MEDIDAS DESCRITIVAS

MEDIDAS DE DISPERSÃO OU VARIABILIDADE

- Os métodos estatísticos dão especial ênfase ao estudo da variação, pois uma medida de posição, como a média por exemplo, por si só é insuficiente para descrever a característica de um conjunto de dados;
- As medidas de dispersão (ou medidas de variabilidade) são medidas que permitem quantificar a variação, ou ainda o grau de afastamento (espalhamento), dos dados em torno da média ou qualquer outro ponto especificado.

MEDIDAS DE DISPERSÃO

As medidas de dispersão mais conhecidas são:

i. Amplitude total
ii. Desvio médio
iii. Variância
iv. Desvio padrão
v. Coeficiente de Variação

De modo geral, considerando a variabilidade em torno de cada medida de posição, as medidas de dispersão servem para verificar a representatividade destas medidas.

MEDIDAS DE DISPERSÃO

Vamos ver cada uma das medidas de dispersão (para dados amostrais) citadas anteriormente.

Amplitude total (A)

É a diferença entre o maior valor e o menor valor do conjunto

$$A = L - l$$

É a mais simples das medidas de dispersão e de grande instabilidade, porque considera somente os valores extremos do conjunto.

DESVIO MÉDIO

Desvio Médio (Dm) para dados não agrupados

É a média aritmética dos valores absolutos dos desvios tomados em relação à média (ou à mediana).

- Considere $X = \{x_1, x_2, \ldots, x_n\}$ os dados amostrais e *n* o tamanho da amostra;
- Sendo d_i os desvios dos valores $x_1, x_2, \ldots, x_n$ em relação à média $\bar{X}$ então D_m é definido por

$$D_m = \frac{\sum_{i=1}^{n} |d_i|}{n} = \frac{\sum_{i=1}^{n} |x_i - \bar{X}|}{n}$$

MEDIDAS DE DISPERSÃO

Exemplo 1

Considere as notas de um determinado candidato A em 6 provas diferentes:

$$6,5 \quad 7,5 \quad 8,0 \quad 8,0 \quad 8,5 \quad 9,5$$

e do candidato B

$$7,5 \quad 8,0 \quad 8,0 \quad 8,0 \quad 8,0 \quad 8,5$$

Optaríamos por qual candidato?

DESVIO MÉDIO

Desvio médio para dados agrupados em tabelas de frequência

$$D_m = \frac{\sum_{i=1}^{k} |d_i F_i|}{n} = \frac{\sum_{i=1}^{k} |(x_i - \bar{X}) F_i|}{n}$$

onde *k* é o número de classes.

Desvio médio para dados agrupados em intervalos

$$D_m = \frac{\sum_{i=1}^{k} |d_i F_i|}{n} = \frac{\sum_{i=1}^{k} |(m_i - \bar{X}) F_i|}{n}$$

onde *k* é o número de classes.

VARIÂNCIA

Variância para dados não agrupados

A variância para dados não agrupados é definida por

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

Exemplo 2

Calcular a variância do peso de 10 alunos em kg

45,3	56,7	48,4	39,9	57,5
60,1	55,3	58,6	57,2	63,0

VARIÂNCIA

Variância para dados agrupados em tabelas de frequências

A variância para dados agrupados em tabelas de frequências é definida por

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{X})^2 F_i}{n-1}$$

onde k é o número de classes.

Exemplo 3

Calcule a variância para o exemplo ao lado.

Tabela: Distribuição de frequências no número de sementes germinadas por vaso

Nº de sementes/germinadas	Nº de vasos
0	2
1	5
2	15
3	8
Total	30

Fonte: Pesquisador

VARIÂNCIA

Variância para dados agrupados em intervalo

A variância para dados agrupados em tabelas de frequências é definida por

$$s^2 = \frac{\sum_{i=1}^{k}(m_i - \bar{X})^2 F_i}{n-1}$$

onde *k* é o número de classes.

Exemplo 4

Considere o tempo de duração de certas lâmpadas em horas apresentados na tabela a seguir:

Tabela: Distribuição da duração em horas de um conjunto de lâmpadas

Duração/horas	Nº de lâmpadas
100 ⊢ 200	60
200 ⊢ 300	25
300 ⊢ 400	10
400 ⊢ 500	8
500 ⊢ 600	2
Total	105

Fonte: Pesquisador

VARIÂNCIA

Calcule a variância desse conjunto de lâmpadas.

DESVIO PADRÃO

- Define-se Desvio Padrão (s) como sendo a raiz quadrada positiva da variância (s^2);
- s é uma das medidas mais úteis da variação, pois como ela é expressa na mesma unidade dos dados, sua interpretação é direta;

Observações

1) Por definição s e s^2 são não-negativos e podem assumir qualquer valor maior ou igual a zero. O valor “zero” é obtido quando todos os dados têm o mesmo valor, isto é, quando não há variação alguma entre os valores dos dados;
2) O desvio padrão tem as mesmas propriedades que a variância, porém a variância tem a dimensão quadrática dos dados.

DESVIO PADRÃO

Exemplo 5

Vamos calcular o desvio padrão para os exemplos 2, 3 e 4.

COEFICIENTE DE VARIAÇÃO

- O Coeficiente de Variação de PEARSON é uma medida relativa que determina o grau de concentração dos dados em torno da média;

- É definida por

$$CV = \frac{s}{\bar{X}}$$

no qual s é o desvio padrão e $\bar{X}$ é a média;

- Para valores do CV $\leq$ 50%, a média é representativa dos dados. Quanto mais próximo de 1 ou de 100%, maior a variabilidade dos dados. Neste caso, CV > 50%, a média não representa bem o conjunto de dados;

- Quanto menor o CV mais representativa será a média do processo.

COEFICIENTE DE VARIAÇÃO

- O CV independe das unidades em que foram medidos os dados e/ou independe do tamanho da amostra;
- O coeficiente de variação é uma medida valiosa para comparar a variação (o grau de homogeneidade) entre dois conjuntos de dados (duas distribuições de Frequências) medidos em unidades diferentes. Ou ainda, quando apresentam médias diferentes, embora suas unidades de medidas sejam iguais.

COEFICIENTE DE VARIAÇÃO

Exemplo 6

Suponha que determinado fornecedor "A" de papeletas tenha enviado ao Departamento de Compras de uma empresa uma amostra de 2000 papeletas, de comprimento variando entre 101 e 113 mm. Um fornecedor "B" apresentou um lote deste mesmo tipo de papeletas com o mesmo número de peças. Efetuaram uma análise e encontrou-se:

Fornecedor	Comprimento Médio	S do comprimento
A	107,9	2,72
B	108,0	1,08

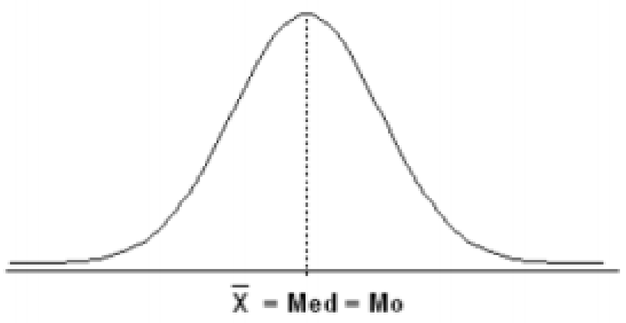
Qual o lote que você escolheria se fosse o comprador?

MEDIDAS DE FORMA

Estas medidas referem-se à forma da curva de uma distribuição de frequência, mais especificamente do polígono de frequência ou do histograma.

Simetria

Em uma distribuição simétrica, média, moda e mediana tem igual valor.



Distribuição simétrica

MEDIDAS DE FORMA

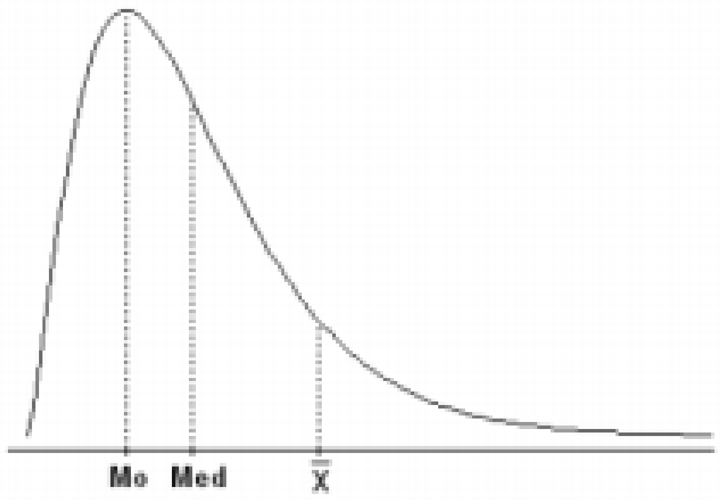
Assimetria

Assimetria é o grau de desvio, afastamento da simetria ou grau de deformação de uma distribuição de frequências.

Toda distribuição deformada é sempre assimétrica. Entretanto, a assimetria pode dar-se na cauda esquerda ou na direita da curva de frequências.

MEDIDAS DE FORMA

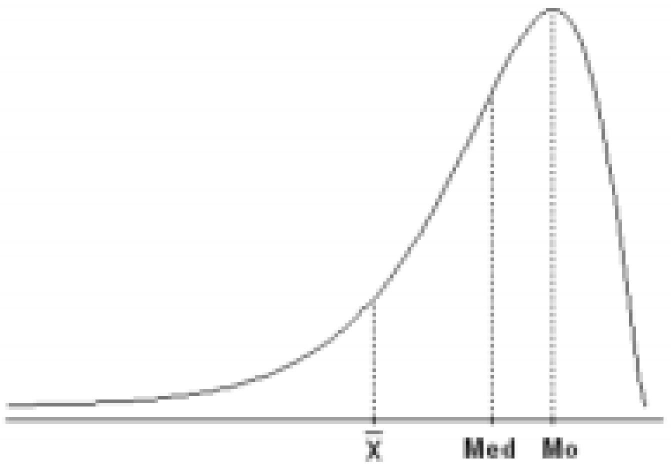
Assimétrica à direita (ou positiva): $M_o < m_d < \bar{X}$



Distribuição assimétrica à direita

MEDIDAS DE FORMA

Assimétrica à esquerda (ou negativa): $\bar{X} < m_d < M_o$



Distribuição assimétrica à esquerda

CÁLCULO DO COEFICIENTE DE ASSIMETRIA

Uma das formas mais utilizadas para calcular o coeficiente de assimetria é o coeficiente de Pearson

$$AS = \frac{\bar{X} - M_o}{s}$$

onde, $\bar{X}$ é a média, M_o é a moda e s é o desvio padrão.

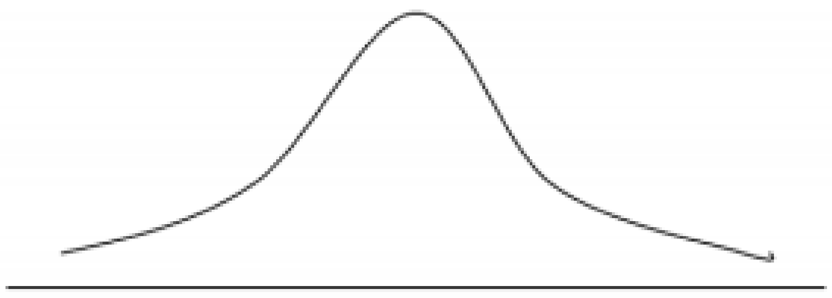
Quando:

- AS = 0, a distribuição é simétrica;
- AS < 0, a distribuição é assimétrica negativa;
- AS > 0, a distribuição é assimétrica positiva.

CURTOSE

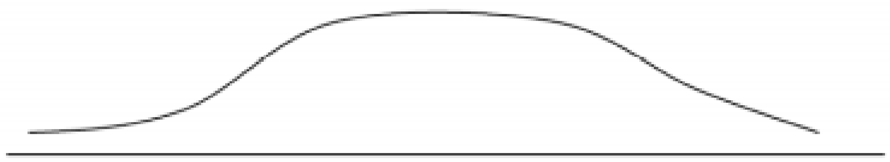
Curtose é o grau de achatamento (ou afilamento) de uma distribuição em comparação com uma distribuição padrão (chamada curva normal). De acordo com o grau da curtose, classificamos três tipos de curvas de frequência:

Mesocúrtica: É uma curva básica de referência chamada curva padrão ou curva normal.



CURTOSE

Platicúrtica: É uma curva mais achatada (ou mais aberta) que uma curva normal.



Leptocúrtica: É uma curva mais afilada que uma curva normal.