

Medidas de Posição e Medidas de Dispersão

O estudo sobre distribuição de frequência permite-nos descrever os grupos de valores que uma variável pode assumir. Dessa forma podemos localizar a *maior concentração* de valores de uma dada distribuição, isto é, se ela se localiza no início, no meio ou no final, ou ainda, se há uma distribuição por igual. Os conceitos que expressam através de números as tendências características de cada distribuição são denominados *elementos típicos da distribuição* e são as:

- a. Medidas de posição
- b. Medidas de variabilidade ou dispersão
- c. Medidas de assimetria
- d. Medidas de curtose

Dentre esses elementos típicos de uma distribuição de frequência daremos uma atenção especial para as medidas de posição e as medidas de variabilidade e dispersão.

Medidas de Posição

Temos as Medidas de Tendência Central e as Separatrizes. Veremos detalhadamente as medidas de tendência central:

- a. Média aritmética
- b. Mediana
- c. Moda

Média Aritmética (\bar{X}): É o quociente da divisão da soma dos valores da variável pelo número deles.

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

Onde:

\bar{X} é a média aritmética

x_i os valores da variável

n o número de valores

A média é utilizada quando desejamos obter a medida de posição de maior estabilidade.

1) Dados não-agrupados: Quando desejamos conhecer a média dos dados não-agrupados, determinamos a média aritmética simples.

Exemplo:

Sabendo-se que a produção leiteira diária da vaca A, durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 litros, temos a produção média da semana:

$$\bar{X} = \frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7} = \frac{98}{7} = 14,$$

A produção média da semana foi de 14 litros.

2) Dados agrupados sem intervalos de classe: Como as frequências são números indicadores da intensidade de cada valor da variável, elas funcionam como fatores de ponderação, o que nos leva a calcular a média aritmética ponderada, dada por

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i}$$

Exemplo:

Considere a distribuição relativa a 34 famílias de quatro filhos, tomando para variável o número de filhos do sexo masculino.

Nº DE MENINOS	f_i
0	2
1	6
2	10
3	12
4	4
	$\Sigma = 34$

O modo mais prático de obtenção da média ponderada é abrir, na tabela, uma coluna correspondente aos produtos $x_i f_i$

Nº DE MENINOS	f_i	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
	$\Sigma = 34$	$\Sigma = 78$

Logo:

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \frac{78}{34} = 2,29 \Rightarrow \bar{X} = 2,3$$

Sendo x uma variável discreta, como interpretar o resultado obtido, 2 meninos e 3 décimos de menino?

O valor médio 2,3 meninos sugere, neste caso, que o maior número de famílias tem 2 meninos e 2 meninas, sendo porém, a tendência geral de uma leve superioridade numérica em relação ao número de meninos.

Moda (Mo): Denominamos moda o valor que ocorre com maior frequência em uma série de valores. Por exemplo, o salário modal dos empregados de uma indústria é o salário mais comum, isto é o salário recebido pelo maior número de empregados dessa indústria. A moda é utilizada quando desejamos obter uma medida rápida e aproximada de posição.

1) Dados não-agrupados: Quando lidamos com valores não-agrupados basta procurar o valor que mais se repete.

Exemplos:

- a) A série de dados 7, 8, 10, 10, 10, 11, 12, 13, 15 tem moda igual a 10.
- b) A série de dados 3, 5, 8, 10, 12 13 não apresenta moda (amodal).
- c) A série de dados 2, 3, 4, 4, 4, 5, 6, 7, 7, 7, 8, 9 tem duas modas: 4 e 7 (bimodal).

2) Dados agrupados sem intervalos de classe: Para dados agrupados a moda é dada pelo valor da variável de maior frequência.

Exemplo:

Se considerarmos a tabela que contém o número de meninos por família com quatro filhos verificamos que a frequência máxima corresponde o valor da variável 3. Logo $Mo = 3$.

Mediana (Md): A mediana de um conjunto de valores ordenados segundo uma ordem de grandeza, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

1) Dados não-agrupados:

Exemplos:

a) Dada a série de valores 5, 13, 10, 2, 18, 15, 6, 16, 9 encontre a sua mediana.

Ordenação (crescente ou decrescente de valores)

2, 5, 6, 9, **10**, 13, 15, 16, 18

Tomamos o valor central, aquele que apresenta o mesmo número de elementos à direita e à esquerda, logo $Md = 10$.

b) Dada a série de valores 2, 6, 7, 10, 12, 13, 18, 21, encontre a sua mediana.

Como temos um número par de elementos calculamos o ponto médio entre os valores centrais. 2, 6, 7, **10, 12**, 13, 18, 21

$$\text{Logo } Md = \frac{10+12}{2} = \frac{22}{2} = 11$$

2) Dados agrupados sem intervalos de classe: Identificamos a frequência acumulada crescente imediatamente superior à metade da soma das frequências. A mediana será aquele valor que corresponde a tal frequência acumulada.

Exemplos:

a) Encontre a mediana para a distribuição:

Nº DE MENINOS	f_i	fac_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\Sigma = 34$	

$$P = \frac{\sum f_i}{2} = \frac{34}{2} = 17$$

A menor frequência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável, sendo este o valor mediano. Logo: $Md = 2$ meninos

Quando $\frac{\sum f_i}{2} = fac_i$ a mediana será dada por $Md = \frac{x_i + x_{i+1}}{2}$, isto é a mediana será dada pela média aritmética entre o valor da variável correspondente a essa frequência acumulada e o seguinte

b) Calcule a mediana para

x_i	f_i	fac_i
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8
	$\Sigma = 8$	

$$\text{Temos } \frac{8}{2} = 4 = fac_3.$$

$$\text{Logo } Md = \frac{15+16}{2} = \frac{31}{2} = 15,5$$

Medidas de dispersão ou variabilidade

A média não pode por si mesma, destacar o grau de homogeneidade ou heterogeneidade que existe entre os valores que compõem o conjunto.

Consideremos os conjuntos e suas médias:

$$X: 70, 70, 70, 70, 70 \text{ com } \bar{X} = \frac{350}{5} = 70$$

$$Y: 68, 69, 70, 71, 72 \text{ com } \bar{Y} = \frac{350}{5} = 70$$

$$Z: 5, 15, 50, 120, 160 \text{ com } \bar{Z} = \frac{350}{5} = 70$$

Os três conjuntos apresentam média igual a 70. É fácil notar que o conjunto X é mais homogêneo que os conjuntos Y e Z. O conjunto Y por sua vez é mais homogêneo que o conjunto Z.

Dispersão ou variabilidade *é a maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central tomado como ponto de comparação*. Para qualificar os valores de uma variável com maior ou menor dispersão entre seus valores e os de sua medida de posição, a Estatística recorre às medidas de dispersão ou variabilidade.

Variância e Desvio Padrão

A variância e o desvio padrão levam em consideração a totalidade dos valores da variável em estudo, o que faz delas índices de variabilidade bastante estáveis e os mais empregados.

A *variância* se baseia nos desvios em torno da média aritmética, determinando a média

aritmética dos quadrados dos desvios, representada por $s^2 = \frac{\sum (x_i - \bar{X})^2}{n}$.

Sendo a variância calculada a partir dos quadrados dos desvios, ela é um número em unidade quadrada em relação à variável em questão, por isso imaginou-se uma nova medida denominada desvio padrão, definida como a raiz quadrada da variância e

representada por $s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n}}$

A fórmula dada para o cálculo do desvio é de fácil compreensão mas não é a mais ideal para o uso pois em geral a média aritmética é um número fracionário o que torna pouco prático o cálculo das quantidades $(x_i - \bar{X})^2$. Essa fórmula também não é muito precisa, pois quando a média não é exata tem que ser arredondada, cada desvio fica afetado ligeiramente do erro, devido a esse arredondamento.

Logo podemos escrever a fórmula do seguinte modo:

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

1) Dados não-agrupados

Exemplo:

Calcule o desvio padrão para o conjunto de valores 40, 45, 48, 52, 54, 62, 70.

A maneira mais prática para se obter o desvio padrão é formar uma tabela com duas colunas, uma para x_i e outra para x_i^2 .

x_i	x_i^2
40	1600
45	2025
48	2304
52	2704
54	2916
62	3844
70	4900
$\Sigma = 371$	$\Sigma = 20293$

$$s = \sqrt{\frac{20293}{7} - \left(\frac{371}{7}\right)^2} = \sqrt{2899 - 53^2} = \sqrt{2899 - 2809} = \sqrt{90} = 9,486$$

Logo, $s = 9,49$

2) Dados agrupados sem intervalos de classe:

Como, neste caso, temos a presença de frequências, devemos levá-las em consideração

$$s = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i}\right)^2}$$

Exemplo:

Considere a tabela de valores seguinte

x_i	f_i
0	2
1	6
2	12
3	7
4	3
	$\Sigma = 30$

Abriremos uma coluna para x_i^2 outra para os produtos $f_i \cdot x_i$ e outra para $f_i \cdot x_i^2$, logo:

x_i	f_i	x_i^2	$f_i x_i$	$f_i x_i^2$
0	2	0	0	0
1	6	1	6	6
2	12	4	24	48
3	7	9	21	63
4	3	16	12	48
	$\Sigma = 30$		$\Sigma = 63$	$\Sigma = 165$

Então

$$s = \sqrt{\frac{165}{30} - \left(\frac{63}{30}\right)^2} = \sqrt{5,5 - 4,41} = \sqrt{1,09} = 1,044, \text{ daí } s = 1,04$$