ESTATÍSTICA

Introdução

Diariamente estamos envolvidos em análises estatísticas, por exemplo, quando você é abordado na rua para responder qual o candidato irá votar na próxima eleição, quando o IBGE faz uma visita a sua casa para o censo... Desta forma você está fazendo parte da estatística. Mas não é só desta forma que você faz parte do infinito mundo da estatística. Quando você está desempregado ou empregado, está fazendo parte da estatística, quando seu salário aumenta ou diminui, faz parte também. Podemos ver que em quase tudo podemos empregar a estatística.

A estatística, como parte da matemática aplicada, trata da coleta, da análise e da interpretação de dados observados. Estudando os mais variados fenômenos das diversas áreas do conhecimento, ela representa um valioso instrumento de trabalho nos dias de hoje.

CONCEITOS ESTATÍSTICOS

Estatística Descritiva

Pode ser definida como os métodos que envolvem a coleta, a apresentação e a caracterização de um conjunto de dados de modo a descrever apropriadamente as várias características deste conjunto.

Embora os métodos estatísticos descritivos sejam importantes para a apresentação e a caracterização dos dados, foi o desenvolvimento de métodos estatísticos de inferência, como um produto de teoria da probabilidade, que levou à ampla aplicação da estatística em todos os campos de pesquisas atuais.

Inferência Estatística

Pode ser definida como os métodos que tornam possível a estimativa de uma característica de uma população ou a tomada de uma decisão referente à população com base somente em resultados de amostras.

Para tornar mais claro esta definição, as definições seguintes são necessárias:

Uma **população** é a totalidade dos itens ou objetos a ser considerado.

Uma **amostra** é a parte da população selecionada para análise.

Um **parâmetro** é a medida calculada para descrever uma característica de toda uma população.

Uma **estatística** é a medida calculada para descrever uma característica de apenas uma amostra da população.

Variáveis Estatísticas

Em estatística, uma variável é um atributo mensurável que tipicamente varia entre indivíduos.

Variável Quantitativa - São aquelas que são numericamente mensuráveis, por exemplo, a idade, a altura, o peso. Estas ainda se subdividem em:

Variável Quantitativa Continua: São aquelas que assumem valores dentro de um conjunto contínuo, tipicamente os números reais. São exemplos, o peso ou a altura de uma pessoa.

Variável Quantitativa Discreta: São aquelas que assumem valores dentro de um tempo finito ou enumerável, tipicamente números inteiros. Um exemplo é o número de filhos de uma pessoa.

Variável Qualitativa - São aquelas que se baseiam em qualidades e não podem ser mensuráveis numericamente.

Exercícios:

Identifique a população e as variáveis e classifique-as como qualitativas ou quantitativas: contínuas (C) ou discretas (D):

a) Altura de precipitação de chuvas em um local.

b) Valores das ações vendidas na bolsa de valores.

c) Quantidade de ações vendidas na bolsa de valores.

d) Número de pétalas em ma flor.

e) Velocidade de um automóvel.

f) Cor dos cabelos dos alunos em uma escola.

g) Número de filhos dos casais residentes em uma cidade.

h) O ponto obtido em cada jogada nas jogadas de um dado.

i) Número de peças produzidas por hora, entre as peças produzidas por certa máquina.

j) Diâmetro externo das peças produzidas por certa máquina.

k) Número de ações negociadas na Bolsa de Valores de São Paulo.

l) Comprimento dos pregos produzidos por certa máquina.

Arredondamento De Dados

Conforme critério universal adotado pela estatística o arredondamento de dados é feito da seguinte forma:

1. Se o 1º algarismo a ser abandonado for menor que "5" o último a permanecer fica inalterado. (arredondamento por falta)

Arredondar para centésimos os números abaixo.

a) 47,3227 ≅ 47,32   b) 0,29364 ≅ 0,29
c) 53,77474 ≅ 53,77   d) 30,00132 ≅ 30,00

2. O último algarismo a permanecer será acrescido de uma unidade se o 1º algarismo a abandonar for maior que "5". (arredondamento por excesso)

Arredondar para décimos os números abaixo.

a) 1,4632 ≅ 1,5   b) 23,09425 ≅ 23,1
c) 38,97777 ≅ 39,0   d) 74,28583 ≅ 74,3

3. Quando o "5" for o 1º algarismo a ser abandonado teremos duas soluções:
   a) Quando o "5" for o único algarismo ou se a ele só se seguirem zeros o último algarismo a permanecer se for ímpar, será acrescido de uma unidade.
   b) Se após o "5" houver em qualquer casa um número diferente de zero, o último algarismo a permanecer será acrescido de uma unidade.

Ex: Arredondar para milésimos os números abaixo:

a) 13,474503≅ 13,475   b) 29,87350 ≅ 29,874
c) 5,55555 ≅ 5,556   d) 0,138500 ≅ 0,138
e) 20,797504 ≅ 20,798   f) 99,99950 ≅ 100,000

Exercícios

01- Os dados abaixo são os índices (em %) alcançados por algumas escolas, em relação à taxa de ocupação nas escolas públicas. Arredondamento para inteiros, unidade.

a) 85,4 ____________   b) 75,7 ___________   c) 85,0 ______________
d) 75,99____________   e) 85,5 ___________   f) 75,55 _____________
g) 95,05 ___________   h) 65,6 ___________   i) 65,3 ______________

02- Em uma pesquisa sobre o tempo, em minutos, gasto pelos alunos da Escola X para realizar uma atividade em sala de aula observou-se os seguintes dados. Faça os arredondamentos para décimos.

a) 35,94   b) 18,09   c) 18,009
d) 19,55   e) 19,93   f) 29,97
g) 10,05   h) 10,55   i) 16,66
j) 18,88   l) 10,00   m) 26,06
n) 16,04   o) 17,65   p) 17,75

03- Utilize os dados do exercício 2 e faça o arredondamento para inteiros, unidade.

AMOSTRAGEM

É o estudo de um pequeno grupo de elementos retirado de uma população que se pretende conhecer. Esses pequenos grupos retirados da população são chamados de Amostras.

Como a amostragem considera apenas parte da população, diferentemente de um censo, o tempo para análise e o custo são menores, além de ser mais fácil e gerar resultados satisfatórios.

Quando não se deve realizar um estudo por amostragem?

Quando o tamanho da amostra é grande em relação ao tamanho da população, ou quando se exige o resultado exato, ou quando já se dispõe dos dados da população, é recomendado realizar um censo, que considera todos os elementos da população.

A partir das três perguntas anteriores, vamos aprender a realizar um estudo por amostragem, conhecendo suas diferentes técnicas.

Para realizar um estudo por amostragem, a amostra deve ser representativa da população estudada. Para isso, existem técnicas adequadas para cada tipo de situação.

Técnicas Probabilísticas (aleatórias)

As técnicas probabilísticas garantem a possibilidade de realizar afirmações sobre a população com base nas amostras. Normalmente, todos os elementos da população possuem a mesma probabilidade de serem selecionados. Assim, considerando N como o tamanho da população, a probabilidade de cada elemento ser selecionado será 1/N. Estas técnicas garantem o acaso na escolha.

São técnicas probabilísticas:

<u>Amostragem Aleatória Simples</u>

É o processo mais elementar e freqüentemente utilizado. Pode ser realizado numerando-se os elementos da população de **1** a **n** e sorteando-se, por meio de um dispositivo aleatório qualquer, **X** números dessa seqüência, que corresponderão aos elementos pertencente à amostra.

Exemplo

Obter uma amostra representativa, de 10%, de uma população de 200 alunos de uma escola.

1º) Numerar os alunos de 1 a 200;

2º) Escrever os números de 1 a 200 em pedaços de papel e colocá-los em uma urna;

3º) Retirar 20 pedaços de papel, um a um, da urna, formando a amostra da população.

Nesta técnica de amostragem, todos os elementos da população têm a mesma probabilidade de serem selecionados: 1/N, onde N é o número de elementos da população.

<u>Amostragem Estratificada</u>

Quando a população possui características que permitem a criação de subconjuntos, as amostras extraídas por amostragem simples são menos representativas. Nesse caso, é utilizada a amostragem estratificada.

Como a população se divide em subconjuntos, convém que o sorteio dos elementos leve em consideração tais divisões, para que os elementos da amostra sejam proporcionais ao número de elementos desses subconjuntos.
Exemplo:
Em uma população de 200 alunos, há 120 meninos e 80 meninas. Extraia uma amostra representativa, de 10%, dessa população.
Nesse exemplo, há uma característica que permite identificar 2 subconjuntos, a característica Sexo. Considerando essa divisão, vamos extrair a amostra da população.

SEXO	POPULAÇÃO	AMOSTRA (10%)
Masculino	120	12
Feminino	80	8
Total	200	20

Portanto, a amostra deve conter 12 alunos do sexo masculino e 8 do sexo feminino, totalizando 20 alunos, que correspondem a 10% da população.
Para selecionar os elementos da população para formar a amostra, podemos executar os seguintes passos:
1º) Numerar os alunos de 1 a 200, sendo os meninos numerados de 1 a 120 e as meninas, de 121 a 200;
2º) Escrever os números de 1 a 120 em pedaços de papel e colocá-los em uma urna A;
3º) Escrever os números de 121 a 200 em pedaços de papel e colocá-los em uma urna B;
4º) Retirar 12 pedaços de papel, um a um, da urna A, e 8 da urna B, formando a amostra da população.
São exemplos desta técnica de amostragem as pesquisas eleitorais por região, cidades pequenas e grandes, área urbana e área rural, sexo, faixa etária, faixa de renda, etc.

Amostragem Sistemática

Esta técnica de amostragem em populações que possuem os elementos ordenados, em que não há a necessidade de construir um sistema de referência. Nesta técnica, a seleção dos elementos que comporão a amostra pode ser feita por um sistema criado pelo pesquisador.

Exemplo

Obter uma amostra de 80 casas de uma rua que contém 2000 casas. Nesta técnica de amostragem, podemos realizar o seguinte procedimento:
1º) Como 2000 dividido por 80 é igual a 25, escolhemos, por um método aleatório qualquer, um número entre 1 e 25, que indica o primeiro elemento selecionado para a amostra.
2º) Consideramos os demais elementos, periodicamente, de 25 em 25.
Se o número sorteado entre 1 e 25 for o número 8, a amostra será formada pelas casas: 8ª, 33ª, 58ª, 83ª, 108ª, etc.
Apesar de esta técnica ser de fácil execução, há a possibilidade de haver ciclos de variação, que tornariam a amostra não-representativa da população.

Amostragem por Conglomerados

Esta técnica é usada quando a identificação dos elementos da população é extremamente difícil, porém pode ser relativamente fácil dividir a população em conglomerados (subgrupos) heterogêneos representativos da população global.

A seguir, é descrito o procedimento de execução desta técnica:

1º) Seleciona uma amostra aleatória simples dos conglomerados existentes;

2º) Realizar o estudo sobre todos os elementos do conglomerado selecionado.

São exemplos de conglomerados: quarteirões, famílias, organizações, agências, edifícios, etc.

Exemplo:

Estudar a população de uma cidade, dispondo apenas do mapa dos quarteirões da cidade.

Neste caso, não temos a relação dos moradores da cidade, restando o uso dos subgrupos heterogêneos (conglomerados). Para realizar o estudo estatístico sobre a cidade, realizaremos os seguintes procedimentos:

1º) Numerar os quarteirões de **1** a **n**;

2º) Escrever os números de **1** a **n** em pedaços de papel e colocá-los em uma urna;

3º) Retirar um pedaço de papel da urna e realizar o estudo sobre os elementos do conglomerado selecionado.

Técnicas Não-Probabilísticas (não-aleatórias)

São técnicas em que há uma escolha deliberada dos elementos da população, que não permite generalizar os resultados das pesquisas para a população, pois amostras não garantem a representatividade desta.

São técnicas não-probabilísticas:

Amostragem Acidental

Trata-se da formação de amostras por aqueles elementos que vão aparecendo. Este método é utilizado, geralmente, em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos.

Exemplo: Pesquisas de opinião em praças públicas, ruas movimentadas de grandes cidades, etc.

Amostragem Intencional

De acordo com determinado critério, é escolhido intencionalmente um grupo de elementos que comporão a amostra. O pesquisador se dirige intencionalmente a grupos de elementos dos quais deseja saber a opinião.

Exemplo: Em uma pesquisa sobre preferência por determinado cosmético, o pesquisador entrevista os freqüentadores de um grande salão de beleza.

Tamanho da amostra

Exemplos:

a) Obter uma amostra representativa para uma pesquisa da estatura de noventa alunos de uma escola. Sugestão usar a 18ª linha da tabela de números aleatórios.

b) Supondo que destes 90 alunos, 54 sejam meninos e 36 meninas, obter a amostra proporcional estratificada.

SEXO	POPULAÇÃO	10%	AMOSTRA
M	54	10X54:100=5,4	5
F	36	10X36:100=3,6	4
TOTAL	90	10X90:100=9,0	9

2) Uma cidade X apresenta a seguinte tabela em relação às suas escolas de Ensino Fundamental:

ESCOLAS	Nº DE ESTUDANTES	
	MASCULINO	FEMININO
A B C D E F	80 102 110 134 150 300	95 120 92 228 130 290
Total	876	955

Obtenha uma amostra proporcional estratificada de 120 estudantes

3) Em uma escola existem 250 alunos, sendo 35 na 1ª série, 32 na 2ª, 30 na 3ª, 28 na 4ª, 35 na 5ª, 32 na 6ª, 31 na 7ª e 27 na 8ª. Obtenha uma amostra de 40 alunos e preencha a tabela seguinte:

Séries	População	Cálculo Proporcional	Amostra
1ª	35	35x49:250=5,6	6
2ª			
3ª			
4ª	28		
5ª			
6ª			
7ª		31x40:250=	
8ª			
Total	250		40

4) Uma escola de 1º grau abriga 124 alunos. Obtenha uma amostra representativa correspondente a 15% da população. Sugestão: Use a 8ª, 9ª e 10ª colunas, a partir da 5ª linha da Tabela de números aleatórios.

5) Uma população é formada por 140 notas resultantes da aplicação de um teste de inteligência:

62	129	95	123	81	93	105	95	96	80	87	110	139	75
123	60	72	86	108	120	57	113	65	108	90	137	74	106
109	84	121	60	128	100	72	119	103	128	80	99	149	85
77	91	51	100	63	107	76	82	110	63	131	65	114	103
104	107	63	117	116	86	115	62	122	92	102	113	74	78
69	116	82	95	72	121	52	80	100	85	117	85	102	106
94	84	123	42	90	91	81	116	73	79	98	82	69	102
100	79	101	98	110	95	67	77	91	95	74	90	134	94
79	92	73	83	74	125	101	82	71	75	101	102	78	108
125	56	86	98	106	72	117	89	99	86	82	57	106	90

Obtenha uma amostra formada de 26 elementos, tomando inicialmente a 1ª linha da esquerda para a direita.

Dados Absolutos: Dados resultantes da coleta direta da fonte, sem qualquer outra manipulação que a contagem ou medição da característica ou fenômeno em estudo.

Dados Relativos: São os resultados de comparações por quociente (ou razão) que se estabelecem entre dados absolutos. Objetiva realçar ou facilitar as comparações entre quantidades, e são normalmente expressos por *PORCENTAGENS*, *ÍNDICES*, *COEFICIENTES* e *TAXAS*.

Índices: São razões entre duas grandezas tais que uma não inclui a outra.

Exemplos: Densidade Demográfica = $\frac{\text{População}}{\text{Superfície}}$

Renda per Capita= $\frac{\text{Renda}}{\text{População}}$

Quociente de Inteligência=$\frac{\text{Idade Mental}}{\text{Idade Cronológica}}$

Coeficientes: São razões entre o número de ocorrências e o número total de casos (ocorrências+não ocorrências). Observe-se que o conseqüente da razão (denominador) inclui o antecedente (numerador).

Exemplos: Coeficiente de Natalidade = $\frac{\text{Nº Nascimentos}}{\text{População}}$

Coeficiente de Aproveitamento = $\frac{\text{Nº Alunos Aprovados}}{\text{Nº Final de Matrículas}}$

Coeficiente de evasão escolar = $\frac{\text{Nº de alunos evadidos}}{\text{Total da População}}$

Taxas: São coeficientes multiplicados por uma potência de 10. No caso específico da segunda potência de 10 (100) temos a ***PORCENTAGEM*.**

Exemplos: Taxa de Mortalidade = Coeficiente de Mortalidadex1.000

Taxa de evasão escolar = coeficiente de evasão escolar X 100

Exercícios:

1) Uma escola apresentava, no final do ano, o seguinte quadro

Séries	Matrículas		Evadidos	Taxa de Evasão
	Março	Novembro		
1ª 2ª 3ª 4ª	480 458 436 420	475 456 430 420		
Total	1794	1781		

a) Calcule a taxa de evasão por série;

b) Calcule a taxa de evasão da escola.

2) Considere a tabela abaixo:

Meses	Valor (U$ milhões)	Taxas percentuais
Janeiro Fevereiro Março Abril	33,3 54,1 44,5 52,9	
Total	184,8	

Evolução das receitas de café industrializado

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?

3) São Paulo tinha em 1989, uma população projetada de 32361700 habitantes. Sabendo que sua área terrestre é de 248256 Km², calcule a sua densidade demográfica.

4) Considerando que Minas Gerais, em 1988, apresentou:
população projetada: 15345800 habitantes
superfície: 586624 Km²
nascimentos: 337889
casamentos: 110473
Calcule:

a) o índice da densidade demográfica;

b) a taxa de natalidade;

c) a taxa de nupcialidade.

GRÁFICOS ESTATÍSTICOS

Representar graficamente significa fazer um desenho que sintetize de maneira clara o comportamento de uma ou mais variáveis.
Existem vários tipos de gráficos. Os melhores são os que primam pela simplicidade e clarezas:

Principais Gráficos

Diagrama Por Linha Poligonal - é a representação gráfica de uma série estatística, por meio de segmentos de retas que une em sequencia os pontos de um sistema cartesiano.
Exs.:
1)

VENDAS

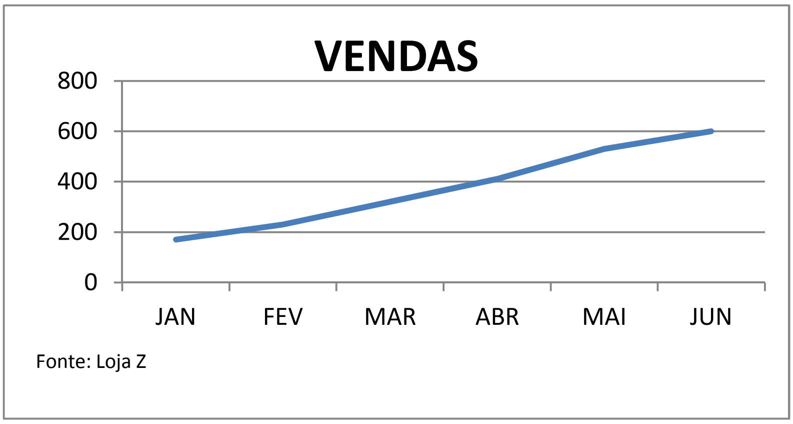
MESES	R$
JAN	170
FEV	230
MAR	320
ABR	410
MAI	530
JUN	600

Fonte: Loja Z

2)

COMÉRCIO EXTERIOR		
	US$ (bilhões)	
ANOS	Importação	Exportação
1996	12	14
1997	15	17,2
1998	17	24,3
1999	19,2	20,4
2000	21,2	22,5
2001	25,4	20,3

Fonte:BB
* Previsão



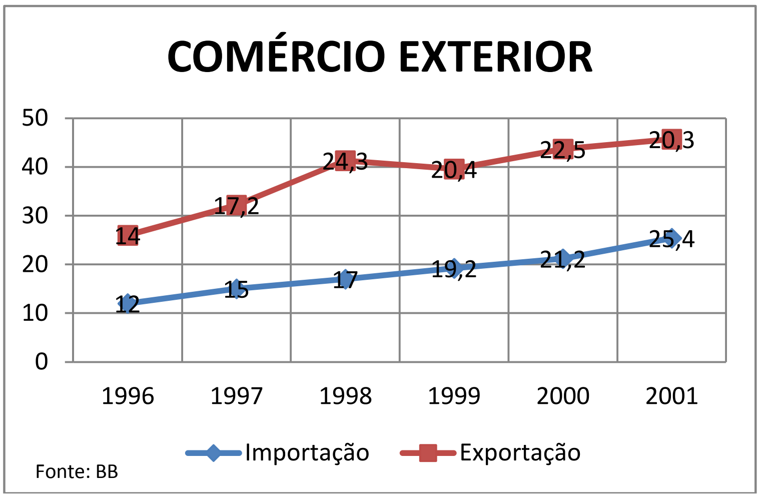
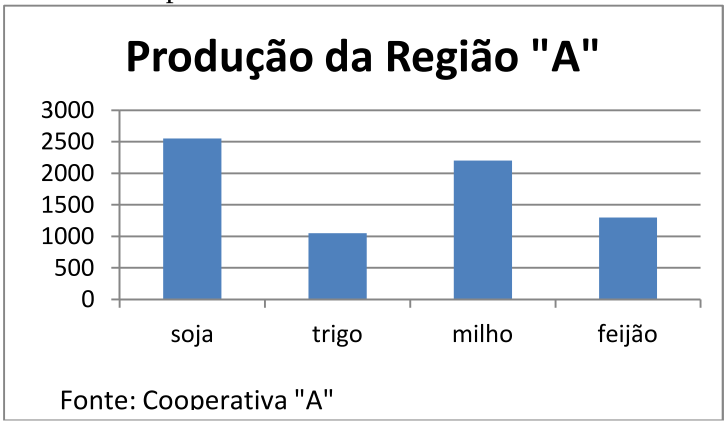


Gráfico Em Colunas: é a representação gráfica de uma série estatística por meio de retângulos dispostos na vertical, com espaços entre eles.

PRODUÇÃO DA REGIÃO "A"

Produtos	Toneladas
soja	2550
trigo	1050
milho	2200
feijão	1300

Fonte: Cooperativa "A"



Gráficos Em Barras: é a representação gráfica de uma série estatística em ordem crescente ou decrescente por meio de retângulos dispostos na horizontal.

MOVIMENTO DA LOJA "IMPORTADOS" - AGO/2000

Modelos	Unidades
BMW	30
HONDA	23
FERRARI	7
GOLF	32
BESTA	48

Fonte: Loja "Importados"

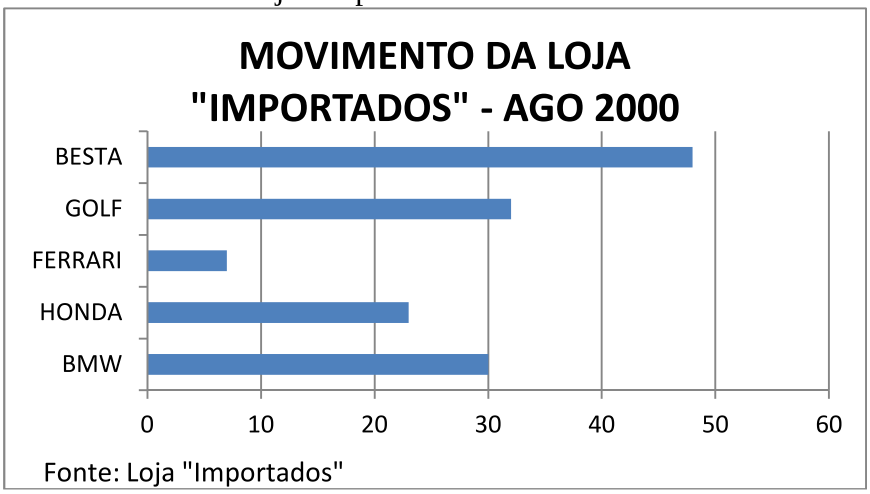
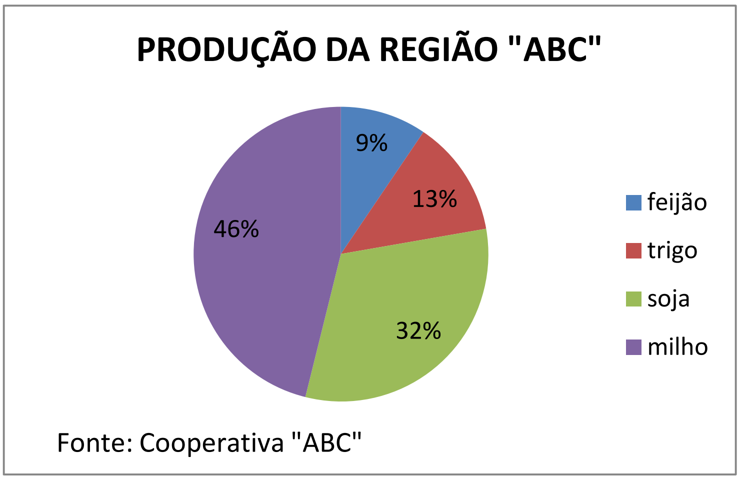


Gráfico Em Setores – é a representação gráfica de uma série estatística por meio de superfícies setoriais.
Ex.:

PRODUÇÃO DA REGIÃO "ABC"

Produtos	Toneladas
feijão	170
trigo	230
soja	570
milho	830

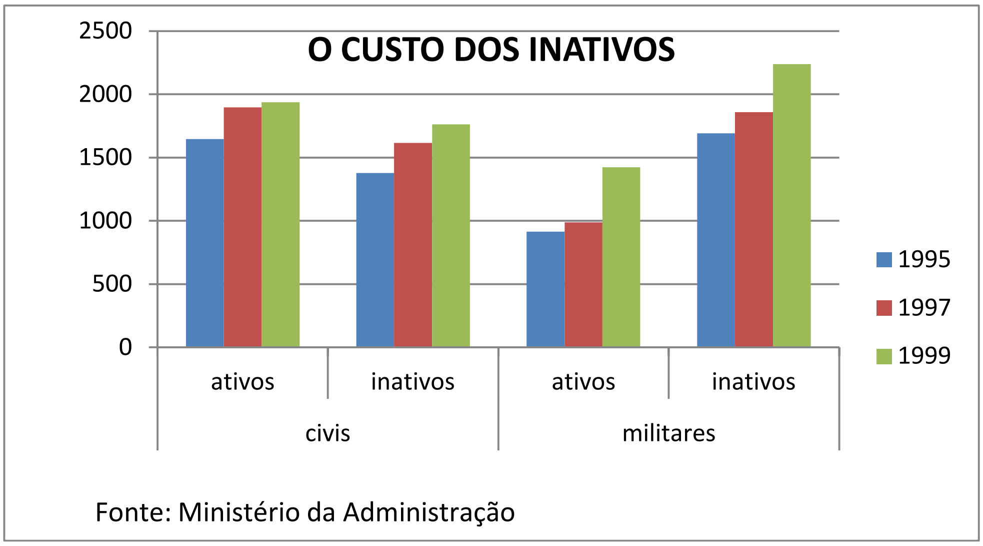
Fonte: Cooperativa "ABC"



Exercícios propostos:
Para as tabelas e quadros abaixo construa os gráficos que melhor representá-los:
1)

O CUSTO DOS INATIVOS				
	civis		militares	
Ano	ativos	inativos	ativos	inativos
1995	1647	1378	914	1692
1997	1897	1617	988	1860
1999	1938	1762	1423	2239

Fonte: Ministério da Administração
*média salarial dos servidores civis e militares em R$



2)

COMPUTADOR/ AUTOMÓVEL		
Ano	Computador	Automóvel
94	2,2	1,2
95	2,8	1,3
96	4	1,5
97	5,6	1,7
98	6,8	1,5

Fonte: Revista Exame/98
*em milhões

3)

QUEM COMPRA PCs?	
Indústria e Comércio	18%
Governo	15%
Setor Financeiro	27%
Pequenos Consumidores	40%

Fonte: Fenasoft/Simonsen Associados

4)

O CLUBE DOS SEM*	
os sem-telefone	108,7
os sem-esgoto	84,6
os sem-casa própria	43,7
os sem-terra	20
os sem-mulher	16,7
os sem-marido	14,1
os sem-trabalho	4,5

Fonte Veja - 23/04/97
*em bilhões

5)

NÚMERO DE CASOS DE AIDS NA REGIÃO ABC

MESES	N. DE CASOS
NOV	13
DEZ	17
JAN	23
FEV	32
MAR	20
ABR	15
MAI	18

Fonte: Secretaria de Saúde

DISTRIBUIÇAO DE FREQUÊNCIA

Tabela primitiva e rol de dados

É o tipo de tabela cujos elementos não foram numericamente organizados.
Exemplo:

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

Tabela 1: Estatura de 40 alunos do colégio A
Fonte: Crespo, 2002

A tabela primitiva após a ordenação recebe o nome de rol
Exemplo:

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

Tabela 2: Estatura de 40 alunos do colégio A
Fonte: Crespo, 2002

Distribuição de frequência

Denomina-se frequência o número de alunos que fica relacionado a um determinado valor da variável. Exemplo:

Estatura (cm)	Frequência
150	1
151	1
152	1
153	1
154	1
155	4
156	3
157	1
158	2
160	5
161	4
162	2
163	2
164	3
165	1
166	1
167	1
168	2
169	1
170	1
172	1
173	1
Total	40

Tabela 3: Estatura de 40 alunos do colégio A/ Fonte: Crespo, 2002

Pode-se ainda simplificar o processo, agrupando os valores da variável em intervalos, que chamamos de classes. Assim, a tabela 4 representa a estatura dos alunos através de uma distribuição com intervalos de classes:

Estatura (cm)	Frequência
150 |-- 154	4
154 |-- 158	9
158 |-- 162	11
162 |-- 166	8
166 |-- 170	5
170 |-- 174	3
Total	40

Tabela 4: Estatura de 40 alunos do colégio A agrupados
Fonte: Crespo, 2002

Ao agruparmos os dados ganhamos em simplicidade, mas perdemos em pormenores.

Elementos de uma distribuição de frequência

<u>Classe</u>

São intervalos de variação da variável. São representadas por i, sendo i = 1, 2, 3, ..., k.
Assim, na tabela 4, o intervalo 154 |-- 158, define a segunda classe, i = 2, sendo k = 6, já que a tabela possui 6 classes.

<u>Limites de classe</u>

São os extremos de cada classe. O menor número é o limite inferior (l_i) e o maior número é o limite superior (L_i)
Na segunda classe por exemplo, tem-se:
$l_i = 154$
$L_i = 158$

<u>Amplitude de um intervalo de classe</u>

É a medida do intervalo que define a classe. É obtida pela diferença entre os limites superior e inferior dessa classe e é indicado por h_i.
Assim:
$h_i = L_i - l_i$

Assim, a amplitude da segunda classe é:
$h_2 = L_2 – l_2$
$h_2 = 158 – 154 = 4$ cm

<u>Amplitude total da distribuição</u>

É a diferença entre o limite superior da última classe e o limite inferior da primeira classe.
AT= L (máx) – l (min)

No exemplo da tabela 4: AT = 174 – 150 = 24 cm

Amplitude amostral

É a diferença entre o valor máximo e valor mínimo da amostra.
AA = x(máx) – x (min)

No exemplo da tabela 2: AA = 173 – 150 = 23 cm
Observe que a amplitude total da distribuição não coincide com a da amplitude amostral.

Ponto médio de uma classe

É o ponto que divide o intervalo de classe em duas partes iguais.

$$x_i = \frac{l_i + L_i}{2}$$

Assim, o ponto médio da segunda classe, é:

$$x_2 = \frac{l_2 + L_2}{2}$$

$$x_2 = \frac{154 + 158}{2} = 156$$

Frequência simples ou absoluta

É o número de observações correspondentes a classe ou ao valor.
A frequência simples é simbolizada por f_i (lemos: f índice i ou frequência da classe i)
Assim, no exemplo anterior:
$f_1 = 4$, $f_2 = 9$, $f_3 = 11$, $f_4 = 8$, $f_5 = 5$ e $f_6 = 3$
A soma de todas as frequências é representada por:

$$\sum_{i=1}^{k} f_i$$

É evidente que:

$$\sum_{i=1}^{k} f_i = n$$

Para a distribuição em estudo, tem-se:

$$\sum_{i=1}^{6} f_i = 40$$

Representa-se da seguinte forma:

i	Estatura (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
	Total	$\sum f_i = 40$

Exercícios:

1) As notas obtidas por 50 alunos de uma classe foram:

1	2	3	4	5	6	6	7	7	8
2	3	3	4	5	6	6	7	8	8
2	3	4	4	5	6	6	7	8	9
2	3	4	4	5	6	6	7	8	9
2	3	4	4	5	6	7	7	8	9

Complete a distribuição de frequência abaixo:

i	Notas	x_i	f_i
1	0 |-- 2	1	1
2	2 |-- 4	____	____
3	4 |-- 6	____	____
4	6 |-- 8	____	____
5	8 |-- 10	____	____
	Total		$\sum f_i = 50$

E, responda:

a) Qual a amplitude amostral?

b) Qual a amplitude da distribuição?

c) Qual o número de classes da distribuição

d) Qual o limite inferior da quarta classe?

e) Qual o limite superior da classe de ordem 2?

f) Qual a amplitude do segundo intervalo de classe?

Complete:

a) h_3 =

b) n =

c) l_1 =

d) L_3 =

e) x_2 =

f) f_5 =

Tipos de Frequências

<u>Frequência simples ou absoluta (f_i)</u>

São os valores que realmente representam o número de cada classe.

$$\sum_{i=1}^{k} f_i = n$$

<u>Frequência relativa (fr_i)</u>

São os valores das razões entre as frequências simples e a frequência total:

$$fr_i = \frac{f_i}{\sum f_i}$$

A frequência relativa da terceira classe em nosso exemplo será:

$$fr_3 = \frac{f_3}{\sum f_i}$$

$$fr_3 = \frac{11}{40} = 0{,}275$$

A soma das frequências relativas será sempre = 1 ou seja de 100%

<u>Frequência acumulada (F_i)</u>

É a total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe:

$F_k = f_1 + f_2 + ... + f_k$ ou $F_k = \sum f_i$ (i = 1, 2, ... , k)

Assim, a frequência acumulada da terceira classe será:

$F_3 = f_1 + f_2 + f_3$ $F_3 = 4 + 9 + 11 = 24$

O que significa que existem 24 alunos com estatura inferior a 162 cm, limite superior do intervalo da terceira classe.

<u>Frequência acumulada relativa (Fr_i)</u>

É a frequência acumulada da classe, dividida pela frequência total da distribuição:

$$Fr_i = \frac{F_i}{\sum f_i}$$

Assim, no exemplo para a terceira classe:

$$Fr_3 = \frac{F_3}{\sum f_i}$$

$$Fr_3 = \frac{24}{40} = 0{,}600$$

Vamos montar uma tabela com os dados do exemplo inicial da estatura dos alunos:

i	Estatura (cm)	x_i	f_i	F_i	fr_i	Fr_i
1	150 |-- 154	______	______	______	______	______
2	154 |-- 158	______	______	______	______	______
3	158 |-- 162	______	______	______	______	______
4	162 |-- 166	______	______	______	______	______
5	166 |-- 170	______	______	______	______	______
6	170 |-- 174	______	______	______	______	______
	Total		$\sum f_i = 40$		$\sum fr_i = 1$	$\sum f_i = 40$

Exercícios:

1) Complete a distribuição abaixo determinando as frequências simples:

i	x_i	f_i	F_i
1	2	______	2
2	3	______	9
3	4	______	21
4	5	______	29
5	6	______	34
		$\sum f_i = 34$	

2) Conhecidas as notas de 50 alunos:

84	68	33	52	47	73	68	61	73	77
74	71	81	91	65	55	57	35	85	88
59	80	41	50	53	65	76	85	73	60
67	41	78	56	94	35	45	55	64	74
65	94	66	48	39	69	89	98	42	54

Obtenha a distribuição de frequência, tendo 30 para limite inferior da primeira classe e 10 para intervalo de classe.

3) Os resultados do lançamento de um dado 50 vezes foram os seguintes:

6	5	2	6	4	3	6	2	6	5
1	6	3	3	5	1	3	6	3	4
5	4	3	1	3	5	4	4	2	6
2	2	5	2	5	1	3	6	5	1
5	6	2	4	6	1	5	2	4	3

Forme uma distribuição de frequência sem intervalos de classe.

4) Considerando as notas de um teste de inteligência aplicado a 100 alunos:

64	78	66	82	74	103	78	86	103	87
73	95	82	89	73	92	85	80	81	90
78	86	78	101	85	98	75	73	90	86
86	84	86	76	76	83	103	86	84	85
76	80	92	102	73	87	70	85	79	93
82	90	83	81	85	72	81	96	81	85
68	96	86	70	72	74	84	99	81	89
71	73	63	105	74	98	78	78	83	96
95	94	88	62	91	83	98	93	83	76
94	75	67	95	108	98	71	92	72	73

Forme uma distribuição de frequência.

5) A tabela abaixo apresenta apresenta as vendas diárias de um determinado aparelho elétrico, durante um mês, por uma firma comercial:

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

Forme uma distribuição de frequência sem intervalos de classes.

6) Complete a tabela abaixo:

i	Classes	x_i	f_i	F_i	fr_i	Fr_i
1	0 |-- 8	______	4	______	______	______
2	8 |-- 16	______	10	______	______	______
3	16 |-- 24	______	14	______	______	______
4	24 |-- 32	______	9	______	______	______
5	32 |-- 40	______	3	______	______	______
	Total		$\sum f_i = 40$		$\sum fr_i = 1$	

7) Dada a distribuição de frequência:

x_i	3	4	5	6	7	8
f_i	2	5	12	10	8	3

Determine:

a) $\sum f_i$
b) As frequências relativas
c) As frequências acumuladas
d) As frequências relativas acumuladas

8) A tabela abaixo apresenta uma distribuição de frequência das áreas de 400 lotes:

Áreas m^2	300 |-- 400	|-- 500	|-- 600	|-- 700	|-- 800	|-- 900	|-- 1000	|-- 1100	|-- 1200
Número de lotes	14	46	58	76	68	62	48	22	6

Com referência a essa 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 frequência da quarta classe;
g) a frequência relativa da sexta classe;
h) a frequência 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;
k) a percentagem de lotes cuja área não atinge 600 m^2;
l) a percentagem de lotes cuja área seja maior ou igual a 900 m^2;
m) a percentagem de lotes cuja área é de 500 m^2, no mínimo, mas inferior a 1000 m^2;
n) a classe do 72° lote;
o) até que classe estão incluídos 60% dos lotes

9) A distribuição abaixo indica o número de acidentes ocorridos com 70 motoristas de uma empresa de ônibus:

Determine:

a) o número de motoristas que não sofreram nenhum acidente;
b) o número de motoristas que sofreram pelo menos 4 acidentes;
c) o número de motoristas que sofreram menos de 3 acidentes;
d) o número de motoristas que sofreram no mínimo 3 e no máximo 5 acidentes;
e) o número de motoristas que não sofreram no máximo 2 acidentes;

10) Complete os dados que faltam na distribuição de frequência:

a)

i	x_i	f_i	fr_i	F_i	Fr_i
1	0	1	0,05	______	______
2	1	______	0,15	4	______
3	2	4	______	______	______
4	3	______	0,25	13	______
5	4	3	0,15	______	______
6	5	2	______	18	
7	6	______	______	19	
8	7	______	______	______	
		$\sum f_i = 20$	$\sum fr_i = 1$		

b)

i	Classes	x_i	f_i	F_i	fr_i	Fr_i
1	0 |-- 2	______	4	______	______	______
2	2 |-- 4	______	8	______	______	______
3	4 |-- 6	______	______	30	______	______
4	6 |-- 8	______	27	______	______	______
5	8 |-- 10	______	15	72	______	______
6	10 |-- 12	______	______	83	______	______
7	12 |-- 14	______	10	93	______	______
8	14 |-- 16	______	______	______	______	______
	Total		$\sum f_i = 40$		$\sum fr_i = 1$	

MEDIDAS DE TENDÊNCIA CENTRAL OU DE POSIÇÃO

MÉDIA

Quando a distribuição é simétrica, a média e a mediana coincidem.
A mediana não é tão sensível, como a média, às observações que são muito maiores ou muito menores do que as restantes. Por outro lado a média reflete o valor de todas as observações.
Assim, não se pode dizer em termos absolutos qual destas medidas de localização é preferível, dependendo do contexto em que estão a ser utilizadas.

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

A média é o quociente da divisão da soma dos valores das variáveis pelo número deles:
Ex. para dados não agrupados: Sabendo-se que a produção diária da vaca A, durante uma semana, foi de 10, 14, 13, 15, 16, 18 e 12 litros, calcule a produção média na semana.

<u>Desvio em relação a média</u> – é a diferença entre cada elemento de um conjunto de valores e a média aritmética.
Ex. Determinar o desvio em relação a média no exemplo anterior.

Cálculo da média para dados agrupados (média ponderada)

<u>Sem intervalo de classe</u>

$$\bar{x} = \frac{\sum xifi}{\sum fi}$$

Ex.:
Considere a distribuição relativa a 34 famílias de quatro filhos, tomando para variável o número de filhos do sexo masculino:

Nº de meninos	fi	xifi
0	2	
1	6	
2	10	
3	12	
4	4	
	∑=34	∑=

Determine a média aritmética ponderada:

Exercício:

Determine a média aritmética da distribuição:

xi	fi	xifi
1	2	
2	4	
3	6	
4	8	
5	3	
6	1	
	∑=	∑=

<u>Com intervalo de classe</u>

Ex.:

<table>
<tr><th>i</th><th>Estaturas (cm)</th><th>fi</th><th>xi</th><th>xifi</th></tr>
<tr><td>1</td><td>150|--154</td><td>4</td><td></td><td></td></tr>
<tr><td>2</td><td>154|--158</td><td>9</td><td></td><td></td></tr>
<tr><td>3</td><td>158|--162</td><td>11</td><td></td><td></td></tr>
<tr><td>4</td><td>162|--166</td><td>8</td><td></td><td></td></tr>
<tr><td>5</td><td>166|--170</td><td>5</td><td></td><td></td></tr>
<tr><td>6</td><td>170|--174</td><td>3</td><td></td><td></td></tr>
<tr><td></td><td></td><td>∑=40</td><td></td><td>∑=6440</td></tr>
</table>

Exercício: Complete o esquema para o cálculo da média aritmética da distribuição de frequência:

<table>
<tr><td colspan="2">Custos</td><td colspan="2">450|--550|--650|--750|--850|--950|--1050|--1150</td></tr>
<tr><td colspan="2">fi</td><td colspan="2">8 10 11 16 13 5 1</td></tr>
<tr><td>i</td><td>xi</td><td>fi</td><td>xifi</td></tr>
<tr><td>1</td><td></td><td>8</td><td>4000</td></tr>
<tr><td>2</td><td></td><td>10</td><td></td></tr>
<tr><td>3</td><td></td><td>11</td><td></td></tr>
<tr><td>4</td><td></td><td>16</td><td></td></tr>
<tr><td>5</td><td></td><td></td><td></td></tr>
<tr><td>6</td><td></td><td></td><td></td></tr>
<tr><td>7</td><td>1100</td><td></td><td></td></tr>
<tr><td></td><td></td><td>∑=</td><td>∑=</td></tr>
</table>

R.:755

MODA (Mo)

Para um conjunto de dados, define-se moda como sendo:

O valor que surge com mais frequência *se os dados são discretos, ou,* ***o intervalo de classe com maior frequência*** *se os dados são contínuos.*
Assim, da representação gráfica dos dados, obtém-se imediatamente o valor que representa a moda ou a classe modal.

Cálculo da moda para dados agrupados

Sem intervalo de classe: Esta medida é especialmente útil para reduzir a informação de um conjunto de dados qualitativos, apresentados sob a forma de nomes ou categorias, para os quais não se pode calcular a média e por vezes a mediana (se não forem susceptíveis de ordenação).
Ex.: 7, 8, 9, 10, 10, 10, 11, 12, 13, 15   Mo=10

Série **amodal:** nenhum valor aparece mais vezes que outro. Ex.: 3, 5, 8, 10, 12, 13

Série **bimodal:** mais de um valor aparece mais vezes que outros. Ex.: 3, 4, 4, 4, 5, 7, 7, 7, 8, 10, 12, 13
Mo= 4 e 7

Cálculo da moda para dados agrupados

Com intervalo de classe

$$\mathbf{Mo = \frac{l + L}{2}}$$

l = limite inferior da classe modal
L = limite superior da classe modal

Ex.: Determine a moda

i	Estatura (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
	Total	$\sum f_i = 40$

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

Exercício: Determine a moda:

i	Estaturas (cm)	fi
1	450 |--550	8
2	550 |--650	10
3	650 |--750	11
4	750 |--850	16
5	850 |--950	13
6	950 |--1050	5
7	1050|--1150	1
		∑=64

MEDIANA

A mediana, **m**, é uma medida de localização do centro da distribuição dos dados, definida do seguinte modo: *Ordenados os elementos da amostra, a mediana é o valor (pertencente ou não à amostra) que a divide ao meio, isto é, 50% dos elementos da amostra são menores ou iguais à mediana e os outros 50% são maiores ou iguais à mediana.*

Para a sua determinação utiliza-se a seguinte regra, depois de ordenada a amostra de n elementos:

Se n é ímpar, a mediana é o elemento médio.

Se n é par, a mediana é a semi-soma dos dois elementos médios.

Ex.: Dada a série: 5, 13, 10, 2, 18, 15, 6, 16, 9, determine a mediana.

Dados não-agrupados

Sem intervalo de classe

Nº de meninos	fi	Fi
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	∑=34	

34/2=17, MD=2 meninos

Ex. 2

xi	fi	Fi
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8

Md = 15,5

Exercícios:

Qual é a mediana nas distribuições?

a)

xi	2	4	6	8	10
fi	3	7	12	8	4

Md =

b)

xi	0	1	2	3	4	5
fi	2	5	9	7	6	3

Md=

<u>Dados agrupados</u>

Deve-se seguir os seguintes passos:

1) Determinar as frequências acumuladas
2) Calcular $\frac{\sum f_i}{2}$
3) Marcar a classe correspondente a frequência acumulada imediatamente superior à $\frac{\sum f_i}{2}$– classe mediana e em seguida empregar a fórmula:

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

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

Exemplo: Determinar a mediana da distribuição de frequência:

i	Estatura (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
	Total	$\sum f_i = 40$

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

Exercício:

1) Determinar a mediana das distribuições de frequência:

i	Custo	fi
1	450 |--550	8
2	550 |--650	10
3	650 |--750	11
4	750 |--850	16
5	850 |--950	13
6	950 |--1050	5
7	1050|--1150	1
		$\sum$=64

i	classes	f_i
1	0 |-- 10	1
2	10 |-- 20	3
3	20 |-- 30	9
4	30 |-- 40	7
5	40 |-- 50	4
6	50 |-- 60	2
	Total	$\sum f_i = 26$

Calcule a média e a mediana das distribuições:

i	toneladas	f_i
1	300 |-- 400	17
2	400 |-- 500	22
3	500 |-- 600	38
4	600 |-- 700	20
5	700 |-- 800	15
	Total	$\sum f_i = 112$

i	cm	f_i
1	140 |-- 150	17
2	150 |-- 155	28
3	155 |-- 160	43
4	160 |-- 165	52
5	165 |-- 170	40
6	170 |-- 180	30
	Total	$\sum f_i = 210$

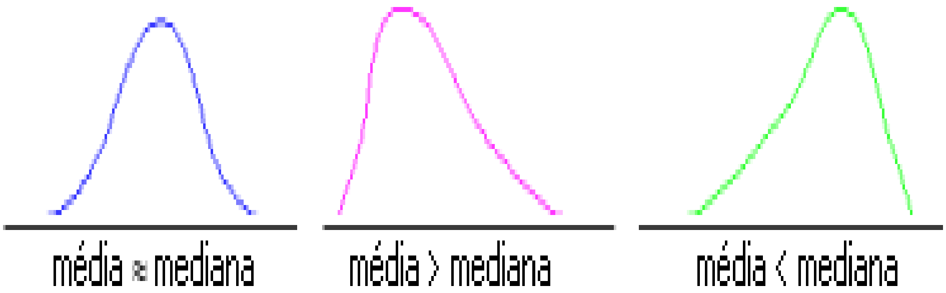
Resumindo, como a média é influenciada quer por valores muito grandes, quer por valores muito pequenos, se a **distribuição dos dados**:

1. for aproximadamente simétrica, a média aproxima-se da mediana

2. for enviesada para a direita (alguns valores grandes como "outliers"), a média tende a ser maior que a mediana

3. for enviesada para a esquerda (alguns valores pequenos como "outliers"), a média tende a ser inferior à mediana.

Representando as ditribuições dos dados (esta observação é válida para as representações gráficas na forma de diagramas de barras ou de histograma) na forma de uma mancha, temos, de um modo geral:



SEPARATRIZES

São medidas teóricas que permitem a separação da distribuição em grupos.
Posição das separatrizes:

Quartis

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

$Q_i = \frac{k\sum f_i}{4}$ com i = 1, 2 ou 3

Decis

Denomina-se decis os valores de uma série que a dividem em 10 partes iguais.

$D_i = \frac{k\sum f_i}{10}$ com i = 1, 2, ..., 9

Percentis

Denomina-se decis os valores de uma série que a dividem em 100 partes iguais.

$C_i = \frac{k\sum f_i}{100}$ com i = 1, 2 ou 3

Quando os dados são agrupados, usa-se a mesma técnica do cálculo da mediana. Por exemplo, para calcularmos o terceiro quartil:

$$Q_3 = l^* + \frac{\left[\frac{3\sum f_i}{2} - F(ant)\right] h^*}{f^*}$$

Exemplo:

1) Calcule o 1° e 3° quartil, 3° e 6° decil e o 32° percentil:11, 13, 13, 15, 17, 17, 17, 18, 18, 21, 23, 23, 25, 26, 28, 30, 31, 33, 36, 37, 37, 38, 38, 39, 39, 40, 41, 42, 42, 42, 45, 46, 48, 49, 51, 52.
2) Calcular os quartis inferior e superior e os 15° e 25° centis:

Vida útil (horas)	f_i
0 |-- 100	13
100 |-- 200	45
200 |-- 300	54
300 |-- 400	125
400 |-- 500	96
500 |-- 600	27
Total	$\sum f_i =$

3) A avaliação de desempenho de um treinamento aplicado aos funcionários da empresa Um Dois Três de Oliveira Quatro obedecia o seguinte critério: Notas de 65 a 100 Aprovados; de 50 até 65 Recuperação Parcial; de 34 a 50, Recuperação Plena e abaixo de 34 Reprovados. Os resultados obtidos foram expressos na distribuição abaixo:

Notas	f_i
10 |-- 25	14
25 |-- 40	20
40 |-- 55	26
55 |-- 70	34
70 |-- 85	60
85 |-- 100	26
Total	$\sum f_i = 180$

Pede-se:

a) A quantidade de funcionários reprovados;
b) A quantidade de funcionários em recuperação plena;
c) A quantidade de funcionários recuperação parcial;
d) A quantidade de funcionários aprovados.

Representação Gráfica De Uma Distribuição
Histograma E Sua Interpretação

Uma leitura atenta do histograma deve responder a questões como:

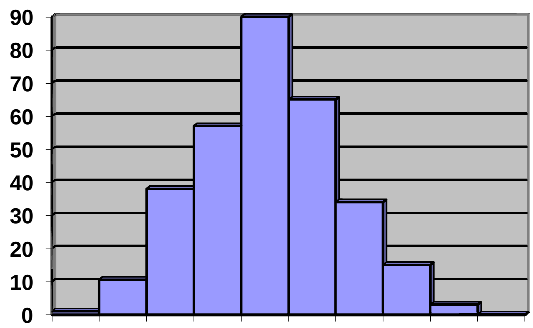
1. Qual é a forma da distribuição?
2. Existe um ponto central bem definido?
3. Quão grande é a variação?
4. Qual é a amplitude dos dados?
5. Existe apenas um pico?
6. A distribuição é simétrica?
7. Existem barras isoladas?
8. Quais conclusões que você pode tirar sobre o desempenho do processo em relação à característica estudada?
9. O histograma é conclusivo ou seu aspecto sugere a necessidade de estratificação para buscar as causas das anomalias encontradas?

Tipos De Histogramas:

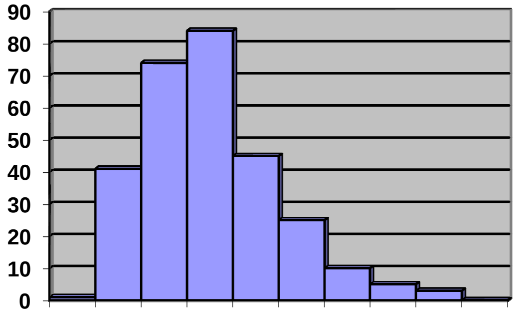
❖ Histograma simétrico, tipo distribuição Normal:

Característica: a frequência é mais alta no centro e decresce gradualmente para as caudas de maneira simétrica (forma de sino). A média e a mediana são aproximadamente iguais e localizam-se no centro do histograma (ponto de pico).

Quando ocorre: forma usualmente observada em processos padronizados, estáveis, em que a característica de qualidade é contínua e não apresenta nenhuma restrição teórica nos valores que podem ocorrer.



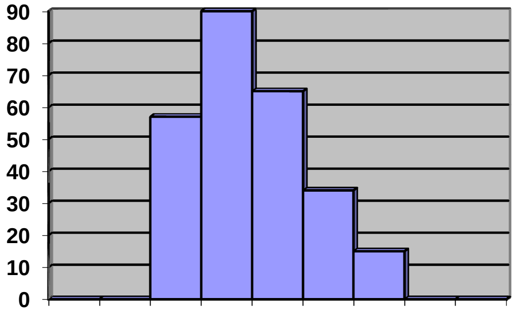
❖ Histograma assimétrico e com apenas um pico:

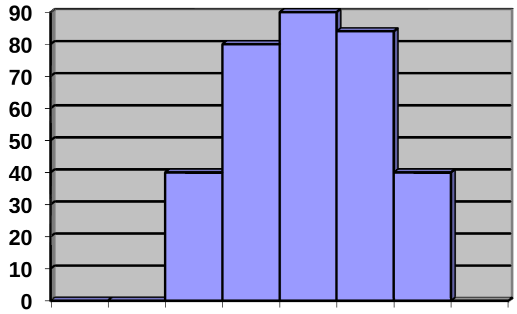


Características: a frequência decresce bruscamente em um dos lados de forma gradual no outro, produzindo uma calda mais longa em um dos lados. A média localiza-se fora do meio da faixa de variação. Quando a assimetria é à direita a mediana é inferior a média. Quando a assimetria é à esquerda a mediana é superior à média.

Quando ocorre: possivelmente a característica de qualidade possui apenas um limite de especificação e é controlada durante o processo, de modo que satisfaça a essa especificação.

❖ Histograma tipo "despenhadeiro":

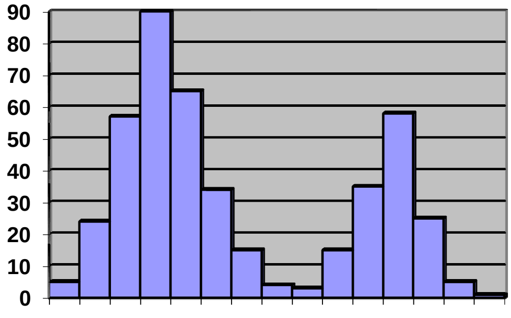




Característica: o histograma termina abruptamente de um ou dos dois lados, dando a impressão de faltar um pedaço na figura.

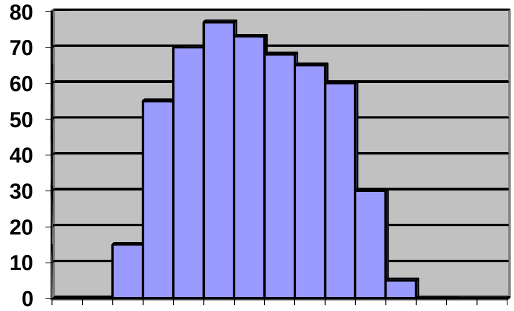
Quando ocorre: possivelmente foram eliminados dados por uma inspeção 100%; nesse caso o "corte" coincide com os limites de especificação.

❖ Histograma com dois picos:



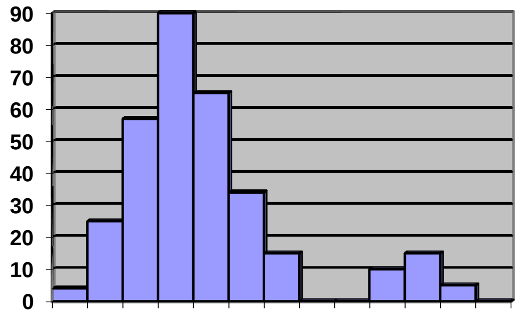
Característica: ocorrem dois picos e a frequência é baixa entre eles
Quando ocorre: em situações em que há mistura de dados com médias diferentes obtidos em duas condições distintas. Por exemplo, dois tipos de matérias primas, duas máquinas ou dois operadores. A estratificação dos dados segundo esses fatores poderá confirmar ou não tais conjecturas.

❖ Histograma do tipo "platô"



Característica: classes centrais possuem aproximadamente a mesma frequência.
Quando ocorre: aspecto possível quando há mistura de várias distribuições com médias diferentes

❖ Histograma com uma pequena "ilha" isolada:



Característica: algumas faixas de valores da característica de qualidade observada ficam isoladas da grande maioria dos dados, gerando barras ou pequenos agrupamentos separados.
Quando ocorre: possivelmente ocorreram anormalidades temporárias no processo, erros de medição, erros de registro ou transcrição dos dados, produzindo alguns resultados muito diferentes dos demais.

MEDIDAS DE DISPERSÃO

VARIÂNCIA: Define-se a variância, e representa-se por $\mathbf{s^2}$, como sendo a medida que se obtém somando os quadrados dos desvios das observações da amostra, relativamente à sua média, e dividindo pelo número de observações da amostra menos um:

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

DESVIO PADRÃO: Uma vez que a variância envolve a soma de quadrados, a unidade em que se exprime não é a mesma que a dos dados. Assim, para obter uma medida da variabilidade ou dispersão com as mesmas unidades que os dados, tomamos a raiz quadrada da variância e **obtemos o desvio padrão**:

$$s = \sqrt{\sum_{i=1}^{n} \frac{(xi - \bar{x})^2}{n-1}}$$

O desvio padrão é uma medida que **só pode assumir valores não negativos** e quanto maior for, maior será a dispersão dos dados.
Algumas propriedades do desvio padrão, que resultam imediatamente da definição, são:

- o desvio padrão é sempre não negativo e será tanto maior, quanta mais variabilidade houver entre os dados.
- se s = 0, então não existe variabilidade, isto é, os dados são todos iguais.

A informação que o desvio padrão dá sobre a **variabilidade** deve ser entendida como a variabilidade que é apresentada relativamente a um ponto de referência - a média, e não propriamente a variabilidade dos dados, uns relativamente aos outros.

Ex.: Calcular o desvio padrão par ao conjunto de valores: 40, 45, 48, 52, 54, 62, 70

xi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2
40	53	-13	169
45	53	-8	64
48	53	-5	25
52	53	-1	1
54	53	1	1
62	53	9	81
70	53	17	289
∑=371			630

S=

Exercício:

Complete o esquema para o cálculo do desvio padrão, dados os valores da variável: 8, 10, 11, 15, 16, 18

xi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2
8			
10			
11			
15			
16			
18			
∑=			

S=

Dados agrupados

Sem intervalo de classes ou com intervalo de classe

$$s = \sqrt{\sum_{i=1}^{n} \frac{(xi - \bar{x})^2 \, fi}{n-1}} \text{ amostral}$$

xi	fi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2	(xi- $\bar{x}$)^{2}fi
0	2				
1	6				
2	12				
3	7				
4	3				
∑=					

S=

Exercício:

Calcule o desvio padrão da distribuição:

xi	1 2 3 4 5 6	
fi	2 5 8 6 3 1	

xi	fi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2	(xi- $\bar{x}$)^{2}fi
1					
2					
3					
4					
5					
6					

S=

Ex.:

i	Estaturas (cm)	fi	xi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2	(xi- $\bar{x}$)^{2}fi
1	150|--154	4	152				
2	154|--158	9	156				
3	158|--162	11	160				
4	162|--166	8	164				
5	166|--170	5	168				
6	170|--174	3	172				

S=

Exercício: Calcule o desvio padrão da distribuição:

i	intervalo	fi	xi	$\bar{x}$	xi - $\bar{x}$	(xi- $\bar{x}$)^2	(xi- $\bar{x}$)^{2}fi
1	30|--50	2					
2	50|--70	8					
3	70|--90	12					
4	90|--110	10					
5	110|--130	5					
		∑=		∑=	∑=		

Exercícios complementares:

1) Os tempos despendidos por 12 alunos (N = 12), em segundos, para percorrer certo trajeto, sem barreira, foram 16, 17, 16, 20, 18, 16, 17, 19, 21, 22, 16, 23. Determine o valor, sem agrupar os dados: da moda, mediana e média; da variância absoluta, do desvio padrão.

2) Considerando uma população, de tamanho 16 (N = 16), constituída de alunos, cuja variável de interesse X é o número de faltas de cada aluno, obteve-se: 0, 0, 1, 2, 2, 2, 4, 4, 4, 3, 0, 5, 4, 4, 3 e 2. Sem agrupar os dados, determine o valor: da moda, mediana e média; da variância absoluta, do desvio padrão.

3) Considere uma população de 40 profissionais liberais que foram, questionados sobre o número de revistas e/ou jornais que os mesmos são assinantes, obteve-se a seguinte tabela:

Nº de Publicações	Nº de Profissionais
0	6
1	8
2	12
3	10
4	4
$\sum$	40

Pede-se:
a) A percentagem de profissionais que tem menos de 3 revistas e/ou jornais (publicações).
b) valor da moda, da mediana e da média aritmética simples.
c) valor da variância absoluta, do desvio padrão.

4) Em certo dia foi realizado um levantamento a respeito das idades dos alunos de um curso noturno, obtendo-se a tabela abaixo:

Idades (anos)	Nº de Alunos
16 |- 20	8
20 |- 24	16
24 |- 28	12
28 |- 32	4
$\sum$	40

Considerando esta turma como uma população, determine:
a) A percentagem de alunos com menos de 24 anos.
b) O valor da média aritmética simples e a moda.
c) O valor da variância absoluta, do desvio padrão.

5) Em um levantamento realizado, em maio de 1983 nos 200 funcionários da empresa XK, em relação a variável expressa em unidades monetárias (u.m.), obteve-se a seguinte tabela:

Salário (u.m.)	Nº de Funcionários
0 |- 2	26
2 |- 4	32
4 |- 6	34
6 |- 8	40
8 |- 10	28
10 |- 12	22
12 |- 14	18
$\sum$	200

Considerando os 200 funcionários como de uma população, determine:

a) A percentagem de funcionários que recebem salário maior ou igual a 2 u.m. e menor que 4 u.m.

b) A porcentagem de funcionários que recebem menos de 8 u.m.

c) O valor da moda e da média dos salários.

d) O valor da variância absoluta, do desvio padrão.

6) Considerando que foi extraída uma amostra aleatória simples de 10 alunos de uma grande escola, cuja variável em estudo é a nota obtida em Matemática, obteve-se: 5, 7, 8, 6, 5, 4, 8, 9, 10 e 6. Determine a média da amostra, a variância da amostra e o desvio padrão da amostra.

7) Considerando que as três distribuições hipotéticas apresentam os valores indicados abaixo:

Valores obtidos em três distribuições hipotéticas

DISTRIBUIÇÃO		
A	**B**	**C**
N = 200	N = 50	N = 400
$\sum f.X = 4000$	$\sum f.X = 500$	$\sum f.X = 3200$
$\sum f.X^2 = 85000$	$\sum f.X^2 = 5450$	$\sum f.X^2 = 32000$

Determine os indicadores: média aritmética, variância absoluta, desvio padrão.

8) Uma empresa de informática possui 10 vendedores e cada um deles trabalha com diferentes cargas horárias. As cargas horárias dos vendedores são dadas abaixo:

5 4 8 8 7 6 6 8 8 12

Calcule a média, a mediana, a moda e desvio padrão das cargas horárias desses vendedores.

9) Uma pesquisa sobre a idade (em anos), de uma classe de calouros do curso de Computação de certa faculdade, revelou os seguintes valores:

17	17	17	18	18	18	18	18	18	18
18	18	18	18	18	18	18	18	18	18
18	19	19	19	19	19	19	19	19	19
19	19	19	19	19	19	19	19	20	20
20	20	20	20	20	20	21	21	21	21

Construa uma distribuição de frequência e em seguida determine a média, a mediana, a moda e desvio padrão das idades.

10) Um produto é condicionado em lotes contendo cada um deles 10 unidades. Considere os produtos que compõe um determinado lote com seus respectivos pesos (em kg):

3	4	3,5	5	3,5	4	5	5,5	4	5

Determine:

a) O peso médio dos produtos;
b) A mediana correspondente ao peso dos produtos;
c) A Moda correspondente ao peso dos produtos;
d) A variação dos pesos dos produtos.

11) Considere as seguintes distribuições A, B e C, que representam a satisfação do cliente em relação ao atendimento ao usuário:

Distribuição A		**Distribuição B**		**Distribuição C**	
Satisfação do Cliente	f_i	Satisfação do Cliente	f_i	Satisfação do Cliente	f_i
$0 \vdash 2$	2	$0 \vdash 2$	5	$0 \vdash 2$	7
$2 \vdash 4$	4	$2 \vdash 4$	8	$2 \vdash 4$	12
$4 \vdash 6$	9	$4 \vdash 6$	11	$4 \vdash 6$	9
$6 \vdash 8$	15	$6 \vdash 8$	8	$6 \vdash 8$	5
$8 \vdash 10$	7	$8 \vdash 10$	5	$8 \vdash 10$	4
	$\Sigma f_i = 37$		$\Sigma f_i = 37$		$\Sigma f_i = 37$

Calcular a média, mediana, moda e desvio padrão das distribuições A, B e C;

Coeficientes De Variação (Cv)

Quando temos dois ou mais conjuntos com média e desvio padrão diferentes entre eles, como podemos determinar o mais homogêneo?

Exemplo:

Conjunto C	Conjunto D
$\bar{x}$ = 100	$\bar{x}$ = 450
s = 10	s = 35

Podemos usar a idéia de: Quanto 10 representa em 100? E quanto 35 representa em 450?
Este cálculo, expresso em porcentagem é o coeficiente de variação, ou seja mede percentualmente a relação entre desvio padrão e a média aritmética, sendo pois uma medida adimensional.

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

MEDIDAS DE ASSIMETRIA E CURTOSE

Momentos

Se $x_1, x_2, \ldots, x_n$ são os n valores assumidos pela variável X, define-se a quantidade

$$\bar{X}^r = \frac{x_1^r + x_2^r + \cdots + x_n^r}{n} = \frac{\sum_{i=1}^n x_i^r}{n},$$

como o momento de ordem r em relação a origem. Nota-se que o primeiro momento em relação a origem $(\bar{X}^1)$ é a média de X.

O momento de ordem r em relação a uma origem k, qualquer, é dado por:

$$M_r'(k) = \frac{\sum_{i=1}^n (x_i - k)^r}{n},$$

O momento de ordem r em relação a média $\bar{X}$ é dado por:

$$M_r'(\bar{X}) = \frac{\sum_{i=1}^n (x_i - \bar{X})^r}{n},$$

Nota-se que o segundo momento em relação a média é a variância.

Para o caso dos dados encotrarem-se agrupados, na forma de uma distribuição de freqüências, as expressões para o cálculo dos momentos serão:

$$M_r'(k) = \frac{\sum_{i=1}^n (x_i - k)^r F_i}{\sum_{i=1}^n F_i},$$

em que:

x_i é o ponto médio da *i-ésima* classe, e

F_i = freqüência absoluta da *i-ésima*.

Formato de uma distribuição

O terceiro e quarto momentos[1] de uma distribuição são freqüentemente usados para estudar a “aparência” de uma distribuição, em especial sua assimetria e sua curtose. Em outras palavras, a distribuição dos dados pode ser simétrica ou não, ou ainda achatada ou pontiaguda e, isso, dará um formato à curva de distribuição.

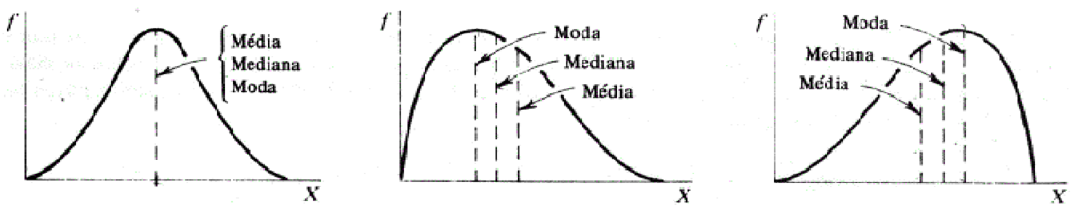
Medida De Assimetria

Denomina-se assimetria o grau de afastamento da simetria de uma distribuição de dados. Em uma distribuição (a) simétrica, tem-se igualdade dos valores da média, mediana e moda. Entretanto, se numa distribuição ocorrer:

b) $\overline{X} \geq Md \geq Mo$: existirão mais dados da série menores do que a média, porém a curva da distribuição terá uma cauda mais longa para os dados maiores do que a média, isto é, diz-se que a distribuição tem assimetria positiva.

c) $\overline{X} \leq Md \leq Mo$: existirão mais dados da série maiores do que a média, porém a curva de distribuição terá uma cauda mais longa para os dados menores do que a média, isto é, diz-se que a distribuição tem assimetria negativa;

A



(a) Simétrica (b) Positivamente assimétrica (c) Negativamente assimétrica

Interpretação

a) Cs = 0: se o resultado for zero, a distribuição é **<u>simétrica</u>**,

b) Cs < 0: se o valor for negativo, a distribuição é **<u>assimétrica negativa</u>** (inclinada para a esquerda) e,

c) Cs > 0: se o resultado for positivo, a distribuição é **<u>assimétrica positiva</u>** (inclinada para a direita).

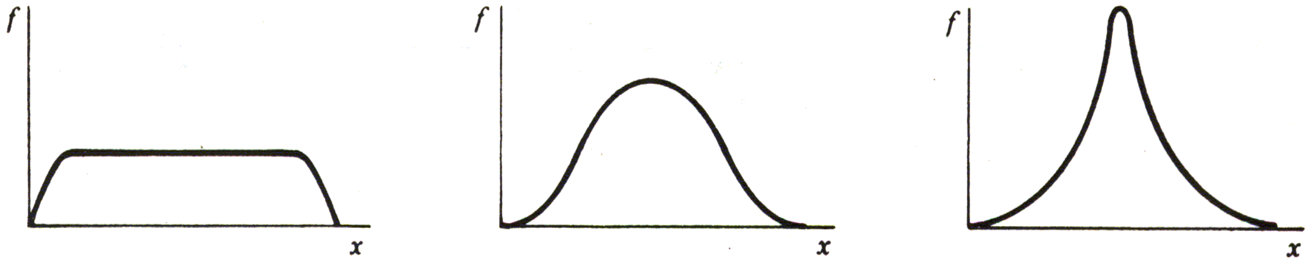
A estimativa do coeficiente de assimetria pode ainda ser encontrado através do primeiro e segundo coeficiente de assimetria de Pearson, dados respectivamente por:

$$Cs = \frac{\bar{x} - Mo}{s} \quad \text{e} \quad Cs = \frac{3(\bar{x} - Md)}{s}$$

[1] O primeiro momento é a média e o segundo momento é a variância.

Medida De Curtose

Denomina-se curtose o grau de achatamento da distribuição.



Platicúrtica Mesocúrtica Leptocúrtica

O resultado pode ser assim definido: $Ck = \frac{M'_4}{(s^2)^2}$

<u>Ck = 3 - Mesocúrtica</u> – a distribuição de frequências é a própria distribuição normal;
<u>Ck > 3 - Platicúrtica</u> – a distribuição é achatada (alta variabilidade);
<u>Ck < 3 - Leptocúrtica</u> – a distribuição é concentrada em torno da média (alta homogeneidade).

Obs: A assimetria positiva surge quando a média aritmética é aumentada por algum valor extraordinariamente elevado e, a assimetria negativa ocorre quando a média é reduzida por algum valor extremamente baixo. Os dados são simétricos quando não existem valores realmente extremos em uma direção específica, de modo que os valores baixos e altos se equilibram entre si.
Ou também pela fórmula:

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

Se

$C = 0,263$ - curva mesocúrtica
$C < 0,263$ - curva leptocúrtica
$C > 0,263$ - curva platicúrtica

Exercícios:

1) Determinar :
   a) O primeiro, o segundo e o terceiro momento para o conjunto de números 2, 3, 7, 8, 10.

   b) O segundo momento em relação à média.

2) Os dados a seguir referem-se ao número de partos/dia ocorridos num determinado hospital durante o mês de março de 2007:

X (partos/dia)	*F (número de dias)*
12	1
14	4
16	6
18	10
20	7
22	2
Total	30

a) Obter o primeiro momento à origem
b) Segundo, terceiro e quarto momentos em relação à média
c) Obter os coeficientes de assimetria e curtose utilizando o terceiro e quarto momento encontrado.

3) Para as distribuições de frequência abaixo determine: a media, a moda, a mediana, o desvio padrão os coeficientes de assimetria de Person e os coeficientes de curtose.

DISTRIBUIÇÃO A

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

DISTRIBUIÇÃO B

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

DISTRIBUIÇÃO C

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