UNIVERSIDADE DO ESTADO DO PARÁ
DEPARTAMENTO DE MATEMÁTICA, ESTATÍSTICA E INFORMÁTICA
LICENCIATURA EM MATEMÁTICA

CCSE

CENTRO DE CIÊNCIAS SOCIAIS E EDUCAÇÃO

ESTATÍSTICA

Fabrício Martins da Costa

Marília Brasil Xavier
REITORA

Prof. Rubens Vilhena Fonseca
COORDENADOR GERAL DOS CURSOS DE MATEMÁTICA

MATERIAL DIDÁTICO

EDITORAÇÃO ELETRONICA

Odivaldo Teixeira Lopes

ARTE FINAL DA CAPA

Odivaldo Teixeira Lopes

REALIZAÇÃO

UEPA
UNIVERSIDADE DO ESTADO DO PARÁ

Dados Internacionais de Catalogação na Publicação (CIP)

C834e Costa, Fabrício Martins da
Estatística / Fabrício Martins da Costa – Belém:
UEPA / Centro de Ciências Sociais e Educação, 2011.
76 p.; il.

ISBN: 978-85-88375-63-5

1.Estatística. I. Universidade Estadual do Pará. II. Título.

CDU: 519.22
CDD: 519.5

Índice para catálogo sistemático
1. Estatística: 519.22

Belém - Pará - Brasil

- 2011 -

APRESENTAÇÃO.

Prezado (a) aluno (a), este livro didático foi elaborado com muito cuidado visando uma aprendizagem autônoma. O que é Estatística? Qual a utilidade? Essas e outras perguntas tentaremos responder no decorrer do livro.

Hoje em dia percebemos que a Estatística está praticamente ao nosso redor de uma forma tão forte, por exemplo, fala-se no noticiário de televisão que houve um aumento nas *estatísticas de acidentes*, ou que *a probabilidade de rebaixamento de um clube para série B* do campeonato brasileiro de futebol é 70%. O ENEM (Exame nacional do ensino médio) adotará neste ano um novo formato, usar-se á Estatística mais precisamente a Teoria de Resposta ao Item (TRI).

Quando os meios de comunicação noticiam um assunto com relação à estatística, será que você entende por completo ou fica algo sem compreensão? Por exemplo, a nota média do Enem aumentou com relação ao ano passado. Primeiro, as provas dos dois anos estão na mesma escala?
Se as provas não estão na mesma escala não pode haver comparações! Logo, temos que ficar atentos a tudo que se fala nos meios de comunicação, pois, vivemos a era da globalização e o advento da internet. Precisamos estar antenados para não cometer nenhum tipo de erro grosseiro.

Este material didático não tem como objetivo formar estatísticos, mas formar profissionais que possam utilizar corretamente as técnicas estatísticas em uma pesquisa mais especificamente análise dos dados.

Sabemos que atualmente os problemas científicos necessitam de vários profissionais para serem resolvidos. Por exemplo, para realizar um Censo, podemos ter a presença de pelo menos os seguintes profissionais: administradores, estatísticos, matemáticos, sociólogos, etc. Isto ocorre devido a nossa sociedade que a cada dia torna-se mais complexa. E isto requer pessoas capacitadas para resolver os problemas da atualidade.

Logo, dominar os conceitos básicos de Estatística neste momento lhe proporcionará no futuro uma grande vantagem em relação aos demais profissionais, pois, a Estatística está presente em várias as áreas. Por exemplo, Áreas de Aplicação da Estatística:
Pesquisa (Artes, Arqueologia, Ciências Agrárias, Ciências Exatas, Ciências Sociais, Literatura, Meio Ambiente, Mercado, Petróleo), **Indústria e Negócio** (Controle de Qualidade, Previsão de Demanda, Gerenciamento Eficiente, Mercado e Finanças), **Medicina** (Diagnóstico, Prognóstico, Ensaios Clínicos), **Direito** (Evidência estatística, teste de DNA, investigação criminal), **Economia** (Técnicas Econométricas e Séries Temporais).

Estude cuidadosamente este material. Refaça os exemplos apresentados e busque apoio nas indicações fornecidas no tópico *pesquisando*. Depois de fazer isto, faça as questões propostas que estão no final de cada unidade que estão de acordo com os exemplos apresentados.

Tenham um bom estudo!

SUMÁRIO

UNIDADE 1 – CONHECENDO A ESTATÍSTICA

Nesta unidade, estaremos apresentando a ciência estatística. Mostraremos sua importância no estudo cientifico e como ela está presente no nosso cotidiano.

Inicialmente estudaremos os seguintes itens:

1.1 Introdução
1.2 Definições
1.3 Conceitos atuais
1.4 Conceitos importantes
1.5 Tipos de variáveis

1.1– INTRODUÇÃO

Na verdade a Estatística é uma ciência muito útil nos dias atuais**,** entretanto ela precisa ser entendida de forma correta. Por exemplo, muitas pessoas acham que Estatística composta apenas de gráficos e tabelas, e assumem este conceito errado de forma tão concreta que são incapazes de aceitar algo contrário. Isto se dá em virtude de vários fatores, por exemplo, o modo como uma noticia é dada na mídia, em muitas ocasiões o meio de comunicação desconhece a estatística e fazem afirmações sem fundamento a respeito do assunto.

A palavra *Estatística* vem de Status (**Estado**, em Latim). Então ela significa estudo do estado.

1.2 – DEFINIÇÕES

Abaixo temos algumas definições sobre Estatística.

- A Estatística está interessada nos métodos científicos para a coleta, organização, resumo, apresentação e análise de dados, bem como na obtenção de conclusões válidas e na tomada de decisões razoáveis, baseadas em tais análises. (SPIEGEL, 1994, p. 1).
- Entendemos a Estatística como um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar *dados* oriundos de estudos ou experimentos, realizados em qualquer área de conhecimento. (MAGALHÃES, 2002, p.1).

1.3 – CONCEITOS ATUAIS

Estatística: é a ciência que se preocupa com coleta, análise, interpretação e apresentação dos dados, permitindo-nos a obtenção de conclusões válidas a partir destes dados, bem como a tomada de decisões razoáveis baseadas nessas conclusões. A Estatística se divide didaticamente em duas partes:

Estatística Descritiva: é aquela que se preocupa com a coleta, análise, interpretação e apresentação dos dados estatísticos.

Estatística Indutiva: também conhecida como amostral ou inferencial, é aquela que partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões fundamentando-se na teoria das probabilidades.

1.4 – CONCEITOS IMPORTANTES

População: é todo conjunto, finito ou infinito, que possui ao menos uma característica em comum entre todos os seus elementos componentes.

População Finita: é aquela população em que é possível enumerar todos os seus elementos componentes.

Exemplos: Idade dos alunos da UEPA; as notas dos alunos da disciplina Estatística ou o número de consumidores de algum produto.

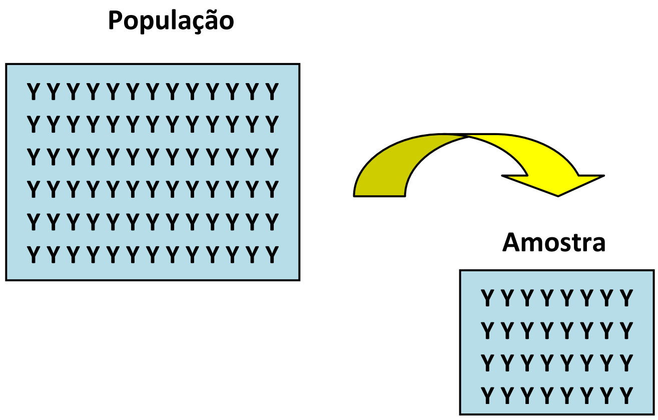
População Infinita: é aquela população em que não é possível enumerar todos os seus elementos componentes.

Exemplo: O número de astros no universo.

Censo: é o conjunto dos dados estatísticos dos habitantes de uma cidade, estado, etc, com todas as suas características, num determinado período de tempo. É a coleta exaustiva das informações de todas as "*N*" unidades da população.

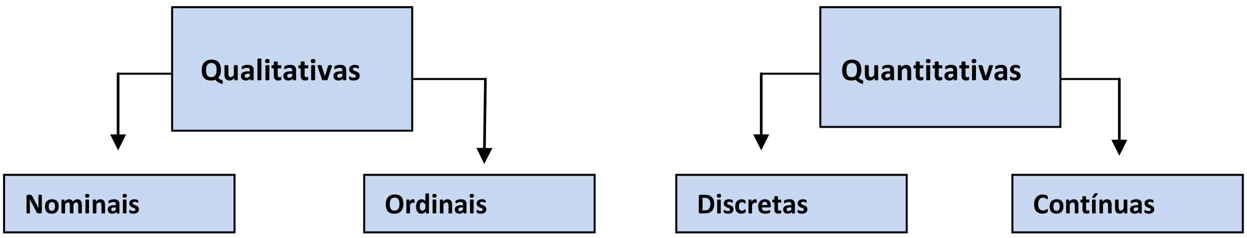
Amostra: é um subconjunto, uma parte selecionada da totalidade de observações abrangidas pela população da qual se quer inferir alguma coisa. (parte representativa da população)

Amostragem: é o processo de coleta das informações de parte da população - "*n*" - chamada amostra, mediante métodos adequados de seleção destas unidades.



1.5 – TIPOS DE VARIÁVEIS

Na Estatística nos deparamos com diversas variáveis, antes de realizar algum estudo estatístico é necessário conhecer as variáveis, pois para cada tipo de variável existe um teste estatístico especifico, por exemplo, Suponha que 150 crianças foram expostas a três tipos de comerciais de TV, sobre cereais para café da manhã. Após a exposição foi solicitado a cada criança para indicar qual dos comerciais ela gostou mais. O que se deseja saber é se a escolha do comercial está relacionada ao gênero da criança: pois se suspeita de que o gênero pode estar influenciando na escolha do comercial. Neste caso utilizaremos o *teste de Independência.* Pois, as variáveis são qualitativas.



Classificação das variáveis

1.5.1 – Variáveis qualitativas

Quando seus valores são expressos por atributos, de forma não numérica.
Ex: Estado civil, sexo, raça, cor dos cabelos, nível de instrução, classe social

As variáveis qualitativas se dividem em Nominal e Ordinal. Variável **Nominal** é aquela para a qual não existe nenhuma ordenação nas prováveis realizações.

Ex.: População: Alunos do Ensino a Distância
Variáveis: Sexo, religião, naturalidade, cor, tipagem sanguínea.

A variável **ordinal** é aquela para a qual existe uma ordem ou hierarquia nos possíveis resultados.
Ex.: População: Funcionários das empresas paraenses.
Variáveis: Classe social, grau de instrução.

1.5.2 – Variáveis quantitativas

Quando seus valores são expressos por números.

Ex: idade, peso, temperatura, número de filhos, volume, tempo, massa. As variáveis quantitativas ainda são classificadas como:

– discreta: quando os seus valores podem enumerados. Ex. Número de acertos na Mega Sena (0, 1, 2, 3, 4, 5 e 6).

– contínua: quando os seus valores podem ser qualquer um num intervalo.

Ex. Alturas dos jogadores de um time de futebol (1,5 *m*; 2,0 *m*; 1,79 *m*...)

1.6 – FASES DO MÉTODO ESTATÍSTICO

Agora estudaremos os seguintes itens:

1.6.1 Definição do problema
1.6.2 Planejamento
1.6.3 Coleta de dados
1.6.4 Crítica dos dados
1.6.5 Apresentação dos dados
1.6.6 Análise e interpretação dos dados

1.6.1 – Definição do Problema

Para chegar ao estágio de interpretação de dados, que é o objetivo final de uma pesquisa, é preciso passar por algumas etapas, denominadas fases do método estatístico.

Saber exatamente o que se pretende pesquisar é o mesmo que definir corretamente o problema. Portanto, a primeira fase consiste em uma definição ou formulação correta do problema a ser estudado.

Por exemplo:

A nota média no ENEM dos alunos do estado do Pará é menor do que as dos alunos dos outros estados?

1.6.2 – Planejamento

Nele se determina o procedimento necessário para resolver o problema, como levantar informações sobre o assunto objeto do estudo.

Nesta fase é importante a escolha das perguntas, que, na medida do possível, devem ser fechadas. No caso de um experimento, deve-se atentar para os objetivos que se pretende alcançar.

O levantamento de dados pode ser de dois tipos:

- Censitário (quando envolve toda a população)
- Por amostragem (quando é utilizada uma fração da população)

  Outros elementos do planejamento de uma pesquisa são: cronograma das atividades, custos envolvidos, exame das informações disponíveis, delineamento da amostra, etc..

1.6.3 – Coleta de dados.

É o registro sistemático de dados, com um objetivo determinado.

Dados primários: Quando são publicados pela própria pessoa ou organização que os haja recolhido.

Dados secundários: Quando são publicados por outra organização. <u>Ex:</u> quando determinado jornal publica estatísticas referentes ao censo demográfico extraídas do IBGE.

É mais seguro trabalhar com fontes primárias. O uso da fonte secundária traz o grande risco de erros de transcrição. Consiste na busca ou compilação dos dados. Quanto ao tempo, ela pode ser classificada em:

a) Contínua: quando realizada permanentemente. Ex.: inflação, registros de nascimentos e óbitos.
b) Periódica: quando é feita em intervalos de tempo. Ex.: Inflação mensal, censo.
c) Ocasional: quando efetuada sem época preestabelecida. Ex.: pesquisa de mercado, pesquisa eleitoral.

1.6.4 – Crítica dos dados

Objetiva a eliminação de erros capazes de provocar futuros enganos. Faz-se uma revisão crítica dos dados, suprimindo os valores estranhos ao levantamento.

1.6.5 – Apresentação dos dados

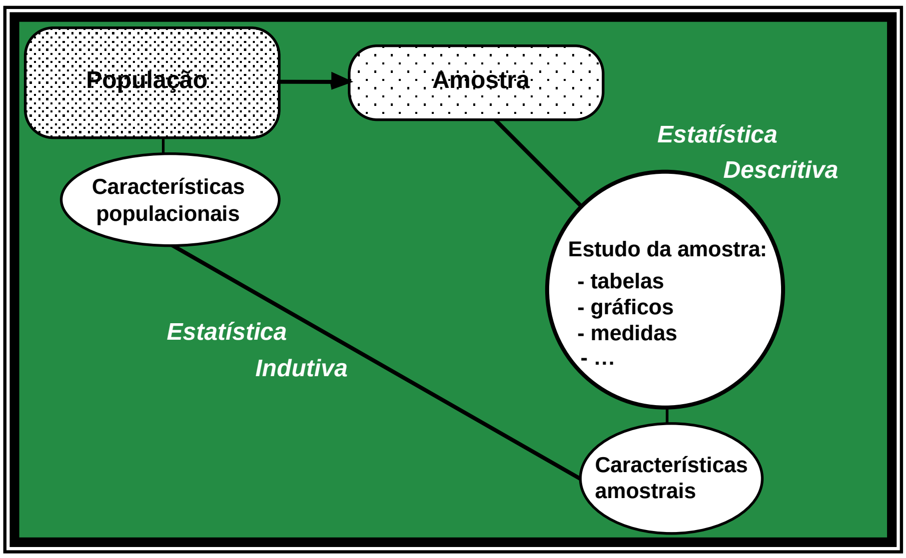
A organização dos dados denomina-se Série Estatística. Sua apresentação pode ocorrer por meio de tabelas ou gráficos.

1.6.6 – Análise e interpretação dos dados

Esta fase consiste em tirar conclusões que auxiliem o pesquisador a resolver seu problema, descrevendo o fenômeno através do cálculo de medidas estatísticas, especialmente as de posição e as de dispersão. O objetivo último da estatística é tirar conclusões sobre o todo (população) a partir de informações fornecidas por parte representativa do todo (amostra). Assim, realizadas as fases anteriores, fazemos uma análise dos resultados obtidos através dos métodos da estatística indutiva ou inferencial, que tem por base a indução ou a inferência, e tiramos desses resultados conclusões e previsões.

O seguinte esquema pretende resumir as diferentes etapas que normalmente são seguidas num procedimento estatístico:

ALGUNS EXEMPLOS DE APLICAÇÃO DA ESTATÍSTICA



Teoria da resposta ao item – Uma Universidade pretende estudar a proficiência dos estudantes submetidos ao processo seletivo e os parâmetros dos itens (questões) da prova.

População – Conjunto de todos os estudantes submetidos ao exame.

Amostra – Conjunto de alguns estudantes submetidos ao exame, que foram selecionados.

Problema – Estudar a proficiência dos estudantes e os parâmetros dos itens (questões) pelo método da Máxima verossimilhança marginal.

Medicina – Pretende-se estudar o efeito de um novo medicamento para curar determinada doença. É selecionado um grupo de 20 doentes, administrando-se o medicamento a 10 desses doentes escolhidos ao acaso, e o medicamento habitual aos restantes.

População – Conjunto de todos os doentes com a doença especifica.

Amostra – Conjunto dos 20 doentes selecionados.

Problema – Pretende-se, a partir dos resultados obtidos, realizar um *teste de hipóteses* para tomar uma decisão sobre qual dos medicamentos é melhor.

Controle de qualidade – O administrador de uma fábrica de parafusos pretende assegurar-se de que a percentagem de peças defeituosas, não excede um determinado valor, a partir do qual determinada encomenda poderia ser rejeitada.

População – Conjunto de todos os parafusos fabricados pela fábrica.

Amostra – Conjunto de alguns parafusos, escolhidos ao acaso, de entre o lote de produzidos.

Problema – A partir da percentagem de parafusos defeituosos presentes na amostra, estimar a percentagem de defeituosos em toda a produção.

Pedagogia – Um conjunto de pedagogos desenvolveu uma técnica nova para a aprendizagem da leitura na escola primária, a qual, segundo dizem diminuem o tempo de aprendizagem relativamente ao método habitual.

População – Conjunto dos alunos que entram para a escola primária sem saber ler.

Amostra – Conjunto de alunos de algumas escolas, selecionadas para o estudo. Os alunos foram separados em dois grupos para se aplicarem as duas técnicas em confronto.

Problema – A partir dos tempos de aprendizagem obtidos verificar se existe evidência significativa para afirmar que os tempos com a nova técnica são menores.

ATIVIDADES

1) Classifique as variáveis abaixo em quantitativas ou qualitativas:

   a) Número de livros em uma estante de biblioteca.
   b) Frequência cardíaca.

c) Diâmetro de artéria.
d) Raça
e) QI (Quociente de inteligência).
f) Diâmetro de esferas.
g) Número de casas de uma cidade sem rede de esgoto.
h) Classificação de um paciente quanto ao estagio de uma determinada doença.
i) Nota em uma prova de Estatística
j) Classificação em um concurso

2) Classifiquem em variáveis qualitativas ou quantitativas, as variáveis que estão no texto abaixo.

O Grupo de Estudos em Pesquisas Estatísticas e Computacionais (GEPEC) e o Laboratório de Sistemas de Informação e Georeferenciamento (LASIG), ambos da UFPA, realizaram mais uma pesquisa amostral durante a VII Parada do Orgulho GLBT, que aconteceu no dia 14 deste mês, em Belém. A iniciativa do GEPEC e do LASIG de estudarem o universo dos gays, lésbicas, bissexuais e transgêneros começou na Parada do Orgulho GLBT de 2007, dando continuidade este ano.

Os entrevistados foram abordados em diferentes pontos do evento e responderam a questões pessoais como idade, renda familiar, raça, grau de escolaridade. Quando o assunto foi saúde, por exemplo, os entrevistados disseram se usam ou não camisinha. Abordando assuntos sociais, os pesquisadores perguntaram sobre adoção de criança, participação em movimentos sociais e conhecimentos de projetos de leis que beneficiem a classe GLBT. Somada a esses assuntos, a questão da violência sofrida por eles também foi abordada.

Para a coordenadora da pesquisa, a professora Sílvia Almeida, os principais resultados da avaliação deste ano, a serem observados, referem-se ao fato de 48,78% dos GLBT´s declararem que já sofreram algum tipo de homofobia. Quanto ao tipo de homofobia, pode-se destacar que 75% dos participantes declararam que sofreram discriminação. Quanto ao tipo de discriminação sofrida, destaca-se que 30,31% dos GLBT´s declararam terem recebido tratamento diferenciado; 18,18% foram excluídos ou marginalizados em ambiente familiar, todos por conta de sua orientação sexual.

A importância geral deste trabalho consiste em contribuir para futuras políticas sociais e de saúde pública, no que se refere aos homossexuais. Além de contribuir para o combate das discriminações e das violências sofridas pela classe GLBT", ressaltou Sílvia Almeida, avaliando também os benefícios deste trabalho para a comunidade acadêmica. "Além da formação de banco de dados para futuras publicações científicas, tanto na área de estatística quanto nas outras áreas, como, por exemplo, Ciências Sociais, Política, entre outras", finalizou.

(Texto: Dandara Almeida - Assessoria de Comunicação UFPA .Foto: Divulgação Google)

3) Classifique cada uma das variáveis abaixo em qualitativa (nominal / ordinal) ou quantitativa (discreta / contínua):
a) Turma a que o aluno foi alocado (A ou B);
b) Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos, além de *não sei*).
c) Perda de peso de maratonistas na Corrida de São Silvestre, em quilos.
d) Tolerância ao cigarro (indiferente, incomoda pouco, incomoda muito)
e) Grau de satisfação da população brasileira com relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
f) Tolerância ao cigarro (indiferente, incomoda pouco, incomoda muito).
g) Intensidade da perda de peso de maratonistas na Corrida de São Silvestre (leve, moderada, forte).
h) Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (*sim* ou *não* são possíveis respostas para esta variável).

4) Qual a principal etapa da fase do método estatístico?

5) Elabore uma situação pratica e aplique as fases do método estatístico.

6) Cite uma situação do seu dia a dia, em que se observa o uso da estatística.

PESQUISANDO

Aprenda mais sobre a história da estatística e os conceitos estudados, acessando os seguintes sites:

http://www.natalest.hpg.ig.com.br/historia.html

http://www.estatisticapr.hpg.ig.com.br/historia.html

http://www.im.ufrj.br/~lpbraga/prob1/historia_estatistica.pdf

http://www.ufpa.br/gepec/

http://www.somatematica.com.br/estat/basica/pagina2.php

http://alea-estp.ine.pt/

http://www.leg.ufpr.br/~silvia/CE055/

http://feferraz.net/br/Docs/Tags/Teoria_da_Resposta_ao_Item_-_TRI

http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=4463

http://pepsic.bvs-psi.org.br/pdf/avp/v2n2/v2n2a02.pdf

http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=2717

UNIDADE 2 – REPRESENTAÇÃO TABULAR

Nesta unidade, trataremos da questão das tabelas e gráficos estatísticos. Também observaremos as séries estatísticas que são de fundamental importância no estudo descritivo. Pois, em todo estudo estatístico os dados observados necessitam serem organizados para que se faça a análise dos mesmos.

Estudaremos nesta unidade os seguintes itens:

2 Tabelas estatísticas
2.1 Tabela
2.2 Séries estatísticas
2.3 Representação gráfica
2.4 Tipos de gráficos
2.5 Distribuição de frequências

2 – TABELAS ESTATÍSTICAS

A apresentação tabular é uma apresentação numérica dos dados. Consiste em dispor os dados em linhas e colunas distribuídos de modo ordenado, segundo algumas regras práticas adotadas pelos diversos sistemas estatísticos. As regras que prevalecem no Brasil foram fixadas pelo Conselho Nacional de Estatística.

2.1 – Tabela: é um quadro que resume um conjunto de observações. As tabelas têm a vantagem de conseguir expor, sinteticamente em um só local, os resultados sobre determinado assunto, de modo a se obter uma visão global mais rápida daquilo que se pretende analisar.

Uma tabela compõe-se de: título, cabeçalho, corpo, rodapé e colunas (indicadoras e numéricas).

Título: O que? (fato) Onde?(lugar) Quando? (tempo)

Corpo			**Cabeçalho**
da	**Coluna**	**Coluna**	O cruzamento de linha com a coluna
Tabela	**indicadora**	**Numérica**	chama-se casa ou célula.

Rodapé: fonte, notas, observações.

Obs:

1) Recomenda-se *não delimitar* (**fechar**) por traços verticais, os extremos da tabela, à direita e à esquerda;
2) Usa-se um traço horizontal (**-**) quando o dado for nulo, inexisti o fenômeno;
3) Usa-se (**...**) quando não se dispuser dos dados, embora ele possa ser quantificado;
4) Usa-se zero (**0**) quando o valor é muito pequeno para ser expresso pela unidade utilizada.
5) Usa-se uma interrogação (**?**) quando o valor é duvidoso.

2.2 - SÉRIES ESTATÍSTICAS

É um conjunto de dados estatísticos referenciados aos seguintes fatores: tempo, local e fenômeno.

2.2.1– Série Temporal ou Cronológica: *Nesta série o elemento de variação é o tempo (dia, mês, ano, etc)*

Produção de café no Brasil em 1991 - 1995 → ***Título***

Anos	Produção (1.000 t)
1991	1.221
1992	2.234
1993	1.254
1994	1.445
1995	1.112

Fonte: IBGE → ***Rodapé***

Cabeçalho

Coluna indicadora

Coluna numérica

Corpo

2.2.2 – Série Geográfica*: O elemento de variação é o lugar (município, bairro, escola, etc.)*

Rebanhos brasileiros - 1992.

Espécie	*Quantidade* (1.000 cabeças)
Bovinos	*154.440*
Ovinos	*19.955*
Caprinos	*12.159*
Suínos	*34.532*

Fonte: IBGE

2.2.3 – Série Especificativa

O elemento de variação é a espécie (material escolar, produto de uma fabrica, remédios, etc.)

Produção de veículos de Autopropulsão
Brasil -1993

Tipos	Quantidade
Automóveis	1.100.278
Comerciais Leves	224.387
Comerciais Pesados	66.771

Fonte: ANFAVEA

2.2.4 – Série Mista: *É a junção de duas ou mais séries simples (geográfica, especificativa ou temporal).*

Nº de Casos de Malária por Município no período de 1993 a 1996

Município	Anos			
	1993	1994	1995	1996
Abaetetuba	1	3	2	2
Ananindeua	2	1	1	1
Barcarena	3	2	4	2
Belém	1	2	1	1

Fonte: Pesquisa de Campo do Curso de Farmácia-UFPA, fevereiro de 1997.

2.3 – REPRESENTAÇÃO GRÁFICA

Em relação a uma tabela estatística um gráfico estatístico possibilita uma impressão visual mais rápida da distribuição dos valores em estudo. Isto não significa que a representação tabular seja de pouca, mas a representação gráfica vem para complementá-la. Os gráficos estatísticos propiciam uma idéia inicial mais satisfatória da concentração e dispersão dos valores, uma vez que através deles os dados estatísticos se apresentam em termos de grandezas visualmente interpretáveis.

O que é um gráfico estatístico?
É uma forma de apresentação dos dados estatísticos, cujo 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.

2.3.1 – Requisitos fundamentais de uma representação gráfica:

a) **Simplicidade:** Deve possibilitar a análise rápida do fenômeno em estudo. Deve conter apenas o essencial.

b) **Clareza**: Deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.

c) **Veracidade**: Deve expressar a verdade sobre o fenômeno em estudo.

2.4 – TIPOS DE GRÁFICOS.

a) **Diagramas:** São gráficos geométricos de no máximo duas dimensões. Para sua construção usa-se o sistema cartesiano.

b) **Cartogramas**: É a representação sobre uma carta geográfica, sendo muito usado na Geografia, História e Demografia.

c) **Pictogramas:** A representação gráfica consta de figuras representativas do fenômeno. Desperta logo a atenção do público.

2.4.1 – Gráfico em linha ou em curva: Este tipo de gráfico utiliza a linha poligonal para representar a série estatística. Constitui uma aplicação do processo de representação das funções num sistema de coordenadas cartesianas. Neste sistema faz-se uso de duas retas perpendiculares; as retas são os **eixos coordenados** e o ponto de intersecção, a **origem**. O eixo horizontal é denominado **eixo das abscissas** (ou eixo dos **x**) e o vertical, **eixo das ordenadas** (ou eixo dos **y**). Considere a série abaixo:

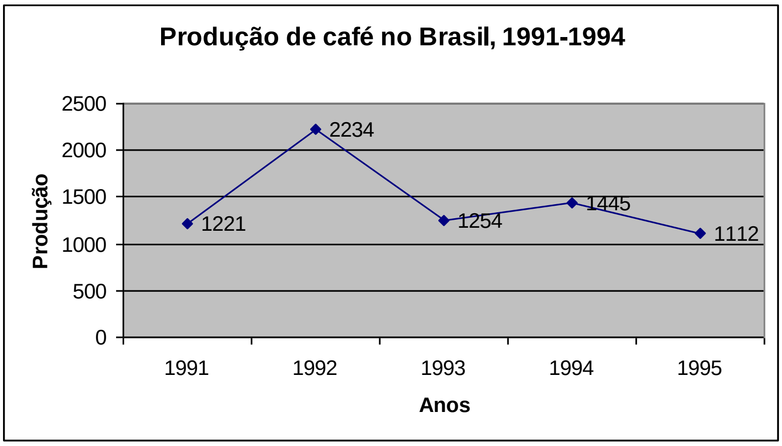
Produção de café no Brasil em 1991 - 1995

Anos	Produção (1.000 t)
1991	1.221
1992	2.234
1993	1.254
1994	1.445
1995	1.112

Fonte: IBGE

*Toma-se a coluna dos **ANOS** como abscissas e a coluna de **PRODUÇÃO** como ordenadas. Desta forma, um ano dado (**x**) e a respectiva quantidade da produção (**y**) formam um par ordenado (**x,y**), que pode ser representado num sistema cartesiano.*

*Determinados, graficamente, todos os pontos da série, usando as coordenadas, liga-se todos estes pontos, dois a dois, por segmentos de reta, que irão dar uma **poligonal**, que é o **gráfico em linha** ou **em curva** correspondente ao gráfico abaixo.*



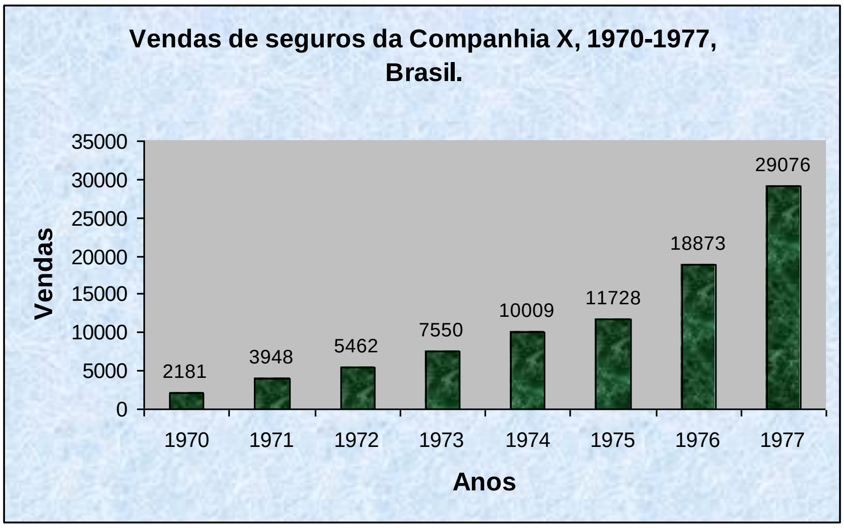
2.4.2 – Gráfico em coluna ou em barras: É a representação de uma série por meio de **retângulos**, dispostos **verticalmente** (em colunas) ou **horizontalmente** (em barras).

Quando em colunas, os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados. E Quando em barras, os retângulos têm a mesma altura e os comprimentos são proporcionais aos respectivos dados. Dada a série abaixo.

Anos	Vendas
1970	2.181
1971	3.948
1972	5.462
1973	7.550
1974	10.009
1975	11.728
1976	18.873
1977	29.076

Fonte: Departamento de marketing da Companhia X

Abaixo temos a representação gráfica da série acima.



NOTAS:

Sempre que os dizeres a serem inscritos forem extensos, deve-se dar preferência ao gráfico em barra (séries geográficas e específicas). Se ainda assim preferir o gráfico em coluna, os dizeres deverão ser dispostos de baixo para cima, nunca ao contrário.

A ordem a ser observada é a cronológica, se a série for histórica, e a decrescente, se for geográfica ou categórica.

À distância entre as colunas (ou barras), por questões estéticas, não deverá ser menor que a metade nem maior que os dois terços da largura (ou da altura) dos retângulos.

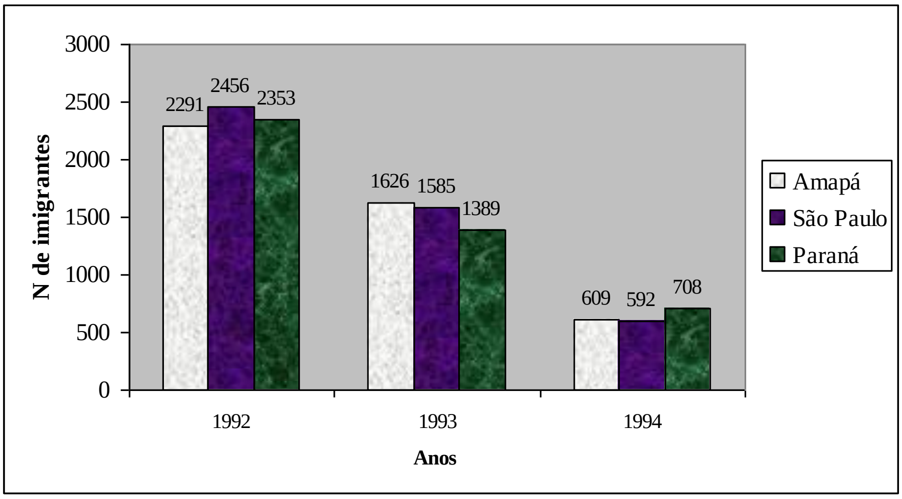
2.4.3 – Gráfico em coluna ou em barras múltiplas: Este tipo de gráfico é geralmente empregado quando se deseja representar, simultaneamente, dois ou mais fenômenos estudados com o propósito de comparação.

Entrada de migrantes em três Estados do Brasil - 1992-1994

Anos	Número de migrantes			
	Total	Estados		
		Amapá	São Paulo	Paraná
1992	4.526	2.291	1.626	609
1993	4.633	2.456	1.585	592
1994	4.450	2.353	1.389	708

Fonte: Fictícia

Abaixo temos a representação gráfica da série acima.



2.4.4 – Gráfico em setores: Este gráfico é construído com base em um círculo, e é empregado sempre que se deseja ressaltar a participação do dado no total.

O total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes.

Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série.

Obtém-se cada setor por meio de uma regra de três simples e direta, lembrando que o total da série corresponde a 360º.

Vacinas	Quantidade
BCG	3000
Sabin	5000
Tríplice	1500
Sarampo	600
Hepatite	400
Total	**10500**

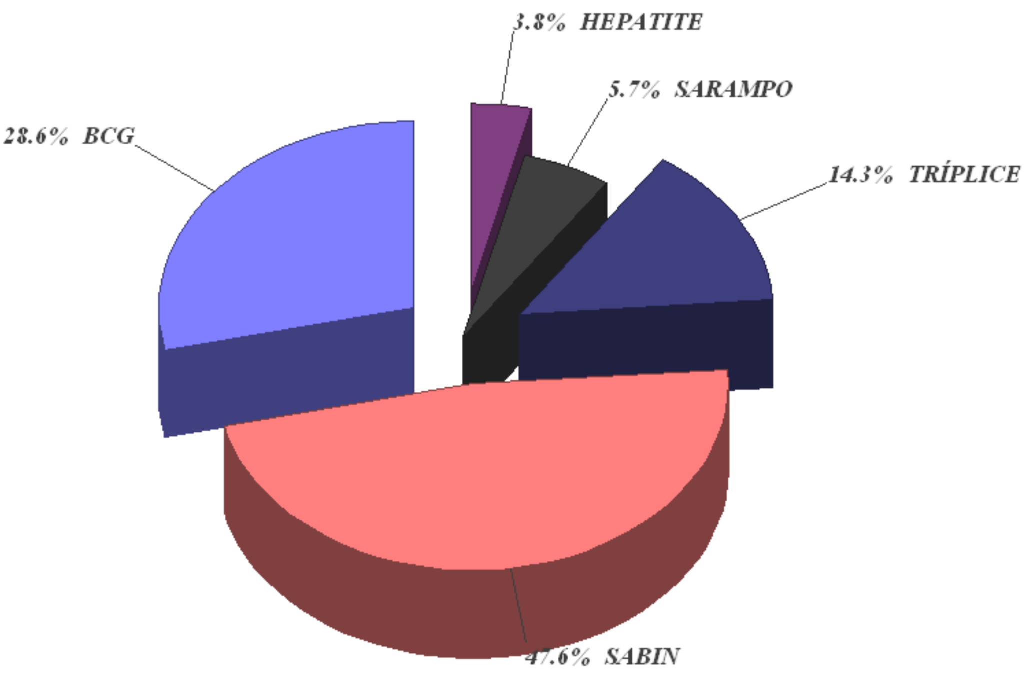
Fonte: Sespa-PA

Temos: 10500– 360º
3000 – X_1 $\Longrightarrow$ $X_1 = 102,86 \Rightarrow X_1 =$ **103º**

$X_2 = 171,43 \Rightarrow X_2 =$ ***172º;*** $X_3 = 51,43 \Rightarrow X_3 =$ ***51º;*** $X_4 = 20,57 \Rightarrow X_4 =$ ***20º;*** $X_5 = 13,72 \Rightarrow X_5 =$ **14º**

Com esses dados (valores em graus), marca-se num círculo de raio arbitrário, com um transferidor, os arcos correspondentes, obtendo o gráfico:

Vacinação de crianças de 0 a 1 ano de idade, Castanhal-PA, 1999.



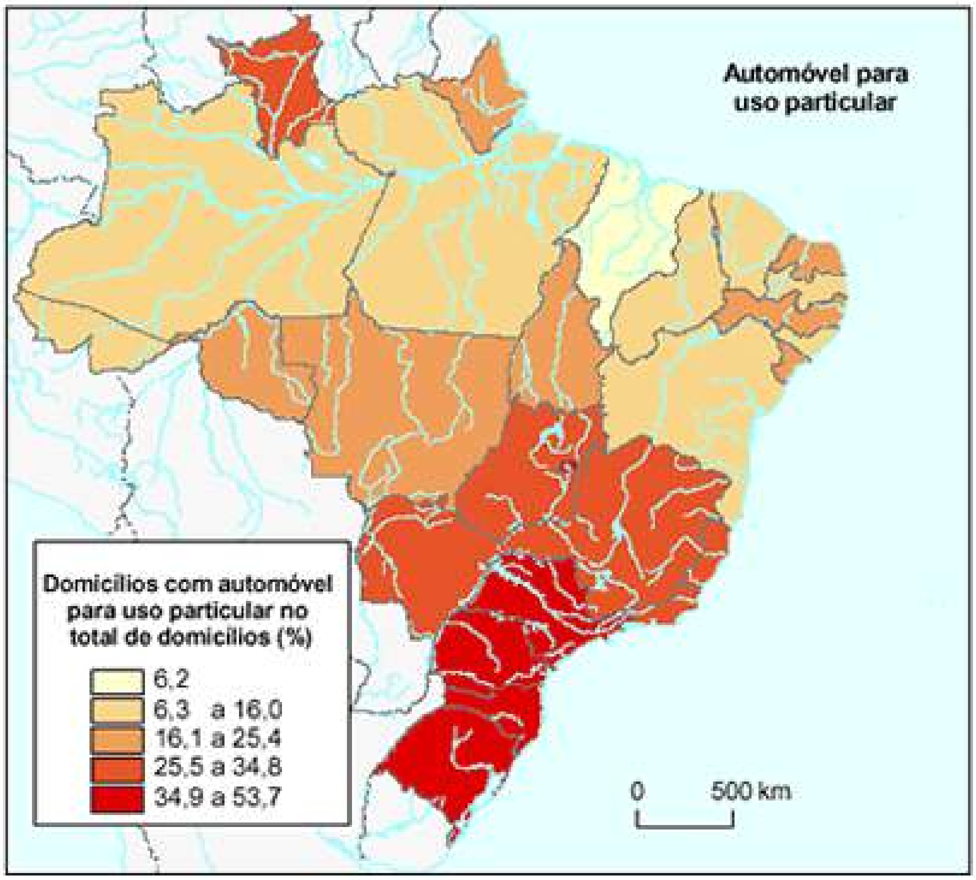
NOTAS:

O gráfico em setores só deve ser empregado quando há, no máximo, sete dados.

Se a série já apresenta os dados percentuais, obtêm-se os respectivos valores em graus multiplicando o valor percentual por 3,6.

2.4.5 – Cartograma: É a representação sobre uma carta geográfica. Este gráfico é empregado com o objetivo de representar dados estatísticos diretamente relacionados com áreas geográficas ou políticas.

Exemplo.

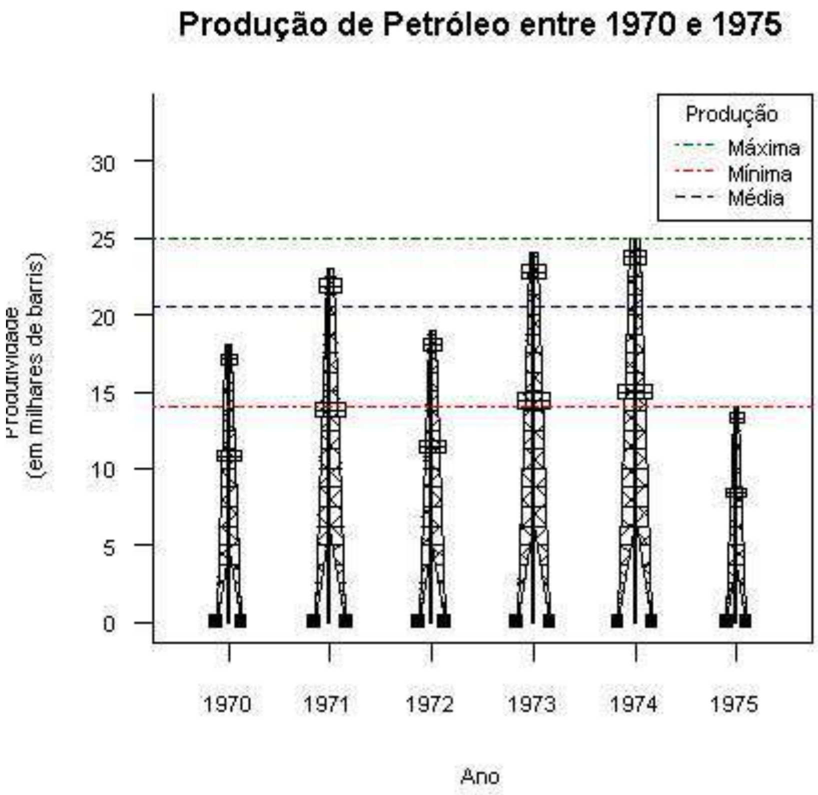


2.4.6 – Pictograma: Constitui um dos processos gráficos que melhor fala ao público, pela sua forma ao mesmo tempo atraente e sugestiva. A representação gráfica consta de **figuras** que lembrem o fato considerado.

Exemplo: Em meados de 1969, uma região situada ao sul do estado do Texas, no Estados Unidos, tinha sua economia baseada quase que exclusivamente no comércio varejista e no turismo. No entanto, em 1970, descobriu-se que nesta região havia petróleo. Estudos geológicos indicavam que ali, encontrava-se a maior reserva petrolífera já encontrada em solo norte-americano. Desde então, várias empresas começaram a explorar este petróleo.

O gráfico abaixo mostra a evolução da produção, entre os anos de 1970 e 1976 (em milhares de barris).

Obs: A linha verde indica a produção máxima, e a linha vermelha, a produção mínima.



Obs: O exemplo acima foi retirado do site: http://estatisticaesteouaquele.blogspot.com/2007/10/pictograma.html, em agosto de 2009.

2.5 – DISTRIBUIÇÕES DE FREQUÊNCIA

Na estatística trabalha-se, habitualmente, com grande número de informações, resultados de medições realizadas. Que podem ser dados discretos (o valor inteiro que não pode ser partido) ou contínuo (em intervalos).

Assim, é quase impossível examiná-los, mesmo que arrolados em ordem crescente ou decrescente (*Rol*). Daí, a necessidade de organizá-los em tabelas de distribuições de frequência.

2.5.1 – Tipos de distribuição de frequência

a) ***Sem Classe:*** *Notas Escola Jardim Encantado /97*

Notas(X)	*N° alunos(fi)*
3	2
4	2
5	3
6	4
7	4
8	4
9	1
Total	20

b) ***Com Classe:*** *Idade de 40 alunos Escola Atlas/98*

Classe	*Frequência (fi)*
5 I— 7	7
7 I— 9	5
9 I— 11	6
11 I— 13	4
13 I— 15	8
15 I— 17	7
17 I—I 19	3
Total	40

Fonte: IBGE

Observa-se que a primeira tabela é composta de duas colunas onde na primeira coluna encontram-se os valores obtidos da variável em estudo (notas dos alunos), apresentados de forma ordenada, cada nota correspondendo a uma classe; na segunda coluna encontram-se os números de alunos que obtiveram as respectivas notas (frequências – fi). O número total de alunos é a soma dos alunos em cada nota. 2 + 2 + 3 + 4 + 4 + 4 +10 = 20.
A primeira tabela de dados na forma acima é chamada **Distribuição de Frequências Sem Classe** *ou* **Por Pontos**. *Quando o conjunto de dados possui muitas observações diferentes ou quando a variável em estudo é continua, é conveniente construir uma distribuição de frequência em intervalos de classe (por intervalo). Para tanto alguns passos devem ser seguidos:*

2.5.2 – Distribuição de frequência por intervalos

CONCEITO: é uma série estatística na qual a variável observada está dividida em subintervalos do intervalo total observado e o tempo, a espécie e a região permanecem fixas

I – ELEMENTOS DE UMA DISTRIBUIÇÃO DE FREQUÊNCIAS

a) Convenções

|-----	Intervalo fechado à esquerda e aberto à direita: apenas o limite inferior pertence ao intervalo;
-----|	Intervalo aberto à esquerda e fechado à direita: apenas o limite superior pertence ao intervalo;
|----|	Intervalo fechado de ambos os lados: os dois limites pertencem ao intervalo;
-----	Intervalo aberto em ambos os lados: os dois limites não pertencem ao intervalo.

Observação: Um símbolo como 5 **----|** 10 é chamado intervalo de classe. O intervalo de classe apresenta dois limites: um inferior e outro superior. Para esse intervalo o limite inferior é 5 e o limite superior é 10.

a) **Limite inferior da distribuição de frequência (LI):** é o valor a partir do qual são contadas as observações na distribuição de frequências.

b) **Limite superior da distribuição de frequência (LS)**: é o valor até o qual são contadas as observações na distribuição de frequências.

c) **d) Amplitude total da distribuição de frequência (AT)**: é a diferença existente entre o maior e o menor valor observado da distribuição de frequência.

$$AT = LS - LI$$

d) **Classes de uma distribuição de frequência**: são os subintervalos nos quais são contadas as observações da variável.

Observação: o número de classes (**K**) é calculado a partir de uma das expressões mostradas a abaixo.

K = 1 + 3,322 log. N $\Rightarrow$ fórmula de STURGES.

Método Prático: se n < 25 utilize k = 5 se n $\geq$ 25 utilize $k = \sqrt{n}$.

Observação: existem *"n"* maneiras de calcularmos o número de classes, depende da sensibilidade do pesquisador.

e) **Limite Inferior de Classe (li)**: é o valor a partir do qual são contadas as observações dentro da classe.

f) **Limite Superior de Classe (ls)**: é o valor até o qual são contadas as observações dentro da classe.

g) **Amplitude de Classe (at)**: é a diferença entre o maior e o menor valor observado dentro da classe.

Observação: A amplitude de classe é obtida através da seguinte equação: $at = ls - li$.

h) **Frequência Simples ou Frequência Absoluta da Classe (fi)**: é o número de observações contadas dentro da classe.

i) **Frequência Absoluta acumulada de Classe (Fi)**: é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer, das frequências simples ou absoluta das classes.

j) **Frequência Relativa de Classe (fr)**: é a relação existente entre a frequência absoluta ou simples de classe e o número de observações da variável.

Observação: Obtém-se a frequência relativa de cada classe a partir da seguinte equação:

$$fr = f_i / \sum f_i .$$

k) **Frequência Relativa Acumulada (Fr)**: é a acumulação sucessiva, a partir da primeira classe até uma classe qualquer das frequências relativas das classes.

l) **Ponto Médio de Classe (xi)**: é a média aritmética calculada entre o limite inferior e o superior da classe.

Observação: Obtém-se o ponto médio de cada classe a partir da seguinte equação: $X_i = \frac{ls+li}{2}$.

m) **Intervalo de Classe ou Amplitude do intervalo de Classe (h):** é o comprimento da classe.

Observações:

Obtém-se o intervalo de cada classe a partir da seguinte equação: $h = \frac{AT}{K}$.

Convém arredondar o número correspondente à amplitude do intervalo de classe para facilitar os cálculos.

As séries de dados grupados (distribuição de frequências por intervalos e por pontos) são também chamadas de "séries de magnitude de variável".

II – CONSTRUÇÃO DA DISTRIBUIÇÃO DE FREQUÊNCIAS

1º Passo: Colocar os dados em forma de Rol. Isto é, organizá-los de forma crescente ou decrescente. Aqui se recomenda colocá-los em ordem crescente;

2º Passo: Identificar o valor máximo e o valor mínimo do conjunto de dados e encontrar a Amplitude Total (AT). Definimos por Amplitude Total (At) a diferença entre o maior e o menor valor do conjunto de dados:

$$AT = Valor \quad Máximo - Valor \quad Mínimo = LS - LI$$

3º Passo: Determinar o Número de Classes (K) que irão formar uma distribuição de frequências. Embora não exista uma formula precisa para esse número K, pode-se orientar a seguinte prática:

K = $\sqrt{n}$ aproximadamente.

4º Passo: Calcular o comprimento ou a amplitude que deve ter o Intervalo de Classe (h), que é obtido dividindo-se a amplitude total pelo número de classe, ou seja: $h = \dfrac{AT}{K}$

Exemplo:

Precisamos organizar as notas da primeira avaliação da disciplina estatística dos 25 alunos matriculados regularmente. As notas são dadas abaixo:
10-10-9,5-8-5-5-5-6-1-0-0-1-3-2-5-7-10-8-6,5-8,5-7,5-4,5-4-9-8 (dados brutos)

1º Passo: Organizar em Rol - 0-0-1-1-2-3-4-4,5-5-5-5-5-6-6,5-7-7,5-8-8-8-8-8,5-9-9,5-10-10.
2º Passo: Obter a Amplitude Total. $AT = 10 - 0 = 10$
3º Passo: Calcular o número de classes (K). K = $\sqrt{25} = 5$
4º Passo: obter Intervalo de Classe (h) e escrever os intervalos da tabela.

$h = \frac{10}{5} = 2$ (Intervalo de classe)

Acompanhe todos os passos!

1) Comece pela 1ª classe, e escreva o menor valor observado.

2) Agora some o menor valor com o h=2 , assim encontramos o limite superior da classe. [0+2=2].

Notas	Freqüência
0 |-- 2	4
2 |-- 4	2
4 |-- 6	6
6 |-- 8	4
8 |-- 10	9
Total	25

3) A partir da 2ª classe usa-se a regra, o limite superior da classe anterior será o limite inferior da classe subsequente, e o limite superior é o resultado da soma do limite inferior com o h (intervalo de classe).

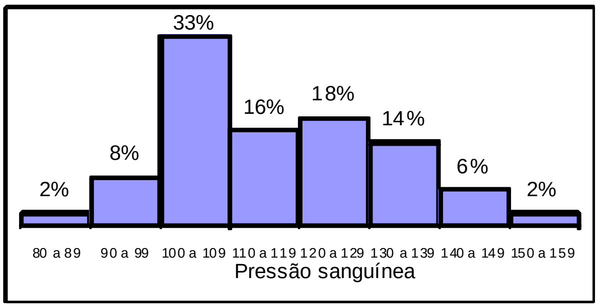
4) Agora temos que contar os elementos que pertencem a cada intervalo. Por exemplo, na 1ª classe temos: 0 |-- 2 , que significa, intervalo fechado à esquerda e aberto à direita. Os valores deste intervalo se aproximam de 2, isto é, não pertencem ao intervalo.

5) Os elementos: 0-0-1-1, pertencem a 1ª classe, então colocamos o valor 4 na 2ªcoluna(Frequência), pois, temos quatro elementos que pertencem ao intervalo. Já os elementos: 2-3, pertencem a 2ª classe, então na 2ªcoluna(Frequência) colocamos o valor 2. Assim é feito para as demais classes, observa-se o Rol

III – REPRESENTAÇÂO GRÀFICA DA DISTRIBUIÇÃO DE FREQUÊNCIAS

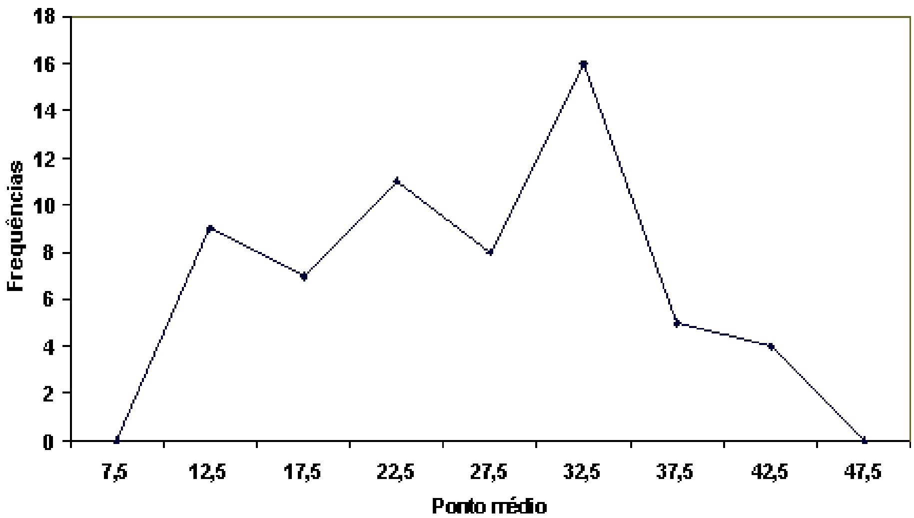
a) **Histograma:** é a representação gráfica de uma distribuição de frequência, através de retângulos justapostos onde a base é colocada no eixo das abscissas corresponde ao intervalo das classes, e a altura é dada pela frequência absoluta (ou relativa) das classes.

Pressão arterial, em milímetros de mercúrio, de 50 cães anestesiados



b) **Polígono de Frequência:** é outro tipo de apresentação bastante comum para dados quantitativos, ou seja, é um sumário gráfico que pode ser preparado para dados que tenham sido sumariamente sintetizados numa distribuição de frequência. Utilizando-se os pontos médios de cada classe para a construção do mesmo, ou seja, é um gráfico em linhas, sendo que as frequências são marcadas no eixo vertical e no eixo horizontal são colocados os pontos médios dos intervalos de cada classe. Utilizando o exemplo da tabela abaixo obtemos os seguintes gráficos

Polígono de frequências da pontuação de 60 alunos da escola Aberlado Condurú em 1999



Agora mostraremos a aplicação da frequência acumulada na distribuição de frequências, que será muito útil na obtenção da mediana e das medidas separatrizes, que serão estudadas na próxima unidade:

CLASSE	*fi*	*Fac*
50 |-------- 54	4	4
54 |-------- 58	9	13
58 |-------- 62	11	24
62 |-------- 66	8	32
66 |-------- 70	5	37
70 |-------- 74	3	40
Total	40	

Onde, ***fi*** - frequência simples
Fac- frequência acumulada

Inicialmente repete-se a *fi* da 1ª classe

Soma-se a *fi* desta classe com a acumulada anterior: 9+4=13

O mesmo procedimento é feito para as demais classes;

Agora estudaremos um item muito simples que faz a diferença quando o nosso objetivo é calcular a média aritmética em distribuições de frequências, estamos falando do ponto médio, *xi*, que é a média aritmética calculada entre o limite inferior e o superior da classe.

CLASSE	*fi*	*xi*
50 |-------- 54	4	52
54 |-------- 58	9	56
58 |-------- 62	11	60
62 |-------- 66	8	64
66 |-------- 70	5	68
70 |-------- 74	3	72
Total	40	

$$xi = \frac{50+54}{2} = 52$$

$$xi = \frac{54+58}{2} = 56$$

$$xi = \frac{58+62}{2} = 60$$

$$xi = \frac{62+66}{2} = 64$$

$$xi = \frac{66+70}{2} = 68$$

$$xi = \frac{70+74}{2} = 72$$

Agora nosso foco será na construção da frequência relativa que é um tópico bastante interessante. A obtenção da frequência relativa é a seguinte: divide-se o valor de pela frequência total, como podemos observar abaixo.

CLASSE	*fi*	*fri*
50 |-------- 54	4	0,100
54 |-------- 58	9	0,225
58 |-------- 62	11	0,275
62 |-------- 66	8	0,200
66 |-------- 70	5	0,125
70 |-------- 74	3	0,075
Total	40	1,000

$$fri = \frac{4}{40} = 0{,}100$$

$$fri = \frac{9}{40} = 0{,}225$$

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

$$fri = \frac{8}{40} = 0{,}200$$

$$fri = \frac{5}{40} = 0{,}125$$

$$fri = \frac{3}{40} = 0{,}075$$

Obs: Se o interesse for apresenta a frequência relativa em termos percentuais, basta multiplicar cada frequência relativa por 100.

ATIVIDADES

1) Crie uma série estatística.

2) Pesquise em livros ou na internet uma série estatística.

3) Quais que elementos que compõe uma tabela estatística?

4) Qual a série estatística que é a junção de outras duas?

5) Uma tabela estatística pode ser fechada, ou seja, ser delimitada nos extremos? Justifique.

6) Quais as perguntas que uma tabela estatística precisa responder?

7) Qual a vantagem de um gráfico estatístico com relação a uma série estatística?

8) Quais os requisitos fundamentais de uma representação gráfica?

9) Quais os tipos de gráficos?

10) Faça uma representação gráfica da série que você criou na questão 1.

11) Fazer uma tabela estatística para representar o movimento religioso de certo município no período 1975-1977, que apresentou os seguintes dados: em 1975, houve 56.738 habitantes batizados, 15.884 casamentos e 13.678 extremas-unções. Em 1976, houve 63483 batizados; os casamentos foram em número de 17.032 e as extremas-unções 14.328. Em 1977, realizou-se um total de 71.232 batizados; as extremas-unções foram 16.107 e os casamentos 16.774. Classifique esta série estatística e faça sua representação gráfica.

12) Quais os tipos de distribuição de frequências?

13) Quais os passos para a construção de uma Tabela de frequências?

14) O que é o rol? Para que serve?

15) 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 frequência acumulada;
b) Os pontos médios;
c) A frequência relativa;
d) A percentagem de alunos com menos de 24 anos.

16) Construa um diagrama de setores, percentual, correspondente aos empregados da Martins Ltda que possui a seguinte distribuição por área de trabalho: Diretoria (3 pessoas), Assessoria (6 pessoas), Transporte (18 pessoas), Administração (5 pessoas), Área técnica (15 pessoas) e Área operacional (33 pessoas).

17) Completar os dados que faltam:

Valores	fi	Fac	fr
1	4		0,08
2	4		
3		16	0,16

4	7		0,14
5	5	28	
6		38	
7	7	45	0,14
8			

18) **(AFRF-2002)** Em um ensaio para o estudo da distribuição de um atributo financeiro (X) foram examinados 200 itens de natureza contábil do balanço de uma empresa. Esse exercício produziu a tabela de frequências abaixo. A coluna *Classes* representa intervalos de valores de X em reais e a coluna *P* representa a frequência relativa acumulada. Não existem observações coincidentes com os extremos das classes.

Classes	Porcentagem Acumulada
70|- 90	5
90|- 110	15
110|- 130	40
130|- 150	70
150|- 170	85
170|- 190	95
190|- 210	100

Qual a estimativa da frequência relativa de observações de X menores ou iguais a 145?

19) Faça uma representação gráfica da série dada abaixo:

Vendas de Seguros da companhia Vera Cruz 1970 - 1977 Brasil

Ano	Vendas
1970	2.181
1971	3.948
1972	5.462
1973	7.550
1974	10.009
1975	11.728
1976	18.873
1977	29.076

Fonte: Departamento de marketing da Companhia

20) Qual a diferença entre o gráfico de barras e o Histograma?

21) Qual a diferença entre o gráfico de linhas e o polígono de frequências?

22) Classifique a série dada abaixo.

Duração média de estudos Superiores na Europa em 1998.

Países	Nº de anos
Itália	7,5
Alemanha	6,5
França	5,5
Holanda	4,0

Fonte: Revista Veja

23) Para os dados abaixo construa duas distribuições de frequências: uma com classe e a outra sem classe:

45, 41, 42, 41, 42 43, 44, 41, 50, 46, 50, 46, 60, 54, 52, 58, 57, 58, 60, 51

24) Crie uma série de dados e faça a duas distribuições de frequências: uma adotando a fórmula de Sturges e a outra não. Compare as diferenças.

25) O número de gols marcados no último campeonato da Federação Paulista de Futebol pelos 20 clubes participantes nos seus 38 jogos é uma variável com os seguintes valores:

Clubes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Gols	32	42	73	35	79	57	37	52	35	25	55	70	42	41	68	36	74	29	47	53

a) Classifique a variável em estudo. Construa uma distribuição (tabela) de frequência para essa variável agrupando as observações em intervalos de comprimento 10 a partir de 20.
b) Obtenha o histograma correspondente.
c) Que porcentagem dos clubes marcou mais de 38 gols?

26) O posto de saúde de certo bairro mantém um arquivo com o número de crianças nas famílias que utilizam o posto. Os dados são os seguintes: 3, 4, 3, 4, 1, 5, 6, 3, 4, 5, 3, 4, 3, 3, 4, 3 , 5, 5, 5, 5, 6, 11, 10, 2, 1, 2, 3, 1, 5 e 2.
a) Organize uma tabela de frequência;
b) Construa um histograma.

27) Uma indústria de laticínios está planejamento redirecionar seus produtos. Para tanto, decidiu investigar a quantidade de leite (em litros) consumida diariamente por diferentes famílias. Os dados (ordenados) estão a seguir: 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4,4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 6, 7, 7, 8 e 15.
a) Obtenha a distribuição de frequências;
b) Obtenha a frequência relativa percentual.

PESQUISANDO

Para saber mais sobre a unidade que acabamos de estudar sugiro que pesquise:

- Acesse o site:
http://www.ime.usp.br/~mae

http://alea-estp.ine.pt

www.estatistica.ccet.ufrn.br

http://www.ibge.gov.br

http://www.heliorocha.com.br/graduacao/sisinfo/download/PES/DistribuicaoDeFrequencia.pdf

- Os seguintes livros:

BUSSAB, W. O. , MORETTIN, P.A, **Estatística Básica** 5ª ed. São Paulo: SARAIVA, 2002.

VIEIRA, Sonia. *Princípios de Estatística. 1ª reimpr. da 1ª ed.* São Paulo: Editora Pioneira Thomson Learning, 2003.

UNIDADE 3 – MEDIDAS DE TENDÊNCIA CENTRAL

Nesta unidade, conheceremos as principais medidas de posição. Será nosso foco a obtenção dessas medidas através das distribuições de frequências que foram estudadas na unidade anterior. Também abordaremos as medidas separatrizes que são medidas que ocupam determinados lugares na distribuição de frequências.

Agora estudaremos os seguintes itens abaixo:

3 Introdução
3.1 Conceito
3.2 Média aritmética Simples
3.3 Média aritmética Ponderada
3.4 Média geométrica
3.5 Média harmônica
3.6 Mediana
3.7 Moda
3.8 Medidas separatrizes

3 – INTRODUÇÃO: A Estatística também é chamada de a ciência das médias, portanto vamos estudar nesta unidade as medidas de posição que são bastante utilizadas na prática. Torna-se necessário, após a tabulação dos resultados e da representação gráfica, encontrar valores que possam representar a distribuição como um todo. São as chamadas medidas de tendência central ou medidas de posição.

3.1 – CONCEITO: as medidas de posição ou também conhecidas como medidas de tendência central compõem-se de um número que representa um conjunto particular de informações. Geralmente se localizam em torno do centro da distribuição, onde a maior parte das observações tende a concentra-se.

3.2 – MÉDIA ARITMÉTICA SIMPLES: consiste em somar todas as observações ou medidas dividindo-se o resultado pelo número total de valores.

Observação: têm-se duas formas de calcular uma média aritmética:

1. Quando estamos trabalhando com dados brutos:

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

onde:

$\overline{X}$ – é a média aritmética;

x_i – são os valores das observações;

n – número de observações.

Exemplo 1: seja o conjunto X = {8, 9, 10, 10, 8}, calcule a média.

$$\overline{X} = \frac{(8+9+10+10+8)}{5} = 9{,}0$$

3.2.1 – PROPRIEDADES DA MEDIA ARITMÉTICA

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

$$\sum (x_i - \overline{X}) = 0.$$

Ex: voltando ao exemplo 1, temos que a média é 9, então calculando os desvios em torno da média tem-se:

$$\sum (x_i - \overline{X}) = (8-9)+(9-9)+(10-9)+(10-9)+(8-9) = -1+0+1+1-1 = 0.$$

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

$$\frac{\sum (x_i + k)}{n} = \frac{\sum x_i + \sum k}{n} = \frac{\sum x_i + nk}{n} = \overline{X} + k.$$

c) 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.

$$\frac{\sum kx_i}{n} = \frac{k\sum x_i}{n} = k\overline{X}.$$

3.3 – MÉDIA ARITMÉTICA PONDERADA: é uma média aritmética na qual será atribuído um peso a cada valor da série.

$$\overline{X} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

$\overline{X}$ – **é a média aritmética ponderada;**
f_i – **frequência simples;**
n – **número de observações**.

Observação: se os valores forem distribuídos em classes o x_i será o ponto médio de cada classe.

Exemplo: Uma escola adotou os seguintes pesos para as notas bimestrais:

1º bimestre peso 1 — 3º bimestre peso 3
2º bimestre peso 2 — 4º bimestre peso 4

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

$$\overline{X} = \frac{5.1 + 4.2 + 3.3 + 2.4}{10} = \frac{5+8+9+8}{10} = \frac{30}{10} = 3$$

3.4 – MÉDIA GEOMÉTRICA:

É apropriada para aqueles casos em que comportamento dos valores da série que se está estudando, possuem um comportamento progressivo, tendendo a uma progressão geométrica.
Para dados brutos temos que, a fórmula da média geométrica é:

$$Mg = \sqrt[n]{x_1.x_2.x_n} = \sqrt[n]{\prod x_i} = \prod x_i^{1/n}$$

Exemplo: A média geométrica dos seguintes dos números: 1, 6, 36 é obtida da seguinte forma:

$$Mg = \sqrt[3]{1.6.36} = \sqrt[3]{216} = 6$$

Para dados tabelados temos que, a fórmula da média geométrica é:

$$Mg = \sqrt[f_1+f_2+\ldots+f_n]{x_1^{f_1}.x_2^{f_2}.x_n^{f_n}} = \sqrt[\sum f_i]{\prod x_i^{f_i}} = \prod x_i^{f_i/\sum f_i}$$

Observação: se os valores forem distribuídos em classes o x_i será o ponto médio de cada classe.
Exemplo: Um aluno realiza três provas, obtendo as seguintes notas: 10, 8 e 7, onde os pesos atribuídos a elas são os seguintes, 1ªprova peso 2, 2ªprova peso 3 e 3ªprova peso 4 . Calcule a média geométrica.

$$Mg = \sqrt[6+4+5]{2^6.3^4.4^5} = \sqrt[15]{5.308.416} \cong 2,8$$

3.5 – MÉDIA HARMÔNICA: É definida como o inverso da média aritmética dos inversos dos valores da série.
Para dados brutos temos que, a fórmula da média harmônica é:

$$Mh = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}} = \frac{n}{\sum(\frac{1}{x_i})}$$

Exemplo: Determinar a média geométrica dos números: 2,4 e 8.

$$Mh = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \frac{3}{0,875} \cong 3,43$$

Para dados tabelados temos que, a fórmula da média harmônica é:

$$Mh = \frac{f_1+f_2+\ldots+f_n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_n}{x_n}} = \frac{n}{\sum(\frac{f_i}{x_i})}$$

Exemplo: Calcular a média harmônica da distribuição de frequências dada abaixo:

Notas dos alunos da disciplina Estatística
no Instituto Datavox, ano de 2002

NOTAS	f_i	x_i
1 |-- 3	5	2
3 |-- 5	7	4
5 |-- 7	8	6
Total	20	20

$$Mh = \frac{5+7+8}{\frac{5}{2}+\frac{7}{4}+\frac{8}{6}} = \frac{20}{\frac{67}{12}} \cong 3,58$$

3.6 – MEDIANA (*Me*): é o valor central em um rol, ou seja, a mediana de um conjunto de valores ordenados, ou ainda a mediana divide a distribuição ao meio.

3.6.1– Mediana de valores brutos

Ordenar os valores em ordem crescente (*Rol*);
Verifica se o número de elementos é par ou ímpar;
Se ***n*** for ímpar, posição da mediana no conjunto, será o valor localizado na posição dada por: $\boxed{P = \frac{n+1}{2}}$
Se ***n*** for par, o conjunto terá dois valores centrais, neste caso, a mediana será igual à média aritmética dos valores centrais, cujas posições são dadas por:

$P_1 = n / 2$ e $P_2 = (n / 2) + 1$

Exemplo 1: calcule a mediana dos valores: 2 ; 5; 7; 15; 13; 4; 10.
Rol: 2; 4; 5; 7; 10; 13; 15.
n = 7 (ímpar)
Posição da mediana: ***P* = (7 + 1) / 2 = 4**
***Me* = 7**

Exemplo 2: Em um grupo de 6 pessoas cujas as alturas medidas em centímetros fossem as seguintes: 183 cm, 170 cm, 165 cm, 180 cm, 185 e 160 cm, qual a altura mediana deste grupo de pessoas?
Rol: 160; 165; 170; 180; 183; 185.
n = 6 (par)
Posição da mediana:
$P_1 = 6 / 2 = 3$
$P_2 = (6 / 2) + 1 = 4$
A mediana será a média aritmética das posições P_1 e P_2 , então:
Me= (170+180)/2= 175

3.5.2 – Mediana de valores tabelados

$$Me = l_i + \left[\frac{\frac{\sum_{i=1}^{n} f_i}{2} - FAA}{f_i} \right] \times h$$

l_i = Limite inferior da classe da mediana;
FAA = Frequência acumulada anterior da classe da *Me*;
f_i = Frequência simples da classe da mediana;
h = Intervalo de classe.

1º passo calcula-se a posição

$$\boxed{p = \frac{\sum_{i=1}^{n} f_i}{2};}$$

2º passo: identifica-se a classe $Mediana_i$ pela coluna das Frequências Acumuladas;

3º passo: Aplica-se a fórmula:

Exemplo: Calcule a mediana da distribuição dada abaixo:

NOTAS	f_i
1 |-- 3	5
3 |-- 5	7
5 |-- 7	8
Total	20

Resolução:

a) Calcule a Mediana.

1º passo: calcula-se a posição

$$p=\frac{\sum_{i=1}^{n} f_i}{2}=\frac{20}{2}=10;$$

2º passo: identifica-se a classe *Mediana$_i$* pela coluna das Frequências Acumuladas;
Comparamos o valor da posição *P* com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta Fac é maior ou igual a *P*?".**

NOTAS	f_i	*Fac*	
1 |-- 3	5	5	→ **5 é maior ou igual a 10? NÃO!**
3 |-- 5	**7**	**12**	→ **12 é maior ou igual a 10? SIM! Então esta é classe da *Mediana*.**
5 |-- 7	8	20	
Total	20		

3º passo: Aplica-se a fórmula:

$$Me = 3+\left[\frac{10-5}{7}\right]\times 2 = 3+1{,}4285 \cong 4{,}43$$

Logo, a Mediana é aproximadamente 2,43.

Observação: a mediana é muito empregada em pesquisas onde não interessam valores extremos, por terem pouca significação para o conjunto em geral.

3.7 – MODA (*Mo*): é aquilo que está em evidencia, o valor que mais aparece num conjunto de informações ou o de maior frequência em uma tabela.

Observação: a moda pode não ser única ou ate mesmo pode não existir.

3.7.1 – Moda de valores brutos: basta observar o valor que mais aparece no conjunto.
Exemplo: 3 ; 3 ; 6 ; 8 ; 10 ; 10; 10; 11; 11; 12 *Mo* = 10.

3.7.2 – Moda de valores tabelados: numa distribuição de frequência chamamos classe modal à classe que possui maior frequência. Como o ponto médio é representativo de qualquer classe de frequências, chamamos moda bruta ao ponto médio da classe modal.

MODA PELO PROCESSO DE *CZUBER*

$$Mo = l_i + \left[\frac{f_{Mo} - f_{ant}}{2f_{Mo} - (f_{ant} + f_{post})}\right] \times h$$

***l_i* = Limite inferior da classe modal;**
***fpost* = Frequência simples posterior à classe modal;**
***fant* = Frequência simples anterior à classe modal;**
***h* = Intervalo de classe;**
***f_{MO}* = Frequência modal.**

Observação: É valida a seguinte relação empírica

$$\bar{x} = \frac{3Md - M_O}{2}$$

Exemplo: Calcular a moda da distribuição dada abaixo:

NOTAS	f_i
1 |-- 3	5
3 |-- 5	7
5 |-- 7	**8**
Total	20

Primeiramente, observamos a coluna das frequências simples e verificamos qual o maior valor.
Neste caso a última classe apresenta o maior valor, 8, em relação às outras classes. Agora é só usar a fórmula pelo processo de ***CZUBER*** .

$$\boxed{Mo = 5 + \left[\frac{8-7}{2\times 8 - (7+0)}\right] \times 2 = 5 + \frac{2}{9} \cong 5{,}22}$$

Então, a Moda é aproximadamente 5,22.

3.8 – MEDIDAS SEPARATRIZES

3.8.1 – <u>Quartis</u> (Q_i): são os valores que dividem um conjunto de dados em quatro partes iguais, representados por Q_1, Q_2 e Q_3 denominam-se primeiro, segundo e terceiro quartis, respectivamente, sendo o valor de Q_2 igual à mediana. Assim, temos;

0%	-----------------	25%	-----------------	50%	-----------------	75%	-----------------	100%
		Q_1		Q_2		Q_3		

A fórmula para determinação dos quartis para dados agrupados é semelhante à usada para o cálculo da mediana.

a) Quartis para dados brutos:

Vamos abordar este assunto por meio de dois exemplos:
Exemplo 1: calcule o Q_1

Dados: 1,9 – 2,0 – 2,1 – 2,5 – 3,0 – 3,1– 3,3– 3,7– 6,1 – 7,7 ⟶ n=10

Posição do $\mathbf{Q_1}$= 0,25(n+1).
Então, a posição de $\mathbf{Q_1}$ é: 0,25 (11) = 2,75
Como o valor encontrado não é inteiro, temos que o 1º Quartil ($\mathbf{Q_1}$) será a média aritmética entre o 2º e 3º elementos:

$$Q_1 = (2 + 2{,}1)/2 = 2{,}05$$

Portanto, o 1º Quartil ($\mathbf{Q_1}$) é igual a **2,05.**
Exemplo 2: calcule o Q_1

Dados: 0,9 – 1,0 – 1,7 – 2,9 –3,1– 5,3– 5,5– 12,2– 12,9 – 14,0 – 33,6 ⟶ n=11

Posição do $\mathbf{Q_1}$= 0,25(n+1).
Então, a posição de $\mathbf{Q_1}$ é: 0,25 (11+1) = 3;
Portanto, o 1º Quartil ($\mathbf{Q_1}$) é igual a **1,7.**
Obs: para calcular o Q_3 a fórmula da posição seria: 0,75(n+1), pois, temos 75% dos dados abaixo dele.

b) Quartis para dados tabelados:
Determinação de Q_i:
1º passo: calcula-se a posição

$$\boxed{p = \frac{i \times \sum_{i=1}^{n} f_i}{4};}$$

2º passo: identifica-se a classe Q_i pela coluna das Frequências Acumuladas;
3º passo: Aplica-se a fórmula:
onde:

$$Q_i = l_{iQ_i} + \left[\frac{\frac{i \times \sum_{i=1}^{n} f_i}{4} - FAA}{f_{iQ_i}} \right] \times h$$

***I* =** **Ordem do quartil, *i* =1 ou 2 ou 3;**
l_{iQi} = **Limite inferior da classe do quartil de ordem *i*.**
***FAA* =** **Frequência acumulada anterior da classe do quartil de ordem *i*;**
f_{iQi} = **Frequência simples da classe do quartil de ordem *i*;**
***h* =** **Intervalo de classe.**

Exemplo: considerando a tabela abaixo, calcule o Q_1, Q_2 e Q_3.

Idades dos alunos da disciplina Estatística
no Instituto Datavox, ano de 2002.

IDADES	f_i
17 |-- 19	8
19 |-- 21	12
21 |-- 23	20
23 |-- 25	6
25 |-- 27	4
Soma	50

Fonte: dados hipotéticos

Resolução:

a) Calcule Q_1

1º passo: calcula-se a posição

$$p = \frac{i \times \sum_{i=1}^{n} f_i}{4} = \frac{1 \times 50}{4} = 12{,}5$$

2º passo: identifica-se a classe Q_i pela coluna das Frequências Acumuladas;

Comparamos o valor da posição *P* com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta Fac é maior ou igual a *P*?".**

IDADES	f_i	*Fac*	
17 |-- 19	8	8	**→ 8 é maior ou igual a 12,5? NÃO!**
19 |-- 21	12	20	**→ 20 é maior ou igual a 12,5? SIM! Então esta é classe do 1º Quartil (Q_1)**
21 |-- 23	20	40	
23 |-- 25	6	46	
25 |-- 27	4	50	
Soma	50		

3º passo: Aplica-se a fórmula:

$$Q_1 = 19 + \left[\frac{12{,}5 - 8}{12}\right] \times 2 = 19 + 0{,}75 = 19{,}75$$

Portanto, o 1º Quartil é igual a 19,75.

b) Calcule Q_2

1º passo: calcula-se a posição

$$p = \frac{i \times \sum_{i=1}^{n} f_i}{4} = \frac{2 \times 50}{4} = 25$$

2º passo: identifica-se a classe Q_i pela coluna das Frequências Acumuladas;

Comparamos o valor da posição P com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta *Fac* é maior ou igual a *P*?".**

IDADES	f_i	*Fac*	
17 |-- 19	8	8	**→ 8 é maior ou igual a 25? NÃO!**
19 |-- 21	12	20	**→ 20 é maior ou igual a 25? NÃO!**
21 |-- 23	20	40	**→ 40 é maior ou igual a 25? SIM! Então esta é classe do 2º Quartil (Q_2)**
23 |-- 25	6	46	
25 |-- 27	4	50	
Soma	50		

3º passo: Aplica-se a fórmula:

$$Q_2 = 21 + \left[\frac{25 - 20}{20}\right] \times 2 = 21 + 0{,}5 = 21{,}5$$

Portanto, o 2º Quartil é igual a 21,5.

c) Calcule Q_3

1º passo: calcula-se a posição

$$p = \frac{i \times \sum_{i=1}^{n} f_i}{4} = \frac{3 \times 50}{4} = 37{,}5$$

2º passo: identifica-se a classe Q_i pela coluna das Frequências Acumuladas;
Comparamos o valor da posição P com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta *Fac* é maior ou igual a *P*?".**

IDADES	f_i	*Fac*	
17 |-- 19	8	8	**→ 8 é maior ou igual a 37,5? NÃO!**
19 |-- 21	12	20	**→ 20 é maior ou igual a 37,5? NÃO!**
21 |-- 23	20	40	**→ 40 é maior ou igual a 37,5? SIM! Então esta é classe do 3º Quartil (*Q_3*)**
23 |-- 25	6	46	
25 |-- 27	4	50	
Soma	50		

3º passo: Aplica-se a fórmula:

$$Q_3 = 21 + \left[\frac{37{,}5 - 20}{20}\right] \times 2 = 21 + 1{,}75 = 22{,}75$$

Portanto, o 3º Quartil é igual a 22,75.

3.8.2 – <u>Decis</u> (D_i): são as medidas separatrizes que dividem a série em 10 partes iguais, e são representadas por $D_1, D_2, ..., D_9$. O quinto decil corresponde à mediana.

0%	--	10%	--	20%	--	30%	--	40%	--	50%	--	60%	--	70%	--	80%	--	90%	--	100%
		D_1		D_2		D_3		D_4		D_5		D_6		D_7		D_8		D_9		

a) Decis para dados brutos:
Vamos abordar este assunto por meio de dois exemplos:
Exemplo 1: calcule o D_8

Dados: 1,9 – 2,0 – 2,1 – 2,5 – 3,0 – 3,1– 3,3– **3,7– 6,1** – 7,7 → *n*=10

Posição do D_8= 0,8(*n*+1).
Então, a posição de D_8 é: 0,8 (10+1) = 8,8
Como o valor encontrado não é inteiro, temos que o 8º Decil (D_8) será a média aritmética entre o 8º e 9º elementos:

D_8=(3,7+6,1)/2=4,9

Portanto, o 8º Decil (D_8) é igual a **4,9.**
Exemplo 2: calcule o D_8

Dados: 0,9 – 1,0 – 1,7 – 2,9 –3,1– 5,3– 5,5– 12,2– **12,9 – 14,0** – 33,6 → *n*=11

Posição do D_8= 0,8(*n*+1).
Então, a posição de D_8 é: 0,8 (11+1) = 9,6;
Como o valor encontrado não é inteiro, temos que o 8º Decil (D_8) será a média aritmética entre o 9º e 10º elementos:

D_8=(12,9+14)/2=13,45

Portanto, o 8º Decil (D_8) é igual a **13,45**

b) Decis para dados tabelados:
A fórmula, neste caso, é semelhante à das separatrizes anteriores.
Determinação de D_i:

1º passo: calcula-se a posição

$$\boxed{p = \frac{i \times \sum_{i=1}^{n} f_i}{10}}$$

2º passo: identifica-se a classe D_i pela Frequência acumulada.

3º Passo: aplica-se a fórmula:

$$D_i = l_{iD_i} + \left[\frac{\frac{i \times \sum_{i=1}^{n} f_i}{10} - FAA}{f_{iD_i}} \right] \times h$$

i = **Ordem do decil, *i* =1; 2; 3; 4; 5; 6; 7; 8 ou 9;**
l_{iDi} = **Limite inferior da classe do decil de ordem *i*.**
FAA = **Frequência acumulada anterior da classe do decil de ordem *i*;**
f_{iDi} = **Frequência simples da classe do decil de ordem *i*;**
h = **Intervalo de classe.**

Exemplo: Considerando a tabela abaixo, calcule o D_3.

Idades dos alunos da disciplina Estatística
no Instituto Datavox, ano de 2002.

IDADES	f_i
17 |-- 19	8
19 |-- 21	12
21 |-- 23	20
23 |-- 25	6
25 |-- 27	4
Soma	50

Fonte: dados hipotéticos

Resolução:
1º passo: calcula-se a posição

$$\boxed{p = \frac{3 \times 50}{10} = \frac{150}{10} = 15}$$

2º passo: identifica-se a classe D_i pela Frequência acumulada. Comparamos o valor da posição P com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta *Fac* é maior ou igual a *P*?".**

IDADES	f_i	*Fac*	
17 |-- 19	8	8	→ **8 é maior ou igual a 15? NÃO!**
19 |-- 21	12	20	→ **20 é maior ou igual a 15? SIM! Então esta é classe do 3º Decil (D_3)**
21 |-- 23	20	40	
23 |-- 25	6	46	
25 |-- 27	4	50	
Soma	50		

3º Passo: aplica-se a fórmula:

$$D_3 = 19 + \left[\frac{15-8}{12} \right] \times 2 = 19 + \frac{14}{12} \cong 20{,}167$$

Portanto, o 3º Decil é aproximadamente 20,167.

3.8.3 – Percentis (P_i): são as medidas separatrizes que dividem a série em 100 partes iguais, e são representadas por P_1, P_2, ...,P_{99}. O quinquagésimo centil corresponde à mediana.

0%	--	1%	--	2%	--	3%	--	--	--	50%	--	--	--	97%	--	98%	--	99%	--	100%
		P_1		P_2		P_3				P_{50}				P_{97}		P_{98}		P_{99}		

a) Percentis para dados brutos:

Vamos abordar este assunto por meio de dois exemplos:

Exemplo 1: calcule o P_{30}

Dados: 1,9 – 2,0 – **2,1 – 2,5** – 3,0 – 3,1– 3,3– 3,7– 6,1 – 7,7 ⟶ n=10

Posição do P_{30}= 0,3(n+1).

Então, a posição de P_{30} é: 0,3 (10+1) =3,3

Como o valor encontrado não é inteiro, temos que o 30º Percentil (P_{30}) será a média aritmética entre o 3º e 4º elementos:

$$Q_1=(2,1+2,5)/2=2,3$$

Portanto, o 30º Percentil (P_{30}) é igual a **2,3**

Exemplo 2: calcule o P_{30}

Dados: 0,9 – 1,0 – **1,7 – 2,9** –3,1– 5,3– 5,5– 12,2– 12,9 – 14,0 – 33,6 ⟶ n=11

Posição do P_{30}= 0,3(n+1).

Então, a posição de P_{30} é: 0,3 (11+1) = 3,6;

Como o valor encontrado não é inteiro, temos que o 30º Percentil (P_{30}) será a média aritmética entre o 3º e 4º elementos:

$$P_{30}=(1,7+2,9)/2=2,3$$

Portanto, o 30º Percentil (P_{30}) é igual a **2,3.**

b) Percentis para dados tabelados:

A fórmula, neste caso, é semelhante à das separatrizes anteriores.

Determinação de P_i:

1º passo: calcula-se a posição

$$\boxed{p = \frac{i \times \sum_{i=1}^{n} f_i}{100}}.$$

2º passo: identifica-se a classe de P_i pela Frequência acumulada.

3º Passo: aplica-se a fórmula:

$$P_i = l_{iP_i} + \left[\frac{\frac{i \times \sum_{i=1}^{n} f_i}{100} - FAA}{f_{iP_i}} \right] \times h$$

***i* =** **Ordem do percentil, *i* = 1;2;3;...,97;98 e 99.**
***l_iPi* =** **Limite inferior da classe do percentil de ordem *i*.**
***FAA* =** **Frequência acumulada anterior da classe do percentil de ordem *i*;**
***f_iPi* =** **Frequência simples da classe do percentil de ordem *i*;**
***h* =** **Intervalo de classe.**

Exemplo: Considerando a tabela abaixo, calcule o C_{70}.

Idades dos alunos da disciplina Estatística
no Instituto Datavox, ano de 2002.

IDADES	f_i
17 |-- 19	8
19 |-- 21	12
21 |-- 23	20
23 |-- 25	6
25 |-- 27	4
Soma	50

Fonte: dados hipotéticos

Resolução:

1º passo: calcula-se a posição

$$\boxed{p = \frac{70 \times 50}{100} = \frac{3500}{100} = 35}$$

2º passo: identifica-se a classe D_i pela Frequência acumulada. Comparamos o valor da posição P com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta *Fac* é maior ou igual a *P*?".**

IDADES	f_i	*Fac*	
17 |-- 19	8	8	**→ 8 é maior ou igual a 35? NÃO!**
19 |-- 21	12	20	**→ 20 é maior ou igual a 35? NÃO!**
21 |-- 23	**20**	**40**	**→ 40 é maior ou igual a 35? SIM!Então esta é classe do 70º Centil (C_{70})**
23 |-- 25	6	46	
25 |-- 27	4	50	
Soma	50		

3º Passo: aplica-se a fórmula:

$$\boxed{P_i = 21 + \left[\frac{35-20}{20}\right] \times 2 = 21 + 1,5 = 22,5}$$

Portanto, o 70º Percentil é igual a 22,5.

ATIVIDADES

1) Foi organizado um churrasco para comemorar a conclusão do Curso de Engenharia Mecânica. Foram compradas as seguintes carnes aos respectivos preços:

10 Kg de filé mignon	R$ 12,00 o Kg
20 Kg de linguiça	R$ 7,00 o Kg
10 Kg de picanha	R$ 16,00 o Kg

Qual o valor médio do Kg de carne adquirida

2) Uma escola possui 18 professores. Um deles aposenta-se e é substituído por um professor de 22 anos. Com isso, a média das idades dos professores diminui 2 anos. Qual a idade do professor que se aposentou?

3) Os dados a seguir foram obtidos em indivíduos contaminados pelo veneno de um certo tipo de inseto e submetidos a tratamento. A variável de interesse Recup é definida como o tempo (em horas) entre a administração do tratamento e a recuperação do indivíduo. Os valores de Recup são os seguintes: 3, 90, 23, 46, 2, 42, 47, 37, 12, 51, 11, 1, 3, 3, 45, 3, 4, 11, 2, 8, 56, 39, 22, 16, 5 e 52.
   a) Construa a Tabela de frequências para a variável Recup.
   b) Obtenha a média.
   c) O primeiro quartil
   d) O terceiro quartil

4) Em um time de futebol, o jogador mais velho entre os onzes titulares foi substituído por um jogador de 16 anos. Isto fez com que a média de idade dos 11 jogadores diminuísse 2 anos. Calcule a idade do jogador mais velho que foi substituído.

5) Durante um jogo de futebol entre Vasco e Flamengo foi feita uma pesquisa de idades das duas torcidas. Constatou-se que a idade média da torcida em geral era 27 anos (independente da preferência). Qual a idade média dos torcedores do Flamengo, sabendo-se que se constituem 60% da torcida presente no estádio e que os torcedores do Vasco têm em média 30 anos?

6) 65% dos alunos de uma escola para adultos têm média 20 anos. Considerando que 15% tem em média 30 anos. A média geral é de 27,5 anos. Qual a média dos demais alunos?

7) O salário pago aos funcionários de uma empresa X é de R$ 710,00. Os salários médios pagos aos funcionários especializados e não especializados correspondem respectivamente a R$ 800,00 e R$ 500,00. Pede-se, determinar o percentual de empregados especializados e não especializados da empresa.

8) Em uma classe de 50 alunos, as notas obtidas formaram a seguinte distribuição:

Notas	2	3	4	5	6	7	8	9	10
N^o de alunos	1	3	6	10	13	8	5	3	1

   Calcule:
   a) A nota média.
   b) A nota mediana.
   c) A nota modal.

9) **(AFTN-98)** Os dados seguintes, ordenados do menor para o maior, foram obtidos de uma amostra aleatória, de 50 preços (Xi) de ações, tomada numa bolsa de valores internacional. A unidade monetária é o dólar americano. 4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 12, 12, 13, 13,14, 15, 15, 15, 16, 16, 18, 23. Obtenha: a média, a mediana e a moda.

10) Considere os valores dos pesos de 32 alunos de uma classe apresentados abaixo:

64 68 63 67 65 64 67 64 66 67 70 67 67 66 69 66
70 62 71 64 69 65 71 66 63 70 68 69 71 68 68 68

Determine:
a) a média ;
b) a mediana;
c) a moda.

11) Dos salários de um grupo de 4000 funcionários da empresa são conhecidos os seguintes parâmetros. C_{95}= R$ 3.600,00; Md= R$ 2.200,00; D_3 = R$ 2.000,00; Q_1= R$ 1600,00.
Responda
a) O n° de funcionários que recebem entre 1600 a 2.200 reais;
b) O n° de funcionários que recebem mais R$ 3.600,00;
c) O n° de funcionários que recebem entre 2.000 e 2.200 reais;
d) O percentual de funcionários que recebem entre 1600 e 2.200 reais

12) O preço médio (aritmético) de produto químico produzido por uma empresa é igual a 50 reais, o preço geométrico é de 40 reais. Qual o preço médio na forma harmônica?

13) Um jogo completo de xadrez é composto das seguintes peças 08 peões, 02 torres, 02 cavalos, 01 dama, e 01 rei. Atribuem-se os seguintes valores comparativos para as peças rei =24 pontos, dama=10 pontos, cavalo=03 pontos, torres=05 pontos, e peão=01 ponto. Se neste sistema 04 ponto é valor médio de uma peça, levando em conta o jogo completo. Qual o valor comparativo da peça bispo?

14) Dada a distribuição de frequências dada abaixo:

Idades dos alunos da disciplina Estatística no Instituto Datavox, ano de 2002.

IDADES	f_i
17 |-- 19	8
19 |-- 21	12
21 |-- 23	20
23 |-- 25	6
25 |-- 27	4
Soma	50

Fonte: dados hipotéticos

Calcule:
a) a média;
b) a moda
c) a mediana

15) Em certa empresa trabalham 4 analistas de mercado, 2 supervisores, 1 chefe de seção e 1 gerente que ganham, respectivamente: R$ 1.300,00; R$ 1.600,00; R$ 2.750,00, R$ 5.000,00. Qual o valor do salário médio desses funcionários?

16) Uma distribuidora de refrigerantes fez um levantamento sobre o consumo semanal (em litros) por pessoa, em jan/2002, em uma cidade do litoral, obtendo a tabela abaixo:

Consumo	Nº de Pessoas

0,0 |-- 0,5	10
0,5 |-- 1,0	25
1,0 |-- 1,5	9
1,5 |-- 2,0	7
2,0 |-- 2,5	6

a) Determine e interprete o consumo médio.
b) Qual o percentual de pessoas que consomem menos de 1 litro por semana?
c) Determine os intervalos que contém o consumo modal e o consumo mediano.

17) A poluição causada por óleo em mares e oceanos estimula o crescimento de certos tipos de bactérias. Uma contagem de microorganismos presentes no petróleo (número de bactérias por 100 mililitros), em 10 porções de água do mar, indicou as seguintes medidas:

49 70 54 67 59 40 71 67 67 52

Determine a média, a mediana e a moda.

18) Abaixo temos as notas de 15 alunos na 1^a avaliação da disciplina Análise Real:

7,0 –7,5 – 5,3 – 6,8 – 5,5 – 4,2 – 8,0 – 7,0 – 7,5 – 6,5 – 5,9 – 8,0 – 2,5 – 5,7 –5,0

Determine:
a) o 1^o quartil;
b) o 8^o decil;
c) 90^o percentil;

19) O preço geométrico de produto F produzido por uma empresa é igual a 9 reais, o preço harmônico é de 3 reais. Qual o preço médio na forma aritmética?

20) **(FISCAL DE TRIBUTOS DE MG-96)** A estatura média dos sócios de um clube é165 cm, sendo a dos homens 172 cm e a das mulheres 162 cm. Qual a porcentagem de mulheres no clube?

21) **(Auditor do Tesouro Municipal - Recife – 2003)** Em uma amostra, realizada para se obter informação sobre a distribuição salarial de homens e mulheres, encontrou-se que o salário médio vale R$ 1.200,00. O salário médio observado para os homens foi de R$ 1.300,00 e para as mulheres foi de R$ 1.100,00. Assinale a opção correta e mostre os cálculos.
a) O número de homens na amostra é igual ao de mulheres.
b) O número de homens na amostra é o dobro do de mulheres.
c) O número de homens na amostra é o triplo do de mulheres.
d) O número de mulheres é o dobro do número de homens.
e) O número de mulheres é o quádruplo do número de homens

PESQUISANDO

Para você aprofundar os conteúdos apresentados nesta unidade, sugiro que pesquise:
- Os seguintes livros:

COSTA NETO, Pedro Luiz de Oliveira. *Estatística. 12. ed*. São Paulo: Editora Edgard Blücher Ltda., 1992.
BUSSAB, W. O. , MORETTIN, P.A, **Estatística Básica** 5ª ed. São Paulo: SARAIVA, 2002.
FONSECA, J. S. **Curso de Estatística**. São Paulo: Atlas, 1980.

- acesse os sites:

www.ime.usp.br/~mae116

http://alea-estp.ine.pt

http://www.somatematica.com.br/estatistica.php

http://www.aleixomkt.com.br/fisio/62.pdf

http://www.heliorocha.com.br/graduacao/turismo/download/EAT/MedidasDePosicao.pdf

UNIDADE 4 – MEDIDAS DE DISPERSÃO

Nesta unidade estudaremos as medidas de dispersão que tem um papel importantíssimo na análise dos dados, pois avaliam a variabilidade em torno da média. Verificaremos também que as medidas de dispersão avaliam a representatividade das medidas de posição. Também estudaremos as medidas de Assimetria e Curtose. Na Assimetria estudaremos o grau de afastamento de uma distribuição da unidade de simetria, enquanto que na Curtose estudaremos o grau de achatamento ou afunilamento de uma distribuição.

Caro(a) aluno(a) eis o s itens que estudaremos nesta unidade

4.1 Amplitude
4.2 Desvio médio
4.3 Variância
4.4 Desvio padrão
4.5 Coeficiente de variação
4.6 Assimetria
4.8 Curtose

4.1 – AMPLITUDE TOTAL

Um modo mais simples de se ter uma indicação da dispersão dos valores de uma amostra ou população é comparar o valor máximo com o mínimo. Entretanto a Amplitude Total não nos fornece qualquer indicação do que ocorre no interior do conjunto.

AT = Valor máximo – Valor mínimo

Temos três situações para obtenção da Amplitude Total:

a) **Para valores brutos:**
Por exemplo, notas em matemática em cinco avaliações, 7; 7,5; 8; 8,5; 10.
A amplitude total será: **AT= 10 – 7 = 3.**

b) **Para distribuição de frequências sem classe (sem intervalo de classe)**

xi	fi
0	2
1	6
3	5
4	3

A amplitude total será: **AT = 4 - 0 = 4**

c) Para distribuição de frequências com classe (com intervalo de classe).

Neste caso a Amplitude Total será a diferença entre o limite superior da última classe (LS) e o limite inferior da primeira classe (LI). **AT = LS – LI**

Classes	*fi*
6 |– 8	6
8 |– 10	2
10|– **12**	3

A amplitude total será: **AT = 12 - 6 = 6**

4.2 – DESVIO MÉDIO

É a medida de dispersão ou o grau de concentração dos valores em torno da média. Quando estamos calculando o desvio médio estamos medindo a dispersão entre cada x_i e a média $\bar{x}$. Temos dois tipos de Desvio médio:

4.2.1 Desvio Médio para dados brutos

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

onde:

$\bar{X}$ **– é a média aritmética;**

x_i **– são os valores das observações;**

$|x_i - \bar{X}|$ **– é o valor absoluto do desvio de** x_i **em relação à** $\bar{X}$ **.**

n **– número de observações**.

Exemplo: Calcular o Desvio médio para os dados abaixo:

3 – 5 – 7 – 4 – 6.

Primeiro calcula-se a média

$$\bar{X} = \frac{(3+5+7+4+6)}{5} = 5$$

Segundo monta-se a tabela:

Calculamos as diferenças entre os valores da série e a média

Aplicamos o módulo nas diferenças.

X_i	$\bar{X}$	$(X_i - \bar{X})$	$\lvert X_i - \bar{X} \rvert$
3	5	-2	2
5	5	0	0
7	5	2	2
4	5	-1	1
6	5	1	1
Total	**-**		**6**

Aplica-se a fórmula:

$$DM = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{X} \rvert}{n} = \frac{2+0+2+1+1}{5} = \frac{6}{5} = 1,2$$

Logo, o Desvio médio é igual a 1,2.

4.2.2 – Desvio Médio para dados tabelados

$$DM = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{X} \rvert \; f_i}{n}$$

onde:

$\bar{X}$ **– é a média aritmética;**

x_i **– são os valores das observações;**

f_i **– frequência simples**

$\lvert x_i - \bar{x} \rvert$ **– é o valor absoluto do desvio de** x_i **em relação à** $\bar{x}$ **.**

n **– número de observações.**

Exemplo: Calcular o Desvio médio da distribuição de frequências abaixo:

Notas	*Fi*	*xi*
0 |– 2	3	1
2 |– 4	3	3
4 |– 6	8	5
6 |– 8	3	7
8 |– 10	8	9
Total	**25**	

Primeiro calcula-se a média

$$\bar{X} = \frac{3\times1 + 3\times3 + 8\times5 + 3\times7 + 8\times9}{25} = \frac{145}{25} = 5{,}8$$

Calculamos as diferenças entre os pontos médios e a média

Aplicamos o módulo nas diferenças.

Segundo monta-se a tabela:

Multiplicamos as diferenças pelas freqüências.

Notas	*fi*	*Xi*	$\overline{X}$	(*Xi*- $\overline{X}$)	|*Xi*- $\overline{X}$ |	|*Xi*- $\overline{X}$ | *fi*
0 |–2	3	1	5,8	-4,8	4,8	14,4
2 |–4	3	3	5,8	-2,8	2,8	8,4
4 |–6	8	5	5,8	-0,8	0,8	6,4
6 |–8	3	7	5,8	1,2	1,2	3,6
8 |–10	8	9	5,8	3,2	3,2	25,6
Total	**25**					**58,4**

Agora, aplica-se a fórmula:

$$DM = \frac{\sum_{i=1}^{n} |x_i - \overline{X}|\ f_i}{n} = \frac{58{,}4}{25} = 2{,}336 \cong 2{,}34$$

Logo, o Desvio médio é aproximadamente 2,34.

4.3 – VARIÂNCIA:

É a média quadrática das somas dos desvios em relação à média aritmética. É uma medida de dispersão bastante estudada no meio cientifico. Quando o estudo for feito na amostra a variância é simbolizada por: S^2. E quando estudamos a variância de uma população, o símbolo usado é σ^2.

4.3.1 – Fórmula para dados brutos

Processo longo

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

Processo breve ou simplificado

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{X}^2}{n-1}$$

Observação: Quando se tratar de população, dividi-se apenas por "*n*".
Exemplo: Calcular a variância para os dados abaixo:

3 – 5 – 7 – 4 – 6.

Primeiro calcula-se a média

$$\overline{X} = \frac{(3+5+7+4+6)}{5} = 5$$

Segundo monta-se a tabela:

Calculamos as diferenças entre os valores da série e a média

Elevamos ao quadrado as diferenças.

X_i	$\overline{X}$	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
3	5	-2	4
5	5	0	0
7	5	2	4
4	5	-1	1
6	5	1	1
Total	**-**		**10**

Aplica-se a fórmula:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} = \frac{10}{5-1} = \frac{10}{4} = 2{,}5$$

Logo, a Variância é igual a 2,5.
Vamos calcular a Variância do exemplo anterior pelo **Processo Breve.**
Os dados são: 3 – 5 – 7 – 4 – 6.
Qual o valor da Variância?

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{X}^2}{n-1} = \frac{(3^2+5^2+7^2+4^2+6^2)-(5\times 5^2)}{5-1} = \frac{135-125}{4} = \frac{10}{4} = 2{,}5$$

Logo, a Variância é igual a 2,5.

4.3.2 - Fórmula para dados tabelados:

Processo longo

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

Processo breve ou simplificado

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i - n\bar{X}^2}{n-1}$$

Observação: Quando se trata de população dividimos apenas por "*n*".
Exemplo: Calcular a Variância da distribuição de frequências abaixo:

Notas	*Fi*	*xi*
0|–2	3	1
2|–4	3	**3**
4|–6	8	5
6|–8	3	**7**
8|–10	8	9
Total	**25**	

Primeiro calcula-se a média

$$\bar{X} = \frac{3\times1 + 3\times3 + 8\times5 + 3\times7 + 8\times9}{25} = \frac{145}{25} = 5{,}8$$

Segundo monta-se a tabela:

Calculamos as diferenças entre os pontos médios e a média

Elevamos ao quadrado as diferenças.

Multiplicamos os quadrados das diferenças pelas *freqüências*.

Notas	*fi*	*Xi*	$\bar{X}$	*(Xi-* $\bar{X}$ *)*	*(Xi-* $\bar{X}$ *)*^{2}	*(Xi-* $\bar{X}$ *)*^{2} *fi*
0|–2	3	1	5,8	-4,8	23,04	69,12
2|–4	3	3	5,8	-2,8	7,84	23,52
4|–6	8	5	5,8	-0,8	0,64	5,12
6|–8	3	7	5,8	1,2	1,44	4,32
8|–10	8	9	5,8	3,2	10,24	81,92
Total	**25**	**-**	**-**	**-**	**-**	**184,00**

Aplica-se a fórmula:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2 . f_i}{n-1} = \frac{184}{25-1} = \frac{184}{24} \cong 7{,}67$$

Logo, a Variância é aproximadamente igual 7,67.
Agora calcularemos a Variância do exemplo anterior pelo processo breve:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i - n\bar{X}^2}{n-1} = \frac{(1^2 \times 3 + 3^2 \times 3 + 5^2 \times 8 + 7^2 \times 3 + 9^2 \times 8) - (25 \times 5{,}8^2)}{25-1} = \frac{1025 - 841}{24} = \frac{184}{24} \cong 7{,}67$$

Logo, a Variância é aproximadamente igual 7,67.

4.3.3 – Propriedades da variância:

A variância absoluta de uma constante é igual a zero;

Somando-se ou diminuindo-se a todos os valores da série um valor constante $K \neq 0$, a nova variância será igual à anterior, isto é, não se altera.

Multiplicando-se ou dividindo-se todos os valores de uma série por um valor constante, $K \neq 0$, a nova variância calculada será igual à variância absoluta original multiplicada ou dividida pelo quadrado da constante utilizada.

4.4 – DESVIO PADRÃO:

É a raiz quadrada da variância. É a medida mais informativa da variação dos dados. O Desvio Padrão nos fornece uma indicação do que ocorre entre os dois extremos. Portanto, o Desvio Padrão é a medida de quanto os valores observados variam em torno da média.

O Desvio Padrão amostral é dado por:

$$\boxed{S = \sqrt{S^2}}$$

O desvio padrão populacional é dado por:

$$\boxed{\sigma = \sqrt{\sigma^2}}$$

Exemplo 1: Calcule o Desvio padrão dos dados brutos abaixo: 3 – 5 – 7 – 4 – 6.
Como já vimos anteriormente o cálculo da variância para os dados acima, basta extrair a raiz quadrada da Variância:
Como a Variância foi igual a 2,5, então:

$$\boxed{S = \sqrt{S^2} = \sqrt{2{,}5} = 1{,}58}$$

Então, o Desvio padrão é igual a 1,58.
Exemplo 2: Calcular o Desvio padrão da distribuição de frequências abaixo:

Notas	*Fi*	*xi*
0|–2	3	1
2|–4	3	**3**
4|–6	8	5
6|–8	3	**7**
8|–10	8	9
Total	**25**	

Como já vimos anteriormente o cálculo da variância para os dados acima, basta extrair a raiz quadrada da Variância:
Como a Variância foi aproximadamente igual a 7,67, então:

$$\boxed{S = \sqrt{S^2} = \sqrt{7{,}67} = 2{,}77}$$

Então, o Desvio padrão é igual a 2,77.
Resumindo: Para o cálculo do desvio padrão, deve-se primeiramente determinar o valor da variância e, em seguida, extrair a raiz quadrada desse resultado.

4.5 – COEFICIENTE DE VARIAÇÃO: é o valor positivo da raiz quadrada da variância relativa.

SÍMBOLO: $\boxed{CV = \dfrac{S}{\overline{X}} \times 100}$

onde:

S: Desvio padrão;
$\overline{X}$ *: Média*

Observação 1: Será considerada a série mais homogênea, aquela que apresentar menor valor do coeficiente de variabilidade.

Observação 2: É uma medida estatística que serve para avaliar a homogeneidade de séries estatísticas, que é o grau de concentração dos valores observados em torno da sua média aritmética.

Observação 3: O seu valor numérico pode ser expresso em percentual.

MEDIDAS DE ASSIMETRIA E CURTOSE
Agora trataremos que são as medidas de assimetria e curtose. Elas referem-se à forma da curva de uma distribuição de frequência, mais especificamente do polígono de frequência ou do histograma.

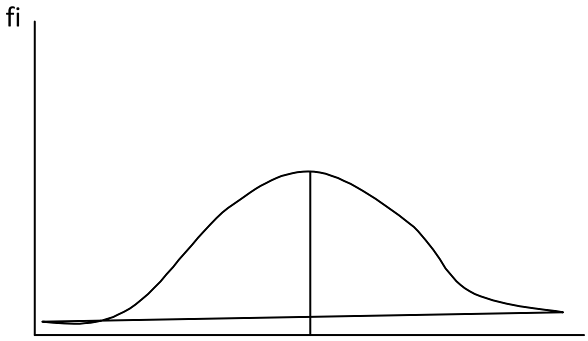
4.6 – ASSIMETRIA

E o que são as medidas de assimetria?
É o grau de afastamento de uma distribuição da unidade de simetria

Destacamos que as distribuições de frequências não diferem apenas quanto ao valor médio e a variabilidade, como também quanto a sua forma.

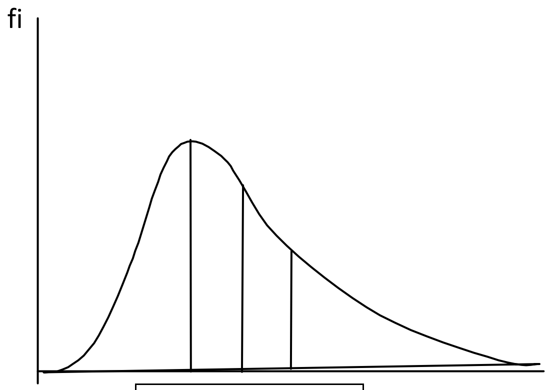
4.6.1 – Tipos de curvas:

a) Simétrica: Quando a média, mediana e a moda são iguais, isto é, apresentam o mesmo valor.



$\overline{X}$ = *Me* = *Mo*

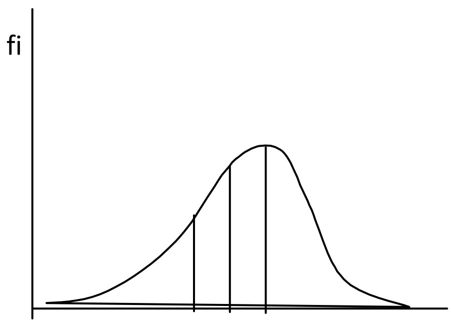
b) Assimetria positiva: Neste caso, a média aritmética apresentará um valor maior do que a mediana, e esta, por sua vez, apresentará um valor maior do que a moda.



Mo < Me < $\overline{X}$

A cauda é mais alongada à direita da ordenada máxima (ordenada correspondente à moda). Nas distribuições assimétricas à direita, há uma predominância de valores superiores a moda.

c) Assimetria negativa: Neste caso a média aritmética será menor do que a mediana, e esta, por sua vez, é menor do que a moda.



$\overline{X}$ < Me < Mo

A cauda é mais alongada à esquerda da ordenada máxima. Nas distribuições assimétricas negativas, predominam valores inferiores à moda.

4.7 – COEFICIENTES DE ASSIMETRIA

Há vários coeficientes de assimetria na literatura, entretanto utilizaremos os de Pearson, por serem mais conhecidos. Temos dois tipos de coeficientes que são dados abaixo:

4.7.1 – Primeiro coeficiente de assimetria de Pearson:

$$\boxed{A = \frac{(\bar{x} - Mo\ )}{S}}$$

4.7.2 – Segundo coeficiente de assimetria de Pearson

$$\boxed{A = \frac{3(\bar{x} - Me\ )}{S}}$$

onde:

A - Coeficiente de Assimetria;
$\bar{x}$ - Média aritmética;
Mo - Moda;
S - Desvio-padrão.

O coeficiente de assimetria, A, pode ser classificado:

se $A = 0 \Rightarrow$ Simétrica
se $A > 0 \Rightarrow$ Assimetria Positiva
se $A < 0 \Rightarrow$ Assimetria Negativa

Exemplo 1: Calcular o coeficiente de assimetria para os dados da série abaixo:

3 – 8– 7 – 4 – 3.

a) Iniciamos obtendo a Moda:
   Neste caso **3** é a Moda.

b) Calculamos a Média.

$$\boxed{\bar{X} = \frac{(3+8+7+4+3)}{5} = \frac{25}{5} = 5}$$

c) Obter o Desvio padrão, antes, porém precisamos calcular a Variância:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{(3^2+8^2+7^2+4^2+3^2)-(5\times 5^2)}{5-1} = \frac{147-125}{4} = \frac{22}{4} = 5{,}5}$$

Então, o Desvio padrão será:

$$S = \sqrt{S^2} = \sqrt{5,5} \cong 2,35$$

Agora, aplicamos a fórmula:

$$A = \frac{(\bar{x} - Mo\ )}{S} = \frac{5-3}{2,35} \cong 0,85$$

Portanto, o coeficiente de Assimetria é igual a 0,85. A distribuição é Assimétrica Positiva.

Exemplo 2: Calcular o coeficiente de Assimetria da distribuição de frequências abaixo:

Notas	*Fi*	*xi*
0|–2	3	1
2|–4	3	**3**
4|–6	8	5
6|–8	3	**7**
8|–10	8	9
Total	**25**	

a) Calcula-se a Mediana
1º passo: calcula-se a posição

$$p = \frac{\sum_{i=1}^{n} f_i}{2} = \frac{25}{2} = 12,5$$

2º passo: identifica-se a classe *Mediana*, pela coluna das Frequências Acumuladas;
Comparamos o valor da posição *P* com os valores da *Fac*, iniciando da *Fac* da primeira classe e fazendo a seguinte pergunta: **"esta Fac é maior ou igual a *P*?".**

Notas	*fi*	*Fac*	
0|–2	3	1	→**1** é maior ou igual a **12,5**? NÃO!
2|–4	3	**4**	→**4** é maior ou igual a **12,5**? NÃO!
4|–6	8	9	→**9** é maior ou igual a **12,5**? NÃO!
6|–8	**3**	**16**	→**16** é maior ou igual a **12,5**? SIM! **Então esta é classe da *Mediana*.**
8|–10	8	25	
Total	**25**		

3º passo: Aplica-se a fórmula:

$$Me = 6 + \left[\frac{12,5-9}{3}\right] \times 2 = 6 + 2,33 \cong 8,33$$

Logo, a Mediana é aproximadamente 8,33.

b) Calcula-se a Média.

$$\bar{X} = \frac{3\times1 + 3\times3 + 8\times5 + 3\times7 + 8\times9}{25} = \frac{145}{25} = 5,8$$

c) Obter o Desvio padrão, antes, porém precisamos calcular a Variância:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i - n\overline{X}^2}{n-1} = \frac{(1^2 \times 3 + 3^2 \times 3 + 5^2 \times 8 + 7^2 \times 3 + 9^2 \times 8) - (25 \times 5{,}8^2)}{25-1} = \frac{1025 - 841}{24} = \frac{184}{24} \cong 7{,}67$$

Logo, a Variância é aproximadamente igual 7,67.

Então, o Desvio padrão será:

$$S = \sqrt{S^2} = \sqrt{7{,}67} \cong 2{,}77$$

Agora, aplicamos a fórmula:

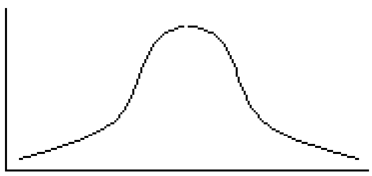
$$A = \frac{3(\overline{x} - Me)}{S} = \frac{3(5{,}8 - 8{,}33)}{2{,}77} \cong -2{,}74$$

Portanto, o coeficiente de Assimetria é igual a -2,74. A distribuição é Assimétrica Negativa.

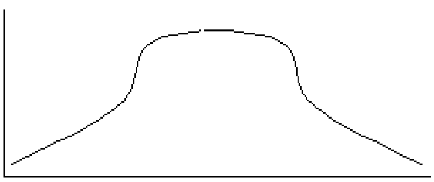
4.8 – CURTOSE

O que é curtose?
É o estudo do grau de achatamento de uma distribuição, considerado em relação a uma distribuição normal. De acordo com o grau de curtose, podemos ter três tipos de curvas de frequência.

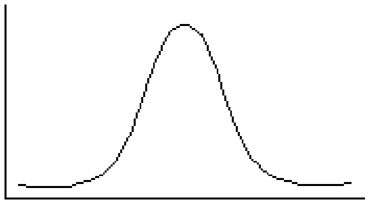
4.8.1 – **Curva Mesocúrtica:** Quando a curva de frequências apresenta um grau de achatamento equivalente ou da curva normal.



4.8.2 – **Curva Platicúrtica:** Quando uma curva de frequências apresenta um alto grau de achatamento, superior ao da normal.



4.8.3 – **Curva Leptocúrtica:** Quando uma curva de frequências apresenta um alto grau de afilamento, superior ao da normal.



Para avaliar o grau de curtose de uma curva de frequência, usaremos o coeficiente percentílico de curtose:

$$K = \frac{D_q}{(D_9 - D_1)} = \left(\frac{Q_3 - Q_1}{2.(C_{90} - C_{10})} \right)$$

onde: **K = coeficiente percentílico de curtose**

D_q = Desvio quartílico = $(Q_3 - Q_1) / 2$

D_9 = 9º decil

D_1 = 1º decil

C_{90} = 90º Centil

C_{10} = 10º Centil

k = 0,263 $\Rightarrow$ Curva ou distribuição mesocúrtica
k > 0,263 $\Rightarrow$ Curva ou distribuição platicúrtica
k < 0,263 $\Rightarrow$ Curva ou distribuição leptocúrtica

Exemplo: determinar o coeficiente percentilico de curtose do resumo estatístico dado abaixo:
Q_1=45;
Q_3=95;
D_1=15;
D_9=105;
Então, aplica-se a formula:

$$K == \left(\frac{Q_3 - Q_1}{2.(C_{90} - C_{10})} \right) = \frac{(95 - 45)}{2(105 - 15)} = \frac{50}{2(90)} = \frac{50}{180} \cong 0{,}277$$

Então, podemos classificar o grau de curtose como Platicurtica, pois, k > 0,263.

ATIVIDADES

1) Dada a tabela abaixo, Pede-se: A Variância, o Desvio médio, o Desvio Padrão e o Coeficiente de variação.

NOTAS	f_i
1 |-- 3	5
3 |-- 5	7
5 |-- 7	8
Σ	20

Fonte: dados hipotéticos

2) Quinze pacientes de uma clínica de ortopedia foram avaliados quanto ao número de meses previstos de fisioterapia, se haverá (S) ou não (N) sequelas após o tratamento e o grau de

complexidade da cirurgia realizada: alto (A), médio(M) ou baixo(B). Os dados são apresentados na tabela abaixo

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fisioterapia	7	8	5	6	4	5	7	7	6	8	6	5	5	4	5
Sequelas	S	S	N	N	N	S	S	N	N	S	S	N	S	N	N
Cirurgia	A	M	A	M	M	B	A	M	B	M	B	B	M	M	A

a) Classifique cada uma das variáveis;
b) Divida os pacientes em dois grupos: com (S) e sem (N) sequelas. Compare os grupos em relação ao tempo médio de fisioterapia.
c) Qual dos dois grupos é mais homogêneo (menor dispersão) em relação ao tempo de fisioterapia?

3) São apresentados abaixo o diâmetro (em polegadas), a altura (em pés) e o volume (em pés cúbicos) de uma amostra de 10 cerejeiras:

Diâmetro	8,3	10,5	10,8	11,1	12,0	13,3	14,5	16,3	17,5	18,0
Altura	70	72	83	80	75	86	74	77	82	80
Volume	10,3	16,4	19,7	22,6	19,1	27,4	38,3	42,6	55,7	51,5

Qual das três variáveis tem maior variabilidade?

4) **(AFC-94)** Entre os funcionários de um órgão do governo, foi retirada uma amostra de dez indivíduos. Os números que representam as ausências ao trabalho registradas para cada um deles, no último ano, são: 0, 0, 0, 2, 2, 2, 4, 4, 6 e 10. Sendo assim, qual o valor do desvio padrão desta amostra.
5) Um experimento é conduzido para comparar dois regimes alimentares no que diz respeito ao aumento de peso. Vinte indivíduos são distribuídos ao acaso entre dois grupos em que ao primeiro deles foi dada a dieta A e ao segundo a dieta B. Decorrido certo período verifica-se que o ganho de peso em Kg para os indivíduos da amostra foram os seguintes

A	-1,0	0,0	2,1	3,1	3,3	4,3	5,0	5,2	5,5	6,8
B	2,5	3,0	4,0	5,7	6,0	6,9	7,0	7,2	7,3	8,1

a) Calcule a média, mediana e desvio padrão da variável ganho de peso para cada dieta.
b) Obtenha o coeficiente de variação e comente o resultado.

6) Em um ensaio para o estudo da distribuição de um atributo financeiro (X), foram examinados 200 itens de natureza contábil do balanço de uma empresa. Esse exercício produziu a tabela de frequência abaixo. A coluna *Classes* representa intervalos de valores de X em reais e a coluna P representa a frequência relativa acumulada. Não existem observações coincidentes com os extremos das classes

Classes	P(%)
70 – 90	5
90 – 110	15
110 – 130	40
130 – 150	70
150 – 170	85
170 – 190	95
190 – 210	100

Calcule:

a) Coeficiente de assimetria;
b) Classifique o coeficiente de assimetria;
c) Coeficiente de curtose;
d) Classifique o coeficiente de curtose.

7) O atributo do tipo contínuo X, observado como um inteiro, numa amostra de tamanho 100 obtida de uma população de 1000 indivíduos, produziu a tabela de frequências seguinte:

Classes	Frequência (fi)
29,5 – 39,5	4
39,5 – 49,5	8
49,5 – 59,5	14
59,5 – 69,5	20
69,5 – 79,5	26
79,5 – 89,5	18
89,5 – 99,5	10

Calcule:

a) Coeficiente de Assimetria.
b) Coeficiente de Curtose.

8) Obtenha a variância, o Desvio Padrão e o Coeficiente de variação da *distribuição de frequências* dada abaixo:

Produção de 100 empregados da empresa Schalcher, Belém/PA, 2004.

Produção	**Frequência**
30 - 39	4
40 - 49	14
50 - 59	29
60 - 69	35
70 - 79	18
Total	**100**

Fonte: fictícia.

9) Um estudo foi realizado por um professor em três turmas, obtendo a média e o desvio padrão das notas de sua disciplina, conforme abaixo. Qual a turma com menor variabilidade? Justifique adequadamente.

TURMA	A	B	C
MÉDIA	6,5	8,0	8,0
DESVIO PADRÃO	2,2	1,7	2,0

10) O Desvio Padrão pode ser zero? O que isso significa?

11) Na empresa Mercury Ltda. Foi observada a distribuição de funcionários do setor de serviços gerais com relação ao salário semanal, conforme mostra a distribuição de frequência

Salário Semanal (em US$)	fi
25|--30	10
30|--35	20
35|--40	30
40|--45	15
45|--50	40
50|--55	35
Total	150

Calcule:
a) O Desvio médio;
b) O Desvio padrão.

12) Classifique corretamente os coeficientes de Assimetria e Curtose dados abaixo:
a) A= 2,45 e K= -0,367;
b) A=-2,45 e K= 0,123;
c) A= 0,35 e K= 0,267;
d) A= -0,45 e K= -0,256;
e) A= -1,77 e k= 0,479.

12) Um banco tem à disposição dos seus clientes duas zonas de atendimento e cada uma com duas máquinas multibanco. Na zona Z_1, os clientes formam fila única e na zona Z_2 fazem duas filas separadas, uma para cada máquina. Registaram-se os seguintes tempos de espera de 10 clientes

Z_1	4,8	4,8	4,9	5,1	5,4	5,5	5,7	5,8	5,8	5,8
Z_2	2,0	3,5	4,1	4,5	5,1	5,8	5,8	5,8	8,4	8,6

a) Com base nestes dados, que conselho daria ao banco quanto ao método a usar, uma fila única ou filas separadas?
b) Calcule o desvio padrão e o coeficiente de variação para Z_1 e Z_2.

13) Como parte de uma avaliação médica em uma certa universidade, foi medida a frequência cardíaca dos alunos do primeiro ano. Os dados serão apresentados em seguida.
a) Obtenha a Amplitude Total.
b) Obtenha a média.
c) Qual o Desvio médio?
d) Encontre a Variância?
e) Qual o valor do Desvio padrão?
f) Calcule o coeficiente de variação.

Frequência cardíaca	Frequência
60|-- 65	11
65|-- 70	35
70|-- 75	68
75|-- 80	20
80|-- 85	12
85|-- 90	10
90|-- 95	1
95|-- 100	3

14) Alunos da Escola de Educação Física foram submetidos a um treinamento de resistência por um período de 2 meses. Antes de iniciarem o treinamento, foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiram correr sem parar. Depois de 4 meses de treinamento, foram novamente submetidos ao mesmo teste. Os dados estão apresentados a seguir.

Faixas	Frequências	
	Antes do treinamento	Depois do treinamento
0|-- 2	442	80
2|-- 4	211	205
4|-- 8	128	297
8|--12	25	184
12|--16	11	45
16|-- 22	3	9

a) Calcule o Desvio médio para ambos os grupos.
b) Obtenha a Variância para ambos os grupos.
c) Obtenha o Desvio padrão Para ambos os grupos
d) Qual o grupo mais homogêneo?

15) Um laboratório clínico precisa escolher, dentre três aparelhos (A, B, C) para dosagem de sangue, qual deverá comprar. Para isto o responsável pelas análises preparou uma substância de concentração conhecida (10 mg/ml) e extraiu várias amostras para serem dosadas pelos três aparelhos. Os resultados obtidos em cada um deles foram os seguintes:

A	5	10	7	15	16	12	4	8	10	13
B	10	9	20	9	11	8	9	7	8	9
C	10	11	9	10	10	9	11	12	8	10

Qual instrumento lhe parece recomendável? Justifique sua resposta.

PESQUISANDO

Para melhorar sua aprendizagem nesta unidade, sugiro que você pesquise:
- Os seguintes livros:

CRESPO, Antonio Arnot. *Estatística Fácil.* São Paulo: Editora Saraiva, 1998.

MAGALHÃES, Marcos Nascimento e LIMA, Antonio Carlos Pedroso de. *Noções de Probabilidade e Estatística.* São Paulo: 5a. ed. Editora da Universidade de São Paulo, 2002.

SPIEGEL, M. R., Estatística, São Paulo: Makron Books, 1993.

STEVENSON, Willian J. *Estatística Aplicada à Administração.* São Paulo: Ed. Harbra, 1981.

- acesse os sites:

www.ime.usp.br/~mae116

http://estatisticax.blogspot.com/2008/02/medidas-de-disperso.html

http://arquivos.unama.br/nead/gol/gol_adm_2mod/estatistica/pdf/ESTA_impresso_aula05.pdf

http://alea-estp.ine.pt/html/nocoes/html/cap5_3_1.html

http://educacao.uol.com.br/matematica/ult1705u28.jhtm

RESPOSTAS

Unidade 1

1)

a) Quantitativa
b) Quantitativa
c) Quantitativa
d) Qualitativa
e) Quantitativa
f) Quantitativa
g) Quantitativa
h) Qualitativa
i) Quantitativa
j) Quantitativa

2)

Idade (quantitativa)
Renda familiar (quantitativa)
Raça (qualitativa)
Grau de escolaridade (qualitativa)
Uso da camisinha (qualitativa)
Participação em movimentos sociais (qualitativa)
Violência sofrida (qualitativa)
Discriminação sofrida (qualitativa)

3)

a) Qualitativa nominal
b) Qualitativa nominal
c) Quantitativa contínua
d) Qualitativa ordinal
e) Qualitativa ordinal
f) Qualitativa ordinal
g) Qualitativa nominal
h) Qualitativa nominal

4) Definição do problema.

Unidade 2

3) Titulo, Cabeçalho, coluna indicadora, coluna numérica, corpo, rodapé.
4) Série mista.
5) Não pode ser fechada, pois, se assim for teremos um quadro e não uma tabela.
6) O que? Onde? e Quando?
7) É que um gráfico explica melhor que uma tabela, ou seja, é mais fácil compreendê-lo.
8) Clareza, simplicidade e veracidade.
9) Diagramas, pictogramas e cartogramas.

12) Com classe e sem classe.
14) É a Organização os dados de forma crescente ou decrescente.
15)

a) frequência acumulada

Fac
8
24
36
40

b) os pontos médios

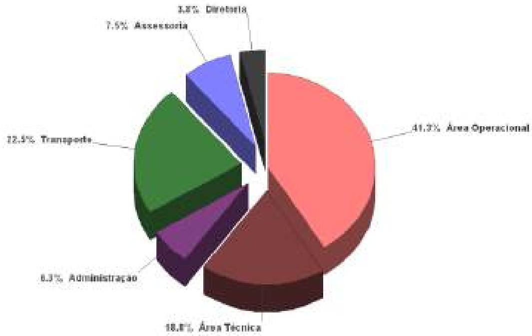
Xi
18
22
26
30

c) frequência relativa

Fr
0,2
0,4
0,3
0,1

d) 60%.

16)



17)

Valores	Fi	Fac	fr
1	4	**4**	0,08
2	4	**8**	**0,08**
3	8	16	0,16
4	7	**23**	0,14
5	5	28	0,10
6	**10**	38	0,20
7	7	45	0,14
8	5	50	0,10

18)62,5%.

20) No Histograma as colunas são justapostas, enquanto que no gráfico de barras as colunas são separadas.

21) No polígono de frequências o eixo das abscissas é cortado pela linha, já o gráfico de linhas não.
22) Série especificativa.

23) Com classe

Classes	**Frequências**
41|-45	7
45|-49	3
49|-53	4
53|-57	1
57|-61	5
Soma	**20**

Sem classe:

Dados	Frequência
41	3
42	2
43	1
44	1
45	1
46	2
50	1
51	1
52	1
54	1
57	1
58	1
60	1
Total	**20**

25)
a) Número de gols – variável quantitativa discreta.

Classes	Frequências
20|-30	2
30|-40	5
40|-50	4
50|-60	4
60|-70	1
70|-80	4
Soma	20

b) 70%

26)
a)

Classes	Frequências
1|-3	5
3|-5	14
5|-7	9
7|-9	0
9|-|11	2
Soma	30

27)
a) Obtenha a distribuição de frequências.

Classes	Frequências
1|-3	14
3|-5	25
5|-7	7
7|-9	3
9|-11	0
11|-13	0
13|-|15	1
Soma	50

b) Obtenha a frequência relativa percentual.

Classes	Frequências	Frequência Relativa (%)
1|-3	14	28
3|-5	25	50
5|-7	7	14
7|-9	3	6
9|-11	0	0
11|-13	0	0
13|-|15	1	2
Soma	50	100

Unidade 3

1) 10,5.
2) 58 anos.
3)
a)

Classes	fi	xi	Percentual (%)
1 |— 16	13	8.5	50.00
16 |— 31	3	23.5	11.54
31 |— 46	4	38.5	15.38
46 |— 61	5	53.5	19.23
61 |— 76	0	68.5	0.00
76|— 91	1	83.5	3.85
Total	26		100

b) Média = 26,38;
c) Primeiro Quartil = 8,5;
d) Terceiro Quartil = 44,125.

4) 38 anos.

5) 25 anos.
6) 50 anos.
7)
Especializados 70%;
Não Especializados 30%.
8)
a) 5,92;
b) 6;
c) 6.
9)
a) a média é 10,36;
b) a mediana é 9,57;
c) a moda é 8,826.
10)
a) 67,10;
b) 67,40;
c) 68,90.
11)
a) 1000 funcionários;
b) 200 funcionários;
c) 800 funcionários;
d) 25%.
12) 32 reais.
13) O valor comparativo é 3.
14)
a) média =21,44;
b) a moda=21,73;
c) a mediana =21,50.

15) R$ 2018,75.
16)
a)1,0219;
b)61,40%;
c)

Consumo modal	0,5|-- 1,0
Consumo mediano	0,5|-- 1,0

17) A Média é 59,6; a Mediana é 63; a Moda é 67.
18)
a) o 1º quartil é 5,15; b) o 8º decil é 7,5;
c) 90º percentil é 7,75 ;
19) 27 reais.
20)
Mulheres 70% e homens 30%;
21) Letra A

Unidade 4

1) Variância é 2,64; Desvio padrão é 1,63; Coeficiente de variação é 37,90%; Desvio médio é 1,36
2)
a)
Paciente: variável quantitativa discreta
Fisioterapia: variável quantitativa discreta
Sequelas: variável qualitativa nominal
Cirurgia: variável qualitativa ordinal.
b)
Grupo com sequelas:
Média= 6,57
Grupo sem sequelas:

Média= 5,25
O grupo com sequelas apresentou uma média maior em relação ao grupo sem sequelas.
c)
Cv com sequelas= 19,48%
Cv sem sequelas = 20,40%. Então, o grupo mais homogêneo é o com sequelas.

3) O Volume é que tem maior variabilidade.
4) Desvio padrão é 3,16
5)
a)Grupo A
Média =3,43
Desvio padrão= 2,48
Mediana=3,8

Grupo B
Média= 5,77
Desvio padrão= 1,95
Mediana= 6,45

b)
CV_A= 72,30%
CV_B=33,79%
O grupo mais homogêneo é o B, pois, tem Coeficiente de variação inferior ao grupo A.

6)
a) 0,1035;
b) Assimetria Positiva;
c) 0,242;
d) Leptocúrtica.

7)
a)Coeficiente de Assimetria. -0,272;
b)Coeficiente de Curtose. 0,26.

8)
Variância =114,13
Desvio padrão = 10,68;
Coeficiente de variação = 17,98%.

9) A menor variabilidade é da turma B.

10) Sim, significa que não houve variação nos dados, todos tinham o mesmo valor.

11)
a) 7;
b) 7,99.

12)
a) Assimétrica positiva; Leptocúrtica.
b) Assimétrica Negativa; Leptocúrtica
c) Assimétrica positiva; Platicúrtica.
d) Assimétrica Negativa; Leptocúrtica.
e) Assimétrica negativa; Platicúrtica.

13)
a) Os resultados revelam que os métodos têm a mesma média, então será necessário calcular o Desvio Padrão que é uma medida de dispersão, pois, o método que apresentar menor desvio padrao será o melhor.
b)
Desvio padrão Z_1= 0,4248;
Coeficiente de variação Z_1=7,92%;
Desvio padrão Z_2=2,04;
Coeficiente de variação Z_2=38%;

14)
a) 40;
b) 73,625;
c) 5,17;
d) 50,29;
e) 7,09;
f) 9,63%;

15)
a)
Desvio médio = 1,95 (ANTES);
Desvio médio = 2,82(DEPOIS);

b)
S^2(Variância)= 7,64 (ANTES);
S^2 (Variância)= 13,62(DEPOIS);

c)
S(Desvio padrão)= 2,76(ANTES);
S(Desvio padrão)= 3,69(DEPOIS);

d)
Coeficiente de variação = 98,57(ANTES)
Coeficiente de variação = 59,13(DEPOIS)

16)
Como as médias são iguais é recomendável usar o desvio padrão por ser uma medida bastante representativa. Neste caso, temos:
Desvio padrão de A é 4,055
Desvio padrão de B é 3,68
Desvio padrão de C é 1,155

Então, o aparelho C é aparentemente o mais homogêneo.

UNIVERSIDADE DO ESTADO DO PARÁ