

MEDIDAS de POSIÇÃO e DISPERSÃO

Prof. Rogério de Melo Costa Pinto

Universidade Federal De Uberlândia

Faculdade de Matemática

Medidas de posição

A medida de posição tem o objetivo de resumir o conjunto de dados em alguns valores que possam representar a variável

As principais medidas de posição são:

- Média Aritmética
- Mediana
- Moda

Média

É a mais utilizada devido:
ser a mais comum
facilidade de compreensão
simplicidade de cálculo
boas propriedades algébricas.

Média Aritmética

μ = média populacional

\bar{x} = média amostral

Média Aritmética (dados não agrupados)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Exemplos de cálculo

A seguir temos notas de 8 pessoas em um teste.

50 80 30 100 80 70 70 60

Calcular a nota média.

Deseja-se saber a média diária de visitas realizadas por enfermeiros de um hospital aos seus pacientes nas suas respectivas residências

5 3 4 2 2 6 3 6 5 4

Media aritmética → dados agrupados

$$\bar{x} = \frac{\sum X_i f a_i}{\sum f a_i} = \sum X_i f r_i \quad X_i = \text{ponto médio da classe } i$$

Exemplo: A tabela a seguir representa a distribuição de freqüências do tempo (min) gasto em uma determinada tarefa.

Distribuição de freqüências tempo (min) gasto na tarefa X.	
Tempo (min)	fa
20 ----25	8
25 ----30	12
30 ----35	20
35 ----40	8
40 ----45	2

Calcular a média.

Propriedades da média aritmética

A soma algébrica dos desvios em relação à média é nula.

A soma de quadrados dos desvios de um conjunto de dados, em relação a uma constante qualquer K, será mínima se e somente se $K = \bar{x}$.

Somando-se (ou subtraindo-se) uma constante (c) a todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

▪ **Ex1: número de acidentes semanais.**

▪ **2, 4, 4, 3, 2, 40 média = 9**

▪ **Ex2: Salários**

▪ **Xi: 1 10 100**
 50 2 1 média = 3,2

OUTROS TIPOS DE MÉDIAS

• **ponderada** => utilizada quando os elementos da amostra recebem pesos diferentes. Um exemplo clássico é o cálculo de média escolar.

valores: y_1, y_2, \dots, y_k pesos: w_1, w_2, \dots, w_k

$$\bar{y}_p = \frac{y_1 w_1 + y_2 w_2 + \dots + y_k w_k}{w_1 + w_2 + \dots + w_k} = \frac{\sum_{i=1}^k y_i w_i}{\sum_{i=1}^k w_i}$$

Exemplo: As notas obtidas pelo aluno nas disciplinas A, B, C, D e E, foram, respectivamente, 80; 85; 90 e 70. Se os pesos destas disciplinas são: 3, 5, 2, 1, respectivamente, determine a média ponderada do aluno.

Tempo (segundos) que um funcionário do hemocentro leva para fazer uma análise do sangue do doador.

Segundos (Xi)	fa	fr	fa .Xi
59	4	0,4	236
60	2	0,2	120
61	4	0,4	244
Total	10		600

Mediana

- **50% abaixo e 50% acima de um conjunto de dados ordenados.**
- **Posição da mediana**
- **Dados não agrupados**
- **Dados agrupados em classes**

Dados não agrupados

mediana de um conjunto de valores é o valor que ocupa a posição $(n + 1)/2$ (**ordenados**)

Se $(n + 1)/2$ for fracionário => média dos dois valores que estão nas posições imediatamente abaixo e acima de $(n + 1)/2$.

Exemplo - Notas das três turmas em uma prova

Turma	Valores
A	4 5 5 6 6 7 7 8
B	1 2 4 6 6 9 10 10
C	0 6 6 7 7 7 7,5 7,5

Dados agrupados

Sem intervalos de classe.

Xi	fa	Fac
2	5	5
4	10	15
6	45	60
8	12	72
10	5	77
12	3	80
Total	80	

Dados agrupados em classes

$$Md = li + \left(\frac{(n/2) - f_{antac.}}{f_{med}} \right) \times c$$

Opção alternativa → usar o ponto médio da classe mediana (Hipótese Tabular Básica)

Exemplo – Taxa de equipamentos a serem consertados em um laboratório

Classes	Freq.	freq. Acum.
9,9 – 18,62	10	10
18,62 – 27,34	13	23
27,34 – 36,06	6	29
36,06 – 44,78	4	33
44,78 – 53,5	0	33
53,5 – 62,2	1	34
Total	34	-

Observações com relação a mediana

- O valor da mediana coincide com um elemento da série só se o n° de elementos da série é ímpar.
- A média aritmética e a mediana não têm, necessariamente, o mesmo valor.
- A mediana depende da posição, e não dos valores dos elementos na série ordenada. Essa é uma diferença marcante entre ela e a média aritmética, que se deixa influenciar pelos valores extremos.

Veja:

$$5, 7, 10, 13, 15 \rightarrow \bar{x} = 10 \text{ e } Md = 10$$

$$5, 7, 10, 13, 65 \rightarrow \bar{x} = 20 \text{ e } Md = 10$$

Moda

- Valor que ocorre com maior frequência
- Distribuições amodal, unimodal, multimodal
- Dados não agrupados
- Dados agrupados em classes

▪ (1, 5, 3, 7, 5, 8, 3, 5, 9, 5)

▪ (1, 3, 5, 8, 7, 9)

▪ (1, 5, 3, 8, 5, 9, 3)

Qual a temperatura modal da distribuição de frequências de temperatura de um determinado mês.

Temperatura (°C)	Frequência
20	3
21	9
22	12
23	6

Dados agrupados em classes

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \times c$$

Alternativa → Usar o ponto médio da classe modal (HTB)

Exemplo – Taxa de equipamentos a serem consertados em um laboratório

Classes	Freq.
9,9 – 18,62	10
18,62 – 27,34	13
27,34 – 36,06	6
36,06 – 44,78	4
44,78 – 53,5	0
53,5 – 62,2	1
Total	34

Características da Moda

- Não é afetada pelos valores extremos da distribuição, desde que estes valores não constituam a classe modal.
- É uma medida bastante utilizada na área econômica.
sindicalista x patrão

SEPARATRIZES (quartil, decil, percentil)

- Quartil → separação dos dados em 4 partes iguais
- Decil → separação dos dados em 10 partes iguais
- Percentil → separação dos dados em 100 partes iguais

Quartis: Q1 → primeiro quartil; Q2 → segundo quartil e Q3 → terceiro quartil

Decis: D1 → primeiro decil; ...; D9 → nono decil

Percentis: P1 → primeiro percentil; ...; P99 → percentil 99

Note que, abaixo de Q1 temos 25% dos dados, entre Q1 e Q2 temos 25%, entre Q2 e Q3 temos 25% e acima de Q3 também 25%. De maneira análoga pode-se fazer com decil (10% para cada decil) e percentil (1% para cada percentil).

Note também que: Q1 = P25; Q2 = P50 = Md = D5; Q3 = P75

Exemplo: Determinar os quartis para:

1. Idade em anos de pessoas aprovadas no vestibular
18 19 17 18 21 20 24 22 18 20 19

2. Dados de temperatura mínima mensal (°C) em uma certa região.
28 27 30 25 22 15 14 21 25 26 29 29

3. Número de casas com foco de dengue

Distribuição de freqüências do número de casas com foco da dengue

Número de casas	fa
0	125
1	45
2	15
3	10
4	4
5	1
Total	200

Considerações sobre as medidas de posição

- é expressiva quando existe uma considerável concentração das freqüências em torno da medida
- não descreve o grau de variação
- não podemos estar seguro se ela é ou não uma medida expressiva da tendência central dos dados.
- Se a distribuição é assimétrica, a mediana ou a moda podem ser mais apropriadas para refletir a tendência central dos dados do que a média aritmética

Exercício 1. As amostras de sangue de três diferentes laboratórios apresentaram os níveis de creatina (mg%) mostrados no quadro a seguir:

Exame →	1	2	3	4	5	6	7
Laboratório A	0,6	0,4	0,5	0,8	0,2	0,8	-
Laboratório B	0,7	0,8	0,6	0,9	0,5	1,1	0,3
Laboratório C	0,6	0,7	2,0	0,5	0,8	0,9	0,9

Creatina (mg%); Dados hipotéticos

O nível de creatina normal está entre 0,4 – 1,3mg% e um valor central de 0,85mg%

- Calcular a média, a mediana e a moda para os dados do Laboratório A;
- Calcular a média, a mediana e a moda para os dados do Laboratório B;
- Calcular a média, a mediana e a moda para os dados do Laboratório C;
- Discorrer sobre as vantagens e as desvantagens de cada uma das medidas de tendência central (MTC) no caso das amostras de cada laboratório

- Calcular a média global dos dados;
- Calcular a mediana para os dados dos três laboratórios;
- Calcular a moda de todos os níveis de creatina;
- Os problemas verificados com a média, no caso dos dados do Laboratório C, continuam no caso da média de todos os dados?;
- Qual seria a MTC mais representativa do conjunto de dados dos três laboratórios?

Ex4. O professor de educação física procedeu um levantamento a respeito das estaturas de seus alunos, verificando que, das 22 medições efetuadas, as alturas situaram-se entre 148cm e 182cm, acusando média aritmética igual a 164cm. Posteriormente, ele constatou que a fita métrica utilizada apresentava 3cm a menos. Em vista da retificação que deverá ser feita, pergunta-se:

- Será necessário medir novamente todos os alunos?
- Qual a nova média?

Ex5. Foi calculada a média aritmética dos QI de 8 indivíduos, encontrando-se 78 pontos. Sabendo que os resultados foram 85, 68, 92, 64, 80, 76, 86 e 73, verifique se a média está correta, sem calculá-la pela fórmula.

Medidas de Dispersão

Medidas de Dispersão

A média é uma medida de centro da distribuição, porém, nada informa com relação a dispersão dos valores em torno do centro. Quanto maior a variação dos dados menor a representatividade da média. Assim, dizemos que as medidas de dispersão servem para qualificar a média. Quanto menor a dispersão, mais confiável é a média.

As características das medidas de dispersão são:

- Medem a variabilidade de um conjunto de dados
- Sempre devem estar associadas a medidas de posição
- Deve ser compreensível e apresentar facilidade de cálculo
- Deve estar exposta o menos possível as flutuações das amostras

Exemplo: suponha que desejamos comparar três amostras diferentes (X, Y e Z):

X: 70, 70, 70, 70, 70. ($\bar{x} = 70$)

Y: 50, 60, 70, 80, 90. ($\bar{y} = 70$)

Z: 5, 15, 50, 120, 160. ($\bar{z} = 70$)



Medem o grau de afastamento dos dados em torno da média

AMPLITUDE TOTAL

A maneira mais natural e simples de se medir a variabilidade de uma amostra é calcular a diferença entre o maior e o menor valor.

$$A = R = \text{maior} - \text{menor valor} = \text{Max} - \text{Min}$$

- é limitada, representada apenas pelos valores extremos da amostra
- é instável, não sendo afetada pela dispersão dos valores internos

Exemplo: Determinar a amplitude total para:

a) Volume de venda (milhares de toneladas) de adubo químico

100 130 120 120 110 105 118 95

b) Notas

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

Vamos pensar em uma medida de dispersão mais adequada

Desvio

$$desvio = d = \sum_{i=1}^n (x_i - \bar{x}) \quad \text{onde} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- problema do zero

$$desvio = d = \sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n x_i - n \frac{\sum_{i=1}^n x_i}{n} = 0$$

$$|d| = \sum_{i=1}^n |(x_i - \bar{x})|$$

Desvio médio

Para evitar que amostras de tamanhos diferentes, que possuíssem mesma dispersão, apresentassem desvios diferentes, ponderou-se o desvio por n

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

Exemplo: Determinar o desvio e desvio médio das notas do exemplo anterior.

Problemas:

- dificuldade de cálculo e de interpretação em algumas situações
- Não é uma medida mínima de erro

Desvio quadrático médio

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

- É uma média mínima de erro
- Apresenta facilidade de cálculo e de interpretação
- Apresenta boas propriedades algébricas
- A medida de dispersão definida como acima apresenta o inconveniente de ser tendenciosa, ou seja, de afastar do valor real da população, entretanto, a partir desse desvio podemos definir as medidas de dispersão mais usadas na estatística

Variância e desvio padrão

Variância

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

Desvio padrão => raiz quadrada positiva da variância

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

Exemplo 1: Os dados a seguir representam o tempo de espera na fila (min) de pacientes na UAI X.

20 15 15 18 17 16 15

Determinar a média e o desvio padrão e interpretar.

Dados agrupados em classes

$$s^2 = \frac{SQD}{(\sum f_i) - 1} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 f_i}{(\sum f_i) - 1} = \frac{\sum X_i^2 f_i - \frac{(\sum X_i f_i)^2}{\sum f_i}}{(\sum f_i) - 1}$$

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

Exemplo – Taxa de equipamentos a serem consertados em um laboratório

Classes	Freq.
9,9 -- 18,62	10
18,62 -- 27,34	13
27,34 -- 36,06	6
36,06 -- 44,78	4
44,78 -- 53,5	0
53,5 -- 62,2	1
Total	34

Calcular a média, o desvio padrão e a variância

Propriedades e características

- O desvio padrão é um valor mínimo de erro, pois os desvios são calculados em relação à média aritmética
- São estatísticas que utilizam todas as observações no cálculo e também sofrem pouca influência de mudanças amostrais
- Valores extremos exercem maior influência que os valores centrais (próximos à média)

- Variância e desvio padrão são sempre positivos
- São estatísticas de grande importância na descrição das populações e nas inferências
- A variância de uma constante é sempre igual a zero
- A variância de uma soma de variáveis será igual a soma das variâncias de cada variável, se e somente se as variáveis forem independentes

- Se somarmos ou subtrairmos uma constante a cada valor observado e calcularmos a nova variância e o novo desvio padrão, verificaremos que estes não se alteram em relação aos dados iniciais.
- A multiplicação ou divisão de cada observação por uma constante, faz com que a nova variância seja multiplicada ou dividida pela constante ao quadrado e novo desvio padrão fique multiplicado ou dividido pela constante.

Comparando a variabilidade entre duas amostras

	Temperatura (°C)	Precipitação (mm)
Média	22	800
s	5	100

Coefficiente de Variação

$$CV = 100 \cdot \frac{s}{\bar{X}}$$

Usado para comparar variabilidade se uma das condições abaixo não forem satisfeitas:

- **Mesmo número de observações**
- **Unidades iguais**
- **Médias iguais**

ERRO PADRÃO DA MÉDIA

Expressa a precisão com que a média foi estimada

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Observação: maior tamanho da amostra implica em menor erro padrão

Exemplo: Determinar o erro padrão para o tempo de espera na fila do banco, usando uma amostra de tamanho 7 e uma amostra de tamanho 15

(A) 20 15 15 18 17 16 15

(B) 15 18 15 18 17 20 22 14 16 15 15 17 19 16 15