



# CIENTISTA DE DADOS

## | Nov22

# REGRESSÃO

# Análise de Regressão

## ❑ OBJETIVO:

Estudar o relacionamento entre uma variável dependente (alvo/resposta) e uma ou mais variáveis independente (preditor).

Ajuda a entender como o valor da variável dependente está mudando quando levamos em consideração uma variável independente enquanto outras variáveis independentes são mantidas fixas.

## ❑ PROCEDIMENTO:

Obter uma relação funcional (modelo matemático) que associe a variável resposta com as variáveis preditoras (uma ou mais). Ou seja, partir do conhecimento do valor das variáveis preditoras, possamos obter o máximo possível de informações sobre a variável resposta, isto é:  $Y = f(x)$ .

# ANÁLISE DE REGRESSÃO

## Alguns usos

Usado principalmente para predição, previsão, modelagem de série temporal e determinação da relação causal-efeito entre as variáveis.

- Explicar situações que se quer entender: Por que as ligações para o atendimento ao cliente caíram no mês passado?
- Prever algo sobre o futuro: Como serão as vendas nos próximos seis meses?
- O risco de uma doença cardíaca, dado o histórico familiar e uma série de características físicas e psicológicas.
- Em finanças, calcular o Beta (volatilidade dos retornos em relação ao mercado geral) para uma ação.



# ANÁLISE DE REGRESSÃO

## Modelos de Regressão

### QUESTÕES:

- Qual a “forma” de  $f(x)$ ?

$$Y = a + bx \quad \text{OU} \quad Y = a + bx_1 + cx_2 \dots$$

- A relação funcional é linear ou não linear (nos parâmetros)?

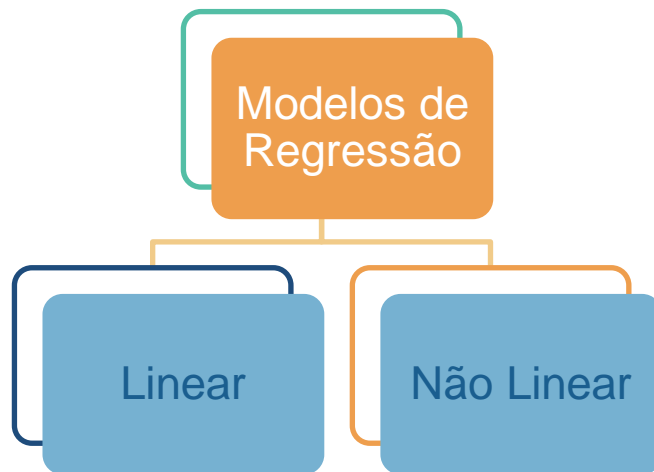
$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

OU

$$Y_i = \frac{\beta_1}{1 + \beta_2 e^{\beta_3 x}} \dots$$

# ANÁLISE DE REGRESSÃO

## Modelos de Regressão



# ANÁLISE DE REGRESSÃO

## Modelos de Regressão

### QUESTÕES:

- ❑ Se LINEAR, qual o número de variáveis respostas?

- UMA  $\rightarrow$  Univariada

$$Y_{n \times 1} = f(x)$$

**Ex:** A partir do valor do Peso, prever valor da Altura

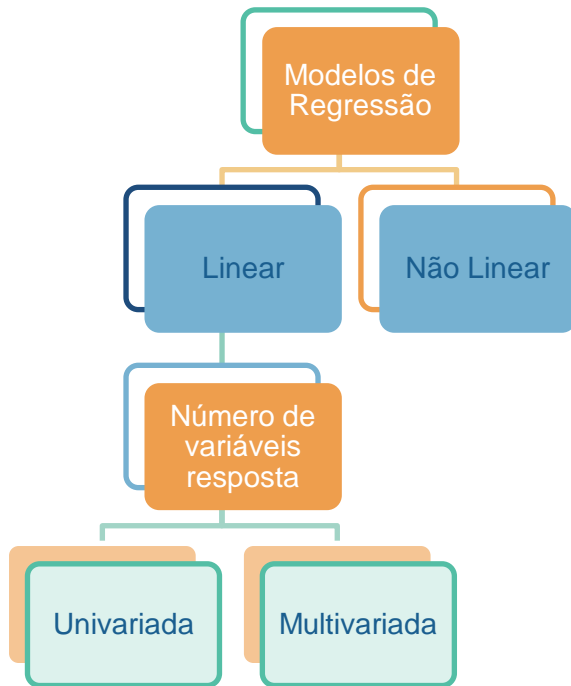
- DUAS OU MAIS  $\rightarrow$  Multivariada

$$Y_{n \times p} = f(x)$$

**Ex:** A partir do valor do Peso, prever valor da Altura e da Circunferência da cintura.

# ANÁLISE DE REGRESSÃO

## Modelos de Regressão





# ANÁLISE DE REGRESSÃO

## Modelos de Regressão

### QUESTÕES:

- Se UNIVARIADO, qual o número de variáveis preditoras?

- UMA → **Simples**

$$Y_{nx1} = f(x_{nx2})$$

**Ex:** A partir do valor do Peso, prever valor da Altura

- DUAS OU MAIS → **Múltipla**

$$Y_{nx1} = f(x_{nxp}), p > 2$$

- **Ex:** A partir do valor do Peso, da Circunferência da cintura, do Gênero e da Idade, prever o valor da Altura.

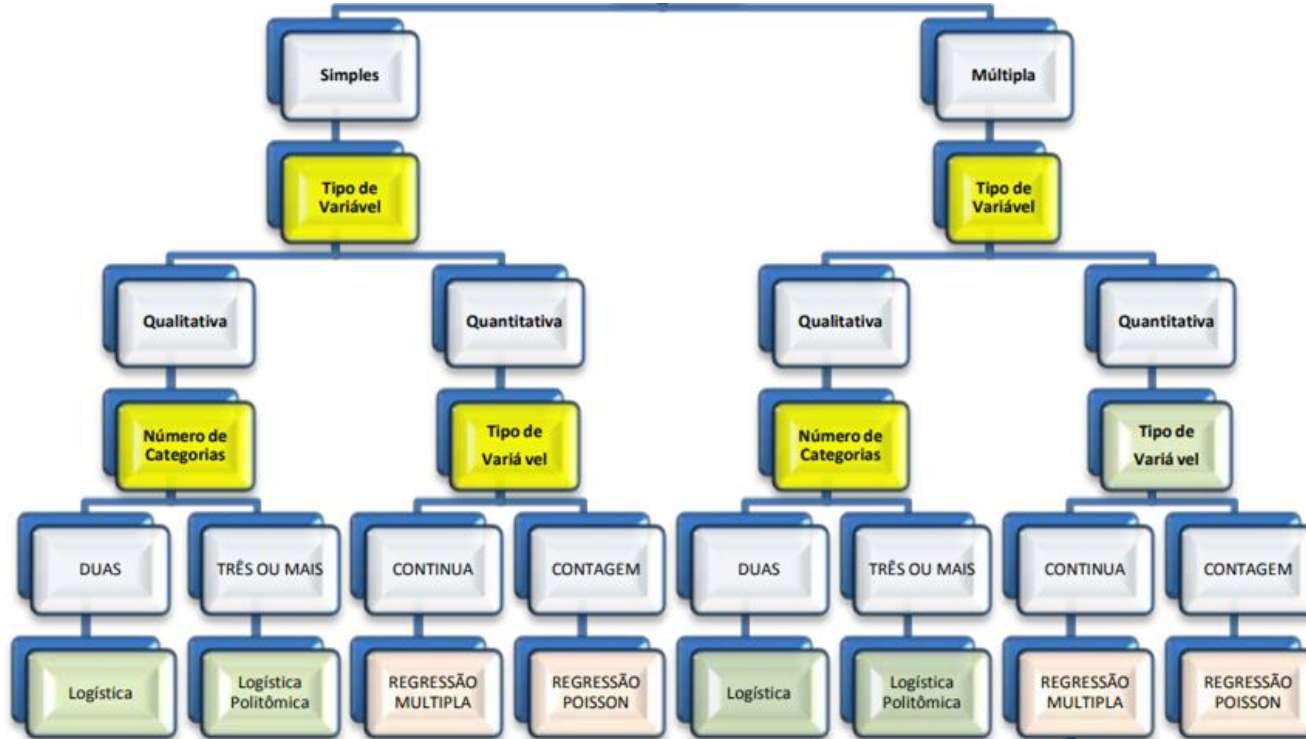
# ANÁLISE DE REGRESSÃO

## Modelos de Regressão



# ANÁLISE DE REGRESSÃO

## Modelos de Regressão

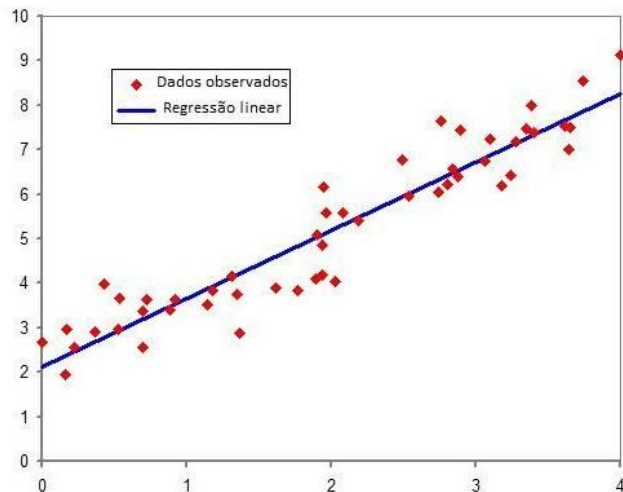


# ANÁLISE DE REGRESSÃO

## Regressão Linear Simples

Reta traçada a partir de uma relação em um diagrama de dispersão. Essa reta resume uma relação entre os dados de duas variáveis e também pode ser utilizada para realizar previsões.

Ou seja, dado  $X$  e  $Y$ , quanto que  $X$  explica  $Y$ , estudando o relacionamento entre **uma** variável resposta ( $Y$ ) e **uma** variável preditora ( $X$ ).



# ANÁLISE DE REGRESSÃO

## Regressão Linear Simples - Exemplo



- 1) Relação entre peso e altura de um homem adulto (x: altura; y: peso).
- 2) Relação entre preço do vinho e o montante da colheita em cada ano (X: Montante da Colheita; Y: Preço do Vinho).

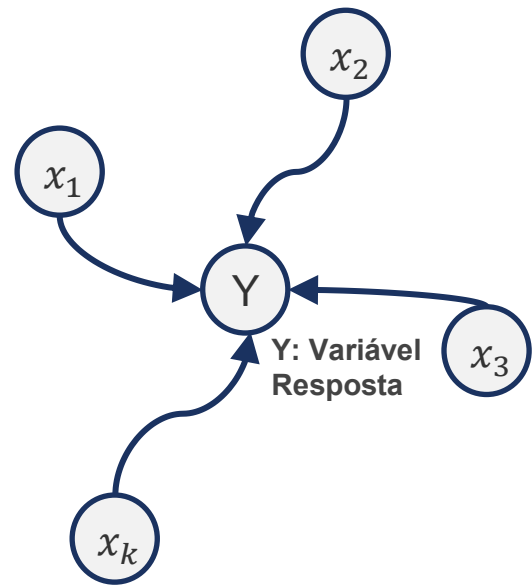
# ANÁLISE DE REGRESSÃO

## Regressão Linear Múltipla

Uma única variável dependente (resposta), porém duas ou mais variáveis independentes (explicativas).

A finalidade das variáveis independentes adicionais é melhorar a capacidade de predição da regressão linear simples.

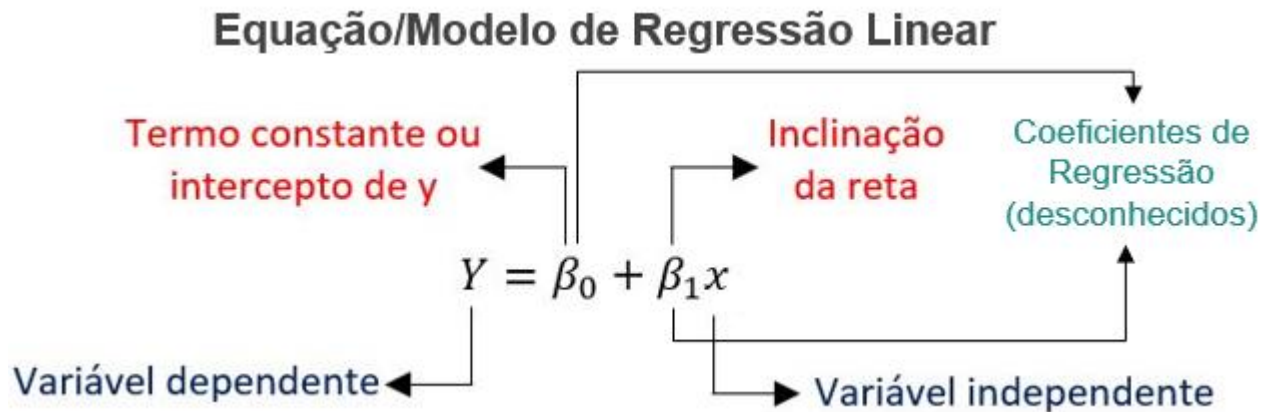
**CUIDADO:** O ideal é obter o mais alto relacionamento explicatório com o mínimo de variáveis independentes, principalmente por conta do custo na obtenção de dados para muitas variáveis.



$x_1, x_2, x_3, \dots, x_k$ : Variáveis Explicativas

# ANÁLISE DE REGRESSÃO

## Como interpretar



**Ex:**

# Batatas quebradas =  $4,251 - 0,909 \text{ \% de batata} + 0,02231 \text{ Temp. de cozimento}$