

D1EAD – Análise Estatística para Ciência de Dados 2021.1



Regressão Linear Múltipla

Prof. Ricardo Sovat

sovat@ifsp.edu.br

Prof. Samuel Martins (Samuka)

samuel.martins@ifsp.edu.br



Equação da Reta: Exemplos

Regressão Linear Simples

$$\overset{\text{escalar}}{\vdots} h(x_1) = \overset{\text{variável dependente}}{\hat{y}} = \overset{\text{intercepto}}{\vdots} b_0 + \underbrace{b_1}_{\text{inclinação da reta}} * \overset{\text{variável independente}}{\overbrace{x}}$$

modelo linear
(hipótese)

Equação da Reta: Exemplos

Regressão Linear Simples

$$h(x_1) = \hat{y} = b_0 + b_1 * x$$

Diagram labels for Simple Linear Regression:

- $h(x_1)$: modelo linear (hipótese)
- \hat{y} : variável dependente
- b_0 : intercepto
- b_1 : inclinação da reta
- x : variável independente
- escalar: points to $h(x_1)$

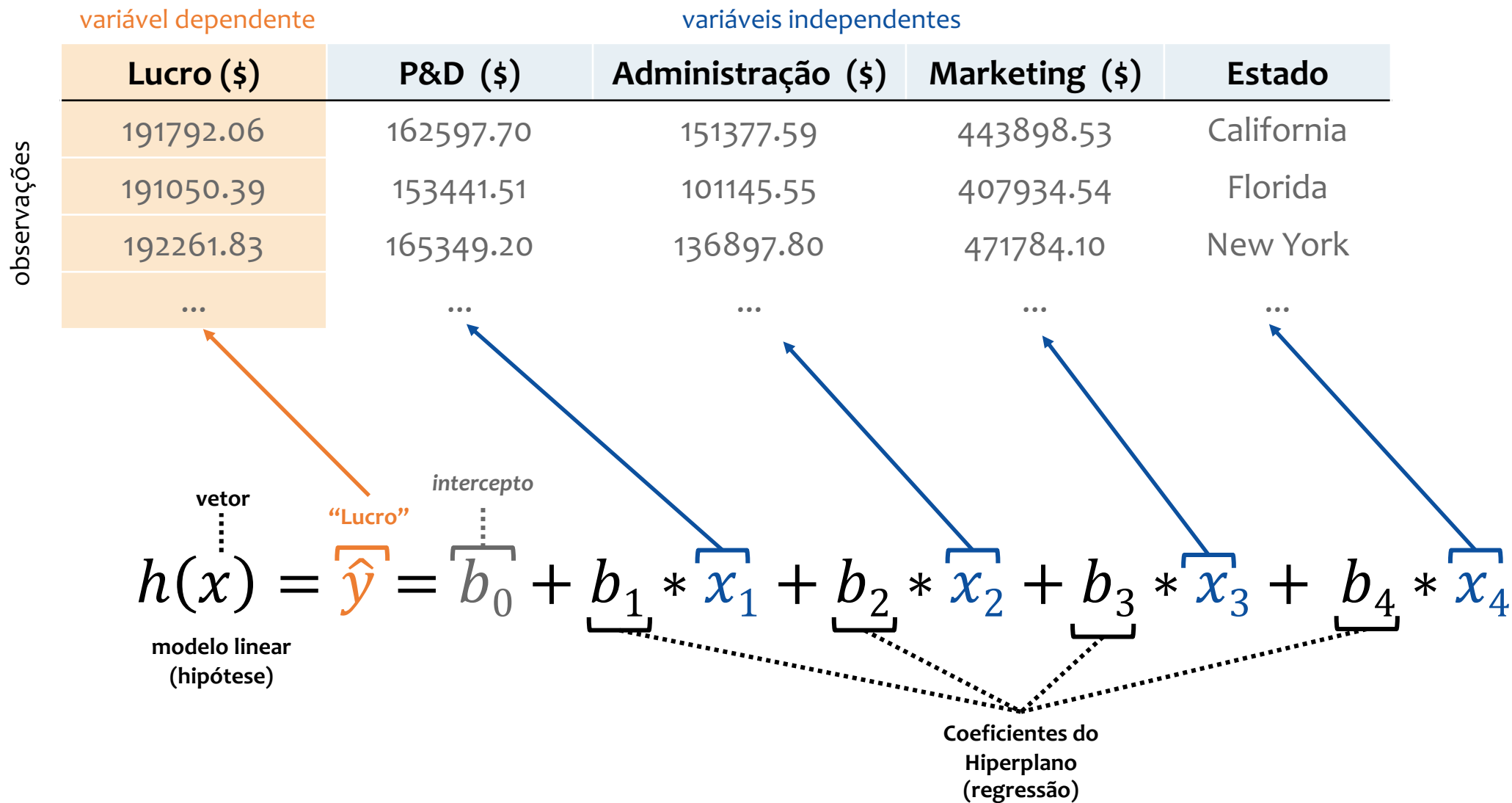
Regressão Linear Múltipla

$$h(x) = \hat{y} = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Diagram labels for Multiple Linear Regression:

- $h(x)$: modelo linear (hipótese)
- \hat{y} : variável dependente
- b_0 : intercepto
- b_1, b_2, \dots, b_n : Coeficientes do Hiperplano (regressão)
- x_1, x_2, \dots, x_n : variáveis independentes
- vetor: points to $h(x)$

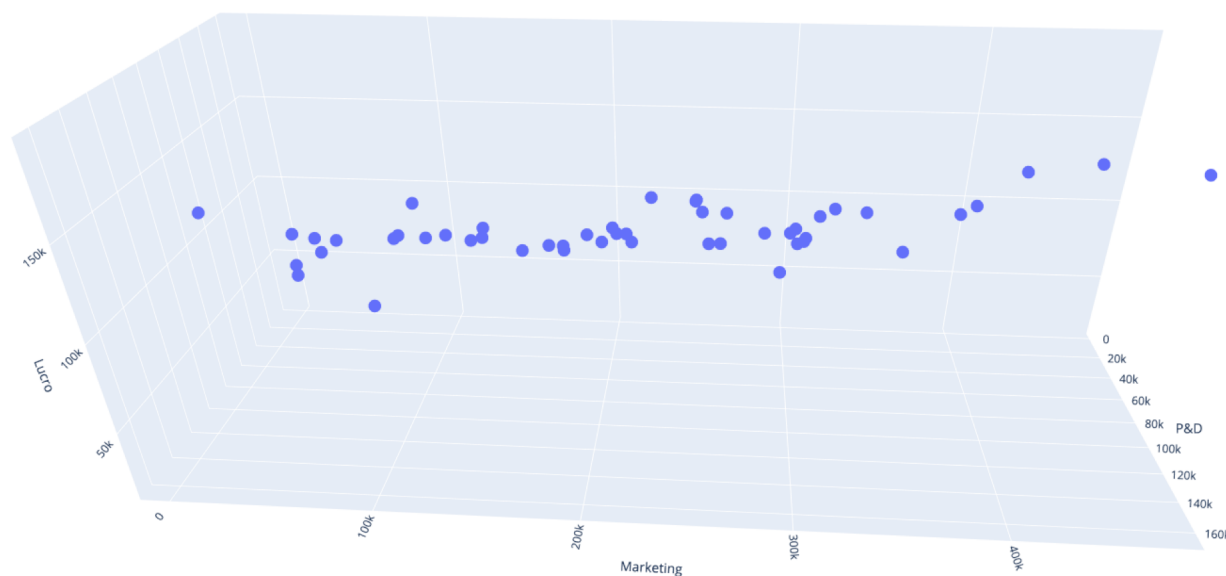
Regressão Linear



Visualizando VD e duas VI

observações	variável dependente	variáveis independentes	
	Lucro (\$)	P&D (\$)	Marketing (\$)
	191792.06	162597.70	443898.53
	191050.39	153441.51	407934.54
	192261.83	165349.20	471784.10

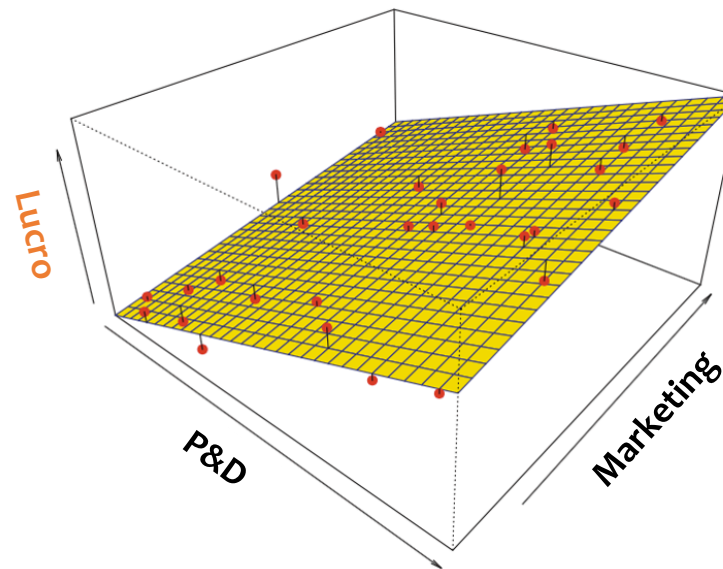
$$h(x) = \hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$$



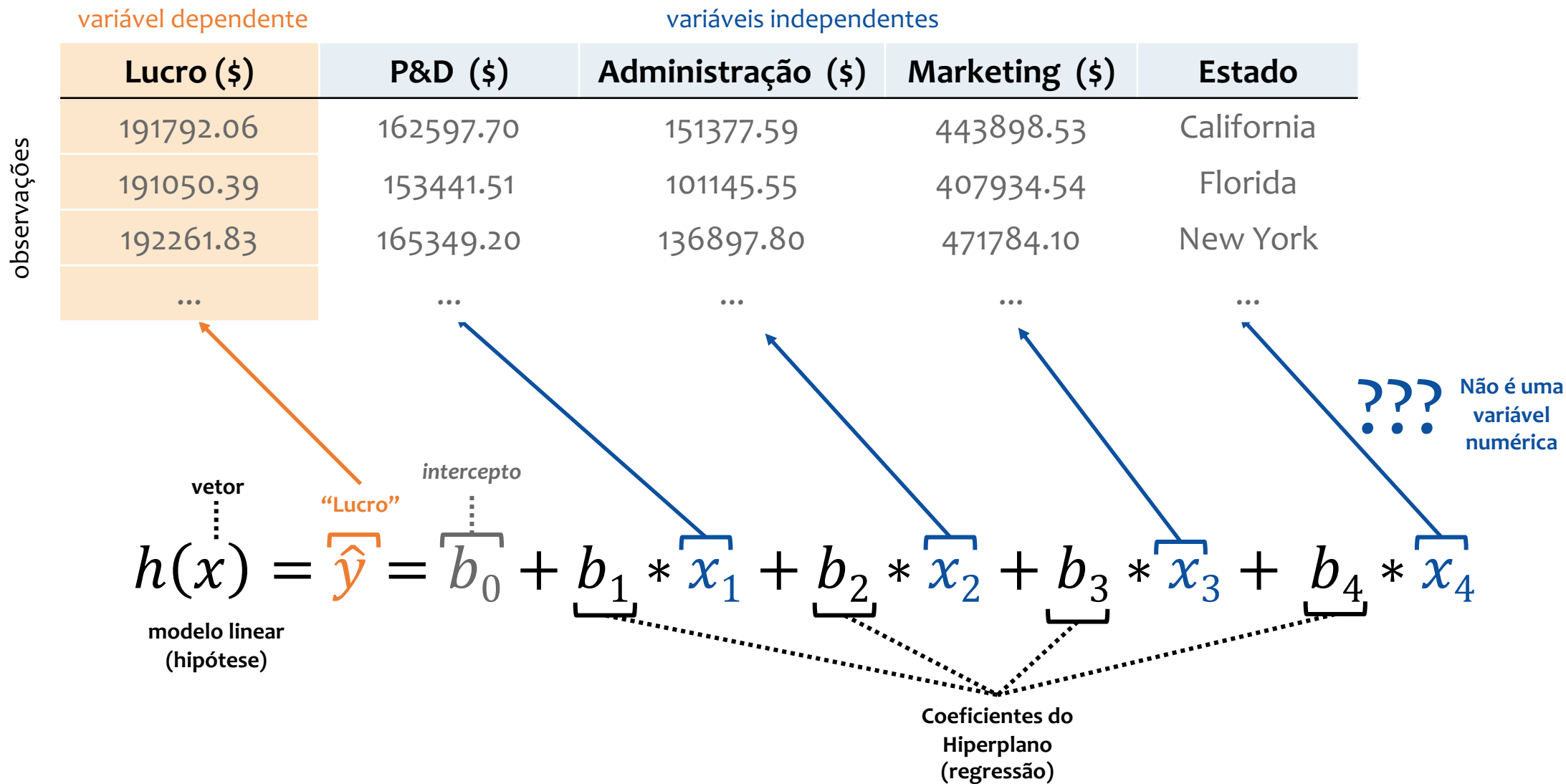
Visualizando VD e duas VI

observações	variável dependente	variáveis independentes	
	Lucro (\$)	P&D (\$)	Marketing (\$)
	191792.06	162597.70	443898.53
	191050.39	153441.51	407934.54
	192261.83	165349.20	471784.10

$$h(x) = \hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$$



Dummy Variables



Dummy Variables

variável dependente		variáveis independentes				
Lucro (\$)	P&D (\$)	Administração (\$)	Marketing (\$)	California	Florida	New York
191792.06	162597.70	151377.59	443898.53	1	0	0
191050.39	153441.51	101145.55	407934.54	0	1	0
192261.83	165349.20	136897.80	471784.10	0	0	1
...

vetor

“Lucro”

intercepto

modelo linear
(hipótese)

Coeficientes do
Hiperplano
(regressão)

$$h(x) = \hat{y} = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 + b_5 * x_5 + b_6 * x_6$$

Dummy Variable Trap

- Incluir **variáveis categóricas** em Regressão Linear, usando **Dummy Variables**, é uma maneira poderosa de incluir dados não-numéricos ao modelo de regressão.
- Entretanto, isso pode causar a **Dummy Variable Trap**
 - Variáveis Independentes são **multi-colineares**: duas ou mais variáveis são **altamente correlacionadas**
 - **New York = 1 – California – Florida**

observações	Lucro (\$)	P&D (\$)	Administração (\$)	Marketing (\$)	California	Florida	New York
	191792.06	162597.70	151377.59	443898.53	1	0	0
	191050.39	153441.51	101145.55	407934.54	0	1	0
	192261.83	165349.20	136897.80	471784.10	0	0	1

Dummy Variable Trap

variável dependente		variáveis independentes				
observações	Lucro (\$)	P&D (\$)	Administração (\$)	Marketing (\$)	Florida	New York
	191792.06	162597.70	151377.59	443898.53	0	0
	191050.39	153441.51	101145.55	407934.54	1	0
	192261.83	165349.20	136897.80	471784.10	0	1

vetero

intercepto

“Lucro”

modelo linear (hipótese)

Coeficientes do Hiperplano (regressão)

$$h(x) = \hat{y} = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 + b_5 * x_5$$

Intepretação dos Coeficientes

Variável Dependente

Lucro

$$\hat{y} = b_0 + \underbrace{b_1}_{\text{intercepto}} * \underbrace{x_1}_{\text{P\&D}} + \underbrace{b_2}_{\text{Administração}} * \underbrace{x_2}_{\text{Marketing}} + \underbrace{b_3}_{\text{Florida}} * \underbrace{x_4}_{\text{New York}} + \underbrace{b_5}_{\text{New York}} * \underbrace{x_5}_{\text{New York}}$$

Coeficientes do
Hiperplano (regressão)

Como escolher os melhores atributos para construir o modelo

1. All-in
2. Backward Elimination
3. Forward Selection
4. Bidirectional Elimination
5. Score Comparison