Medidas-Resumo

Medidas de Posição

- **A *moda*** é definida como a realização mais frequente do conjunto de valores observados.

Tabela 2.5 Frequências e porcentagens dos empregados da seção de orçamentos da Companhia MB, segundo o número de filhos.

Nº de filhos z_i	Frequência n_i	Porcentagem $100 f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Fonte: Tabela 2.1.

Moda =2

- **A mediana é a realização que ocupa a posição central da série de observações, quando estão ordenadas em ordem crescente.**

Exemplo:

i) 3, 4, 7, 8 e 8, a mediana = 7;

ii) 3, 4, 7, 8, 8 e 9, a mediana = (7+8)/2 = 7,5.

- **A média aritmética, é a soma das observações dividida pelo número delas. Assim, a média aritmética de 3, 4, 7, 8 e 8 é**

  (3 + 4 + 7 + 8 + 8)/5 = 6.

Exemplo 3.1 Usando os dados da Tabela 2.5, já encontramos que a moda da variável *Z* é 2. Para a mediana, constatamos que esta também é 2, média aritmética entre a décima e a décima primeira observações.

Tabela 2.5 Frequências e porcentagens dos empregados da seção de orçamentos da Companhia MB, segundo o número de filhos.

Nº de filhos z_i	Frequência n_i	Porcentagem $100 f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Fonte: Tabela 2.1.

A média aritmética será

$$\frac{4 \times 0 + 5 \times 1 + 7 \times 2 + 3 \times 3 + 5 \times 1}{20} = \frac{33}{20} = 1,65.$$

Vamos formalizar os conceitos introduzidos acima. Se $x_1, ..., x_n$ são os n valores (distintos ou não) da variável X, a média aritmética, ou simplesmente média, de X pode ser escrita

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.1}$$

Agora, se tivermos n observações da variável X, das quais n_1 são iguais a x_1, n_2 são iguais a x_2 etc., n_k iguais a x_k, então a média de X pode ser escrita

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i. \tag{3.2}$$

Se $f_i = n_i/n$ representar a frequência relativa da observação x_i, então (3.2) também pode ser escrita

$$\overline{x} = \sum_{i=1}^{k} f_i x_i. \tag{3.3}$$

Consideremos, agora, as observações ordenadas em ordem crescente. Vamos denotar a menor observação por $x_{(1)}$, a segunda por $x_{(2)}$, e assim por diante, obtendo-se

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n-1)} \leq x_{(n)}. \tag{3.4}$$

Por exemplo, se $x_1 = 3$, $x_2 = -2$, $x_3 = 6$, $x_4 = 1$, $x_5 = 3$, então $-2 \leq 1 \leq 3 \leq 3 \leq 6$, de modo que $x_{(1)} = -2$, $x_{(2)} = 1$, $x_{(3)} = 3$, $x_{(4)} = 3$ e $x_{(5)} = 6$.

As observações ordenadas como em (3.4) são chamadas *estatísticas de ordem*.

Com essa notação, a mediana da variável X pode ser definida como

$$\text{md}(X) = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ ímpar;} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ par.} \end{cases} \tag{3.5}$$

Exemplo:

Tabela 2.6 Distribuição de frequências da variável S, salário dos empregados da seção de orçamentos da Companhia MB.

Classes de salários	Ponto médio s_i	Frequência n_i	Porcentagem $100 f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	–	36	100,00

Fonte: Tabela 2.4.

A moda, mediana e média para os dados da Tabela 2.6 são, respectivamente,

$$\mathrm{mo}(\mathrm{S}) \simeq 10{,}00,$$

$$\mathrm{md}(\mathrm{S}) \simeq 10{,}00,$$

$$\overline{s} \simeq \frac{10 \times 6{,}00 + 12 \times 10{,}00 + 8 \times 14{,}00 + 5 \times 18{,}00 + 1 \times 22{,}00}{36} = 11{,}22.$$

Medidas de Dispersão: um valor que resuma a variabilidade de um conjunto de dados.

Por exemplo, suponhamos que cinco grupos de alunos submeteram-se a um teste, no qual obtiveram as seguintes notas:

Grupo A (variável X):	3, 4, 5, 6, 7.
Grupo B (variável Y):	1, 3, 5, 7, 9.
Grupo C (variável Z):	5, 5, 5, 5, 5.
Grupo D (variável W):	3, 5, 5, 7.
Grupo E (variável V):	3, 5, 5, 6, 6.

$$\overline{x} = \overline{y} = \overline{z} = \overline{w} = \overline{v} = 5,0$$

Duas medidas são as mais usadas: **desvio médio e variância.**

O princípio básico é analisar os desvios das observações em relação à média dessas observações.

Para o grupo A teríamos, respectivamente,

$$\sum_{i=1}^{5} |x_i - \bar{x}| = 2 + 1 + 0 + 1 + 2 = 6,$$

$$\sum_{i=1}^{5} (x_i - \bar{x})^2 = 4 + 1 + 0 + 1 + 4 = 10.$$

O uso desses totais pode causar dificuldades quando comparamos conjuntos de dados com números diferentes de observações, como os conjuntos A e D acima. Desse modo, é mais conveniente exprimir as medidas como médias, isto é, o desvio médio e a variância são definidos por

$$\mathrm{dm}(X) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}, \tag{3.6}$$

$$\mathrm{var}(X) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}, \tag{3.7}$$

respectivamente. Para o grupo A temos

$$\mathrm{dm}(X) = 6/5 = 1,2,$$
$$\mathrm{var}(X) = 10/5 = 2,0,$$

enquanto para o grupo D temos

$$\mathrm{dm}(W) = 4/4 = 1,0,$$
$$\mathrm{var}(W) = 8/4 = 2,0.$$

Sendo a variância uma medida de dimensão igual ao quadrado da dimensão dos dados (por exemplo, se os dados são expressos em cm, a variância será expressa em cm^2), pode causar problemas de interpretação. Costuma-se usar, então, o desvio padrão, que é definido como a raiz quadrada positiva da variância.

Ambas as medidas de dispersão (dm e dp) indicam, em média, qual será o “erro”(desvio) cometido ao tentar substituir cada observação pela medida resumo do conjunto de dados (no caso, a média).

Exemplo 3.3 Calcular as medidas de dispersão acima para a variável Z = número de filhos, resumida na Tabela 2.5.

Tabela 2.5 Frequências e porcentagens dos empregados da seção de orçamentos da Companhia MB, segundo o número de filhos.

Nº de filhos z_i	Frequência n_i	Porcentagem $100\,f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Fonte: Tabela 2.1.

A média aritmética será

$$\frac{4\times 0+5\times 1+7\times 2+3\times 3+5\times 1}{20}=\frac{33}{20}=1,65.$$

Os desvios são $z_i-\overline{z}$: –1,65; –0,65; 0,35; 1,35; 3,35. Segue-se que

$$\mathrm{dm}(Z)=\frac{4\times(1,65)+5\times(0,65)+7\times(0,35)+3\times(1,35)+1\times(3,35)}{20}=0,98.$$

Também,

$$\mathrm{var}(Z)=\frac{4(-1,65)^2+5(-0,65)^2+7(0,35)^2+3(1,35)^2+1(3,35)^2}{20}=1,528.$$

Consequentemente, o desvio padrão de Z é

$$\mathrm{dp}(Z)=\sqrt{1,528}=1,24.$$

Suponha que observemos n_1 vezes os valores x_1 etc., n_k vezes o valor x_k da variável X. Então,

$$\mathrm{dm}(X) = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}|}{n} = \sum_{i=1}^{k} f_i |x_i - \bar{x}|, \tag{3.8}$$

$$\mathrm{var}(X) = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{n} = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2, \tag{3.9}$$

$$\mathrm{dp}(X) = \sqrt{\mathrm{var}(X)}. \tag{3.10}$$

Exemplo 3.4 Consideremos a variável S = salário.

Tabela 2.6 Distribuição de frequências da variável S, salário dos empregados da seção de orçamentos da Companhia MB.

Classes de salários	Ponto médio s_i	Frequência n_i	Porcentagem $100 f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

Fonte: Tabela 2.4.

$$\bar{s} \simeq \frac{10 \times 6,00 + 12 \times 10,00 + 8 \times 14,00 + 5 \times 18,00 + 1 \times 22,00}{36} = 11,22.$$

$$\mathrm{var}(S) \simeq \Big[10(6,00 - 11,22)^2 + 12(10,00 - 11,22)^2 + 8(14 - 11,22)^2$$
$$+5(18,00 - 11,22)^2 + 1(22,00 - 11,22)^2\Big]\Big/36 = 19,40$$

e, portanto,

$$\mathrm{dp}(S) \simeq \sqrt{19,40} = 4,40.$$

É fácil ver que $\mathrm{dm}(S) \simeq 3,72$.

Uma maneira computacionalmente mais eficiente de calcular a variância é

$$\operatorname{var}(X) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2, \tag{3.11}$$

e, no caso de observações repetidas,

$$\operatorname{var}(X) = \sum_{i=1}^{k} f_i x_i^2 - \overline{x}^2. \tag{3.12}$$