

Medidas de Localização

ESTAT0011 – Estatística Aplicada

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

Medidas de Localização

- São estatísticas que resumem a informação contida nos dados.
- Elas estabelecem valores em torno dos quais os dados se distribuem ou determinam pontos importantes na distribuição dos dados.
- **Exemplos:** média, mediana, moda, separatrizes.

Média

- A **média aritmética** (ou simplesmente *média*) é a medida de localização mais utilizada, representando o valor central de um conjunto de dados.
- Ela é calculada como a soma de todos os valores dividida pelo número de observações.

Média

Média Populacional

- Se temos uma população com N elementos (X_1, X_2, \dots, X_N) , a média é dada por

$$\mu = \frac{\sum_{i=1}^N X_i}{N},$$

onde μ é o símbolo padrão para *média populacional*.

- A média populacional é a média **verdadeira**, calculada quando temos **TODOS** os elementos do grupo de interesse.
- **Exemplo:** Se uma universidade tem 1.000 alunos e registramos a nota de cada um, a média dessas 1000 notas é μ .

Média

Média Amostral

Exemplo

Suponha que uma analista de TI queira calcular a média de falhas diárias no servidor de uma empresa durante uma semana. Os registros foram:

Dia	Seg	Ter	Qua	Qui	Sex
Número de falhas	5	3	7	4	6

Temos que $n = 5$ e obtemos

$$\bar{X} = \frac{\sum_{i=1}^5 X_i}{5} = \frac{5 + 3 + 7 + 4 + 6}{5} = \frac{25}{5} = 5.$$

Ou seja, há uma média de 5 falhas por dia.

Vantagens da Média

- V1.** É a medida mais conhecida e de maior uso;
- V2.** É facilmente calculável;
- V3.** Serve para compararmos conjuntos semelhantes;
- V4.** Suas propriedades são bem compreendidas.

Desvantagens da Média

- D1.** É uma medida sensível a observações extremas. Ou seja, é fortemente influenciada pelos valores extremos (muito grandes ou muito pequenos) do conjunto;
- D2.** Só deve ser utilizada quando a distribuição dos dados for simétrica (normal ou Gaussiana).

Média Ponderada

- Calculada multiplicando cada valor X_i por um peso w_i e dividindo sua somatória pela soma dos pesos:

$$X_w = \frac{\sum_{i=1}^n (w_i \cdot x_i)}{\sum_{i=1}^n w_i}.$$

Exemplos:

- No cálculo da média de dados de múltiplos sensores, reduza o peso dos menos precisos.
- Atribuir maior peso a grupos sub-representados e menor peso a grupos super-representados.

Exemplo

Um analista de TI precisa calcular o tempo médio de resposta de um sistema, considerando o peso definido com base no tráfego de cada servidor:

Servidor	Tempo de resposta (ms)	Peso (w_i)
Primário	120	0,6
Backup 1	200	0,3
Backup 2	150	0,1

Temos:

$$\bar{X}_w = \frac{(120 \times 0,6) + (200 \times 0,3) + (150 \times 0,1)}{0,6 + 0,3 + 0,1} = 147 \text{ ms}$$

Exemplo

- A média simples seria 156,7 ms, mas a ponderada (147 ms) reflete melhor a experiência do usuário, já que a maioria acessa o servidor mais rápido (Primário).

Mediana

- É o valor que divide os dados ordenados em duas partes iguais.
- Podemos encontrar a mediana de um conjunto de dados das seguintes formas:
 - **Se n é ímpar:** a mediana será o valor central dos dados ordenados, ou seja, o valor na posição $\frac{n+1}{2}$.
 - **Se n é par:** a mediana será a média entre os dois valores centrais, isto é, a média dos valores nas posições $\frac{n}{2}$ e

$$\left(\frac{n}{2} + 1\right).$$

Vantagens da Mediana

- V1.** Não é influenciada por valores extremos de um conjunto de dados;
- V2.** É utilizada especialmente para distribuições assimétricas, mas pode ser usada para dados com distribuição simétrica também.

Desvantagens da Mediana

- D1.** Suas prioridades não são bem compreendidas;
- D2.** Não é levada em consideração na maior parte dos testes estatísticos.

Obs.: Quando os dados são simétricos, a média e a mediana coincidem.

Exemplo

Um estudante precisa calcular o tempo de resposta (em segundos) de dois servidores diferentes para identificar qual tem desempenho mais consistente:

Servidor A	5,2	3,8	4,5	6,1	2,9	
<hr/>						
Servidor B	4,1	5,7	2,3	3,6	6,0	1,8

- Para calcular a mediana, primeiro ordenamos os dados:

Servidor A	2,9	3,8	4,5	5,2	6,1	
<hr/>						
Servidor B	1,8	2,3	3,6	4,1	5,7	6,0

Exemplo

- Como o tamanho da amostra do servidor A é ímpar ($n = 5$), a mediana é o valor central, isto é, o valor na posição $\frac{n+1}{2} = 3$. A mediana então é 4, 5.
- Para o servidor B foi obtida uma amostra de tamanho par ($n = 6$), logo a mediana corresponderá à média dos valores nas

posições $\frac{n}{2} = 3$ e $\left(\frac{n}{2} + 1\right) = 4$. Assim, a médiana é

$$\frac{3,6 + 4,1}{2} = 3,85.$$

Moda

- É o valor que ocorre com maior frequência.
- A moda pode não existir ou assumir mais de um valor.

Exemplos:

- 1; 2; 3; 3; 4; 4; 4; 4; 5; 5; 6; 6; 6; 6; 6; 6; 7; 7; 8
 - Moda = 6
- 1; 1; 2; 2; 3; 3; 4; 4; 5; 5
 - Moda não existe (conjunto de dados **amodal**).
- 2; 2; 3; 3; 3; 4; 4; 5; 5; 5
 - Moda = 3 e 5

Vantagem da Moda

V1. Não é influenciada por valores extremos de um conjunto de dados;

Desvantagem da Moda

D1. Não depende de todos os valores do conjunto de dados, podendo mesmo não se alterar com a modificação de alguns deles.

Exemplo 2.1

Um professor registra o número de faltas de seus alunos em um determinado semestre. Em uma amostra aleatória, os dados são:

2 4 2 0 40 2 4 3 6

- Calcule a média, a mediana e a moda.
- Suponha que o aluno com 40 faltas abandone o curso. Como ficam a média, a mediana e a moda?

Separatrizes

- São valores que ocupam determinadas posições considerando os dados ordenados.
- As separatrizes podem ser classificadas em:
 - **Mediana:** divide o conjunto de dados duas em partes iguais (M_d).
 - **Quartis:** Dividem os dados em quatro partes iguais (Q_1, Q_2, Q_3).
 - **Decis:** Dividem os dados em dez partes iguais ($D_1, D_2, D_3, \dots, D_9$).
 - **Percentis:** Dividem os dados em cem partes iguais ($P_1, P_2, P_3, \dots, P_{99}$).

Exemplo 2.2

Obtenha o primeiro e o terceiro quartil (Q_1 e Q_3) dos dados abaixo (note que já estão ordenados).

9	15	19	22	24	25	30	34	35	35
<hr/>									
36	36	37	38	42	43	46	48	54	55
<hr/>									
56	56	59	62	69	70	82	82	89	139

Fim