CURSO DE ESTATÍSTICA BÁSICA

MÓDULO 2

CAPÍTULO 1 – Medidas de Posição

1- Introdução

O estudo feito sobre distribuições de frequência permite descrever, de maneira geral, os grupos dos valores que uma variável pode assumir. Desse modo, pode-se localizar a maior concentração de valores de uma distribuição, isto é, se ela se localiza no início, no meio ou no final, ou, ainda, se há uma distribuição por igual.

No entanto, para ressaltar as tendências características de cada distribuição, isoladamente, ou em comparação com outras, necessita-se introduzir conceitos que se expressem através de números, que possibilitem traduzir essas tendências. Esses conceitos denominam-se elementos típicos da distribuição e são as

- medidas de posição.
- medidas de variabilidade ou dispersão.
- medidas de assimetria.
- medidas de curtose.

A maioria das vezes em que os dados estatísticos são analisados procura-se obter um valor para representar um conjunto de dados. Esse valor deve sintetizar o comportamento do conjunto do qual ele é proveniente. Nem sempre os dados estudados têm um bom comportamento, isto pode fazer com que um único valor possa representa-lo ou não perante o grupo.

As medidas de posição mais relevantes são as medidas de tendência central, que recebem tal denominação porque os dados observados tendem, em geral, a se agrupar em torno dos valores centrais.

Dentre as **medidas de tendência central**, destacam-se as seguintes: **Médias, Moda** e **Mediana**. Cada uma com um significado diferenciado, no entanto tendo como utilidade representar um conjunto de dados.

As outras medidas de posição são as separatrizes, que englobam: a própria mediana, os quartis e os percentis.

2- Médias

2.1-Média Aritmética Simples ($\bar{x}$)

É o quociente da divisão da soma dos valores de todos os dados do conjunto pela quantidade deles.

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

$$\boxed{\overline{X} = \frac{X_1 + X_2 + X_3 + \dots + \dots X_n}{n} \quad ou \ \overline{X} = \frac{\sum X_i}{n}}$$

Sendo:

$\bar{x}$ → a média aritmética

x_i → os valores da variável

n→ o número de valores

Exemplo: Média de dados não-agrupados

Sabendo-se que as vendas diárias da empresa A, durante uma semana, foram de 10, 14, 13, 15, 16, 18 e 12 unidades. Determinar a média de vendas nesta semana feitas pela empresa A.

$x_1 = 10$, $x_2 = 14$, $x_3 = 13$, $x_4 = 15$, $x_5 = 16$, $x_6 = 18$ e $x_7 = 12$ e $n = 7$.

$$\boxed{\overline{X} = \frac{\sum X_i}{n} \quad \frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7} = \frac{98}{7} = 14}$$

Obs.: Às vezes, a média pode ser um número diferente de todos os da série de dados que ela representa. É o que acontece quando temos os valores 2,4, 6 e 8, para os quais a média é 5. Esse será o número representativo dessa série de valores, embora não esteja representado nos dados originais. Neste caso, costuma-se dizer que a média não tem existência concreta.

2.2-Média Aritmética Ponderada ($\overline{x_p}$)

É uma média aritmética na qual será atribuído um peso a cada valor da série.

Exemplo:

O capital da empresa está sendo formado pelos acionistas, por financiamentos e por debêntures. Cada tipo tem um custo diferente para a empresa, definido pela sua taxa anual. Calcule a taxa de juros média do capital da empresa, considerando os dados apresentados na tabela seguinte.

Capital da Empresa	Participação	Taxas de Juros
Acionista	**R$1000000,00**	**12%**
Financiamento	**R$600000,00**	**8%**
Debêntures	**R$400000,00**	**14%**

$$\overline{x_p} = \frac{x_1.p_1 + x_2.p_2 + x_3.p_3}{p_1 + p_2 + p_3}$$

$$\overline{x_p} = \frac{12\%.\ 1.000.000 + 8\%.600.000 + 14\%.400.00}{1.000.000 + 600.00 + 400.00} = 11,20\%$$

2.3-Média Aritmética Para Dados Agrupados sem Intervalos de Classes ($\bar{x}$)

As frequências são as quantidades de vezes que a variável ocorre na coleta de dados, elas funcionam como fatores de ponderação, o que leva a se calcular a média aritmética ponderada.

$$\bar{x} = \frac{\sum x_i.f_i}{n} \quad \textbf{Ou} \quad \bar{x} = \frac{\sum x_i.f_i}{\sum f_i}$$

Exemplos:

1-Após ter sido realizado um trabalho bimestral, o professor efetuou o levantamento das notas obtidas pelos alunos, observou a seguinte distribuição e calculou a média de sua turma.

Notas dos alunos - x_i	Número de alunos - f_i	$x_i . f_i$
1	1	1
2	3	6
3	5	15
4	1	4
Total ($\sum$)	n = 10	26

$$\boxed{\overline{X} = \frac{\sum X_i . f_i}{n} = \frac{26}{10} = 2,6}$$

2-Consideremos 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
	$\Sigma = 34$

Nesse caso, 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 pela fórmula.

$$\bar{x} = \frac{\sum x_i . f_i}{\sum f_i}$$

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

x_i	f_i	$x_i f_i$
0	2	0
1	6	6
2	10	20
3	12	36
4	4	16
	$\Sigma = 34$	$\Sigma = 78$

Temos, então:

$\sum x_i.f_i = 78$ e $\sum f_i = 34$

Logo:

$$\bar{x} = \frac{\sum x_i.f_i}{\sum f_i} = \bar{x} = \frac{78}{34} = 2{,}29 \rightarrow \bar{x} = 2{,}3$$

Isto é: $\bar{x} = 2{,}3$ meninos.

Observação: Sendo x uma variável discreta, como interpretar o resultado obtido, 2 meninos e 3 décimos de meninos? O valor médio 2,3 meninos sugere, nesse 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.

2.4-Média Aritmética Para Dados Agrupados com Intervalos de Classes ($\bar{x}$)

Convenciona-se que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio, e determina-se a média aritmética ponderada por meio da fórmula.

$$\bar{x} = \frac{\sum x_i.f_i}{n} \qquad \textbf{ou} \qquad \bar{x} = \frac{\sum x_i.f_i}{\sum f_i}$$

Onde x_i é o ponto médio da classe.

Exemplos:

1- Determine a renda média familiar, de acordo com os dados da tabela.

Classes - Renda familiar	xi	fi - Número de famílias	xifi
2 ├ 4	3	5	15
4 ├ 6	5	10	50
6 ├ 8	7	14	98
8 ├ 10	9	8	72
10 ├ 12	11	3	33
Total Σ		**n = 40**	**268**

$$\boxed{\overline{X} = \frac{\sum X_i.f_i}{n} = \frac{268}{40} = 6{,}7}$$

2- Consideremos a distribuição

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		Σ = 40

Vamos, inicialmente, abrir uma coluna para os pontos médios e o outra para os produtos $x_i.f_i$.

i	ESTATURAS (cm)	f_i	x_i	x_if_i
1	150 ⊢ 154	4	152	608
2	154 ⊢ 158	9	156	1.404
3	158 ⊢ 162	11	160	1.760
4	162 ⊢ 166	8	164	1.312
5	166 ⊢ 170	5	168	840
6	170 ⊢ 174	3	172	516
		Σ = 40		Σ = 6.440

$$\bar{x} = \frac{\sum x_i.f_i}{\sum f_i} \rightarrow \bar{x} = \frac{6440}{40} = 161 \rightarrow \bar{x} = 161\ cm.$$

2.5-Desvio em Relação à Média

Denomina-se desvio em relação à média, a diferença entre cada elemento de um conjunto de valores e a média aritmética.

Designando o desvio por d_i, tem – se: $d_i = x_i - \bar{x}$

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, tem-se, para produção média da semana:

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

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

$d_1 = x_1 - \bar{x} \rightarrow d_1 = 10 - 14 = -4$

$d_2 = x_2 - \bar{x} \rightarrow d_2 = 14 - 14 = 0$

$d_3 = x_3 - \bar{x} \rightarrow d_3 = 13 - 14 = -1$

$d_4 = x_4 - \bar{x} \rightarrow d_4 = 15 - 14 = 1$

$d_5 = x_5 - \bar{x} \rightarrow d_5 = 16 - 14 = 2$

$d_6 = x_6 - \bar{x} \rightarrow d_6 = 18 - 14 = 4$

$d_7 = x_7 - \bar{x} \rightarrow d_7 = 12 - 14 = -2$

2.6- Propriedades da Média

1ª propriedade:

A soma algébrica dos desvios tomados em relação à média é nula.

$$d_i = \sum(x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{k} d_i = 0$$

No exemplo anterior, tem-se:

$$\sum_{i=1}^{7} d_i = (-4) + 0 + (-1) + 1 + 2 + 4 + (-2) = (-7) + 7 = 0$$

2ª propriedade:

Somando-se (ou subtraindo-se) uma constante (c) a de todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

$$y_i = x_i \pm c \rightarrow \bar{y} = \bar{x} \pm c$$

Somando 2 a cada um dos valores da variável do exemplo anterior, tem-se: y_1=12, y_2=16, y_3=15, y_4=17, y_5=18, y_6=20 e y_7=14.

Daí:

$$\sum_{i=1}^{7} y_i = 12 + 16 + 15 + 17 + 18 + 20 + 14 = 112$$

Como n = 7, vem:

$$\bar{y} = \frac{112}{7} = 16 \ \rightarrow \bar{y} = 16 \ = 14 + 2 \ \rightarrow \bar{y} = \ \bar{x} + 2$$

3ª propriedade:

Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

$$y_i = x_i . c \rightarrow \bar{y} = \bar{x} . c$$

Ou

$$y_i = \frac{x_i}{c} \rightarrow \bar{y} = \frac{\bar{x}}{c}$$

Multiplicando por 3 cada um dos valores da variável do exemplo dado, obtém-se: y_1=3, y_2=42, y_3=39, y_4=45, y_5=48, y_6=54 e y_7=36.

Daí:

$$\sum_{i=1}^{7} y_i = 30 + 42 + 39 + 45 + 48 + 54 + 36 = 294$$

Como n = 7 , tem-se:

$$\bar{y} = \frac{294}{7} = 42 \ \rightarrow \bar{y} = 42 \ = 14 \, . \, 3 \ \rightarrow \bar{y} = \ \bar{x} \, . \, 3$$

Obs.:

A média é utilizada quando

- deseja-se obter a medida de posição que possui maior estabilidade.
- houver necessidade de um tratamento algébrico ulterior.

3- Moda (Mo)

Denomina-se **moda** o valor que ocorre com maior frequência em uma série de valores.

Dessa forma, o **salário modal** dos empregados de uma empresa é o salário mais comum, isto é, o salário recebido pelo maior número de empregados dessa empresa.

3.1-Moda (Mo) Para Dados Não-Agrupados

Primeiramente os dados devem ser ordenados (colocados em rol) para, em seguida, se observar o valor que tem maior frequência. Quando se trata de valores não-agrupados, a moda é facilmente reconhecida: basta, de acordo com a definição, procurar o valor que mais se repete.

Exemplos:

Calcular a moda dos seguintes conjuntos de dados.

- X = (4, 5, 5, 6, 6, 6, 7, 7, 8, 8) → Mo = 6 (0 valor mais frequente)

Esse conjunto é unimodal, pois apresenta apenas uma moda.

- Y = (1, 2, 2, 2, 3, 4, 4, 4, 5, 5, 6) → Mo = 2 e Mo = 4 (valores mais frequentes)

Esse conjunto é bimodal, pois apresenta duas modas.

- Z = (1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5) →Mo = 2, Mo = 3 e Mo = 4 (valores mais frequentes)

Esse conjunto é plurimodal, pois apresenta mais de duas modas.

Pode-se, entretanto, encontrar séries nas quais não exista valor modal, isto é, nas quais nenhum valor apareça mais vezes que outros. É o caso da série: 3,5,8,10,12,13 que não apresenta moda (**amodal**).

- W = (1, 2, 3, 4, 5, 6) →Esse conjunto é amodal porque não apresenta um valor predominante, ou seja, não tem moda.

Observação:

- A moda é utilizada quando se deseja obter uma medida rápida e aproximada de posição ou quando a medida de posição deva ser o valor mais típico da distribuição. É uma medida pouco utilizada.
- Já a média aritmética é a medida de posição que possui maior confiabilidade numérica, além de ser mais intuitiva, do ponto de vista matemático.

3.2-Moda (Mo) Para Dados Agrupados sem Intervalos de Classe

Uma vez agrupados os dados, é possível determinar imediatamente a moda: basta fixar o valor da vaiável de maior frequência.

Na distribuição da tabela abaixo, à frequência máxima (12) corresponde o valor 3 da variável. Logo Mo = 3.

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

3.3-Moda (Mo) Para Dados Agrupados Com Intervalos de Classe

A classe que apresenta a maior frequência é denominada **classe modal**. Pela definição, pode-se afirmar que a moda, neste caso, é o valor dominante que está compreendido entre os limites da classe modal.

O método mais simples para o cálculo da moda consiste em tomar o ponto médio da classe modal. Esse valor denomina-se **moda bruta**. Tem-se:

$$Mo = \frac{l^* + L^*}{2}$$

Onde:

l* é o limite inferior da classe modal.

L* é o limite superior da classe modal.

Assim para a distribuição

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11 ←
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\Sigma = 40$

temos que a classe modal é i = 3, l* = 158 e L* = 162.

Como

$$Mo = \frac{l^* + L^*}{2}$$

vem

$$Mo = \frac{158 + 162}{2} = \frac{320}{2} = 160$$

Logo: Mo= 160 cm.

Observação:

Existem, para o cálculo da moda, outros métodos mais elaborados, como por exemplo, o que faz uso da **fórmula de Czuber**

$$Mo = l^* + \frac{D_1}{D_1 + D_2}\ x\ h^*$$

na qual:

l* é o limite inferior da classe modal

h* é a amplitude da classe modal

D_1 = f* – f (ant)

D_2 = f* – f(post)

sendo:

f* a frequência simples da classe modal

f(ant) a frequência simples da classe anterior à classe modal

f(post) a frequência simples da classe posterior à classe modal.

Assim para a distribuição

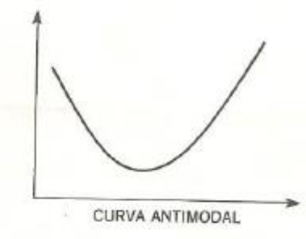
i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11 ←
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\Sigma = 40$

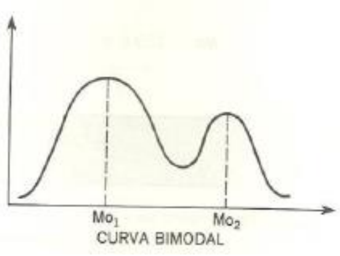
D1 = 11 – 9 = 2 e D2 = 11 – 8 = 3

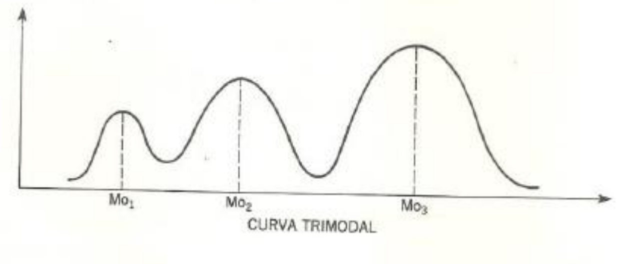
$$Mo = l^* + \frac{D_1}{D_1 + D_2}\ x\ h^* \to Mo = 158 + \frac{2}{2+3}\ x\ 4 \to Mo = 158 + \frac{8}{5} \to$$
$$Mo = 158 + 1{,}6 = 159{,}6$$

Logo: Mo = 159,6 cm.

3.4-As Expressões Gráficas da Moda







4- Mediana (Md)

A mediana é a medida de posição definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Ou seja, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é um valor situado de tal modo no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

4.1-Mediana (Md) Para Dados Não-Agrupados

Para se descobrir o elemento mediano de uma série devem-se seguir os procedimentos abaixo:

- Se **N** (número de elementos do conjunto) for **ímpar** a mediana é o termo de ordem P dado pela razão:

$$\boxed{P = \frac{N+1}{2}}$$

- Se **N** (número de elementos do conjunto) for **par,** a mediana é a média aritmética dos termos de ordem, em um primeiro passo de P1 (média aritmética simples) e, em seguida, pela razão P2 (termo subsequente da ordem P):

$$\boxed{P_1 = \frac{N}{2} \quad \text{e} \quad P_2 = \frac{N}{2} + 1}$$

Exemplos:

1- Determine o valor da mediana da série que é composta pelos seguintes elementos: 56,65, 58, 62, e 90.

Colocar os elementos em ordem: 56,58,62,65 e 90

N = 5 (ímpar)

$$P = \frac{N+1}{2} = \frac{5+1}{2} = 3 \rightarrow 3^{\underline{o}}\ elemento\ \rightarrow Md = 62$$

2- Em uma pesquisa realizada a respeito de erros por folha, cometidos por digitadores, revelou as seguintes quantidades: 12,13.15,13,12,18,16,20. Determinar a quantidade mediana de falhas.

Colocar os dados em ordem: 12,12,13,13,15,16,18 e 20.

N=8 (par)

$$P_1 = \frac{N}{2} = \frac{8}{2} = 4 \rightarrow 4^{\underline{o}}\ elemento \rightarrow Md = 13$$

$$P_2 = \frac{N}{2} + 1 = 4 + 1 = 5 \rightarrow 5^{\underline{o}}\ elemento \rightarrow Md = 15$$

Logo a mediana será $Md = \frac{13+15}{2} = \frac{28}{2} = 14$

Nota:

Ou ainda, de modo simplista, dividimos os dados em limites à esquerda e à direita, destacando os valores centrais.

1°. passo: organizamos o Rol:

Rol: 12, 12, 13, 13, 15, 16, 18, 20.

2°. passo: destacamos os valores centrais (metade à esquerda, metade à direita):

12, 12, 13, $\boxed{13,\ 15,}$ **16, 18, 20**

3°. passo: efetuamos a média aritmética:

$$\frac{13+15}{2} = \frac{28}{2} = 14$$

Observações:

- O valor da mediana pode coincidir ou não com um elemento da série. Quando o número de elementos da série é ímpar, há coincidência. O mesmo não acontece, porém, quando esse número é par.
- Se uma lista de valores tem um número ímpar de elementos, a mediana é o valor do meio, quando a lista se apresenta ordenada; se a lista tem número par de elementos, então a mediana é a média dos dois números mais próximos do meio.

- A mediana e a média aritmética não têm, necessariamente, o mesmo valor.

Exemplo: Dada a série de valores 5,13,10, 2, 18, 15, 6, 16,9, determine a média e a mediana:

Colocando os elementos em ordem: 2,5,6,9,10,13,15,16,18

Média: $\bar{x} = \frac{2+5+6+9+10+13+15+16+18}{9} = \frac{94}{9} = 10{,}4$

Mediana: $Md = 10$

- A mediana depende da posição dos valores dos elementos na série ordenada. Essa é uma das diferenças marcantes entre a mediana e a média (que se deixa influenciar, e muito, pelos valores extremos.

  Exemplos:

  5,7,10,13,15 →$\bar{x} = 10\ e\ Md = 10$

  5,7,10,13,65→$\bar{x} = 20\ e\ Md = 10$

  Obs.: A média do segundo conjunto de valores é maior do que a do primeiro, por influência dos valores extremos, ao passo que a mediana permanece a mesma.
- A mediana é designada, muitas vezes, por valor mediano.

4.2-Mediana (Md) Para Dados Agrupados sem Intervalos de Classe

Se os dados se agrupam em uma distribuição de frequência, o cálculo da mediana se processa de modo muito semelhante àquele dos dados não agrupados, implicando, porém, a determinação prévia das frequências acumuladas. Precisa-se determinar um valor tal que divida a distribuição em dois grupos que contenham o mesmo número de elementos. Para o caso de uma distribuição, porém, a ordem, a partir de qualquer um dos extremos, é dada por $\frac{\sum f_i}{2}$.

Neste caso, é o bastante identificar a frequência acumulada imediatamente superior à metade da soma das frequências. A mediana será o valor da variável que corresponde à tal frequência acumulada.

Exemplo:

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

Sendo $\frac{\sum f_i}{2} = \frac{34}{2} = 17$, 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.

Observação:

- No caso de existir uma frequência acumulada (F_i), tal que $F_i = \frac{\sum f_i}{2}$, a mediana será dada por $Md = \frac{x_i + x_i + 1}{2}$, isto é, a mediana será a média aritmética entre o valor da variável correspondente a essa frequência acumulada e o seguinte.

Exemplo:

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

Temos: $\frac{\sum f_i}{2} = \frac{8}{2} = 4 = F_3$

Logo: $Md = \frac{x_i + x_i + 1}{2} \rightarrow Md = \frac{15+16}{2} = \frac{31}{2} = 15{,}5$

Donde: Md = 15,5.

4.3-Mediana (Md) Para Dados Agrupados com Intervalos de Classe

Neste caso, o problema consiste em determinar o ponto do intervalo em que está compreendida a mediana.

Para tento, temos inicialmente que determinar a classe na qual se acha a mediana – classe mediana. Tal classe será, evidentemente, aquela correspondente à frequência acumulada imediatamente superior a $\frac{\sum f_i}{2}$.

Feito isto, um problema de interpolação resolve a questão, admitindo-se, agora, que os valores se distribuam uniformemente em todo o intervalo de classe.

Assim, considerando a distribuição da tabela abaixo, acrescida das frequências acumuladas,

i	ESTATURAS (cm)	f_i	F_i	
1	150 ⊢ 154	4	4	
2	154 ⊢ 158	9	13	
3	158 ⊢ 162	11	24	← classe mediana
4	162 ⊢ 166	8	32	
5	166 ⊢ 170	5	37	
6	170 ⊢ 174	3	40	
		$\Sigma = 40$		

temos: $\frac{\sum f_i}{2} = \frac{40}{2} = 20.$

Como há 24 valores incluídos nas três primeiras classes da distribuição e como se pretende determinar o valor que ocupa o 20º lugar, a partir do início da série, observa-se que este deve estar localizado na terceira classe (i = 3), supondo que as frequências dessas classes estejam uniformemente distribuídas.

Como existem 11 elementos nessa classe e o intervalo de classe é igual a 4, deve-se tomar, a partir do limite inferior, a distância : $\frac{20-13}{11}\, x\, 4 = \frac{7}{11}\, x\, 4$, e a mediana será dada por: $Md = 158 + \frac{7}{11}\, x\, 4 = 158 + \frac{28}{11} = 158 + 2{,}54 = 160{,}54.$

Logo: Md = 160,5 cm.

Passos para o cálculo da mediana:

1º) Determinam-se as frequências acumuladas.

2º) Calcula-se $\frac{\sum f_i}{2}$.

3º) Marca-se a classe correspondente à frequência acumulada imediatamente superior à $\frac{\sum f_i}{2}$ - classe mediana – e, em seguida, emprega-se a fórmula

$$Md = l^* + \frac{\left[\frac{\sum f_i}{2} - F(ant)\right] h^*}{f^*}$$

na qual:

l* é o limite inferior da classe mediana.

F(ant) é a frequência acumulada da classe anterior à classe mediana.

f* é a frequência simples da classe mediana.

h* é a amplitude do intervalo da classe mediana.

Exemplo:

i	ESTATURAS (cm)	f_i	F_i	
1	150 ⊢ 154	4	4	
2	154 ⊢ 158	9	13	
3	158 ⊢ 162	11	24	← classe mediana
4	162 ⊢ 166	8	32	
5	166 ⊢ 170	5	37	
6	170 ⊢ 174	3	40	
		$\Sigma = 40$		

$$\frac{\sum f_i}{2} = \frac{40}{2} = 20$$

Logo, a classe mediana é a de ordem 3. Então: l*=158, F(ant)=13, f*=11 e h*=4.

Substituindo esses valores na fórmula, obtém-se:

$Md = 158 + \frac{(20-13).4}{11} = 158 + \frac{28}{11} = 158 + 2{,}54 = 160{,}54$

Isto é :Md = 160,5 cm.

Observações:

- No caso de existir uma frequência acumulada exatamente igual a $\frac{\sum f_i}{2}$, a mediana será o limite superior da classe correspondente.
- Quando uma distribuição de números é razoavelmente simétrica, sem valores extremamente altos ou baixos, os valores da média e da mediana, em geral, são muito próximos um do outro.
- Há ocasiões em que a mediana constitui melhor medida de tendência central do que a média.

Exemplo:

Suponha que os valores abaixo representem vendas de pizza de muçarela por um período de 9 dias.

36,35,37,29,39,36,340,35,36

Observe que, certo dia, um grande ônibus com amantes de pizza de muçarela chegou ao estabelecimento; as vendas desse tipo de pizza foram muito maiores naquele dia.

Calculando a média desses valores obtém-se: $\frac{623}{9} = 69{,}22.$

Entretanto, nenhum dos valores está próximo de 69,22. Ordenando então a relação: 340 – 39 – 37 – 36 – 36 – 36 – 35 – 35 – 29 .

Verifica-se então que a mediana é 36. Nesse caso, o valor da mediana dá uma ideia muito melhor do número provável das vendas em determinado dia. Em geral, quando uma relação de valores contém um valor extremo (muito acima ou muito abaixo dos outros valores da lista), a média não é uma medida muito representativa. A mediana constitui melhor medida de tendência central. Mas a média é mais fácil de calcular, sendo, por isso, utilizada com maior frequência.

4.4-Emprego da Mediana

Emprega-se a mediana quando

- deseja-se obter o ponto que divide a distribuição em partes iguais.
- existem valores extremos que afetam de uma maneira acentuada a média.
- a variável em estudo é salário.

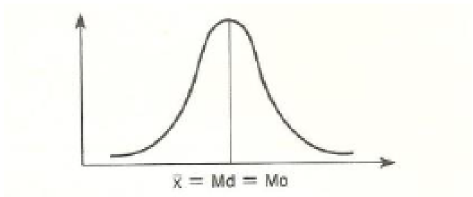
5-Posição Relativa da Média, Mediana e Moda.

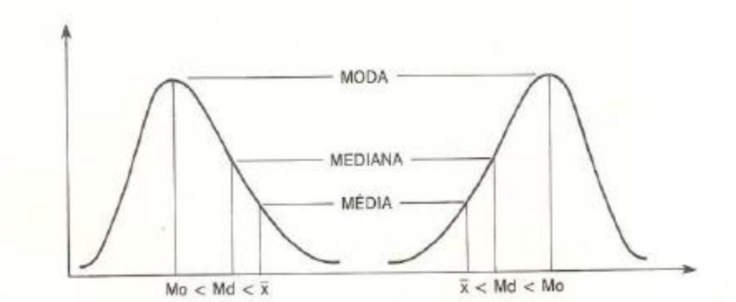
Quando uma distribuição é simétrica, as três medidas coincidem. Porém, a assimetria torna-as diferentes e essa diferença é tanto maior quanto maior é a assimetria. Assim, em uma distribuição em forma de sino, tem-se:

$$\bar{x} = Md = Mo, no\ caso\ da\ curva\ simétrica$$

$$Mo < Md < \bar{x}, no\ caso\ da\ curva\ assimétrica\ positiva$$

$$\bar{x} < Md < Mo, no\ caso\ da\ curva\ assimétrica\ negativa$$





6- Comparação Entre Média, Mediana e Moda

Medida	Definição	Vantagens	Desvantagens
Média	Centro de distribuição de frequências.	• Reflete cada valor. • Possui propriedades matemáticas atraentes.	• É afetada por valores extremos.
Mediana	Metade dos valores são maiores, metade menores.	• Menos sensíveis a valores extremos do que a média.	• Difícil de determinar quando há uma grande quantidade de dados.
Moda	Valor mais frequente.	• Valor "típico": maior quantidade de valores concentrados neste ponto.	• Não se presta à análise matemática. • Pode não ter moda para certos conjuntos de dados.

7-As Separatrizes

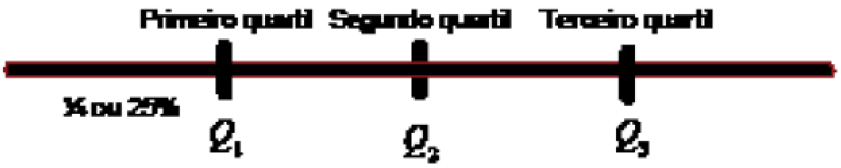
Outras medidas de posição, como os quartis, os decis e os percentis, embora sejam medidas de posição, possuem uma característica muito especial: separam os conjuntos em quantidades de iguais valores. Por isso, essas medidas podem ser chamadas de separatrizes.

Alguns estudiosos de Estatística preferem chamar as separatrizes de medidas de posição e a média, a mediana e a moda (que também são medidas de posição), preferem chamar de medidas de tendência central. Os autores não concordam quanto a melhor maneira de considerá-las.

Quartis, decis e percentis são medidas de posição, isto é, semelhantes às medidas de tendência central, indicam uma determinada localização em relação ao conjunto de dados em estudo. Entretanto, separam o conjunto em 4 partes iguais (quartis), 10 partes iguais (decis) ou 100 partes iguais (percentis), ou seja, em partes que apresentam o mesmo número de valores. Por isso, alguns autores preferem as medidas de posição (quartis, decis e percentis) de separatrizes (juntamente com a mediana).

7.1-Os Quartis

Denominam-se quartis os valores de uma série que a dividem em quatro partes.



Existem, portanto, três quartis:

a) O **primeiro quartil** (**Q_1**) – valor situado de tal modo na série que uma quarta parte (25%) dos dados é menor que ele e as três quartas partes restantes (75%) são maiores.

b) O **segundo quartil** (**Q_2**) – evidentemente, coincide com a mediana (**Q_2** = Md).

c) O **terceiro quartil** (**Q_3**) – valor situado de tal modo que as três quartas partes (75%) dos termos são menores que ele e uma quarta parte (25%) é maior.

Quando os dados são agrupados, para determinar os quartis usa-se a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana, $\frac{\sum f_i}{2}$ por:

$$\frac{k\sum f_i}{4}$$

$$\frac{k\sum f_i}{4},\ sendo \begin{cases} k \text{ o número de ordem do quartil} \\ \sum f_i \text{ a soma total das freqüências simples} \end{cases}$$

sendo K o número de ordem do quartil. Assim, tem-se:

$$Q_1 = l^* + \frac{\left[\frac{\sum f_i}{4} - F(ant)\right] h^*}{f^*}$$

$$Q_2 = l^* + \frac{\left[\frac{\sum f_i}{2} - F(ant)\right] h^*}{f^*}$$

$$Q_3 = l^* + \frac{\left[\frac{3\sum f_i}{4} - F(ant)\right] h^*}{f^*}$$

$$Q = l^* + \frac{\left[\frac{k\sum f_i}{4} - F(ant)\right] h^*}{f^*}$$

Onde,

k é o número de ordem do *quartil* (1, 2 ou 3);

l^* é o limite inferior da *classe mediana*;

$F(ant)$ é a freqüência acumulada da classe *anterior* à classe mediana;

f^* é a freqüência simples da classe mediana;

h^* é a amplitude do intervalo da classe mediana.

Exemplo: Calcular o primeiro, o segundo e o terceiro quartis da distribuição de frequência abaixo.

Altura dos alunos da Turma A

Estaturas (cm)	f_i	F_i
[150,154[	4	4
[154,158[	9	13
[158,162[	11	24
[162,166]	8	32
[166,170[	5	37
[170,174[	3	40
	$\sum = 40$	

Os quartis, como já foi falado, são valores que dividem os conjuntos em 4 partes iguais. O resultado encontrado quando se aplica a fórmula, lamentavelmente, não fornece de imediato, a posição do quartil; no entanto, indica em que linha de classe se encontra.

Solução

Sabemos que $\sum f_i = 40$ Então,

Primeiro quartil (k = 1) Segundo quartil (k = 2) Terceiro quartil (k = 3)

$$\frac{1 \times \sum f_i}{4} = \frac{40}{4} = 10 \qquad \frac{2 \times \sum f_i}{4} = \frac{80}{4} = 20 \qquad \frac{3 \times \sum f_i}{4} = \frac{120}{4} = 30$$

Qual é o significado, por exemplo, da posição 20 para Q_2?

O segundo quartil divide o conjunto em duas partes iguais. Ainda não se sabe que valor é esse; porém o resultado 20 indica a linha (ou classe) em que se encontra.

Observação:

1) se o valor encontrado existir na linha da Freqüência Acumulada (no nosso exercício esse valor é 20), então, esta será a classe *quartil* (a linha que estou procurando);

2) caso o valor não exista, a classe *quartil* será aquela que contiver a Freqüência Acumulada, imediatamente, superior. No nosso caso, não existe a Freqüência acumulada 20, portanto, a imediatamente superior é 24. Essa é a linha que estamos procurando.

Assim, como o segundo quartil se encontra na posição 20. Então, ele só pode estar na 3ª linha da Tabela de Distribuição de Frequência.

Altura dos alunos da Turma A

Estaturas (cm)	f_i	F_i	
[150,154[	4	4	
[154,158[	9	13	← Q_1
[158,162[	11	24	← Q_2
[162,166[	8	32	← Q_3
[166,170[	5	37	
[170,174[	3	40	
	$\sum = 40$		

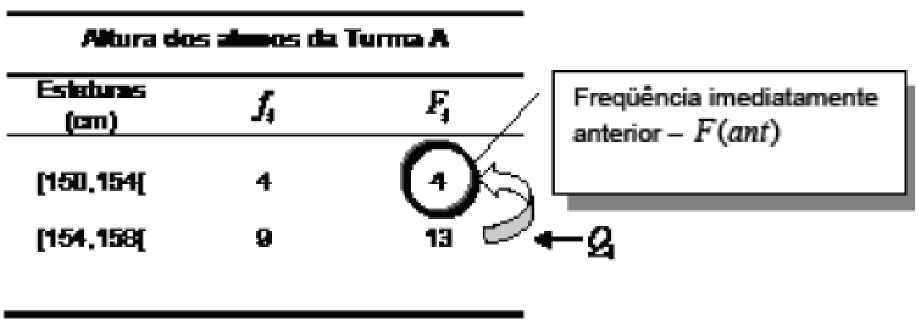
Uma vez descobertas as classes do primeiro, segundo e terceiro quartis, pode-se destacar a linha da classe do primeiro quartil.

Altura dos alunos da Turma A		
Estaturas (cm)	f_i	F_i
[154,158[	9	13

Na linha de classe de Q_1, as estaturas variam de 154 cm a 158 cm: o limite inferior ($l*$), isto é, o menor valor é 154. Na linha de classe de Q_2, o limite inferior da classe é 158. E para Q_3, $l* = 162$.

Quartil	$\frac{k\sum f_i}{4}$	$l*$	$F(ant)$	$h*$	$f*$	Resultado
Q_1	10	154				
Q_2	20	158				
Q_3	30	162				

Agora, para encontrar a frequência acumulada F(ant), uma vez determinada a linha Q_1, basta observar a frequência acumulada da linha de cima. Para Q_1, a frequência acumulada anterior será 4



.

Quartil	$\frac{k\sum f_i}{4}$	$l*$	$F(ant)$	$h*$	$f*$	Resultado
Q_1	10	154	4			
Q_2	20	158	13			
Q_3	30	162	24			

A determinação da amplitude do intervalo de classe é imediata. Localizada a linha quartil, basta subtrair o maior valor do menor valor do intervalo de classe. Desse modo, Q_1 pertence à 2ª linha e o intervalo de classe é [154,158[; a amplitude do intervalo será dada por: 158 – 154 = 4. Efetuando o cálculo para Q_2 e Q_3 será encontrado o mesmo resultado.

Quartil	$\frac{k\sum f_i}{4}$	$l*$	$F(ant)$	$h*$	$f*$	Resultado
Q_1	10	154	4	4		
Q_2	20	158	13	4		
Q_3	30	162	24	4		

Consultando a tabela, identifica-se a frequência simples de cada quartil. Assim, tem-se: 9,11 e 8, respectivamente para Q_1,Q_2 e Q_3.

Quartil	$\frac{k\sum f_i}{4}$	$l*$	$F(ant)$	$h*$	$f*$	Resultado
Q_1	10	154	4	4	9	
Q_2	20	158	13	4	11	
Q_3	30	162	24	4	8	

Solução:

Primeiro Quartil.

$$Q_1 = l^* + \frac{\left[\frac{k\sum f_i}{4} - F(ant)\right]h^*}{f^*} = 154 + \frac{[10-4]\times 4}{9} = 156,66$$

Segundo Quartil.

$$Q_2 = l^* + \frac{\left[\frac{k\sum f_i}{4} - F(ant)\right]h^*}{f^*} = 158 + \frac{[20-13]\times 4}{11} = 160,54$$

Terceiro Quartil.

$$Q_3 = l^* + \frac{\left[\frac{k\sum f_i}{4} - F(ant)\right]h^*}{f^*} = 162 + \frac{[30-24]\times 4}{8} = 165$$

7.2-Os Percentis

Denominam-se percentis os noventa e nove valores que separam uma série em 100 partes iguais.

Indicam-se: $P_1, P_2, P_3, \ldots, P_{25}, \ldots, P_{50}, \ldots, P_{75}, \ldots, P_{99}$

É evidente que : $P_{50} = Md$; $P_{25} = Q_1$ e $P_{75} = Q_3$

O cálculo de um percentil segue a mesma técnica do cálculo da mediana, no entanto, a fórmula $\frac{\sum f_i}{2}$ será substituída por: $\frac{k\sum f_i}{100}$, sendo k o número de ordem do percentil.

Exemplo:

Calcular o oitavo percentil considerando a tabela de distribuição de frequência abaixo.

Altura dos alunos da Turma A		
Estaturas (cm)	f_i	F_i
[150,154[	4	4
[154,158[	9	13
[158,162[	11	24
[162,166]	8	32
[166,170[	5	37
[170,174[	3	40
$\sum = 40$		

Solução:

$$P = l^* + \frac{\left[\frac{k\sum f_i}{100} - F(ant)\right]h^*}{f^*}$$

$$\frac{k\sum f_i}{100} = \frac{8\times 40}{100} = 3,2$$

Como não existe na coluna de frequência acumulada o valor 3,2; o valor imediatamente acima dele é 4. Portanto, o percentil (P_8) encontra-se na 1ª linha (ou classe).

$$P = l^* + \frac{\left[\frac{k\sum f_i}{100} - F(ant)\right]h^*}{f^*}$$

$$P_8 = 150 + \frac{[3,2-0]\times 4}{4} = 153,2$$

Logo, P_8 = 153,2 cm. Significa que 8% possuem estatura inferior a 153,2%.

7.3-Os Decis

São valores que dividem o conjunto de dados ordenados (rol) em 10 (dez) partes iguais.

Primeiro Decil (D_1) - valor situado de tal modo na série de dados que 10% das observações são menores que ele e 90% são maiores.

Segundo Decil (D_2) - valor situado de tal modo na série de dados que 20% das observações são menores que ele e 80% são maiores.

Nono Decil (D_9) - valor situado de tal modo na série de dados que 90% das observações são menores que ele e 10% são maiores.

Para encontrar as posições dos decis utilizam-se fórmulas semelhantes às da mediana,

$$\frac{k.\sum f_i}{10}$$

$$D = l^* + \frac{\left[\frac{k\sum f_i}{10} - F(ant)\right]h^*}{f^*}$$

Exemplo:

Os salários (em salários mínimos) de 160 profissionais de uma empresa estão distribuídos conforme a tabela a seguir. Calcule Q1, D4 e P85 e interprete os resultados.

Salário	*N. ° de prof. (fi)*	*fac*
01 ├-- 03	20	20
03 ├-- 05	40	60
05 ├-- 07	60	120
07 ├-- 09	30	150
09 ├-- 11	10	160
Total	160	-----

Solução:

1º passo: Determinar as frequências acumuladas da distribuição.

2º passo: Calcular a posição do Quartil, Decil ou Percentil desejado.

$\frac{k\sum f_i}{4} = \frac{1x160}{4} = 40º\ elemento$ → Quartil

$\frac{k\sum f_i}{10} = \frac{4x160}{10} = 64º\ elemento$ → Decil

$\frac{k\sum f_i}{100} = \frac{85x160}{100} = 136º\ elemento$ → Percentil

3º passo: Identificar a classe que contém o quartil, o Decil ou o Percentil desejado por meio da frequência acumulada simples. Quartil (2ª classe); Decil (3ª classe); Percentil (4ª classe).

4º passo: Calcular o Quartil, o Decil ou o Percentil desejado por meio das fórmulas.

$$Q_1 = l^* + \frac{\left[\frac{\sum f_i}{4} - F(ant)\right]h^*}{f^*} \rightarrow Q_1 = 3 + \frac{(40-20)}{40}.2 = 4 \rightarrow Q_1 = 4$$

$$D_4 = l^* + \frac{\left[\frac{\sum f_i}{10} - F(ant)\right]h^*}{f^*} \rightarrow D_4 = 5 + \frac{(64-60)}{60}.2 = 5{,}13 \rightarrow D_4 = 5{,}13$$

$$P_{85} = l^* + \frac{\left[\frac{\sum f_i}{100} - F(ant)\right]h^*}{f^*} \rightarrow P_{85} = 7 + \frac{(136-120)}{30}.2 = 8{,}07 \rightarrow Q_{85} = 8{,}07$$

- Interpretação: 25% dos profissionais da empresa ganham até 4 salários mínimos ou 75% dos profissionais ganham mais de 4 salários mínimos.
- Interpretação: 40% dos profissionais da empresa ganham até 5,13 salários mínimos ou 60% dos profissionais ganham mais de 5,13 salários mínimos.
- Interpretação: 85% dos profissionais da empresa ganham até 8,07 salários mínimos ou 15% dos profissionais ganham mais de 8,07 salários mínimos.

CAPÍTULO 2 – Medidas de Dispersão ou de Variabilidade

1- Dispersão ou Variabilidade

As medidas de dispersão ou variabilidade são empregadas para descobrir o grau de variabilidade ou dispersão dos valores observados em torno da média aritmética. Servem para medir a representatividade da média e destacam o nível de homogeneidade ou heterogeneidade dentro de cada grupo estatístico analisado.

Quando se trata de interpretar dados estatísticos é necessário ter-se uma ideia retrospectiva de como se apresentavam esses mesmos dados nas tabelas. Assim, não é o bastante dar uma das medidas de posição para caracterizar perfeitamente um conjunto de valores.

"Se uma pessoa comeu dois salgadinhos e outra não comeu nenhum, em média cada uma comeu um salgadinho."

Essa frase, que tem relação com a Estatística, não agradou muito àquele que ficou com fome. Ao se fazer a média, há sempre informação que se perde. A média, apesar de ser uma medida muito usada em Estatística, é muitas vezes insuficiente para caracterizar uma distribuição. A moda e a mediana também são medidas que não informam muito sobre como as variáveis se alteram. Por isso, foi preciso encontrar outro indicador que informasse sobre a maneira como os dados se distribuem em torno da média.

Exemplo:

Um empresário deseja comparar o desempenho de dois empregados, com base na produção diária de determinada peça, durante cinco dias.

Empregado A: 70, 71, 69, 70, 70 → = 70

Empregado B: 60, 80, 70, 62, 83 → = 71

O desempenho médio do empregado A é de 70 peças produzidas diariamente, enquanto que a do empregado B é de 71 peças. Com base na média aritmética, verifica-se que o desempenho de B é melhor do que o de A. No entanto, observando bem os dados, percebe-se que a produção de "A" varia apenas de 69 a 71 peças, ao passo que a de “B” varia de 60 a 83 peças, o que revela que o desempenho de A é bem mais uniforme do que o de B.

Portanto, para qualificar os valores de uma dada variável, ressaltando a maior ou menor dispersão ou variabilidade entre esses valores e a sua medida de posição, a Estatística recorre às medidas de dispersão. Dessas medidas, serão destacadas neste estudo a **amplitude total**, a **variância**, o **desvio padrão** e o **coeficiente de variação**.

2- Amplitude Total

2.1- Dados não-agrupados

A amplitude total é a diferença entre o maior e o menor valor observado:
AT = x(máx) – x(mín)

Exemplos:

a) Para os valores: 40,45,48,52,54,62 e 70

Tem-se: AT = 70 – 40 = 30

b) Para a situação sugerida anteriormente.

Empregado A: 70, 71, 69, 70, 70 → = 70

Empregado B: 60, 80, 70, 62, 83 → = 71

- Empregado A → A T = 71 — 69 = 2
- Empregado B → A T = 83 — 60 = 23

Resumo:

A amplitude total é a medida mais simples de dispersão.

- A desvantagem desta medida de dispersão é que ela leva em conta apenas os valores mínimo e máximo do conjunto. Se ocorrer qualquer variação no interior do conjunto de dados, a amplitude total não nos dá qualquer indicação dessa mudança.
- A amplitude total também sofre a influência de um valor “atípico” na distribuição (um valor muito elevado ou muito baixo em relação ao conjunto).

2.2- Dados agrupados

2.2.1- Sem intervalos de classe

Neste caso, tem-se: AT = x(máx) – x(mín)

Considerando a tabela abaixo.

x_i	0	1	2	3	4
f_i	2	6	12	7	3

AT = 4 – 0 = 4

2.2.2- Com intervalos de classe

Neste caso, a amplitude total é a diferença entre o limite superior da última classe e o limite inferior da primeira classe.

$$\boxed{AT = L(\text{máx.}) - \ell(\text{mín.})}$$

Considerando a tabela abaixo.

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\Sigma = 40$

AT= 174 – 150 = 24, logo AT = 24 cm.

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 variabilidade.

Faz-se uso da amplitude total quando se quer determinar a 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.

3- Desvio Médio (D_m)

Vamos verificar o desvio do valor que representa a produção diária de cada empregado em relação à média aritmética.

O desvio médio é calculado pela média aritmética dos valores absolutos dos desvios.

Exemplo:

Empregado A: 70, 71, 69, 70, 70 → = 70

Empregado B: 60, 80, 70, 62, 83 → = 71

$$D_m = \frac{|70-70| + |71-70| + |69-70| + |70-70| + |70-70|}{5}$$

$$= \frac{0+1+1+0+0}{5} = \frac{2}{5} = 0{,}4$$

$$D_m = \frac{|60-71| + |80-71| + |70-71| + |62-71| + |83-71|}{5}$$

$$= \frac{11+9+1+9+12}{5} = \frac{42}{5} = 8{,}4$$

Há duas medidas estatísticas, a variância e o desvio padrão, que informam sobre a maior ou menor dispersão dos dados em torno da média. Para obter essas medidas de dispersão, parte-se da diferença que cada valor tem em relação à média. Essa diferença chama-se desvio.

O significado do desvio em Estatística é o mesmo atribuído a esse termo na linguagem comum. Quando se diz, por exemplo, que um navio desviou de sua rota, isso significa que havia um percurso a ser seguido e que o navio se desviou dele. Em

Estatística, considerando a Média Aritmética como referência, ela seria o valor provável para todos os dados, mas eles se **desviam** da média.

4- Variância (V_{ar})

O desvio médio é uma boa medida de dispersão porque dá a distância média de cada número em relação à média. No entanto, para muitas finalidades, é mais conveniente elevar ao quadrado cada desvio e tomar a média de todos esses quadrados. Essa grandeza é chamada **variância**.

Exemplo:

Empregado A: 70, 71, 69, 70, 70 → = 70

Empregado B: 60, 80, 70, 62, 83 → = 71

$$V_{ar} = \frac{|70-70|^2 + |71-70|^2 + |69-70|^2 + |70-70|^2 + |70-70|^2}{5}$$

$$= \frac{0^2 + 1^2 + 1^2 + 0^2 + 0^2}{5} = \frac{2}{5} = 0{,}4$$

$$V_{ar} = \frac{|60-71|^2 + |80-71|^2 + |70-71|^2 + |62-71|^2 + |83-71|^2}{5}$$

$$= \frac{11^2 + 9^2 + 1^2 + 9^2 + 12^2}{5} = \frac{121 + 81 + 1 + 81 + 144}{5} = \frac{428}{5}$$

$$= 85{,}6$$

Notas:

- Dado um conjunto de dados, a variância é uma medida de dispersão que mostra o quão distante cada valor desse conjunto está do valor central (médio).
- Quanto menor é a variância, mais próximos os valores estão da média; mas quanto maior ela é, mais os valores estão distantes da média.

Observações:

- Se os valores dos dados se repetirem em todas as amostras, então a variância da amostra será zero.

- Se os dados estiverem muito espalhados, então a variância da amostra acusará um número positivo elevado. Assim, uma grande variância significará uma grande dispersão dos dados em relação à média.

- A variância é uma medida que tem pouca utilidade na estatística descritiva, porém é extremamente importante na inferência estatística e em combinações de amostras.

5- Desvio Padrão (D_P)

5.1-Introdução

A variância é uma boa medida de dispersão, mas tem uma desvantagem: é difícil interpretar o valor numérico da variância. Uma variância de 85,6 significa uma grande dispersão ou uma pequena dispersão? Parte do problema se deve à questão das unidades: a variância é medida em uma unidade que é o quadrado da unidade de medida. Em geral, é mais conveniente calcular a raiz quadrada da variância, chamada **desvio padrão**.

Quanto maior for o desvio padrão, maior será a heterogeneidade entre os valores que estão sendo analisados. Isso significa, portanto, que quanto maior for o desvio padrão, maior será a variação entre os valores.

Exemplo:

Empregado A: 70, 71, 69, 70, 70 → = 70

Empregado B: 60, 80, 70, 62, 83 → = 71

$$V_{ar} = \frac{|70-70|^2 + |71-70|^2 + |69-70|^2 + |70-70|^2 + |70-70|^2}{5}$$

$$= \frac{0^2 + 1^2 + 1^2 + 0^2 + 0^2}{5} = \frac{2}{5} = 0{,}4$$

$$D_P = \sqrt{V_{ar}} = \sqrt{0{,}4} \cong 0{,}63$$

$$V_{ar} = \frac{|60-71|^2 + |80-71|^2 + |70-71|^2 + |62-71|^2 + |83-71|^2}{5}$$

$$= \frac{11^2 + 9^2 + 1^2 + 9^2 + 12^2}{5} = \frac{121 + 81 + 1 + 81 + 144}{5} = \frac{428}{5}$$

$$= 85{,}6$$

$$D_P = \sqrt{V_{ar}} = \sqrt{85{,}6} \cong 9{,}25$$

Observações:

- Quanto menor o desvio padrão, mais os valores da variável se aproximam de sua média.
- Quanto maior o desvio padrão, mais significativa a heterogeneidade entre os elementos de um conjunto, ou seja, maior será a variação entre os valores.

5.2-Aplicação prática do desvio padrão

O desvio padrão é um parâmetro muito usado em Estatística e indica o grau de variação de um conjunto de elementos.

Exemplificando:

Se medirmos a temperatura máxima durante três dias em uma cidade e obtivermos os seguintes valores, 28º, 29º e 30º, podemos dizer que a média desses três dias foi 29º.

Em outra cidade, as temperaturas máximas nesses mesmos dias podem ter sido 22º, 29º e 35º. No segundo caso, a média dos três dias também foi 29º.

As médias têm o mesmo valor, mas os moradores da primeira cidade viveram três dias de calor, enquanto os da segunda tiveram dois dias de calor e um de frio.

Para diferenciar uma média da outra, foi criada a noção de desvio padrão, que serve para dizer o quanto os valores dos quais se extraiu a média são próximos ou distantes da própria média. No exemplo acima, o desvio padrão da segunda cidade é muito maior que o da primeira.

Uma das aplicações mais comuns do desvio padrão é para cálculo da classificação no vestibular. Se dois candidatos ao mesmo curso tiram nota 7 em provas diferentes, o peso desse resultado vai depender do desvio padrão de cada exame. Digamos que a média das notas nas duas provas tenha sido 5. Aquele que obteve 7 na prova cujo desvio padrão foi menor, será mais considerado porque significa que ele conseguiu um 7 em um exame em que quase todo mundo ficou próximo a 5. Enquanto o outro conquistou um 7 em uma prova na qual muitos outros também tiraram notas altas.

5.3-Aplicação da Fórmula de Desvio Padrão

$$D_P = \sqrt{V_{ar}} \;\; \rightarrow \;\; D_P = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

Tanto o desvio padrão como a variância são usados como medidas de dispersão ou variabilidade. O uso de uma ou de outra dependerá da finalidade que se tenha em vista. A variância é uma medida que tem pouca utilidade como estatística descritiva, porém é extremamente importante na inferência estatística e em combinações de amostras.

Se bem que a fórmula dada para o cálculo do desvio seja a que torna mais fácil a sua compreensão, ela não é uma boa fórmula para fins de computação, pois em geral, a média aritmética $(\bar{x})$ é um número fracionário, o que torna pouco prático o cálculo das quantidades $(x_i - \bar{x})^2$.

Os cálculos podem ser simplificados fazendo uso da igualdade:

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

Assim, substituindo $\sum(x_i - \bar{x})^2$ por seu equivalente obtém-se:

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

5.4- Desvio Padrão (D_P) com dados não-agrupados

Exemplo: Considerar o conjunto de valores da variável x: 40,45, 48, 52, 54, 62,70.

O modo mais prático para se obter o desvio padrão é formar uma tabela com duas colunas: uma para x_i e outra para x_i^2. Assim:

x_i	x_i^2
40	1.600
45	2.025
48	2.304
52	2.704
54	2.916
62	3.844
70	4.900
$\Sigma = 371$	$\Sigma = 20.293$

Como n = 7, tem – se:

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

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

Logo, o desvio padrão é 9,49.

5.5- Desvio Padrão (D_P) com dados agrupados

5.5.1- Sem intervalos de classe

Como, neste caso, tem-se a presença de frequências, deve-se levá-las em consideração, resultando a fórmula:

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

Exemplo:

Considerando a distribuição da tabela abaixo, calcular o desvio padrão.

x_i	0	1	2	3	4
f_i	2	6	12	7	3

O modo mais prático para se calcular o desvio padrão é abrir, na tabela dada, uma coluna para os produtos $f_i x_i$ e outra para $f_i\ x_i^2$, lembrando que para se obter $f_i\ x_i^2$, basta multiplicar cada $f_i x_i$ pelo seu respectivo x_i . Assim:

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

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

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

Logo: $D_P = 1.04$.

5.5.2- Com intervalos de classe

Exemplo:

Considerando a distribuição da tabela abaixo, calcular o desvio padrão.

i	ESTATURAS (cm)	f_i	x_i	$f_i x_i$	$f_i x_i^2$
1	150 — 154	4	152	608	92.416
2	154 — 158	9	156	1.404	219.024
3	158 — 162	11	160	1.760	281.600
4	162 — 166	8	164	1.312	215.168
5	166 — 170	5	168	840	141.120
6	170 — 174	3	172	516	88.752
		$\Sigma = 40$		$\Sigma = 6.440$	$\Sigma = 1.038.080$

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

$$D_P = \sqrt{\frac{1038080}{40} - \left(\frac{6440}{40}\right)^2} = \sqrt{25952 - 25921} = \sqrt{31} = 5{,}567$$

Logo, o desvio padrão é 5,57cm.

5.6- Desvio Padrão (D_P) pelo processo breve

Baseados na mudança da variável x por outra y, tal que

$$y_i = \frac{x_i - x_0}{h}$$

pode-se obter um processo breve de cálculo com a aplicação da seguinte fórmula.

$$D_P = \ h.\sqrt{\frac{\sum f_i {y_i}^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

Exemplo:

Considerando a distribuição da tabela abaixo, calcular o desvio padrão.

i	ESTATURAS (cm)	f_i	x_i	y_i	f_iy_i	$f_iy_i^2$
1	150 ⊢ 154	4	152	−2	−8	16
2	154 ⊢ 158	9	156	−1	−9	9
3	158 ⊢ 162	11	160	0	0	0
4	162 ⊢ 166	8	164	1	8	8
5	166 ⊢ 170	5	168	2	10	20
6	170 ⊢ 174	3	172	3	9	27
	h = 4	Σ = 40			Σ = 10	Σ = 80

$$\boldsymbol{D_P = 4\ .\sqrt{\frac{80}{40} - \left(\frac{10}{40}\right)^2}\ = 4.\sqrt{2 - 0{,}0625} = 4.\sqrt{1{,}9375}\ = 4\ x\ 1{,}3919}$$

$$\boldsymbol{= 5{,}5676}$$

Logo DP = 5,57 cm.

Fases para o cálculo do desvio padrão pelo processo breve:

1) Abrimos uma coluna para os valores x_i (ponto médio).

2) Escolhemos um dos pontos médios (de preferência o de maior frequência) para o valor de x_0.

3) Abrimos uma coluna para os valores de y_1 e escrevemos zero na linha correspondente a classe onde se encontra o valor de x_0; a sequência -1, -2, -3, ..., logo acima do zero, e a sequência 1, 2, 3, ..., logo abaixo.

4) Abrimos uma coluna para os valores do produto f_iy_i, conservando os sinais + ou -, e, em seguida, somamos algebricamente esses produtos.

5) Abrimos uma coluna para os valores do produto $f_i y_i^2$, obtidos multiplicando cada $f_i y_i$ pelo seu respectivo y_i, e, em seguida, somamos esses produtos.

6) Aplicamos a fórmula.

Exemplos.

1-Em uma escola, a direção decidiu observar a quantidade de alunos que apresentam todas as notas acima da média em todas as disciplinas. Para analisar melhor, a diretora Ana resolveu montar uma tabela com a quantidade de notas "azuis" em uma amostra de quatro turmas ao longo de um ano. Observe a seguir a tabela organizada pela diretora.

Turmas	Quantidade de alunos acima da média			
	1° Bimestre	2° Bimestre	3° Bimestre	4° Bimestre
6° ano	5	8	10	7
7° ano	8	6	6	12
8° ano	11	9	5	10
9° ano	8	13	9	4

Antes de calcular a variância, é necessário verificar a média aritmética **(x)** da quantidade de alunos acima da média em cada turma.

$$\textbf{6° ano} \rightarrow x = \frac{5 + 8 + 10 + 7}{4} = \frac{30}{4} = 7{,}50.$$

$$\textbf{7° ano} \rightarrow x = \frac{8 + 6 + 6 + 12}{4} = \frac{32}{4} = 8{,}00.$$

$$\textbf{8° ano} \rightarrow x = \frac{11 + 9 + 5 + 10}{4} = \frac{35}{4} = 8{,}75.$$

$$\textbf{9° ano} \rightarrow x = \frac{8 + 13 + 9 + 4}{4} = \frac{34}{4} = 8{,}50.$$

Para calcular a variância da quantidade de alunos acima da média em cada turma, utilizamos uma **amostra**, por isso empregamos a fórmula da **variância amostral**.

$$\text{Var. amostral} = \frac{(x_1 - x)^2 + (x_2 - x)^2 + (x_3 - x)^2 + \dots + (x_n - x)^2}{n - 1}$$

$$\textbf{6° ano} \rightarrow \text{Var} = \frac{(5 - 7{,}50)^2 + (8 - 7{,}50)^2 + (10 - 7{,}50)^2 + (7 - 7{,}50)^2}{4 - 1}$$

$$\text{Var} = \frac{(-2{,}50)^2 + (0{,}50)^2 + (2{,}50)^2 + (-0{,}50)^2}{3}$$

$$\text{Var} = \frac{6{,}25 + 0{,}25 + 6{,}25 + 0{,}25}{3}$$

$$\text{Var} = \frac{13{,}00}{3}$$

Var = 4,33

7° ano → $$\text{Var} = \frac{(8 - 8{,}00)^2 + (6 - 8{,}00)^2 + (6 - 8{,}00)^2 + (12 - 8{,}00)^2}{4 - 1}$$

$$\text{Var} = \frac{(0{,}00)^2 + (-2{,}00)^2 + (-2{,}00)^2 + (4{,}00)^2}{3}$$

$$\text{Var} = \frac{0{,}00 + 4{,}00 + 4{,}00 + 16{,}00}{3}$$

$$\text{Var} = \frac{24{,}00}{3}$$

Var = 8,00

8° ano → $$\text{Var} = \frac{(11 - 8{,}75)^2 + (9 - 8{,}75)^2 + (5 - 8{,}75)^2 + (10 - 8{,}75)^2}{4 - 1}$$

$$\text{Var} = \frac{(2{,}25)^2 + (0{,}25)^2 + (-3{,}75)^2 + (1{,}25)^2}{3}$$

$$\text{Var} = \frac{5{,}06 + 0{,}06 + 14{,}06 + 1{,}56}{3}$$

$$\text{Var} = \frac{20{,}74}{3}$$

Var = 6,91

9° ano → $$\text{Var} = \frac{(8 - 8{,}50)^2 + (13 - 8{,}50)^2 + (9 - 8{,}50)^2 + (4 - 8{,}50)^2}{4 - 1}$$

$$\text{Var} = \frac{(-0{,}50)^2 + (4{,}50)^2 + (0{,}50)^2 + (-4{,}50)^2}{3}$$

$$\text{Var} = \frac{0{,}25 + 20{,}25 + 0{,}25 + 20{,}25}{3}$$

$$\text{Var} = \frac{41{,}00}{3}$$

Var = 13,66

Conhecida a variância de cada turma, vamos calcular agora o desvio padrão.

6° ano	7° ano	8° ano	9° ano
dp = √var **dp = √4,33** **dp ≈ 2,08**	**dp = √var** **dp = √8,00** **dp ≈ 2,83**	**dp = √var** **dp = √6,91** **dp ≈ 2,63**	**dp = √var** **dp = √13,66** **dp ≈ 3,70**

Para concluir sua análise, a diretora pode apresentar os seguintes valores que indicam a quantidade média de alunos acima da média por turma pesquisada.

6° ano: 7,50 ± 2,08 alunos acima da média por bimestre.
7° ano: 8,00 ± 2,83 alunos acima da média por bimestre.
8° ano: 8,75 ± 2,63 alunos acima da média por bimestre.
9° ano: 8,50 ± 3,70 alunos acima da média por bimestre.

2-Considerar os três conjuntos abaixo com seus respectivos valores.

X: 70, 70, 70, 70, 70.

Y: 68, 69, 70, 71, 72.

Z: 5, 15, 50, 120, 160.

- Calcular a média aritmética dos três conjuntos.

Para X: $\bar{x} = \frac{70+70+70+70+70}{5} = \frac{350}{5} = 70$

Para Y: $\bar{x} = \frac{68+69+70+71+72}{5} = \frac{350}{5} = 70$

Para Z: $\bar{x} = \frac{5+15+50+120+160}{5} = \frac{350}{5} = 70$

Obs.: Observa-se que os três conjuntos possuem a mesma média.

Mas também, pode-se observar que o conjunto X é mais homogêneo do que os conjuntos Y e Z; o conjunto Y, por sua vez, é mais homogêneo que o conjunto Z; por fim o conjunto Z é mais heterogêneo de todos. Mesmo possuindo a mesma média, os conjuntos apresentam comportamentos muito diferentes. Isso é chamado dispersão.

No nosso exercício acima, os conjuntos **X**, **Y** e **Z** apresentam como ponto de tendência central para fins de comparação a *média*. Essa *média* é a mesma para os três conjuntos: 70. Assim, o conjunto **X** apresenta *dispersão* nula, pois não há variação dos valores do conjunto em relação a essa média; o conjunto **Y** apresenta *dispersão* menor que o conjunto **Z**; isso porque os valores de **Y** estão mais próximos que os do conjunto **Z**.

Os conjuntos X,Y e Z são de dados não agrupados que serão dispostos em tabelas para melhor visualização.

Tabela X

x_i	x_i^2
70	490
70	490
70	490
70	490
70	490
$\sum = 350$	$\sum = 24500$

Tabela Y

x_i	x_i^2
68	4624
69	4761
70	4900
71	5041
72	5184
$\sum = 350$	$\sum = 24510$

Tabela Z

x_i	x_i^2
5	25
15	225
50	2500
120	14400
160	25600
$\sum = 350$	$\sum = 42750$

Cálculo do desvio padrão:

Para o conjunto X:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 2450 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24500}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4900 - 4900} = 0$$

Para o conjunto Y:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 24510 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24510}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4902 - 4900} = \sqrt{2} = 1,4$$

Para o conjunto Z:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 42750 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{42750}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{8550 - 4900} = \sqrt{3650} = 60,4$$

Obs.:

- O desvio padrão do conjunto **X** é igual a 0. De fato, isso significa que não há variação alguma no conjunto, portanto, é um conjunto *homogêneo*;
- O desvio padrão do conjunto **Y** é igual a 1,4 e o do conjunto **Z** é igual a 60,4. Comparando-se os dois conjuntos, vemos que há uma pequena variação em **Y** (1,4) e uma alta variação em **Z** (60,4). Na prática, significa que os valores do conjunto **Y** estão mais próximos da *média*, ao passo que, em **Z**, os valores do conjunto estão muito distantes da *média*.

3- Pesquisa realizada pela Secretaria de Saúde de uma cidade, visando conhecer os hábitos de higiene bucal da população, identificou num de seus itens o tipo de creme dental mais consumido e tabelou os seguintes dados.

Tipo de creme dental	Flúor	Bicarbonato de sódio	Menta e flúor	Flúor e bicarbonato de sódio
Número de pessoas	80	20	60	40

Calcule

a) a média aritmética dos valores.

$$Média = \frac{80 + 20 + 60 + 40}{4} = 50$$

b) o desvio médio.

$$D_M = \frac{|80 - 50| + |20 - 50| + |60 - 50| + |40 - 50|}{4} = 20$$

c) a variância

$$V_{ar} = \frac{(80 - 50)^2 + (20 - 50)^2 + (60 - 50)^2 + (40 - 50)^2}{4} = 500$$

d) o desvio padrão

$$D_P = \sqrt{500} \cong 22{,}3$$

4- Uma distribuidora pesquisou o consumo de refrigerantes entre diferentes faixas etárias, para melhor direcionar a sua campanha publicitária.

Idade dos consumidores	Número de consumidores
10 ⊢ 14	60
14 ⊢ 18	100
18 ⊢ 22	130
22 ⊢ 26	90
26 ⊢ 30	20
Total	400

Baseado nos dados, calcule

a) a idade média dos consumidores.

Idade dos consumidores	Ponto Médio	Frequência
10|--14	12	60
14|--18	16	100
18|--22	20	130
22|--26	24	90
26|--30	28	20
Total	-	400

Considerando a frequência de cada intervalo e o respectivo ponto médio, temos:

$$Média = \frac{12.60 + 16.100 + 20.130 + 24.90 + 28.20}{400} = 19{,}1$$

b) 0 desvio médio

$$D_M = \frac{60.|12 - 19{,}1| + 100.|16 - 19{,}1| + 130.|20 - 19{,}1| + 90.|24 - 19{,}1| + 20.|28 - 19{,}1|}{400} = 3{,}68.$$

c) a variância

$$V_{ar} = \frac{60(12 - 19{,}1)^2 + 100(16 - 19{,}1)^2 + 130(20 - 19{,}1)^2 + 90(24 - 19{,}1)^2 + 20(28 - 19{,}1)^2}{400} \cong 19{,}59$$

d) o desvio padrão

$$D_P = \sqrt{V_{ar}} = \sqrt{19{,}59} \cong 4{,}4$$

5- A tabela a seguir mostra as notas dos alunos de uma classe de Ensino Médio em Matemática.

Notas	F
0 ⊢ 2,0	3
2,0 ⊢ 4,0	9
4,0 ⊢ 6,0	16
6,0 ⊢ 8,0	8
8,0 ⊢ 10,0	4
Total	40

Calcule

a) a média aritmética.

b) a variância.

c) o desvio padrão.

d) construa um polígono de frequência e analise a zona de normalidade dos dados.

a) Para calcular a média aritmética, precisamos inicialmente encontrar os pontos médios das classes.

Nota	f	Pm
0 ⊢ 2,0	3	1,0
2,0 ⊢ 4,0	9	3,0
4,0 ⊢ 6,0	16	5,0
6,0 ⊢ 8,0	8	7,0
8,0 ⊢ 10,0	4	9,0

$$\bar{x} = \frac{3 \cdot 1 + 9 \cdot 3 + 16 \cdot 5 + 8 \cdot 7 + 4 \cdot 9}{40} = \frac{202}{40} = 5{,}05$$

b) Para calcular a variância, primeiramente calculamos o desvio, fazendo

d= ponto médio – média aritmética

$$1 - 5{,}05 = -4{,}05 \cong -4$$

$$3 - 5{,}05 = -2{,}05 \cong -2$$

$$5 - 5{,}05 = -0{,}05 \cong 0$$

$$7 - 5{,}05 = 1{,}95 \cong 2$$

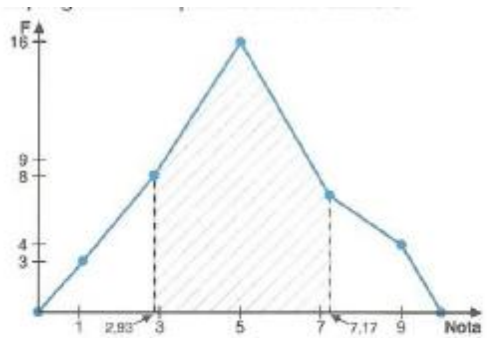
$$9 - 5{,}05 = 3{,}95 \cong 4$$

A variância então será dada por

$$V_{ar} = \frac{3.(-4)^2 + 9.(-2)^2 + 16.0^2 + 8.2^2 + 4.4^2}{40} = \frac{180}{40} = 4{,}5$$

c) $D_P = \sqrt{4{,}5} \cong 2{,}12$

d)



A zona de normalidade é uma região que se define em torno da média quando fazemos:
desvio padrão – média = 5,05 – 2,12 = 2,93
desvio padrão – média = 5,05 + 2,12 = 7,17
e é representada no gráfico pela parte hachurada.

6- Coeficiente de Variação (CV)

O desvio padrão por si só não nos diz muita coisa. Assim, um desvio padrão de duas 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. Além disso, o fato de o desvio padrão ser expresso na mesma unidade dos dados limita o seu emprego quando desejamos comparar duas ou mais séries de valores, relativamente à sua dispersão ou variabilidade, quando expressas em unidades diferentes.

Para contornar essas dificuldades e limitações, podemos caracterizar a dispersão ou variabilidade dos dados em termos relativos a seu valor médio, medida essa denominada **coeficiente de variação** (**CV**).

Exemplo:

Considerando a distribuição da tabela abaixo, na qual a média é 161 cm e o desvio padrão é igual a 5,57 cm, calcular o coeficiente de variação (CV).

i	ESTATURAS (cm)	f_i	x_i	y_i	$f_i y_i$	$f_i y_i^2$
1	150 ⊢ 154	4	152	−2	−8	16
2	154 ⊢ 158	9	156	−1	−9	9
3	158 ⊢ 162	11	160	0	0	0
4	162 ⊢ 166	8	164	1	8	8
5	166 ⊢ 170	5	168	2	10	20
6	170 ⊢ 174	3	172	3	9	27
	h = 4	Σ = 40			Σ = 10	Σ = 80

$$CV = \frac{Desvio\ Padrão}{Média}\ x\ 100 \rightarrow CV = \frac{5,57}{161}\ x\ 100 = 0,03459\ x\ 100 = 3,459$$

Logo CV = 3,5%.

Exemplo:

Considerar os resultados das medidas das estaturas e dos pesos de um mesmo grupo de indivíduos.

	X	s
ESTATURAS	175 cm	5,0 cm
PESOS	68 kg	2 kg

Tem-se

CV_E = 5 / 175 x 100 = 2,85%

CV_P = 2 / 68 x 100 = 2,94%

Logo, nesse grupo de indivíduos, os pesos apresentam maior grau de dispersão que as estaturas.

Nota:

- Se bem que, para qualificar a dispersão de uma distribuição, seja mais proveitoso o coeficiente de variação, não devemos deduzir daí que a variância e o desvio padrão careçam de utilidade. Pelo contrário, são medidas muito úteis no tratamento de assuntos relativos à inferência estatística.

Observações:

O coeficiente de variação fornece a variação dos dados obtidos em relação à média. Quanto menor for o seu valor, mais homogêneos serão os dados. O coeficiente de variação é considerado baixo (apontando um conjunto de dados bem homogêneos) quando for menor ou igual a 25%. O fato de o coeficiente de variação ser dado em valor relativo nos permite comparar séries de valores que apresentam unidades de medida distintas.

Exemplo:

Compare a variabilidade relativa do tempo de reação de um analgésico A com a variabilidade do peso das pessoas que se submeteram à dosagem desse analgésico. As médias e os desvios padrão foram.

Analgésico A: Média =3 min e D_P = 0,71

Peso das pessoas: Média =58,25 kg e D_P = 5,17

Solução: Vamos calcular o coeficiente de variação para cada item observado.

Cálculo para o tempo de reação do analgésico.

$$Cv = 100 \cdot \frac{0,71}{3} = 23,67\%$$

Cálculo para o peso das pessoas.

$$Cv = 100 \cdot \frac{5,17}{58,25} = 8,88\%$$

Comparando o coeficiente de variação do tempo de reação do analgésico e o do peso das pessoas, podemos concluir que os dados referentes ao peso são mais consistentes que os dados referentes ao tempo de reação do analgésico, ou ainda, que os dados referentes ao peso são mais homogêneos que os do tempo de reação do analgésico.

Observações:

Diz- se que uma distribuição tem:

Baixa dispersão: CV $\leq$ 15%
Média dispersão: 15%< CV<30%
Alta dispersão: CV $\geq$ 30%

Um coeficiente de variação maior ou igual a 30% revela que a série é heterogênea e a média tem pouco significado. Se o coeficiente de variação for menor que 30%, portanto a série é homogênea e a média tem grande significado.

CAPÍTULO 3 – Medidas de Assimetria / Medidas de Curtose

1- Medidas de Assimetria

1.1-Introdução

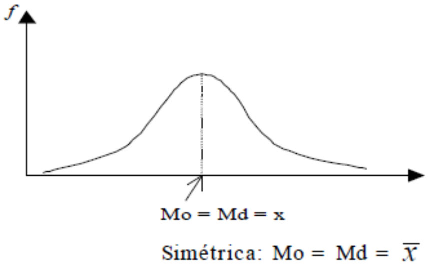
As medidas de assimetria indicam o grau de assimetria de uma distribuição de frequências unimodal em relação a uma linha vertical que passa por seu ponto mais elevado.

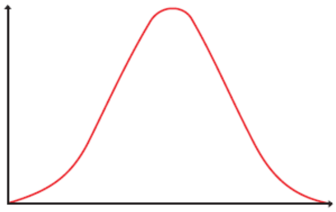
Sendo a distribuição simétrica, a média e a moda coincidem; sendo a distribuição assimétrica à esquerda ou negativa, a média é menor que a moda; e sendo assimétrica à direita ou positiva, a média é maior que a moda.

- Uma distribuição de frequências com classes é simétrica quando:

Média = Mediana = Moda.

Graficamente, uma distribuição simétrica tem associada a si uma curva de frequências unimodal apresentando duas “caudas” simétricas em relação a uma linha vertical que passa por seu ponto mais alto (eixo de simetria).

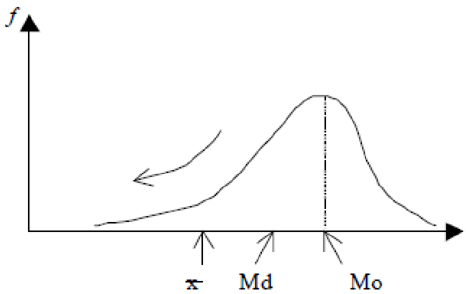


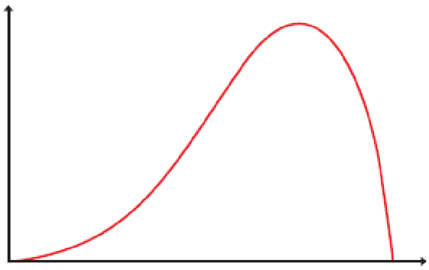


Uma distribuição assimétrica tem associada a si uma curva de frequências unimodal que apresenta a partir do seu ponto mais alto, uma “cauda” mais longa para a direita (assimetria positiva) ou para a esquerda (assimetria negativa).

Nas distribuições assimétricas os valores da moda, da mediana e da média divergem sendo que a média sempre estará do mesmo lado que a cauda mais longa.

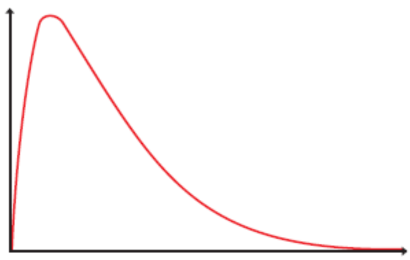
- Uma distribuição de frequências com classes é assimétrica à esquerda ou negativa quando: **Média < Mediana < Moda.**

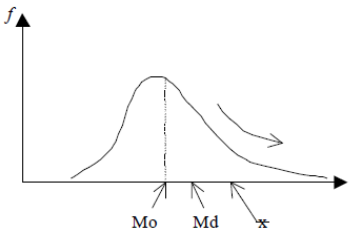




Assimétrica à esquerda (Negativa): .... $\bar{x}$ à *esquerda* da Mo ($\bar{x}$ < Md < Mo)

- Uma distribuição de frequências com classes é assimétrica à direita ou positiva quando: **Média > Mediana > Moda.**





Assimétrica à direita (Positiva): $\overline{X}$ à ***direita*** da Mo (Mo < Md < $\overline{X}$)

Para fazer a classificar do tipo de assimetria através de cálculo, tem-se a seguinte relação:

$$\overline{x} = Mo$$

Se: $\overline{x}$ - Mo = 0 ⟶ assimetria nula ou distribuição simétrica.
$\overline{x}$ - Mo < 0 ⟶ assimetria negativa ou à esquerda.
$\overline{x}$ - Mo > 0 ⟶ assimetria positiva ou à direita.

Exemplos:

1-Determinar os tipos de assimetria das distribuições abaixo.

Distribuição A	
Classes	fi
2 ⊢ 6	6
6 ⊢ 10	12
10 ⊢ 14	24
14 ⊢ 18	12
18 ⊢ 22	6
Total Σ	60

Distribuição B	
Classes	fi
2 ⊢ 6	6
6 ⊢ 10	12
10 ⊢ 14	24
14 ⊢ 18	30
18 ⊢ 22	6
Total Σ	78

Distribuição C	
Classes	fi
2 ⊢ 6	6
6 ⊢ 10	30
10 ⊢ 14	24
14 ⊢ 18	12
18 ⊢ 22	6
Total Σ	78

Calcular a média ($\bar{x}$), a mediana (Md), a moda (Mo) e o desvio padrão(s) para cada distribuição.

Distribuição A ⟶ $\overline{x}$ = 12, Md = 12, Mo = 12 kg e s = 4,42

Distribuição B ⟶ $\overline{x}$ = 12,9, Md = 13,5 , Mo = 16 kg e s = 4,20

Distribuição C ⟶ $\overline{x}$ = 11,1, Md = 10,5 , Mo = 8 kg e s = 4,20

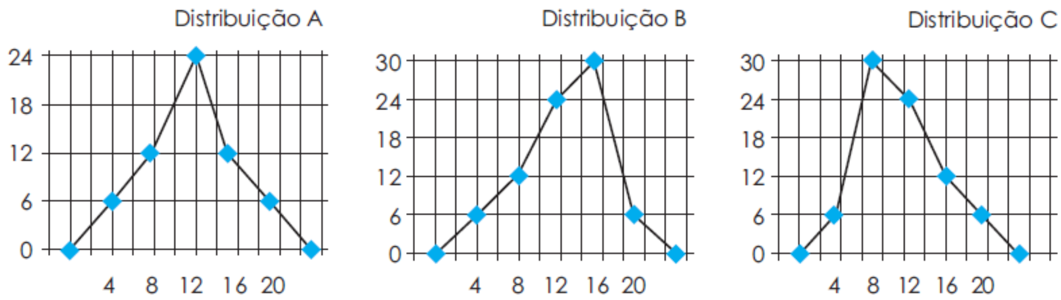
Calcular o tipo de assimetria.

Distribuição A ⟶ $\overline{x}$ - Mo = 12 - 12 = 0 ⟶ a distribuição é simétrica.

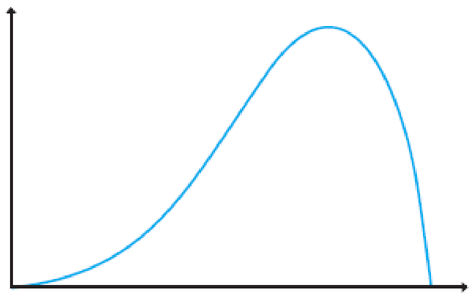
Distribuição B ⟶ $\overline{x}$ - Mo = 12,9 – 16 = - 3,1 ⟶ a distribuição é assimétrica negativa.

Distribuição C ⟶ $\overline{x}$ - Mo = 11,1 – 8 = 3,1 ⟶ a distribuição é assimétrica positiva.

Construir os gráficos das distribuições anteriores.



2- Dada a figura a seguir, pode-se afirmar que



A) a moda é maior do que a mediana e menor do que a média.

B) a moda é menor do que a mediana e maior do que a média.

C) a moda é menor do que a mediana e menor do que a média.

D) a mediana é maior do que a média e menor do que a moda.

E)nda

Resposta: Letra B.

1.2- Coeficiente de Assimetria

Um coeficiente de assimetria quantifica o desvio de uma distribuição em relação a uma distribuição simétrica e o sinal resultante do seu cálculo indica o tipo de assimetria da distribuição.

Coeficiente de Assimetria de Person

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

Ou

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

Obs.: Desvio Padrão: Dp ou s

Se 0,15 < |As| < 1, a assimetria é considerada **moderada**; |As| > 1, é **forte**.

Exemplo:

Calcular o coeficiente de assimetria.

DISTRIBUIÇÃO A

PESOS (kg)	f_i
2 ⊢ 6	6
6 ⊢ 10	12
10 ⊢ 14	24
14 ⊢ 18	12
18 ⊢ 22	6
	Σ = 60

$\bar{x}$ = 12 kg
Md = 12 kg
Mo = 12 kg
s = 4,42 kg

DISTRIBUIÇÃO B

PESOS (kg)	f_i
2 ⊢ 6	6
6 ⊢ 10	12
10 ⊢ 14	24
14 ⊢ 18	30
18 ⊢ 22	6
	Σ = 78

$\bar{x}$ = 12,9 kg
Md = 13,5 kg
Mo = 16 kg
s = 4,20 kg

DISTRIBUIÇÃO C

PESOS (kg)	f_i
2 ⊢ 6	6
6 ⊢ 10	30
10 ⊢ 14	24
14 ⊢ 18	12
18 ⊢ 22	6
	Σ = 78

$\bar{x}$ = 11,1 kg
Md = 10,5 kg
Mo = 8 kg
s = 4,20 kg

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

$$As_A = \frac{3(12 - 12)}{4{,}42} = 0 \Rightarrow \text{simetria}$$

$$As_B = \frac{3(12{,}9 - 13{,}5)}{4{,}20} = -0{,}429 \Rightarrow \text{assimetria negativa}$$

$$As_C = \frac{3(11{,}1 - 10{,}5)}{4{,}20} = 0{,}429 \Rightarrow \text{assimetria positiva}$$

2- Medidas de Curtose

2.1- Introdução

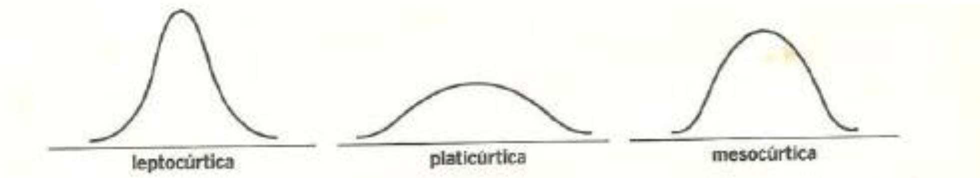
Denomina-se curtose o grau de achatamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal (curva correspondente a uma distribuição teórica de probabilidade).

Embora seja comum explicar a curtose como o "grau de achatamento" de uma distribuição de frequências, o que as medidas de curtose buscam indicar realmente é o grau de concentração de valores da distribuição em torno do centro desta distribuição.

Numa distribuição unimodal, quanto maior for a concentração de valores em torno do centro da mesma, maior será o valor de sua curtose. Graficamente, isso será associado a uma curva com a parte central mais afilada, mostrando um pico de frequência simples mais destacado, mais pontiagudo, caracterizando a moda da distribuição de forma mais nítida.

Uma distribuição de frequência pode ser classificada como

- **Mesocúrtica** – quando apresenta uma medida de curtose igual à da distribuição normal.
- **Platicúrtica** – quando apresenta uma medida de curtose menor que a da distribuição normal.
- **Leptocúrtica** – quando apresenta uma medida de curtose maior que a da distribuição normal.



2.2- Coeficiente de Curtose

Uma fórmula para a medida da curtose é

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Essa fórmula é conhecida como **coeficiente percentílico de curtose**. Esse coeficiente é definido como o quociente entre a amplitude semi-interquatílica e a amplitude entre o 10º e o 90º percentis.

O valor deste coeficiente para a curva normal é 0,26367...., C $\cong 0{,}263$.

Assim sendo, ao se calcular o coeficiente percentílico de curtose de uma distribuição qualquer tem-se

- Quando C $\cong 0{,}263 \rightarrow$ diz-se que a distribuição é mesocúrtica.
- Quando C $< 0{,}263 \rightarrow$ diz-se que a distribuição é platicúrtica.
- Quando C > $0{,}263 \rightarrow$ diz-se que a distribuição é leptocúrtica.

Exemplo:

Sabendo-se que uma distribuição apresenta as medidas a seguir:

Q_1=24,4 cm; Q_3 = 41,2 cm ; P_{10} = 20,2 cm e P_{90} = 49,5 cm, tem – se:

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

$$C = \frac{41{,}2 - 24{,}4}{2(49{,}5 - 20{,}2)} = \frac{16{,}8}{58{,}6} = 0{,}2866 \rightarrow C = 0{,}287$$

Como

0,287 > 0,263, conclui-se que a distribuição é platicúrtica, em relação à normal.

CAPÍTULO 4 – Probabilidade

1-Introdução

Embora o cálculo das probabilidades pertença ao campo da Matemática, sua inclusão neste curso se justifica pelo fato de a maioria dos fenômenos de que trata a Estatística ser de natureza aleatória ou probabilística. Consequentemente, o conhecimento dos aspectos fundamentais do cálculo de probabilidades é uma necessidade essencial para o estudo de Estatística Indutiva ou Inferencial. Lembrando, também, que a Estatística é um ramo da Matemática.

O estudo das probabilidades foi motivado inicialmente pelos jogos, encontrando posteriormente aplicações em outros campos, como a genética, a medicina, a economia, a política e outros setores da atividade humana em que há necessidade de prever a ocorrência de determinado fato.

2- A linguagem das probabilidades

A teoria das Probabilidades utiliza uma linguagem muito própria que é preciso assimilar.

- **Experimento aleatório**

Experimento aleatório é todo experimento que, mesmo repetido, várias vezes, sob condições semelhantes, apresenta resultados imprevisíveis, dentre os resultados possíveis.

Exemplos:

a) Lançamento de uma moeda.
b) Lançamento de um dado.
c) Loteria de números.
d) Extração de uma carta de baralho.
e) Abrir um livro ao acaso e ver o número de página.
f) Escolher um aluno ao acaso e perguntar-lhe quantos irmãos tem.

- **Espaço Amostral**

Espaço amostral de um experimento aleatório é o conjunto de todos os resultados possíveis desse experimento. **Notação: S**

Exemplos:

a) No lançamento de uma moeda, temos S = { cara, coroa}
b) No lançamento de um dado (perfeito), temos S = {1,2,3,4,5,6}

- **Evento**

Evento é todo subconjunto de um espaço amostral S de um experimento aleatório. **Notação: E**

Exemplos:

a) No lançamento de duas moedas

E_1: aparecem faces iguais

$E_1 = \{(c,c), (k,k)\}$; portanto, o número de elementos do evento E_1 é $n(E_1) = 2$.

E_2: aparece cara em pelo menos 1 face

E_2 = {(c,c), (c,k), (k,c), }; portanto, o número de elementos do evento E_2 é n (E_2) = 3.

b) No lançamento de dois dados

E_1: aparecem números iguais

E_1 = {(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)}; portanto, o número de elementos do evento E_1 é n (E_1) = 6.

E_2: o primeiro número é menor ou igual a 2

E_2 = {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)}; portanto, o número de elementos do evento E_2 é n (E_2) = 12.

Alguns eventos particulares, através de exemplos:

- ✓ **Evento certo**: evento que possui os mesmos elementos do espaço amostral, (E = S).

  E_3: a soma dos resultados nos dois dados é menor ou igual a 12.

- ✓ **Evento impossível**: evento igual ao conjunto vazio.

  E_4: o número do primeiro dado é igual a 7.

  E_4 = **Ø**

- ✓ **Evento simples (ou elementar)**: evento que possui um único elemento.

  E_5: a soma dos resultados nos dois dados é igual a 12.

  E_5 = {(6,6)}

- ✓ **Evento complementar**: se A é um evento de um espaço amostral S, o evento complementar de A indicado por A', A^c ou $\overline{A}$ é tal, que $A^c = S - A$.

  Ex.: a) A probabilidade de se realizar um evento é 1/5, a probabilidade de que ele não ocorra é 1 – 1/5 = 4/5.

  b) Sabendo que a probabilidade de tirar o 4 no lançamento de um dado é 1/6. Logo, a probabilidade de não tirar o 4 no lançamento de um dado é 1 – 1/6 = 5/6.

- ✓ **Eventos mutuamente exclusivos:** quando a realização de um evento exclui a realização do(s) outro(s). Assim, no lançamento de uma moeda, o evento “tirar cara” e o evento “tirar coroa” são mutuamente exclusivos, já que, ao se realizar um deles, o outro não se realiza. Se dois eventos são mutuamente exclusivos, a probabilidade de que um **ou** outro se realize é igual à soma das probabilidades de que cada um se realize. $p = p_1 + p_2$.

  Ex.: Lançamos um dado. A probabilidade de se tirar o 3 ou o 5 é: p = 1/6 + 1/6 = 2/6 = 1/3, pois os dois eventos são mutuamente exclusivos.
  Dois eventos são mutuamente exclusivos quando a ocorrência de um implica a não-ocorrência do outro. Se A e B são eventos mutuamente exclusivos, então $A \cap B = \emptyset$.

- ✓ **Eventos independentes:** Dois eventos independentes quando a realização ou a não-realização de um dos eventos não afeta a probabilidade da realização do outro e vice-versa. Quando lançamos dois dados, o resultado obtido em um deles independe do resultado no outro. Se dois eventos são independentes, a probabilidade de que se realizem simultaneamente é igual ao produto das probabilidades de realização dos dois eventos. $p = p_1 \times p_2$. Ex.: No lançamento de dois dados, qual é a probabilidade de se obter 1 no primeiro dado e 5 no segundo dado? p = 1/6 x 1/6 = 1/36.

3- Probabilidade

Dado um experimento aleatório, sendo S o seu espaço amostral, vamos admitir que todos os elementos de S tenham a mesma chance de acontecer, ou seja , que S é um conjunto equiprovável.

Chamamos de probabilidade de um evento A (A C S) o número real P(A), tal que $P(A) = \frac{n(A)}{n(S)}$, onde:
n(A) = é o número de elementos do evento A.
n(S) = é o número de elementos do espaço amostral S.

Ou melhor, se em um fenômeno aleatório as possibilidades são igualmente prováveis, então a probabilidade de ocorrer um evento A é

$$P(A) = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}}$$

Exemplos:

Considerando o lançamento de um dado, calcular

a) a probabilidade do evento A "obter um número ímpar na face superior".

S= {1,2,3,4,5,6} →n(S)= 6

A= {1,3,5} → n(A)= 3

$P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{3}{6} = \frac{1}{2}$ ou 50%

b) a probabilidade do evento B "obter um número menor ou igual a 6 na face superior".

S= {1,2,3,4,5,6} →n(S)= 6

B= {1,2,3,4,5,6} → n(B)= 6

$P(B) = \frac{n(B)}{n(S)} \rightarrow P(B) = \frac{6}{6} = 1$

c) a probabilidade do evento C "obter o número 4 na face superior".

S= {1,2,3,4,5,6} →n(S)= 6

C= {4} → n(C)= 1

$P(C) = \frac{n(C)}{n(S)} \rightarrow P(C) = \frac{1}{6}$

d) a probabilidade do evento D "obter um número maior que 6 na face superior".

S= {1,2,3,4,5,6} →n(S)= 6

D= { } ou ∅ → n(D)= 0

$P(D) = \frac{n(D)}{n(S)} \rightarrow P(D) = \frac{0}{6} = 0$

Pelos exemplos, pode-se concluir que, sendo n(S) = n:

a) a probabilidade de um **evento E qualquer** (E C S) é um número P(E), tal que 0≤P(E)≤1.

b) a probabilidade do **evento certo** é igual a 1: P(S) = 1.

c) a probabilidade de um **evento elementar E qualquer** é, lembrando que n(E) 1: $P(E) = \frac{1}{n}$.

d) a probabilidade do **evento impossível** é igual a zero: P(∅)= 0.

4-Exercícios Resolvidos

1- Qual é a probabilidade de sair o ás de ouros quando retiramos uma carta de um baralho de 52 cartas?

Como só há um ás de ouros, o número de elementos do evento é 1. Logo:

n(S)= 52 n(A)= 1 $P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{1}{52}$

2- Qual é a probabilidade de sair um rei quando retiramos uma carta de um baralho de 52 cartas?

Como há 4 reis (de copas, de ouros, de paus , de espadas), o número de elementos do evento é 4. Logo:

n(S)= 52 n(A)= 4 $P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{4}{52} = \frac{1}{13}$

3- Em um lote de 12 peças, 4 são defeituosas. Sendo retirada uma peça, calcule:

a) a probabilidade de essa peça ser defeituosa.

n(S)= 12 n(A)= 4 $P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{4}{12} = \frac{1}{3}$

b) a probabilidade de essa peça não ser defeituosa.

Sendo este evento e o anterior complementares, temos: $P = 1 - \frac{1}{3} = \frac{2}{3}$.

4- No lançamento de dois dados, calcule a probabilidade de se obter soma igual a 5.

n(S) = 6 x 6 = 36

n(A) = {(1,4),(2,3),(3,2),(4,1)} → n(A) = 4

$$P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{4}{36} = \frac{1}{9}$$

5- De dois baralhos de 52 cartas retiram-se, simultaneamente, uma carta do primeiro baralho e uma carta do segundo. Qual a probabilidade de a carta do primeiro baralho ser um rei e a do segundo ser o 5 de paus?

$P_1 = \frac{4}{52} = \frac{1}{13}$ e $P_2 = \frac{1}{52}$

Como esses dois acontecimentos são independentes e simultâneos, vem:

$P = \frac{1}{13} . \frac{1}{52} = \frac{1}{676}$

6- Uma urna A contém: 3 bolas brancas, 4 pretas, 2 verdes; uma urna B contém: 5 bolas brancas, 2 pretas, 1 verde; uma urna C contém: 2 bolas brancas, 3 pretas, 4 verdes. Uma bola é retirada de cada urna. Qual é a probabilidade de as três bolas retiradas da primeira, segunda e terceira urnas serem, respectivamente, branca, preta e verde?

$P_1 = \frac{3}{9} = \frac{1}{3} \quad P_2 = \frac{2}{8} = \frac{1}{4} \quad P_3 = \frac{4}{9}$

Como os três eventos são independentes e simultâneos, vem:

$$P = \frac{1}{3} . \frac{1}{4} . \frac{4}{9} = \frac{1}{27}$$

7- De um baralho de 52 retiram-se, ao acaso, duas cartas sem reposição. Qual é a probabilidade de a primeira carta ser o ás de paus e a segunda ser o rei de paus?

A probabilidade de sair o ás de paus na primeira carta é: $P_1 = \frac{1}{52}$.

Após a retirada da primeira carta, restam 51 cartas no baralho, já que a carta retirada não foi reposta. Assim, a probabilidade de a segunda ser o rei de paus é : $P_2 = \frac{1}{51}$.

Como esses eventos são independentes, temos:

$$P = \frac{1}{52} . \frac{1}{51} = \frac{1}{2652}$$

8- Qual é a probabilidade de sair uma figura quando retiramos uma carta de um baralho de 52 cartas?

$P_{reis} = \frac{4}{52} = \frac{1}{13} \quad P_{damas} = \frac{4}{52} = \frac{1}{13} \quad P_{valetes} = \frac{4}{52} = \frac{1}{13}$

Como os eventos são mutuamente exclusivos, vem:

$$P = \frac{1}{13} + \frac{1}{13} + \frac{1}{13} = \frac{3}{13}$$

Obs.: Este problema pode ser resolvido, ainda, com o seguinte raciocínio. Como em um baralho temos 12 figuras (4 damas, 4 valetes, 4 reis), vem: $P = \frac{12}{52} = \frac{3}{13}$

9-Qual é a probabilidade de sair uma carta de copas ou de ouros quando retiramos uma carta de um baralho de 52 cartas?

$P_{copas} = \frac{13}{52} = \frac{1}{4} \qquad P_{ouros} = \frac{13}{52} = \frac{1}{4}$

Como os eventos são mutuamente exclusivos, vem:

$$P = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}$$

10-No lançamento de um dado, qual é a probabilidade de se obter um número não-inferior a 5?

A probabilidade de se obter um número não inferior a 5 é a probabilidade de se obter 5 ou 6. Assim:

$$P = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

11- Qual é a probabilidade de obter ao menos um 6 na jogada de dois dados?

Sejam, o evento A, 6 na primeira jogada, e o evento B, 6 na segunda jogada. Então o evento ao menos 1 seis é A U B. Sabe-se que P(A) = P(B) = 6/36. Como A ∩ B é o evento 6 em ambos os dados, sabe-se que sua probabilidade é 1/36. Aplicando a fórmula P(A U B) = P(A) + P(B) – P(A ∩ B), obtém-se:

$$P(AUB) = \frac{6}{36} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}$$

12- Lançando um dado duas vezes, observa-se os pares ordenados de números das faces superiores. Qual é a probabilidade de ocorrência do número 5 em pelo menos uma vez.

n(S) = 6 x 6 = 36

A= {(1,5), (2,5),(3,5),(4,5),(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,5)} → n(A) = 11

$$P(A) = \frac{n(A)}{n(S)} \rightarrow P(A) = \frac{11}{36}$$

CAPÍTULO 5 – Distribuição Binomial

Considerando-se um experimento aleatório, observa-se a probabilidade de ocorrer um evento E (sucesso), assim como seu complementar E’ (insucesso), em n tentativas independentes. A probabilidade de ocorrerem k sucessos e (n – k) fracassos é dada pelo termo geral do Binômio de Newton $(p + q)^n$.

$P = \binom{n}{k}. p^k. q^{n-k}$, sendo p a probabilidade de sucesso em cada tentativa e q =1– p a probabilidade de fracasso.

“Sucesso” e “fracasso” aqui apenas apresentam ocorrências que se excluem e se complementam.

- Se ocorre um sucesso, não ocorre um fracasso, e vice-versa.
- Sucesso e fracasso cobrem todas as possibilidades, não há ocorrência diferente dessas.

Exemplos:

a) Uma moeda é lançada 6 vezes seguidas. Determine a probabilidade de se obter cara 4 vezes nos 6 lançamentos.

Sendo p a probabilidade de se obter cara, p = ½ , e q a probabilidade de se obter coroa, q = 1 – ½ = ½ . Como n = 6 e k = 4, tem – se :

$$P = \binom{n}{k}.p^k.q^{n-k} \to P = \binom{6}{4}.\left(\frac{1}{2}\right)^4.\left(\frac{1}{2}\right)^2 \to P = \frac{6!}{4!\,(6-4)!}.\frac{1}{16}.\frac{1}{4}$$

$$\to P = \frac{6.5.4!}{4!\,2!}.\frac{1}{64} \to P = \frac{15}{64}$$

b) Qual é a probabilidade de uma família com seis filhos ter dois filhos homens, supondo-se que a probabilidade de que nasça menino ou menina seja igual:

Dados: n= 6 k=2 p= ½

$$P = \binom{n}{k}.p^k.q^{n-k} \to P = \binom{6}{2}.\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^4 \to P = \frac{6!}{2!\,(6-2)!}.\frac{1}{4}.\frac{1}{16}$$

$$\to P = \frac{6.5.4!}{2!\,4!}.\frac{1}{64} \to P = \frac{15}{64}$$

c) Jogando 5 vezes um dado honesto, qual a probabilidade de ocorrer só três vezes o resultado 2?

Dados: n=5 k=3 p= 1/6 q= 1 – 1/6 = 5/6

$$P = \binom{n}{k}.p^k.q^{n-k} \to P = \binom{5}{3}.\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^2 \to P = \frac{5!}{3!\,(5-3)!}.\frac{1}{216}.\frac{25}{36}$$

$$\to P = \frac{5.4.3!}{3!\,2!}.\frac{25}{7776} \to P = \frac{500}{15552} = \frac{125}{3888} = 0{,}032 = 3{,}2\%$$

d) Dois times de futebol A e B disputam 6 partidas. Qual é a probabilidade de o time A ganhar 4 partidas?

Dados: n=6 k=4 p= 1/3 e q=2/3

$$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{6}{4}.\left(\frac{1}{3}\right)^4.\left(\frac{2}{3}\right)^2 \rightarrow P = \frac{6!}{4!\,(6-4)!}.\frac{1}{81}.\frac{4}{9}$$

$$\rightarrow P = \frac{6.5.4!}{4!\,2!}.\frac{4}{729} \rightarrow P = \frac{120}{1458} = \frac{60}{729} = \frac{20}{243} = 0{,}0823 = 8{,}23\%$$

e) Sabendo-se que a probabilidade de uma pessoa acertar um tiro no alvo é ¼ , qual é a probabilidade de acertar pelo menos um tiro em 4 tentativas?

Dados: n = 4 p= ¼ q= ¾

- 1 acerto, k = 1

$$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{4}{1}.\left(\frac{1}{4}\right)^1.\left(\frac{3}{4}\right)^3 \rightarrow P_1 = \frac{108}{256}.$$

- 2 acertos, k = 2

$$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{4}{2}.\left(\frac{1}{4}\right)^2.\left(\frac{3}{4}\right)^2 \rightarrow P_2 = \frac{54}{256}$$

- 3 acertos, k = 3

$$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{4}{3}.\left(\frac{1}{4}\right)^3.\left(\frac{3}{4}\right)^1 \rightarrow P_3 = \frac{12}{256}$$

- 4 acertos, k = 4

$$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{4}{4}.\left(\frac{1}{4}\right)^4.\left(\frac{3}{4}\right)^0 \rightarrow P_4 = \frac{1}{256}$$

$$\boldsymbol{P} = \frac{108}{256} + \frac{54}{256} + \frac{12}{256} + \frac{1}{256} = \frac{175}{256} = 0{,}683 = 68{,}3\%$$

f) Suponha que numa linha de produção, a probabilidade de se obter uma peça defeituosa é igual a 0,1. Toma-se uma amostra de 10 peças para serem inspecionadas. Qual é a probabilidade de se obter uma peça defeituosa?

Dados: n=10 k=1 p= 0,1 q= 1 – 0,1 = 0,9

$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{10}{1}.(0,1)^1.(0,9)^9 \rightarrow P = 10\ x\ 0,1\ \ x\ 0,3874$ = 0,3874= 38,74%

g) Suponha que numa linha de produção, a probabilidade de se obter uma peça defeituosa é igual a 0,1. Toma-se uma amostra de 10 peças para serem inspecionadas. Qual é a probabilidade de se obter nenhuma peça defeituosa?

Dados: n=10 k=0 p= 0,1 q= 1 – 0,1 = 0,9

$P = \binom{n}{k}.p^k.q^{n-k} \rightarrow P = \binom{10}{0}.(0,1)^0.(0,9)^{10} \rightarrow P = \frac{10!}{0!(10-0)!}\ x\ 1\ \ x\ 0,3486$

= 1 x 1x 0,3486 = 0,3486 = 34,86%

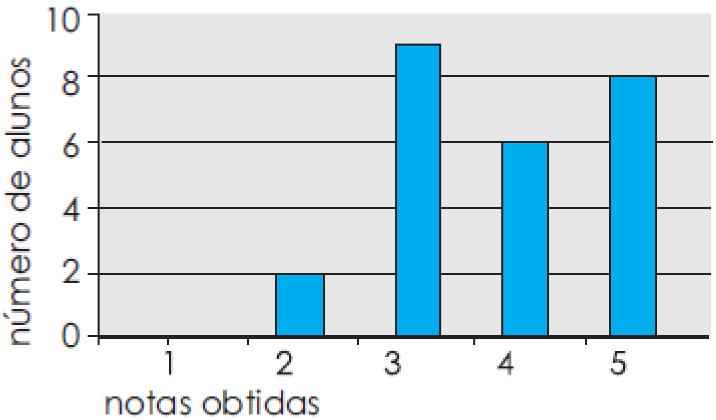
Avaliação

Esta avaliação corresponde a 100% da nota do segundo módulo.

Cursista: ____

1ª Questão

O professor Carlos aplicou uma prova de Matemática a 25 alunos, contendo 5 questões, valendo 1 ponto cada. Após fazer a correção, o professor construiu o gráfico abaixo, que relaciona o número de alunos às notas obtidas por eles.



Observando o gráfico, conclui-se que a média, a moda e mediana das notas obtidas pelos alunos correspondem, respectivamente, a

A) 3,0; 3,0; 3,0

B) 3,8; 3,0; 4,0

C) 3,8; 3,8; 3,8

D) 4,0; 4,0; 4,0

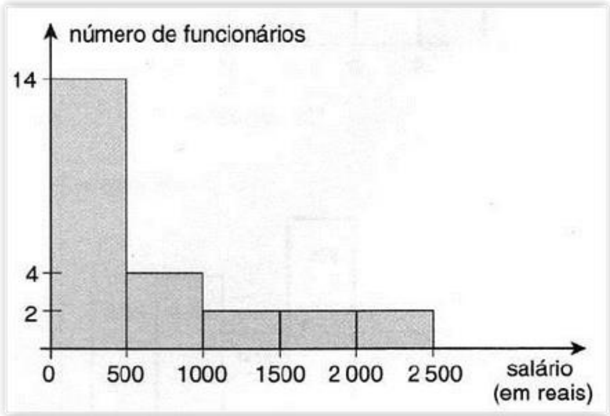
E) nda

2ª Questão

Alê, Bia e Célia têm a mesma idade. A soma dessas idades com as de Dora (13 anos), Érica (18 anos) e Fausto (20 anos) é 96 anos. Determine a moda, a média aritmética e a mediana dessas seis idades.

3ª Questão:

O gráfico abaixo apresenta a distribuição de frequência das faixas salariais numa pequena empresa.



Com os dados disponíveis, pode-se concluir que a média aritmética desses salários é, aproximadamente

A) R$420,00

B) R$536,00

C) R$562,00

D) R$640,00

E) R$708,00

4ª Questão:

Quando a mediana é melhor do que a média como medida do valor típico em um grupo? Responda e justifique com um exemplo.

5ª Questão:

Ao calcular a média aritmética das notas dos Testes Físicos (TF) de suas três turmas, um professor de Educação Física anotou os seguintes valores.

TURMA	Nº DE ALUNOS	MÉDIA DO TF
A	20	9
B	40	7,5
C	30	8

A média aritmética das notas do TF dos 90 alunos das turmas A, B e C é
A) 8,0 B) 8,1 C) 8,2 D) 8,3 E)nda

6ª Questão:

Considerar a seguinte situação: O dono de uma empresa pretende saber, em média, quantos produtos são produzidos por cada funcionário em um dia. O chefe tem conhecimento que nem todos conseguem fazer a mesma quantidade de peças, mas pede que seus funcionários façam um registro de sua produção em uma semana de trabalho. Ao fim desse período, chegou-se à seguinte tabela.

Funcionários	Quantidade de peças produzidas por dia				
	Segunda	Terça	Quarta	Quinta	Sexta
A	10	9	11	12	8
B	15	12	16	10	11
C	11	10	8	11	12
D	8	12	15	9	11

Calcular a média aritmética relativa a cada funcionário.

7ª Questão:

Uma caixa contém 30 bolinhas numeradas de 1 a 30. Retirando-se ao acaso uma bolinha da urna, qual a probabilidade de essa bolinha ter um número múltiplo de 4 ou de 3?

8ª Questão:

Você é o gerente de uma pizzaria e mantém o controle das vendas dos diversos tipos de pizza. Suponha que tenha observado os seguintes valores de vendas diárias de pizzas do tipo calabresa, durante um período de 9 dias: 40, 56, 38, 38, 63, 59, 52, 49,46. Calcule o coeficiente de variação para verificar se a dispersão é muito grande em relação à média.

9ª Questão:

Determinada editora pesquisou o número de páginas das revistas mais vendidas em uma cidade.

Revistas	A	B	C	D	E	F
Nº.de páginas	62	90	88	92	110	86

Calcular

a) o número médio de páginas.

b) a variância.

c) o desvio padrão.

10ª Questão:

Após um ano de funcionamento, uma maternidade registrou o nascimento de 720 crianças, em parto normal. Os dados referentes à altura dessas crianças permitiu a construção da tabela abaixo.

Calcular

a) a altura média.

b) o desvio médio.

c) a variância.

d) o desvio padrão.

Altura (em cm)	Número de crianças
45 ⊢— 47	80
47 ⊢— 49	260
49 ⊢— 51	200
51 ⊢— 53	160
53 ⊢— 55	20
Total	720