

B&S 002 - BIOSTATISTICS & STATISTICS - A FASCINATING SCIENCE.

# BIOESTATÍSTICA

CONCEITOS E FUNDAMENTOS

AUTHOR: JOB TOLENTINO JUNIOR (PhD)

E-MAIL: [jobtjr2000@yahoo.com](mailto:jobtjr2000@yahoo.com)

PHONE: +5521984803221 (COM WHATZAAP)

B&S 002 - BIOSTATISTICS & STATISTICS - A FASCINATING SCIENCE.

# BIOESTATÍSTICA

CONCEITOS E FUNDAMENTOS

AUTHOR: JOB TOLENTINO JUNIOR (PhD)

- 1 – ORCID: <http://orcid.org/0000-0002-8054-3237>
- 2 – Centro Universitário Redentor (UNIRENTOR – Itaperuna/RJ)
- 3 – Centro de Tecnologia Mineral (CETEM-RJ) / Laboratório de Argilas Aplicadas (LAA)
- 4 – Universidade Federal Fluminense (UFF) / Núcleo de Pesquisa e Extensão em Educação e Saúde Comunitária (NUPEESC)
- 5 – Universidade Federal Fluminense (UFF) / Grupo Saúde Integral da Mulher e do Recém Nascido
- 6 – Universidade Federal Fluminense (UFF) / Escola de Enfermagem Aurora de Afonso Costa (EEAAC/UFF) / Curso de Controle de Infecção em Assistência À Saúde (CIAS)

# CAPÍTULO 2: REVISITAR O BÁSICO É IMPORTANTE?

---

# ANTES QUE ENTREMOS NO ASSUNTO LEMBRE-SE:

---

**O ano de 2020 pode ser considerado um divisor de águas na história humana, pois nunca antes a modelagem matemática foi tão amplamente divulgada com o objetivo de informar e esclarecer a população leiga sobre o comportamento da difusão de uma doença específica, denominada PANDEMIA COVID-19.**



# MAS PARA QUE POSSAMOS ENTENDER DE FATO ESTES MODELOS, ALGUMAS DEFINIÇÕES FUNDAMENTAIS PRECISAM SER ESCLARECIDAS

**Mas para isso precisamos classifica-las  
em quatro grupos:**

- ❖ PRIMEIRO GRUPO DE DEFINIÇÕES,
- ❖ SEGUNDO GRUPO DE DEFINIÇÕES,
- ❖ TERCEIRO GRUPO DE DEFINIÇÕES,
- ❖ QUARTO GRUPO DE DEFINIÇÕES.





# MAS PARA QUE POSSAMOS ENTENDER DE FATO ESTES MODELOS, ALGUMAS DEFINIÇÕES FUNDAMENTAIS PRECISAM SER ESCLARECIDAS

## ❖ PRIMEIRO GRUPO DE DEFINIÇÕES:

- UNIVERSO AMOSTRAL
- POPULAÇÃO
- CENSO
- AMOSTRA
- AMOSTRA REPRESENTATIVA
- UNIDADE AMOSTRAL



# PRIMEIRO GRUPO DE DEFINIÇÕES:

---

- **UNIVERSO AMOSTRAL (OU UNIVERSO)**

- são todos os elementos que desejamos estudar ou avaliar

- **POPULAÇÃO**

- São os elementos do universo que possuem uma característica comum (fundamental) para o tema em estudo

# PRIMEIRO GRUPO DE DEFINIÇÕES:

- **CENSO (OU RECENSEAMENTO)**

- **Estudo de um ou mais variáveis (descritores) de todo o universo amostral**

- **AMOSTRA**

- **É a parte da população do qual serão obtidos os dados a serem pesquisados**

# PRIMEIRO GRUPO DE DEFINIÇÕES:

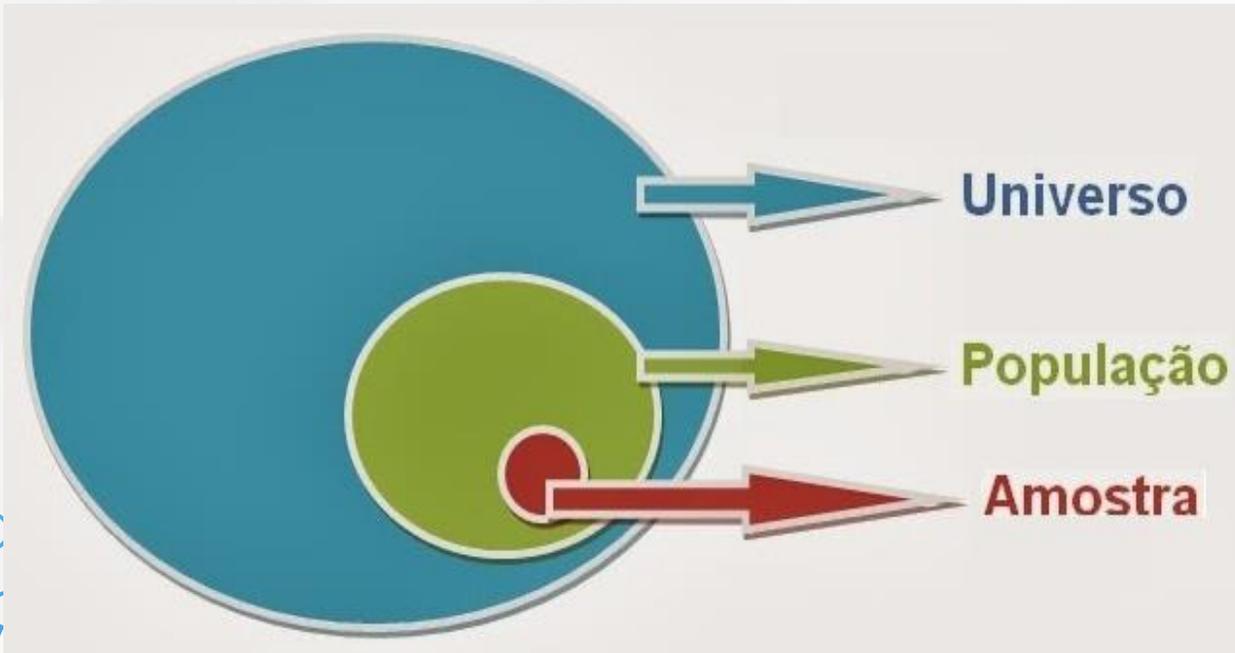
- **AMOSTRA REPRESENTATIVA**

- É a amostra que será escolhida aleatoriamente, garantindo a aleatoriedade na seleção de cada membro da amostra a ser inclusa no grupo amostral.

- **UNIDADE AMOSTRAL**

- é o objeto de estudo

# PRIMEIRO GRUPO DE DEFINIÇÕES:



- Universo =  $N = 10$** 
  - *total de indivíduos na população*
- Amostra =  $n \leq 10$** 
  - *uma parte representativa da população*
- Se  $N = n$ , então é Censo.**
- Se  $n < N$ , então é Amostra.**

# SEGUNDO GRUPO DE DEFINIÇÕES:

## • TIPOS DE AMOSTRAS

Tipo	Característica
Casual simples ou equiprobabilística	Cada elemento da população tem a mesma chance de entrar na amostra. Por exemplo, se desejamos tomar uma amostra casual de 5 cartas de um baralho de 52 cartas, uma maneira seria embaralhar cuidadosamente e então tirar 5 cartas sem olhar a face das mesmas. Aqui o baralho é a população, e as 5 cartas as amostras.
Sistemática	Na amostragem sistemática, seleciona-se qualquer unidade amostral e, a partir desta, escolhem-se as seguintes segundo o intervalo de seleção. Por exemplo: Imagine uma cidade em que cada quarteirão tem 10 casas. Um pesquisador quer analisar 1 casa de cada quarteirão. Ele pode sortear um número de 1 a 10. Se o número sorteado for 7, ele vai pesquisar somente a sétima casa de cada quarteirão.
Estratificada	Uma amostra pode ser estratificada de acordo com algum fator como sexo, idade ou nível econômico. Para uma amostra estratificada de acordo com o sexo, por exemplo, divide-se a população em dois estratos: homens e mulheres, e em seguida escolhe-se uma amostra casual dentro de cada estrato.

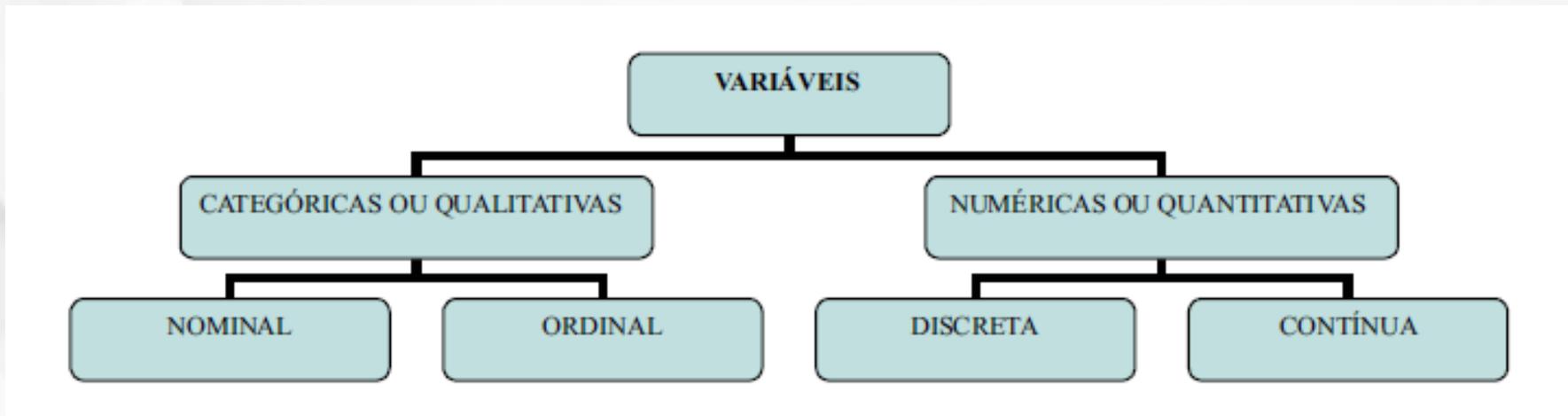
# SEGUNDO GRUPO DE DEFINIÇÕES:

## • TIPOS DE VARIÁVEIS

Tipo	Característica
Contínuas	São aquelas que podem assumir qualquer valor dentro de um intervalo de interesse. Os dados advindos deste tipo de variável são ditos contínuos. Ex.: peso, estatura, distância percorrida em um teste de esforço etc. Em geral estão associadas a medidas que tenham unidade (m, kg, l, m/s etc.)
Discretas	São aquelas que só podem assumir valores inteiros dentro de um intervalo de interesse. Os dados discretos são resultados da contagem do número de itens referente à variável. Ex.: número de repetições executadas em um exercício, número de gols marcados em uma partida de futebol, quantidade de saltos dados por um jogador em uma partida de voleibol etc.
Nominais	São aquelas que só podem assumir alguns estados ou categorias e geralmente não são numéricas: Os dados nominais surgem quando se definem categorias e se conta suas observações. Ex.: Sexo de uma população (masculino e feminino), queixas de dor lombar (sim e não), cor dos olhos de uma população (azuis, castanhos, pretos, verdes) etc.
Ordinais	São aquelas que se relacionam a avaliações subjetivas segundo preferência ou desempenho. Os dados ordinais constituem valores relativos, atribuídos para denotar ordem. Ex.: Avaliação do estado de saúde de um paciente excelente, bom ou reservado.

# SEGUNDO GRUPO DE DEFINIÇÕES:

- TIPOS DE VARIÁVEIS



# TERCEIRO GRUPO DE DEFINIÇÕES:

---

- **MEDIDAS DE TENDÊNCIA**

- **É um valor numérico que representa a tendência de um conjunto de variáveis numéricas presentes em uma amostra representativa.**
- **As três medidas mais usadas são: média, mediana e moda**

# TERCEIRO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE TENDÊNCIA**



## Média Aritmética

Medida de tendência central mais utilizada;

É definida como soma dos valores teóricos de todas as observações (observação é um elemento de uma amostra) dividida pelo número de observações;

O símbolo  $\mu$  (mu) será usado para denotar média de uma população;

O símbolo  $\bar{x}$  será usado para denotar a média de uma amostra;

$$\bar{x} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

Exemplo: Uma amostra constituída de 3 pesos de recém-nascidos: 2,75kg, 3,25kg e 3,80kg. Aqui, n, o tamanho da amostra, é igual a 3.  $x_1$ , primeira observação, é 2,75kg;  $x_2$ , segunda observação é 3,25kg;  $x_3$  é 3,80kg.

$$\bar{x} = \frac{(2,75 + 3,25 + 3,80)}{3} = 9,80/3 = 3,27, \text{ ou seja, peso médio é } 3,27 \text{ kg.}$$

# TERCEIRO GRUPO DE DEFINIÇÕES:

## • MEDIDAS DE TENDÊNCIA

### Mediana

Outra medida usada para indicar o centro de uma distribuição;

A mediana de uma série de observações é o número que fica exatamente no meio da série quando os dados estão ordenados e o número de observações é ímpar; ou a média aritmética de dois números do meio, quando o número de observações é par;

Isto significa que para um conjunto de dados, se os mesmos forem ordenados, a mediana ocupará o centro deste conjunto.

Ex.: Dada a variável  $x = \{1, 3, 0, 2, 4\}$ , a mediana é 2.

Para se calcular a mediana a mediana de um conjunto de dados deve-se:

- 1) ordenar o conjunto; no exemplo acima:  $x = \{0, 1, 2, 3, 4\}$ ;
- 2) verificar se há um número par ou ímpar de valores no conjunto; no exemplo acima: 5 observações - ímpar;
- 3) se for ímpar a mediana será o valor que ocupa a posição central e se for par será a média entre as duas posições centrais.

# TERCEIRO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE TENDÊNCIA**

## **Moda**

A moda é o valor que aparece com maior frequência em uma distribuição.

Exemplos:

Seja  $x = \{0\ 1\ 0\ 2\ 3\ 4\ 4\ 0\ 3\ 2\ 5\ 6\}$ , a moda é 0.

Seja  $x = \{3\ 1,5\ 2\ 3\ 3\ 4\ 5\ 1,5\ 2\ 1,5\ 0\ 4\ 1,5\ 1,5\ 6\}$  a moda é 1,5.

# TERCEIRO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE TENDÊNCIA**
- **OBS:**

Portanto, é importante saber que a média é a melhor e mais importante medida de tendência central, devido a sua maior estabilidade amostral, utilidade e facilidade de cálculos. Porém, em distribuições assimétricas, a média não descreve adequadamente o centro. Nesses casos, prefere-se a mediana. Por exemplo, numa distribuição: ( 2, 4, 6, 9, 10, 11, 2000) na qual um dos valores(2000) destoa bastante dos outros. Nesse caso, temos:

\* média:  $2+4+6+9+10+11+2000 / 7 = 291,7$

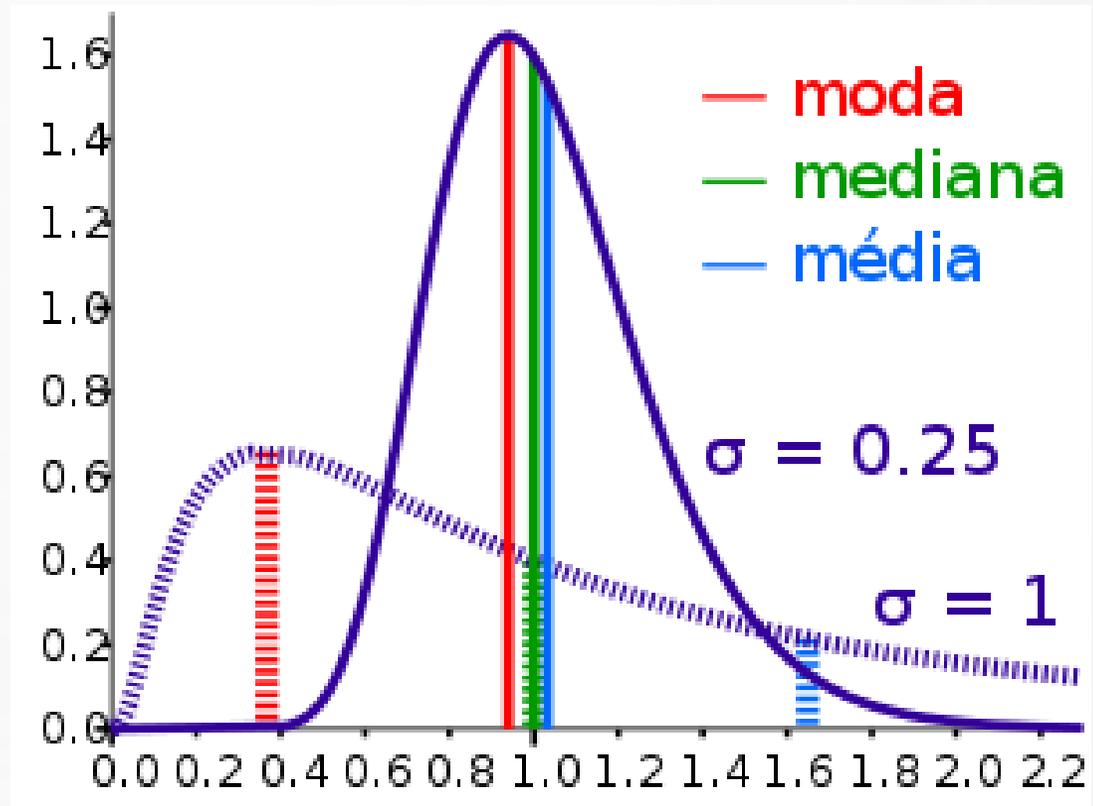
\* mediana: temos sete elementos nesse conjunto. Portanto, a mediana é o elemento central desse conjunto ordenado – 4º elemento ( $n+1/2$ , ou seja,  $7 + 1 / 2 = 4$ ). O quarto elemento desse conjunto corresponde à observação 9.

Mediana = 9.

Nesse caso, verifica-se facilmente que a mediana representa melhor o centro da distribuição.

# TERCEIRO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE TENDÊNCIA**
- **OBS:**



# QUARTO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE VARIABILIDADE**

- São medidas que nos dão idéia da concentração de dados em torno de um valor (dispersão).
- As características do espalhamento (dispersão) das medidas são: amplitude, desvio padrão e variância

# QUARTO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE VARIABILIDADE**



## Amplitude

Amplitude é definida como a diferença do menor ao maior valor de um conjunto de dados. Na série 7-3-4-6-1-6-7-6-5, a amplitude é  $6-(1)$ ; É considerada uma medida de dispersão muito satisfatória para grupos pequenos de dados.



# QUARTO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE VARIABILIDADE**



## Desvio padrão e variância

Nem me perguntem o que significa desvio padrão. Porém, o importante nesse momento é saber para que serve essa medida de dispersão.

Considere essa série: 7-3-4-6-1-6-7-6-5. A média  $\bar{x}$  é igual a 5. Para o cálculo do desvio padrão e da variância é necessário diminuir cada observação pela média ( $x - \bar{x}$ ). Essa diferença é chamada de desvio médio ( $x - \bar{x}$ ). Depois eleva-se o valor do desvio médio ao quadrado  $(x - \bar{x})^2$ . Feito isso, soma-se todos os desvios médios ao quadrado. Divide-se esse valor pelo número de observações - 1 ( $n - 1$ ), obtendo-se, portanto a variância de uma amostra. Veja exemplo.



# QUARTO GRUPO DE DEFINIÇÕES:

## • MEDIDAS DE VARIABILIDADE



Cálculo da variância e do desvio padrão

x	(x- $\bar{x}$ )	(x- $\bar{x}$ ) <sup>2</sup>
7	+2	4
3	-2	4
4	-1	1
6	+1	1
1	-4	16
6	+1	1
7	+2	4
6	+1	1
5	0	0
<hr/>	<hr/>	<hr/>
45	0	32

$$\text{variância} = s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = 32/8 = 4$$

$$\text{desvio padrão} = s = \sqrt{4} = 2$$

# QUARTO GRUPO DE DEFINIÇÕES:

- **MEDIDAS DE VARIABILIDADE**

A fórmula da variância em uma amostra ( $n$ = número de elementos de uma amostra) é diferente da fórmula de variância em uma população ( $N$ =números de elementos de uma população). A variância de uma amostra é usualmente denotada por  $s^2$  e é definida pela fórmula:  $\text{variância} = s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ ,

A variância de uma população é representada por  $\sigma^2$  e é dada pela fórmula:  $\text{variância} = \sigma^2 = \frac{\sum(x-\bar{\mu})^2}{N}$

O desvio padrão nada mais é do que a raiz quadrada(positiva) da variância. O desvio padrão de uma amostra é denotado por  $s$  e o da população por  $\sigma$  (sigma).

Portanto, desvio padrão mede dispersão. E o faz mais ou menos do mesmo modo que o desvio médio, ou seja, medindo o afastamento médio dos dados em relação à média do conjunto. A diferença é que, tomando o quadrado dos desvios, o desvio padrão obtém uma espécie de média ponderada desses desvios, onde desvios maiores entram com peso maior que os desvios menores.

# REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ALMEIDA P.M., AnEx: Analytical Expertise & Cenários, Amostra e Universo, FESPSP, disponível em <https://slideplayer.com.br/slide/3687249/> , acessado em 01/07/2020.
- [2] FERRAZ E. S., Universidade Federal de Alagoas - UFAL - Disciplina: Estatística Educacional, disponível em <http://estatisticaeducacionalufal.blogspot.com/2013/09/turno-vespertino-relatorio-do-dia-19-de.html>, acessado em 01/07/2020,
- [3] ROCHA, H.A.L., COELHO, E.R., Conceitos básicos em epidemiologia e bioestatística, disponível em: <http://www.epidemioufc.br/files/ConceitosBasicosemEpidemiologiaeBioestatistica.pdf>, acessado em 01/07/2020,
- [4] WIKIPEDIA., Comparação entre média, mediana e moda de duas distribuições log-normal com diferentes dispersões, disponível em [https://pt.wikipedia.org/wiki/Mediana\\_\(estat%C3%ADstica\)#/media/Ficheiro:Comparison\\_mean\\_median\\_mode\\_-\\_PT.svg](https://pt.wikipedia.org/wiki/Mediana_(estat%C3%ADstica)#/media/Ficheiro:Comparison_mean_median_mode_-_PT.svg), acessado em 01/07/2020