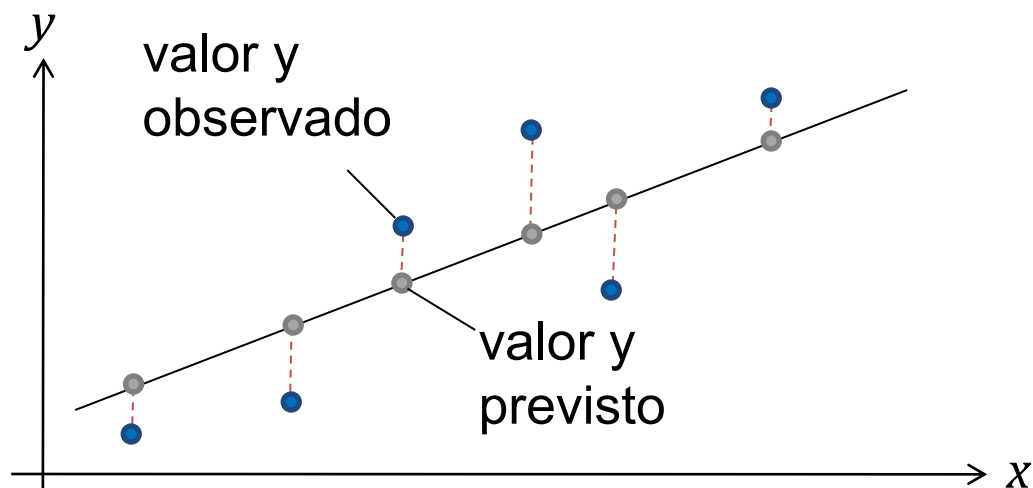
# Reta de regressão

- Após verificar se a correlação linear entre duas variáveis é significativa, o próximo passo é determinar a equação da linha que melhor modela os dados (linha de regressão).
- Pode ser usada para prever o valor de  $y$  para um dado valor de  $x$ .



# Reta de regressão

- Método dos mínimos quadrados: A reta para a qual a soma dos quadrados dos resíduos é um mínimo.



# Equação de regressão estimada

$$\hat{Y} = a + bX$$

- Interpretação de **b** (inclinação da reta):  
**Para cada aumento de uma unidade em X, temos um aumento médio de b unidades em Y.**
- Interpretação de a (intercepto):  
**É o valor esperado de y quando x = 0 (caso x = 0 faça parte do suporte do problema).**

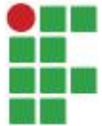


# Equação de regressão estimada

Os coeficientes a e b são calculados da seguinte maneira:

$$b = \frac{\sum[(x - \bar{x})(y - \bar{y})]}{\sum(x - \bar{x})^2}$$

$$a = \bar{y} - b \cdot \bar{x}$$





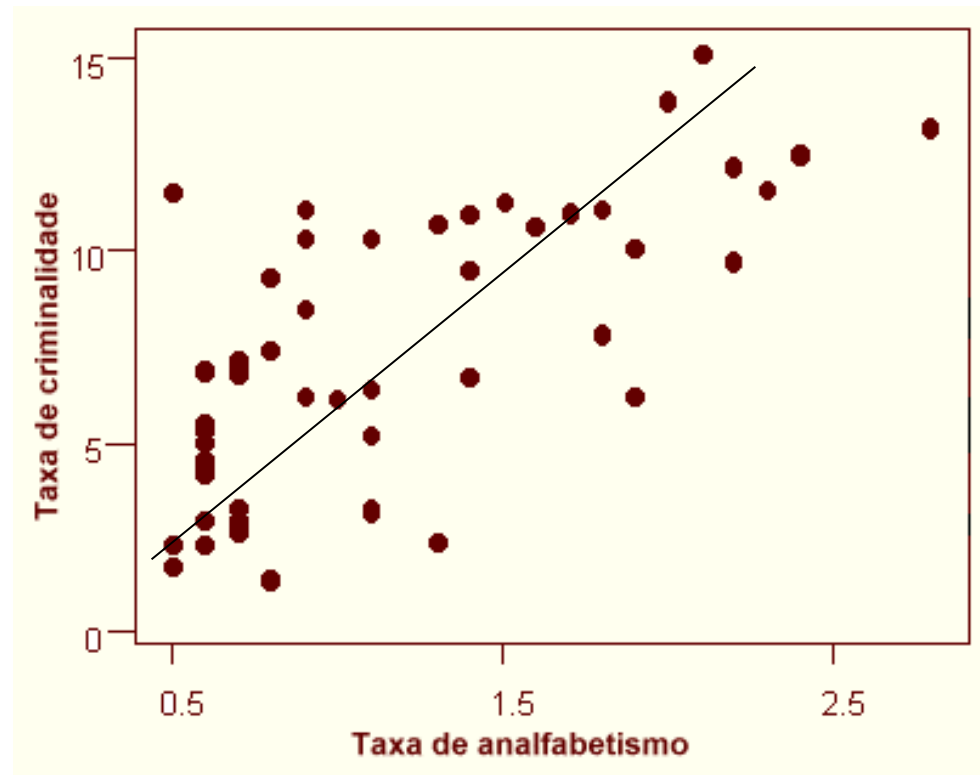
# Exemplo 1

- Criminalidade e Analfabetismo

Considere duas variáveis observadas em 50 estados:

Y: taxa de criminalidade

X: taxa de analfabetismo



# Exemplo 1

---

- Criminalidade e Analfabetismo

a reta ajustada é:  $\hat{Y} = 2,4 + 4,3 \cdot X$

Sendo:  $\hat{Y}$ : valor predito para a taxa de criminalidade  
X: taxa de analfabetismo

- Interpretação de b:

Para um aumento de uma unidade na taxa do analfabetismo (X), a taxa de criminalidade (Y) aumenta, em média, 4,3 unidades.



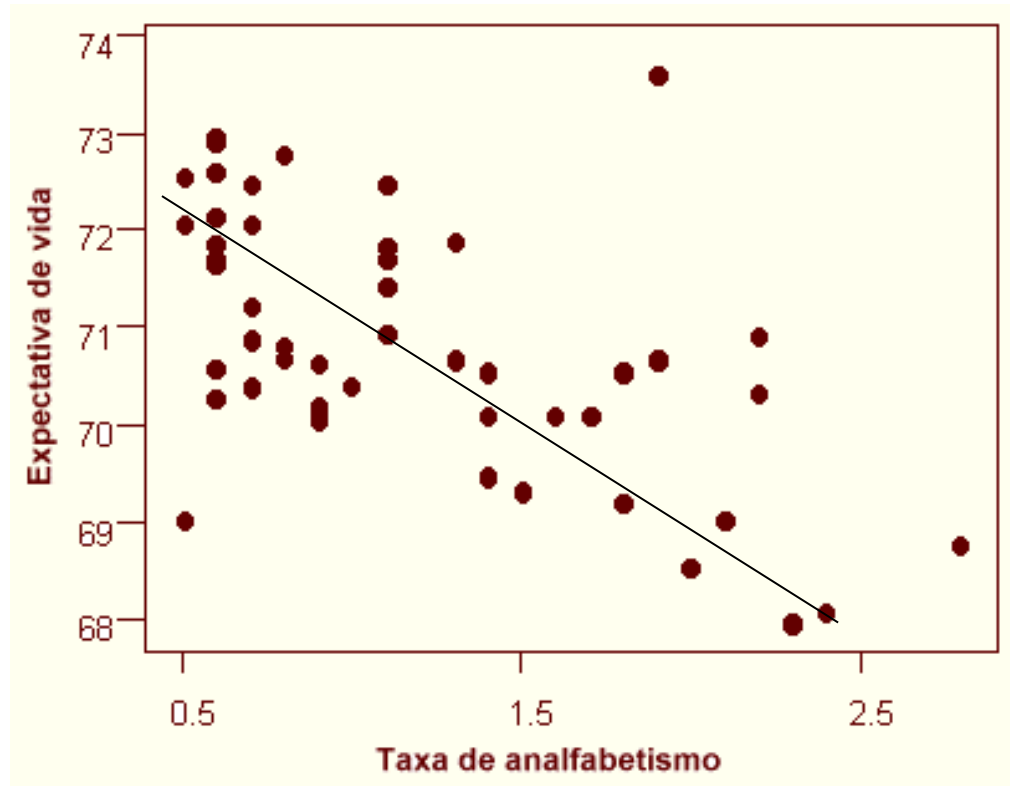
## Exemplo 2

- Expectativa de vida e Analfabetismo

Considere duas variáveis observadas em 50 estados:

Y: expectativa de vida

X: taxa de analfabetismo



# Exemplo 2

---

- Expectativa de vida e Analfabetismo

a reta ajustada é:  $\hat{Y} = 72,4 - 1,3 \cdot X$

Sendo:  $\hat{Y}$ : valor predito para a expectativa de vida  
X: taxa de analfabetismo

- Interpretação de b:

Para um aumento de uma unidade na taxa do analfabetismo (X), a expectativa de vida (Y) diminui, em média, 1,3 anos.



# Exemplo 3

- Consumo de cerveja e temperatura

a reta ajustada é:  $\hat{Y} = 217,4 + 4,7 \cdot X$

Sendo:  $\hat{Y}$ : valor predito para o consumo de cerveja (litros por mil habitantes)  
X: temperatura (°C)

- Interpretação de b:

Aumentando-se um grau de temperatura (X), o consumo de cerveja (Y) aumenta, em média, 4,74 litros por mil habitantes.



# Exemplo 4

---

A equação de regressão para os dados sobre gastos com propaganda (em milhares de dólares) e vendas da empresa (em milhares de dólares) é:  $\hat{y} = 104,061 + 50,729 \cdot x$

Use essa equação para prever as vendas *esperadas* da empresa para os seguintes gastos com propaganda:

- 1,5 mil dólares
- 1,8 mil dólares
- 2,5 mil dólares



# Exemplo 4

---

- 1,5 mil dólares

$$\hat{y} = 104,061 + 50,729x$$

$$\hat{y} = 104,061 + 50,729(1,5) \approx 180,155$$

Quando os gastos com propaganda são de \$1500, as vendas da empresa são cerca de \$180,155.

- 1,8 mil dólares

$$\hat{y} = 104,061 + 50,729(1,8) \approx 195,373$$

Quando os gastos com propaganda são de \$1800, as vendas da empresa são cerca de \$195,373.



# Exemplo 4

- 2,5 mil dólares

$$\hat{y} = 104,061 + 50,729(2,5) \approx 230,884$$

Quando os gastos com propaganda são de \$2500, as vendas da empresa são cerca de \$230,884.

Valores de previsão são significantes somente para valores  $x$  na (ou próximos à) faixa dos dados. Os valores  $x$  do conjunto original de dados variam de 1,4 a 2,6. Portanto, não seria apropriado usar a equação de regressão para prever as vendas da empresa por gastos com propaganda, tais como 0,5 (\$ 500) ou 5,0 (\$ 5.000).



# Exemplo 5

---

X : tempo de estudo (em horas)

Y : nota da prova

**Pares de observações ( $X_i$  ,  $Y_i$ ) para cada estudante**

Tempo (X)	Nota (Y)
3,0	4,5
7,0	6,5
2,0	3,7
1,5	4,0
12,0	9,3



# No exemplo 1:

Tempo (X)	Nota (Y)	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$
3,0	4,5	-2,1	-1,1	2,31	$(-2,1)^2$
7,0	6,5	1,9	0,9	1,71	$(1,9)^2$
2,0	3,7	-3,1	-1,9	5,89	$(-3,1)^2$
1,5	4,0	-3,6	-1,6	5,76	$(-3,6)^2$
12,0	9,3	6,9	3,7	25,53	$(6,9)^2$
25,5	28,0	0	0	41,2	78,2

$$\bar{X} = 5,1$$

$$\bar{Y} = 5,6$$

$$b = \frac{\sum[(x - \bar{x})(y - \bar{y})]}{\sum(x - \bar{x})^2}$$

$$b = \frac{41,2}{78,2} = 0,53$$

$$a = \bar{y} - b \cdot \bar{x}$$

$$a = 5,6 - (0,53 \cdot 5,1) = 2,90$$

$$\hat{Y} = 2,90 + 0,53 \cdot X$$



Rio Grande do Sul  
Campus Erechim

**Aumentando-se uma hora de estudo (X), a nota (Y) aumenta, em média, 0,53 pontos.**