

Estatística e Probabilidade

Carolina Silva Pena

Exemplo Base de Dados

Id	Turma	Sexo	Idade	Horas_estudo	ENEM_MAT	Notas_EST
1	B	M	21	Até 1h	346,82	2,79
2	B	F	22	Até 1h	634,38	4,30
3	B	M	21	Mais de 1h e até 3h	561,03	4,46
4	B	F	22	Mais de 3h e até 5h	602,91	7,64
5	A	F	21	Mais de 3h e até 5h	751,11	7,70
6	B	F	22	Até 1h	441,68	2,09
7	A	M	24	Até 1h	571,72	4,94
8	A	M	23	Mais de 1h e até 3h	606,97	5,78
9	A	M	21	Mais de 5h	731,62	8,33
10	B	F	18	Mais de 3h e até 5h	709,59	7,45
11	B	M	21	Até 1h	502,09	5,28
12	B	F	20	Mais de 5h	779,17	10,00
13	B	M	23	Mais de 5h	640,57	7,80
14	A	F	17	Mais de 1h e até 3h	564,35	5,56
15	B	F	20	Até 1h	504,76	4,15

Medidas de Posição (Tendência Central)

A tabela de frequência e os gráficos vistos na aula anterior permitem transformar em informação os dados extraídos da base original.

No entanto, há situações em que precisamos resumir ainda mais os dados, utilizando um ou mais valores que representem a informação contida em toda a série de dados.

Se for necessário utilizar apenas um valor para resumir toda a série de dados, é usual empregarmos uma das seguintes medidas de tendência central:

- Média;
- Mediana;
- Moda.

Notação Matemática

Antes de definirmos as medidas, é preciso que o estudante compreenda a seguinte notação matemática:

- X : variável aleatória (Maiúsculo);
- x_i : o valor da variável X **observada** no indivíduo i ;
- i varia de 1 até n , sendo n o tamanho da amostra;
- Média amostral: \bar{x} .
- Média populacional: μ .

Média amostral

- A média amostral é calculada da seguinte forma:

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

- Exemplo: Calcule a média amostral do desempenho na disciplina Estatística e Probabilidade utilizando os dados mostrados no slide 2 (Notas_EST).

$$\begin{aligned}\bar{x} &= \frac{1}{15} \times (2,79 + 4,3 + 4,46 + 7,64 + 7,7 + 2,09 + 4,94 + \\ &\quad 5,78 + 8,33 + 7,45 + 5,28 + 10 + 7,8 + 5,56 + 4,15) \\ &= \frac{88,27}{15} = 5,88\end{aligned}$$

Média amostral

A média amostral também pode ser calculada para dados organizados em tabelas de frequência.

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^k n_i x_i = \sum_{i=1}^k f_i x_i$$

Idade	Frequência (n_i)	Proporção (f_i)	Percentual ($100 \times f_i$)
17	1	0,0667	6,67%
18	1	0,0667	6,67%
20	2	0,1333	13,33%
21	5	0,3333	33,33%
22	3	0,2	20%
23	2	0,1333	13,33%
24	1	0,0667	6,67%
Total	15	1	100%

$$\bar{x} = \frac{1}{15} \times (17 \times 1 + 18 \times 1 + 20 \times 2 + \dots + 24 \times 1) = 21,0667$$

Média amostral (aproximação)

O cálculo da média amostral para uma variável contínua, fornecida por meio de uma tabela de frequência, exige que sejam feitas aproximações, uma vez que não dispomos das observações da amostra.

Nota ENEM	Frequência (n_i)	Proporção (f_i)	Ponto Médio
300 ┤ 400	1	0,0667	350
400 ┤ 500	1	0,0667	450
500 ┤ 600	5	0,3333	550
600 ┤ 700	4	0,2667	650
700 ┤ 800	4	0,2667	750
Total	15	1	-

$$\bar{x} \approx \frac{1}{15} \times (350 \times 1 + 450 \times 1 + 550 \times 5 + 650 \times 4 + 750 \times 4) \approx 610$$

Mediana

A mediana (md) é o valor que ocupa a posição central dos dados ordenados.

- Pelo menos 50% dos valores são menores ou iguais à mediana;
- Pelo menos 50% dos valores são maiores ou iguais à mediana.

Para calcular a mediana é necessário ordenar os dados em ordem crescente. Seja:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

em que $x_{(1)}$ representa o menor valor observado, $x_{(2)}$ o segundo menor valor e assim sucessivamente.

- A mediana da variável X pode então ser definida como:

$$md(X) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar;} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Mediana

- Exemplo: Calcule a mediana do desempenho na disciplina Estatística e Probabilidade utilizando os dados mostrados no slide 2 (Notas_EST).

Dados: {2,79, 4,3, 4,46, 7,64, 7,7, 2,09, 4,94, 5,78, 8,33, 7,45, 5,28, 10, 7,8, 5,56, 4,15}

- Passo 1: ordenar os dados

{2,09, 2,79, 4,15, 4,3, 4,46, 4,94, 5,28, 5,56, 5,78, 7,45, 7,64, 7,7, 7,8, 8,33, 10}

- Passo 2: Com $n = 15$, temos a situação em que n é ímpar. Logo $md(X) = X_{(\frac{15+1}{2})} = X_{(8)} = 5,56$.

Mediana

- Vamos refazer o exemplo anterior considerando que temos apenas 14 observações e não 15.

Dados: $\{2,79, 4,3, 4,46, 7,64, 7,7, 2,09, 4,94, 5,78, 8,33, 7,45, 5,28, 10, 7,8, 5,56\}$

- Passo 1: ordenar os dados

$\{2,09, 2,79, 4,3, 4,46, 4,94, 5,28, 5,56, 5,78, 7,45, 7,64, 7,7, 7,8, 8,33, 10\}$

- Passo 2: Com $n = 14$, temos a situação em que n é par.

$$md(X) = \frac{X_{(\frac{14}{2})} + X_{(\frac{14}{2}+1)}}{2} = \frac{X_{(7)} + X_{(8)}}{2} = \frac{5,56 + 5,78}{2} = 5,67.$$

Moda

A moda (*mo*) é a realização mais frequente do conjunto de dados observados.

- Exemplo: Para a variável Idade, a moda é igual a 21 anos.

Idade	Frequência (n_i)	Proporção (f_i)	Percentual ($100 \times f_i$)
17	1	0,0667	6,67%
18	1	0,0667	6,67%
20	2	0,1333	13,33%
21	5	0,3333	33,33%
22	3	0,2	20%
23	2	0,1333	13,33%
24	1	0,0667	6,67%
Total	15	1	100%

Limitações para aplicação

- A moda pode ser calculada para variáveis **qualitativas** e **quantitativas**.
- A mediana pode ser calculada para variáveis **qualitativas ordinais** e para variáveis **quantitativas**.
- A média pode ser calculada apenas para variáveis **quantitativas**.