ESTATÍSTICA DESCRITIVA

SUMÁRIO:

FONTE:
CRESPO, ANTÔNIO ARNOT. ESTATÍSTICA BÁSICA FACIL. 15ª ED. SARAIVA .SÃO PAULO
Com adaptações.

UNIDADE I

1 - A ESTATÍSTICA

Exprimindo por meio de números as observações que se fazem de elementos com, pelo menos, uma característica comum (por exemplo: os alunos do sexo masculino de uma comunidade), obtemos os chamados **dados** referentes a esses elementos.

Podemos dizer, então, que:

A **Estatística** é uma parte da matemática aplicada que fornece métodos para coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões.

2 - FASES DO MÉTODO ESTATÍSTICO

Podemos distinguir no método estatístico as seguintes fases:

2.1 - Coleta de dados

A coleta pode ser direta e indireta.

A coleta é **direta** quando feita sobre elementos informativos de registro obrigatório (nascimentos, casamentos e óbitos, importação e exportação de mercadorias), ou, ainda, quando os dados são coletados pelo próprio pesquisador através de inquéritos e questionários, como é o caso das notas de verificação e de exames, do censo demográfico etc.

A coleta direta de dados pode ser classificada relativamente ao fator tempo em:

a) **contínua** (registro) **-** quando feita continuamente, tal como a de nascimentos e óbitos e a de freqüência dos alunos às aulas;

b) **periódica -** quando feita em intervalos constantes de tempo, como os censos (de 10 em 10 anos) e as avaliações mensais dos alunos;

c) **ocasional -** quando feita extemporaneamente, a fim de atender uma conjuntura ou a uma emergência, como no caso de epidemias que assolam ou dizimam rebanhos inteiros.

A coleta se diz **indireta** quando é inferida de elementos conhecidos (coleta direta) e/ou do conhecimento de outros fenômenos relacionados com o fenômeno estudado. Como por exemplo, podemos citar a pesquisa sobre a mortalidade infantil, que é feita através de dados colhidos por uma coleta direta.

2.2 - Crítica de dados

Obtidos os dados, eles devem ser cuidadosamente criticados, à procura de possíveis falhas e imperfeições, a fim de não incorremos em erros grosseiros ou de certo vulto, que possam influir sensivelmente nos resultados.

A crítica é **externa** quando visa às causas dos erros por parte do informante, por distração ou má interpretação ou má interpretação das perguntas que lhe foram feitas; é **interna** quando visa observar os elementos originais dos dados da coleta.

2.3 - Apuração dos dados

Nada mais é do que a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação. Pode ser manual, eletromecânica ou eletrônica.

2.4 - Exposição ou apresentação dos dados

Por mais diversa que seja a finalidade que se tenha em vista, os dados devem ser apresentados sob a forma adequada (tabelas ou gráficos), tornando mais fácil o exame daquilo que está sendo objeto de tratamento estatístico.

2.5 - Análise dos resultados

Após a exposição de dados, fazemos uma análise dos resultados obtidos, e tiramos desses resultados conclusões e previsões.

UNIDADE II

1 - VARIÁVEIS

A cada fenômeno corresponde um número de resultados possíveis. Assim por exemplo:

- para o fenômeno "sexo" são dois os resultados possíveis: sexo masculino e sexo feminino;

- para o fenômeno "número de filhos" há um número de resultados possíveis expresso através dos números naturais:

0, 1, 2, 3, ...,n;

- para o fenômeno "estatura" temos uma situação diferente, pois os resultados podem tomar um número infinito de valores numéricos dentro de um determinado intervalo.

Variável é, convencionalmente, o conjunto de resultados possíveis de um fenômeno.

Os exemplos acima nos dizem que uma variável pode ser:

a) **qualitativa -** quando seus valores são expressos por atributos: sexo (masculino - feminino), cor da pele (branca ,preta) etc.;

b) **quantitativa -** quando seus valores são expressos em números (salários, idade etc.) Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de **variável contínua;**

uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de **variável discreta.**

Assim , o número de alunos de uma escola pode assumir qualquer um dos valores do conjunto N = {1,2,3, ...,50, ...}, mas nunca valores como 2,5 ou 3,78 ou 4,325 etc. Logo, é uma **variável discreta.** Já o peso desses alunos é uma variável contínua, pois um dos alunos tanto pode pesar 72 Kg, como 72,5 Kg, como 72,54 Kg etc., dependendo esse valor da precisão da medida.

De modo geral, as medições dão origem a variáveis contínuas e as contagens ou enumerações, a variáveis discretas.

2 - POPULAÇÃO E AMOSTRA

Ao conjunto de entes portadores de, pelo menos, uma característica comum denominamos **população estatística** ou **universo estatístico**.

Assim, os estudantes, por exemplo, constituem uma população, pois apresentam pelo menos uma característica comum: são os que estudam.

Quando limitamos as observações referentes a uma determinada pesquisa a apenas uma parte da população, a essa parte proveniente da população em estudo denominamos **amostra.**

Uma **amostra** é um subconjunto finito de uma população.

É necessário que a amostra seja **representativa** da população, isto é, a amostra deve possuir as mesmas características básicas da população, no que diz respeito ao fenômeno que desejamos pesquisar. É preciso, pois, que a amostra ou as amostras que vão ser usadas sejam obtidas por processos adequados.

3 - AMOSTRAGEM

É o método de seleção dos elementos de uma população, de modo a se obter uma amostra representativa da população.

3.1 - Amostragem casual ou aleatória simples

Este tipo de amostragem é equivalente a um sorteio lotérico.

Na prática, a amostragem casual ou aleatória simples pode ser realizada numerando-se a população de **1** a **n** e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, **k** números dessa seqüência, os quais corresponderão aos elementos pertencentes à amostra.

Exemplo:

Vamos obter uma amostra representativa para a pesquisa da estatura de 90 alunos de uma escola:

a) Numeramos os alunos de 01 a 90.
b) Colamos os números em uma urna, e depois retiramos, um a um, nove números que formarão a amostra. Neste caso, 10% da população.

3.2 - Amostragem proporcional estratificada

Muitas vezes a população se divide em subpopulações **- estratos.**

Como é provável que a variável em estudo apresente, de estrato em estrato, um comportamento heterogêneo e, dentro de cada estrato, um comportamento homogêneo convém que o sorteio dos elementos da amostra leve em consideração tais estratos.

É exatamente isso que fazemos quando empregamos a **amostragem proporcional estratificada**, que além de considerar a existência dos estratos, obtém os elementos da amostra proporcional ao número de elementos dos mesmos.

Exemplo:

Supondo, no exemplo anterior, que, dos 90 alunos, 54 sejam meninos e 36 sejam meninas, vamos obter a amostra proporcional estratificada.

São, portanto, dois estratos (sexo masculino e sexo feminino) e queremos uma amostra de 10% da população. Logo, temos:

SEXO	POPULAÇÃO	10%	AMOSTRA
M	54	$\frac{10 \times 54}{100} = 5{,}4$	5
F	36	$\frac{10 \times 36}{100} = 3{,}6$	4
total	90	$\frac{10 \times 90}{100} = 9{,}0$	9

3.3 - Amostragem sistemática

Quando os elementos da população já se acham ordenados, não há necessidade de construir o sistema de referência. São exemplos os prontuários médicos, os prédios de uma rua, as linhas de produção etc. Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador. A esse tipo de amostragem denominamos **sistemática.**

Assim, no caso de uma linha de produção, podemos, a cada 10 itens produzidos, retirar um para pertencer a uma amostra da produção diária. Neste caso, estaríamos fixando o tamanho da amostra em 10% da população.

Exemplo:

Suponhamos uma rua contendo 900 prédios, dos quais desejamos obter uma amostra formada de 50 prédios. Podemos, neste caso, usar o seguinte procedimento:

Como $\frac{900}{50} = 18$, escolhemos por sorteio casual um número de 1 a 18 (inclusive), o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, se o número sorteado fosse o 4, tomaríamos, pelo lado direito da rua, o 4º prédio, 0 22º, o 40º etc., até voltarmos ao início da rua, pelo lado esquerdo.

EXERCÍCIOS

1) A coleta se diz ___________ quando é inferida de elementos conhecidos (coleta direta) e/ ou do conhecimento de outros fenômenos relacionados com o fenômeno em estudo

a) contínua
b) periódica
c) ocasional
d) indireta

2) Nada mais é do que a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação

a) apuração de dados
b) exposição de dados
c) apresentação de dados
d) análise dos resultados

3) É o conjunto de resultados possíveis de um fenômeno

a) dados
b) variável
c) amostra
d) rol

4) A graduação de um militar é uma variável

a) quantitativa
b) contínua
c) qualitativa
d) discreta

5) Ao nascer, os bebês são pesados e medidos, para se saber se estão dentro das tabelas de peso e altura esperadas. Estas duas variáveis são:

a) qualitativas
b) ambas discretas
c) ambas contínuas
d) Uma discreta e outra contínua

6) Ao conjunto de entes portadores de, pelo menos, uma característica comum denominamos

a) amostragem b) rol
c) universo estatístico d) dados brutos

7) Numa população de determinada cidade desejou-se estudar a aceitação do controle da natalidade pesquisando 10% de indivíduos de cada religião. A técnica de amostragem utilizada foi?

a) estratificada
b) casual
c) sistemática
d) outra

8) (EAGS - 86) Os dados estatísticos, após sofrerem transformações são transmitidos ao público através de quadros e gráficos. Como é chamado esta fase do processo estatístico?

a) apuração
b) apresentação
c) planejamento
d) interpretação

9) A coleta direta de dados ____________é feita extemporaneamente, a fim de atender a uma conjuntura ou a um emergência, como no caso de epidemias que assolam ou dizimam rebanhos inteiros.

a) contínua
b) periódica
c) ocasional
d) extraordinária

10) Uma população encontra-se dividida em 3 três estratos, com tamanhos, respectivamente, n1 = 40, n2 = 100 e n3 = 60. Sabendo que, ao ser realizada uma amostragem estratificada proporcional, nove elementos da amostra foram retirados do 3º estrato, determine o número total de elementos da amostra.

a) 18
b) 27
c) 30
d) 36

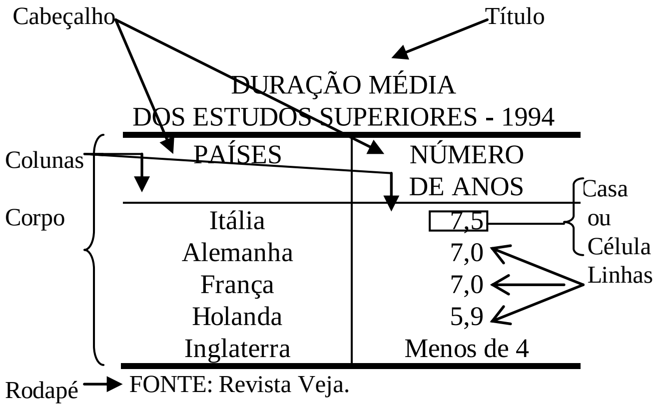
UNIDADE III

1- TABELAS

Tabela é um quadro que resume um conjunto de observações.

Uma tabela compõe-se de:

a) Corpo - conjunto de linhas e colunas que contém informações sobre a variável em estudo;
b) Cabeçalho - parte superior da tabela que especifica o conteúdo das colunas;
c) Coluna indicadora - parte da tabela que específica o conteúdo das linhas;
d) Linhas - retas imaginárias que facilitam a leitura, no sentido horizonte, de dados que se inscrevem nos seus cruzamentos com as colunas;
e) Casa ou célula - espaço destinado a um só número;
f) Título - conjunto de informações, as mais completas possíveis de uma tabela.



De acordo com a Resolução 886 do IBGE, nas casas ou células devemos colocar:

- Um traço horizontal (__) quando o valor é o zero, não só quanto à natureza das coisas, como quanto ao resultado do inquérito;
- Três pontos (...) quando não temos os dados;
- Um ponto de interrogação (?) quando temos dúvida quanto à exatidão de determinado valor:
- Zero (0) quando o valor é muito pequeno para ser expresso pela unidade utilizada. Se os valores são em numerais decimais, precisamos acrescentar à parte decimal um número correspondente de zeros (0, 0; 0,00; 0,000 ...).

2 - SÉRIES ESTATÍSTICAS

Denominamos série estatística toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie.

Numa série estatística observamos a existência de três elementos ou fatores: o tempo, o espaço e a espécie. Conforme varie um dos elementos da série, podemos classificá-la em **histórica, geográfica e específica.**

2.1 - Séries históricas, cronológicas, temporais ou marchas

Descrevem os valores da variável, em determinado local, discriminados segundo intervalos de tempo variáveis.
Exemplo:

PREÇO DO ACÉM NO VAREJO
SÃO PAULO - 1989-94

ANOS	PREÇO MÉDIO (US$)
1989	2,24
1990	2,73
1991	2,12
1992	1,89
1993	2,04
1994	2,62

FONTE: APA

2.2 - Séries geográficas, espaciais, territoriais ou de localização

Descrevem os valores da variável, em determinado instante, discriminados segundo regiões.
Exemplo:

DURAÇÃO MÉDIA
DOS ESTUDOS SUPERIORES - 1994

PAÍSES	NÚMERO DE ANOS
Itália	7,5
Alemanha	7,0
França	7,0
Holanda	5,9
Inglaterra	Menos de 4

FONTE: Revista Veja.

2.3 - Séries específicas ou categóricas

Descrevem os valores da variável, em determinado tempo e local, discriminados segundo especificações ou categorias.

REBANHOS BRASILEIROS
1992

ESPÉCIES	QUANTIDADE (1.000 CABEÇAS)
Bovinos	154.440,8
Bubalinos	1.423,3
Eqüinos	549,5
Asininos	47,1
Ovinos	19.955,9
Caprinos	12.159,6

FONTE: IBGE

3 - SÉRIES CONJUGADAS (TABELAS DE DUPLA ENTRADA)

Quando conjugamos duas séries em uma única tabela, obtemos uma tabela de dupla entrada. Em uma tabela desse tipo ficam criadas duas ordens de classificação: uma horizontal (linha) e uma vertical (coluna).
Exemplo:

TERMINAIS TELEFÔNICOS EM SERVIÇO
1991-93

REGIÕES	1991	1992	1993
Norte	342.938	375.658	403.494
Nordeste	1.287.813	1.379.101	1.486.649
Sudeste	6.234.501	6.729.467	7.231.634
Sul	1.497.315	1.608.989	1.746.232
Centro-Oeste	713.357	778.925	884.822

FONTE: MINISTÉRIO DAS COMUNICAÇÕES

A conjugação, no exemplo dado, foi série **geográfica-série histórica**, que dá origem à série geográfico-histórica ou geográfica-temporal.

EXERCÍCIOS:

11 - (EAGS - 99) De acordo com a Resolução 886 da fundação IBGE, nas casas ou células devemos colocar

a) Zero (0) quando não temos os dados.
b) Três pontos (...) quando não temos os dados.
c) Um traço horizontal (-) quando temos dúvidas quanto à exatidão de determinado valor.
d) Um ponto de interrogação (?) quando o valor é zero, não só quanto a natureza das coisas, como quanto ao resultado do inquérito.

12 - Considerando-se a tabela a seguir indicada, pode-se concluir que seus dados refletem uma série:

1972

PRODUTOS	QUANTIDADE (TONELADAS)
CAFÉ	400.000
AÇÚCAR	200.000
MILHO	100.000
FEIJÃO	20.000

FONTE: Dados fictícios

a) especificativa ou específica
b) geográfica
c) temporal
d) evolutiva

4 - DADOS ABSOLUTOS E DADOS RELATIVOS

Os dados estatísticos resultantes da coleta direta da fonte, sem outra manipulação senão a contagem ou medida, são chamados **dados absolutos.**

Dados relativos são o resultado de comparações por quociente (razões) que se estabelecem entre dados absolutos e têm por finalidade realçar ou facilitar as comparações entre quantidades.

Traduzem-se os dados relativos, em geral, por meio de percentagens, índices, coeficientes e taxas.

4.1 - As percentagens

Dada uma razão qualquer $\frac{a}{b}$, chamamos de **porcentagem** do valor **b** a todo valor de **a** que estabeleça uma proporção com alguma razão centesimal.

$$\boxed{\frac{a}{b} = \frac{r}{100} = r\,\%}$$

Exemplo:

$$\underbrace{\frac{9}{100}}_{\text{razão centesimal}} = 9\% \text{ (nove por cento)} \Rightarrow \textbf{Taxa percentual}$$

Observe a tabela abaixo

MATRÍCULAS NA ESCOLA REPETÊNCIA

CATEGORIAS	**N.º DE ALUNOS**	**%**
1º Grau	19. 286	91,0
2º Grau	1.681	7,9
3º Grau	234	1,1
Total	21.201	100,0

A terceira coluna da tabela representa as percentagens dos alunos de cada grau:

1º grau $\rightarrow \dfrac{19.286 \text{ X } 100}{21.201} = 90,96 = 91,0$

2º grau $\rightarrow \dfrac{1.681 \text{ X } 100}{21.201} = 7,92 = 7,9$

3º grau $\rightarrow \dfrac{234 \text{ X } 100}{21.201} = 1,10 = 1,1$

Os valores da terceira coluna nos dizem que, de cada 100 alunos da escola, 91 estão matriculados no 1º grau, 8, aproximadamente, no 2º grau e 1 no 3º no grau.

O emprego da percentagem é de grande valia quando é nosso intuito destacar a participação da parte no todo.

4.2 - Índices

Os índices são razões entre duas grandezas tais que uma não inclui a outra.

Exemplos:

Densidade demográfica = $\dfrac{\text{população}}{\text{superfície}}$

Renda *per capita* = $\dfrac{\text{renda}}{\text{população}}$

4.2 - Coeficientes

Os coeficientes são razões entre o número de ocorrências e o número total (número de ocorrências e número de não-ocorrências).

Exemplos:

Coeficiente de natalidade = $\dfrac{\text{números de nascimentos}}{\text{população total}}$

Coeficiente de evasão escolar = $\dfrac{\text{N}^\circ \text{ de alunos evadidos}}{\text{N}^\circ \text{ inicial de matrícula}}$

4.3 - Taxas

As taxas são os coeficientes multiplicados por uma potência de 10 (10, 100, 1.000 etc.) para tornar o resultado mais inteligível.

Exemplos:

Taxa de mortalidade = coeficiente de mortalidade X 1.000
Taxa de evasão escolar = coef. De evasão escolar X 100

EXERCÍCIO RESOLVIDO

O estado **A** apresentou 733.986 matrículas na 1ª série, no início do ano de 1994, e 683.816 no fim do ano. O estado **B** apresentou, respectivamente, 436.127 e 412.457 matrículas. Qual o Estado que apresentou maior evasão escolar?

A $\rightarrow$ TEE = $\dfrac{733.986 - 683.816}{733.986} x100 = 6,83 = 6,8$ %

B $\rightarrow$ TEE = $\dfrac{436.127 - 412.457}{436.127} x100 = 5,42 = 5,4$ %

O estado que apresentou maior evasão escolar foi **A.**

EXERCÍCIOS:

13) Considere a tabela abaixo:

EVOLUÇÃO DAS RECEITAS DO CAFÉ INDUSTRIALIZADO - 94

MESES	**VALOR (US$ MILHÕES)**
Janeiro	33,3
Fevereiro	54,1
Março	44,5
Abril	52,9
Total	184,8

Dados fictícios

a) Complete-a com uma coluna de taxas percentuais.
b) Como se distribuem as receitas em relação ao total?
c) Qual o desenvolvimento das receitas de um mês para o outro?
d) Qual o desenvolvimento das receitas em relação ao mês de janeiro?

14) Considerando que Minas Gerais, em 1992, apresentou (dados fornecidos pelo IBGE):

- População: 15.957,6 mil habitantes;
- Superfície: 586.624 Km^2;
- Nascimentos: 292.036;
- Óbitos: 99.281.

Calcule:
a) o índice da densidade demográfica;
b) a taxa de natalidade;
c) a taxa de mortalidade.

15) Uma frota de 40 caminhões, transportando, cada um, 8 toneladas, dirige-se a duas cidades A e B. Na cidade A são descarregados 65% desses caminhões, por 7 homens, trabalhando 7 horas. Os caminhões restantes seguem para a cidade B, onde 4 homens gastam 5 horas para o seu descarregamento. Em que cidade se obteve melhor produtividade?

RESPOSTAS DAS UNIDADES I A III:
1) d 2) a 3) b 4) c 5) c 6) c 7) a 8) b 9) c 10) c 11) b 12) a 13) b) 18; 29,3 ; 24,1 ; 28,6 ; 100 c) 162,5 ; 82,3 ; 118,9 d) 100,0 ; 162,5 ; 133,6 158,9 14)a) 27,2 h/Km^2 b) 1,83 % c) 0,62 % 15) B

UNIDADE IV

1 - GRÁFICOS ESTÁTISTICOS

O gráfico estatístico é a forma de apresentação dos dados estatísticos, cujo o objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que as séries.

A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil:

a) **Simplicidade -** o gráfico deve ser destituído de detalhes e traços de importância secundária.
b) **Clareza** - O gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
c) **Veracidade -** o gráfico deve expressar a verdade sobre o fenômeno em estudo.

Os principais tipos de gráficos são os **diagramas** os **cartogramas** e os **pictogramas**.

2 - DIAGRAMAS

Os diagramas são gráficos geométricos de , no máximo, duas dimensões; para sua construção, em geral, fazemos uso do sistema cartesiano.

Principais diagramas:

2.1 Gráfico em linha ou em curva

Este tipo de gráfico se utiliza da linha poligonal para representar a série estatística no plano cartesiano.

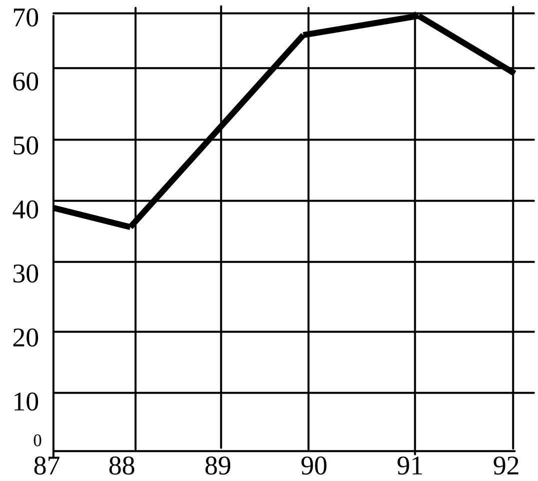
Exemplo:

PRODUÇÃO BRASILEIRA DE ÓLEO DE DENDÊ
1987-92

ANOS	QUANTIDADE (1.000 t)
1987	39,3
1988	39,1
1989	53,9
1990	65,1
1991	69;1
1992	59,5

FONTE: Agropalma.

PRODUÇÃO BRASILEIRA
DE ÓLEO DE DENDÊ
1987 - 92
mil
toneladas



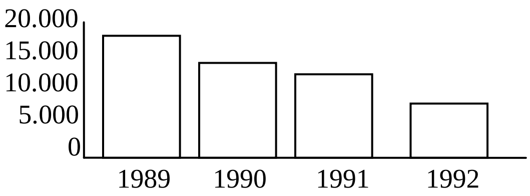
FONTE: Agropalma

2.2 Gráficos em colunas ou em barras

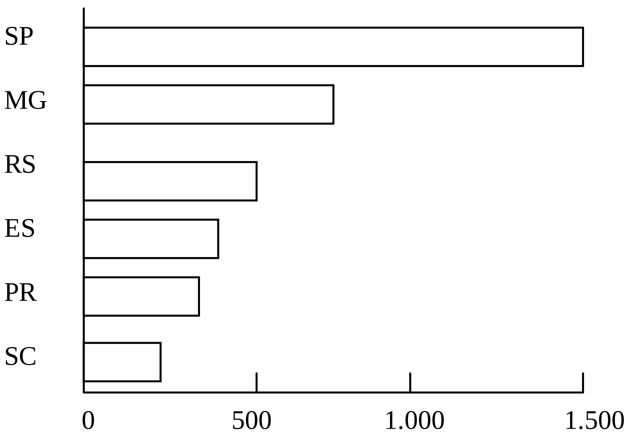
É a representação de uma serie por meio de retângulos, dispostos verticalmente (em colunas) ou horizontalmente (em barras).

Exemplos:

PRODUÇÃO BRASILEIRA DE CARVÃO 1989 -92



EXPORTAÇÕES BRASILEIRAS
MARÇO - 1995



- Outro tipo de diagrama é o **gráfico em setores**.
- O **cartograma** é a representação sobre uma carta geográfica.
- O **pictograma** é a representação gráfica por figuras.

UNIDADE V

1 - TABELA PRIMITIVA E ROL

Suponhamos termos feito uma coleta de dados relativos às estaturas de quarenta alunos, que compõem uma amostra dos alunos de um colégio **A** , resultando a seguinte tabela de valores:

TABELA 5.1

ESTATURA DE 40 ALUNOS DO COLÉGIO A

166	161	162	165	164
162	168	156	160	164
155	163	155	169	170
154	156	153	156	158
160	150	160	167	160
161	163	173	155	168
152	160	155	151	164
161	172	157	158	161

A esse tipo de tabela, cujos elementos não foram numericamente organizados, denominamos **tabela primitiva.**

A maneira mais simples de organizar os dados é através de uma certa ordenação (crescente ou decrescente). A tabela obtida após a ordenação dos dados recebe o nome de **rol**.

Vamos observar o **rol** abaixo:

TABELA 5.2
ESTATURA DE 40 ALUNOS DO COLÉGIO A

150	155	160	162	166
151	156	160	162	167
152	156	160	163	168
153	156	160	163	168
154	157	161	164	169
155	158	161	164	170
155	158	161	164	172
155	160	161	165	173

Agora, podemos analisar, com relativa facilidade os dados da tabela.

2 - DISTRIBUIÇÃO DE FREQÜENCIAS

A representação tabular é uma das modalidades mais utilizadas para a apresentação dos dados estatísticos coletados na amostragem. Para isto, nós, freqüentemente, listamos os valores encontrados na amostra. Uma classificação metodológica usual é verificar a natureza dos dados estatísticos. Vejamos, o tratamento tabular dado aos dados discretos e contínuos

2.1 - REPRESENTAÇÃO DE DADOS DISCRETOS

Seja a seguinte amostra: 3,4,4,4,5,5,5,5,5, 6,6,6,6,6,6,7,7,7,7,8 (já em forma de ROL).

2.1.1 - Classe (i)

Classes de freqüência ou, simplesmente, classes são intervalos de variação da variável.

No nosso exemplo acima nós temos as variáveis ($\mathbf{x_i}$): 3,4,5,6,7 e 8 perfazendo total de 6 classes, onde **i** representa é o número total de classes, isto é; **i** = 6.

2.1.2 - Freqüência simples ou absoluta ($\mathbf{f_i}$)

É o número de observações correspondente a um determinado valor. Por exemplo: a freqüência do valor 3 é 1, do valor 4 é 3, conforme podemos ver no quadro abaixo.

VALOR	FREQÜENCIA
3	1
4	3
5	5
6	6
7	4
8	1
TOTAL	20

Representando a distribuição acima sob a forma de tabela, temos:

i	$\mathbf{x_i}$	$\mathbf{f_i}$
1	3	1
2	4	3
3	5	5
4	6	6
5	7	4
6	8	1
		$\sum f_i = 20$

2.2 - REPRESENTAÇÀO DE DADOS CONTÍNUOS

Na representação de dados contínuos utiliza-se a forma de intervalos (distribuição intervalar) em virtude desses dados serem obtidos, normalmente, através de medidas. Isto melhora a distribuição dos erros cometidos (erro do observador, do método utilizado, do equipamento de medida, etc.).

Dada a tabela abaixo como exemplo, vamos representá-la sob a forma de **distribuição intervalar** (intervalos de classe).

TABELA 5.2
ESTATURA DE 40 ALUNOS DO COLÉGIO A

150	155	160	162	166
151	156	160	162	167
152	156	160	163	168
153	156	160	163	168
154	157	161	164	169
155	158	161	164	170
155	158	161	164	172
155	160	161	165	173

1º PASSO: Determinação do números de classe (i)

Utilizaremos **a regra de sturges**, que nos dá o número de classes em função do número de valores da variável:

$$\boxed{i \cong 1 + 3{,}3 \,.\, \log_{10} n}$$

onde i é o número de classe
n é o número total de dados

assim n = 40, temos:
$i \cong 1 + 3{,}3 \,.\, \log_{10} 40$
$i \cong 1 + 3{,}3 \,.\, 1{,}60$
$i \cong 6{,}29$
$i \cong 6$

2º PASSO: Determinar a **Amplitude amostral da distribuição (AA).**
É a diferença entre o maior e o menor valor da amostra.
AA = 173 - 150 → AA = 23

3º PASSO: Determinar a **Amplitude do intervalo de classe (h).**

Achamos aplicando a fórmula $\boxed{h = \dfrac{AA}{i}}$

$$h = \frac{173 - 150}{6} = 3{,}8 \cong 4$$

4º PASSO: Escrever os intervalos de classe.
Os intervalos de classe sempre serão escritos, em termos **desta quantidade até menos aquela,** empregando, para isso, o símbolo |— , (inclusão do 1º extremo e exclusão do 2º extremo).
Exemplo: a 1ª classe seria 150 |— 154 (150 + 4), a 2ª seria 154 |— 158 (154 + 4), e assim sucessivamente.

5º PASSO: Construir a tabela de freqüências:

TABELA 5.3

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
		$\sum f_i = 40$

Obs.: Analisando a tabela acima, o indivíduo com uma estatura de 158 cm está incluído na terceira classe (i = 3) e não na segunda.

3 - OUTROS ELEMENTOS DA DISTRIBUIÇÃO

3.1 - Limites de classe
São extremos de cada classe, isto é; o menor número , limite inferior (ℓ_i) e o maior número, limite superior (L_i).
Exemplo: na segunda classe temos $\ell_2 = 154$ e $L_2 = 158$

3.2 - Amplitude de um intervalo de classe
Amplitude de um intervalo de classe ou, simplesmente, intervalo de classe é a medida do intervalo que define a classe.

Assim $\boxed{h_i = L_i - \ell_i}$

Na tabela 5.3 temos: $h_2 = L_2 - \ell_2 = 158 - 154 = 4$ cm

3.3 - Amplitude total da distribuição
Amplitude total da distribuição (AT) é a diferença entre o limite superior da última classe (limite superior máximo) e o limite inferior da primeira classe (limite inferior mínimo).

Assim AT = L(máx.) - ℓ(mín.)

Na tabela 5.3 temos AT = 174 - 150 = 24 cm

3.4 - Ponto médio de uma classe (x_i)

Ponto médio de uma classe é, como o próprio nome indica, o ponto que divide o intervalo de classe em duas partes iguais.

Assim $\boxed{x_i = \dfrac{\ell_i + L_i}{2}}$

Na tabela 5.3 temos: $x_2 = \dfrac{\ell_2 + L_2}{2} = \dfrac{154+158}{2}$ = 156 $\rightarrow x_2$ = 156 cm

4 - TIPOS DE FREQUENCIAS

Além da freqüências simples temos as seguintes freqüências:

4.1 - Freqüências relativas (fr_i)

São os valores das razões entre as freqüências simples e a freqüência total:

Assim $fr_i = \dfrac{f_i}{\sum f_i}$

Na tabela 5.3 temos $fr_3 = \dfrac{f_3}{\sum f_i} \Rightarrow fr_3 = \dfrac{11}{40}$ = 0,275 ,

donde $\sum fr_i$ = 1 ou 100 %

4.2 - Freqüências acumulada (F_i)

É o total das freqüências de todos os valores inferiores ao limite superior do intervalo de uma classe.

Assim $F_K = f_1 + f_2 + ... + f_K$

Na tabela 5.3 temos $F_3 = f_1 + f_2 + f_3 = 4 + 9 + 11 \rightarrow F_3 = 24$

4.3 - Freqüências acumulada relativa (Fr_i)

É a freqüência acumulada da classe, dividida pela freqüência total da distribuição

Assim $\boxed{Fr_i = \dfrac{F_i}{\sum f_i}}$

Na tabela 5.3 temos $Fr_3 = \dfrac{F_3}{\sum f_i} \Rightarrow Fr_3 = \dfrac{24}{40}$ = 0,600 ,

Fr_3 = 0,600

Considerando a tabela 5.3, podemos montar a seguinte tabela com as freqüências estudadas:

TABELA 5.4

i	ESTATURA (cm)	f_i	x_i	fr_i	F_i	Fr_i
1	150 |— 154	4	152	0,100	4	0,100
2	154 |— 158	9	156	0,225	13	0,325
3	158 |— 162	11	160	0,275	24	0,600
4	162 |— 166	8	164	0,200	32	0,800
5	166 |— 170	5	168	0,125	37	0,925
6	170 |— 174	3	172	0,075	40	1,000
		$\sum$=40		$\sum$=1,000		

EXERCÍCIOS:

1) Forme uma distribuição de freqüência sem intervalo de classe, com os dados abaixo.

14	12	11	13	14	13
12	14	13	14	11	12
12	14	10	13	15	11
15	13	16	17	14	14

A tabela abaixo apresenta uma distribuição de freqüência das áreas de 400 lotes:

AREAS (m^2)	N.º de Lotes
300 |— 400	14
400 |— 500	46
500 |— 600	58
600 |— 700	76
700 |— 800	68
800 |— 900	62
900 |— 1.000	48
1.000 |— 1.100	22
1.100 |— 1.200	6

2) Com referência a tabela, determine:

a) a amplitude total
b) o limite superior da quinta classe
c) o limite inferior da oitava classe
d) o ponto médio da sétima classe
e) a amplitude do intervalo da segunda classe
f) a freqüência da quarta classe
g) a freqüência relativa da sexta classe
h) a freqüência acumulada da quinta classe
i) o número de lotes cuja área não atinge 700 m^2
j) o número de lotes cuja área atinge e ultrapassa 800 m^2
l) a percentagem dos lotes cuja área não atinge 600 m^2
m) a percentagem dos lotes cuja área seja maior ou igual a 900 m^2
n) a percentagem dos lotes cuja área é de 500 m^2, no mínimo, mas inferior a 1.000 m^2
o) a classe do 72º lote
p) até que classe estão incluídos 60% dos lotes

5 - REPRESENTAÇÃO GRÁFICA DE UMA DISTRIBUIÇÃO

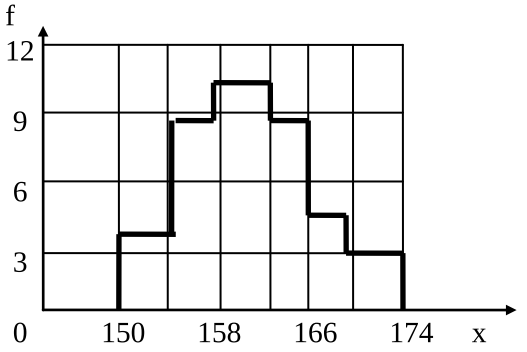
Uma distribuição de freqüência pode ser representada graficamente pelo **histograma**, pelo **polígono de freqüência** e pelo **polígono de freqüência acumulada (ogiva de Galton).**

5.1 - Histograma

O histograma é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de modo que seus pontos médios coincidam com os pontos médios dos intervalos de classes.

As larguras dos retângulos são iguais às amplitudes dos intervalos de classe.

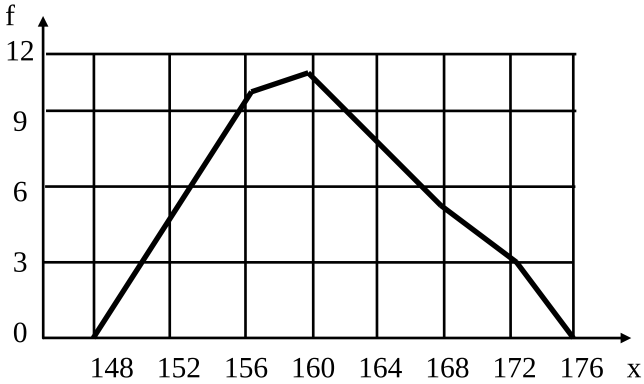
As alturas dos retângulos devem ser proporcionas às freqüências das classes, sendo a amplitude dos intervalos iguais. Isso nos permite tomar as alturas numericamente iguais às freqüências.



5.2 - Polígono de freqüência

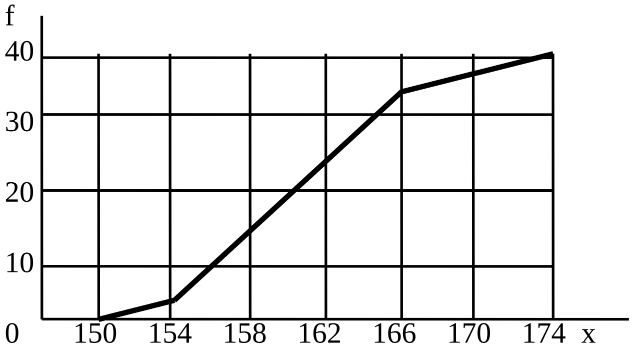
É um gráfico em linha, sendo as freqüências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe.

Para realmente obtermos um polígono (linha 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.



5.3 - Polígono de freqüência acumulada

É traçado marcando-se as freqüências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.



6 - A CURVA DE FREQÜÊNCIA

A partir do polígono de freqüências podemos representar contornos mais suaves (polígono de freqüências polido), utilizando-se curvas para chegarmos a uma das representações de grande utilidade para a Estatística que é a curva de freqüências.

7 - AS FORMAS DAS CURVAS DE FREQÜÊNCIA

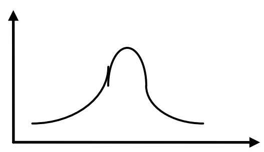
7.1 - Curvas em forma de sino

Caracterizam-se pelo fato de apresentarem um valor máximo na região central.

Distinguimos a curva em forma de sino **simétrica** e a **assimétrica.**

- **Curva simétrica**

Esta curva caracteriza-se por apresentar o valor máximo no ponto central e os pontos eqüidistantes desse ponto terem a mesma freqüência.

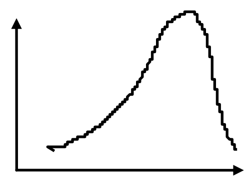


- **Curva assimétrica**

Se a cauda mais alongada fica à direita, a curva é chamada **assimétrica positiva ou enviesada à direita**. Se a cauda se alonga à esquerda, a curva é chamada **assimétrica negativa ou enviesada à esquerda.**

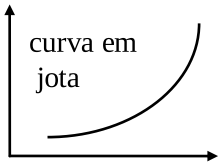
Assimétrica positiva:

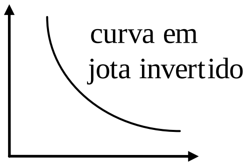
Assimétrica negativa:



7.2 - Curvas em forma de jota

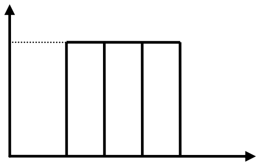
São relativas a distribuições extremamente assimétricas, caracterizadas por apresentarem o ponto de ordenada máxima em uma das extremidades.

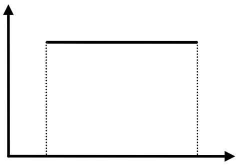




7.3 - Distribuição retangular

Essa distribuição, muito rara na verdade, apresenta todas as classes com a mesma freqüência. Tal distribuição seria representada por um histograma em que todas as colunas teriam a mesma altura ou por um polígono de freqüência reduzido a um segmento de reta horizontal.





EXERCÍCIOS

13) Denominamos ______________ toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou espécie.

a) Rol
b) Série estatística
c) Histograma
d) gráfica

14) Numa Série estatística observamos a existência de três elementos ou fatores:

a) tempo, espaço e a espécie
b) local, tempo e a espécie
c) variável, rol e coordenadas
d) tempo, variável e rol

15) As séries que descrevem os valores da variável, em determinado local, discriminados segundo intervalos de tempo variáveis:

a) Séries geográficas
c) Séries históricas
b) Séries categóricas
d) Séries conjugadas

16) (EAGS) Que tipo de gráfico estatístico é o histograma?

a) Diagrama
b) Cartograma
c) Estereograma
d) Poligonal característica

17) Marque a segunda coluna de acordo com a primeira:

1 - Índice
2 - Coeficientes
3 - Taxas

( ) São as razões entre duas grandezas tais que uma não inclui a outra.
( ) São as razões entre o número de ocorrências e o número total (número de ocorrências e o número de não-ocorrências).
( ) São os coeficientes multiplicado por uma potência de 10.

18) Os gráficos próprios de uma distribuição de freqüência são:

a) colunas, curva de freqüência e histograma
b) polígono de freqüência e histograma
c) colunas, curva de freqüência e polígono de freqüência
d) gráfico em setor, gráfico em barra, curva de freqüência e curva normal
e) colunas, barra, setor e curva de freqüência

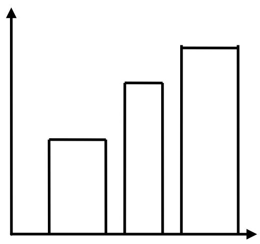
19) São diagramas <u>EXCETO</u> :

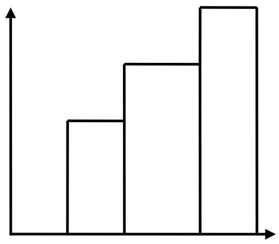
a) gráficos em linha e em barras
b) gráficos em linha e em barras múltiplas
c) gráficos em colunas e em setores
d) gráficos em setores e cartogramas

20) As curvas que caracterizam-se por pelo fato de apresentarem um valor máximo na região central:

a) curva simétrica
b) curva assimétrica
c) curva em forma de jota
d) curva em forma de U

21)





Estes dois gráficos são respectivamente:

a) gráficos em colunas
b) histogramas
c) gráficos em colunas e polígono de freqüência
d) gráfico em colunas e histogramas

UNIDADE VI

1 - MEDIDAS DE POSIÇÃO

As medidas de posição são elementos estatísticos que representam uma série de dados orientando-nos quanto a posição da distribuição em relação ao eixo horizontal (eixo das abcissas).

As medidas de posição mais importantes **são as medidas de tendência central**, que recebem tal denominação pelo fato de os dados observados tenderem, em geral, a se agrupar em torno dos valores centrais. Dentre as medidas de tendência central, destacamos: a média aritmética, a mediana e a moda.

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

1.1 - MÉDIA ARITMÉTICA ($\overline{X}$ ou μ)

1.1.1 - Dados não grupados

É obtida através da **média aritmética simples**, isto é; do quociente entre a soma dos valores observados e o seu número total.

Exemplo:

x = {1,3,5,7,9}

$$x = \frac{1+3+5+7+9}{5} = 5 \quad \therefore \quad \overline{x} = 5$$

1.1.2 - Desvio em relação à média (d_1)

É a diferença entre cada elemento de um conjunto de valores e a média aritmética.

$$\boxed{d_i = x_i - \overline{x}}$$

Para o exemplo dado, temos:

$d_1 = x_1 - \overline{x} = 1 - 5 = -4$

$d_2 = x_2 - \overline{x} = 3 - 5 = -2$

$d_3 = x_3 - \overline{x} = 5 - 5 = 0$

$d_4 = x_4 - \overline{x} = 7 - 5 = 2$

$d_5 = x_5 - \overline{x} = 9 - 5 = 4$

1.1.3 - Propriedades da média

1ª propriedade

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

$$\boxed{\sum d_i = 0}$$

No exemplo anterior temos:

$\sum d_i = (-4) + (-2) + 0 + 2 + 4 = 0$

$\sum d_i = 0$

2ª propriedade

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

$$\boxed{y_i = x_i \pm c \Rightarrow \overline{y} = \overline{x} \pm c}$$

Exemplo:

x = {1,3,5,7,9} com $\overline{x} = 5$

fazendo c = 10 e somando, teremos:

Y = {11,13,15,17,19} com $\overline{y} = 15$

Podemos ver que $\overline{y} = 15$ é igual a $\overline{x} = 5$ acrescida de c = 10, ou seja, $\overline{y} = 5 + 10 = 15$

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:

$$\boxed{y_i = x_i \,.\, c \Rightarrow \overline{y} = \overline{x} \,.\, c} \quad \text{ou} \quad \boxed{y_i = \frac{x_i}{c} \Rightarrow \overline{y} = \frac{\overline{x}}{c}}$$

Exemplo:

x = {100, 200, 300, 400 ,500} com $\overline{x} = 300$

fazendo c = 10 e dividindo, teremos:

Y = {10, 20, 30, 40 ,50} com $\overline{y} = 30$

A mesma forma, podemos ver que $\overline{y} = 30$ é igual a $\overline{x}$ =300 dividido por c = 10, ou seja, $\overline{y} = \frac{300}{10} = 30$

1.1.4 - Dados agrupados

Neste caso, como as freqüências 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 forma: $$\boxed{\overline{x} = \frac{\sum x_i f_i}{\sum f_i}}$$

Ex.:

x_i	f_i	$x_i f_i$
4	1	4 . 1 = 4
5	5	5 . 5 = 25
6	6	6 . 6 = 36
7	5	7 . 5 = 35
8	3	8 . 3 = 24
soma	20	124

$$\overline{x} = \frac{124}{20} \quad \Rightarrow \quad \boxed{\overline{x} = 6{,}2}$$

1.1.5 - Com intervalos de classe

Neste caso, convencionamos que todos os valores incluídos coincidem com o seu ponto médio, e determinamos a média aritmética ponderada por meio da fórmula:

$$\boxed{\overline{x} = \frac{\sum x_i f_i}{\sum f_i}}$$

Exemplo:

Abrindo uma coluna para os pontos médios e outra para os produtos $x_i f_i$:

i	ESTATURA (cm)	f_i	x_i	$x_i f_i$
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
		$\sum$=40		$\sum$=6.440

Temos:

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} \Rightarrow \overline{x} = \frac{6.440}{40} \Rightarrow \overline{x} = 161\,\text{cm}$$

EXERCÍCIOS:

22) Calcule a média aritmética da distribuições abaixo:

a)

Custo (R$)	450|— 550|— 650|— 750|— 850|—950|—1.050|— 1.150
f_i	8 10 11 16 13 5 1

b)

CLASSES	30 |— 50 |— 70 |— 90 |— 110 |—130
f_i	2 8 12 10 5

1.2 - MODA (Mo)

Denominamos moda o valor que ocorre com maior freqüência em uma série de valores.

1.2.1 - Dados não-agrupados

X_1 = {2, 4, 5, 5, **6, 6, 6**, 7, 8, 9}

$\boxed{\text{Mo} = 6}$

X_2 = {2, 2, 3, 3, 4, 4}

$\boxed{\text{CONJUNTO AMODAL}}$, não existe moda

X_3 = {1, **2, 2, 2**, 3, 4, 4, **5, 5, 5**, 6}

$\boxed{\text{CONJUNTO BIMODAL}}$

Mo = 2

Mo = 5

1.2.2 - Dados agrupados

- **Sem intervalos de classe**

Basta fixar o valor da variável de maior freqüência

Exemplo:

x_i	f_i
1	1
2	5
3	6
4	5
5	3
soma	20

Mo = 3

- **Com intervalos de classe**

Utilizamos um dos seguintes métodos :

a) Moda bruta
b) Método de Czuber
c) Método de King

a) **Moda Bruta**

Determinamos a classe modal (de maior freqüência) e depois calculamos o ponto Médio entre os limites de classe.

Exemplo: (Tab. 5.5)

i	ESTATURA (cm)	f_i	
1	150 |— 154	4	
2	154 |— 158	9	
3	158 |— 162	11	→ classe modal
4	162 |— 166	8	
5	166 |— 170	5	
6	170 |— 174	3	

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

Moda Bruta = 160 cm

b) **Método de Czuber**

Após determinar a classe modal, calculamos através da seguinte fórmula $\boxed{Mo = \ell^* + \frac{D_1}{D_1 + D_2} \cdot h^*}$

na qual:

ℓ^* é o limite inferior da classe modal

h* é a amplitude de classe modal

D_1 = f* - f (ant)

D_2 = f* - f (post)

sendo:

f* a freqüência simples da classe modal

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

f (post) a freqüência simples da classe posterior à classe modal

Exemplo utilizando a tabela 5.5:

D_1 = 11 - 9 = 2

D_2 = 11 - 8 = 3

$$Mo = 158 + \frac{2}{2+3} \cdot 4 = 159{,}6$$

Moda de Czuber = 159,6 cm

c) **Moda de King**

$$\boxed{Mo = \ell^* + \frac{f_{post}}{f_{ant} + f_{post}} \cdot h^*}$$

Exemplo utilizando a tabela 5.5:

$$Mo = 158 + \frac{8}{9+8} \cdot 4 = 159{,}9$$

Moda de King = 159,9 cm

Nota: dos três métodos, o método considerado mais preciso é o de **Czuber**.

EXERCÍCIO:

23) Calcule as três modas da distribuição abaixo

Custo (R$)	450 ⊢ 550	550 ⊢ 650	650 ⊢ 750	750 ⊢ 850	850 ⊢ 950	950 ⊢ 1.050	1.050 ⊢ 1.150
f_i	8	10	11	16	13	5	1

1.3 - MEDIANA (Md)

É uma medida de posição que divide uma série ordenada (Rol) de tal maneira que pelo menos a metade ou 50% dos itens sejam iguais ou maiores que ela. Desta maneira é também uma **separatriz** dividindo a série em partes iguais.

1.3.1 Dados não grupados

a) Quando tivermos um número ímpar de valores

Exemplo:

X = {1, 3, 5, 7, 9, 11, 12, 13, 14}, onde n = 9 (ímpar)

Determina-se o elemento central: $\boxed{E = \frac{n+1}{2}}$ =

$$E = \frac{9+1}{2} = 5$$

Verifica-se, então, na seqüência ordenada os valores correspondentes à posição 5 indicada por E. Vemos assim que a mediana será o valor, Md = 9

b) Quando tivermos um número par de valores

Exemplo:

X = {1, 3, 5, 7, 9, 11}

Determinam-se os elementos centrais: $E = \frac{n}{2} = \frac{6}{2} = 3$

Verifica-se, então, na seqüência ordenada os valores correspondentes às posições 3 e 4 (posição seguinte), indicadas por E. Para a determinação da Mediana, calculamos **a média aritmética dos valores centrais.**

$$Md = \frac{5+7}{2} = 6 \;\therefore\; Md = 6$$

1.3.2 - Dados agrupados

- **Sem intervalos de classe**

Primeiramente determinamos previamente as freqüências acumuladas e dividimos a distribuição em dois grupos que contenham o mesmo número de elementos, através da fórmula: $\frac{\sum f_i}{2}$

Depois indicamos a freqüência acumulada imediatamente superior à metade da soma das freqüências. A mediana será aquele valor da variável que corresponde a tal freqüência acumulada.

Exemplo:

Dada a tabela abaixo

N.º DE MENINOS	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\sum$=34	

Sendo:

$$\frac{\sum f_i}{2} = \frac{34}{2} = 17$$

A menor freqüência 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 freqüência 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 freqüência 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
	$\sum$ = 8	

Temos:

$$\frac{8}{2} = 4 = F_3$$

Logo:

$$Md = \frac{15+16}{2} = \frac{31}{2} = 15{,}5$$

Donde:
Md = 15,5

- **Com intervalos de classe**

Para calcular a mediana executamos os seguintes passos:
1º) Determinamos as freqüências acumuladas.
2º) Calculamos $\frac{\sum f_i}{2}$.
3º) Marcamos a classe correspondente à freqüência acumulada imediatamente superior à - **classe mediana** - e, em seguida, empregamos a fórmula:

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

na qual:
ℓ^* é o limite inferior da classe mediana
F (ant) é a freqüência acumulada da classe anterior à classe mediana;
f* é a freqüência simples da classe mediana;
h* é amplitude do intervalo da classe mediana.

Tomando como exemplo a distribuição anterior, temos:

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

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, obtemos:

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

isto é: Md = 160,5 cm

Observação:
No caso de existir uma freqüência acumulada exatamente igual a $\frac{\sum f_i}{2}$, **a mediana será o limite superior da classe correspondente.**

Exemplo:

i	CLASSES	f_i	F_i	
1	0 ├─ 10	1	1	
2	10 ├─ 20	3	4	
3	20 ├─ 30	9	13	←
4	30 ├─ 40	7	20	
5	40 ├─ 50	4	24	
6	50 ├─ 60	2	26	
		$\sum$=26		

Temos: $\frac{\sum f_i}{2} = \frac{26}{2}$ = 13, Logo Md = L* $\Rightarrow$ Md = 30

NOTAS:

A **média** é utilizada quando:

a) desejamos obter a medida de posição que possui a maior estabilidade;
b) houver necessidade de um tratamento algébrico posterior.

A **moda** é utilizada quando:

a) desejamos obter uma medida rápida e aproximada de posição;
b) a medida de posição dever ser o valor mais típico da distribuição.

A **mediana** é utilizada quando:

a) desejamos obter o ponto que divide a distribuição em partes iguais;
b) há valores extremos que afetam de uma maneira acentuada a média;
c) a variável em estudo é salário.

EXERCÍCIOS

24) Calcule a mediana das distribuições abaixo:

a)

x_i	2	4	6	8	10
f_i	3	7	12	8	4

b)

x_i	0	1	2	3	4	5
f_i	2	5	9	7	6	3

c)

Custo (R$)	450 ⊢ 550	550 ⊢ 650	650 ⊢ 750	750 ⊢ 850	850 ⊢ 950	950 ⊢ 1.050	1.050 ⊢ 1.150
f_i	8	10	11	16	13	5	1

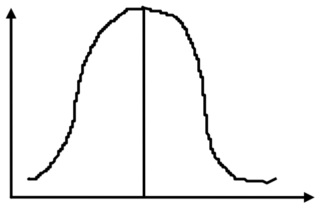
2 - POSIÇÃO RELATIVA DA MÉDIA, MEDIANA E MODA

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

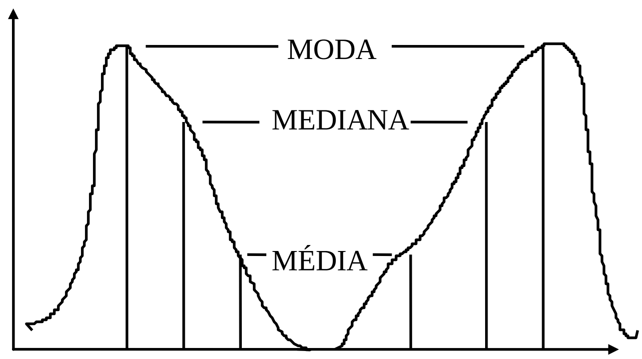
$\bar{x} = Md = Mo$, no caso da **curva simétrica;**

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

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



$\bar{x} = Md = Mo$



$Mo < Md < \bar{x}$ $\qquad$ $\bar{x} < Md < Mo$,

3 - SEPARATRIZES

Além da mediana existem outras medidas que dividem as séries em partes iguais, são os **quartis**, os **percentis** e os **decis**.

3.1 Os Quartis

Denominamos quartis os valores de uma série que a dividem em quatro partes iguais.

Quando os dados são agrupados, para determinar os quartis usamos a mesma técnica do cálculo da mediana, bastando substituir, na fórmula da mediana, $\frac{\sum f_i}{2}$ por $\boxed{\frac{k\sum f_i}{2}}$.

Sendo k o número de ordem do quartil .

Assim temos: $\boxed{Q_k = \ell * + \frac{\left[\frac{k\sum f_i}{4} - F(ant)\right] h *}{f *}}$

Obs.: O segundo quartil coincide com a mediana. (Q2 = Md)

3.2 Os percentis

Denominamos percentis os noventa e nove valores que separam uma série em 100 partes iguais.

O cálculo de um percentil segue a mesma técnica do cálculo da mediana, onde utilizamos:

$$\boxed{\frac{k\sum f_i}{100}} \quad e \quad \boxed{P_k = \ell * + \frac{\left[\frac{k\sum f_i}{100} - F(ant)\right]h*}{f*}}$$

EXERCÍCIO RESOLVIDO

Dada a tabela abaixo calcule o 1º e o 4º quartis, e 8º percentil:

i	ESTATURA (cm)	f_i	F_i
1	150 ├─ 154	4	4
2	154 ├─ 158	9	13
3	158 ├─ 162	11	24
4	162 ├─ 166	8	32
5	166 ├─ 170	5	37
6	170 ├─ 174	3	40
		$\sum$=40	

a) Primeiro quartil

$$\frac{\sum f_i}{4} = \frac{40}{4} = 10$$

$$Q_1 = 154 + \frac{(10-4)4}{9} =$$

= 154 + 2,66 = 156,66

Q_1 = 156,7 cm

b) Terceiro quartil

$$\frac{3\sum f_i}{4} = \frac{3 \times 40}{4} = 30$$

$$Q_3 = 162 + \frac{(30-24)4}{8} =$$

= 162 + 3 = 165

Q_3 = 165 cm

c) oitavo percentil

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

$$P_8 = 150 + \frac{[3{,}2-0]4}{4} = 150 + \frac{12{,}8}{4} = 153{,}2$$

P_8 = 153,2 cm

EXERCÍCIO

25) Dada a distribuição abaixo, calcule:
a) O 1º e o 3º quartis
b) O 10º, 1º, 23º, 15º e o 90º percentis.

i	ESTATURA (cm)	f_i
1	150 ├─ 158	5
2	158 ├─ 166	12
3	166 ├─ 174	18
4	174 ├─ 182	27
5	182 ├─ 190	8
		$\sum$=70

26) A diferença entre cada elemento de um conjunto de valores e a média aritmética
a) mediana
b) desvio
c) assimetria
d) separatriz

27) Quando desejamos obter a medida de posição que possui a maior estabilidade utilizamos:
a) a média
b) a mediana
c) a moda
d) a amplitude

28) A moda considerada mais precisa é:
a) moda bruta
b) moda de King
c) moda de Czuber
d) todas dão valores iguais

29) Quando desejamos obter o ponto que divide a distribuição em partes iguais empregamos
a) a moda bruta
b) a moda de King
c) a mediana
d) a média

30) O Professor bigode, após verificar que toda classe obteve nota baixa, eliminou as questões que não foram respondidas pelos alunos. Com isso, as notas de todos os alunos foram aumentadas de três pontos. Então:
a) a média aritmética ficou alterada, assim como a Md.
b) apenas a média aritmética ficou alterada
c) apenas a mediana ficou alterada.
d) não houve alteração nem na média nem na mediana
e) nada podemos afirmar sem conhecer o número total de alunos.

UNIDADE VII

1 - MEDIDAS DE DISPERSÃO OU DE VARIABILIDADE

São as medidas que qualificam os valores de uma dada variável, ressaltando a maior ou menor dispersão ou variabilidade entre esses valores e sua medida de posição.

Dessas medidas, estudaremos a amplitude total, a variância, o desvio padrão e o coeficiente de variação.

2 - AMPLITUDE TOTAL

A amplitude total é a diferença entre o maior e o menor valor observado:

$$\boxed{AT = x\,(\text{máx.}) - x\,(\text{min.})}$$

2.1 - Dados não grupados

Ex.: {40, 45, 48, 52, 54, 62 e 70}
AT = 70 - 40
AT = 30

2.2 - Dados agrupados

- **Sem intervalos de classe**

Dada a tabela

x	0	1	2	3	4
f	2	6	12	7	3

AT = 4 - 0
AT = 4

- **Com intervalos de classe**

Dada a distribuição 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
		$\sum f_i = 40$

Utilizamos a fórmula AT = L (máx.) - ℓ (min.)
AT = 174 - 150
AT = 24 cm

Observações quanto a amplitude total:

a) Quanto maior a AT, maior a dispersão ou a variabilidade dos valores da variável.
b) Ela é apenas uma indicação aproximada da dispersão ou variabilidade.

3 - VARIÂNCIA/DESVIO PADRÃO

A variância e o desvio padrão são medidas que não se deixam influenciar pelos valore extremos, levam em consideração a totalidade dos valores da variável em estudo, o que faz delas índices de variabilidades bastantes estáveis e, por isso mesmo, os mais geralmente empregados.

A variância é a média aritmética dos quadrados dos desvios, ou seja:

$$\boxed{\sigma^2 = \frac{\sum(x_i - \bar{\mu})^2}{n}} \text{ ou } \boxed{\sigma^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

O desvio padrão é a raiz quadrada da variância, ou seja:

$$\boxed{\sigma = \sqrt{\frac{\sum(x_i - \bar{\mu})^2}{n}}} \text{ ou } \boxed{\sigma = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}}$$

Observações:

1) $n = \sum f_i$.
2) Nas amostras utilizamos n – 1, a letra s no lugar de σ e a letra $\bar{x}$ no lugar de $\bar{\mu}$.

Propriedades do desvio padrão:

1ª) somando-se (ou subtraindo-se) uma constante a (de) todos os valores de uma variável, o desvio padrão não se altera:
$y_i = x_i \pm c \Rightarrow \sigma_y = \sigma_x$
2ª) Multiplicando-se todos os valores de uma variável por uma constante (diferente de zero), o desvio padrão fica multiplicado por essa constante:
$y_i = c\,.\,x_i \Rightarrow \sigma_y = c\,.\,\sigma_x$

3.1 - Desvio padrão de dados não agrupados

Sendo x = {40,45,48,52,54,62,70}. Abre-se duas colunas: uma para x_i e outra para x_i^2:

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
$\sum = 371$	$\sum = 20.293$

Como n = 7, temos:

$$\sigma = \sqrt{\frac{20293}{7} - \left(\frac{371}{7}\right)^2} = \sqrt{90} \Rightarrow s = 9{,}49$$

3.2 - Dados agrupados

3.2.1 Sem intervalos de classe

Ao modo mais prático para se obter 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 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
	$\sum = 30$	$\sum = 63$	$\sum = 165$

Como temos freqüências usamos a fórmula

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

logo:

$$\sigma = \sqrt{\frac{165}{30} - \left(\frac{63}{30}\right)^2} = \sqrt{1{,}09} \Rightarrow \sigma = 1{,}044$$

3.2.2 Com intervalos de classe

Abrimos colunas para x_i (ponto médio), para $f_i x_i$ e para $f_i x_i^2$. Assim:

TABELA 6 (baseada na tab. 5.4, pag. 7)

i	ESTATURA (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
	h = 4	$\sum$=40		$\sum$=6.440	$\sum$=1.038.080

Logo:

$$\sigma = \sqrt{\frac{1.038.080}{40} - \left(\frac{6.440}{40}\right)^2} = \sqrt{25.952 - 25.921} =$$

$$\sqrt{31} \Rightarrow \sigma = 5{,}57\,cm$$

3.3 Processo breve

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

$$y_1 = \frac{x_1 - x_0}{h}$$

aplicamos a seguinte fórmula:

$$\boxed{\sigma = h\sqrt{\frac{\sum f_i y_1^2}{n} - \left(\frac{\sum f_i y_i}{n}\right)^2}}$$

Procedimentos para o cálculo do desvio padrão pelo processo breve:

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

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

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

4º) Abrimos uma coluna para os valores de produto $f_i y_i$, conservando os sinais + ou -, e, em seguida, somamos algebricamente esses produtos.

5º) Abrimos uma coluna para os valores do produto $f_i y_1^2$, obtidos multiplicando cada $f_i y_i$ pelo seu respectivo y_i, e, em seguida, somamos esses produtos.

6º) Aplicamos a fórmula.

Exemplo: (baseada na tab. 5.4, pag. 7)

TABELA 6.1

i	ESTATURA (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	$\sum$=40			$\sum$= 10	$\sum$=80

$$\sigma = 4\sqrt{\frac{80}{40} - \left(\frac{10}{40}\right)^2} = 4\sqrt{2 - 0{,}0625} = 4\sqrt{1{,}9375} =$$

$$4 \times 1{,}3919 = 5{,}5676 \Rightarrow \sigma = 5{,}57\,\text{cm}$$

EXERCÍCIOS

31) Calcule o desvio padrão:

a) 8, 10, 11, 15, 16, 18

b)

x_i	1	2	3	4	5	6
f_i	2	5	8	6	3	1

c)

CLASSES	30 ⊢ 50	50 ⊢ 70	70 ⊢ 90	90 ⊢ 110	110 ⊢ 130
f_i	2	8	12	10	5

4 - COEFICIENTE DE VARIAÇÃO

O **coeficiente de variação (CV)** é a relação entre o desvio padrão e a média das variáveis:

$$\boxed{CV = \frac{\sigma}{\overline{\mu}} \times 100}$$

Quanto menor esse valor, mais homogêneo será o conjunto de dados.

Exemplo 1:
Para a distribuição da tabela 6.1, onde $\overline{\mu}$ = 161 cm e σ = 5,57 cm, temos:

$$CV = \frac{5{,}57}{161} \times 100 = 0{,}03459 \text{ x } 100 = 3{,}459$$

Exemplo 2:
Tomemos os resultados das medidas das estaturas e dos pesos de um mesmo grupo de indivíduos:

	$\overline{\mu}$	σ
ESTATURAS	175 cm	5,0 cm
PESOS	68 kg	2,0 kg

Temos:

$$CV_E = \frac{5}{175} \times 100 = 0{,}0285 \text{ x } 100 = 2{,}85\ \%$$

$$CV_P = \frac{2}{68} \times 100 = 0{,}0294 \text{ x } 100 = 2{,}94\ \%$$

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

EXERCÍCIOS:

32) Sabendo que um conjunto de valores apresenta para média aritmética e para desvio padrão, respectivamente, 18,3 e 1,47, calcule o coeficiente de variação.

33) Um grupo de 85 moças tem estatura média de 160,6 cm, com um desvio padrão igual a 5,97 cm. Outro grupo de 125 moças tem uma estatura média de 161,9 cm, sendo o desvio padrão igual a 6,01 cm. Qual é o coeficiente de variação de cada um dos grupos? Qual o grupo mais homogêneo?

UNIDADE VIII

1 - ASSIMETRIA

Baseando nas relações entre a média e a moda, podemos empregá-las para determinar o tipo de assimetria. Assim, calculando o valor da diferença:

$$\boxed{\overline{\mu} = Mo}$$

se:

$\overline{\mu}$ **- Mo = 0** ⇒ assimetria nula ou distribuição simétrica;

$\overline{\mu}$ **- Mo < 0** ⇒ assimetria negativa ou à esquerda;

$\overline{\mu}$ **- Mo > 0** ⇒ assimetria positiva ou à direita.

Exemplo:

DISTRIBUIÇÃO A:
$\overline{\mu}$ = 12 kg, Md = 12 kg, Mo = 12 kg, σ = 4,42 kg

DISTRIBUIÇÃO B:
$\overline{\mu}$ = 12,9 kg, Md = 13,5 kg, Mo = 16 kg, σ = 4,20 kg

DISTRIBUIÇÃO C:
$\overline{\mu}$ = 11,1 kg, Md = 10,5 kg, Mo = 8 kg, σ = 4,20 kg

Logo:

A. 12 - 12 = 0 ⇒ a distribuição é simétrica.
B. 12,9 - 16 = - 3,1 kg ⇒ a distribuição é assimétrica negativa.
C. 11,1 - 8 = 3,1 kg ⇒ a distribuição é assimétrica positiva.

1.2 - COEFICIENTE DE ASSIMETRIA

Para uma comparação mais precisa de duas distribuições utilizamos o **coeficiente de assimetria de Pearson**, dado por:

$$\boxed{As = \frac{3(\overline{\mu} - Md)}{\sigma}}$$

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

Exemplo:
Considerando as distribuições A, B, e C dadas anteriormente, temos:

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

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

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

EXERCÍCIOS:

34) Uma distribuição de freqüência apresenta as seguintes medidas: $\overline{\mu} = 48{,}1$; e Md = 47,9 e $\sigma = 2{,}12$. Calcule o coeficiente de assimetria.

35) Considerando a distribuição de freqüência relativa aos pesos de 100 operário de uma fábrica:

PESOS	50 ⊢ 58	58 ⊢ 66	66 ⊢ 74	74 ⊢ 82	82 ⊢ 90	90 ⊢ 98
N.º OPER.	10	15	25	24	16	10

determine o grau de assimetria.

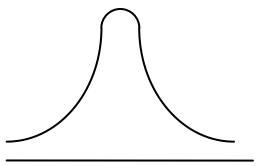
2 - CURTOSE

Denominamos curtose o grau de achatamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal (curva correspondente a uma distribuição teórica de probalidade).

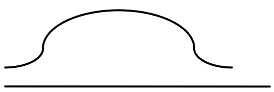
Quando a distribuição apresenta uma curva de freqüência mais fechada que a normal (ou mais aguda em sua parte superior), ela recebe o nome de **leptocúrtica.**

Quando a distribuição apresenta uma curva de freqüência mias aberta que a normal (ou mais achatada na sua parte superior) ela é chamada **platicúrtica.**

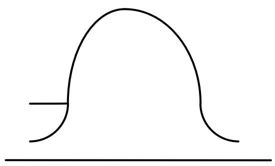
A curva normal, que é a nossa base referencial, recebe o nome de mesocúrtica.



Leptocúrtica



platicúrtica



mesocúrtica

2.1 - Coeficiente de curtose

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

Essa fórmula é conhecida como coeficiente percentílico de curtose.

Relativamente à curva normal, temos:

c = 0,263

Assim:

$c = 0{,}263 \Rightarrow$ curva mesocúrtica

$c < 0{,}263 \Rightarrow$ curva leptocúrtica

$c > 0{,}263 \Rightarrow$ curva platicúrtica

Exemplo:

Sabendo-se que uma distribuição apresenta as seguintes medidas:

$Q_1 = 24{,}4$ cm, $Q_3 = 41{,}2$ cm, $P_{10} = 20{,}2$ cm e $P_{90} = 49{,}5$ cm, temos:

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

C = 0,287

Como:

0,287 > 0,263,

concluímos que a distribuição é platicúrtica, em relação à normal.

EXERCÍCIO:

36) Determine o grau de curtose e classifique a distribuição em relação à curva normal:

PESOS	50 ⊢ 58	58 ⊢ 66	66 ⊢ 74	74 ⊢ 82	82 ⊢ 90	90 ⊢ 98
N.º OPER.	10	15	25	24	16	10

UNIDADE IX

1- ARREDONDAMENTOS DE DADOS

De acordo com a resolução 886/66 da fundação IBGE, o arredondamento é feito da seguinte maneira:

- Quando o primeiro algarismo a ser abandonado é 0,1,2,3 ou 4, fica inalterado o último algarismo a permanecer.

Exemplo: 53,24 passa a 53,2

- Quando o primeiro algarismo a ser abandonado é 6,7,8 ou 9, aumenta-se de uma unidade o algarismo a permanecer.

Exemplos: 42,87 passa a 42,9/ 25,08 passa a 25,1

- Quando o primeiro algarismo a ser abandonado é 5, há duas soluções:

a) Se ao 5 seguir em qualquer casa um algarismo diferente de zero, aumenta-se uma unidade ao algarismo a permanecer.

Exemplos:

2,352 passa a 2,4
25,6501 passa a 25,7
76,250002 para 76,3

b) Se o 5 for o último algarismo ou se ao 5 só se seguirem zeros, o último algarismos a ser conservado só será aumentado de uma unidade se for ímpar.

Exemplos:
24,75 passa a 24,8
24,65 passa a 24,6
24,75000 passa a 24,8
24,6500 passa a 24,6

2 - COMPENSAÇÃO

Suponhamos os dados abaixo, aos quais aplicamos as regras do arredondamento:

	25,32		25,3
	17,85		17,8
	10,44		10,4
+	31,17	+	31,2
	84,78		84,8 (?)
			(84,7)

Verificamos que houve uma pequena discordância: a soma é exatamente 84,7 quando, pelo arredondamento, deveria ser 84,8. Entretanto, para que a apresentação dos resultados, é necessário que desapareça tal diferença, o que é possível pela prática do que denominamos **compensação,** conservando o mesmo número de casas decimais.

Praticamente, usamos "descarregar" a diferença na(s) maior(es) parcela(s). Assim, passaríamos a ter:

	25,3
	17,8
	10,4
+	31,3
	84,8

NOTA:
Convém, ainda, observar que, se a maior parcela é igual ao dobro de qualquer outra parcela (ou maior que esse dobro), descarregamos a diferença (maior que uma unidade) apenas na maior parcela.

EXERCÍCIOS

37) Arredonde para o centésimo mais próximo e compense, se necessário:
0,060 + 0,119 + 0,223 + 0,313 + 0,164 + 0,091 + 0,030 = 1,000

Arredonde para o centésimo mais próximo e compense, se necessário:
4,0 + 7,6 + 12,4 + 27,4 + 11,4 + 8,0 = 70,8

RESPOSTAS DE ALGUNS EXERCÍCIOS:

1)d	7)a
2) a	8)b
3)b	9)c
4)c	10)c
5)c	12)a) 900 b) 800 c) 1.000 d) 950
6)c	e) 100 f) 76 g) 0,155 h) 262 i) 194
	j) 138 l) 29,5 % m) 19 % n)78%
	o) i = 3 p) i = 5
13)b	20)a
14)a	21)e
15)c	22) a) 755 b) 84,3
16)a	26)b
17) (1) (2) (3)	27)a
18)b	28)c
19)d	29)c
	30)a
	31) a) s = 3,56 b) s = 1,24
	c) 21,88