Escola Livre de Inteligência Artificial

Inteligência Artificial ao alcance de todos

Aula 08/06/2021: Regressão Linear Múltipla

Professor: Eng. Rodolfo Magliari de Paiva

Objetivos da Aula

- Compreender o que é uma Análise de Regressão e seus tipos;
- 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 Múltipla;
- Interpretar uma Regressão Linear Múltipla.

Análise de Regressão

Parte da Estatística que estuda a relação entre duas ou mais variáveis (dependentes e independentes), de modo que seja possível identificar quais variáveis possuem maior ou menor impacto em um fenômeno de estudo, além de também permitir a explicação de um fenômeno e prever o futuro.

Para isso, utilizamos os chamados **Modelos de Regressão**.

Os Modelos de Regressão podem ser Lineares (ML) ou Não Lineares (MNL):

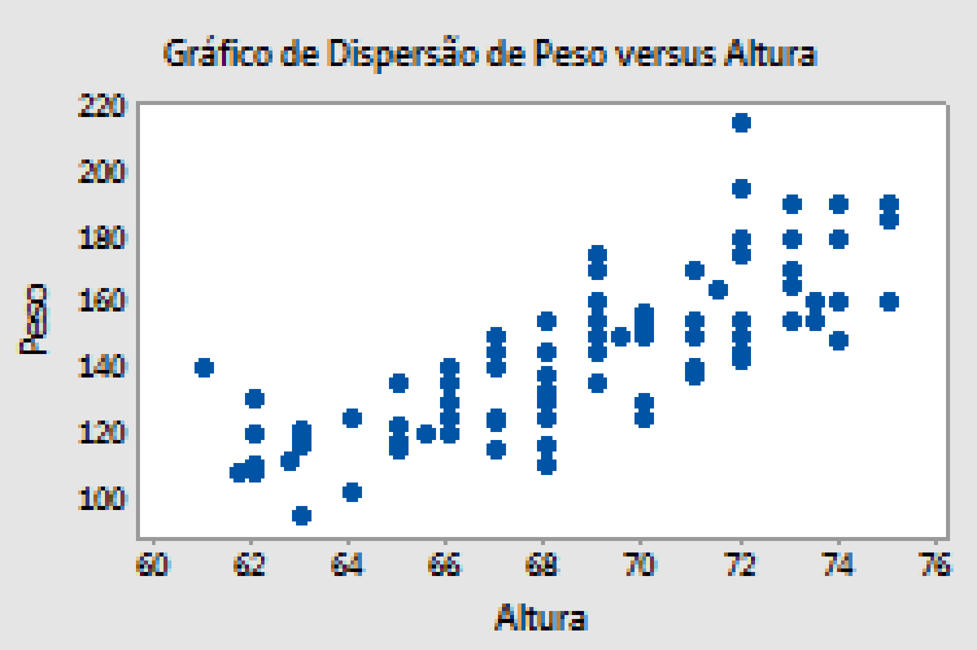
- **Modelos de Regressão Linear:**
  - Regressão Linear Simples;
  - Regressão Linear Múltipla.

- **Modelos de Regressão Não Linear:**
  - Regressão Exponencial;
  - Regressão Logística Simples;
  - Regressão Logística Múltipla;
  - Regressão Poisson;

  …

Gráfico de Dispersão

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



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

Vejamos como podem ser os resultados gráficos!

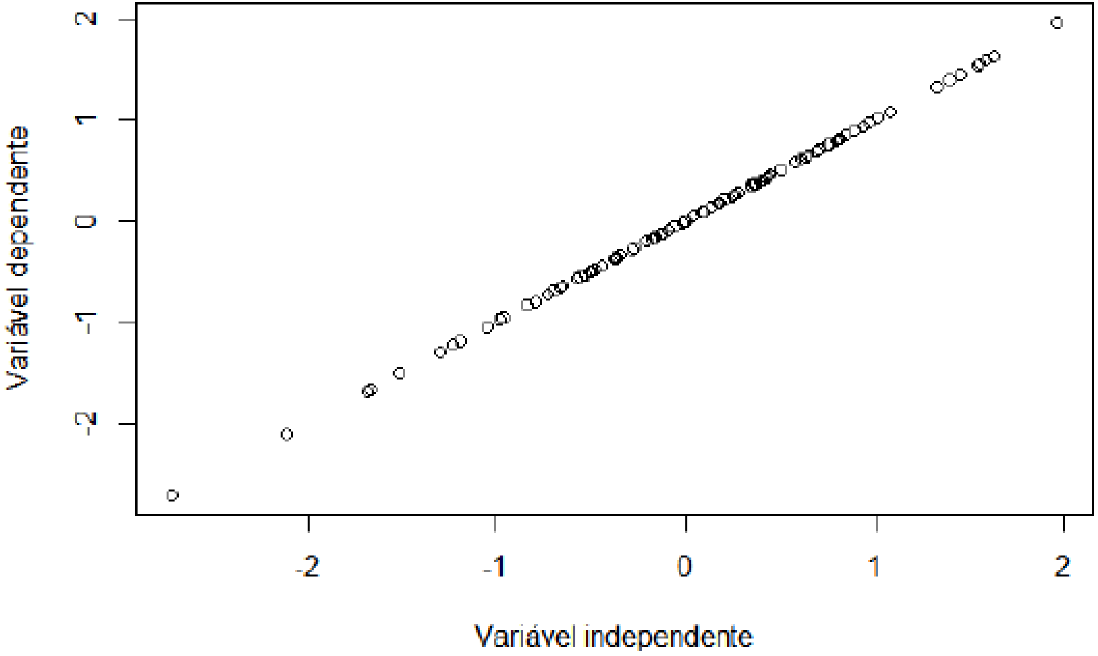


Figura 1: Correlação linear perfeita positiva.

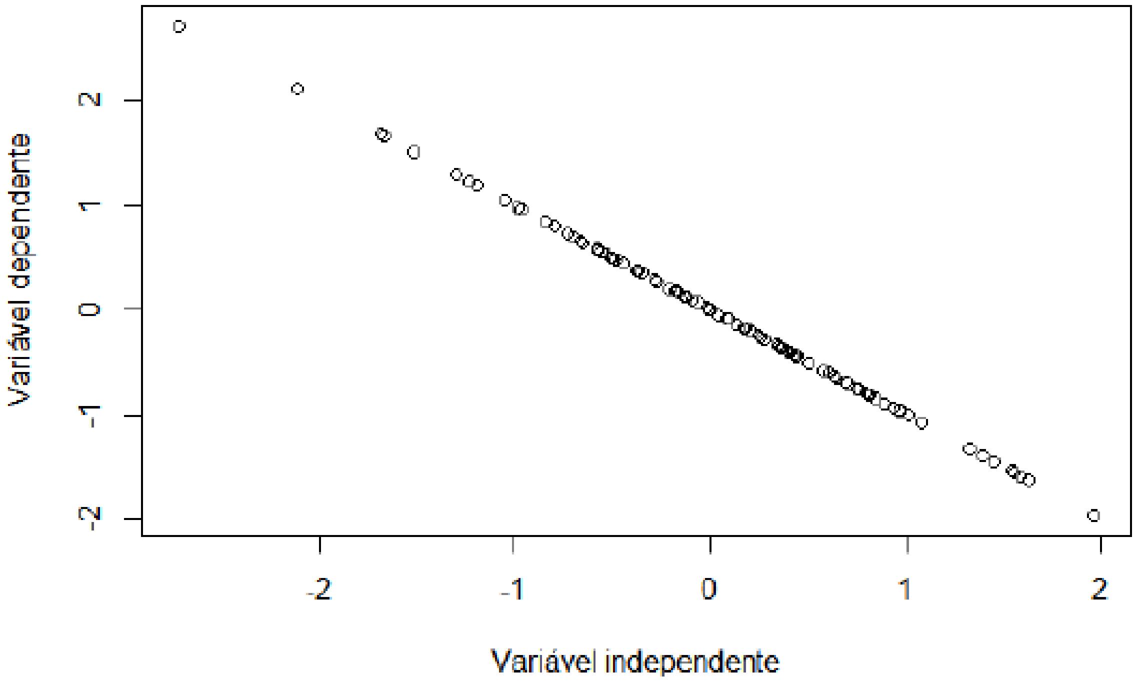


Figura 2: Correlação linear perfeita negativa.

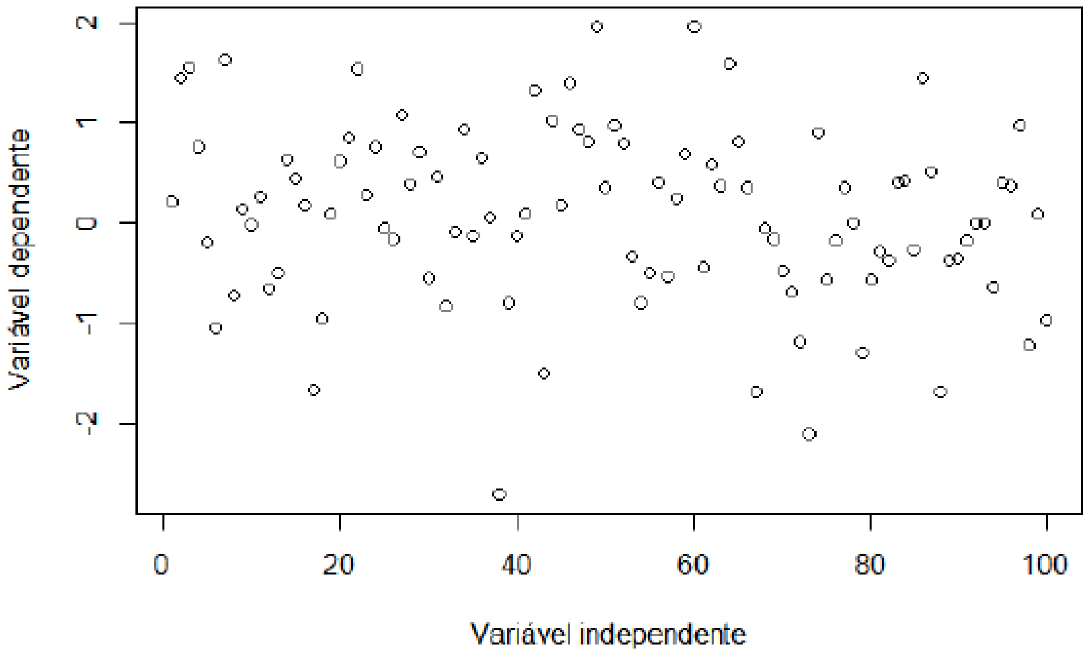


Figura 3: Não existe correlação.

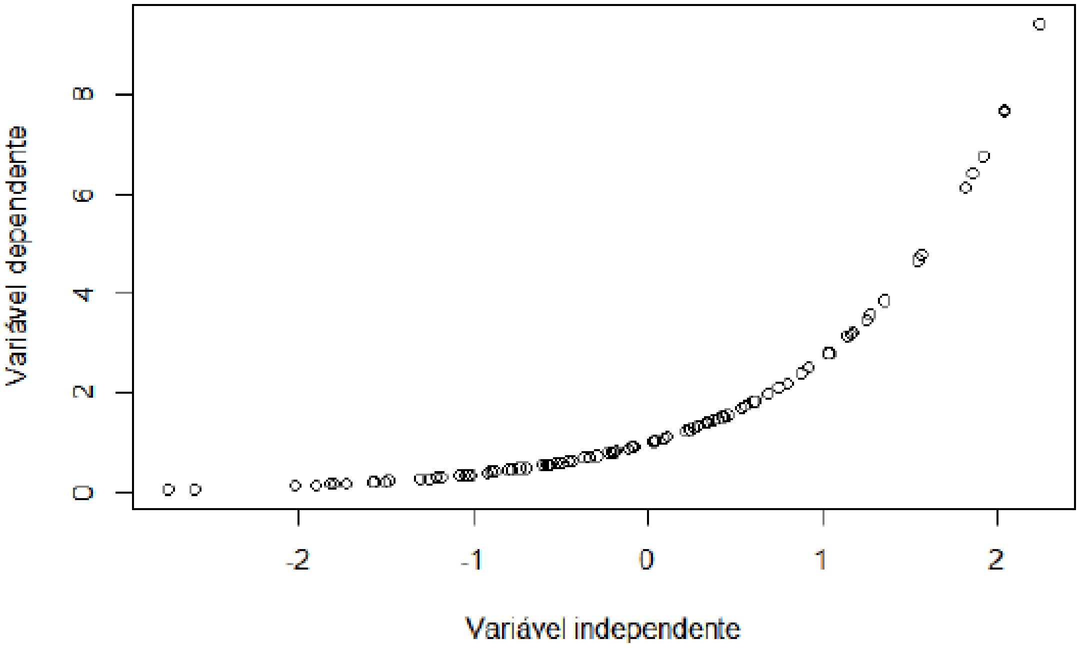


Figura 4: Não existe relação linear.

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

Regressão Linear Múltipla

É uma generalização da Regressão Linear Simples para um situação em que há mais de uma variável preditiva ou regressora (variável independente)

Para aplicarmos o modelo de Regressão Linear Simples, é necessário cumprir alguns pré-requisitos:

- **Linearidade:** A associação entre x e y deve ser linear (reta), verificação por meio do Gráfico de Dispersão;

- **Independência dos valores de y:** Cada valor de y é independente;

- **Distribuição Normal:** A variável y deve ser quantitativa e com Distribuição Normal, verificação por meio do Histograma e/ou por meio de testes de normalidade (como o clássico Teste de Shapiro-Wilk);

- **Homocedasticidade:** Variância de y deve ser a mesma independentemente do valor de x, verificação por meio da análise dos resíduos;

- **Sem correlação entre variáveis preditivas:** Estas não podem estar altamente correlacionadas, verificação por meio da Correlação Linear de Pearson ou por meio do VIF (Fatores de Inflação de Variância).

A equação do modelo de uma Regressão Linear Simples é dada por:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i$$

(ε = Erro, resultado de flutuações aleatórias)

Onde:

β_0 = Intercepto / Coeficiente Angular

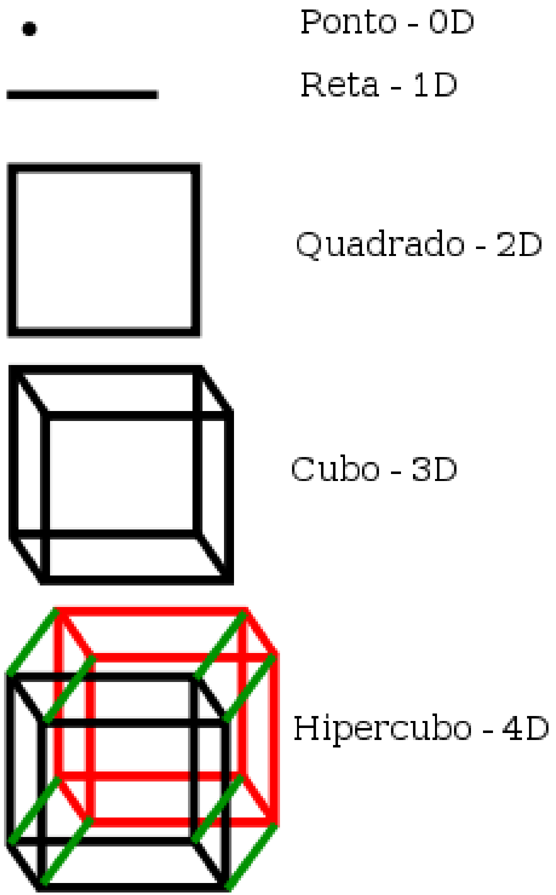
β_1 = Coeficiente Angular

β_2 = Coeficiente Angular

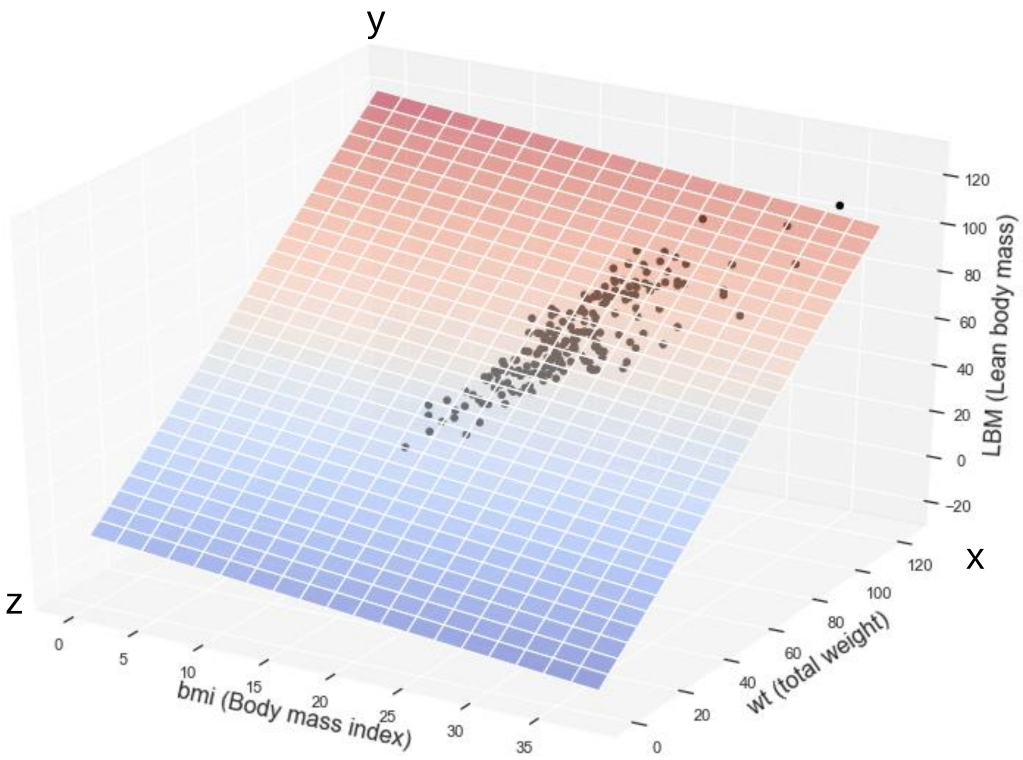
β_p = Coeficiente Angular

O gráfico da Regressão Linear Múltipla é um **hiperplano** no **Espaço k-dimensional**.

Vamos entender o conceito de dimensão:



Para duas variáveis preditivas ou regressoras é um **plano** no **Sistema Triortogonal**:



Vale lembrar:
Eixo y = Eixo das Ordenadas
Eixo x = Eixo das Abscissas
Eixo z = Eixo das Cotas

Após aplicar o modelo é interessante analisar o **R^2 (Coeficiente de Determinação Múltipla)**

É um indicador que mede a qualidade do ajuste na Regressão Linear Múltipla, ou seja, quanto que as variáveis preditivas (juntas), explicam Y.
Seu resultado varia de 0 a 1, sendo que:
- Quanto mais próximo de 0, menos a Regressão Linear Múltipla se ajustou;
- Quanto mais próximo de 1, mais a Regressão Linear Múltipla se ajustou.

O calculamos fazendo:

$$R^2 = \frac{SQR}{SQT} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

Exemplo

Uma imobiliária possui um modelo de Regressão Linear Múltipla para prever o valor dos apartamentos (R$), em função do tamanho dos apartamentos (m²), e em função da idade dos prédios (anos).
A equação do modelo é dada por:

$$Y = 440.107 + 6.772{,}1x_1 - 19.129{,}7x_2 + \varepsilon$$

Sendo x_1 o tamanho do apartamento e x_2 a idade do prédio, qual o valor estimado para um apartamento que possui 55m² e o prédio tem 25 anos de idade?

Resolução:

$$Y = 440.107 + 6.772{,}1x_1 - 19.129{,}7x_2 + \varepsilon$$

$$Y = 440.107 + 6.772{,}1 \cdot 55 - 19.129{,}7 \cdot 25 + \varepsilon$$

$$Y = 440.107 + 372.465{,}5 - 478.242{,}5 + \varepsilon$$

$$Y = 334.330 + \varepsilon$$

Conclusão

Com essas ferramentas da **Análise de Regressão** é possível entender a relação entre duas variáveis e tentar prever o comportamento de uma delas, basta **ter** ou **iniciar** a coleta de dados e na sequência:

APLICAR!

Exercícios

1) Em que tipo de situação devemos realizar uma Regressão Linear Múltipla ao invés de uma Regressão Linear Simples?

2) Para que seja possível realizar uma Regressão Linear Múltipla, quis pré-requisitos devemos cumprir?

3) Explique a Correlação Linear de Pearson.

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

Suprimento de Voltagem	Corrente sem Eletrônicos (mA)
0,66	7,32
1,32	12,22
1,98	16,34
2,64	23,66
3,3	28,06

5) Com relação ao VIF responda:

a) O que é?

b) Para que serve?

c) Como o interpretamos?

6) A equação de Regressão Linear Múltipla de um laboratório de Engenharia para prever a resistência à tração de fios é dada por:

$$Y = 2{,}26379 + 2{,}4427x_1 + 0{,}01253x_2 + \varepsilon$$

Sendo x_1 o comprimento do fio (m), e x_2 a atura da garra (m), qual a resistência (Pa) esperada se o comprimento do fio for de 2 m e a altura da garra for de 4 m?

7) Um modelo de Regressão Linear Múltipla com 3 variáveis apresentou R^2 de 0,67, o que isso significa?

Gabarito

1) Quando existe mais de uma variável preditiva ou regressora para explicar y.

2) Devemos cumprir:
Linearidade, Independência dos valores de y, Distribuição Normal, Homocedasticidade e sem correlação entre variáveis preditivas.

3) Mede a associação entre duas variáveis quantitativas, sendo que:
Quanto mais próximo de 1, as variáveis estão fortemente associadas de forma positiva;
Quanto mais próximo de 0, as variáveis não estão associadas;
Quanto mais próximo de -1, as variáveis estão fortemente associadas de forma negativa.

4) O Coeficiente de Correlação Linear deu 0,99 , logo é possível dizer que as variáveis X e Y estão fortemente associadas de forma positiva.

5)

a) VIF é o Fator de Inflação de Variância

b) Serve para medir a multicolinearidade entre as variáveis preditivas ou regressoras. Avaliando o quanto a variância de um coeficiente de regressão estimado aumenta se as suas preditoras estiverem correlacionadas.

c) Se o resultado for:

1 = Não correlacionado

1 – 5 = Correlação Moderada

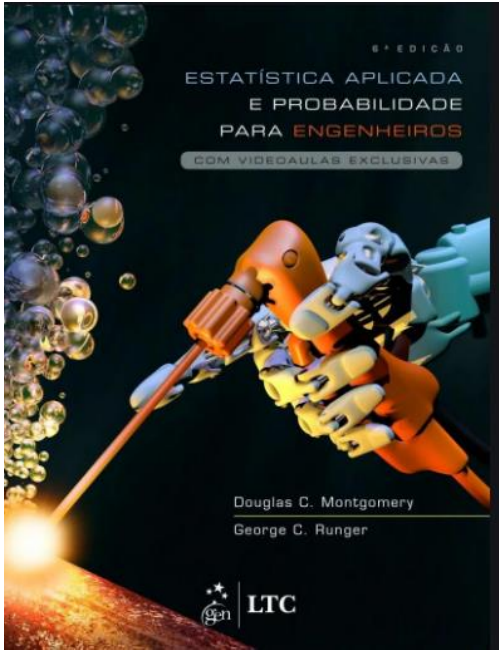
> 5 = Altamente Correlacionado

6) Espera-se que a resistência seja de aproximadamente 7,20 Pa + ε

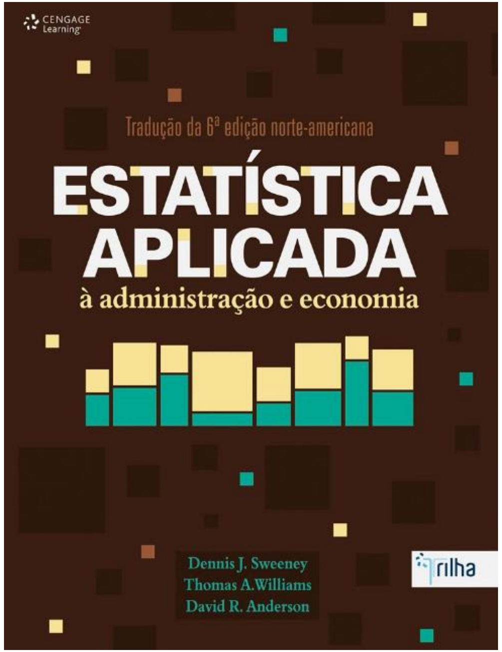
7) Significa que as 3 variáveis juntas explicam 67% a variabilidade de Y.

Bibliografia

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



SWEENEY, Dennis J; WILLIAMS, Thomas A. e ANDERSON, David R. ***Estatística Aplicada à Administração e Economia.***
6ª Edição. São Paulo: Editora Cengage Learning, 2013.



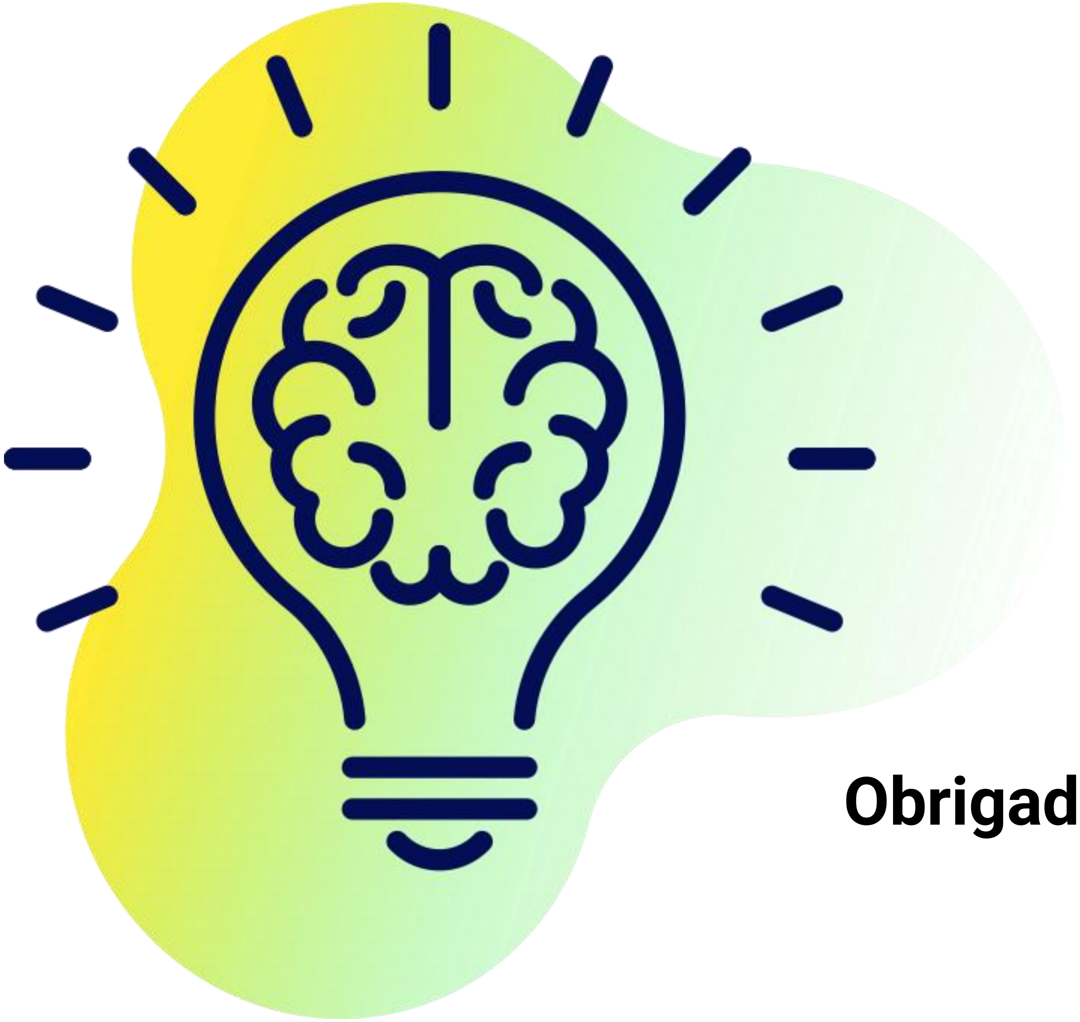
Contatos

Prof. Eng. Rodolfo Magliari de Paiva

Cel.: (11) 9-6866-5501

E-mail: rodolfomagliari@gmail.com

LinkedIn: Rodolfo Magliari de Paiva



Obrigado!