Capítulo 7

Medidas de Dispersão ou de Variabilidade

- Apresentar as principais medidas de dispersão ou de variabilidade.
- Conceituar as medidas de assimetria e de curtose.

1. Dispersão ou variabilidade

Sabemos que as medidas de posição[23] apresentam apenas uma das características dos valores numéricos de um conjunto de observações, o da tendência central. Nenhuma delas informa sobre o grau de variação ou de dispersão dos valores observados. Em qualquer grupo de dados, os valores numéricos não são semelhantes e apresentam desvios variáveis em relação à tendência geral de média.

[23] Medidas de posição: média, mediana e moda.

As medidas de dispersão ou de variabilidade têm como objetivo avaliar o quanto estão dispersos os valores de uma distribuição de freqüência, ou seja, o grau de afastamento ou de concentração entre os valores.

A média que é considerada como um número que representa uma série de valores não pode, por si mesma, destacar o grau de homogeneidade ou de heterogeneidade que há entre eles.

Analisemos, por exemplo, os conjuntos de valores a seguir

- A = {60, 60, 60, 60, 60}
- B = {58, 59, 60, 61, 62}
- C = {5, 10, 40, 100, 145}

Ao calcularmos a mesma média aritmética desses conjuntos:

$$\bar{a} = \frac{\sum ai}{n} = \frac{300}{5} = 60$$

$$\bar{b} = \frac{\sum bi}{n} = \frac{300}{5} = 60$$

$$\bar{c} = \frac{\sum ci}{n} = \frac{300}{5} = 60$$

Verificamos que os três conjuntos apresentam a mesma média aritmética: 60.

Chegamos à conclusão que o conjunto A é mais homogêneo que os conjuntos B e C, visto que todos os valores são iguais à média. O conjunto B, por sua vez, é mais homogêneo que o conjunto C, pois há menor diversificação entre cada um de seus valores e a média representativa. Logo, o conjunto A apresenta dispersão nula, e o conjunto B apresenta uma dispersão menor que C.

As principais medidas de dispersão são

1. Amplitude total.
2. Desvio médio absoluto.
3. Variância.
4. Desvio padrão.
5. Coeficiente de Variação

1.1 Amplitude total

A amplitude total em dados não agrupados é a diferença entre o maior e o menor valor da série de dados, ou seja,

$AT = X_{máximo} - X_{minimo}$

Considerando os valores 30, 45, 48, 62 e 72 teremos

AT = 72 – 30 = 42

A amplitude total tem o inconveniente de só levar em conta os dois valores extremos da série, descuidando do conjunto de valores intermediários, o que quase sempre invalida a idoneidade do resultado. Ela é apenas uma indicação aproximada da dispersão ou da variabilidade.

No caso em que os dados estejam agrupados sem intervalos de classe, ainda teremos

$$AT = X_{máximo} - X_{minimo}$$

Vejamos o exemplo a seguir.

Tabela 1

x_i	f_i
0	2
1	6
3	5
4	3

Dados Hipotéticos

Calculando a amplitude,

$$AT = X_{máximo} - X_{minimo} = 4 - 0 = 4$$

Com intervalos de classe a amplitude total é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Logo,

AT = L máximo - L mínimo.

Tabela 2

Classes	Freqüência
150 ├ 154	4
154 ├ 158	9
158 ├ 162	11
162 ├ 166	8
166 ├ 170	5
170 ├ 174	3

Dados Hipotéticos

Então,

$$AT = 174 - 150 = 24$$

Quanto maior a amplitude total, maior a dispersão ou variabilidade dos valores da variável.

A amplitude total é muito utilizada quando se deseja determinar a amplitude da temperatura em um dia ou no ano, no controle de qualidade ou uma medida de cálculo rápido sem muita exatidão.

1.2 Desvio médio absoluto

Como a amplitude total não leva em consideração todos os valores da série de dados, é preferível se trabalhar com medidas que utilizam toda a informação disponível do conjunto de dados. Uma dessas medidas é o desvio médio, que é a média aritmética dos desvios absolutos dos elementos da série, tomados em relação à sua média aritmética, que é representada por DMA e calculada pela fórmula a seguir.

[24] As barras verticais indicam que são tomados os valores absolutos.

$$DMA = \frac{\sum | xi - \overline{x}}{n}$$

Considerando a série de dados 1, 2, 3, 4, 5, calculamos o desvio médio absoluto, ou simplesmente desvio médio.

Calculando inicialmente a média, teremos

$$x = \frac{(1+2+3+4+5)}{5} = \frac{15}{5} = 3$$

Calculando o numerador da nossa fórmula,

$$\sum |x_i - \bar{x}| = |1-3| + |2-3| + |3-3| + |4-3| + |5-3| = 2+1+0+1+2 = 6$$

Logo,

$$DMA = \frac{6}{5} = 1{,}2$$

1.3 Variância

A variância mede a dispersão dos dados em torno de sua média, levando em consideração a totalidade dos valores da variável em estudo, o que a torna um índice de variabilidade bastante estável. A variância é representada por s^2 e definida como sendo a média dos quadrados dos desvios em relação à média aritmética.

$$s^2 = \frac{\sum (xi - \bar{x})^2}{n}$$

Como exemplo, consideremos que foi aplicado um teste a dois grupos com cinco alunos cada, o grupo A obteve os seguintes pontos 6, 8, 7, 4, 10 e o grupo B 9, 7, 8, 5, 6. Utilizando a variância vamos determinar o grupo mais regular.

Cálculo da variância do grupo A,

$$\bar{x} = \frac{(6+8+7+4+10)}{5} = \frac{35}{5} = 7$$

$$s^2 = \frac{(6-7)^2 + (8-7)^2 + (7-7)^2 + (4-7)^2 + (10-7)^2}{5} =$$

$$\frac{(-1)^2 + (1)^2 + (0)^2 + (-3)^2 + (3)^2}{5} = 4$$

Cálculo da variância do grupo B:

$$\bar{x} = \frac{(9+7+8+5+6)}{5} = \frac{35}{5} = 7$$

$$s^2 = \frac{+(7-7)^2 + (8-7)^2 + (5-7)^2 + (6-7)^2}{5} =$$

$$\frac{(2)^2 + (0)^2 + (1)^2 + (-2)^2 + (-1)^2}{5} = \frac{10}{5} = 2$$

Assim, concluímos que o grupo B demonstra maior regularidade, tendo em vista que a variância foi menor do que a do grupo A. Isso quer dizer que seus valores estão mais próximos da média do grupo.

• Fórmula alternativa para o cálculo da variância

Quando a média é um valor decimal não é exato, a fórmula da variância apresentada anteriormente não é muito prática, uma vez que entrará no cálculo "n" vezes aumentando os erros de arredondamento que ocorrem. Então, é melhor utilizar uma expressão alternativa que é obtida através de algumas manipulações algébricas na fórmula anterior.

$$s^2 = \frac{1}{n-1}[[(\textstyle\sum_{i=1}^{n} xi^2) - n(x)^2]$$

1.4 Desvio padrão

É a medida de dispersão geralmente mais empregada, pois leva em consideração a totalidade dos valores da variável em estudo. O desvio padrão é uma medida de dispersão usada com a média. Mede a variabilidade dos valores à volta da média. O valor mínimo do desvio padrão é 0, indicando que não há variabilidade, ou seja, que todos os valores são iguais à média. O símbolo para o desvio padrão em um conjunto de dados observados é s, e a fórmula para obter o desvio padrão é a seguinte.

$$s = \sqrt{\frac{\sum(xi - \bar{x})^2}{2}}$$

A fórmula anterior é utilizada quando estamos trabalhando com uma distribuição de dados não agrupados.

Considerando os valores 30, 45, 48, 62 e 72, vamos construir a tabela a seguir para facilitar o cálculo do desvio padrão[25]:

[25] O desvio padrão é a raiz quadrada da variância.

Tabela 3

X_i	$\bar{X}_i$	$(X_i - \bar{X}_i)$	$(X_i - \bar{X}_i)^2$
30	51,4	-21,4	457,96
45	51,4	-6,4	40,96
48	51,4	-3,4	11,56
62	51,4	10,6	112,36
72	51,4	20,6	424,36
Total	-	-	1047,2

Como n = 5, temos

$$s = \sqrt{\frac{1047,2}{5}} = 14,47$$

Propriedades do desvio padrão

- Ao adicionarmos ou subtrairmos uma constante a todos os valores de uma variável, o desvio padrão[26] não se altera.
- Ao multiplicarmos ou dividirmos todos os valores de uma variável por uma constante (diferente de zero), o desvio padrão fica multiplicado (ou dividido) por essa constante.

[26] Ao aplicar a fórmula do cálculo do desvio padrão para dados agrupados com intervalo de classe, lembre-se de que o valor de xi é o ponto médio de cada classe.

Quando os dados estão agrupados, a fórmula do desvio padrão será

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

Consideremos como exemplo a distribuição da tabela a seguir:

Tabela 4

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	2	0	0
1	6	6	6
3	5	15	45
4	3	12	48
	∑ = 16	**∑ = 33**	**∑ = 99**

Logo,

$$s = \sqrt{\frac{99}{16} - \left(\frac{33}{16}\right)^2} = \sqrt{6,18 - 2,06} = 2,03$$

1.5 Coeficiente de variação

O desvio padrão é uma medida limitada se utilizada isoladamente. Por exemplo, um desvio padrão de 2 unidades pode ser considerado pequeno para uma série de valores cujo valor médio é 200; no entanto, se a média for igual a 20, o mesmo não pode ser dito.

Quando desejamos comparar duas ou mais unidades relativamente à sua dispersão ou variabilidade, o desvio padrão não é a medida mais indicada, visto que ele se encontra na mesma unidade dos dados.

Assim, contornamos este problema caracterizando a dispersão em termos relativos ao seu valor médio. Essa medida é denominada de coeficiente de variação de Pearson.

O coeficiente de variação de pearson é a razão entre o desvio padrão e a média referentes aos dados de uma mesma série.

$$CV = \frac{s}{x}$$

Consideremos o exemplo da tabela a seguir.

Tabela 5

Discriminação	Média	Desvio padrão
Estaturas	165 cm	3,0 cm
Pesos	55 kg	2,5 kg

Calculando os coeficientes de variação, temos

$$CVe = \frac{3}{165} * 100 = 1,81\%$$

$$CVp = \frac{2,5}{55} * 100 = 4,54\%$$

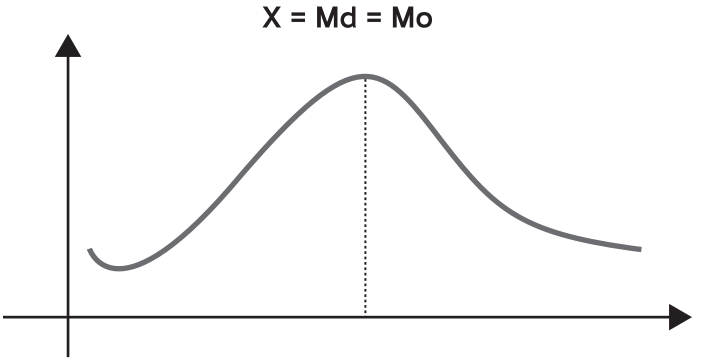
Logo, nesse grupo de indivíduos, os pesos apresentaram maior grau de dispersão em relação às estaturas.

Para Refletir

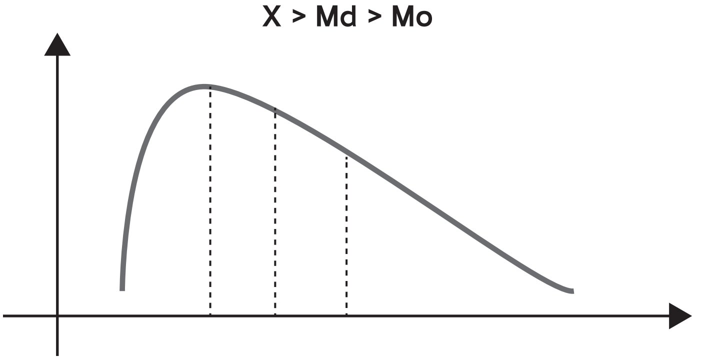
1. Considere o seguinte conjunto de dados: 2, 3, 5, 7, 10. Utilize a fórmula alternativa para calcular a variância, sabendo que a média é 5,4.
2. O desvio padrão e a variância podem ser negativos?
3. Em que situação o desvio padrão e a variância são nulos? Qual e a amplitude neste caso?
4. Calcule do desvio padrão da idade de 5 pessoas I = {10,13,24,47,50}
5. Sabendo que um conjunto de dados apresenta para média aritmética e para desvio padrão, respectivamente, 18,3 e 1,47, calcule o coeficiente de variação.
6. Um grupo de 100 estudantes tem uma estatura média de 163,8 cm, com um coeficiente de variação de 3,3%. Qual o desvio padrão desse grupo?

2. Assimetria

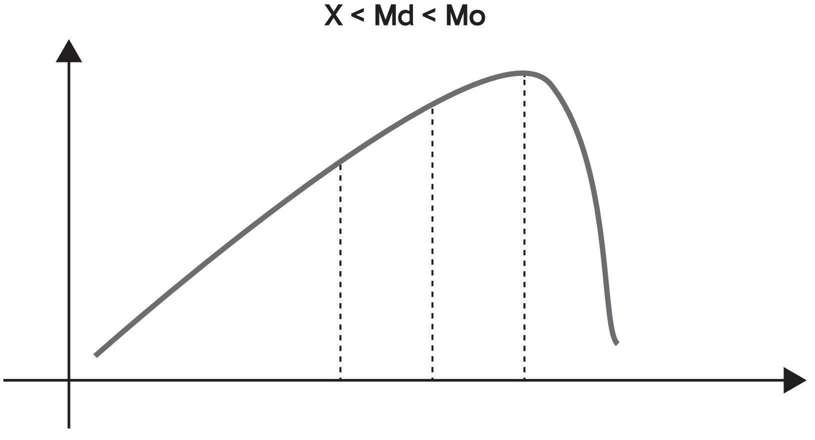
Uma distribuição de frequência pode ser caracterizada como simétrica ou assimétrica. Essa característica está associada à distribuição relativa da suas medidas de posição central: média, moda e mediana.



Distribuição de frequência simétrica



Distribuição de frequência assimétrica positiva



Distribuição de frequência assimétrica negativa

Baseando-se nas relações entre a média e a moda, podemos verificar o tipo de assimétria calculando o valor de suas diferenças.

$x - Mo$

Se

- $x - Mo = 0$ ⇨ assimetria nula ou distribuição simétrica;
- $x - Mo < 0$ ⇨ assimetria negativa ou à esquerda;
- $x - Mo > 0$ ⇨ assimetria positiva ou à direita.

2.1 Coeficiente de assimetria de Person - A_s

A medida da diferença entre a média e a moda, por ser absoluta, sofre das mesmas restrições do desvio padrão, isto é, não permite a possibilidade de comparação entre as medidas de duas distribuições. Portanto, deve ser dada preferência ao coeficiente de assimetria de Person.

$$As = \frac{3(\bar{x} - Md)}{s}$$

Se $0,15 < |As| < 1$, a assimetria é considerada **moderada.**

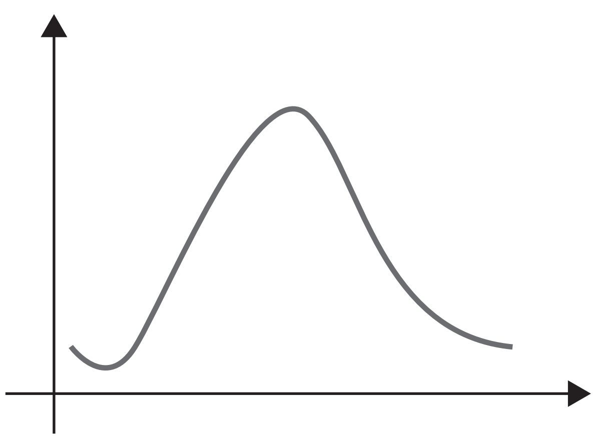
Se $|As| > 1$, a assimetria é considerada **forte.**

3. Curtose

A curtose exprime o grau de "achatamento" de uma distribuição de frequência em relação a uma distribuição padrão denominada curva normal[27].

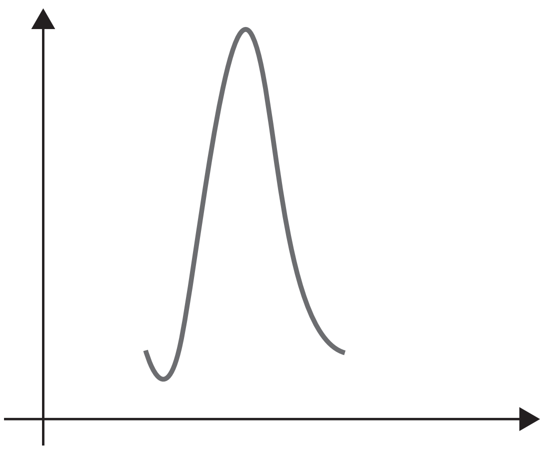
Essa característica está associada à concentração dos resultados. Assim, quanto à curtose, a distribuição de frequência pode ser: leptocúrtica, mesocúrtica e platicúrtica.

[27] Curva normal: curva correspondente a uma distribuição teórica de probabilidade.



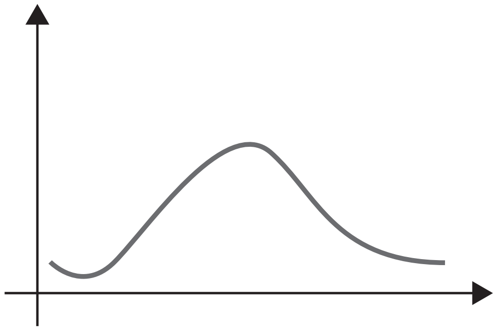
Distribuição de frequência mesocúrtica ou normal

A curva normal é a curva de base referencial recebe o nome de mesocúrtica.



Distribuição de frequência leptocúrtica

Quando a distribuição apresenta uma curva de frequência mais fechada que a normal ou mais aguda em sua parte superior, ela recebe o nome de **leptocúrtica**.



Distribuição de frequência platicúrtica

Quando a distribuição apresenta uma curva de frequência mais aberta que a normal ou mais achatada na sua parte superior ela recebe o nome de **platicúrtica.**

3.1. Coeficiente de Curtose – K

$$k = \frac{\frac{\sum (x_i - \overline{x})^4 * f_i}{n}}{s^4} - 3$$

- Se k = 0 ⇨ curva mesocúrtica.
- Se k > 3 ⇨ curva leptocúrtica.
- Se k < 3 ⇨ curva platicúrtica.

Referências

BARBETA, P. A., REIS, M. M.,BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**, São Paulo: Atlas, 2004.

BOLFARINE, Heleno & BUSSAB, Wilton O. **Elementos de Amostragem**. 1. ed. São Paulo: Edgard Blücher, 2005.

BUSSAB, Wilton de Ol., MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva, 2002.

COSTA, F. S. **Introdução Ilustrada à Estatística**. São Paulo: Harbra, 1998.

CRESPO, A. **Estatística Fácil**. Rio de Janeiro: Saraiva, 1996.

FONSECA, Jairo Simon da. **Introdução a Estatística**. Rio de Janeiro: LTC. 1993.

JAIRO, Simon da Fonseca, MARTINS, Gilberto de Andrade. **Curso de Estatística**, São Paulo: Atlas, 1996.

MARTINS G. A. **Estatística Geral e Aplicada**. São Paulo: Atlas, 2001.

MEYER, Paul L. **Aplicações a Estatística**. Rio de Janeiro: LTC. 1996.

MIRSHAWKA, V. **Probabilidade e Estatística para Engenharia**, São Paulo: Nobel, 1978.

MORETTIN, Pedro A. **Estatística Básica**. São Paulo: Saraiva. 2013.

PEREIRA, Wilson., TANAKA, Oswaldo K. **Estatística – Conceitos Básicos**, Makron Books, 1990.

ROSS, Sheldon. **A First Course in Probability**. 7. ed. New Jersey: Prentice Hall, 2005.

SPIEGEL M. R. **Estatística**. Coleção Schaum. São Paulo: Editora Afiliada, 1993.

TRIOLA M. F. **Introdução à Estatística e Probabilidade -** Exercícios Resolvidos e Propostos. Rio de Janeiro: LTC, 1999.

TRIOLA, Mario F. **Introdução a Estatística**. Rio de janeiro: LTC. 2005.

VIEIRA, S. **Estatística Experimental**. São Paulo: Atlas, 1999.