# Regressão Linear

Disciplina: Modelos Estatísticos
Professora: Jéssica Assunção

# O que temos em comum nos exemplos?

# Regressão Linear Simples

| Variável independente (x) | Variável dependente (y) |
|---|---|
| O tamanho da tela de um monitor | Preço do monitor |
| Número de visitantes em um web site | Quantidade de vendas no web site |

# Regressão Linear Simples

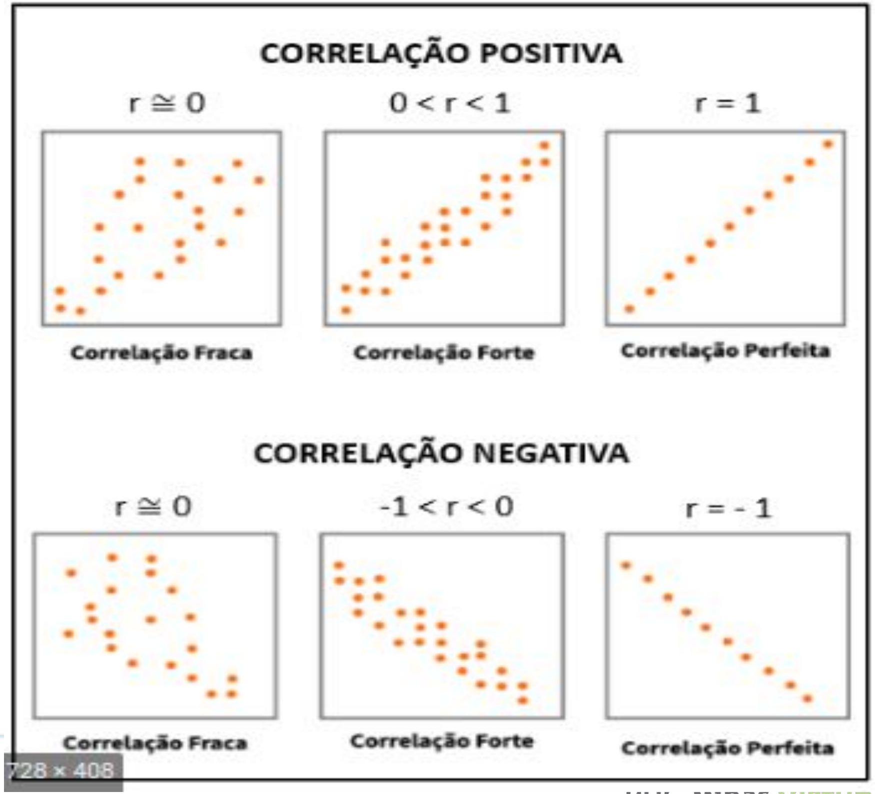
Uma **variável independente x** explica a variação em outra variável, que é chamada de **variável dependente y.**

Este relacionamento existe em apenas uma direção:

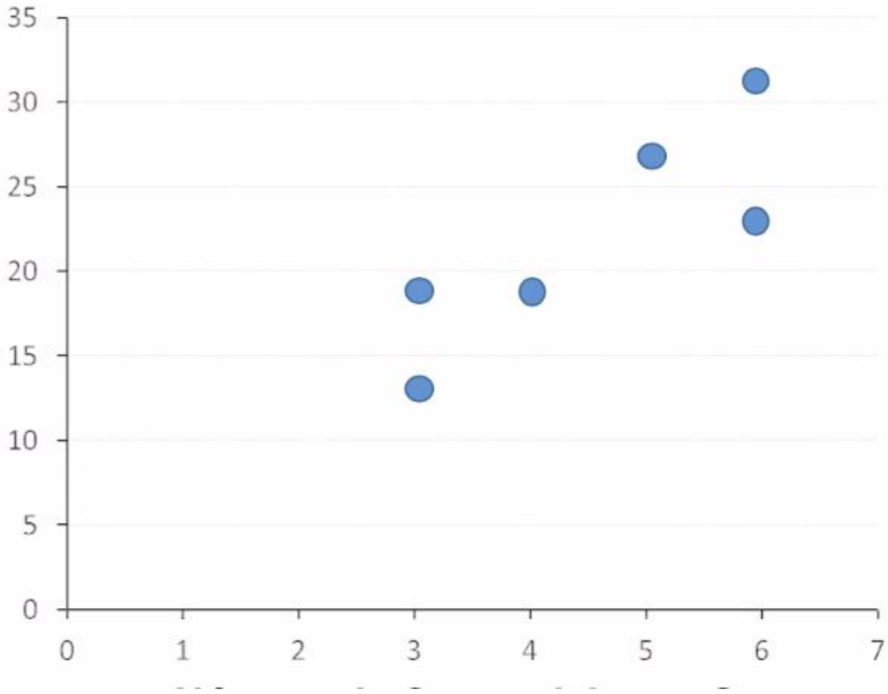
variável independente (x) -> variável dependente (y)

# Regressão Linear Simples

A análise de correlação nos permite medir a força e direção de um relacionamento linear entre duas variáveis.



# Regressão Linear Simples




| Número de anúncios no instagram (x) | Número de bolos vendidos (y) |
|---|---|
| 3 | 13 |
| 6 | 31 |
| 4 | 19 |
| 5 | 27 |
| 6 | 23 |
| 3 | 19 |

# Regressão Linear Simples

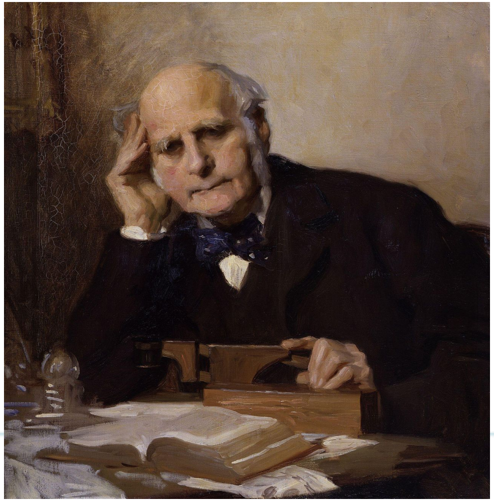
A correlação, isto é, a ligação entre dois eventos, não implica necessariamente uma relação de causalidade, ou seja, que um dos eventos tenha causado a ocorrência do outro.

# Regressão Linear Simples

- Regressão Linear Simples e Múltipla
- Regressão Logística Binária
- Regressão Logística Multinomial
- Regressão Poisson
- Regressão Binomial
- Regressão Ridge
- Regressão Lasso
- Regressão ElasticNet

# Regressão Linear Simples

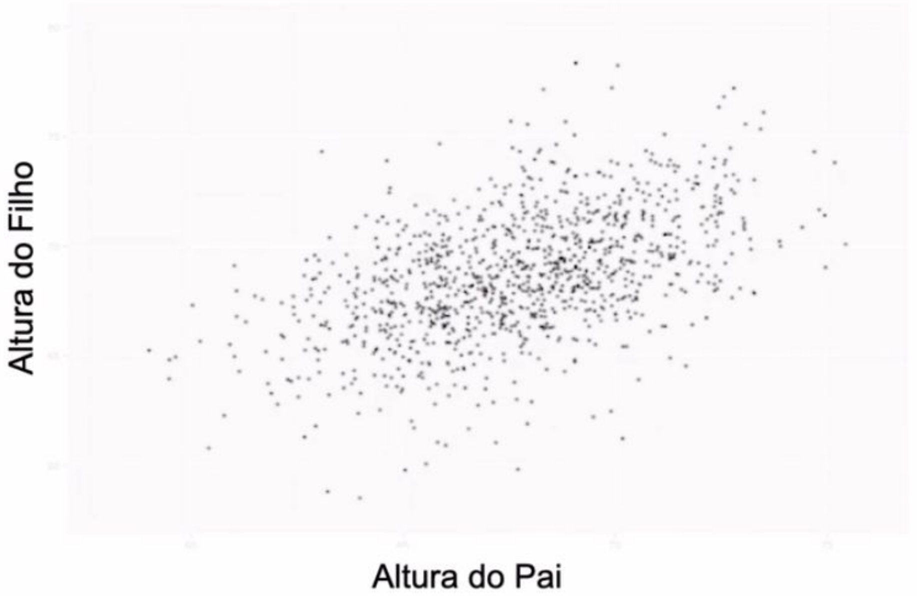
A **análise de regressão** é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas de tal forma que uma variável possa ser predita a partir de outra.



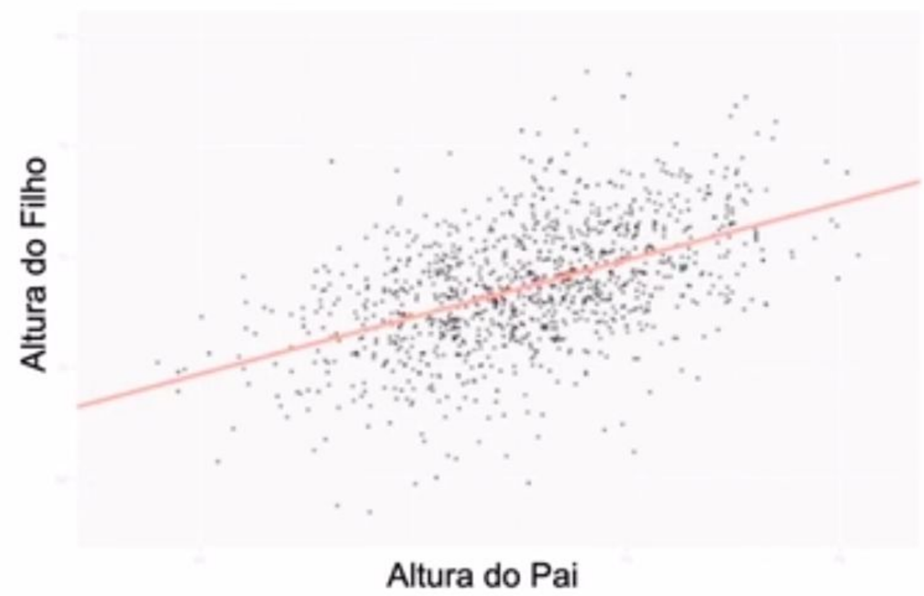
# Regressão Linear Simples

A interpretação moderna da regressão é diferente - ocupa-se do estudo da dependência de uma variável dependente, em relação a uma ou mais variáveis independentes, com o objetivo de estimar e/ou prever a média (da população) ou valor médio de uma variável em termos dos valores conhecidos ou fixos das variáveis independentes.

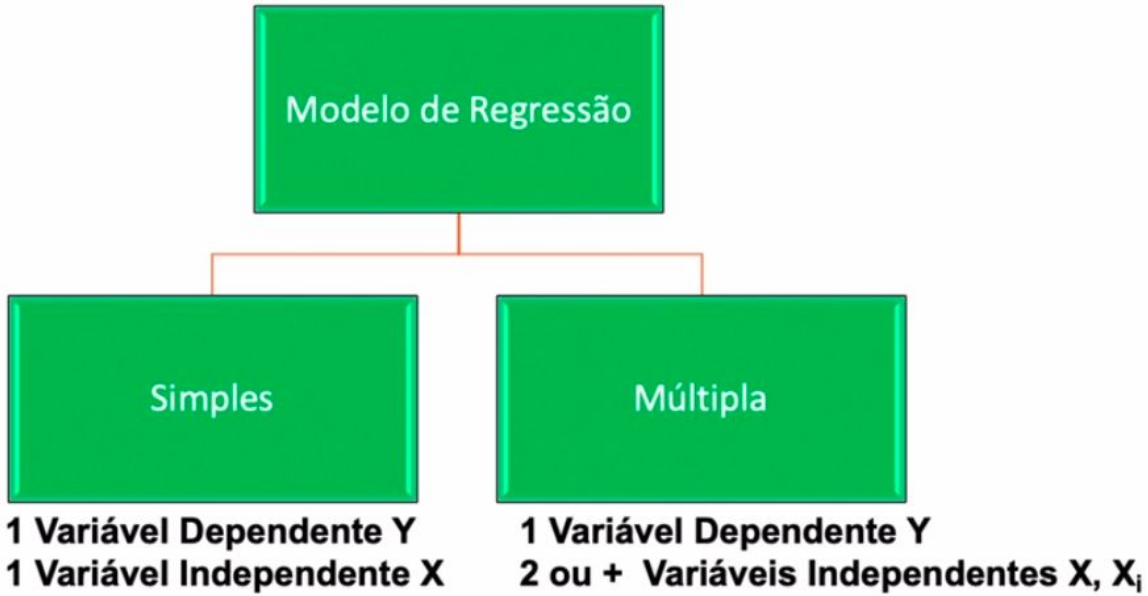
# Regressão Linear Simples



# Regressão Linear Simples



# Regressão Linear Simples



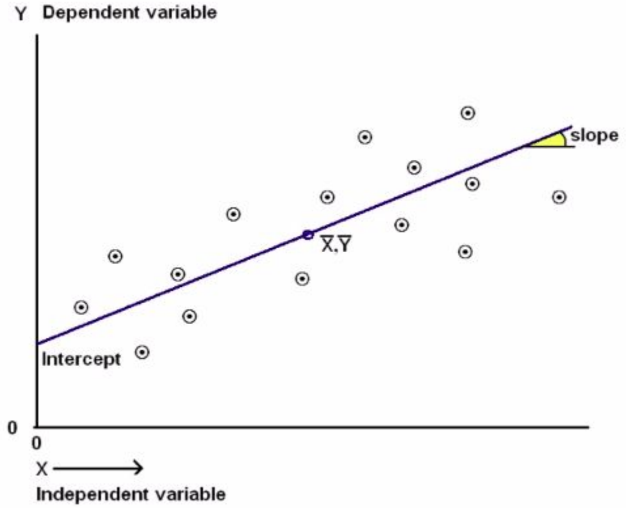
# Regressão Linear Simples

Fórmula para a equação que descreve uma linha reta através de um par ordenado:

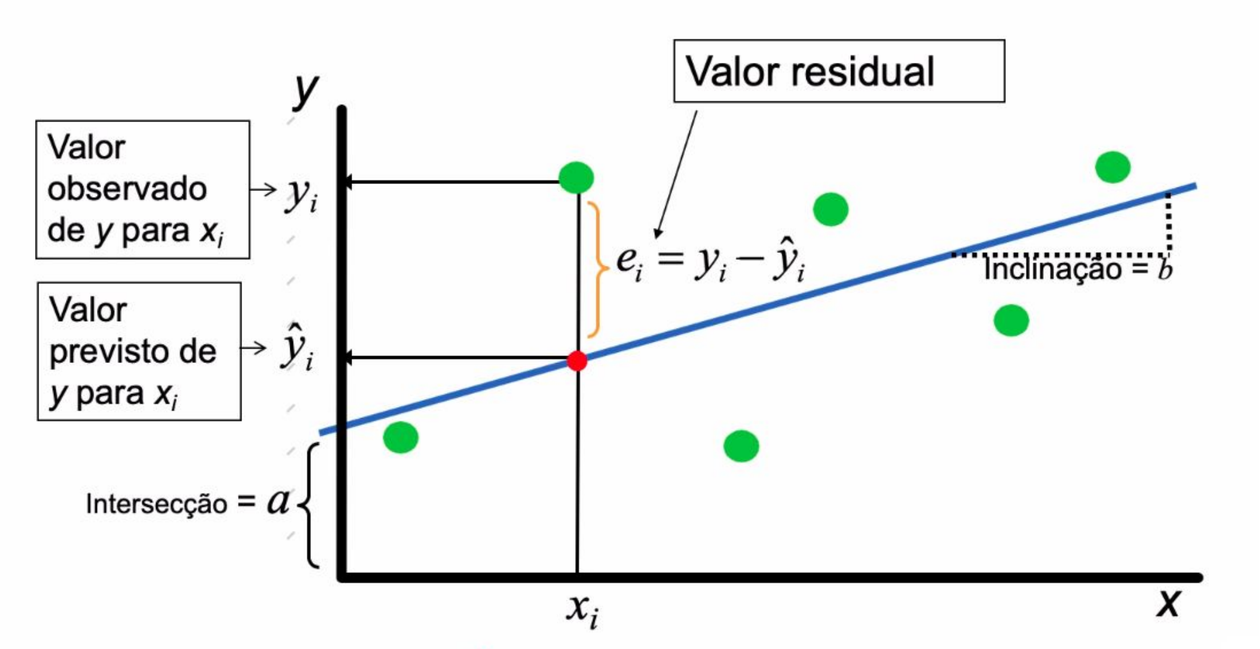
$$\boxed{\hat{y} = a + bx}$$

Onde:

$\hat{y}$ = valor previsto de *y* dado um valor para *x*
***x*** = variável independente
***a*** = ponto onde a linha intercepta o eixo y
***b*** = inclinação da linha reta



# Regressão Linear Simples



# Regressão Linear Simples

Population Y intercept

Population Slope Coefficient

Independent Variable

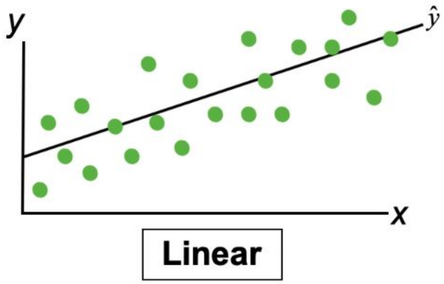
Random Error term

Dependent Variable

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$

# Premissas da Regressão Linear

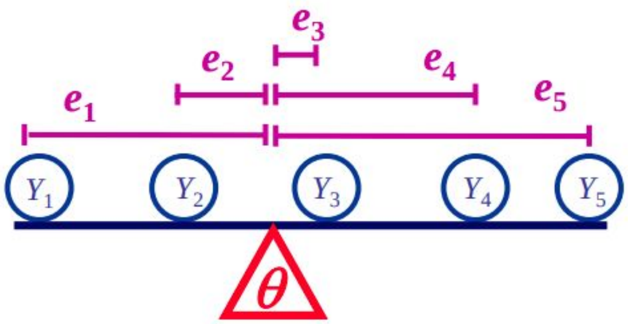
- O relacionamento entre as variáveis independentes e a variável dependente devem ser linear.



# Estimativa dos Coeficientes

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\theta = \beta_0 + \beta_1 X$$




$$e_1 = Y_1 - \theta$$
$$e_2 = Y_2 - \theta$$
$$e_3 = Y_3 - \theta$$
$$e_4 = Y_4 - \theta$$
$$e_5 = Y_5 - \theta$$

# Estimativa dos Coeficientes

**1 ° Passo) Definir Erro Quadrático Total:**

$$EQT = e_1^{\,2} + e_2^{\,2} + e_3^{\,2} + e_4^{\,2} + e_5^{\,2}$$

$$EQT = (Y_1 - \theta)^2 + (Y_2 - \theta)^2 + (Y_3 - \theta)^2 + (Y_4 - \theta)^2 + (Y_5 - \theta)^2$$

$$EQT(\,\theta\,) = \sum_{i=1}^{5} (Y_i - \theta)^2$$

# Estimativa dos Coeficientes

**Para uma amostra de tamanho *n* teremos:**

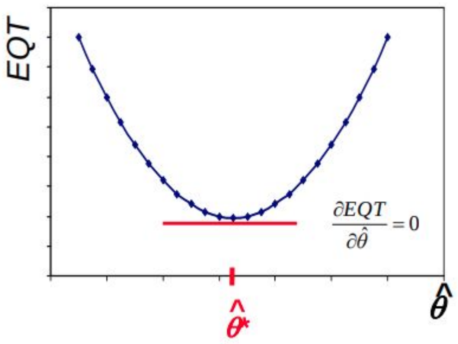
$$Y_i = \hat{\theta} + \hat{e}_i \quad \Rightarrow \quad EQT = \sum_{i=1}^{n} \hat{e}_i^{\,2} = \sum_{i=1}^{n} (Y_i - \hat{\theta})^2$$

**2 ° Passo) Encontrar $\hat{\theta}$ que minimiza *EQT*:**

$$\frac{dEQT}{d\hat{\theta}} = 0$$

$$\frac{dEQT}{d\hat{\theta}} = \sum_{i=1}^{n} 2(Y_i - \hat{\theta})(-1) \quad = -2\sum_{i=1}^{n} Y_i + 2\sum_{i=1}^{n} \hat{\theta}$$

$$\frac{dEQT}{d\hat{\theta}} = -2\sum_{i=1}^{n} Y_i + 2n\hat{\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{n}$$



# Estimativa dos Coeficientes

**Para encontrar os valores que minimizam a função EQT:**

$$\frac{\partial EQT}{\partial \hat{\alpha}} = 2\sum_{i=1}^{n}[Y_i - (\hat{\alpha} + \hat{\beta}X_i)](-1) = 0 \qquad (1) \quad \Rightarrow \quad \hat{\alpha} = \overline{Y} - \hat{\beta}_1\overline{X}$$

$$\frac{\partial EQT}{\partial \hat{\beta}} = 2\sum_{i=1}^{n}[Y_i - (\hat{\alpha} + \hat{\beta}X_i)](-X_i) = 0 \qquad (2) \quad \Rightarrow \quad \hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$

# Estimativa dos Coeficientes

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$\begin{cases} \widehat{\beta}_1 = \dfrac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2} = \dfrac{\widehat{\mathbb{C}\mathrm{ov}}(X,Y)}{\widehat{\mathbb{V}\mathrm{ar}}(X)} \\ \\ \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X} \end{cases}$$

# Premissas da Regressão Linear

- Sem presença de outliers na análise dos resíduos
- Homocedasticidade
- Normalmente distribuído
- Ausência de multicolinearidade e autocorrelação

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

# Como avaliar o modelo?

**Coeficiente de Determinação ($R^2$):**

**Definição:** **Estima a proporção da variabilidade da variável dependente (Y) que é explicada pela(s) variável(eis) independente(s) do modelo de regressão.**

$$R^2 = \frac{SQReg}{STQ} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n}\hat{e}_i^2}{\sum_{i=1}^{n}y_i^2} = \hat{\beta}_1^2 \frac{\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}y_i^2}$$

## Escala de $R^2$:

**0**
**independência linear**

A significância das escalas depende muito da natureza da variável dependente

**1**
**relação linear exata**

# Regressão Linear Múltipla

E quando possuímos mais de uma variável independente?

A solução está na Regressão Linear Múltipla

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + E$$

# Referências Bibliográficas


- Introdução à Estatística - Mário F. Triola
- Noções de Probabilidade e Estatística - Marco Nascimento Magalhães
- Modelos de Regressão em R - Écio Souza Diniz
- Análise de Séries Temporais - Pedro A. Moretin / Clélia M. C. Toloi
- Análise e Previsões de Séries Temporais: Os modelos ARIMA - Reinaldo Castro Souza / Maria Emília Camargo