A primeira Escola presencial gratuita de Inteligência Artificial do Brasil

Aula 17/03/2020: Regressão Linear

Professor: Eng. Rodolfo Magliari de Paiva

Apoio:

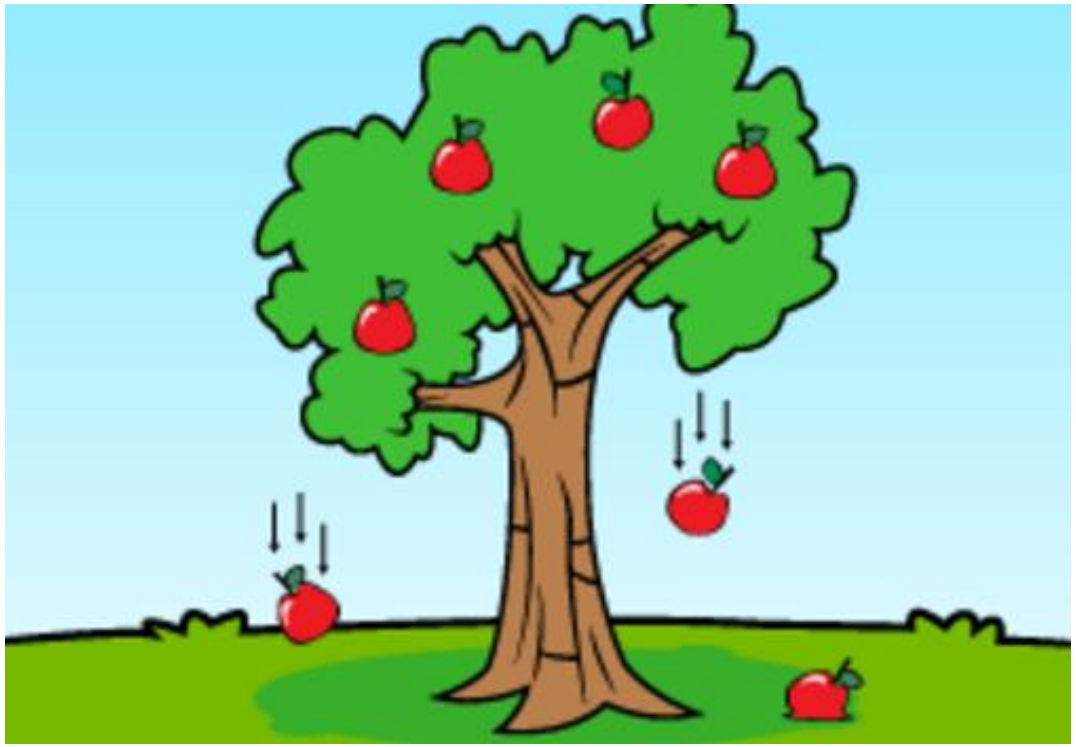
LAMBDA3

Objetivo da Aula

- Entender a diferença entre eventos Determinísticos e Probabilísticos;
- Aprender a interpretar um Gráfico de Dispersão;
- Saber como efetuar e interpretar uma Correlação Linear;
- Compreender o sentido e o objetivo de se efetuar uma Regressão Linear;
- Efetuar uma Regressão Linear.

Eventos

Determinístico x Probabilístico



- Evento Determinístico:

As variáveis são todas conhecidas, sendo possível saber qual será o resultado exato quando efetuarmos o cálculo. *(Matemática)*

- Evento Probabilístico:

Não são conhecidas todas as variáveis, não sendo possível saber o resultado exato, mas sim um resultado esperado. *(Estatística)*

- Evento Determinístico:

Um funcionário de uma empresa ganha R$ 200,00 fixo por mês, mais R$ 150,00 por hora trabalhada na semana.

Quanto ele ganhará se trabalhar 20h em um mês?

f(x) = a.x + b *(Função Polinomial de 1º Grau)*

f(x) = 150.x + 200

f(20) = 150.20 + 200

f(20) = 3200

Ele ganhará R$ 3.200,00

- Evento Probabilístico:

Um motorista de táxi acordou às 7h00 da manhã para ir trabalhar, se ele sair de casa às 7h30 e voltar às 22h, quanto ele ganhará no dia sabendo que ele cobra R$ 5,00 fixo por passageiro e mais R$ 1,50 por Km rodado?

Para responder a essa pergunta precisaríamos responder outras antes:

- Quantas pessoas farão corrida com ele?
- Qual o dia da semana?
- É feriado ou não?
- Houve trânsito?
- Houve acidente no percurso?

...

- Entre outras flutuações aleatórias!

Escola Livre de IA

Saber o futuro não é tarefa fácil!

Porém, é possível tentar se aproximar dele e prever resultados esperados

Algumas técnicas estatísticas que podem ajudar:

- Regressão Linear Simples;
- Regressão Linear Múltipla;
- Regressão Logística;
- Floresta Randômica;
- Árvore de Decisão;
- Redes Neurais;
- Machine Learning;
- ARIMA;
- SARIMA;

...

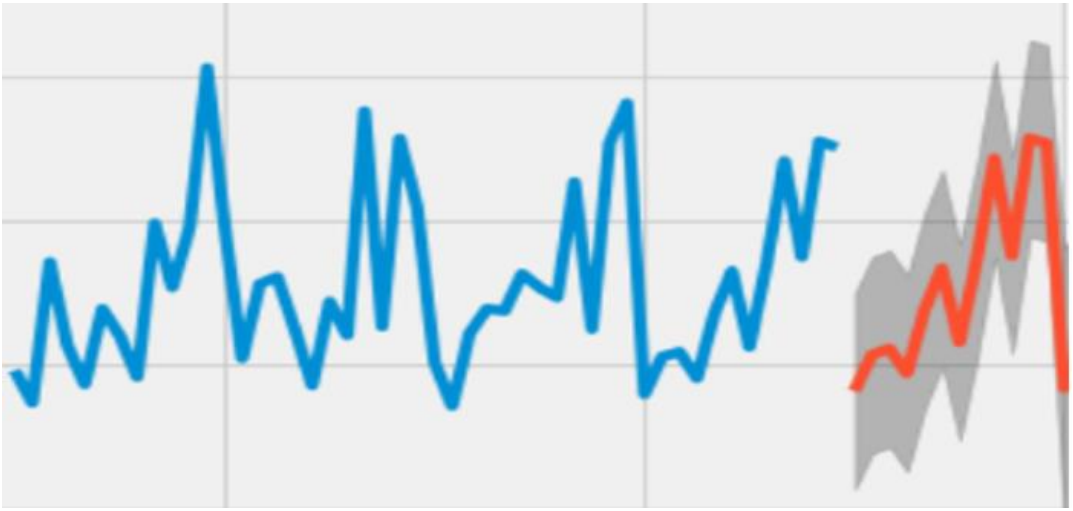
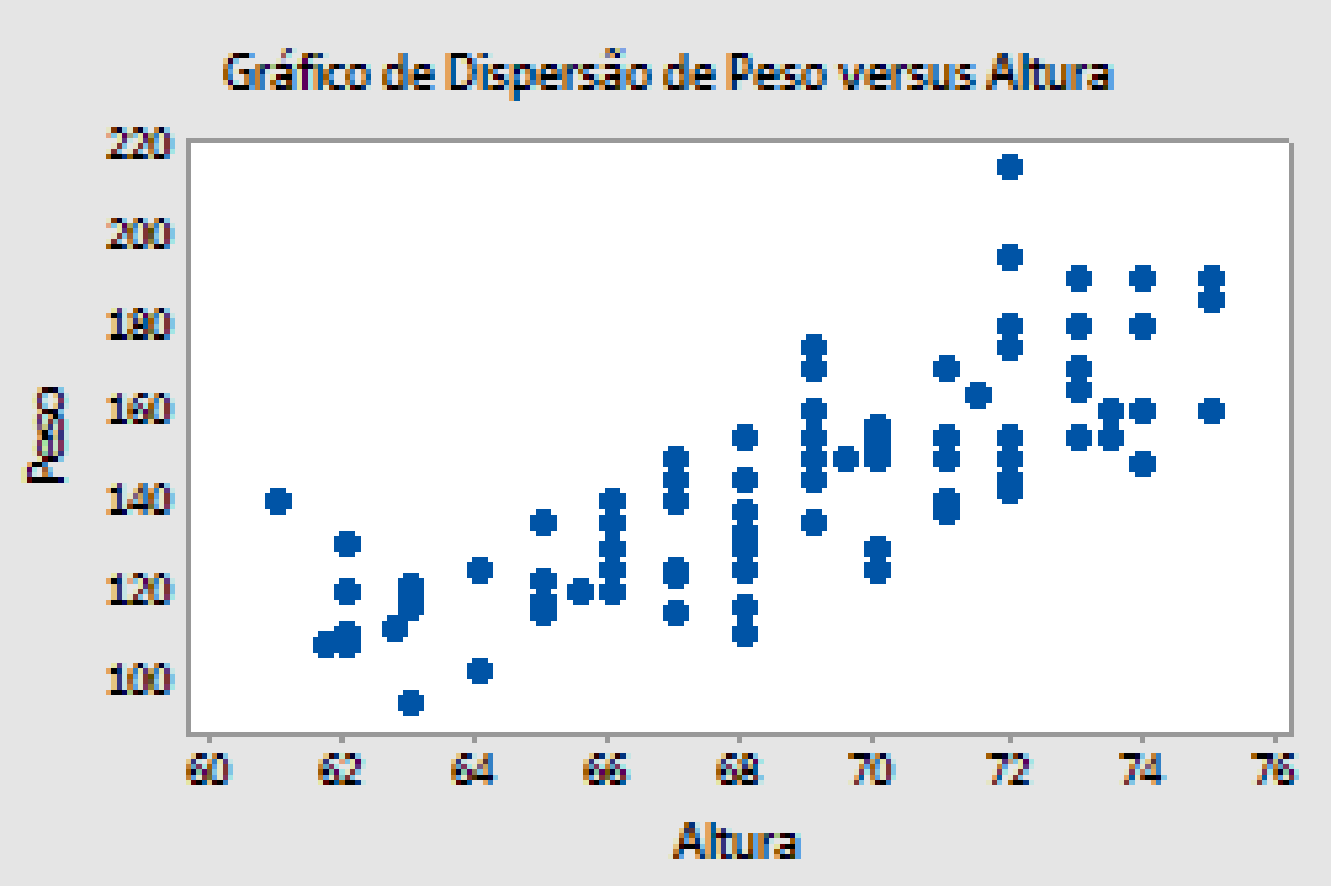


Gráfico de Dispersão

A utilização deste gráfico é muito importante para descobrir se duas variáveis podem estar associadas:



Escola Livre de IA

Correlação Linear de Pearson

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

OBS: Válido apenas para mostrar a associação entre variáveis **quantitativas**.

Interpretação:

$$-1 \leq \rho \leq 1$$

Onde:

$\rho = 1$, correlação linear perfeita positiva;

$\rho = -1$, correlação linear perfeita negativa;

$\rho = 0$, não existe correlação.

Valor de ρ (+ ou −)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Fonte: Shimakura, 2006

Exemplos:

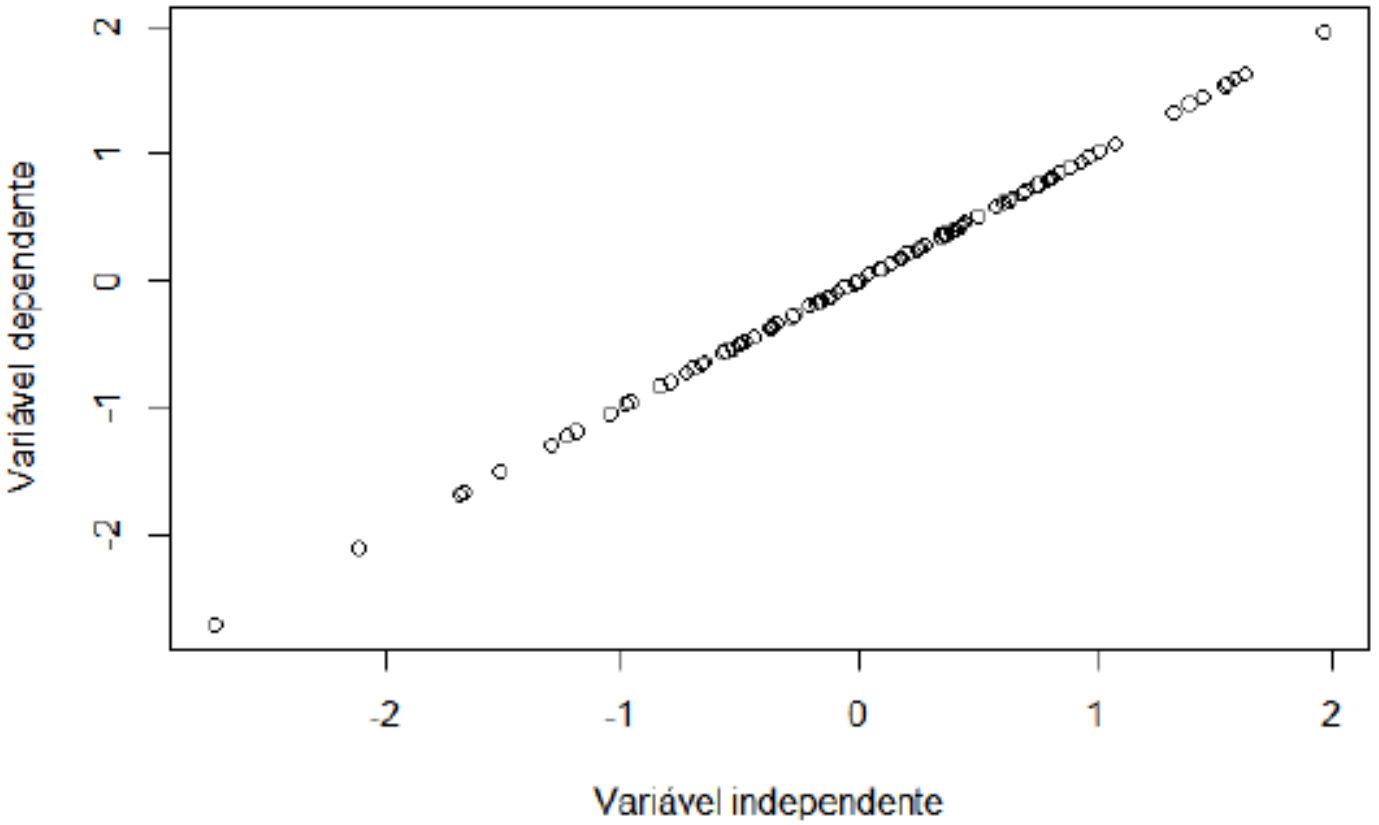


Figura 1: Correlação linear perfeita positiva.

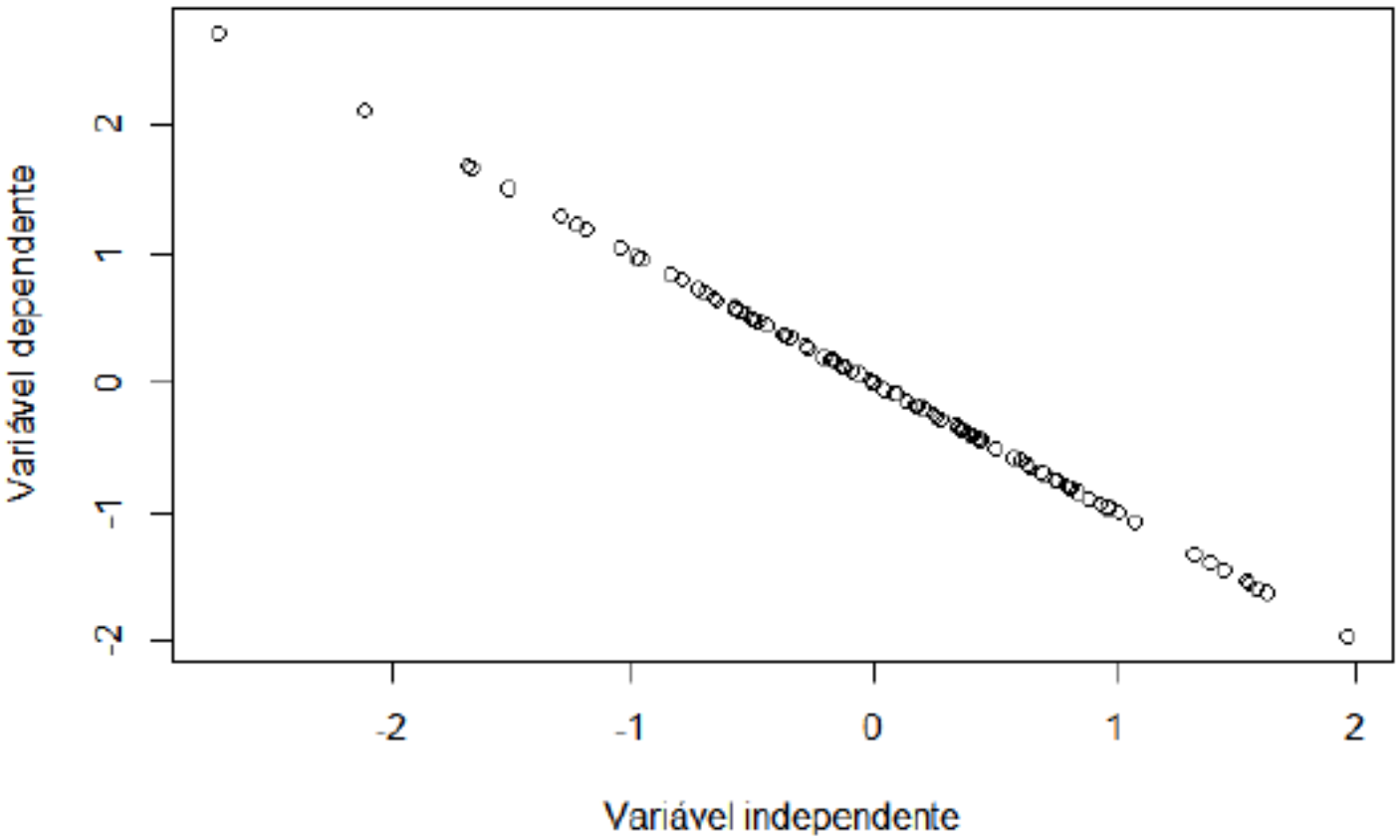


Figura 2: Correlação linear perfeita negativa.

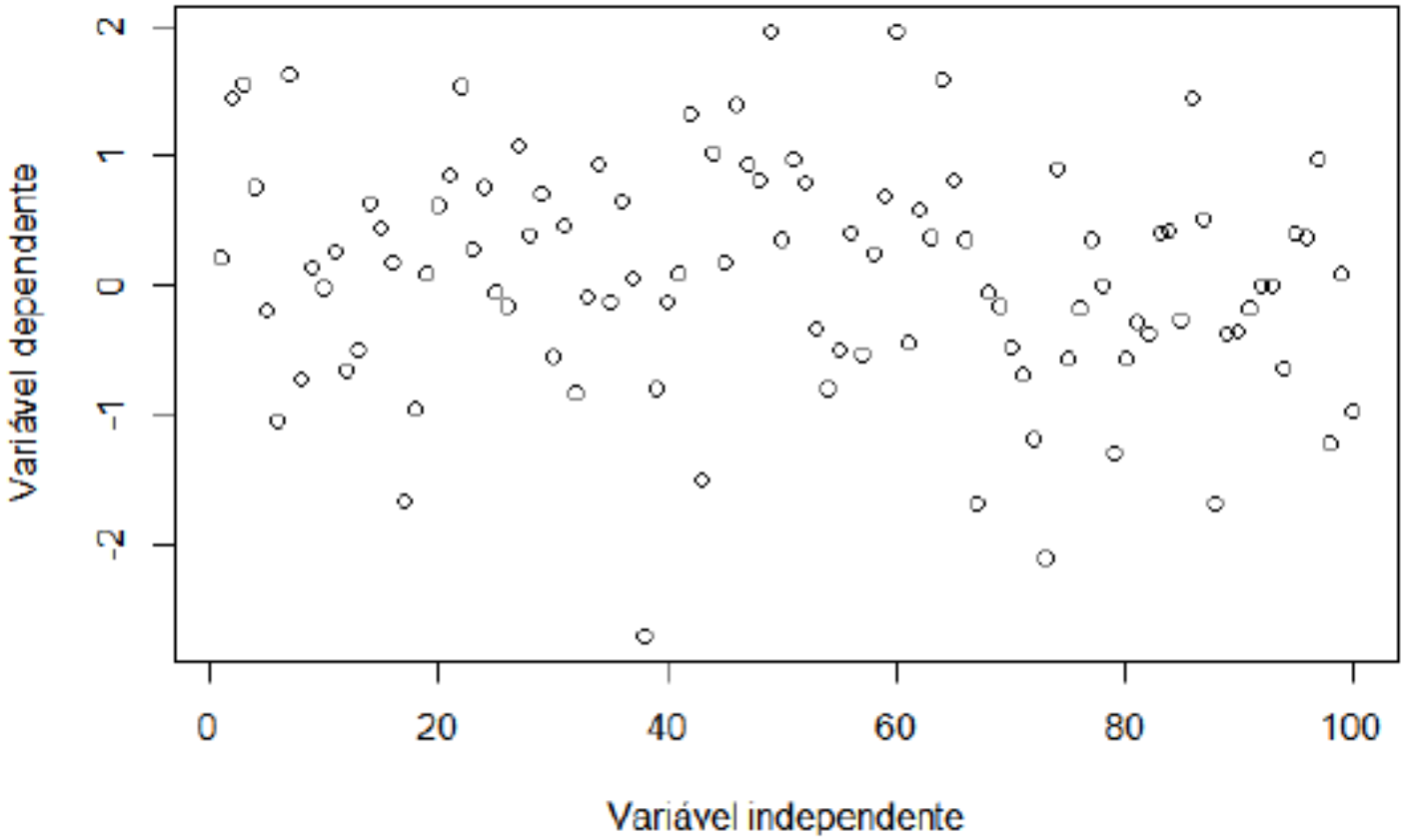


Figura 3: Não existe correlação.

Escola Livre de IA

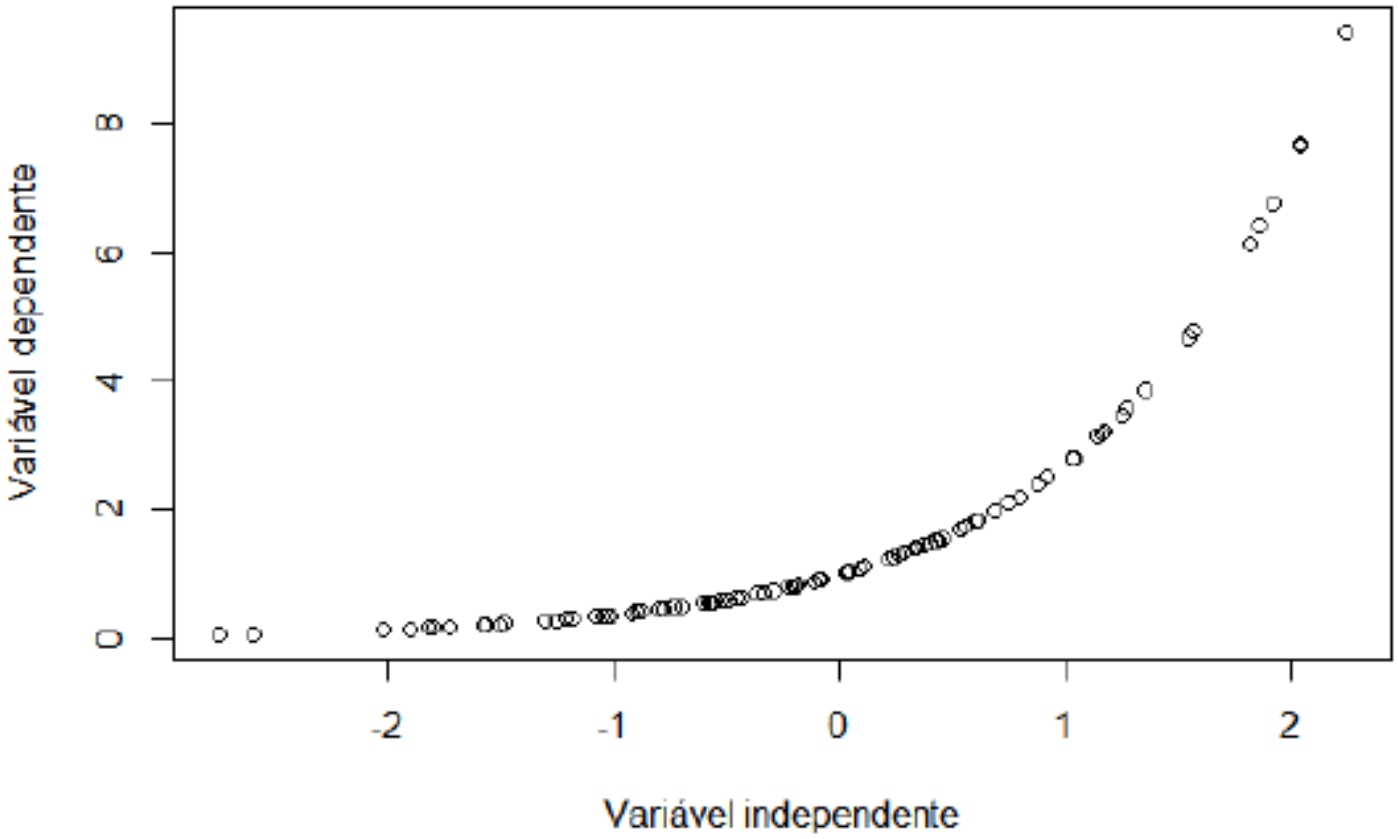


Figura 4: Não existe relação linear.

Atenção

Associação não implica em **Causa e Efeito**!

Regressão Linear Simples

Técnica excelente para ser aplicada quando existe uma boa Correlação Linear (positiva ou negativa), e é dada por:

$$Y = \alpha + \beta X + \varepsilon$$

(Erro, resultado de flutuações aleatórias)

Onde:

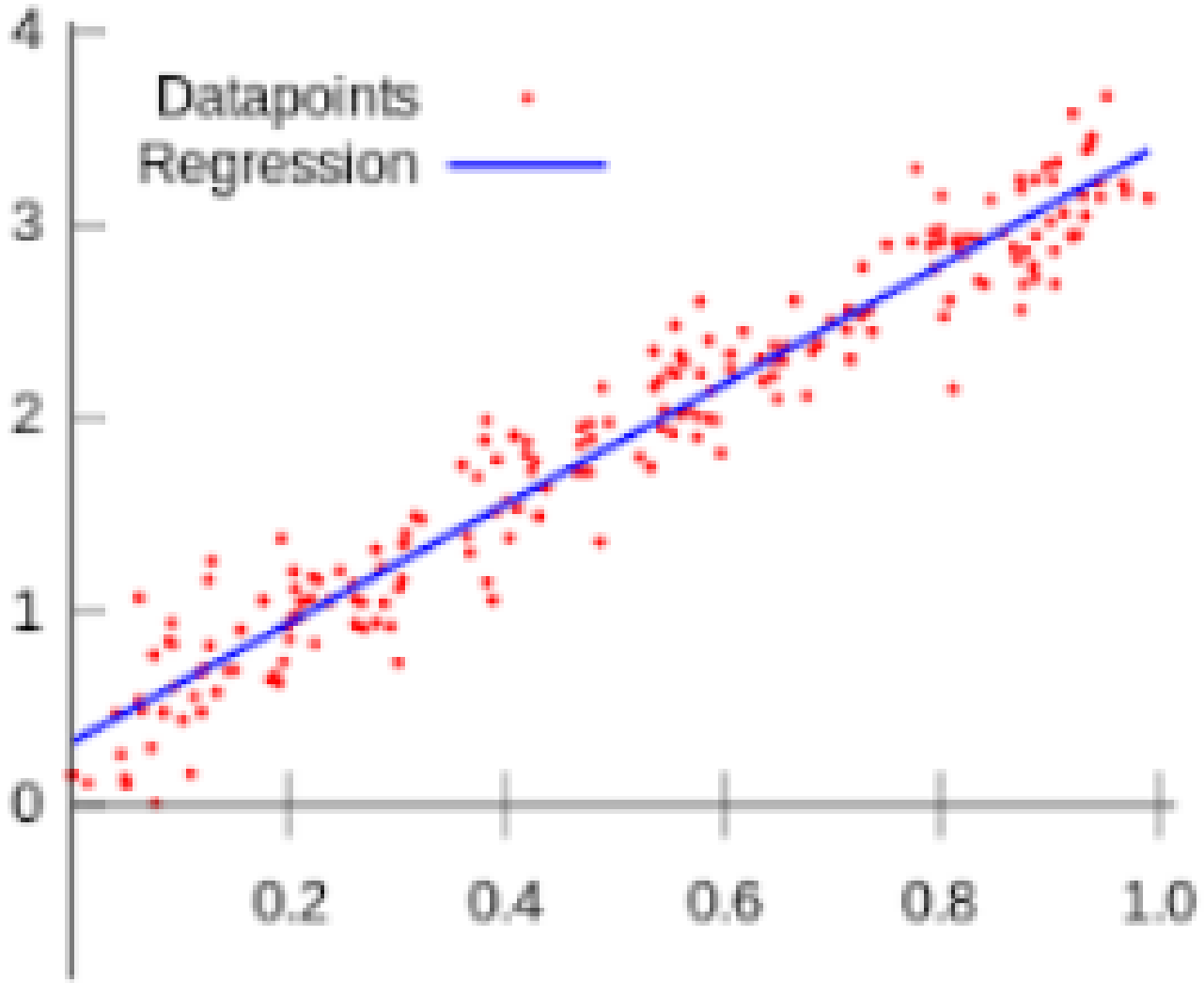
(Coeficiente Angular)

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{(n-1)S_X^2},$$

(Intercepto)

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

Escola Livre de IA



Outro item que é interessante analisar: R^2

- É um indicador que mede a qualidade do ajuste na Regressão Linear. Seu resultado varia de 0 a 1;

- Quanto mais próximo de 0, menos a Regressão Linear se ajustou;

- Quanto mais próximo de 1, mais a Regressão Linear se ajustou.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{predito})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Escola Livre de IA

Exercícios

1-) Qual a diferença entre um evento determinístico e probabilístico? Dê um exemplo para cada.

2-) Qual a importância de elaborar um gráfico de dispersão entre duas variáveis?

3-) O que mede o Coeficiente de Correlação Linear de Pearson?

4-) Dada a tabela a seguir, calcule o coeficiente de Correlação Linear de Pearson e interprete o resultado.

X	Y
4	2
6	3
8	4

5-) Com base no resultado anterior, seria possível realizar uma previsão utilizando a Regressão Linear Simples?

6-) Com base na tabela do exercício **4-)** faça o que se pede:

a-) Dê a Equação da Reta e o Gráfico de Dispersão com o ajuste da reta

b-) Sabendo que o $R^2 = 1$, o que isso significa?

c-) Para uma observação futura de X = 10, qual deve ser o resultado esperado para Y?

Gabarito

1-) Em um evento determinístico as variáveis são conhecidas e o resultado é exato. Já em um evento probabilístico não são conhecidas todas as variáveis e o resultado é sempre algo esperado.
Exemplo (livre): Ao soltar uma caneta de uma determinada altura é 100% de certeza que ela irá cair, porém não é possível determinar a posição que ela irá ficar no chão.

2-) Descobrir se as variáveis possuem alguma tendência a estarem associadas, entendendo o comportamento do fenômeno.

3-) Mede a associação entre duas variáveis quantitativas de forma linear crescente ou decrescente.

4-)

	X	Y	X - Xmédia	Y - Ymédia
	4	2	-2	-1
	6	3	0	0
	8	4	2	1
Média	**6**	**3**	-	-

	(X - Xmédia).(Y - Ymédia)	$(X - Xmédia)^2$	$(Y - Ymédia)^2$	$(X - Xmédia)^2 . (Y - Ymédia)^2$
	2	4	1	16
	0	0	0	$\sqrt{[(X - Xmédia)^2 . (Y - Ymédia)^2]}$
	2	4	1	
Soma	**4**	**8**	**2**	**4**

$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$	4	1
	4	

O Coeficiente de Correlação Linear deu 1, logo é possível dizer que as variáveis X e Y estão fortemente associadas de forma positiva.

5-) Sim, pois é existe uma correlação linear forte.

6-)

a-)

X	Y	X.Y
4	2	8
6	3	18
8	4	32
Soma		**58**

n
3

n.Xmédia.Ymédia
54
X.Y - n.Xmédia.Ymédia
4
$(n-1)S_X^2$
8

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{(n-1)S_X^2}$$

0,5

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

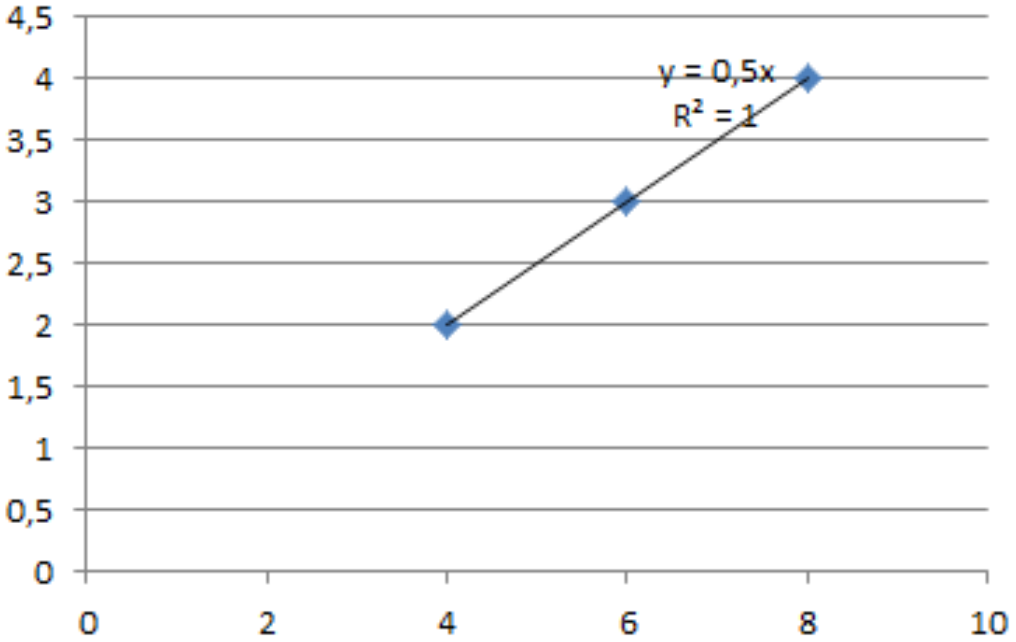
0

Eq. da Reta

$$Y = \alpha + \beta X + \varepsilon$$

Y = 0 + 0,5x + Erro Aleatório

Escola Livre de IA



b-) Significa que a variável X responde 100% pela qualidade do modelo, ou seja, o modelo de Regressão Linear é válido.

c-) Y = 0,5X + Erro Aleatório

Y = 0,5.10 + Erro Aleatório

Y = 5 + Erro aleatório

Bibliografia

MONTGOMERY, Douglas C. e RUNGER, George C.
Estatística Aplicada e Probabilidade para Engenheiros.
6ª Edição. Rio de Janeiro: GEN|LTC, 2016

Escola Livre de IA

Contatos

Prof. Eng. Rodolfo Magliari de Paiva

Cel.: (11) 9-6866-5501

E-mail: rodolfomagliari@gmail.com

LinkedIn: Rodolfo Magliari de Paiva

Escola Livre de IA