



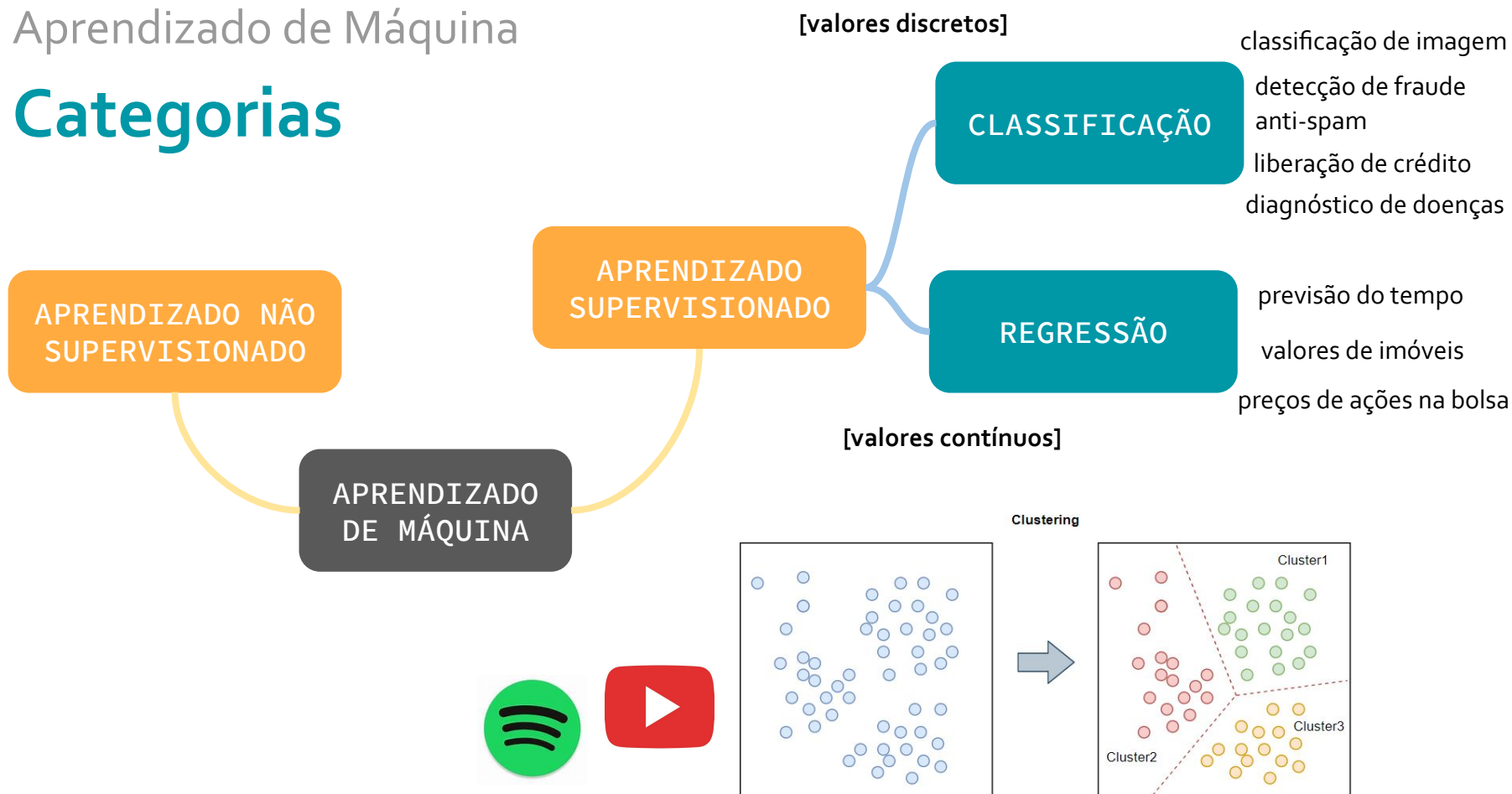
## APRENDIZADO DE MÁQUINA - REGRESSÃO

profs. Lívia Ruback, Raimundo Macário e Marcelo Dib

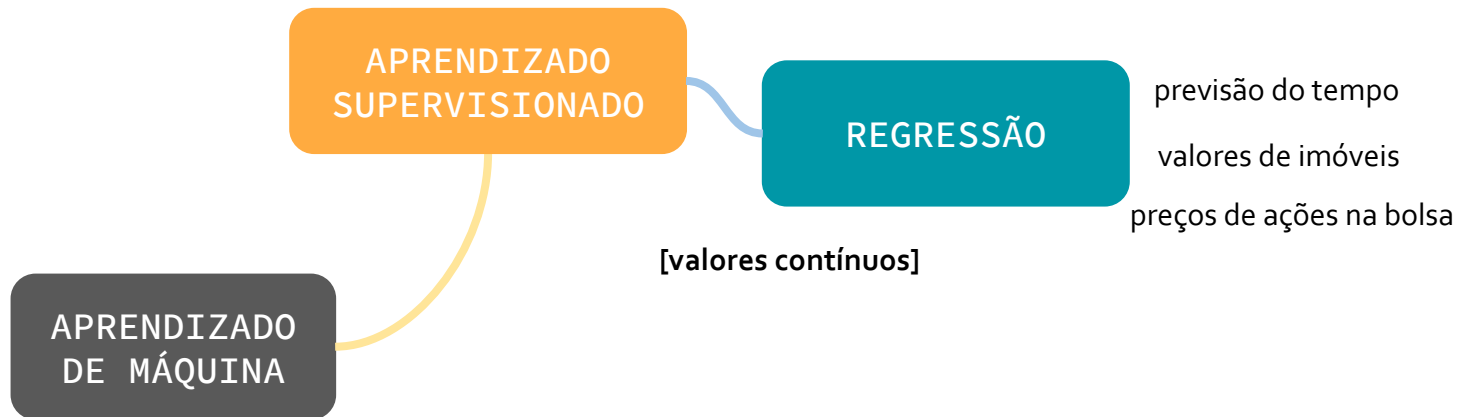
# Aprendizado de Máquina - Regressão

- Aprendizado de Máquina:
  - Classificação x Regressão
- Regressão
- Exemplos
- Medindo o desempenho
- Algoritmos de Regressão
- Parte prática

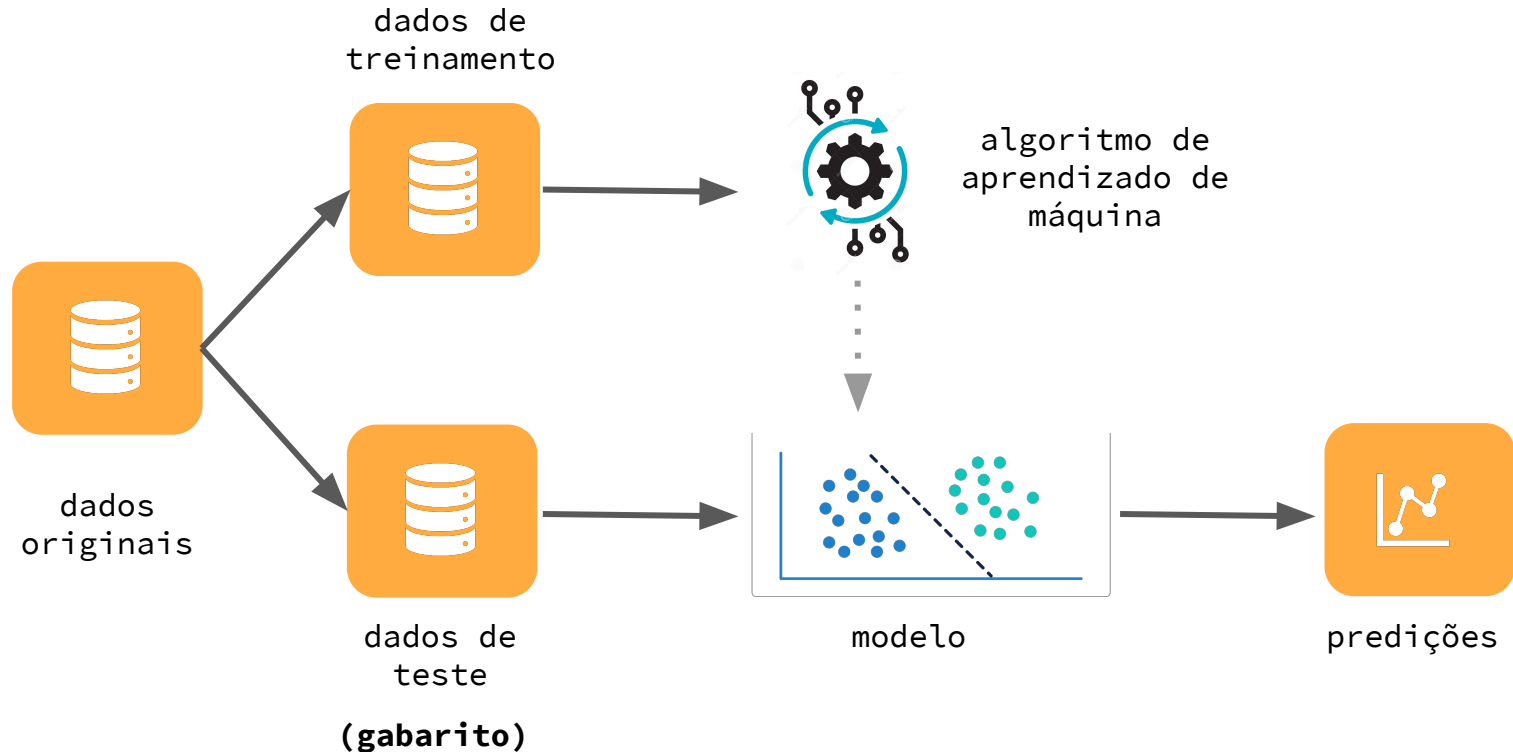
## Categorias



## Categorias



# A.M Supervisionado

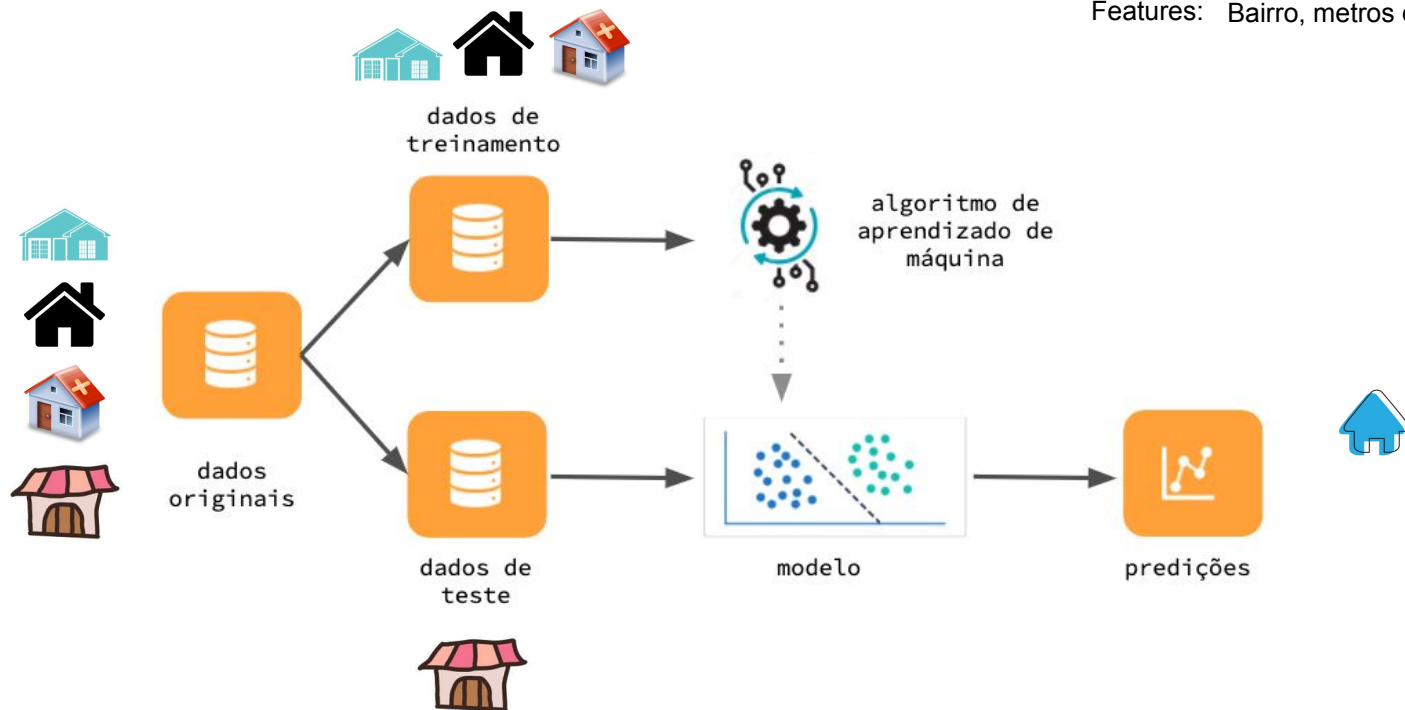


# Aprendizado de Máquina Supervisionado: Regressão

## Regressão: Exemplo 1

Modelo para prever o preço de venda de uma casa

Features: Bairro, metros quadrados, n. quartos



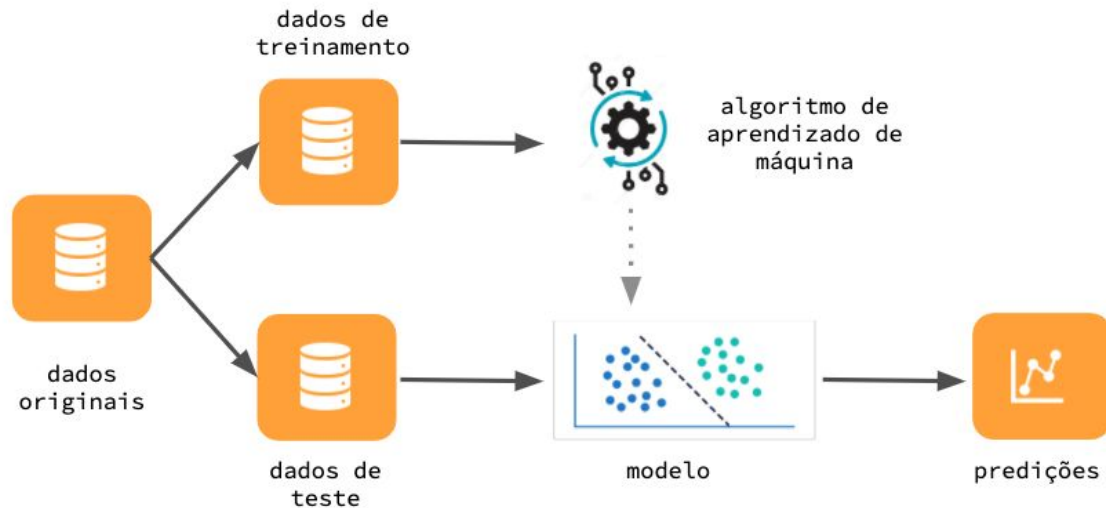
# Aprendizado de Máquina Supervisionado Regressão

## Regressão: Exemplo 2

Modelo para prever consumo de CO2 de motores

Features: Cilindradas do motor (ex.: 1.0, 1.2, 1.8, 2.0)

Classe: Número de ppm (partes por milhão)



Aprendizado de Máquina

# Técnicas de Regressão

REGRESSÃO

- Regressão Linear
- Máquinas de Vetores de Suporte
- Redes Neurais Artificiais
- Árvores de Regressão



Aprendizado de Máquina

# Técnicas de Regressão

REGRESSÃO

- **Regressão Linear**
- Máquinas de Vetores de Suporte
- Redes Neurais Artificiais
- Árvores de Regressão

# Regressão Linear

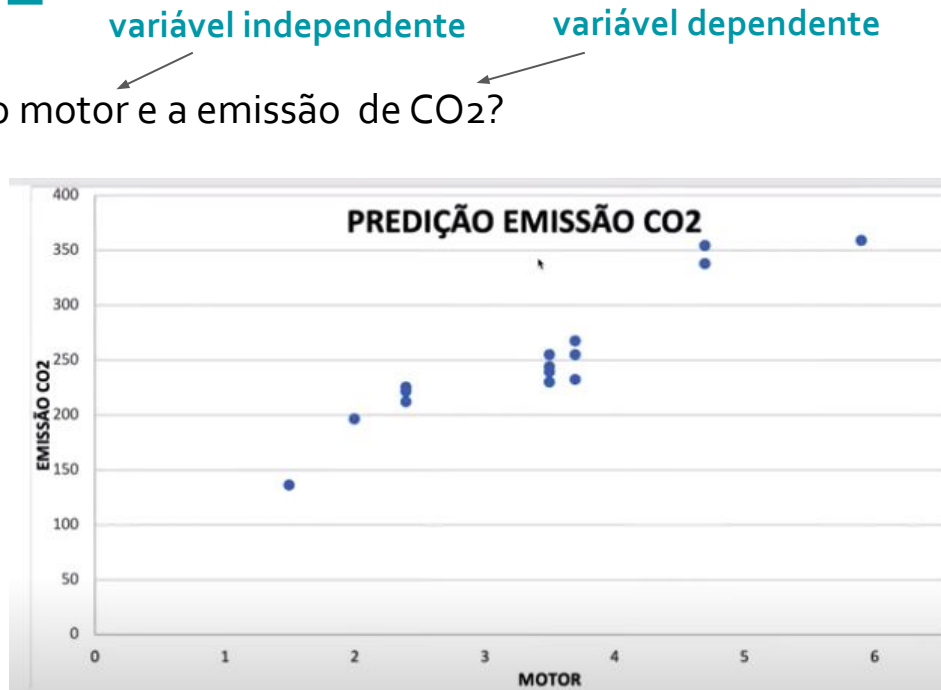
feature

- Estuda a relação entre a variável dependente e a variável independente
- A relação entre elas é representada por um **modelo matemático**
  - Uma variável independente: **Modelo de Regressão Linear Simples (MRLS)**
  - Várias variáveis independentes: **Modelo de Regressão Linear Múltipla (MRLM)**
  - **Correlação**: Representa a força ou o grau da relação entre as variáveis

## Correlação: Exemplo 1

- **PERGUNTA:** Existe relação entre o motor e a emissão de CO<sub>2</sub>?

Motor	Emissão de CO <sub>2</sub>
2	196
2,4	221
1,5	136
3,5	255
3,5	244
3,5	230
3,7	232
3,7	255
3,7	267
2,4	212
2,4	225
3,5	239
5,9	359
5,9	359
4,7	338



## Correlação: Exemplo 2

variável independente

variável dependente

- **PERGUNTA:** Existe relação entre horas de estudo dedicadas e a nota na prova?

Aluno	Horas de Estudo	Nota na Prova
A	6	82
B	2	63
C	1	57
D	5	88
E	3	68
F	2	75



## Correlação: Exemplo 2 (continuação)

variável independente

variável dependente

- **PERGUNTA:** Existe relação entre horas de estudo dedicadas e a nota na prova?



- Podemos observar uma **tendência linear** entre as variáveis:

Quanto **MAIS** horas de estudo,

**MAIOR** é a nota na prova

[hipótese]

## Correlação: Exemplo 3

variável independente

variável dependente

- **PERGUNTA:** Existe relação entre horas de estudo dedicadas e a nota na prova?



- Ou seja...

As horas de estudo determina (ou interfere) na nota na prova?

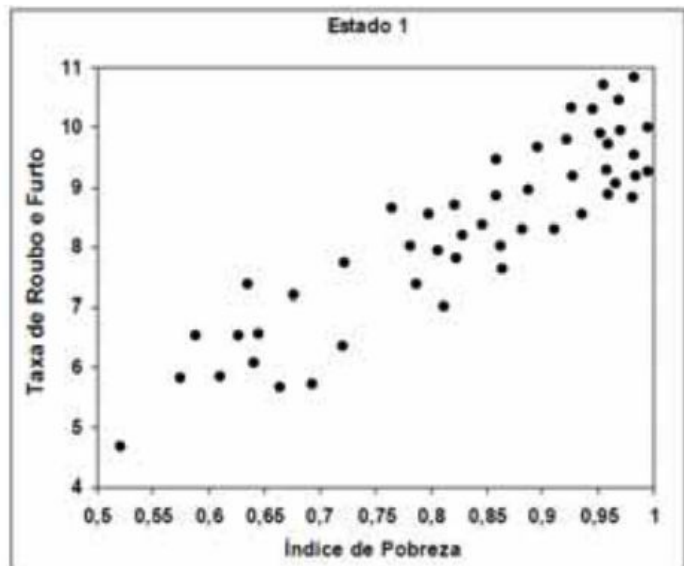
Quão forte é essa relação?

[**correlação**]

## Correlação: Exemplo 4

variável independente      variável dependente

- **PERGUNTA:** Existe relação entre o índice de pobreza e a criminalidade?



- Podemos observar uma **tendência linear** entre as variáveis:

Quanto **MAIOR** o índice de pobreza,  
**MAIOR** é taxa de roubo e furto

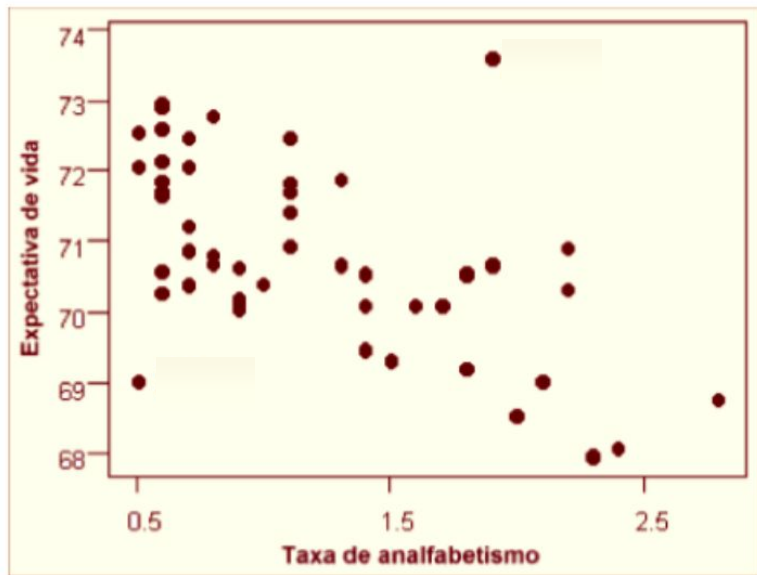
[hipótese]

## Correlação: Exemplo 5

variável independente

variável dependente

- **PERGUNTA:** Existe relação entre a taxa de analfabetismo e a expectativa de vida?



- Podemos observar uma **tendência linear** entre as variáveis:

Quanto **MAIOR** é a taxa de analfabetismo,

**MENOR** é a expectativa de vida da população

[hipótese]

Variáveis observadas em 50 estados norte-americanos

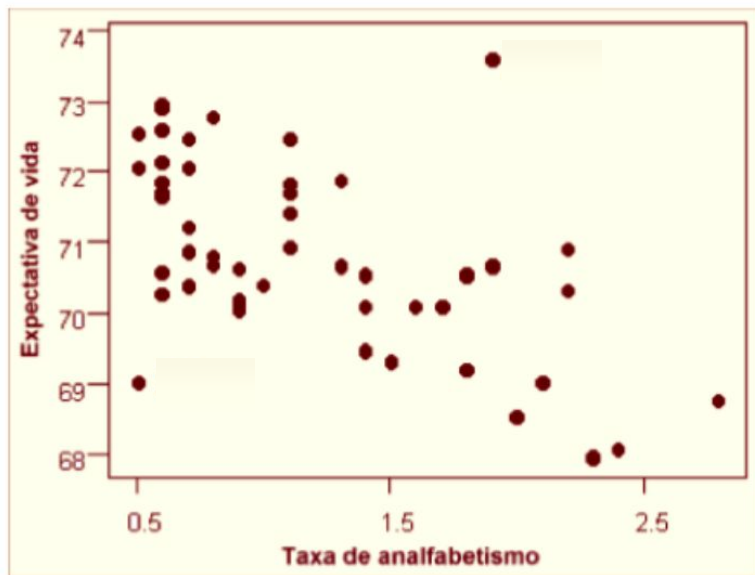


# Correlação: Exemplo 5 (continuação)

variável independente

variável dependente

- **PERGUNTA:** Existe relação entre a taxa de analfabetismo e a expectativa de vida?



- Ou seja...

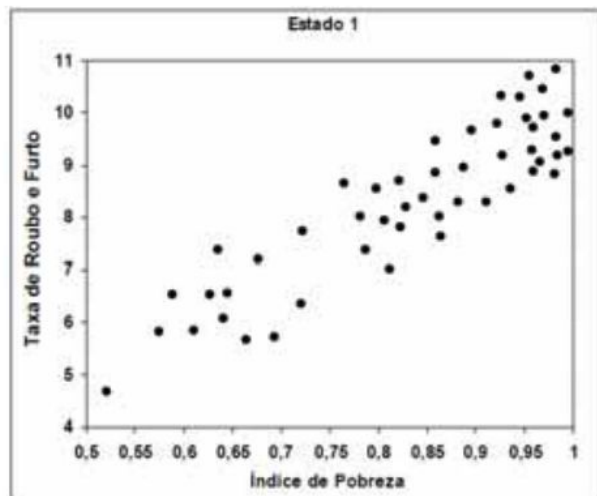
A taxa de analfabetismo nessas regiões determina (ou interfere) na expectativa de vida da população?

Quão forte é essa relação?

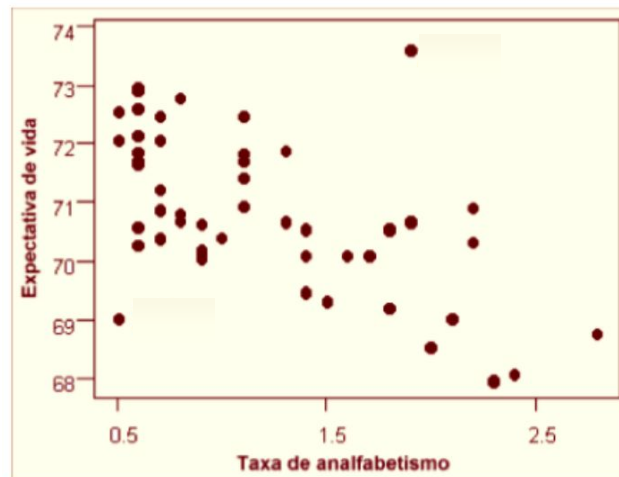
**[correlação]**

Variáveis observadas em 50 estados norte-americanos

# Coeficiente de Correlação Linear

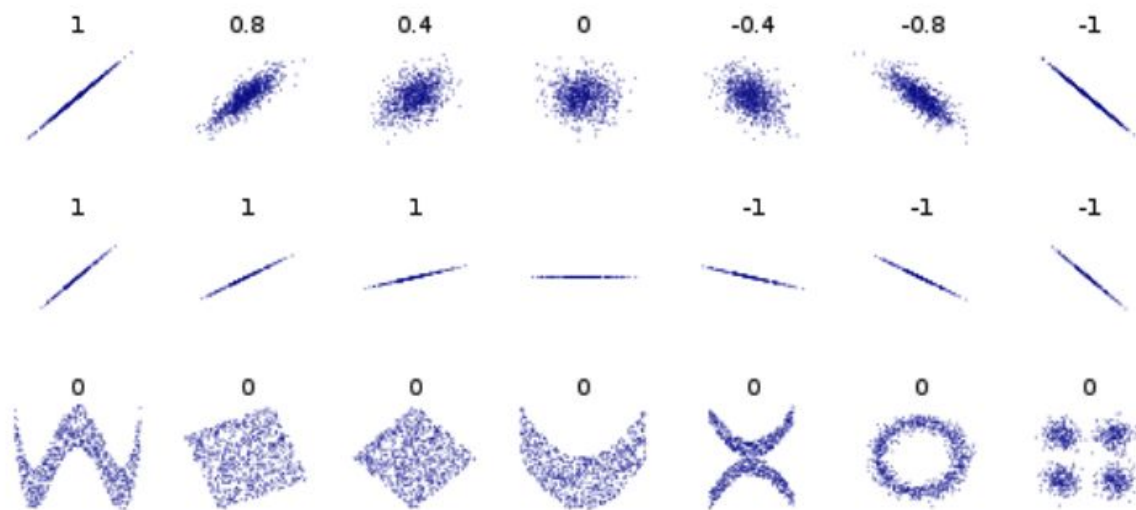


[correlação positiva]



[correlação negativa]

# Coeficiente de correlação linear (r)



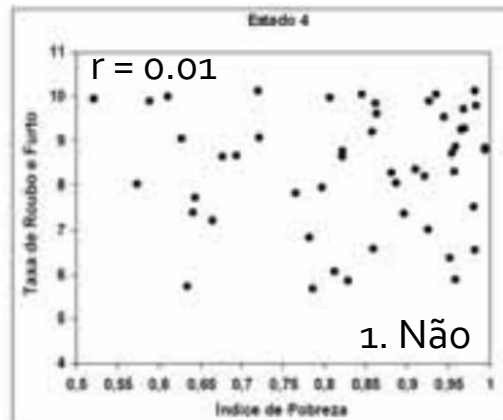
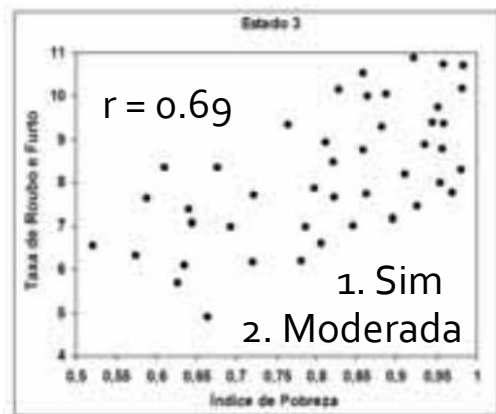
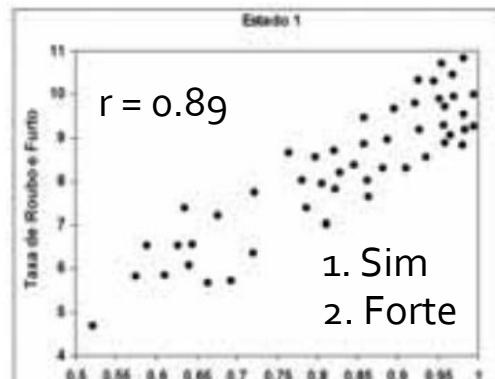
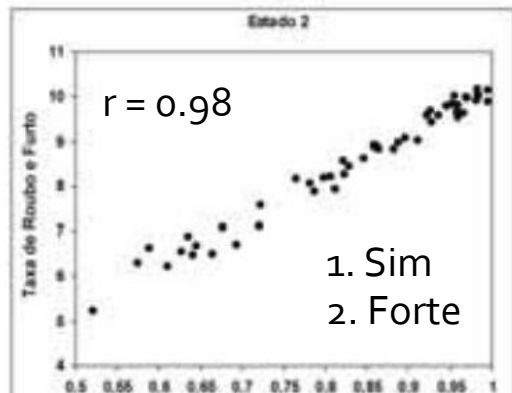
$$-1 \leq r \leq 1$$

$r \approx 1$  -> correlação forte positiva

$r \approx -1$  -> correlação forte negativa

$r \approx 0$  -> não existe cor. linear

# Coeficiente de correlação linear (r)



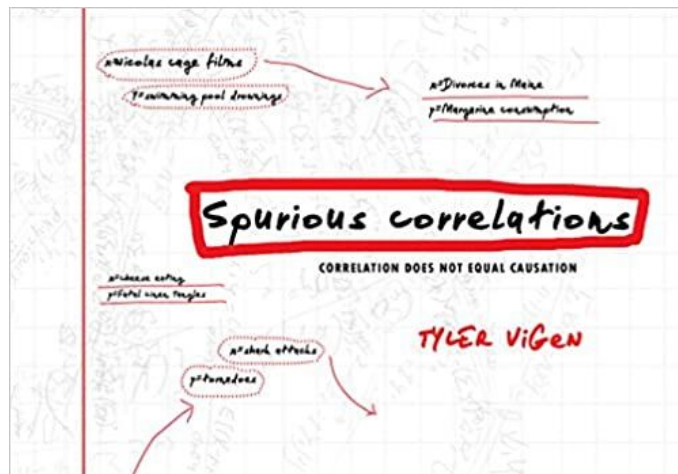
4 possíveis gráficos de dispersão para a análise de correlação entre o índice de pobreza e a criminalidade

## PERGUNTAS:

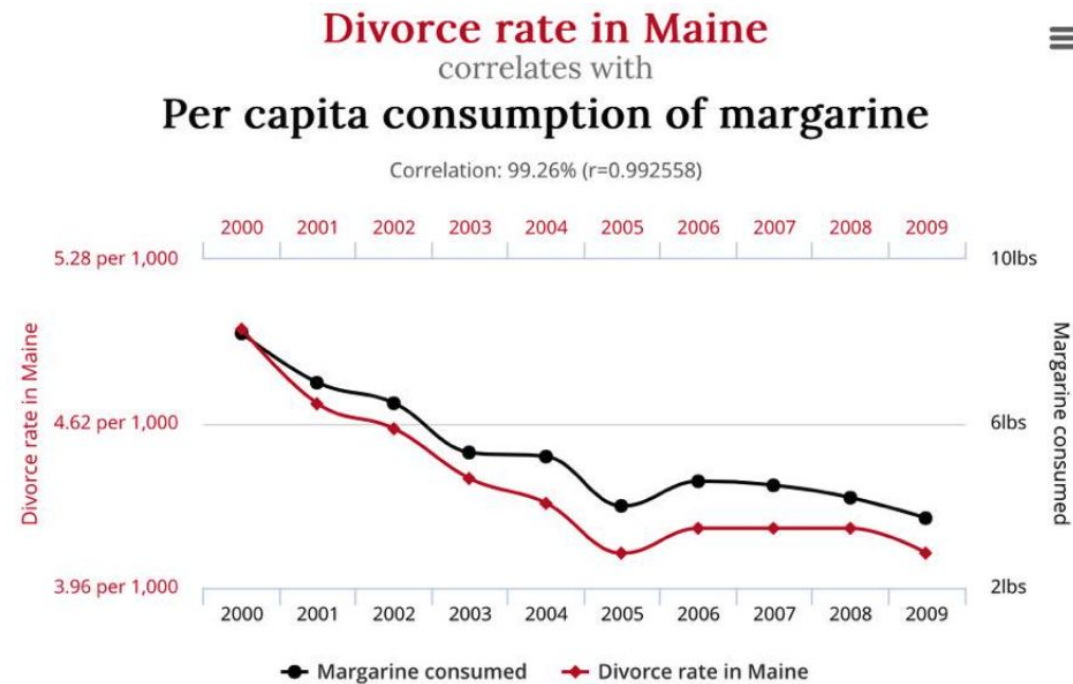
- 1) Existe associação linear entre as variáveis?
- 2) Ela é forte ou fraca?

# ALERTA: Correlação não é causalidade!!

- Não é por que duas variáveis tem correlação, existe uma **relação de causa-efeito** entre elas
- Inferência causal comprova relação causa-efeito

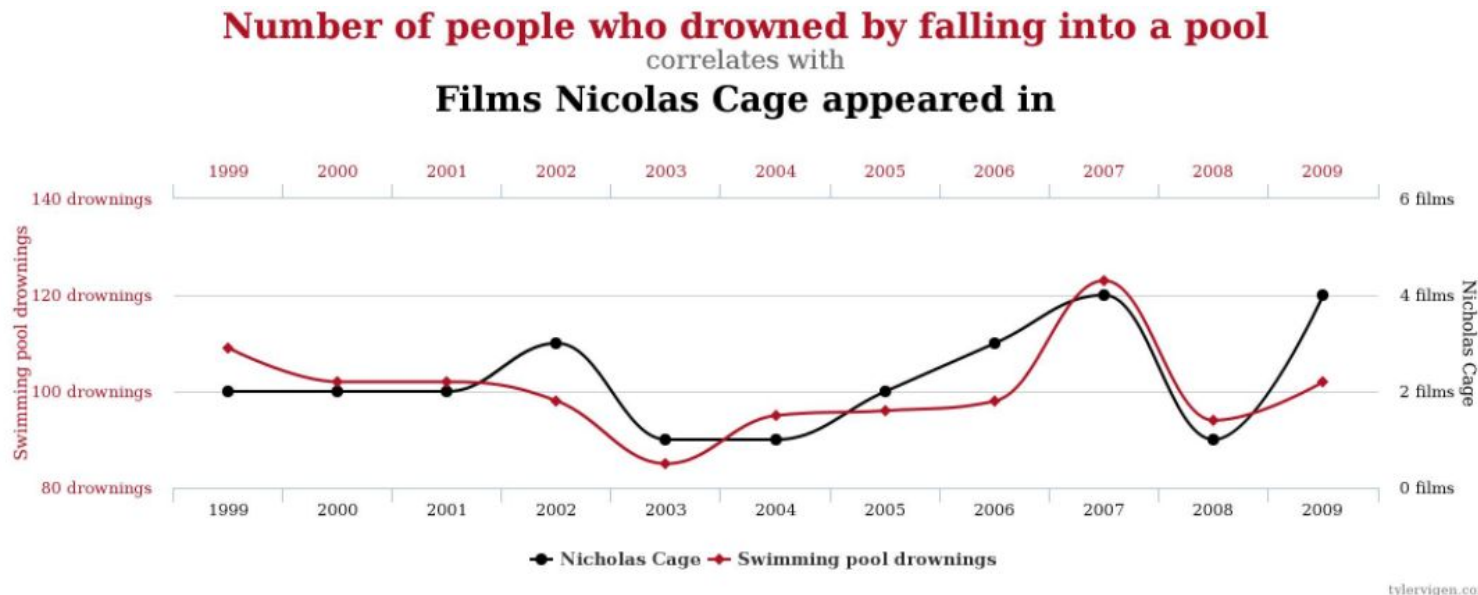


# Exemplos de correlação sem causa-efeito



Taxa de divórcios em Maine  
x  
Consumo per capita de margarina

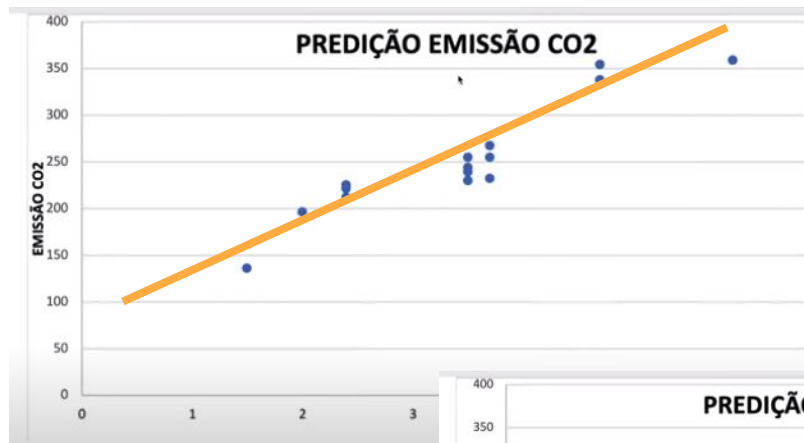
# Exemplos de correlação sem causa-efeito



Número de pessoas que morreram afogadas x Aparições de Nicolas Cage em filmes

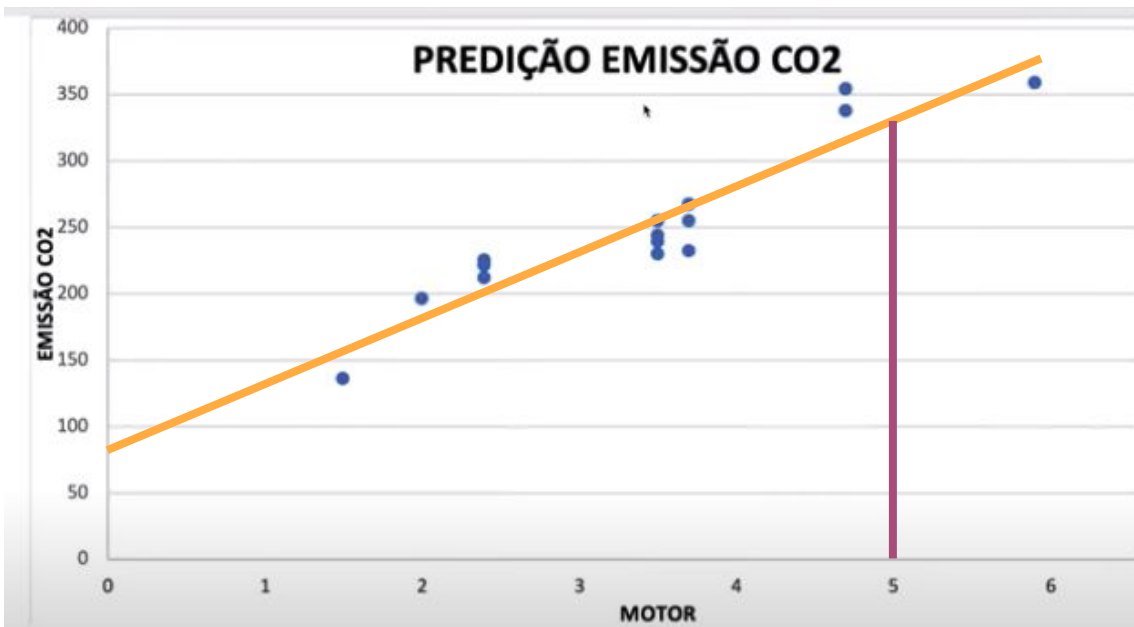
## Aprendizado de Máquina: Regressão

# Como construir a reta?





# Como construir a reta?



$$Y = A + B.X$$

X = Motor

Y = Emissão de CO<sub>2</sub>

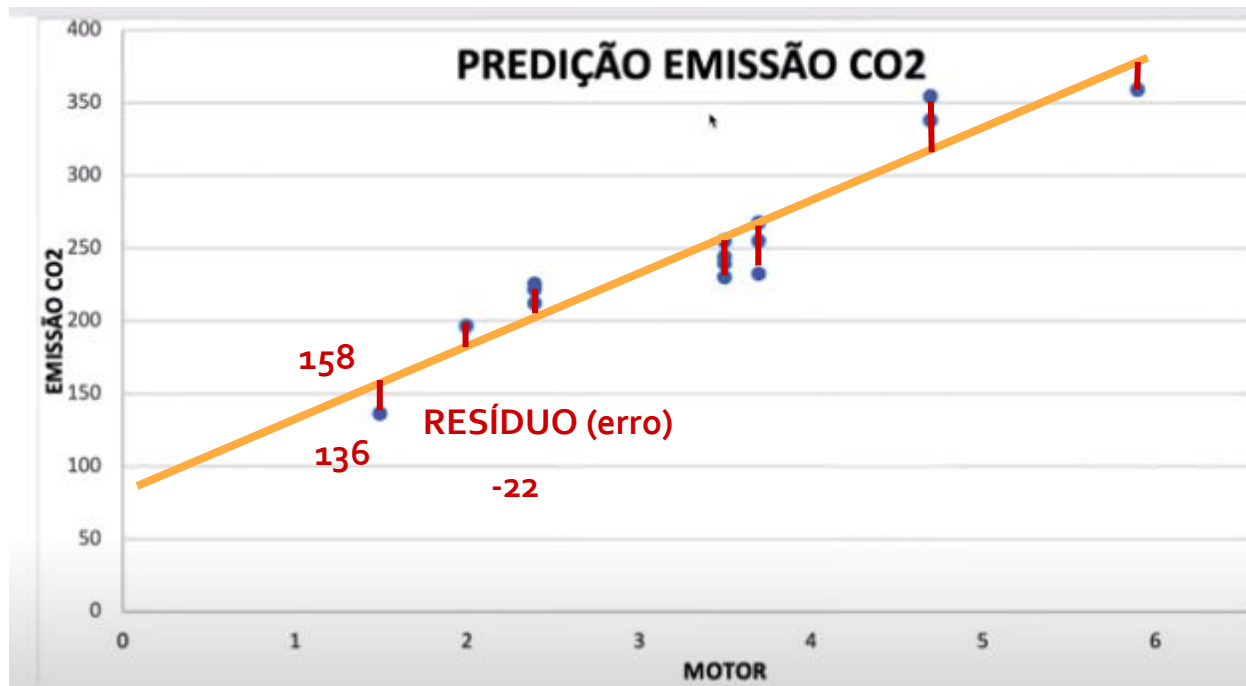
$$Y = 82.4 + 50.14X$$

Para um motor de 5.0, qual seria a emissão de CO<sub>2</sub>?

$$Y = 82.4 + 50.14 * 5.0$$

$$Y = 333.1$$

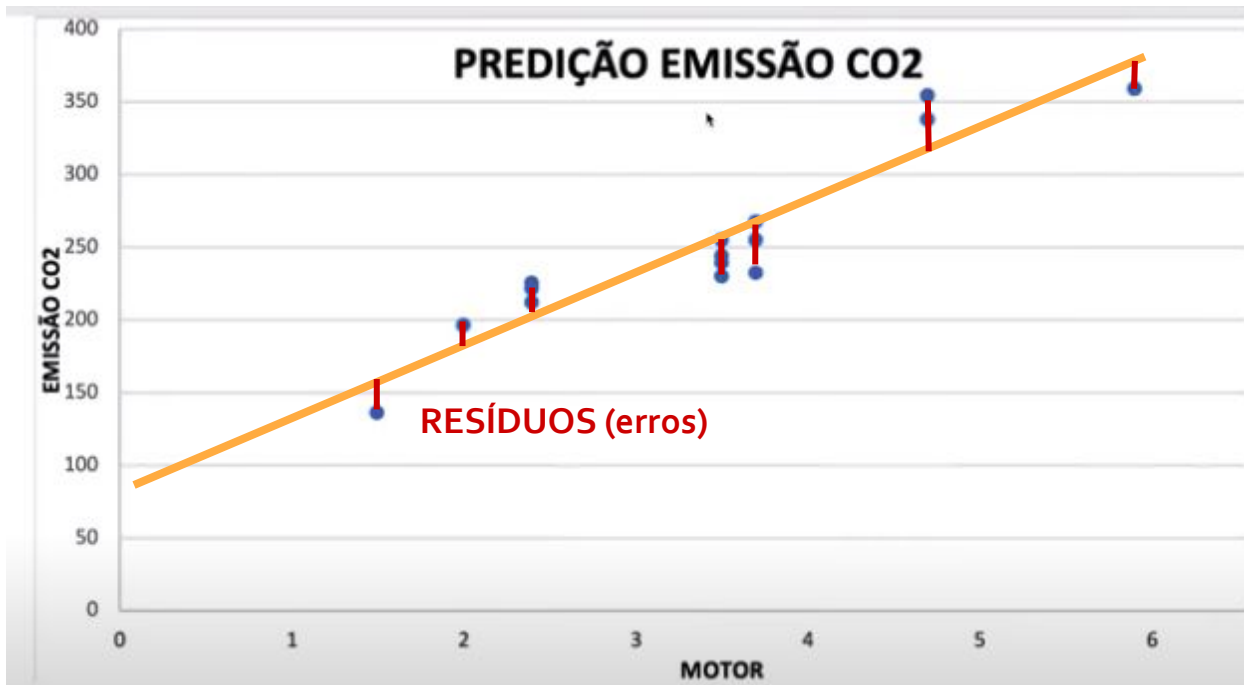
# Como construir a melhor reta?



(quadrados dos)

Soma-se os resíduos (erros)  
de todos os data points!

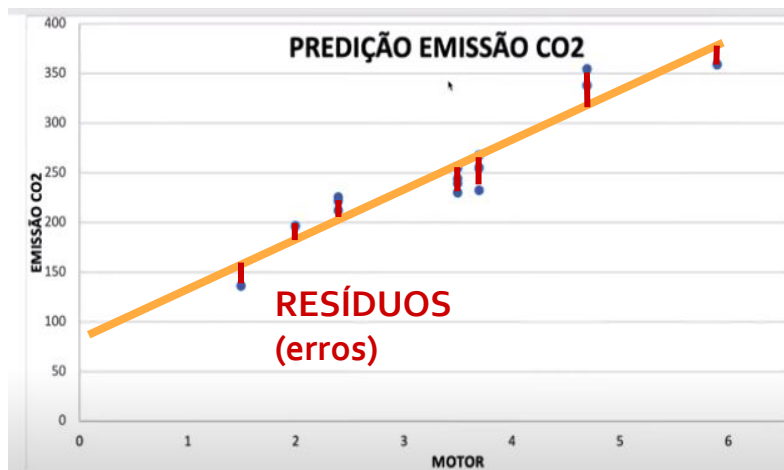
# Método dos Mínimos Quadrados (MMQ)



Ajusta o modelo de modo que a soma dos quadrados das diferenças dos valores observados e previstos seja **minimizada**

$$\min \sum (y_i - \hat{y}_i)^2$$

## Como avaliar o modelo?



<u>Resíduos</u>	<u>Quadrados</u>
-----------------	------------------

-22	505
-----	-----

-1	2
----	---

-13	182
-----	-----

Soma do Erro Quadrático Médio (SSE)	689
-------------------------------------	-----

Erro médio absoluto	12.33
---------------------	-------

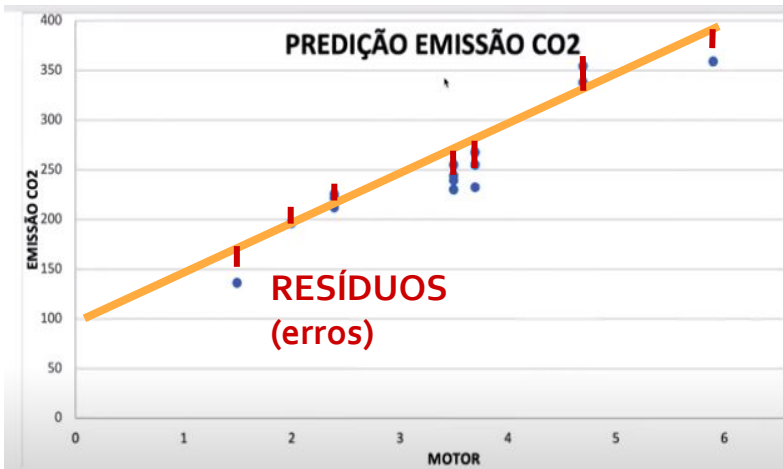
Erro quadrático médio (MSE)	229.66
-----------------------------	--------

Raiz do Erro quadrático médio (RMSE)	15.15
--------------------------------------	-------

R2	0.77
----	------

# Coeficiente de determinação: R<sup>2</sup>

$$r^2 = \frac{SSR}{SST} = \frac{\text{soma dos quadrados da regressão}}{\text{soma total dos quadrados}}$$



R<sup>2</sup> explica a variação da variável dependente em relação a variável independente

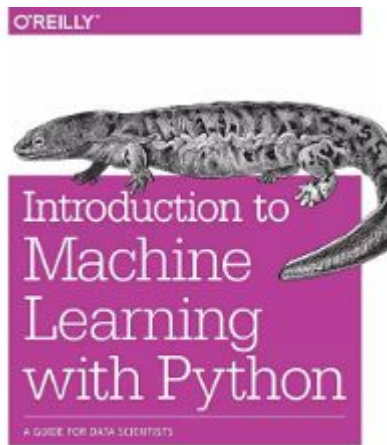
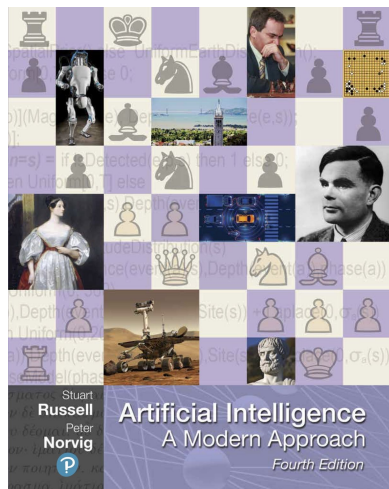
$$R^2 = 0.77$$

O modelo explica 77% da variação de CO<sub>2</sub> em relação ao motor

## Parte prática



# Livros e links!



Andreas C. Müller & Sarah Guido

- [Curso IA para todos do prof. Diogo Cortiz da PUC-SP](#)
- [Livros gratuitos de Ciência de Dados](#)
- [Materiais do prof. Regis da UFC](#)
- [Slides da prof. Erica Castilho Rodrigues - UFOP](#)
- [Slides da prof. Lúgia Henriques-Rodrigues - IME-USP](#)

