

# **Ciência de Dados e Aprendizado de Máquina**

Introdução à Análise de  
Regressão Linear





# Robson Fernandes

## Acadêmico

Mestrando em Matemática, Estatística e Computação Aplicadas (Data Science & Machine Learning) - USP

Especialização em Reconhecimento de Padrões e Análise de Imagens - UNICAMP

Pós-Graduado em Arquitetura de Software Distribuído - PUC-MG

MBA em Engenharia de Software Orientada a Serviços – SOA – METROCAMP

Certificado – JavaScript e HTML5 Developer – W3C INTERNACIONAL

Autor do Livro Gestão da Tecnologia da Informação: Teoria e Prática

## Profissional

Cientista de Dados Sênior – Finch Soluções

Docente Pós-Graduação - MBA em Data Science & Machine Learning - UNIP

Docente Pós-Graduação em Engenharia de Software - USC

Docente Graduação em Ciência da Computação - UNIP

## Site

<http://robsonfernandes.net>

## e-mail

[robson.fernandes@usp.br](mailto:robson.fernandes@usp.br) / [robs.fernandes@outlook.com](mailto:robs.fernandes@outlook.com)



# Análise de Regressão - Introdução

- A **análise de regressão** estuda o relacionamento entre uma ou mais variáveis chamadas **variável independentes (X)** e uma variável chamada **variável dependente (Y)**.



Variável independente  
ou  
Variável explicativa



Variável dependente  
ou  
Variável resposta



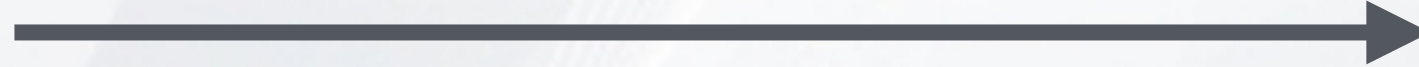


# Análise de Regressão - **Objetivo**

- **Predizer** valores de uma **variável dependente (Y)** em função de uma ou  $N$  **variáveis independentes (X)**.



Variável independente  
ou  
Variável explicativa



Variável dependente  
ou  
Variável resposta





# Análise de Regressão - **Modelo Matemático**

- Este relacionamento é representado por um **modelo matemático**.
- Isto é, por uma equação que associa a **variável dependente (Y)** com as **variáveis independentes (X)**.





# Análise de Regressão - Modelo de Regressão Linear Simples

- Este modelo é designado por **modelo de regressão linear simples**, se define uma relação linear entre a **variável dependente (Y)** e **uma variável independente (X)**.

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

**Y** – Variável dependente ou resposta

**X** – Variável independente ou explicativa

$\beta_0$  e  $\beta_1$  – Parâmetros desconhecidos do modelo (a estimar)

**e** – Variável aleatória residual. Comportamento da variável **Y** que não podem ser explicadas linearmente pelo comportamento da variável **X**



# Análise de Regressão - **Modelo de Regressão Linear Simples**

- **Exemplos**

- **Desenvolver um modelo matemático que busque representar a relação dos exemplos abaixo:**
- **Relação entre o peso e a altura de um homem adulto**  
( **X : altura**; **Y : peso** )
- **Relação entre o preço do vinho e o montante da colheita em cada ano**  
( **X : montante da colheita**; **Y : preço do vinho** )





# Análise de Regressão - Modelo de Regressão Linear Múltipla

- Este modelo é designado por **modelo de regressão linear múltipla**, se define uma relação linear entre a **variável dependente (Y)** e **várias variáveis independentes (X)**.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + e$$

**Y** – Variável dependente ou resposta

**X<sub>1</sub>, ..., X<sub>k</sub>** – Variáveis independentes ou explicativas

**β<sub>0</sub>, ..., β<sub>k</sub>** – Parâmetros desconhecidos do modelo (a estimar)

**e** – Variável aleatória residual. Comportamento da variável **Y** que não podem ser explicadas linearmente pelo comportamento da variável **X<sub>1</sub>, ..., X<sub>k</sub>**



# Análise de Regressão - **Modelo de Regressão Linear Múltipla**

- **Exemplos**
- **Desenvolver um modelo matemático que busque representar a relação dos exemplos abaixo:**
- **Relação entre o volume de vendas (Y) efetuadas durante um dado período de tempo por um vendedor, os seus anos de experiência (X1) e o seu score num teste de inteligência (X2).**





# Análise de Regressão - **Estimação de Betas por MMQ**

- A partir dos dados disponíveis estimamos  $\beta_0, \beta_1, \dots, \beta_k$  pelo **Método dos Mínimos Quadrados (MMQ)** e substituímos estes parâmetros para obter a equação de regressão estimada.

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$





# Análise de Regressão - **Vamos Praticar!**

- Acompanhe o arquivo que consta na pasta “Artigos”
  - **Data Science - Análise de Regressão Linear Aplicado a Previsão de Vendas**
- Ou no link
  - <https://goo.gl/BFNBvY>





# Bibliografia

- SILVA. L. A.; PERES. S. M; BOSCARIOLI C. **Introdução à Mineração de Dados**. Elsevier. 2016
- FACELI, Katti; Lorena, Ana Carolina; Gama, João ; de Carvalho, A. C. P. L. F. (2011). **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 1. ed. Rio de Janeiro: LTC.
- PROVOST, F.; Fawcett, T. **Data Science for Business: What you need to know about data mining and data-analytic thinking** by O'Reilly Media, 2013.
- FLACH, P. (2012). **Machine Learning: The Art and Science of Algorithms that Make Sense of Data**. Cambridge University Press.
- ALPAYDIN, E. (2004). **Introduction to Machine Learning**. MIT Press.