



Escola Livre de Inteligência Artificial

Inteligência Artificial ao alcance de todos

Aula 30/03/2021: Estatística Descritiva – Parte II
Professor: Eng. Rodolfo Magliari de Paiva



Objetivos da Aula

- Efetuar uma boa Análise Exploratória de Dados, entendendo:
Medidas de Tendência Central;
Medidas de Dispersão;
Medidas Separatrizes.
- Aprofundar a interpretação do Boxplot.



Medidas de Tendência Central



São medidas que mostram a convergência dos dados:

- Média;
- Mediana;
- Moda.



- **Média (Esperança):** Representa o ponto de equilíbrio.

$$\bar{x} = \frac{\sum x_i}{n} \quad \mu = \frac{\sum x_i}{n}$$

- **Moda:**

Valor ou classe que mais se repete, podendo ser: Amodal, Unimodal, Bimodal ou Multimodal. Representada por **Mo**

- **Mediana:**

Elemento central
(dados devem estar em rol)

$$M_e = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{se "n" é ímpar} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se "n" é par} \end{cases}$$



Exemplos

1) Uma dermatologista registrou em uma planilha o número de pacientes que ela recebeu durante uma amostra de uma semana e o resultado foi:

Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
28	23	22	27	25	13

Qual a média de pacientes que essa médica atendeu?



Resolução:

$$\bar{X} = \frac{28 + 23 + 22 + 27 + 25 + 13}{6}$$

$$\bar{X} = 23 \text{ pacientes}$$



2) Uma pesquisa em uma cidade selecionou algumas pessoas de forma aleatória para verificar quantos irmãos cada indivíduo possui.

Número de irmãos	Frequência absoluta
0	3
1	6
2	13
3	4
4	3
5	1

Qual a moda?



Resolução:

$M_o = 2$ irmãos (unimodal)



3) Um escritório de Direito registrou o número de ligações que receberam ao longo de alguns dias, e o resultado foi:

25 ; 40 ; 27 ; 38 ; 28 ; 34 ; 30 ; 32

Qual a mediana?



Resolução:

25, 27, 28
3 valores

30, 32
valores centrais

34, 38, 40
3 valores

$$M_d = \frac{30 + 32}{2} = \frac{62}{2} = 31$$



Medidas de Dispersão

São medidas que mostram a variabilidade dos dados:

- Máximo;
- Mínimo;
- Amplitude;
- Variância;
- Desvio Padrão;
- Coeficiente de Variação.



- **Amplitude:**

$$R = X_{(n)} - X_{(1)}$$

- **Variância:**

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}$$

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$$

- **Desvio Padrão (Desvio Médio):**

$$s = \sqrt{s^2}$$

$$\sigma = \sqrt{\sigma^2}$$



- **Coeficiente de Variação (CV):**

$$CV = \frac{\sigma}{\mu} \text{ ou } \frac{s}{\bar{x}}$$

*Normalmente multiplicamos o resultado por **100** para expressá-lo em %.



Exemplos

1) Uma editora de uma revista decidiu coletar uma amostra com as idades de seus leitores que são assinantes, e os dados mostraram os seguintes resultados:

19 20 21 20 19 20 19 20 21

21 21 22 20 21 22 22 23 19

20 21 21 23 20 21 19

Qual o máximo, mínimo e a amplitude das idades? Quais conclusões é possível chegar?



Resolução:

Máximo: 23 anos

Mínimo: 19 anos

Amplitude: $23 - 19 = 4$ anos

Conclusões: Resposta Individual



2) Uma fábrica decidiu verificar o tempo de atraso de um funcionário, pegando aleatoriamente alguns dias e registraram os seguintes tempos de atraso: 6min; 5min; 5min; 3min; 3min; 2min.

Qual a média, a variância e o desvio padrão de atraso do funcionário?



Resolução:

$$\bar{x} = \frac{6 + 5 + 5 + 3 + 3 + 2}{6} \Rightarrow \bar{x} = 4 \text{ min}$$

$$s^2 = (6-4)^2 + (5-4)^2 + (5-4)^2 + (3-4)^2 + (3-4)^2 + (2-4)^2 / 5 \\ = 2,4 \text{ min}^2$$

$$s = \sqrt{2,4} \approx 1,55 \text{ min}$$

Logo, este funcionário atrasa em média $4 \text{ min} \pm 1,55 \text{ min}$.



3) Um carro de Fórmula 1 atingiu as seguintes velocidades médias durante uma corrida:

190km/h ; 198km/h ; 196km/h ;
204km/h ; 202km/h

- a) Qual seu Coeficiente de Variação da velocidade média?
- b) Essa variação foi alta?



Resolução:

a) $\mu = 198\text{km/h}$

$\sigma \sqrt{24} \approx 4,9\text{km/h}$

$$\begin{aligned} CV &= 4,9/198 \\ &= 0,0247 \end{aligned}$$

$$0,0247 \cdot 100 = 2,47\%$$

b) O coeficiente de variação não foi alto.



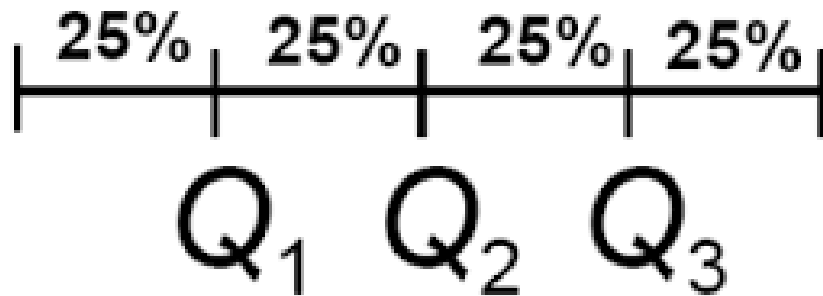
Medidas Separatrizes

Valores que dividem os dados em partes iguais:

- Quartil;
- Decil;
- Centil.

Para o cálculo de todas as medidas separatrizes, os dados sempre devem ser colocados em **rol**!

Normalmente os quartis são os mais usados, vejamos:



Onde:

Q_1 = 25% dos dados abaixo e 75% dos dados acima

Q_2 = 50% dos dados abaixo e 50% dos dados acima

Q_3 = 75% dos dados abaixo e 25% dos dados acima



Exemplos

1) Um sistema de cibersegurança registrou durante uma semana o número de vezes que um site sofreu tentativa de ser hackeado e o resultado obtido foi:

15x ; 5x ; 6x ; 13x ; 6x ; 10x; 9x

Encontre os quartis desse conjunto de dados.



Resolução:

Etapas

1) Dados em rol: 5, 6, 6, 9, 10, 13, 15

2) $Q2 = \text{Md} = 5, 6, 6, \textcolor{red}{9}, 10, 13, 15$

3) $Q1 = 5, \textcolor{red}{6}, 6$

4) $Q3 = 10, \textcolor{red}{13}, 15$

Portanto: $Q1 = 6$; $Q2 = 9$; $Q3 = 13$



2) Os pesos de 9 caixas foram coletados para uma pesquisa de controle da qualidade, e os resultados foram:

73kg ; 74kg ; 77kg ; 52kg ; 85kg ;
59kg ; 73kg ; 84kg ; 92kg

Encontre os valores dos quartis dessa amostra.



Resolução:

Etapas

1) Dados em rol: 52, 59, 73, 73, 74, 77, 84, 85, 92

2) $Q2 = \text{Md} = 52, 59, 73, 73, \textcolor{red}{74}, 77, 84, 85, 92$

3) $Q1 = 52, \textcolor{red}{59}, \textcolor{red}{73}, 73 \Rightarrow (59 + 73)/2 = 66$

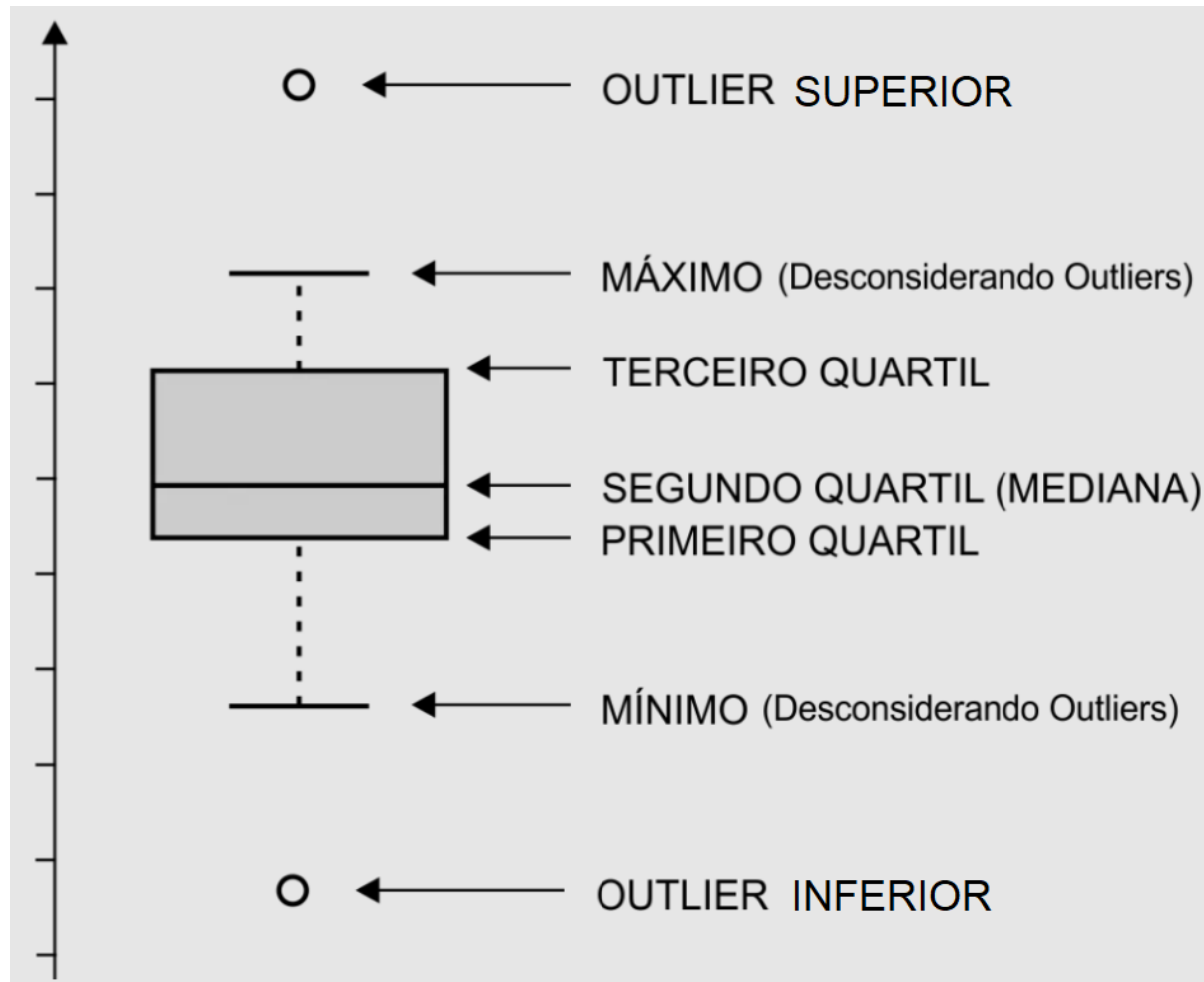
4) $Q3 = 77, \textcolor{red}{84}, \textcolor{red}{85}, 92 \Rightarrow (84 + 85)/2 = 84,5$

Portanto: $Q1 = 66$; $Q2 = 74$; $Q3 = 84,5$



Retomando o Boxplot

Agora que já vimos sobre as medidas de Tendência Central, Medidas de Dispersão e Medidas Separatrizes, é possível olhar o **Boxplot** de uma forma melhor e compreendê-lo!





Outliers: Dados extremamente discrepantes com relação ao conjunto, ultrapassando o limite inferior e/ou superior:

$$\text{Limite Inferior} = Q1 - 1,5.(Q3 - Q1)$$

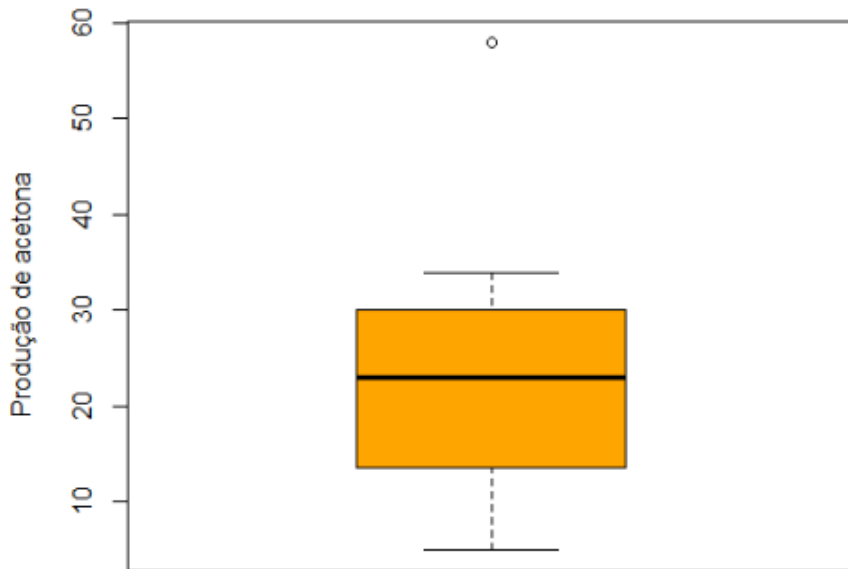
$$\text{Limite Superior} = Q3 + 1,5.(Q3 - Q1)$$

Onde: $Q3 - Q1$ = Amplitude Interquartílica
ou
Intervalo Interquartílico



Exemplo

O boxplot a seguir representa a produção de acetona/dia em uma empresa.



Explique-o.



Conclusão

Com essas ferramentas de análise exploratória de dados da **Estatística Descritiva** é possível entender o comportamento de diversos fenômenos do dia a dia, basta **ter** ou **iniciar** a coleta de dados e na sequência:



APLICAR!





Exercícios

1) Complete a frase com as palavras que faltam:

- a) Utilizamos a _____ quando queremos extrair a média entre elementos que possuem um peso, ou seja uma frequência de aparição para cada elemento.
- b) Dizemos que uma amostra é _____ quando dois elementos aparecem a mesma quantidade de vezes a mais que os outros elementos da amostra.
- c) Entende-se por _____ como sendo o elemento central de uma amostra.



2) Em uma sala de aula do 2º ano do ensino médio existem 20 alunos, dos quais foram tirados aleatoriamente 9 para serem analisadas as suas médias com relação à disciplina de Matemática, e os dados obtidos foram:

10,0	6,0	7,0
5,0	9,0	6,0
6,0	7,0	8,0

Com base nessas informações responda:

- a) Qual a população?
- b) Qual a amostra?
- c) Qual a variável? Classifique-a.
- d) Qual a moda?
- e) Qual a mediana?
- f) Qual a média?
- g) Qual a variância?
- h) Qual o desvio padrão?
- i) Qual o coeficiente de variação?



3) Um mercado registrou em uma tabela o “peso” em kg de determinado produto que chegou ao longo de alguns dias:

2,20	2,28	2,23	2,20	2,35	2,28	2,25	2,30	2,37
------	------	------	------	------	------	------	------	------

Qual a mediana?

4) O conjunto de dados a seguir representam o número de vacinas que um posto de saúde aplicou nas pessoas ao longo de 18 dias.

13 12 11 12 15 14

13 12 11 16 15 14

13 12 11 13 15 11

Qual a moda?



5) Dois jogadores disputam uma partida de ping pong e suas pontuações até o momento são dadas pela tabela:

Vítor	12	15	12	16	14	15
Rafael	12	11	18	9	19	15

- a) Qual a média de pontuação dos jogadores?
- b) Qual apresenta desempenho mais regular?
- c) Em qual você apostaria?

6) Uma pessoa decide registrar quantas vezes ela lava as mãos durante uma pandemia ao longo de alguns dias aleatórios, e os resultados obtidos foram:

$7x$; $11x$; $8x$; $14x$; $10x$

- a) Qual a média de vezes?
- b) Qual o desvio padrão de vezes?
- c) Qual o coeficiente de variação?

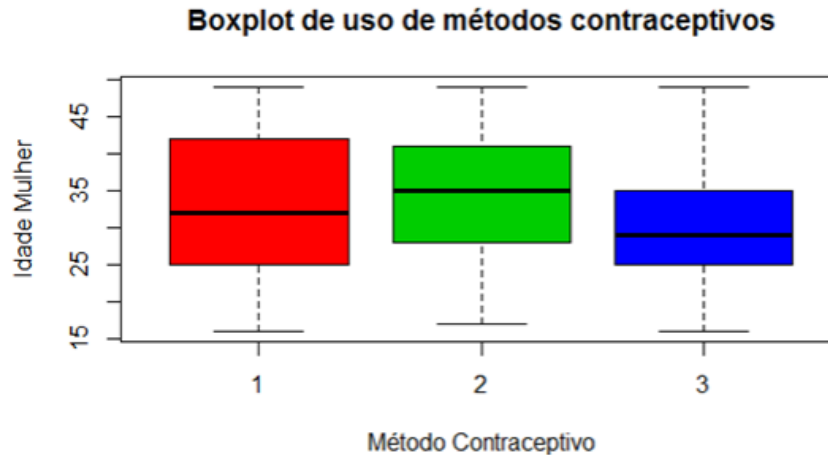


7) Uma câmera está em fase de teste e precisa detectar foragidos da polícia e quando essas pessoas são localizadas, a câmera envia um SMS ao policial mais próximo com a posição dos foragidos. Durante uma simulação, o tempo de resposta da câmera foi:

1,9s ; 7,7s ; 2,0s ; 6,1s ; 2,1s ; 3,7s ; 2,5s ; 3,3s ; 3,0s ; 3,1s

Calcule os quartis da amostra.

8) Três grupos de mulheres, de regiões diferentes, passaram por uma entrevista com um assistente social, para que ele entendesse o perfil das idades que utilizam métodos contraceptivos.

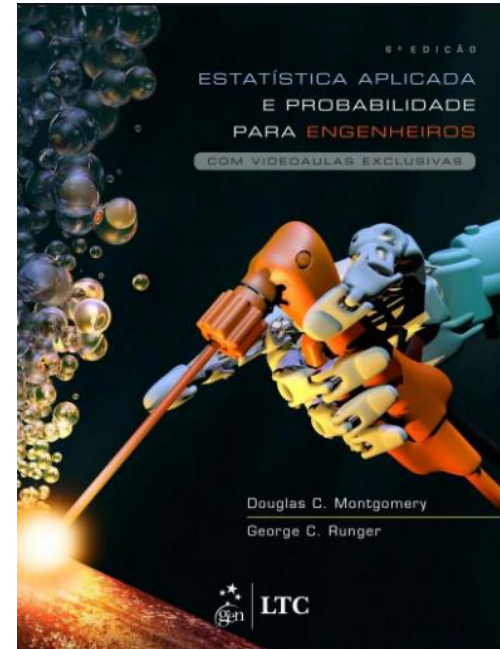


Quais conclusões ele chegará?



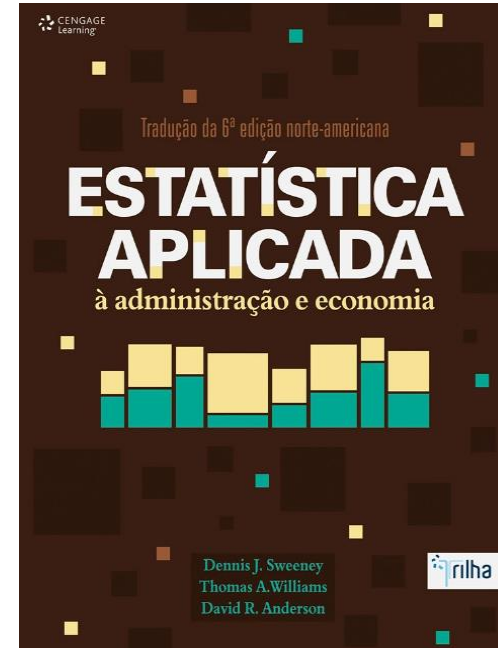
Bibliografia

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



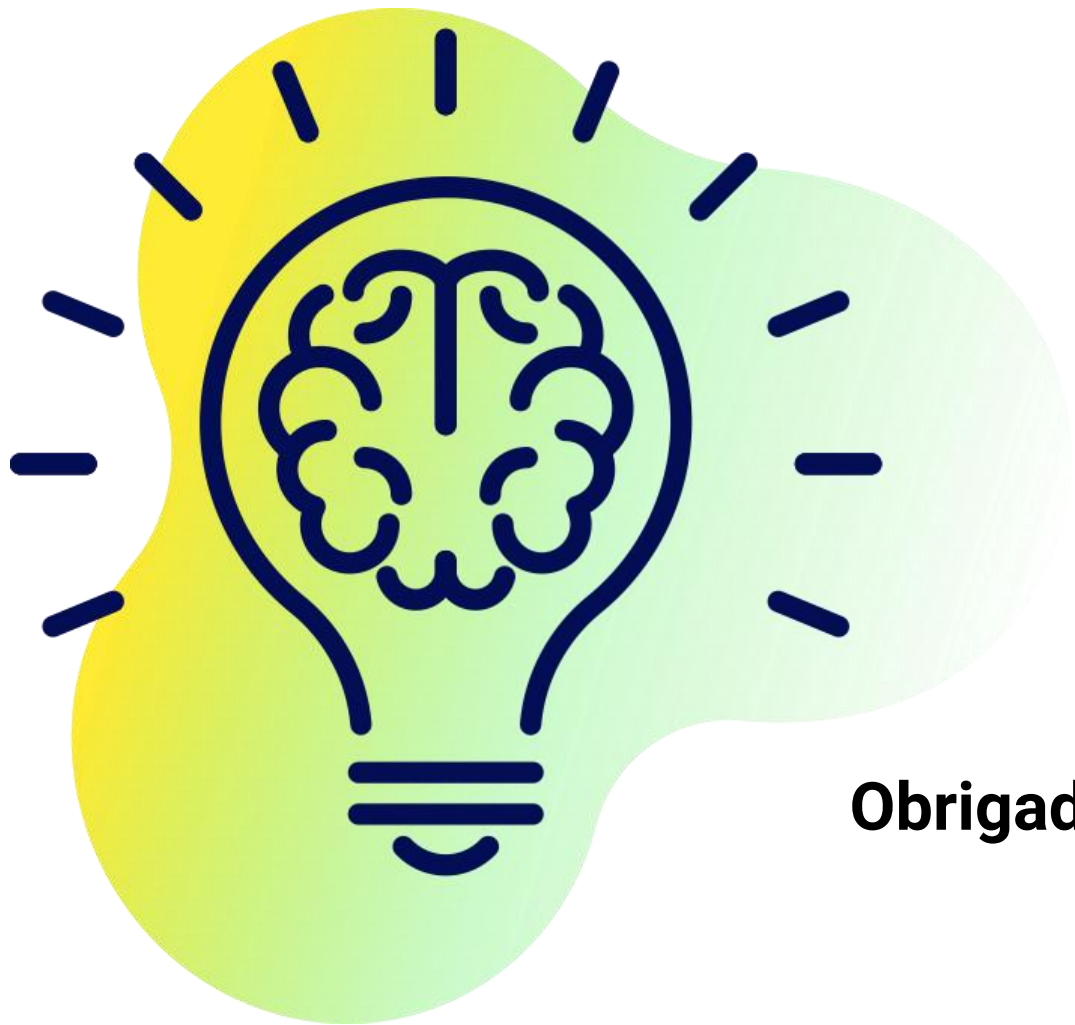
Cel.: (11) 9-6866-5501



E-mail: rodolfomagliari@gmail.com



LinkedIn: Rodolfo Magliari de Paiva



Obrigado!