

Estatística é a parte da Matemática que fornece métodos para a coleta, organização, apresentação, análise e interpretação de dados para a tomada de decisões.

Aplicações da estatística:

- _ Um professor comunica que a nota média da classe foi 7;
- _ O meteorologista informa que a probabilidade de chover hoje é de 30%;
- _ Um fabricante testa a resistência à ruptura, de cintos de segurança de automóveis, sem destruir toda a sua produção.

População: em estatística, é o conjunto formado por todos os elementos (pessoas, objetos, etc.) que contém pelo menos uma característica comum a qual temos interesse em estudar.

Amostra: em estatística, é uma parte da população retirada para ser analisada, a qual permite que se conheça tal população.

Variáveis ou Dados: os dados ou variáveis são as informações obtidas através de observações, medidas, respostas de pesquisas ou contagens em geral. Esses dados ou variáveis podem ser classificados em:

- _ **numéricos** ou **QUANTITATIVOS**;
- _ **categóricos** ou **QUALITATIVOS**.

As variáveis quantitativas podem ser **contínuas** ou **discretas**. E as variáveis qualitativas podem ser **nominais** ou **ordinais**.

Variáveis Quantitativas Contínuas (QC): podem assumir qualquer valor numérico num intervalo contínuo. Os dados referentes a tais variáveis dizem-se **dados contínuos**. Ou seja, quando pode assumir qualquer valor dentro de dois limites definidos, números “quebrados”, por exemplo: pesos de peças fabricadas, temperatura do corpo humano, etc.

Variáveis Quantitativas Discretas (QD): assumem valores numéricos inteiros. Os **dados discretos** são o resultado da contagem do número de itens. Ou seja, quando só pode assumir valores pertencentes a um conjunto enumerável, números inteiros, como por exemplo, quantidade de peças fabricadas, número de filhos, etc.

Variáveis Qualitativas Ordinais (QO): consistem de valores relativos (numéricos ou não) atribuídos para denotar ordem. Os dados referentes a tais variáveis dizem-se **dados ordinais**. Ou seja, apresenta uma ordenação, por exemplo: grau de escolaridade (Fundamental, Médio, Superior), nota obtida numa prova (de ZERO a DEZ ou de A até E ou de MB até I), etc.

Variáveis Qualitativas Nominais (QN): referem-se a avaliações subjetivas. Os dados referentes a tais variáveis dizem-se **dados nominais**. Ou seja, não apresentam ordem, nem estrutura numérica, como por exemplo, religião, cor dos olhos, etc.

A dificuldade para classificar dados, se dá em função da fácil confusão gerada entre uma **variável quantitativa discreta** e uma **variável qualitativa ordinal**.

Por exemplo, num questionário estatístico, a pergunta: grau de importância que você dá ao seu curso (de 0 a 10) é uma variável qualitativa ordinal **QO**. Outro exemplo: soma da renda familiar (até \$ 1.000,00, entre \$ 1.000,00 e \$ 2.000,00, acima de \$ 2.000,00), é variável numérica, mas quando se pede para encaixar numa categoria, é classificada como variável qualitativa ordinal **QO**.

Exemplo 1: Considere as afirmações a seguir e assinale a alternativa correta.

- I. A qualidade de um produto, defeituoso ou não defeituoso, trata de um dado qualitativo.

II. A altura dos atletas do time de basquetebol da escola “Aprender”, trata de um dado qualitativo.

III. O diâmetro dos parafusos produzidos por certa máquina trata de um dado quantitativo.

- a) Todas as afirmações estão corretas.
- b) Apenas a afirmação I está correta.
- c) Apenas as afirmações I e III estão corretas.
- d) Todas as afirmações estão incorretas.
- e) Apenas a afirmação III está correta.

Resposta Correta: C a altura dos atletas do time de basquetebol da escola “Aprender” trata de um dado QUANTITATIVO e não QUALITATIVO.

Exemplo 2: Classifique as variáveis abaixo em Qualitativa Nominal (N), Qualitativa Ordinal (O), Quantitativa Discreta (D) e Quantitativa Contínua (C).

a) profissão; b) idade; c) tempo de vida; d) peso; e) estado civil; f) tipo de escola (pública/particular); g) número de alunos numa classe; h) disciplina que mais gosta; i) rua de uma residência; j) número de uma residência.

Respostas: a)N b)D c)C d)C e)N f)N g)D h)N i)N j)O

ORGANIZAÇÃO DE DADOS

Quando os dados são coletados para uma pesquisa, são chamados de **dados brutos**. Um exemplo de dado bruto corresponde ao valor médio (em dólares) de comercialização nos últimos 10 meses da saca de soja, na Bolsa de Cereais, conforme apresentado abaixo:

9,0 - 8,0 - 8,0 - 2,0 - 6,3 - 6,5 - 6,8 - 7,0 - 7,1 - 7,1

Geralmente, este tipo de dado traz pouca ou nenhuma informação ao leitor, sendo necessário organizar os dados, com o intuito de aumentar sua capacidade de informação.

Rol: A primeira forma de organização que vamos estudar é o Rol, que são os dados organizados em ordem crescente ou decrescente.

2,0 – 6,3 – 6,5 – 6,8 – 7,0 – 7,1 – 7,1 – 8,0 – 8,0 – 9,0

Como podemos observar, a simples organização dos dados em um **rol** aumenta muito a capacidade de informação destes. Pode-se verificar facilmente que **o menor preço observado foi 2 dólares** e **o maior, 9 dólares**, o que nos fornece uma **amplitude total** de variação da ordem de **7 dólares**. Amplitude total corresponde à diferença entre o maior e o menor valor observado em um conjunto de dados, simbolizado por **A**. Outra informação que podemos obter nos dados por meio da organização em rol crescente, é que alguns preços, como **7,1 e 8,0, foram os mais frequentes**, ou seja, os mais citados na pesquisa.

Tabela: Para organizar os dados de uma forma mais eficiente, na qual se possa apresentar uma quantidade maior de informações, podemos usar as tabelas. Os elementos básicos de uma tabela são: o título, o corpo e a fonte. Quando temos poucos valores, podemos agrupá-los numa tabela simples. Por exemplo:

Valor Médio da saca de soja

Valor	Frequência
2,0	1
6,3	1
6,5	1

6,8	1
7,0	1
7,1	2
8,0	2
9,0	1

Fonte: Bolsa de Cereais

Quando temos muitos valores, fica inviável a tabela simples, dessa forma, os agrupamos numa tabela com intervalos de classe. Classes são intervalos nos quais os valores da variável analisada são agrupados (linhas da tabela). Distribuindo-se os dados observados em **classes** e contando-se o número de observações contidas em cada classe, obtém-se a **frequência de classe**. A disposição tabular dos dados agrupados em classes, juntamente com as frequências correspondentes, se denomina **distribuição de frequências**. A partir dos dados do exemplo relativo ao preço da saca de soja, vamos construir uma distribuição de frequência:

Valor Médio da saca de soja

CLASSES		FREQUÊNCIA
2	4	1
4	6	0
6	8	6
8	10	3

Fonte: Bolsa de Cereais

Na tabela acima, dizemos que o limite inferior e o superior da segunda classe são 4 e 6. O **ponto médio (PM)** da primeira classe é 3. O ponto médio é a média aritmética entre o Li: Limite inferior; e o Ls: Limite superior.

TIPOS DE FREQUÊNCIA

Como já vimos, após a coleta dos dados, não temos informações claras. Ou seja, na tabela abaixo, temos os dados brutos ou uma tabela primitiva, pois os dados não estão organizados.

5,1	4,9	4,9	5,1	4,7
5,0	5,0	5,0	5,1	5,4
5,2	5,2	4,9	5,3	5,0
4,5	5,4	5,1	4,7	5,5
4,8	5,1	5,3	5,3	5,0

Na tabela anterior, é difícil averiguar, qual o Menor valor, o Maior valor, a Faixa de valores, a Amplitude, etc. Por isso, é melhor organizarmos a tabela acima, num rol.

4,5	4,7	4,7	4,8	4,9
4,9	4,9	5,0	5,0	5,0
5,0	5,0	5,1	5,1	5,1
5,1	5,1	5,2	5,2	5,3
5,3	5,3	5,4	5,4	5,5

Através do Rol, fica fácil averiguar, o Menor valor (4,5), o Maior valor (5,5), a Faixa de valores (4,5 a 5,5) e a Amplitude (1).

Frequência simples ou absoluta (f): É a quantidade de vezes que um dado aparece, ou seja, a frequência absoluta, ou apenas frequência, de um valor é o número de vezes que uma determinada variável assume esse valor. Ao conjunto das frequências dos diferentes valores

da variável dá-se o nome de distribuição da frequência (ou apenas distribuição). conforme tabela abaixo:

valor	f
4,5	1
4,6	0
4,7	2
4,8	1
4,9	3
5,0	5
5,1	5
5,2	2
5,3	3
5,4	2
5,5	1
total	25

Na tabela acima, temos uma observação direta, do número de vezes (frequência) que cada valor aparece. Quando uma tabela possui muitas linhas, podemos transformá-la de simples em intervalos de classe, conforme abaixo:

Valor	frequência (f)
4,5 4,9	4
4,9 5,3	15
5,3 5,7	6
Total	25

OBS.: _ A escolha do intervalo de classe (0,4) geralmente é arbitrário, embora possa ser definido por diferentes métodos de cálculo, como o método de Sturges.

_ O símbolo $|$ significa intervalo fechado à esquerda e aberto à direita, ou seja:

$4,5| 4,9$ significa $4,5 \leq x < 4,9$ ou $[4,5 ; 4,9 [$

_ Na tabela simples, percebe-se que não há nenhum resultado com 4,6, mas na tabela com intervalos de classe, não observamos este detalhe. Ou seja, a tabela simples é mais detalhada que a tabela com intervalos de classe.

Frequência relativa (fr): São os valores das razões entre as frequências simples e a frequência total $fr = f / \Sigma f$. **Ou seja, a frequência relativa**, é a porcentagem relativa à frequência. O propósito das frequências relativas é o de permitir a análise ou facilitar as comparações, pois multiplicando a frequência relativa por cem, temos o percentual de cada dado.

$$fr = f / \Sigma f \leftrightarrow fr = 1 / 25 \leftrightarrow fr = 0,04$$

valor	f	fr
4,5	1	0,04
4,6	0	0
4,7	2	0,08
4,8	1	0,04
4,9	3	0,12

5,0	5	0,20
5,1	5	0,20
5,2	2	0,08
5,3	3	0,12
5,4	2	0,08
5,5	1	0,04
total	25	1

Através da tabela acima e análise de dados, 8% das amostras apresentam o valor 5,2.

Frequência acumulada (F): também chamada de **fa**, é o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe **$F = \sum f$** . **Ou seja, a frequência acumulada** de um valor, é o numero de vezes que uma variável assume um valor inferior ou igual a esse valor.

$$F_3 = \sum f \leftrightarrow F = 1 + 0 + 2 \leftrightarrow \underline{F = 3}$$

$$F_6 = 1 + 0 + 2 + 1 + 3 + 5 \leftrightarrow \underline{F = 12}$$

valor	f	fr	F
4,5	1	0,04	1
4,6	0	0	1
4,7	2	0,08	3
4,8	1	0,04	4
4,9	3	0,12	7
5,0	5	0,20	12
5,1	5	0,20	17
5,2	2	0,08	19
5,3	3	0,12	22
5,4	2	0,08	24
5,5	1	0,04	25
total	25	1	-

Através da tabela acima e análise de dados, há 3 resultados com valores $\leq 4,7$, há 12 resultados com valores ≤ 5 .

Frequência acumulada relativa (Fr): É a porcentagem relativa à frequência acumulada. Ou seja, a frequência relativa acumulada de uma classe é a frequência acumulada da classe, dividida pela frequência total da distribuição **$Fr = F / \sum f$** .

$$Fr_1 = F / \sum f \leftrightarrow Fr = 1 / 25 \leftrightarrow \underline{Fr = 0,04}$$

$$Fr_4 = 4 / 25 \leftrightarrow \underline{Fr = 0,16}$$

valor	f	fr	F	Fr
4,5	1	0,04	1	0,04
4,6	0	0	1	0,04
4,7	2	0,08	3	0,12
4,8	1	0,04	4	0,16
4,9	3	0,12	7	0,28
5,0	5	0,20	12	0,48
5,1	5	0,20	17	0,68
5,2	2	0,08	19	0,76
5,3	3	0,12	22	0,88

5,4	2	0,08	24	0,96
5,5	1	0,04	25	<u>1</u>
total	25	1	-	-

Através da tabela acima e análise de dados, 96% das amostras apresentam valores $\leq 5,4$.

CONTRIBUIÇÕES PERCENTUAIS

Note que $f\% = fr \cdot 100$ e $F\% = Fr \cdot 100$, pois se multiplicarmos a frequência relativa por 100, obtemos a mesma na forma percentual.

Uma tabela de frequências, para variáveis quantitativas, apresenta, porém, outros conceitos que permitem uma maior profundidade para análise e devem ser adicionadas. São eles o PONTO MÉDIO (PM) e a FREQUENCIA PERCENTUAL ACUMULADA (F%).

PM é o valor médio de cada classe, ou intervalo, é o ponto médio de cada classe. Torna-se o valor representativo de cada classe.

F% é a frequência percentual acumulada, obtida repetindo-se o primeiro valor de $f\%$ e somando aos demais.

Veja como exemplo, uma tabela de frequências para a variável quantitativa “idades” de forma completa:

Classes	PM	f	f%	F%
12 - 14	13	3	15	15
14 - 16	15	5	25	40
16 - 18	17	5	25	65
18 - 20	19	7	35	100
Total	-	20	100	-

Amplitude em tabelas de frequências: a amplitude de um rol é a diferença entre o maior e o menor valor. Numa tabela de frequências, temos a amplitude de cada classe, a amplitude total das classes, a amplitude dos pontos médios e a amplitude das frequências. Por exemplo, na tabela acima, temos que a amplitude de cada classe é $14 - 12 = \underline{2}$, a amplitude total das Classes ou amplitude da distribuição é $20 - 12 = \underline{8}$, a amplitude dos pontos médios é $19 - 13 = \underline{6}$ e a amplitude das frequências é $7 - 3 = \underline{4}$.

Exemplo: um grupo de alunos foi consultado sobre o time paulista de sua preferência, e os votos foram registrados assim: Santos: II, Palmeiras: IIII, Corinthians: IIIIIII, São Paulo: IIIIII. Construa a tabela de frequência correspondente a essa pesquisa.

Times	Contagem	f	f%	F%
Santos	II	2	10	10
Palmeiras	IIII	4	20	30
Corinthians	IIIIIIII	8	40	70
São Paulo	IIIIII	6	30	100
Total		20	100	-

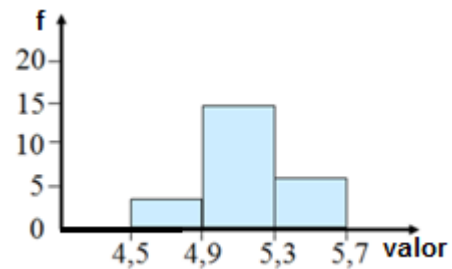
REPRESENTAÇÕES GRÁFICAS

Os resultados de uma pesquisa estatística podem ser apresentados em forma de ROL, de TABELA ou de GRÁFICO.

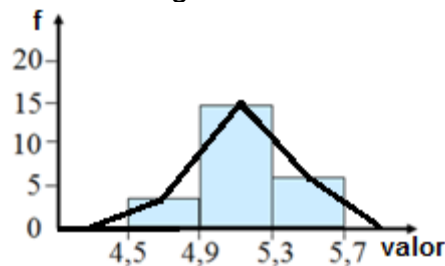
Os Gráficos Estatísticos são importantes ferramentas para analisar e interpretar dados numéricos relativos a uma pesquisa, possibilitando melhor visualização.

Uma tabela de distribuição de frequência pode ser representada através de um gráfico chamado Histograma, conforme abaixo:

valor	frequência (f)
4,5 - 4,9	4
4,9 - 5,3	15
5,3 - 5,7	6
total	25



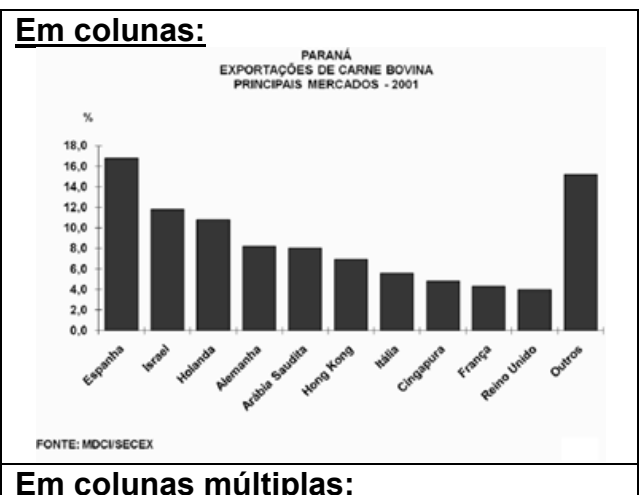
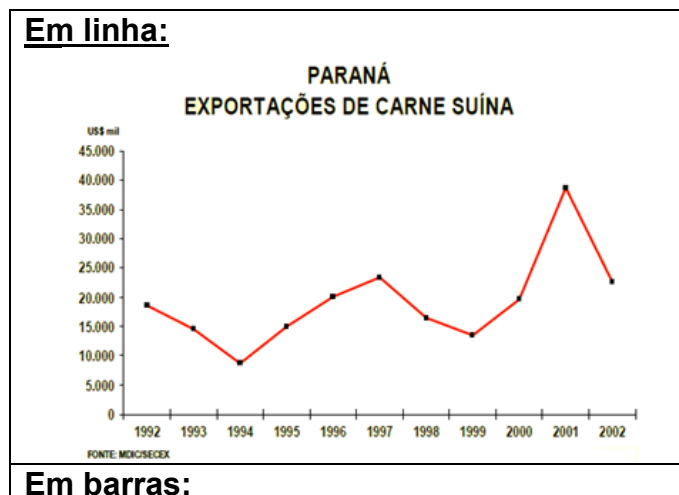
Polígono de frequências: é um gráfico em linha, obtido unindo-se por segmentos de reta os pontos médios das bases superiores dos retângulos de um histograma. Para realmente obtermos um polígono (linha poligonal fechada) devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição. O polígono do histograma será usado em assuntos posteriores.

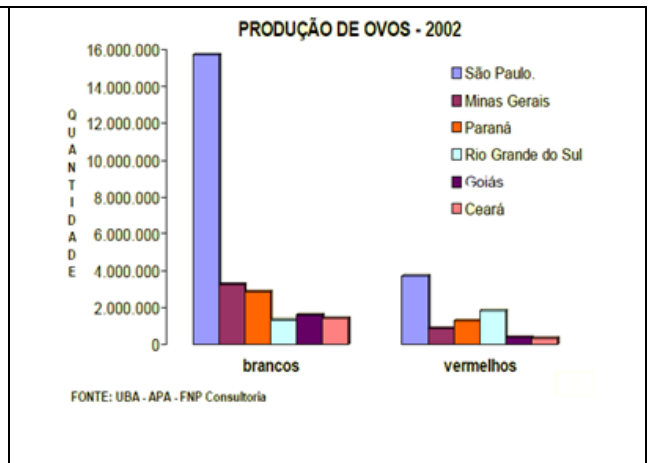
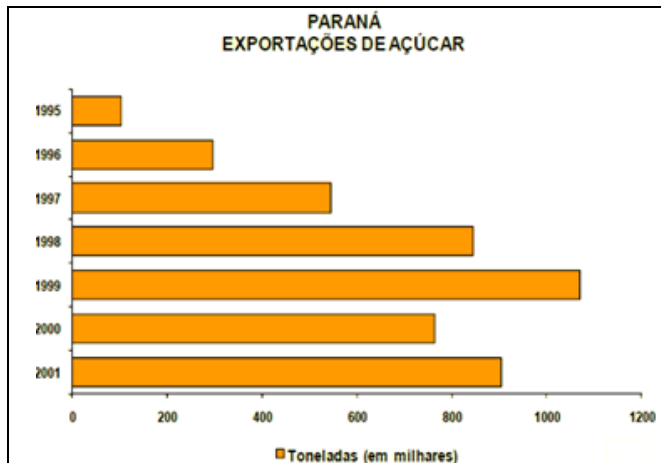


Principais Tipos de Gráficos Estatísticos: os principais tipos de gráficos são:

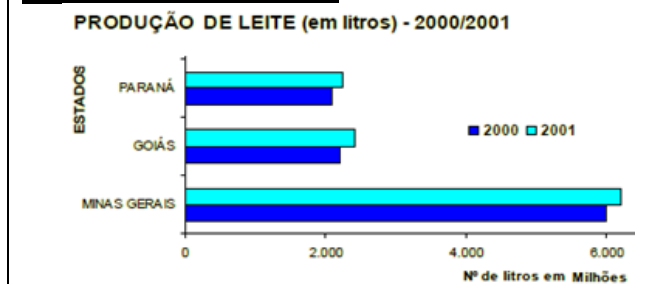
- Gráfico em linha ou em curva;
- Gráfico em colunas ou em barras;
- Gráfico em colunas ou em barras múltiplas;
- Gráfico em setores.

Veja os exemplos a seguir:

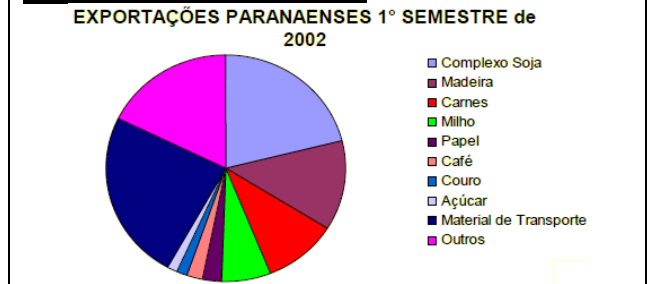




Em barras múltiplas:



Em setores circulares:



MEDIDAS DE TENDÊNCIA CENTRAL

As **Medidas de Posição**, também denominadas de **medidas de tendência central**, são as medidas que representam os fenômenos pelos seus valores médios, em torno dos quais tendem a concentrar-se os dados. São usadas para indicar um valor que resume um conjunto de números. As mais utilizadas são a média, a mediana e a moda.

Média Aritmética (μ ou $\bar{x}$): É a soma de todos os resultados obtidos dividido pela quantidade de valores. Utiliza-se a letra grega mu “ μ ” (leia-se “mi”) para a média de uma população de N elementos. E, a média de uma amostra de n elementos é representada pelo símbolo “ $\bar{x}$ ” (leia-se “xis barra”).

$$\text{Fórmulas: } \bar{x} = \frac{\sum x}{n} \quad \mu = \frac{\sum x}{N}$$

Quando o exercício não mencionar se os dados são amostrais ou populacionais, usaremos o símbolo $\bar{x}$, pois quase a totalidade das estatísticas são feitas através de dados amostrais.

A média possui várias propriedades matemáticas, a que considero mais interessante é: “somando-se uma constante a cada valor, a média ficará aumentada do valor dessa constante. O mesmo ocorre com as operações de subtração, multiplicação e divisão.”

Exemplo 1: Determinar a média aritmética dos valores amostrais: 5, 8, 10, 12 e 15.

$$\bar{X} = \frac{\sum X_i}{n} \Rightarrow \bar{X} = \frac{5+8+10+12+15}{5} \rightarrow \bar{X} = \frac{50}{5} \Rightarrow \bar{X} = 10$$

Exemplo 2: Em uma empresa de componentes eletrônicos, a exportação nos últimos 4 anos, em milhares de dólares, foi US\$ 800,00; US\$ 880,00; US\$ 760,00 e US\$ 984,00. Determine a média de exportações dessa empresa.

$$\bar{X} = \frac{\sum X}{n} \Rightarrow \bar{X} = \frac{800+880+760+984}{4} \rightarrow \bar{X} = \frac{3424}{4} \Rightarrow \bar{X} = 856$$

Média Aritmética Ponderada (μ_p ou $\bar{x}_p$): A fórmula anterior para calcular a média aritmética supõe que cada observação tem a mesma importância. A Média ponderada é uma média aritmética na qual será atribuído um peso a cada valor da série.

$$\bar{X}_p = \frac{\sum_{i=1}^n x_i \cdot p_i}{\sum_{i=1}^n p_i}$$

Exemplo 1: Um professor de Matemática adotou para 2013 os seguintes pesos para as notas bimestrais:

1º bimestre: peso 1 3º bimestre: peso 3
2º bimestre: peso 2 4º bimestre: peso 4

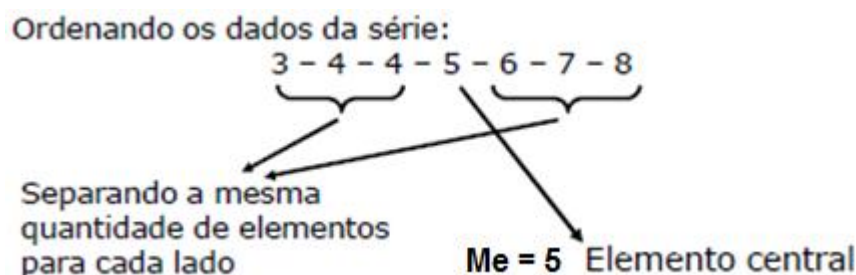
Qual será a média de um aluno que obteve as seguintes notas: 5, 4, 3 e 2 nos respectivos bimestres?

$$\bar{X}_p = \frac{(5.1)+(4.2)+(3.3)+(2.4)}{10} = \frac{5+8+9+8}{10} = \frac{30}{10} = 3$$

Mediana (Me ou Md): É o valor que ocupa a posição central de um conjunto de dados, desde que estejam colocados ordenadamente, seja em ordem crescente ou em ordem decrescente.

Exemplo 1: Calcular a mediana dos dados: 5 ; 8 ; 4 ; 6 ; 7 ; 3 ; 4.

OBS.: Quando a quantidade de dados for ímpar, o valor da mediana será dado pelo valor central da série de dados.



Exemplo 2: Calcular a mediana dos dados: 8; 0; 7; 4; 7; 10; 6; 5.

OBS.: Quando a quantidade de dados for par, o valor da mediana será dado pela média dos dois valores centrais da série de dados.

Ordenando os dados da série:

0 - 4 - 5 - 6 - 7 - 7 - 8 - 10

Separando a mesma
quantidade de elementos
para cada lado

Elementos centrais

$$Me = \frac{6+7}{2} \Rightarrow Me = 6,5$$

Moda (Mo): É o valor que ocorre com maior frequência nos dados de uma pesquisa. Ou seja, É o valor que aparece a maior quantidade de vezes. É a única que pode ser usada para dados nominais.

Exemplos 1: Determinar a moda dos dados: 4 4 5 5 5 6 7 8 9

No grupo de dados acima, o valor que mais aparece é o valor 5, então, a moda vale 5. $Mo = 5$, nesse caso dizemos que a série é unimodal.

Exemplos 2: Determinar a moda dos dados: 10 10 10 15 15 15 17 18 19 19

No grupo de dados acima, os valores que mais aparecem são o 10 e o 15, então, a moda vale 10 e 15. **$Mo = 10$ e 15** , nesse caso dizemos que a série é **bimodal**.

Exemplos 3: Determinar a moda dos dados: 100 200 300 400 500 600 700

No grupo de dados acima, não há repetição de valores, então não existe moda, nesse caso dizemos que a série é **amodal**.

OBS.: Uma série pode ser: amodal, unimodal, bimodal, trimodal e acima disso, polimodal.

Medidas Centrais (Dados Agrupados SEM Intervalos de Classes)

Moda: A Moda para dados agrupados sem intervalos de classes, é calculada observando-se o maior valor da frequência.

Exemplo: Calcular a moda dos valores representados na distribuição de frequências:

x_i	f
5	3
6	5
7	11
8	10
9	5

Moda

Maior frequência

O valor da moda é $Mo = 7$.

Média: A Média para dados agrupados sem intervalos de classes, é calculada de maneira análoga a média ponderada.

Exemplo: Calcular a média das idades representadas na distribuição de frequências:

Idade de um grupo de pessoas

Idade	f
4	4
5	6
6	6
7	4
$\Sigma f = 20$	

Resolução:

$$\bar{X} = \frac{\Sigma(X_i \cdot f_i)}{\Sigma f} \Rightarrow \bar{X} = \frac{4 \cdot 4 + 5 \cdot 6 + 6 \cdot 6 + 7 \cdot 4}{20}$$

$$\bar{X} = \frac{16 + 30 + 36 + 28}{20} \Rightarrow \bar{X} = \frac{110}{20}$$

$$\bar{X} = 5,5$$

Resposta: A média das idades é de 5,5 anos.

Mediana: Para dados agrupados sem intervalos de classe, identifica-se a frequência acumulada imediatamente superior à metade da soma das frequências. A mediana será o valor da variável que corresponde a tal frequência acumulada.

Exemplo: Calcular o valor da mediana da distribuição dada:

Número de meninos	f
0	2
1	6
2	10
3	12
4	4
$\Sigma f = 34$	

Como existem 34 valores, a mediana é calculada através do 17º e do 18º valor, observe que ambos aparecem na variável 2, logo **Me = 2**. Para facilitar a compreensão, vou detalhar o procedimento:

Número de meninos	f
0	2
1	6
2	10
3	12
4	4
$\Sigma f = 34$	

Calculando as frequências acumuladas:

Número de meninos	f	fa
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
$\Sigma f = 34$		

Encontrando a classe mediana:

$$\frac{\Sigma f}{2} = \frac{34}{2} = 17$$

Número de meninos	f	fa
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
		$\Sigma f = 34$

Número de meninos	Variável correspondente à classe mediana	
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
		$\Sigma f = 34$

Resposta: a mediana é igual a 2, ou seja, $Me = 2$.

Medidas Centrais (Dados Agrupados COM Intervalos de Classes)

Média: A média para dados agrupados com intervalos de classes é calculada de maneira análoga a média ponderada, utilizando-se os pontos médios.

Exemplo: Calcular a média das idades representadas na distribuição de frequências:

Idades	f
18 — 21	9
21 — 24	12
24 — 27	12
27 — 30	17
30 — 33	16
33 — 36	14
36 — 39	11
39 — 42	9
$\Sigma f = 100$	

Calculando o valor do ponto médio (Pm)

Idades	f	Pm
18 — 21	9	19,5
21 — 24	12	22,5
24 — 27	12	25,5
27 — 30	17	28,5
30 — 33	16	31,5
33 — 36	14	34,5
36 — 39	11	37,5
39 — 42	9	40,5
$\Sigma f = 100$		

$$\bar{x} = \frac{\sum (Pm_i \cdot f_i)}{n}$$

$$\bar{x} = \frac{19,5 \cdot 9 + 22,5 \cdot 12 + 25,5 \cdot 12 + 28,5 \cdot 17 + 31,5 \cdot 16 + 34,5 \cdot 14 + 37,5 \cdot 11 + 40,5 \cdot 9}{100}$$

$$\bar{x} = \frac{175,5 + 270 + 306 + 484,5 + 504 + 483 + 412,5 + 364,5}{100} \Rightarrow \bar{x} = 30$$

Moda Bruta e Classe Modal: A classe que apresenta a maior frequência é denominada classe **modal**, uma vez agrupados os dados, é possível determinar imediatamente a classe modal, basta fixar o valor da variável de maior frequência.

Exemplo: A tabela abaixo mostra os pesos das crianças em uma classe. Usando esta informação, encontre a classe modal e a moda bruta.

Kg de massa (m)	Frequência
$30 \leq m < 40$	7

A classe modal é a classe que tem a maior frequência. Neste caso a classe modal é 50

$40 \leq m < 50$	6
$50 \leq m < 60$	8
$60 \leq m < 70$	4

$\leq m < 60$. O método mais simples para o cálculo da moda consiste em tomar o ponto médio da classe modal. Damos a esse valor a denominação de **moda bruta**, no caso, **Mo = 55**.

Mediana Bruta e Classe Mediana: A classe que apresenta o valor central das frequências é denominada classe **mediana**, uma vez agrupados os dados, é possível determinar imediatamente a classe mediana, por observação e contagem das frequências acumuladas, conforme já foi visto.

Exemplo: A tabela abaixo mostra os pesos das crianças em uma classe. Usando esta informação, encontre a classe mediana e a mediana bruta.

Kg de massa (m)	Frequência
$30 \leq m < 40$	7
$40 \leq m < 50$	6
$50 \leq m < 60$	8
$60 \leq m < 70$	4

Como o total de crianças é 25, a classe mediana é a classe que tem o 13º valor de frequência. Neste caso a classe mediana é $40 \leq m < 50$. O método mais simples para o cálculo da mediana consiste em tomar o ponto médio da classe mediana. Damos a esse valor a denominação de **mediana bruta**, no caso, **Me = 45**.

A moda bruta é o ponto médio da classe de maior frequência e a mediana bruta é o ponto médio da classe da frequência mediana. Para se obter uma moda e uma mediana mais precisa, para dados agrupados, existem várias fórmulas, de matemáticos como KING e CZUBER, essas fórmulas não serão estudadas nesse curso.

M E D I D A S D E D I S P E R S Ã O

Vimos que um conjunto de valores pode ser convenientemente sintetizado, por meio de procedimentos matemáticos, em poucos valores representativos, tais como: **média aritmética, mediana e moda**.

No entanto, quando se trata de interpretar dados estatísticos, mesmo aqueles já convenientemente simplificados, é necessário ter-se uma idéia retrospectiva de como se apresentavam esses mesmos dados nas tabelas.

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** ou de **variabilidade**. Pois são necessários dois tipos de medidas para descrever adequadamente um conjunto de dados. Além da informação quanto ao “meio” de um conjunto de números (estudado em medidas de tendência central), precisamos saber também a dispersão desses dados. As medidas de dispersão indicam se os valores estão relativamente próximos uns dos outros, ou separados. Essas dispersões tem como ponto de referência as medidas de tendência central. O valor zero indica a ausência de dispersão e quanto maior o valor, maior a dispersão.

Ou seja, as medidas de dispersão ou de afastamento são medidas estatísticas utilizadas para verificar o quanto os valores encontrados em uma pesquisa estão dispersos ou afastados em relação à média ou em relação à mediana. Para avaliar o grau de variabilidade ou de dispersão

são utilizadas as chamadas medidas de dispersão. Dessas medidas, estudaremos a **amplitude**, a **variância**, o **desvio padrão** e o **coeficiente de variação**.

Amplitude Total (A): é a diferença entre o maior e o menor valor de uma série de dados. Quanto maior a amplitude total, maior a dispersão ou variabilidade dos valores da variável.

Exemplo: No conjunto de números 4 , 6 , 8 , 9 , 12 , 17, 20, calcule a Amplitude.

$$A = \text{maior valor} - \text{menor valor}$$

$$A = 20 - 4$$

$$A = 16$$

Alternativamente, pode-se dizer que o intervalo de valores vai de 4 a 20.

No caso de termos uma distribuição de frequência com intervalos de classe, calculamos a Amplitude total, pela diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Exemplo:

Idade	f
18 — 21	9
21 — 24	12
24 — 27	12
27 — 30	17
30 — 33	16
33 — 36	14
36 — 39	11
39 — 42	9
Σ	100

$$A = Ls_{(máx)} - Li_{(mín)}$$

$$A = 42 - 18$$

$$A = 24$$

O fato do intervalo só levar em conta dois valores extremos de um conjunto, nada informando quanto aos outros valores, torna sua utilização bastante limitada.

Variância: É a média dos quadrados dos desvios dos valores a contar da média. Utiliza-se a letra grega sigma minúsculo elevado a 2 “ σ^2 ” para a variância de uma população de N elementos. E, a variância de uma amostra de n elementos é representada pela letra esse minúscula elevado a 2 “ s^2 ”. O símbolo da variância é elevado a 2, porque essa medida de dispersão exprime em quadrados de unidades os valores observados e a média deles, ou seja, se estivermos calculando uma dispersão de comprimento em cm, a variância será obtida em cm^2 . Por isso, também não é muito utilizada como medida de dispersão, mas o cálculo da variância é usado para se obter o desvio padrão, que é a medida de dispersão mais utilizada. Fórmula para a variância amostral:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Substitui-se “ $n-1$ ” por “ n ” no denominador para a variância da população, ou quando a finalidade é apenas descrever os dados e não fazer uma inferência sobre uma população. Nesse curso, usaremos “ n ”, somente quando o exercício mencionar que os dados são

populacionais, ou seja, quando o exercício não mencionar se os dados são populacionais ou amostrais, vamos considerá-los amostrais e usar $n-1$.

Exemplo 1: Calcule a variância para os valores amostrais 5; 7 e 9.

Primeiro calculamos a média aritmética entre os valores:

$$\bar{x} = \frac{\sum x}{n} \rightarrow \bar{x} = \frac{5+7+9}{3} = 7$$

Em seguida aplicamos a fórmula da variância:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{(5 - 7)^2 + (7 - 7)^2 + (9 - 7)^2}{3 - 1} = \frac{4 + 0 + 4}{2} = \frac{8}{2} = 4$$

Resposta: A variância entre os valores 5; 7 e 9 é 4.

Exemplo 2: Calcule a variância para a distribuição de frequência abaixo:

Valor	f
3	6
5	11
9	3
Total	20

Primeiro calculamos a média aritmética entre os valores:

$$\bar{x} = \frac{\sum (x_i \cdot f_i)}{\sum f_i} \rightarrow \bar{x} = \frac{(3 \cdot 6) + (5 \cdot 11) + (9 \cdot 3)}{20} = \frac{18 + 55 + 27}{20} = \frac{100}{20} = 5$$

Em seguida aplicamos a fórmula da variância para a distribuição de frequência:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 F_i}{\sum f_i - 1} = \frac{(3 - 5)^2 \cdot 6 + (5 - 5)^2 \cdot 11 + (9 - 5)^2 \cdot 3}{20 - 1} = \frac{4 \cdot 6 + 0 \cdot 11 + 16 \cdot 3}{19} \\ = \frac{24 + 0 + 48}{19} = \frac{72}{19} = 3,79$$

Se a tabela for com intervalo de classe, basta usar os valores dos pontos médios.

Exemplo 3: Calcule a variância para a distribuição de frequência abaixo:

Valor	f
1 - 5	6
5 - 9	11
9 - 13	3
Total	20

Primeiro calculamos a média: $\bar{x} = \frac{\sum (PM \cdot f_i)}{\sum f_i} \rightarrow \bar{x} = \frac{(3 \cdot 6) + (7 \cdot 11) + (11 \cdot 3)}{20} = \frac{18 + 77 + 33}{20} = \frac{128}{20} = 6,4$

Em seguida, calculamos a variância:

$$s^2 = \frac{\sum (PM - \bar{x})^2 F_i}{\sum f_i - 1} = \frac{(3 - 6,4)^2 \cdot 6 + (7 - 6,4)^2 \cdot 11 + (11 - 6,4)^2 \cdot 3}{20 - 1} = \frac{69,36 + 3,96 + 63,48}{19} \\ = \frac{136,8}{19} = 7,2$$

A **variância** baseia-se nos desvios em torno da média aritmética, porém determinando a **média aritmética dos quadrados dos desvios**. Ela é um número em unidade quadrada em relação à variável em questão, o que, sob o ponto de vista prático, é um inconveniente. Por

isso, imaginou-se uma nova medida que tem utilidade e interpretação práticas, denominada **desvio padrão**.

Desvio Padrão: Utiliza-se a letra grega sigma minúsculo “ σ ” para o desvio padrão de uma população de N elementos. E, o desvio padrão de uma amostra de n elementos é representada pela letra esse minúscula “ s ”. O desvio padrão é simplesmente a raiz quadrada da variância.

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

O desvio padrão é a medida de dispersão mais utilizada, desempenha papel relevante em toda a estatística e a sua unidade é a mesma da média. O desvio padrão dá uma idéia de como os valores de uma amostra estão dispersos em relação à média. Quanto maior o desvio padrão, maior é a dispersão dos valores em relação à média. Um desvio padrão igual a zero indica que todos os valores são iguais à média.

Se você já leu um artigo científico com certeza deve ter percebido que os resultados geralmente são apresentados por meio da média aritmética. E logo em seguida a média, é apresentado um outro número, que curiosamente é precedido pelo símbolo de "mais ou menos". Exatamente como na tabela abaixo:

Velocidade de homens e mulheres numa caminhada

	Faixa etária	Idade (valor médio)	Velocidade (m/s)
Homens (n = 8)	70 a 79	74,4 ± 2,92	1,1 ± 0,13
Mulheres (n = 7)	60 a 79	70,1 ± 2,91	1,07 ± 0,16

Pois bem, este número depois do "mais ou menos" é o Desvio Padrão, que indica a dispersão dos dados dentro da amostra. Isto é: o quanto os resultados diferem da média. Por isso que ele sempre é apresentado junto da média. Um não faz sentido sem o outro.

É importante ter em mente que quanto menor o desvio padrão, mais homogênea é a minha amostra. Em termos de pesquisas científicas, é isso que desejamos em nossos resultados.

Na tabela acima, a média da velocidade da marcha dos homens foi de 1,1m/s e o Desvio Padrão foi de 0,13m/s. Isso significa que, no geral, boa parte da minha amostra caminha com uma velocidade entre 0,97 m/s e 1,23 m/s. Enfim, quando eu adiciono o Desvio Padrão a interpretação dos meus números, eu tenho idéia de quanto a velocidade da minha amostra varia em torno da média.

Coeficiente de Variação (CV): O **Índice de Variabilidade (IV)** ou o **Coeficiente de Variação (CV)** é a razão entre o desvio padrão e a média, o resultado normalmente é multiplicado por 100 para que o coeficiente seja dado em porcentagem. O CV é utilizado quando dois grupos apresentam mesmo desvio padrão e médias diferentes, ou para se comparar duas ou mais séries de valores, quanto a sua dispersão ou variabilidade, quando expressas em unidades diferentes, ou ainda quando duas médias forem muito distantes.

O **Coeficiente de Variação (CV)**, é uma medida relativa de dispersão, onde a variabilidade, através do desvio padrão, é comparada com sua média, através da relação abaixo:

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Onde s é o desvio padrão, $\bar{x}$ é a média aritmética e o fator 100 é utilizado para apresentar a resposta na forma percentual.

Normalmente, dizemos que um CV abaixo de 15% indica um grupo de dados com baixa dispersão. Um CV acima de 30% representa uma alta dispersão dos dados e, entre esses valores, o CV representa uma dispersão média.

Exemplo 1: A análise de dois grupos diferentes de dados foi realizada e eles apresentaram o mesmo desvio padrão, mas valores médios diferentes:

Grupo 1: (3; 1 e 5) $\rightarrow \bar{x} = 3; s^2 = 4$ e $s = 2$

Grupo 2: (55; 57 e 53) $\rightarrow \bar{x} = 55; s^2 = 4$ e $s = 2$

Qual deles possui maior dispersão?

Vamos obter as variabilidades com relação as médias, através do cálculo dos coeficientes de variação para cada grupo:

Grupo 1: $CV = \frac{2}{3} \cdot 100 = 66,7\%$ (o desvio padrão é um percentual grande, comparado com o valor médio)

Grupo 2: $CV = \frac{2}{55} \cdot 100 = 3,64\%$ (o desvio padrão é um percentual pequeno, comparado com o valor médio)

Observe que:

_ Para o Grupo 1, o desvio padrão corresponde a 66,7% da média;

_ Para o Grupo 2, o desvio padrão corresponde a 3,64% da média;

Podemos concluir que:

O Grupo 1 possui maior dispersão do que o Grupo 2

Exemplo 2: (Grupos com unidades diferentes) Ao medir a variabilidade das alturas em cm e comparar com a variabilidade das massas em kg dos alunos. Os resultados foram:

Alturas: $s = 15$ cm e $\bar{x} = 165$ cm

Massas: $s = 10$ kg e $\bar{x} = 65$ kg

Pela comparação direta dos desvios chegaríamos a conclusão que as alturas tem mais variabilidade do que as massas. Mas obtendo o CV:

Alturas: CV = 9,1%

Massas: CV = 15,4%

Concluimos que:

As massas tem maior variabilidade que as alturas.

Exemplo 3: (Grupos com mesmas unidades, porém com médias distantes) Imagine que desejamos comparar a variabilidade das massas de adultos com as de recém-nascidos:

Adultos: $s = 10$ kg , $\bar{x} = 65$ kg e CV = 15,4%

Recém-nascidos: $s = 1$ kg , $\bar{x} = 3$ kg e CV = 33,3%

Analogamente ao exemplo 2, a comparação das variabilidades através do desvio nos levaria a decisão contrária, pois **a maior variabilidade ocorreu entre os recém-nascidos.**