Universidade Federal do Rio Grande do Norte
Instituto Metrópole Digital
IMD0601 – Bioestatística

Medidas de Tendência Central

Renata Lilian Dantas Cavalcante
Mestranda PPG-Bioinformática - UFRN

Por que estudar bioestatística é legal?

Deixando alguns conceitos mais claros...

- Estatística: é a ciência que tem por objetivo **planejar**, **coletar**, **tabular**, **analisar** e **interpretar** informações de dados experimentais.

- Bioestatística: consiste na aplicação da estatística nos campos relacionados às ciências da vida.

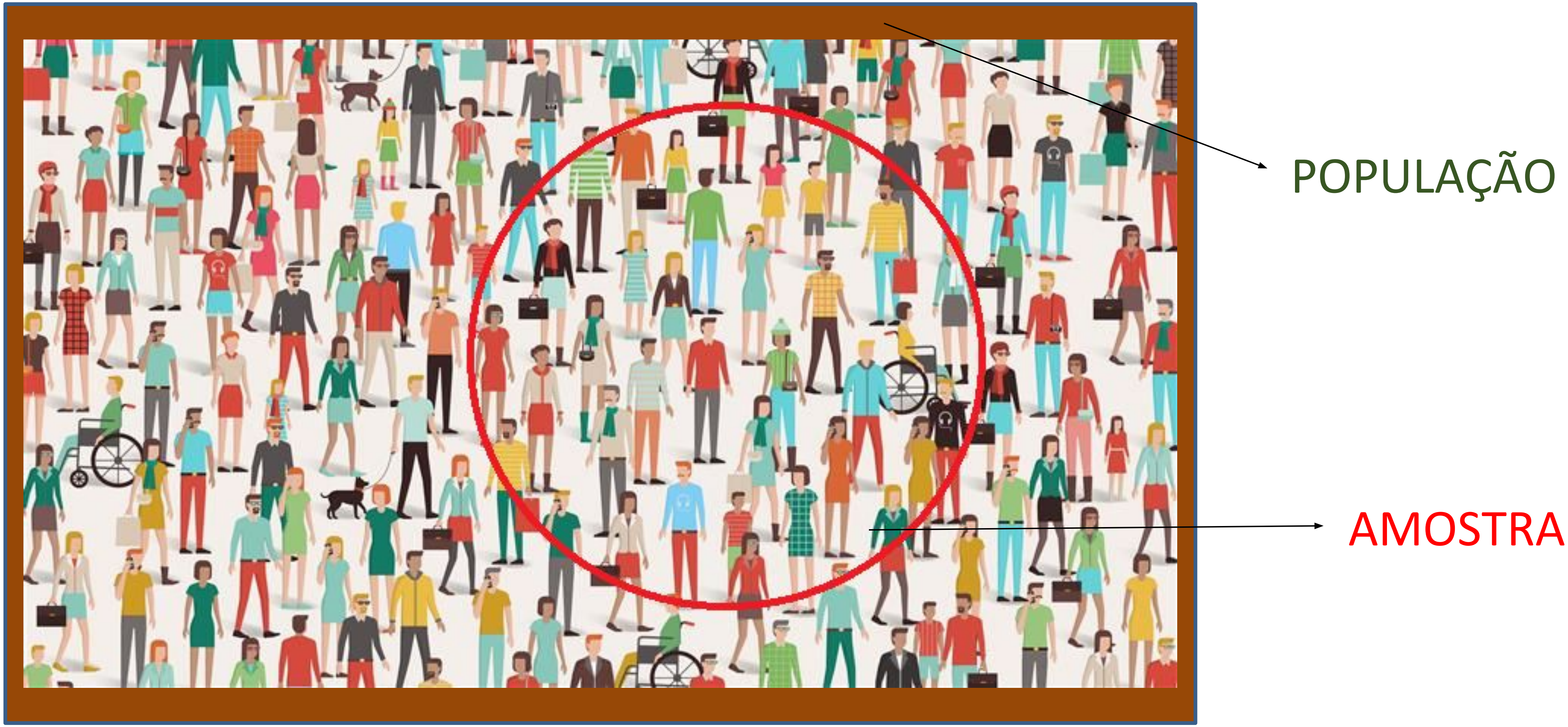
- Áreas:
  - $\longrightarrow$ Estatística Descritiva.
  - $\longrightarrow$ Estatística Inferencial ou Indutiva.

- Estatística descritiva: descreve, analisa e representa um conjunto de dados, utilizando métodos numéricos e gráficos que resumem e apresentam a informação neles contidos.

- Faz-se uso de: tabelas e gráficos, medidas de centralidade (MTC) e dispersão (MD).

- População: conjunto de elementos que possuem pelo menos uma característica em comum de interesse a ser analisada.

- Amostra: subconjunto finito de elementos em uma população, que são representativos para o estudo de uma determinada característica de interesse na população.


POPULAÇÃO
AMOSTRA

Medidas de tendência central

- Costumam representar um conjunto de dados, com valores centrais pelos quais os dados tendem a concentrar-se.

- A sua medida dá uma ideia de onde localiza-se o centro (o ponto médio) de um determinado conjunto de dados.

- São medidas de resumo dos dados.

- Ex.: valores típicos que servem para caracterizar uma população.

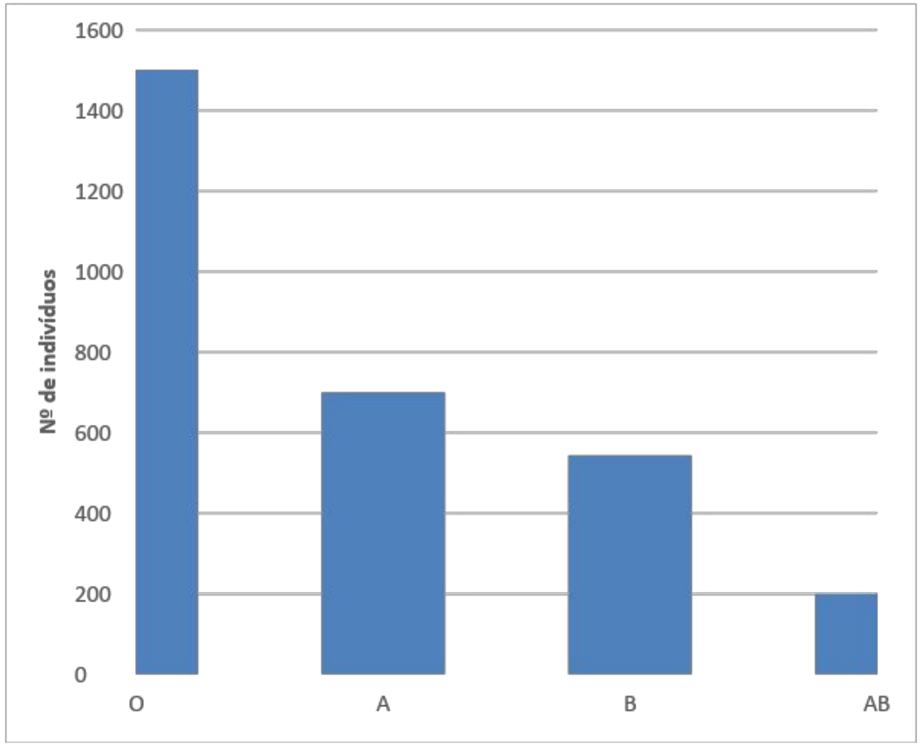
- Existem várias maneiras de expressar essa tendência central. As mais comuns são: **Moda**, **Mediana** e **Média**.

Moda (*Mo*)

- É definida como a observação mais frequente do conjunto de valores observados, ou seja, o valor que mais aparece dentro da amostra.

- Uma amostra pode ser:
  → Amodal: não apresenta uma moda.
  → Unimodal: 1 único valor aparece mais.
  → Bimodal: 2 valores aparecem mais.
  → Multimodal: 3 ou + valores.

Tipo Sanguíneo	Número de indivíduos
O	1500
A	700
B	543
AB	200



Mediana (Md)

- É a observação que indica exatamente o ponto médio de um conjunto de dados quando estes estão ordenados (crescente ou decrescente).

- Sua principal característica é dividir o conjunto de dados em duas partes com o mesmo número de elementos.

- Quantidade ímpar de valores: a Md corresponderá ao termo central do conjunto de dados.

- Quantidade par de valores: a Md corresponderá a média dos dois termos centrais.

Método para determinação da Mediana

Amostra obtida

$$\{ x_1, x_2, \ldots, x_k, \ldots, x_n \}$$

$\longrightarrow$

Amostra ordenada

$$\{ x_{(1)}, x_{(2)}, \ldots, x_{(k)}, \ldots, x_{(n)} \}$$

min k-ésima max

- Série com número par de termos:

[(n/2) +(n/2+ 1)] / 2

Ex.: Calcule a mediana da amostragem de plantas com flores { 1, 3, 0, 2, 4, 1, 2, 5 }

1º - ordenar a série { 0, 1, 1, 2, 2, 3, 4, 5 } n = 8

[(n/2) +(n/2+ 1)] / 2 ⟶ [(8/2) +(8/2+ 1)] / 2 = 4,5
Ou seja, a média entre o 4º e 5º elemento da série ordenada será a mediana.

A mediana será o elemento = **2**

- Série com número ímpar de termos:

(n + 1) / 2

Ex.: Calcule a mediana da amostragem de plantas com flores { 1, 3, 0, 0, 2, 4, 1, 2, 5 }

1º - ordenar a série { 0, 0, 1, 1, 2, 2, 3, 4, 5 } n = 9

(n + 1)/2 ⟶ (9+1) / 2 = 5
Ou seja, o 5º elemento da série ordenada será a mediana.

A mediana será o 5º elemento = **2**

Média

- É o valor médio das observações.

- Tipos:

    → Média Aritmética;

    → Média Ponderada;

    → Média Geométrica;

    → Média Harmônica;

Média Aritmética

- Consiste na soma dos valores observados dividido pelo número de observações.

- É a medida de tendência central mais utilizada.

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

Média Aritmética

Ex.: Calcule a média de crianças internadas com viroses em um hospital por dia, baseando-se nos seguintes dados semanais: {2,1,8,9,6,4,5}.

Média = (2 + 1 + 8 + 9 + 6 + 4 + 5)/7
Média = 35/7
Média = **5 crianças por dia.**

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

Média Ponderada

- Em alguns casos, um determinado valor pode repetir-se várias vezes ao longo de uma amostra e a quantidade de vezes que esse número se repete é denominado **peso.**

- A média ponderada nada mais é que a média aritmética com o uso de pesos.

$$M_p = \frac{x_1 p_1 + x_2 p_2 + \cdots + x_n p_n}{p_1 + p_2 + \cdots + p_n}$$

$$\boxed{M_p = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}}$$

Média Ponderada

Espécie	Frequência (f_i)	Peso
A	8,2	3
B	10	2
C	9,5	4
D	7,8	2
E	10	2
F	9,5	3
G	6,7	4

$$M_p = \frac{3\,.\,8{,}2 + 2\,.\,10{,}0 + 4\,.\,9{,}5 + 2\,.\,7{,}8 + 2\,.\,10 + 3\,.\,9{,}5 + 4\,.\,6{,}7}{3 + 2 + 4 + 2 + 2 + 3 + 4}$$

$$M_p = \frac{24{,}6 + 20 + 38 + 15{,}6 + 20 + 28{,}5 + 26{,}8}{20}$$

$$M_p = \frac{173{,}5}{20}$$

$$M_p = 8{,}7$$

Média Geométrica

- Utilizada principalmente em dados que possuem natureza exponencial.

- Retorna a média do produto dos elementos de um conjunto.

$$\left(\prod_{i=1}^{n} a_i\right)^{1/n} = \sqrt[n]{a_1\ a_2\ a_3 \ldots .a_n}$$

$$M_G = \sqrt[n]{x_1 . x_2 . x_3 \ldots . x_n}$$

- M_G: média geométrica;
- n: número de elementos do conjunto de dados;
- $x_1, x_2, x_3, ..., x_n$: valores dos dados;

Média Geométrica

Exemplo: Qual o valor da média geométrica entre os números 9, 8 e 3?

$$M_G = \sqrt[3]{3\,.8.9} = \sqrt[3]{216} = 6$$

- M_G: média geométrica;
- n: número de elementos do conjunto de dados;
- x_1, x_2, x_3, ..., x_n: valores dos dados;

Média Harmônica

- Utilizada principalmente em dados que estão organizados por taxas e proporções (grandezas inversamente proporcionais).

- O valor é obtido dividindo o número de elementos pela soma dos valores recíprocos de cada elemento.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

→ Considerando os conjuntos A e B, abaixo:

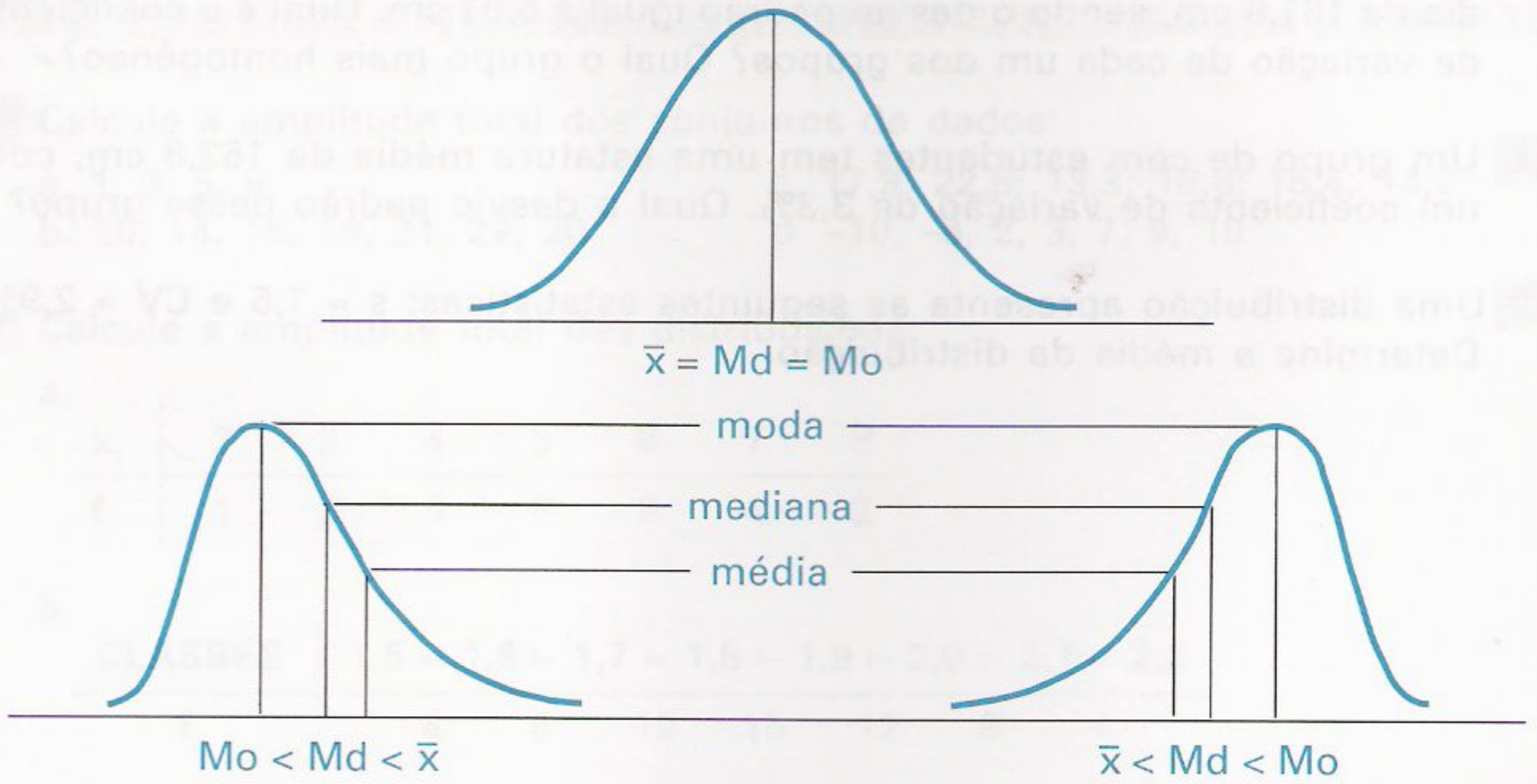
A = (1, 3, 5, 7, 9)

Média = 5
Mediana = 5

B = (2, 3, 5, 7, 58)

Média = 15
Mediana = 5

Enquanto a média é afetada por valores extremos, a mediana é mais "robusta", ou seja, não sofre influência de valores extremos.



$\bar{x}$ = Md = Mo
moda
mediana
média
Mo < Md < $\bar{x}$
$\bar{x}$ < Md < Mo

Dados agrupados (Média)

Idade	Frequência (f_i)
25-30	1
30-35	2
35-40	6
40-45	16
45-50	10
50-55	8
55-60	4
60-65	3
65-70	2
TOTAL	**52**

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

- x_i: ponto médio do intervalo;
- f_i: frequência;
- n: soma das frequências;

1º Calcular o ponto médio de uma classe:

$$\text{Ponto médio} \rightarrow x_i = \frac{LI_{classe} + LS_{classe}}{2}$$

$$x_i = \frac{25 + 30}{2} = 27{,}5$$

Dados agrupados (Média)

Idade	Frequência (f_i)	Ponto Médio (x_i)
25-30	1	27,5
30-35	2	32,5
35-40	6	37,5
40-45	16	42,5
45-50	10	47,5
50-55	8	52,5
55-60	4	57,5
60-65	3	62,5
65-70	2	67,5
TOTAL	**52**	**427,5**

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

$$\text{Média} = \frac{27{,}5*1 + 32{,}5*2 + 37{,}5*6 + 42{,}5*16 + \ldots + 67{,}5*2}{1 + 2 + 6 + 16 + \ldots + 2}$$

Média = 2445/52

Média = **47,02**

Dados agrupados (Média)

Idade	Frequência (f_i)	Ponto Médio (x_i)	$f_i * x_i$	Acumulado
25-30	1	27,5	27,5	1
30-35	2	32,5	65	3
35-40	6	37,5	225	9
40-45	16	42,5	680	25
45-50	10	47,5	475	35
50-55	8	52,5	420	43
55-60	4	57,5	230	47
60-65	3	62,5	187,5	50
65-70	2	67,5	135	52
TOTAL	**52**	**427,5**	**2445**	

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

x_i: ponto médio do intervalo;
f_i: frequência;
n: soma das frequências;

Média = 2445/52
Média = **47,02**

Dados agrupados (Mediana)

Idade	Frequência (f_i)	Acumulado	Posições
25-30	1	1	1ª
30-35	2	3	2ª a 3ª
35-40	6	9	4ª a 9ª
40-45	16	25	10ª a 25ª
45-50	10	35	26ª a 35ª
50-55	8	43	36ª a 43ª
55-60	4	47	44ª a 47ª
60-65	3	50	48ª a 50ª
65-70	2	52	51ª a 52ª
TOTAL	**52**	**--**	--

$$Md = L_{Md} + \left(\frac{\frac{n}{2} - F_{ant}}{f_{Md}} \right) . h$$

- L_{Md}: limite inferior do classe que contém a mediana;
- n: número total de valores;
- F_{ant}: frequência acumulada da classe anterior à classe mediana;
- f_{Md}: frequência do grupo que possui a mediana;
- h: amplitude do intervalo da classe mediana;

1º passo: encontrar a classe mediana → 52/2 = **26ª posição**

Idade	Frequência (f_i)	Acumulado	Posições
25-30	1	1	1ª
30-35	2	3	2ª a 3ª
35-40	6	9	4ª a 9ª
40-45	16	25	10ª a 25ª
45-50	10	35	26ª a 35ª
50-55	8	43	36ª a 43ª
55-60	4	47	44ª a 47ª
60-65	3	50	48ª a 50ª
65-70	2	52	51ª a 52ª
TOTAL	**52**	**--**	--

- $L_{Md} = 45$
- $F_{ant} = 25$
- $f_{Md} = 10$
- $h = 5$
- $n = 52$

$$Md = L_{Md} + \left(\frac{\frac{n}{2} - F_{ant}}{f_{Md}} \right) . h$$

$$Md = 45 + \left(\frac{\frac{52}{2} - 25}{10} \right) * 5$$

Md = 45 + 0,5

Mediana = **45,5**

Dados agrupados (Moda)

Método de Czuber:

$$Mo = L_{Mo} + \left(\frac{D_1}{D_1 + D_2}\right).h$$

$$D_1 = f_{mo} - f_{ant}$$
$$D_2 = f_{mo} - f_{post}$$

- L_{mo}: limite inferior da classe modal.
- f_{mo}: frequência da classe modal.
- f_{ant}: frequência da classe imediatamente anterior à classe modal.
- f_{post}: frequência da classe imediatamente posterior à classe modal.
- h: amplitude da classe modal.

Dados agrupados (Moda)

Método de Czuber Ex.: Calcule a moda para a distribuição de frequências das idades das mulheres diagnosticadas com câncer de mama.

IDADE	fi
25-30	2
30-35	6
40-45	16
45-50	10
50-55	8
55-60	4
60-65	3
65-70	2
TOTAL	**51**

$$Mo = L_{Mo} + \left(\frac{D_1}{D_1 + D_2}\right).h$$

$$D_1 = f_{mo} - f_{ant}$$
$$D_2 = f_{mo} - f_{post}$$

L_{Mo}: 40
f_{mo}: 16
f_{ant}: 6
f_{post}: 10
h: 5

Logo:

D_1: 16-6 = 10

D_2: 16-10 = 6

A moda será:

M_o = 40 + (10/10+6)*5

M_o = 40 + 10/16*5

M_o = **43,13**

Dados agrupados (Moda)

Método de King Ex.: Calcule a moda para a distribuição de frequências das idades das mulheres diagnosticadas com câncer de mama.

IDADE	fi
25-30	2
30-35	6
40-45	16
45-50	10
50-55	8
55-60	4
60-65	3
65-70	2
TOTAL	**51**

$$Mo_{king} = l_i + \left(\frac{f_{post}}{f_{ant} + f_{post}} \right) h$$

- l_i: limite inferior da classe modal;
- f_{ant}: frequência da classe anterior à modal;
- f_{post}: frequência da classe posterior à modal;
- h: amplitude da classe modal;

- l_i: 40
- f_{ant}: 6
- f_{post}: 10
- h: 5

$$Mo_{king} = 40 + \left(\frac{10}{6 + 10} \right) 5$$

$$Mo_{king} = \mathbf{43{,}13}$$

Dados agrupados (Moda)

Método de Pearson:

IDADE	fi
25-30	2
30-35	6
40-45	16
45-50	10
50-55	8
55-60	4
60-65	3
65-70	2
TOTAL	**51**

$$\mathbf{M_o} = 3M_d - 2\overline{X}$$

- $\overline{X}$: valor da média;
- M_d: valor da mediana;

$$M_o = 3 * 45{,}5 – 2*47{,}02$$

$$M_o = 136{,}5 - 94{,}04$$

$$M_o = \mathbf{42{,}5}$$

OBRIGADA