

A primeira Escola presencial gratuita de Inteligência Artificial do Brasil



Aula 03/03/2020: Estatística Básica

Professor: Eng. Rodolfo Magliari de Paiva

Apoio:



Objetivo da Aula

- Compreender a diferença entre Matemática e Estatística;
- Conhecer a História da Estatística;
- Áreas de Aplicação da Estatística;
- Efetuar uma boa Análise Exploratória de Dados, entendendo:
Tipos de Dados, Gráficos Estatísticos, Coleta de Dados, Medidas de Tendência Central, Medidas de Dispersão, Medidas Separatrizes.

Um pouco da História da Matemática e da Estatística

- A **Matemática** é uma ciência bem mais antiga do que a **Estatística** formal que conhecemos hoje.
- É fato que só houveram (na origem), descobertas e avanços matemáticos em regiões que tinham a necessidade da **agricultura**.
- Aos poucos, os profissionais das ciências exatas foram surgindo (corda, balança, astros...), e com eles novas necessidades de descobertas junto com o desenvolvimento das cidades.
- A Matemática Primitiva teve quatro alicerces: Clima, Agricultura, Comércio e Registro.

A Ciência Estatal

Esteve presente entre os hebreus, chineses, egípcios, maias, romanos, hindus, persas e babilônios, povos que se organizaram ao redor de um Estado, que necessitavam de informações censitárias, informações essas que eram colhidas pelo **estadista**.

A Estatística

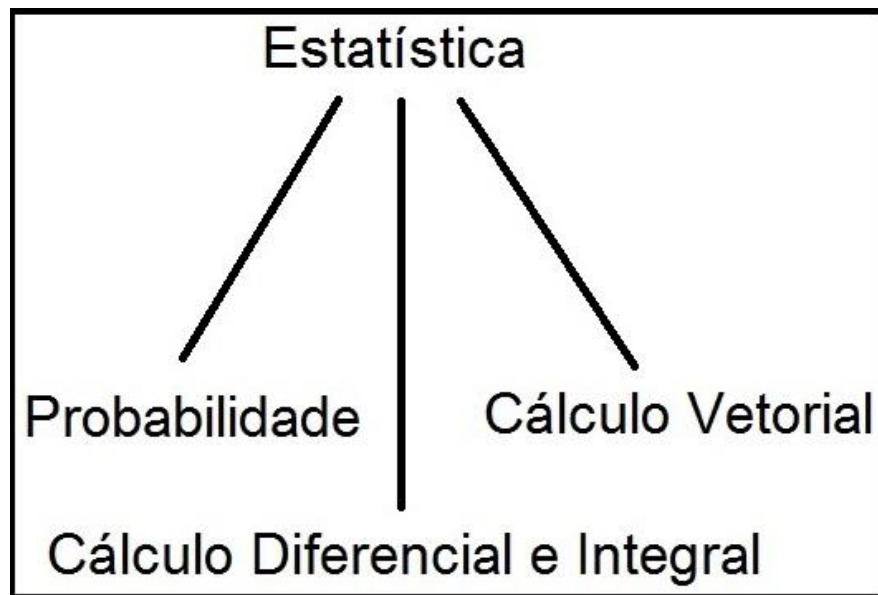
1858 – Florence Nightingale, com o trabalho:

“Notas sobre os Assuntos que Afetam a Eficiência da Saúde e Administração Hospitalar do Exército Britânico”

Nascida em 12 de Maio de
1820, Florence Nightingale
era
enfermeira britânica.

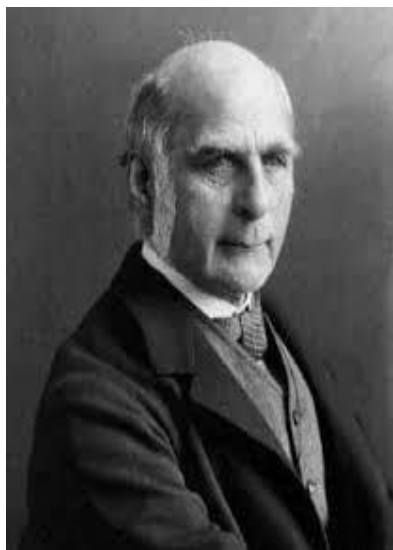


Com o passar do tempo a Estatística foi tomando forma, com os avanços de outros conteúdos matemáticos:



Tendo a atenção de estudiosos como: Christian Huygens, Pierre Fermat, Blaise Pascal, John Graunt, Jacques Bernoulli, Thomas Bayes, Poisson, Mary Somerville, entre outros...

No final do século XIX, com Francis Galton, Francis Ysidro Edgeworth, Karl Pearson e George Udny Yule, a Estatística ganhou sua aplicabilidade e operacionalidade, além do visual moderno.



Importância:

- Tomada de decisão;
- Quantificar incerteza;
- Levantamento de dados;
- Entendimento de um fenômeno;
- Etc...





Embora a Estatística Clássica se fez por conta de conteúdos da Matemática, hoje em dia sabemos que são **ciências distintas**, com algumas coisas em interseção.



Áreas de Aplicação

- Bioestatística;
- Ciência Atuarial;
- Demografia;
- Econometria;
- Epidemiologia;
- Geoestatística;
- Controle da Qualidade;
- Data Science;
- B.I.;
- Etc...



Análise Exploratória de Dados

As ferramentas da Estatística Descritiva
permitem uma boa Análise Exploratória de
Dados de forma rápida e sem muita
complicação!

Tipos de Dados

Os dados (variáveis), são classificados em dois grupos:

Quantitativo X Qualitativo



Quantitativo: Discreto = Números inteiros, contagem, é mensurável
Contínuo = Conjunto dos Reais, também mensurável

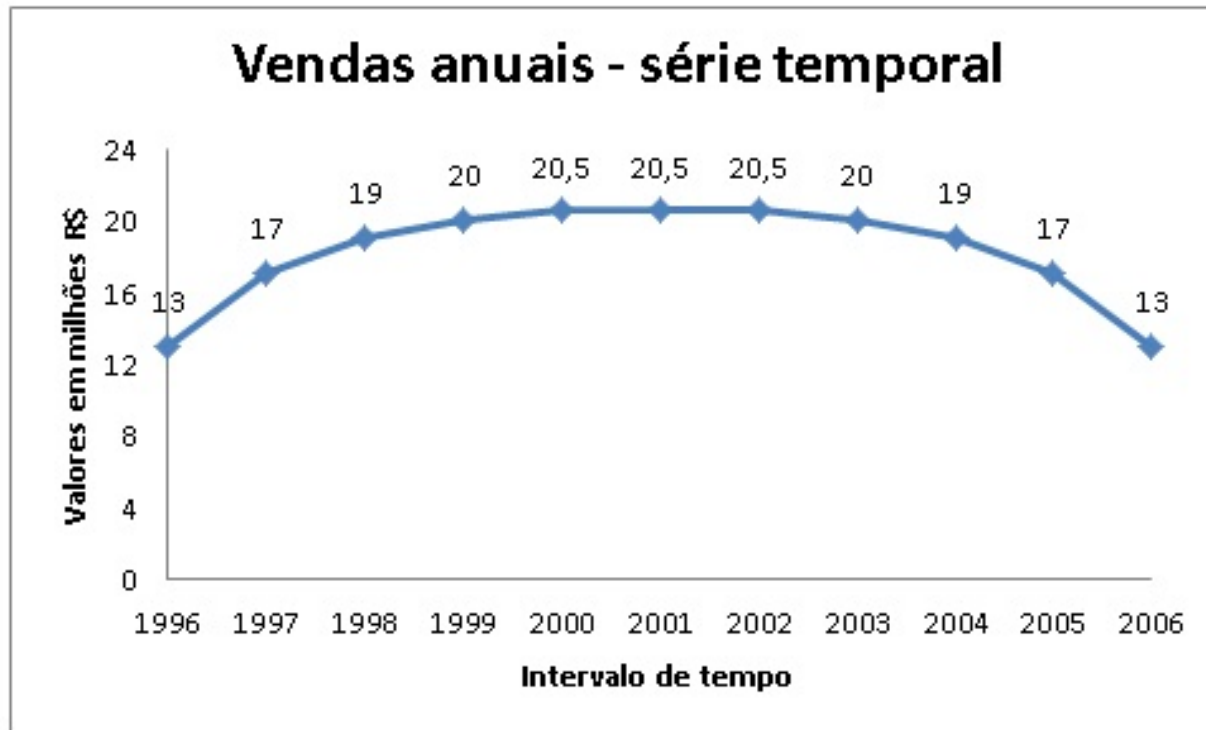


Qualitativo: Ordinal = Apresenta hierarquia
Nominal = Não apresenta hierarquia

Gráficos Estatísticos para Variáveis Qualitativas e Quantitativas

Gráficos auxiliam de uma forma rápida e prática que entendamos o **comportamento de um fenômeno**.

Exemplo:



Gráficos para Variáveis Qualitativas:

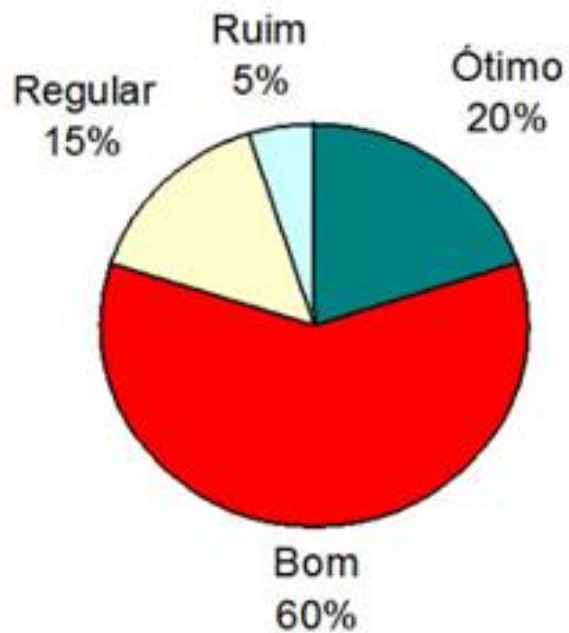
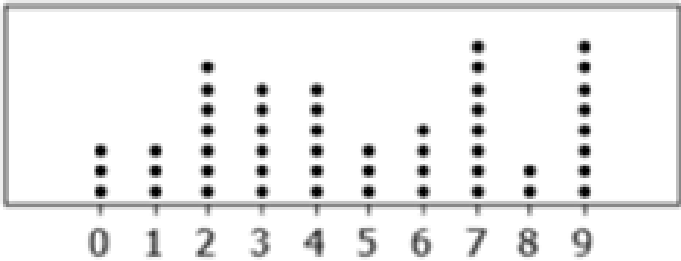


Gráfico de Setores
ou Circular



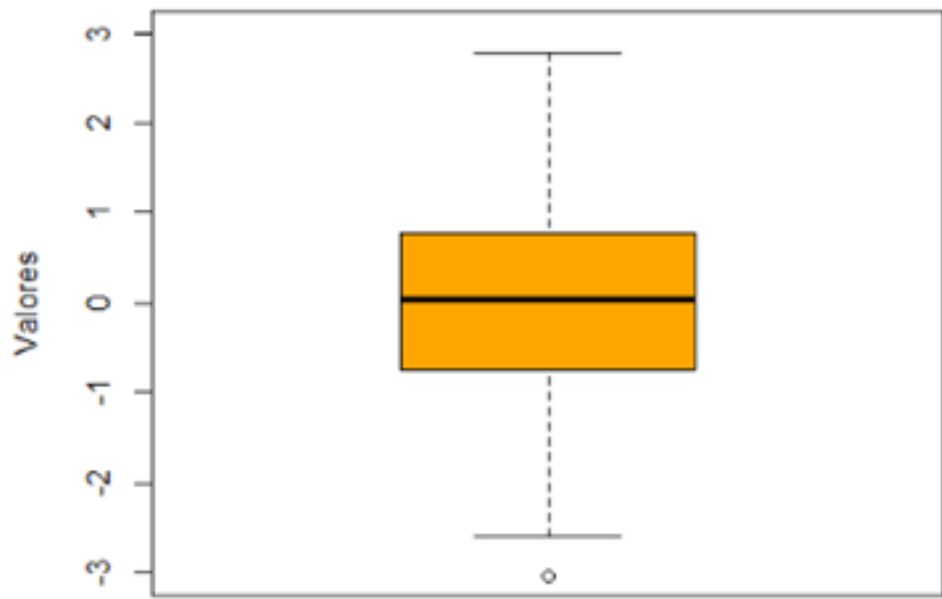
Gráficos para Variáveis Quantitativas:



Dotplot ou Strip Chart



Histograma



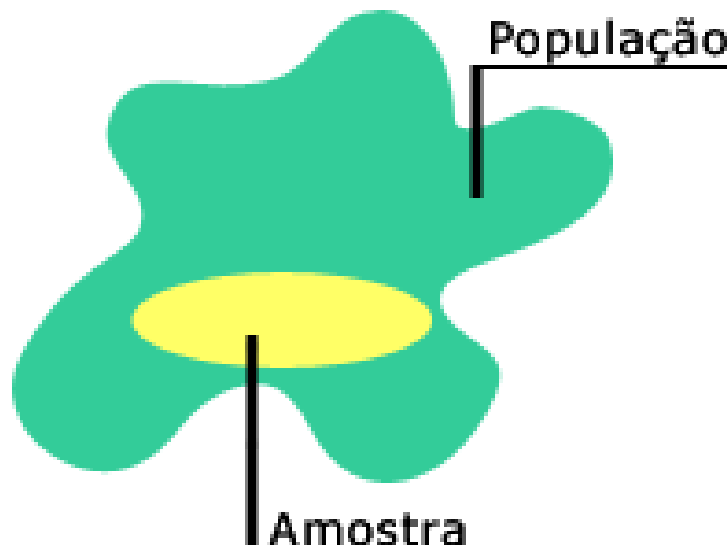
Variável

Boxplot

Coleta e Organização de Dados

A coleta de dados é uma parte extremamente importante, pois sem dados não existe a possibilidade de ser feito qualquer tipo de estudo estatístico.

Para obter dados, a Estatística utiliza amostragens, que podem ser **probabilísticas** e **não probabilísticas**.



Tipos de Amostragem

- **Probabilística:**
 - Aleatória Simples;
 - Sistemática;
 - Estratificada;
 - Conglomerados.
- **Não Probabilística:**
 - Intencional;
 - Cotas;
 - Conveniência.

Uma vez os dados coletados (dados brutos), a forma mais prática de organizá-los é por meio de **tabelas**.

Aparelho	Potência (KW)	Tempo de uso diário (horas)
Ar condicionado	1,5	8
Chuveiro elétrico	3,3	1/3
Freezer	0,2	10
Geladeira	0,35	10
Lâmpadas	0,10	6

Nº de pessoas que doaram	Valor Doador
12	R\$ 5,00
10	R\$ 7,00
8	R\$ 10,00
7	R\$ 12,00
3	R\$ 15,00

Com os dados observados, é possível extrair vários tipos de informações!

Frequência Relativa e Frequência Absoluta:

Estatura dos alunos (cm)	Frequência simples f_i	Frequência relativa fr_i (%)	Frequência acumulada F_i	Frequência acumulada relativa Fr_i (%)
150 --154	4	10,0	4	10,0
154 --158	9	22,5	13	32,5
158 --162	11	27,5	24	60,0
162 --166	8	20,0	32	80,0
166 --170	5	12,5	37	92,5
170 --174	3	7,5	40	100,0
Total Σ	40	100,0	---	---

Medidas de Tendência Central

Média (Esperança):

$$\bar{X} = \mu = \frac{\sum x_i}{n}$$

Moda:

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

Mediana:

Elemento central
(dados devem estar
em rol)

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

Medidas de Dispersão

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 %.

Medidas Separatrizes

Valores que dividem os dados em partes iguais:

- **Quartis:** Dividem os dados em 4 partes iguais;
- **Decis:** Dividem os dados em 10 partes iguais;
- **Centis:** Dividem os dados em 100 partes iguais.

Normalmente os quartis são os mais usados.

Quartis:

n = número total de elementos

$j(n+1)/4$, para $j=1,2$ e 3

k = o maior inteiro $\leq j(n+1)/4$

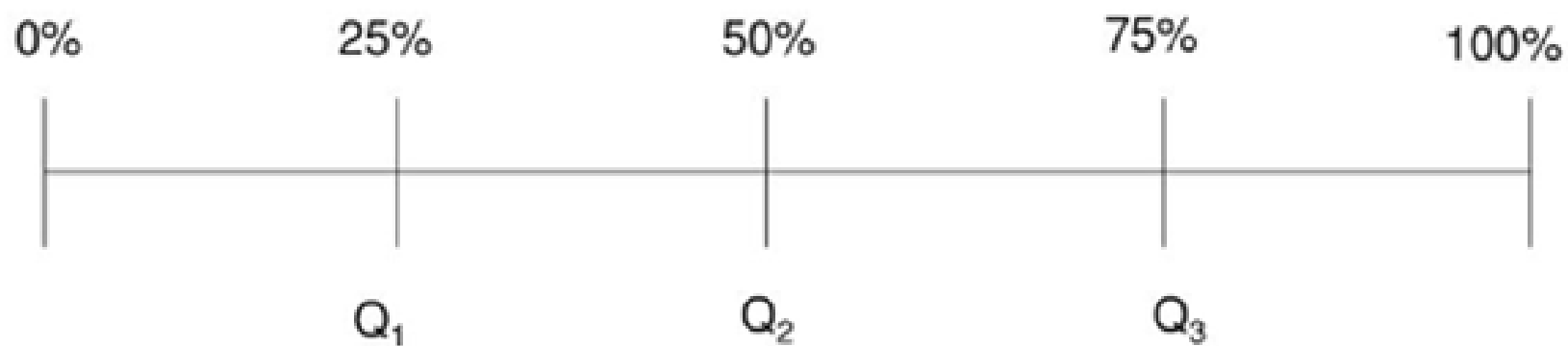
$$Q_j = X_k + \left(\frac{j(n+1)}{4} - k \right) (X_{k+1} - X_k)$$

1° Quartil: $Q_1 = X_1 + \left(\frac{n+1}{4} - k \right) (X_{k+1} - X_k)$

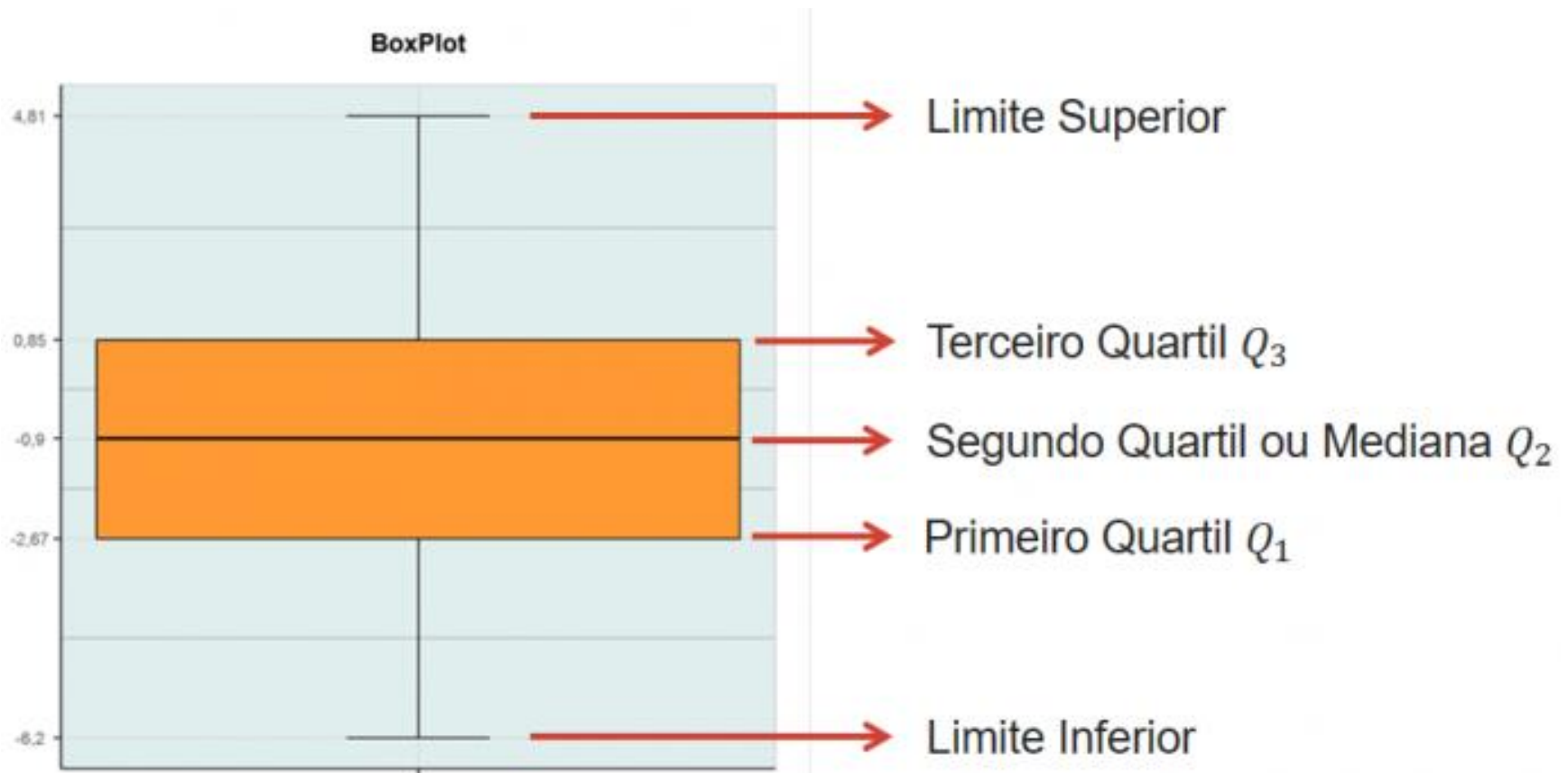
2° Quartil: $Q_2 = X_3 + \left(\frac{2(n+1)}{4} - k \right) (X_{k+1} - X_k)$

3° Quartil: $Q_3 = X_5 + \left(\frac{3(n+1)}{4} - k \right) (X_{k+1} - X_k)$

Representação gráfica:



Retomando um Tópico



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-) Defina com suas palavras, de acordo com o que foi falado em sala de aula, o que é Estatística e para que serve.
- 2-) Qual a diferença entre dados brutos e dados em rol?
- 3-) Como se dividem as Variáveis Quantitativas e Qualitativas? Dê um exemplo para cada.
- 4-) O que é população e amostra dentro da Estatística?

5-) Marque a alternativa correta.

- a-) A apresentação de dados estatísticos pode ser feita através de gráficos, como um histograma, setograma, entre outros.
- b-) Apenas observando as tabelas de frequências é possível dizer qual o Desvio Padrão.
- c-) São exemplos de Medidas de Tendência Central a moda, a mediana e a variância.
- d-) São exemplos de Medidas de Dispersão o desvio padrão, desvio médio e a média.
- e-) Se uma amostra possui um elemento que se repete várias vezes, dizemos que é uma amostra amodal.

6-) 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.

7-) 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?

Bibliografia

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

Contatos

Prof. Eng. Rodolfo Magliari de Paiva



Cel.: (11) 9-6866-5501



E-mail: rodolfomagliari@gmail.com



LinkedIn: Rodolfo Magliari de Paiva