Medidas de Dispersão ou de Variabilidade

ML205 - Estatística I

João Luis Gomes Moreira

DMAT/CCT/UFRR

25 de fevereiro de 2021

Conteúdo

Medidas de Dispersão

Medidas de Dispersão, Variação ou Variabilidade, servem para verificarmos a representatividade das medidas de posição.

Sejam as séries:

a) 20, 20, 20, 20, 20

b) 15, 10, 20, 25, 30

Temos para ambas as séries que: $\bar{x}_a = \bar{x}_b = 20$

Contudo, verificamos que na série 'a' os valores se concentram totalmente na média 20, enquanto que os valores na série 'b' se dispersam em torno da média.

Amplitude Total

A amplitude de um conjunto de dados é a distância entre o maior valor e o menor valor.

$$R = \Delta_{Total} = X_{max} - X_{min}$$

A amplitude total é muito limitada como medida de dispersão, por depender apenas dos valores extremos e não ser afetada pela dispersão dos valores internos.

A amplitude total tem a mesma unidade dos dados.

Intervalo Semi-Interquartílico.

É utilizado para verificar a dispersão em torno da mediana.

$$I_Q = \frac{Q_3 - Q_1}{2}$$

Intervalo Semi-Interquartílico

Exemplo: Para as notas dos alunos em duas disciplinas, foram calculadas as seguintes medidas :

$$\text{Estatística}: \; Q_1 = 3,0, \; Q_3 = 6,5 \text{ e } \tilde{x} = 5$$
$$\text{Matemática}: \; Q_1 = 2,0, \; Q_3 = 7,0 \text{ e } \tilde{x} = 5$$

As duas distribuições apresentaram a mesma mediana,

$$I_{Q\,E} = \frac{6,5 - 3,0}{2} = 1,75 \qquad I_{Q\,M} = \frac{7,0 - 2,0}{2} = 2,5$$

Contudo as notas de Estatísticas apresentaram menor dispersão.

Desvio Médio

O desvio médio (D_M), ou desvio médio absoluto, de um conjunto de dados $x_1, x_2, \cdots, x_n$ é definido por:

$$D_M = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|.F_i}{n}$$

Onde : $d_i = x_i - \bar{x}$

O desvio médio tem a mesma unidade dos dados.

Variância Populacional

A variância de uma população com um conjunto de dados $x_1, x_2, \cdots, x_n$ é definida por:

$$\sigma^2 = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2.F_i}{n}$$

A variância não tem unidade.

Variância Amostral

A variância de uma amostra definida por $x_1, x_2, \cdots, x_n$ é definida por:

$$s^2 = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2.F_i}{n-1}$$

Ao utilizarmos a média amostral como estimador da variância amostral, perdemos 1 grau de liberdade em relação à variância populacional.

Novamente, a variância não tem unidade.

Desvio Padrão Populacional

O desvio padrão de uma população com um conjunto de dados $x_1, x_2, \cdots, x_n$ é definido por:

$$\sigma = \sqrt{\text{Variância}} = \sqrt{\sigma^2}$$

O desvio padrão tem a mesma unidade dos dados.

Desvio Padrão Amostral

O desvio padrão de uma amostra definida por $x_1, x_2, \cdots, x_n$ é definido por:

$$s = \sqrt{\text{Variância}} = \sqrt{s^2}$$

O desvio padrão tem a mesma unidade dos dados.

Propriedades das Medidas de Dispersão

Propriedade 1: Todas as medidas de dispersão são não negativas :

$$\Delta \geq 0,\ D_M \geq 0,\ \sigma^2 \geq 0 \text{ e } \sigma \geq 0$$

Propriedade 2: Somando-se uma constante a todas as observações, as medidas de dispersão não se alteram :

$$y_i = x_i + k \Rightarrow \Delta_y = \Delta_x,\ D_{My} = D_{Mx},\ \sigma_y^2 = \sigma_x^2 \text{ e } \sigma_y = \sigma_x$$

Propriedade 3: Ao multiplicarmos todos os dados por uma constante não nula, temos que:

$$y_i = k.x_i \Rightarrow \Delta_y = |k|.\Delta_x,\ D_{My} = |k|.D_{Mx},$$
$$\sigma_y^2 = |k|.\sigma_x^2 \ \text{ e } \ \sigma_y = |k|.\sigma_x$$

Coeficiente de Variação.

Medida relativa de dispersão. O CV é útil para a comparação, em termos relativos do grau de concentração em torno da média, de séries distintas, inclusive quando expressas em unidades diferentes.

É dado por :

$$CV = \frac{\sigma}{\bar{x}} \qquad \text{ou} \qquad CV = \frac{s}{\bar{x}}$$

Costuma-se considerar que CV superior a 50% indica alto grau de dispersão e, consequentemente, pequena representatividade da média.

Enquanto que para valores inferiores a 50%, a média será tanto mais representativa do fato quanto menor for o valor de seu CV.

Coeficiente de Variação

Exemplo: Numa empresa, o salário médio dos homens é de $R\$4.000,00$, com desvio padrão de $R\$1.500,00$, e o das mulheres é em média $R\$3.000,00$, com desvio padrão de $R\$1.200,00$. Então

$$\begin{aligned} \text{Para os homens} &: CV = \frac{1500}{4000} = 0,375 = 37,5\% \\ \text{Para as mulheres} &: CV = \frac{1200}{3000} = 0,400 = 40,0\% \end{aligned}$$

Conclui-se que os salários das mulheres apresentam maior dispersão relativa que os dos homens.

Escores Reduzidos..

Uma importante utilização do desvio padrão. E mede o afastamento de um valor x_i, em relação a média da distribuição, em unidades do desvio padrão.

É dado por :

$$z_i = \frac{x_i - \bar{x}}{\sigma} \qquad \text{ou} \qquad z_i = \frac{x_i - \bar{x}}{s}$$

Escores Reduzidos.

Exemplo: Um estudante obteve nota $7,5$ em Matemática e $8,0$ em Economia. Em que disciplina sua posição relativa foi melhor, sabendo-se que a média da classe de Matemática foi $6,85$ com desvio padrão de $1,1$, e em Economia foi $7,75$ com desvio padrão de $1,6$?

$$\text{em Matemática} : z = \frac{7,5 - 6,85}{1,1} = 0,59 = 59,0\%$$

$$\text{em Economia} : z = \frac{8,0 - 7,75}{1,6} = 0,16 = 16,0\%$$

Escores Reduzidos

Do resultado

$$z_M = 0,59 = 59,0\% \qquad \text{e} \qquad z_E = 0,16 = 16,0\%$$

Concluimos que : A posição relativa do estudante foi melhor em Matemática do que em Economia, embora tenha tido melhor nota em Economia.

Para valores de z_i negativos, quanto mais próxima de zero melhor a posição relativa.

Considerações.

- A amplitude total, por considerar apenas os dois valores extremos da série, serve apenas como indicação aproximada da dispersão. Um uso prático da amplitude total se dá na determinação da amplitude da temperatura em um dia ou no ano, no controle de qualidade ou como uma medida de cálculo rápido, e quando a compreensão popular é mais importante que a exatidão e a estabilidade.
- A amplitude total, por se deixar influenciar pelos valores extremos, que são na maioria das vezes devidos ao acaso, é dita uma medida instável.

Considerações

- A variância e o desvio padrão, por considerarem a totalidade dos valores da variável em estudo, são chamadas de medidas estáveis.
- A variância é uma medida que tem pouca utilidade como estatística descritiva, porém é extremamente importante na inferência estatística e em combinações de amostra.

That's all, Folks !!!