e-Tec Brasil
Escola Técnica Aberta do Brasil

Colégio Técnico Industrial de Santa Maria

Curso Técnico em Automação Industrial

Estatística

Paulo Roberto da Costa

Colégio Técnico Industrial UFSM
Universidade Federal de Santa Maria 1960
Universidade Federal de Santa Catarina
Ministério da Educação
Brasil Um País de Todos
Governo Federal

PAULO ROBERTO DA COSTA

ESCOLA TÉCNICA ABERTA DO BRASIL – E-TEC BRASIL

CURSO TÉCNICO EM AUTOMAÇÃO INDUSTRIAL

Disciplina: Estatística

COLÉGIO TÉCNICO INDUSTRIAL DE SANTA MARIA
Santa Maria – RS
2009

Presidência da República Federativa do Brasil
Ministério da Educação
Secretaria de Educação a Distância

Este Caderno foi elaborado em parceria entre o Colégio Técnico Industrial de Santa Maria e a Universidade Federal de Santa Catarina para o Sistema Escola Técnica Aberta do Brasil – e-Tec Brasil.

Equipe de Elaboração
Colégio Técnico Industrial de Santa Maria – CTISM

Coordenação Institucional
Paulo Roberto Colusso/CTISM

Professor-autor
Paulo Roberto da Costa/CTISM

Equipe Técnica
Carlos Gustavo Hoelzel/CTISM
Ana Claudia Pavão Siluk/CTISM
Sílvia Nascimento/CTISM

Revisão
Eduardo Lehnhart Vargas/CTISM
Francisca Lima Rodrigues/CTISM
Francine Netto Martins
Gilciano Sala/CTISM
Marcus Vinicius Braun/CTISM

Ilustração
Rafael Cavalli Viapiana/CTISM
Marcel Jacques/CTISM
André Krusser Dalmazzo/CTISM

Comissão de Acompanhamento e Validação
Universidade Federal de Santa Catarina – UFSC

Coordenação Institucional
Araci Hack Catapan/UFSC

Coordenação do Projeto
Silvia Modesto Nassar/UFSC

Coordenação de Design Instrucional
Beatriz Helena Dal Molin/UNIOESTE

Design Instrucional
Doris Roncarelli/UFSC

Web Design
Beatriz Wilges/UFSC

Projeto Gráfico
Beatriz Helena Dal Molin/UNIOESTE
Araci Hack Catapan/UFSC
Elena Maria Mallmann/UFSC
Jorge Luiz Silva Hermenegildo/CEFET-SC
Mércia Freire Rocha Cordeiro Machado/ETUFPR
Silvia Modesto Nassar/UFSC

Supervisão de Projeto Gráfico
Luís Henrique Lindner/UFSC

Diagramação
Bruno César Borges Soares de Ávila/UFSC
Luís Henrique Lindner/UFSC

Ficha catalográfica elaborada por Maristela Eckhardt - CRB-10/737
Biblioteca Central da UFSM

C837e Costa, Paulo Roberto da
Estatística / Paulo Roberto da Costa.– Santa Maria : Universidade Federal de Santa Maria, Colégio Técnico Industrial de Santa Maria, Curso Técnico em Automação Industrial, 2009.
86 p. : il. ; 21 cm.

1. Estatística 2. Probabilidade 3. Programa Escola Aberta do Brasil I. Universidade Federal de Santa Maria. Curso Técnico em Automação Industrial.

CDU 519.22

PROGRAMA E-TEC BRASIL

Amigo(a) estudante!

O Ministério da Educação vem desenvolvendo Políticas e Programas para expansão da Educação Básica e do Ensino Superior no País. Um dos caminhos encontrados para que essa expansão se efetive com maior rapidez e eficiência é a modalidade a distância. No mundo inteiro são milhões os estudantes que frequentam cursos a distância. Aqui no Brasil, são mais de 300 mil os matriculados em cursos regulares de Ensino Médio e Superior a distância, oferecidos por instituições públicas e privadas de ensino.

Em 2005, o MEC implantou o Sistema Universidade Aberta do Brasil (UAB), hoje, consolidado como o maior programa nacional de formação de professores, em nível superior.

Para expansão e melhoria da educação profissional e fortalecimento do Ensino Médio, o MEC está implementando o Programa Escola Técnica Aberta do Brasil (e-Tec Brasil). Espera, assim, oferecer aos jovens das periferias dos grandes centros urbanos e dos municípios do interior do País oportunidades para maior escolaridade, melhores condições de inserção no mundo do trabalho e, dessa forma, com elevado potencial para o desenvolvimento produtivo regional.

O e-Tec é resultado de uma parceria entre a Secretaria de Educação Profissional e Tecnológica (SETEC), a Secretaria de Educação a Distância (SEED) do Ministério da Educação, as universidades e escolas técnicas estaduais e federais.

O Programa apóia a oferta de cursos técnicos de nível médio por parte das escolas públicas de educação profissional federais, estaduais, municipais e, por outro lado, a adequação da infra-estrutura de escolas públicas estaduais e municipais.

Do primeiro Edital do e-Tec Brasil participaram 430 proponentes de adequação de escolas e 74 instituições de ensino técnico, as quais propuseram 147 cursos técnicos de nível médio, abrangendo 14 áreas profissionais. O resultado desse Edital contemplou 193 escolas em 20 unidades federativas. A perspectiva do Programa é que sejam ofertadas 10.000 vagas, em 250 polos, até 2010.

Assim, a modalidade de Educação a Distância oferece nova interface para a mais expressiva expansão da rede federal de educação tecnológica dos últimos anos: a construção dos novos centros federais (CEFETs), a organização dos Institutos Federais de Educação Tecnológica (IFETs) e de seus *campi*.

O Programa e-Tec Brasil vai sendo desenhado na construção coletiva e participação ativa nas ações de democratização e expansão da educação profissional no País, valendo-se dos pilares da educação a distância, sustentados pela formação continuada de professores e pela utilização dos recursos tecnológicos disponíveis.

A equipe que coordena o Programa e-Tec Brasil lhe deseja sucesso na sua formação profissional e na sua caminhada no curso a distância em que está matriculado(a).

Brasília, Ministério da Educação – setembro de 2008.

SUMÁRIO

PALAVRAS DO PROFESSOR-AUTOR

"A questão primordial não é o que sabemos, mas como o sabemos."

Aristóteles

"Não há ramo da Matemática, por mais abstrato que seja que não possa um dia vir a ser aplicado ao ramo do mundo real."

Lobachevsky

Ao elaborarmos este material, procuramos levar em conta as duas afirmações acima, conservando o que julgamos ser fundamental: desenvolvimento claro e compreensível de todos os conceitos básicos, aliado ao rigor matemático, com o objetivo de fazer com que o estudante compreenda as idéias básicas da Estatística e, quando necessário, saiba aplicá-las na resolução de problemas do mundo real.

Esperamos que, este material facilite sua aprendizagem e desenvolva a sua forma de pensar matematicamente. Queremos que ele seja bastante útil na busca de soluções para uma série de problemas do dia-a-dia.

Portanto, faz-se necessário muito estudo para obter sucesso no processo de aprendizagem; mãos à obra!

Que o proveito seja significativo.

Paulo Roberto da Costa

PROJETO INSTRUCIONAL

A disciplina de Estatística apresenta-se na forma a distância.

A carga horária total é de 30 horas, sendo todas as horas no Ambiente Virtual de Ensino-Aprendizagem (AVEA). Este material didático contém as palavras do professor-autor, projeto instrucional, mapa conceitual da disciplina com a composição das unidades de estudo, uma introdução geral às oito unidades, que são as seguintes: Conceitos, Fases do Método Estatístico, Tabelas, Séries Estatísticas, Gráficos Estatísticos, Distribuição de Frequências, Medidas Descritivas e Probabilidade, juntamente com as referências.

Essas unidades serão estudadas de forma sequencial, permitindo ao estudante construir o conhecimento a partir do entendimento teórico da Estatística, relacionando os conhecimentos adquiridos nas unidades estudadas, visando identificar suas diversas aplicações industriais. Cada unidade de estudo apresenta: uma introdução, objetivos e um roteiro de estudo, composto por textos, figuras, exemplos, equações, *links*, mídias integradas, questionamentos, reflexões, lembretes, atividades de aprendizagem e avaliação, e síntese.

As atividades de aprendizagem e avaliação, dispostas ao final de cada unidade, visam reforçar o entendimento dos conceitos, bem como permitir uma reflexão acerca das aplicações do assunto estudado ao cotidiano. Os questionamentos presentes no material didático, bem como as atividades de aprendizagem e avaliação serão objetos de discussão em fóruns específicos no AVEA. Dessa forma, as atividades de aprendizagem e avaliação serão desenvolvidas através da participação em fóruns, bem como pela realização de tarefas, exercícios e pesquisas complementares. Estas atividades serão avaliadas pelo professor e pelos tutores da disciplina, cujos resultados e sugestões serão divulgados aos estudantes. Devido ao estudo da Estatística envolver também análise e ferramentas matemáticas, serão disponibilizadas no AVEA atividades de aprendizagem adicionais. Os critérios de avaliação da disciplina serão disponibilizados no AVEA, assim como as instruções, períodos e prazos.

No AVEA serão utilizados ainda recursos como hipertextos, animações, simulações, vídeos e imagens, vinculando o ambiente virtual ao material didático. Serão disponibilizados também cronogramas de estudo.

ÍCONES E LEGENDAS

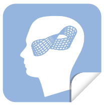
Caro estudante! Oferecemos para seu conhecimento os ícones e sua legenda que fazem parte da coluna de indexação. A intimidade com estes e com o sentido de sua presença no caderno ajudará você a compreender melhor as atividades e exercícios propostos (DAL MOLIN, *et al.*,2008).

Saiba mais

Ex: http://www.etecbrasil.mec.gov.br

Este ícone apontará para atividades complementares ou para informações importantes sobre o assunto. Tais informações ou textos complementares podem ser encontrados na fonte referenciada junto ao ícone.

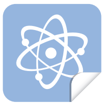
Para refletir...



Ex: Analise o caso... dentro deste tema e compare com..., Assista ao filme...

Toda vez que este ícone aparecer na coluna de indexação indicará um questionamento a ser respondido, uma atividade de aproximação ao contexto no qual você vive ou participa, resultando na apresentação de exemplos cotidianos ou *links* com seu campo de atuação.

Mídias integradas



Ex.: Assista ao filme... e comente-o.

Quando este ícone for indicado em uma dada unidade significa que você está sendo convidado a fazer atividades que empreguem diferentes mídias, ou seja, participar do AVEA, assistir e comentar um filme, um videoclipe, ler um jornal, comentar uma reportagem, participar de um *chat*, de um fórum, enfim, trabalhar com diferentes meios de comunicação.

Ícones e Legendas

Avaliação

Este ícone indica uma atividade que será avaliada dentro de critérios específicos da unidade.

Lembre-se

Ex.: O canal de satélite deve ser reservado com antecedência junto à Embratel.

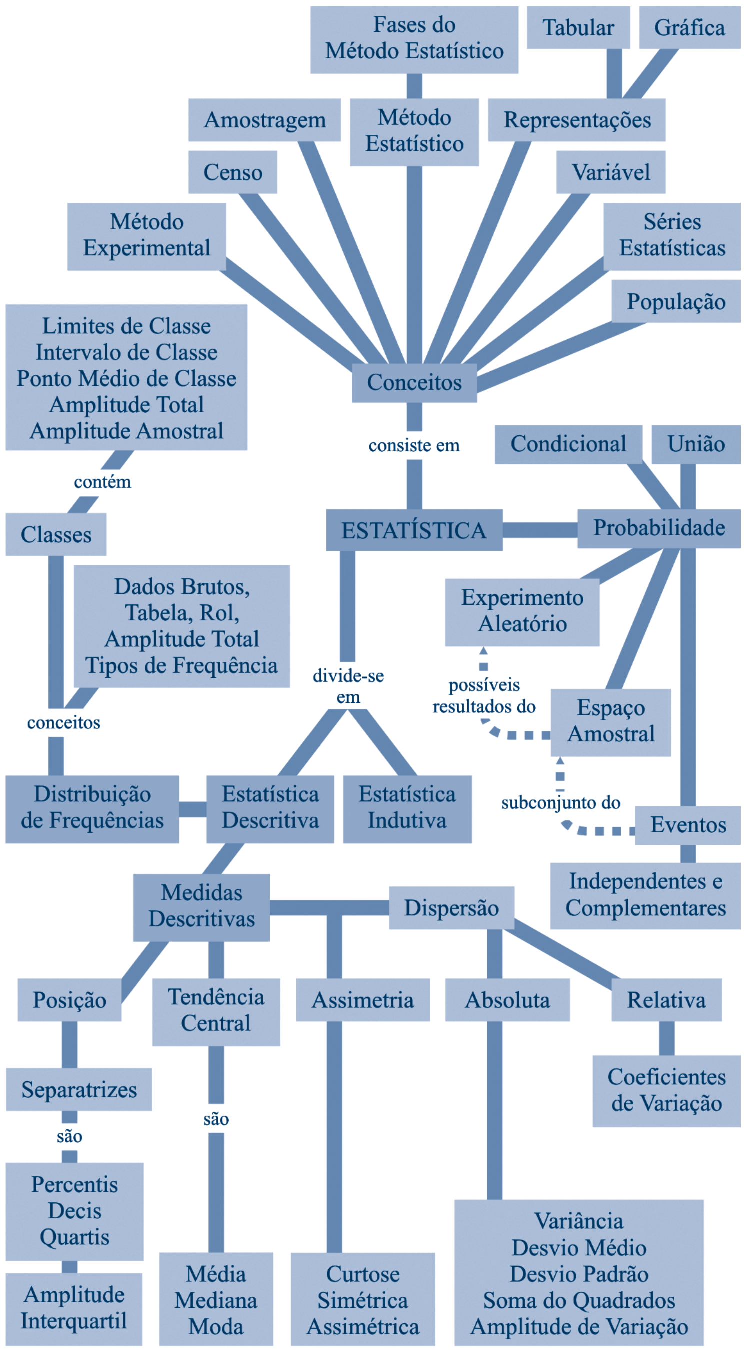
A presença deste ícone ao lado de um trecho do texto indicará que aquele conteúdo significa algo fundamental para a aprendizagem.

Destaque

Retângulo com fundo colorido.

A presença do retângulo de fundo indicará trechos importantes do texto, destacados para maior fixação do conteúdo.

MAPA CONCEITUAL



INTRODUÇÃO

Investigação sugere pesquisa, busca de informações e análise de dados. Tudo isto faz pensar em Estatística.

O relacionamento da Estatística com as demais ciências é cada vez mais intenso e mais importante. Veja, por exemplo, que a Estatística auxilia a Genética, nas questões de hereditariedade; é valiosa na Economia, na análise da produtividade, da rentabilidade e estudos de viabilidade; é básica para as Ciências Sociais nas pesquisas sócio-econômicas; é de aplicação intensa na Engenharia Industrial, no controle de qualidade e na comparação de fabricações.

Quando a Estatística é aplicada a dados provenientes de observações realizadas em diferentes aspectos das Ciências da Vida, como: medicina, psicologia, nutrição, biologia, farmácia, enfermagem, odontologia, veterinária e agronomia, neste caso é utilizado o termo **Bioestatística** para distinguir da aplicação de outras áreas do conhecimento.

A Estatística fornece-nos as técnicas para extrair informação de dados, os quais são muitas vezes incompletos, na medida em que nos dão informação útil sobre o problema em estudo. Sendo assim, é objetivo da Estatística extrair informação dos dados para obter uma melhor compreensão das situações que representam.

Conheça o Estatuto do Conselho Nacional de Estatística, acesse o *site*: http://www.ine.cv/Legislaçao/CNEST/Estatutos do CNEST.pdf

Quando se aborda uma problemática envolvendo métodos estatísticos, estes devem ser utilizados mesmo antes de se recolher a amostra. Isto é, deve-se planejar a experiência que nos vai permitir recolher os dados, de modo que, posteriormente, se possa extrair o máximo de informação relevante para o problema em estudo, ou seja, para a população de onde os dados provêm.

Quando de posse dos dados, primeiramente procura-se agrupá-los e reduzi-los, sob forma de amostra, deixando de lado a aleatoriedade presente.

O objetivo do estudo estatístico pode, também, ser o de estimar uma quantidade ou testar uma hipótese, utilizando-se técnicas estatísticas convenientes. Tais técnicas realçam toda a potencialidade da Estatística na medida em que vão permitir tirar conclusões acerca de uma população, baseando-se numa pequena amostra, dando-nos ainda uma medida do erro cometido.

Concluindo, a Estatística tornou-se uma poderosa ferramenta para a compreensão, análise e previsão de inúmeras situações em nossa vida. As empresas utilizam-se dos modelos estatísticos para calcular fluxo de estoques, de consumo e de produção, objetivando adaptar-se rapidamente a um mercado em constante mutação. Para podermos nos situar de forma mais precisa possível nesse rápido processo de mudanças que enfrentamos, é necessário a utilização dessas ferramentas. Por isso, é necessário saber ler gráficos, interpretá-los, prever situações, analisar dados, etc. A Estatística então nos ajudará nesta tarefa.

Panorama histórico da Estatística

Historicamente, o desenvolvimento da Estatística pode ser entendido a partir de dois fenômenos: a necessidade de governos coletarem dados censitários, e o desenvolvimento da Teoria do Cálculo de Probabilidade.

Dados têm sido coletados ao longo da história. Vejamos:

a) **Antiguidade:** vários povos já registravam o número de habitantes, de nascimentos e de óbitos. Faziam estimativas das riquezas sociais, distribuíam equitativamente terra aos povos, cobravam impostos e realizavam inquéritos quantitativos por processos que, hoje, chamaríamos de "estatísticas".
b) **Idade média:** colhiam-se informações, que geralmente eram tabuladas com finalidades tributárias ou bélicas.
c) **Séc. XVI**: surgiram as primeiras análises sistemáticas, as primeiras tabelas e os números relativos.

Periodicamente, nas civilizações antigas eram feitos inquéritos sobre os quantitativos anuais de trigo e outros produtos e, com base nesses dados, eram estabelecidos impostos. Esta prática também era utilizada na Idade Média.

Até o início do séc. XVII, a Estatística servia apenas para assuntos de Estado e limitava-se a uma simples técnica de contagem, traduzindo numericamente fatos ou fenômenos observados, era a fase da **Estatística Descritiva.**

d) **Séc. XVII**: iniciou-se na Inglaterra uma nova fase de desenvolvimento da Estatística, voltada para a análise dos fenômenos observados, fase da **Estatística Analítica.**

John Graunt (1620-1674) foi quem publicou pela primeira vez, concretamente, um trabalho de Estatística, o qual se reportava a mortalidade dos habitantes de Londres. Esse trabalho foi a base do aparecimento das primeiras tábuas de mortalidade e da elaboração de previsões sobre a duração de vida humana. Nasceu assim a **Demografia**.

e) **Séc. XVIII:** A origem do termo **estatística** surgiu no século XVIII. Quanto à origem da Estatística, a data do seu aparecimento não parece ser encarada com unanimidade. Há quem diga que o seu autor foi o professor Godofredo Achenwall (1719-1772), que usou pela primeira vez o termo **estatística** (*statistik*, do grego *statizein*). Há também quem diga que tem origem na palavra **estado**, do latim *status*, pelo aproveitamento que dela tiravam os políticos e o Estado.

Contudo, muito antes do termo estatística, os romanos asseguravam o recenseamento dos cidadãos, e a Bíblia chega até a testemunhar um desses recenseamentos.

f) Ao longo dos **séculos XVIII e XIX**, a Estatística avançou muito com a associação ao cálculo de probabilidade, que havia se desenvolvido, e a realização de trabalhos de pesquisa científica nos domínios da Botânica, Biologia, Meteorologia, Astronomia, etc. Mais tarde a Estatística deixou de ser mera técnica de contagem de fenômenos para se transformar numa poderosa "alfaia" científica ao serviço dos diferentes ramos do saber, aparece então a fase da **Estatística Aplicada.** É com essas características que a Estatística é hoje reconhecida, pois atualmente informações numéricas são necessárias para cidadãos e organizações de qualquer natureza e de qualquer parte do mundo globalizado.

UNIDADE 1 – CONCEITOS

1.1 Objetivos de aprendizagem

- Definir o que significa e quais os tipos de Estatística;
- Reconhecer os processos relacionados ao estudo da população na Estatística;
- Diferenciar e aplicar os métodos científicos;
- Diferenciar e determinar as variáveis quantitativas e qualitativas.

1.2 Introdução

Quando se faz uma **pesquisa científica**, o procedimento geral é formular hipóteses e testá-las. Inicialmente essas hipóteses são formuladas em termos científicos, dentro da área de estudo, em seguida, as hipóteses devem ser expressas em termos estatísticos.

Existem muitas definições propostas por autores, objetivando estabelecer com clareza a Estatística, mas de uma maneira geral como veremos a seguir, ela visa elaborar métodos capazes e aplicáveis a todas as fases do estudo dos fenômenos de massa.

1.3 Método científico

1.3.1 Método

Conjunto de meios e rotinas dispostos convenientemente para se chegar a um fim que se deseja.

1.3.2 Método experimental

Método que consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que se possam achar seus efeitos.

1.3.3 Método estatístico

Método que admite todas as causas presentes variando-as, dada a impossibilidade de manter as causas constantes, registrando estas variações e procurando determinar as influências que cabem a cada uma delas.

1.4 O que é Estatística?

Figura 1.1 – Pesquisa e Estatística

Estatística é um ramo da Matemática Aplicada, que se destina ao estudo dos processos de obtenção, coleta, organização, apresentação, descrição, análise e interpretação de dados numéricos variáveis, referentes a qualquer fenômeno, sobre uma população, coleção ou conjunto de seres para a utilização dos mesmos na tomada de decisões (Figura 1.1).

Em outras palavras, Estatística é a ciência que tem como objetivo fundamental o estudo de uma população.

Este estudo pode ser feito de duas maneiras:

- Investigando todos os elementos da população.
- Por amostragem, ou seja, selecionando alguns elementos da população.

Assim, a Estatística divide-se em duas grandes áreas que são: **Estatística Descritiva** e **Estatística Indutiva ou Inferencial**.

1.4.1 Estatística descritiva

É aquela que possui um conjunto de técnicas para planejar, organizar, coletar, resumir, classificar, apurar, descrever, comunicar e analisar os dados em tabelas, gráficos ou em outros recursos visuais. Além do cálculo de estimativas de parâmetros representativos desses dados, interpretação de coeficientes e exposição que permitam descrever o fenômeno.

1.4.2 Estatística indutiva ou inferencial

É o conjunto de técnicas que partindo de uma amostra, estabelece hipóteses, tira conclusões sobre a população de origem e formula previsões fundamentando-se na Teoria de Probabilidade, baseia-se na análise e a interpretação dos dados.

O processo de generalização do método indutivo está associado a uma margem de incerteza. Isto se deve ao fato de que a conclusão que se pretende obter para o conjunto de todos os indivíduos (população) analisados, quanto a determinadas características comuns, baseia-se em uma parcela (amostra) de observações.

Na análise estatística de dados pode-se obter os resultados de duas maneiras: através de um censo ou através de uma amostragem, isto é pesquisa em uma amostra (Figura 1.2).

Exemplo:
Pesquisa de mercado, pesquisa de opinião pública e praticamente todo experimento.

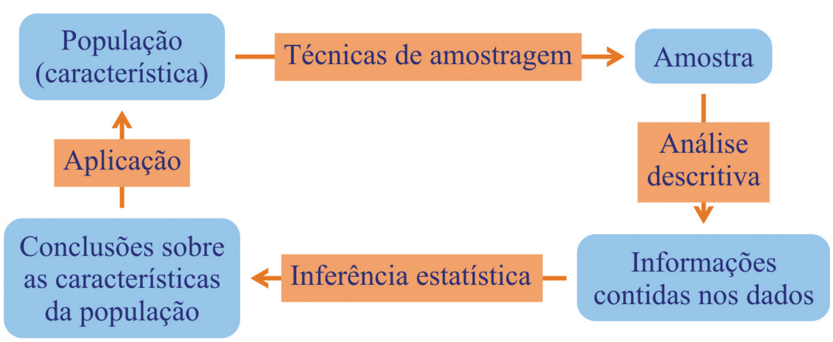


Figura 1.2 – Análise estatística de dados de uma amostra

1.5 População

A população e a amostra estão diretamente relacionadas com a inferência estatística. Geralmente a partir da amostra obtém-se os dados e as estatísticas que, por meio da inferência estatística, nos permitem entender de uma maneira mais clara, a população de referência.

1.5.1 População estatística ou universo estatístico

É o conjunto de informações ou conjunto de entes ou seres portadores de pelo menos uma característica em comum, cujo comportamento interessa-nos analisar. Em outras palavras, é o conjunto de todas as medidas e observações relativas ao estudo de determinado fenômeno.

Exemplo:
Os estudantes constituem uma população com uma característica em comum: matriculados em um determindo curso.

Como em qualquer estudo estatístico objetiva-se conhecer uma ou mais características dos elementos de uma população. É importante definir bem essas características de interesse para que sejam delimitados os elementos que pertencem à população e quais os que não pertencem.

Exemplos:

1. Deseja-se saber se nas indústrias situadas no Estado do Rio Grande do Sul, em 2007, existia algum tipo de controle ambiental.

   População ou universo: indústrias situadas no Estado do Rio Grande do Sul em 2007.

   Características: Existência ou não de algum tipo de controle ambiental na indústria.

2. Deseja-se conhecer o consumo total de energia elétrica em MWh nas residências da cidade de Porto Alegre no ano de 2007.

   População ou universo: todas as residências que estavam ligadas à rede elétrica em Porto Alegre no ano de 2007.

   Características: consumo anual de energia elétrica em MWh.

1.5.1.1 Divisão da população

a) **População finita:** apresenta um número limitado de elementos. É possível enumerar todos os elementos componentes.

Exemplo:
Idade dos alunos do curso de Medicina UFRGS no Estado do Rio Grande do Sul.

População: todos os alunos do curso de Medicina da UFRGS no Estado do Rio Grande do Sul, em 2007.

b) **População infinita:** apresenta um número ilimitado de elementos. Não é possível enumerar todos os elementos componentes. Entretanto, tal definição existe apenas no campo teórico. Uma vez que na prática, nunca encontraremos populações com infinitos elementos, mas sim, populações com grande número de componentes. Nessas circunstâncias, tais populações são tratadas como se fossem infinitas.

Exemplo:
Tipos de bactérias no corpo humano.

População: todas as bactérias existentes no corpo humano.

Em geral, como os universos são grandes então para investigar todos os elementos populacionais, e determinarmos a característica de interesse, pode ocorrer: necessidade de muito tempo, o custo é elevado, o processo de investigação leva à destruição do elemento observado, ou, como no caso de populações infinitas, é impossível observar a totalidade da população. Assim, estudar parte da população constitui-se um aspecto fundamental da Estatística.

1.6 Amostra

É uma parte ou um subconjunto representativo de uma população, selecionado segundo métodos adequados. Os dados de observação, registrados na amostra, fornecem informações sobre a população. O processo pelo qual se tira conclusões sobre a população, com base nos resultados obtidos na amostra, é chamado inferência estatística. As estatísticas obtidas na amostra são denominadas estimativas. Portanto, toda a análise estatística será inferida a partir das características obtidas da amostra, é importante que a amostra seja representativa da população, isto é, que as suas características (amostra) sejam em geral as mesmas que do todo (população). Enfim, a amostra é um subconjunto finito e representativo de uma população.

Muitas vezes, por motivos práticos ou econômicos, limita-se os estudos estatísticos somente a uma parte da população - a amostra. As razões de recorrer às amostras são: menor custo e tempo para o levantamento de dados, e melhor investigação dos elementos observados.

1.7 Amostragem

É a coleta das informações de parte da população, chamada amostra, mediante métodos adequados de seleção destas unidades.

É a técnica especial de escolher amostras, de tal forma que garanta o acaso na escolha. Assim, cada elemento da população tem a mesma chance de ser escolhido, o que garante à amostra um caráter de representatividade da população.

1.7.1 Métodos de amostragem

a) Amostragem de conveniência

Trata-se de um procedimento não-probabilístico, para casos onde é inviável realizar um sorteio entre todos os componentes da população alvo.

Exemplo:
Pesquisas clínicas.

b) Amostragem casual ou aleatória simples

Este tipo de amostragem é baseado no sorteio da amostra. Numera-se a população de 1 a n, e depois utilizando um dispositivo aleatório qualquer escolhe-se k números desta sequência, os quais corresponderão aos elementos da amostra.

c) Amostragem proporcional estratificada

Quando as populações se dividem em subpopulações (**estratos)** é necessário utilizar uma amostragem proporcional estratificada, que considera os estratos (subgrupos) e obtém a amostra proporcional a estes.

Exemplo:
Suponha um Grupo com 90 alunos, 54 sejam meninos e 36 sejam meninas. Neste caso precisamos obter a amostra estratificada. Serão dois estratos (sexo masculino e sexo feminino e queremos uma amostra de 10% da população. Assim,

1. Definimos a amostra em estratos:

Sexo	População	10%	Amostra
Meninas	54	5,4	5
Meninos	36	3,6	4
Total	90	9,0	9

2. Numeram-se os alunos de 1 a 90 sendo que 1 a 54 correspondem a meninos e de 55 a 90, a meninas. Tomando a 2ª coluna desta determinada tabela, de cima para baixo, tem-se: 05, 17, 31, 46, 56, 63, 74, 75 e 90.

3. Neste caso serão obtidas as características dos seguintes alunos:

Meninos: 05, 17, 31 e 46.
Meninas: 56, 63, 74, 75 e 90.

d) Amostragem por meio de conglomerados

Aplicada quando a população apresenta uma subdivisão em pequenos grupos, chamados conglomerados. É possível e muitas vezes conveniente fazer-se a amostragem por meio desses conglomerados, cujos elementos constituirão a amostra. Ou seja, as unidades de amostragem sobre as quais é feito o sorteio passam a ser os conglomerados e não mais os elementos individuais da população. A amostragem por meio de conglomerados é adotada por motivos de ordem prática e econômica.

Exemplo:
Quarteirões de um bairro.

e) Amostragem sistemática

Aplicada quando os elementos da população já estão ordenados, assim não é necessário construir um sistema de referência ou de amostragem.

Exemplo:
Suponha uma rua que tenha 500 prédios e desejamos obter uma amostra de 40 prédios (8%). Como os prédios já estão ordenados na rua, podemos usar o seguinte procedimento:

1. Como $500 \div 40 = 12.5$; então temos de selecionar um prédio para a amostra a cada 12;
2. Sorteamos um número entre 1 e 12, incluindo ambos. Digamos que seja sorteado 5;
3. Vamos amostrando os prédios iniciando pelo 5º e pulando de 12 em 12. Assim, iniciamos pelo prédio 5, depois usamos o prédio $12 + 5$, depois $12 + 12 + 5$, e assim por diante.
4. No final teremos amostrado 40 prédios.

f) Amostragem por estágios múltiplos

Numa amostragem múltipla a amostra é retirada em diversas etapas sucessivas. Dependendo dos resultados observados, etapas suplementares podem ser dispensadas. É bastante empregada na inspeção por amostragem, sendo particularmente importante a amostra dupla. Tem por finalidade diminuir, a longo prazo, o número médio de itens inspecionados, baixando o custo da inspeção.

1.8 Censo

É o exame completo de toda população. Quanto maior a amostra, mais precisas e confiáveis deverão ser as induções feitas sobre a população. Logo, os resultados mais perfeitos são obtidos pelo Censo. Na prática, esta conclusão muitas vezes não acontece, pois, o emprego de amostras, com certo rigor técnico, pode levar a resultados mais confiáveis ou até mesmo melhores do que os que seriam obtidos através de um Censo (Figura 1.3).

Figura 1.3 – Família sendo entrevistada durante o censo das favelas do PAC (Programa de aceleração do Crescimento) na comunidade de Manguinhos, RJ
Fonte: http://www.imprensa.rj.gov.br/SCSSiteImprensa/detalhe_foto.asp?id=16879

Conheça o Portal do Sistema Estatístico Nacional, acesse o *site*: http://www.ine.cv/

A Estatística ocupa-se fundamentalmente das propriedades das populações cujas características são passíveis de representação numérica, como resultado de medições e contagens. Essas características da população são, comumente, chamadas de **variáveis**.

1.9 Variáveis estatísticas

As características ou variáveis podem ser divididas em dois tipos: **variáveis qualitativas** e **variáveis quantitativas.**

1.9.1 Variáveis qualitativas

Quando o resultado da observação é apresentado na forma de qualidade ou atributo. Dividem-se em:

a) **Variáveis nominais***:* quando podem ser separadas por categorias chamadas de não mensuráveis.

Exemplo:
A cor dos olhos, tipo de acomodação, marcas de carro, sexo...

b) Variáveis ordinais: quando os códigos numéricos podem agir como categorias ou ordenações. Como sugere o nome, elas envolvem variáveis que representam algum elemento de ordem. Uma classificação em anos pode ser um exemplo clássico, por exemplo, a faixa etária dos indivíduos. A classificação deste tipo de variáveis geralmente causa confusão ao pesquisador.

> Exemplo:
> 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).

1.9.2 Variáveis quantitativas

Quando o resultado da observação é um número, decorrente de um processo de mensuração ou contagem. Dividem-se em:

a) Variáveis contínuas: são aquelas que podem assumir qualquer valor num certo intervalo (contínuo) da reta real. Não é possível enumerar todos os possíveis valores. Essas variáveis, geralmente, provêm de medições.

> Exemplo:
> A altura dos alunos é uma variável contínua, pois teoricamente, um aluno poderá possuir altura igual a 1.70m, 1.71m, 1.711m, 1.712m. (medições: peso, estatura, etc.)

b) Variáveis discretas: são aquelas que podem assumir apenas valores inteiros em pontos de uma reta. É possível enumerar todos os possíveis valores da variável.

> Exemplo:
> Número de alunos de uma escola, número de mensagens em um e-mail, etc.

As variáveis podem ser resumidas da seguinte maneira:

Quantitativas		Qualitativas	
Contínua	Discreta	Nominal	Ordinal
peso, altura, idade	nº de filhos, nº de carros	sexo, cor dos olhos	classe social, grau de instrução

1.10 Atividades de aprendizagem e avaliação

01. Explique o que você entende por Estatística e como ela se divide.
02. Quais as diferenças entre população, amostra e amostragem?
03. Defina os métodos de amostragem.
04. Qual o significado do censo?

1.11 Síntese

Nesta unidade tratamos do tema relacionado à Estatística de uma forma geral. Apontamos também para sua aplicação em estudos de populações. Estudamos ainda as relações entre população, amostra e amostragem. Finalizamos com as características relacionadas à população, conhecidas também como variáveis.

UNIDADE 2 – FASES DO MÉTODO ESTATÍSTICO

2.1 Objetivos de aprendizagem

- Conhecer e analisar as fases do método estatístico;
- Empregar as fases do método estatístico.

2.2 Introdução

Ao desenvolver um estudo estatístico completo, existem algumas fases do seu método que devem ser desenvolvidas, em sequência, para se chegar aos resultados finais do trabalho. Assim, nesta unidade estudaremos essas várias fases.

2.3 Divisão das fases

As fases do método estatístico são seis:

- Definição do problema;	- Apuração dos dados;
- Planejamento;	- Apresentação dos dados;
- Coleta de dados;	- Análise e interpretação dos dados.

a) 1ª fase - definição do problema: saber exatamente aquilo que se pretende pesquisar é o mesmo que definir corretamente o problema.

b) 2ª fase - planejamento: consiste em planejar o modo como serão realizadas as fases seguintes, determinando o objetivo da pesquisa e os métodos que serão utilizados.

Nesta etapa são definidos os objetivos, as características da amostra, o método de aquisição e de processamento de dados.

A seguir listamos algumas perguntas presentes nesta fase:

> Como levantar informações? Que dados deverão ser obtidos?
> Qual levantamento a ser utilizado? Censitário? Por amostragem?
> E o cronograma de atividades? Os custos envolvidos? Etc.

c) 3ª fase - coleta de dados: fase operacional. É o registro sistemático de dados, com um objetivo determinado.

Após a definição do problema a ser estudado e o estabelecimento do planejamento da pesquisa (forma pela qual os dados serão coletados; cronograma das atividades; custos envolvidos exame das informações disponíveis; delineamento da amostra; etc.), o passo seguinte é a coleta de dados, que podem ser de dois tipos:

1. Coleta direta

A coleta direta de dados é quando os dados são obtidos pelo próprio pesquisador através de levantamento de registros (nascimentos, óbitos, notas fiscal, impostos, etc.) ou coletados diretamente por meio de inquéritos, questionários, etc. Estes dados são chamados **dados primários.**

A coleta direta pode ser classificada quanto ao fator **tempo** como:

Contínua: quando feita de forma contínua, com registro de nascimentos e óbitos, frequência de alunos às aulas, etc.

Periódica: quando feita em intervalos constantes de tempo, como censos (de 10 em 10 anos), avaliações mensais dos alunos, etc.

Ocasional: quando feitas em determinada situação para atender a um objetivo, como pesquisa de mortalidade de um rebanho, pesquisa de um produto no mercado, etc.

Dados Primários: os dados são obtidos diretamente na fonte originária (coleta direta)

Exemplo:
Tabelas do censo demográfico do IBGE.

Métodos de coleta de dados primários: É importante garantir que a coleta de dados primários seja executada de maneira estatisticamente correta, senão os resultados podem ser tendenciosos.

Observação:
O pesquisador não pergunta, observa.

Exemplo:
Pesquisa de observação para diagnosticar as necessidades de trânsito de uma cidade.

Levantamento: É o método mais comum de se coletar dados. O instrumento pode ser um questionário estruturado ou um roteiro de itens em que o entrevistado disserta à vontade sobre cada item da pesquisa.

Resumidamente as vantagens e desvantagens das três principais formas de levantamento de dados são:

Entrevista pessoal: mais flexível e muito cara.

Telefone: mais barato, penetra em segmentos difíceis, mas é de fácil recusa.

Questionário (postal, fax ou e-mail): mais lento, média de retorno das respostas muito baixa, mas sem interferência do pesquisador.

2. Coleta indireta

A **coleta indireta** é realizada a partir de elementos conhecidos, através de uma coleta direta (**dados primários**), ou do conhecimento de características relacionadas ao fenômeno estudado.

Exemplo:
Pesquisa sobre mortalidade infantil que é feita sobre a coleta direta de dados de nascimentos e óbitos.

Na coleta indireta têm-se os **dados secundários,** isto é os dados obtidos de algo já disposto (**dados primários**).

Exemplo:
Quando determinado jornal publica estatísticas referentes ao censo demográfico extraídas do IBGE.

Conheça alguns dados do Brasil, acesse o *site*: http://www.ibge.gov.br

Observação: É mais seguro trabalhar com fontes primárias. O uso da fonte secundária traz o grande risco de erros de transcrição.

d) 4ª fase - apuração dos dados: resumo dos dados através de sua contagem e agrupamento. É a condensação e tabulação de dados. É a etapa de soma e processamento dos dados obtidos mediante critérios de classificação. Pode ser **manual** ou **eletrônica** (via computador).

e) 5ª fase - apresentação dos dados: há duas formas de apresentação, que não se excluem mutuamente. A apresentação tabular, que é uma apresentação numérica dos dados em linhas e colunas distribuídas de modo ordenado, segundo regras práticas fixadas pelo Conselho Nacional de Estatística. E a apresentação gráfica dos dados numéricos, que constitui uma apresentação geométrica permitindo uma visão rápida e clara do fenômeno.

f) 6ª fase - análise e interpretação dos dados: é a última fase do trabalho estatístico, é a mais importante e delicada. Está ligada essencialmente ao cálculo de medidas e coeficientes, cuja finalidade principal é descrever o fenômeno (Estatística Descritiva).

2.4 Atividades de aprendizagem e avaliação

01. Cite quais são as fases do método estatístico.
02. Explique como ocorre a fase de coleta de dados.
03. Como se faz um planejamento através de método estatístico?

2.5 Síntese

Nessa unidade aprendemos que para o desenvolvimento de um trabalho estatístico, devemos seguir uma sequência de procedimentos e tratá-los de forma correta para chegarmos a resultados satisfatórios.

UNIDADE 3 – TABELAS

3.1 Objetivos de aprendizagem

- Conhecer e aplicar a representação tabular;
- Compreender os elementos constituintes de uma tabela.

3.2 Introdução

A representação tabular é uma apresentação numérica dos dados. Nesse tipo de representação, os dados são dispostos em linhas e colunas e distribuídos de modo ordenado, segundo algumas regras práticas ditadas pelo Conselho Nacional de Estatística e pelo IBGE.

Essa integração de valores que temos nas tabelas, nos permite ainda a utilização de representações gráficas, que normalmente são uma forma mais benéfica e elegante de demonstrar as características que estão sendo analisadas.

3.3 Elementos de uma tabela

Corpo	Conjunto de linhas e colunas que contêm informações sobre a variável em estudo.
Título	Conjunto das informações, mais completas possíveis, respondendo às perguntas: O quê? Quando? Onde?. Localizado no topo da tabela.
Cabeçalho	Parte superior da tabela que especifica o conteúdo das colunas.
Coluna Indicadora	Parte da tabela que especifica o conteúdo das linhas.
Casa ou Célula	Espaço destinado a um só número.
Fonte	Indicação da entidade responsável pelo fornecimento dos dados ou pela sua tabela.

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

1. Um traço horizontal (——) quando o valor for zero, não só quanto a natureza das coisas, como quanto ao resultado do inquérito;

2. Três pontos (...) quando não temos os dados;

3. Um ponto de interrogação (?) quando temos dúvida quanto a exatidão de determinado valor;

4. Zero (0) quando o valor é muito pequeno pra ser expresso pela unidade utilizada. Se os valores são expressos em numerais decimais, precisamos acrescentar a parte decimal um número correspondente de zeros (0.0; 0.00; 0.000;...).

Observação:
Os lados direito e esquerdo de uma tabela oficial devem ser abertos.

3.4 Atividades de aprendizagem e avaliação

01. Cite quais os elementos que fazem parte de uma tabela.

02. De acordo com o IBGE, o que devemos colocar nas casas ou células de uma tabela?

3.5 Síntese

Nessa unidade aprendemos que as tabelas têm por finalidade expor, sinteticamente e em um só local, os resultados sobre determinado assunto, de modo a obter uma visão global mais rápida daquilo que está em análise.

UNIDADE 4 – SÉRIES ESTATÍSTICAS

4.1 Objetivos de aprendizagem

- Reconhecer, através da série estatística, a existência de seus fatores;
- Empregar a nomenclatura técnica no estudo e na interpretação da série estatística.

4.2 Introdução

Uma **série estatística** é um conjunto de dados ordenados, que possuem uma característica em comum, sendo apresentadas sob forma de tabela e/ou gráfico.

Numa série estatística observa-se a existência de três elementos ou fatores: o **tempo** (cronologia), o **espaço** (lugar) e a **espécie** (fenômeno). Se houver a variação de um desses elementos, a série estatística classifica-se em **temporal, geográfica** ou **específica.**

Portanto, o nome da série depende dos elementos que variam e podem ser divididos conforme apresentado a seguir.

4.3 Tipos de séries estatísticas

4.3.1 Série temporal, histórica, cronológica ou evolutiva

É a série cujos dados estão em correspondência com o tempo, ou seja, variam com o tempo.

Exemplo:

GAÚCHA VEÍCULOS LTDA,
Vendas no primeiro bimestre de 2008

Período	Unidades vendidas*
Jan / 08	20
Fev / 08	10
Total	30

* Em mil unidades.

Fonte: dados fictícios.

4.3.2 Série geográfica, territorial ou de localidade

É a série cujos dados estão em correspondência com a região geográfica, ou seja, o elemento variável é o fator geográfico (a região).

Exemplo:

GAÚCHA VEÍCULOS LTDA,
Vendas no primeiro bimestre de 2008

Filiais	Unidades vendidas*
São Paulo	13
Rio Grande do Sul	17
Total	30

* Em mil unidades.

Fonte: dados fictícios.

4.3.3 Série específica ou categórica

É a série cujos dados estão em correspondência com a espécie, ou seja, variam com o fenômeno.

Exemplo:

GAÚCHA VEÍCULOS LTDA,
Vendas no primeiro bimestre de 2008

Marca	Unidades vendidas*
Fiat	18
Chevrolet	12
Total	30

* Em mil unidades.

Fonte: dados fictícios.

4.3.4 Séries mistas ou conjugadas

As combinações entre as séries anteriores constituem novas séries, que são denominadas séries compostas ou mistas e são apresentadas em tabelas de dupla entrada.

Exemplo:

GAÚCHA VEÍCULOS LTDA,
Vendas no primeiro bimestre de 2008

Filiais	Janeiro / 08	Fevereiro / 08
São Paulo	10	3
Rio Grande do Sul	12	5
Total	22	8

* Em mil unidades.

Fonte: dados fictícios.

4.4 Atividades de aprendizagem e avaliação

01. Qual a definição de série estatística e quais os fatores que a compõe?

02. Faça um exemplo de série geográfica.

4.5 Síntese

Nessa unidade foi possível observarmos a existência dos fatores que fazem a distinção das séries estatísticas, os quais são de fundamental importância para a sua compreensão.

UNIDADE 5 – GRÁFICOS ESTATÍSTICOS

5.1 Objetivos de aprendizagem

- Conhecer e identificar os tipos de gráficos estatísticos;
- Interpretar e comparar os valores colocados nos gráficos.

5.2 Introdução

Gráfico é toda a forma de representação das séries estatísticas que se baseia no desenho, devendo ser atraente para cumprir sua finalidade de mostrar os resultados. Assim como, deve ser bem construído a fim de que permita a análise do fenômeno exposto. Enfim, podemos dizer que os gráficos são representações visuais dos dados estatísticos, cujo objetivo é produzir uma impressão mais rápida e viva do fenômeno em estudo.

Os gráficos devem ser correspondentes às tabelas estatísticas, mas nunca substituí-las. Assim, a apresentação gráfica é um complemento importante da apresentação tabular.

A vantagem de um gráfico sobre a tabela está em possibilitar uma rápida impressão visual da distribuição dos valores ou das frequências observadas.

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

Simplicidade: o gráfico deve ser destituído de detalhes e traços desnecessários.

Clareza: o gráfico deve possuir uma correta interpretação dos valores representativos do fenômeno em estudo.

Veracidade: o gráfico deve expressar a verdade sobre o fenômeno em estudo.

5.3 Formas de apresentação dos gráficos

5.3.1 Gráficos de informação

São gráficos destinados principalmente ao público em geral, objetivando proporcionar uma visualização rápida e clara. São gráficos tipicamente expositivos, dispensando comentários explicativos adicionais. As legendas podem ser omitidas, desde que as informações desejadas estejam presentes.

5.3.2 Gráficos de análise

São gráficos que servem principalmente ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser também informativos. Os gráficos de análise frequentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes um texto explicativo, chamando a atenção do leitor para os pontos principais revelados pelo gráfico.

5.4 Principais tipos de gráficos

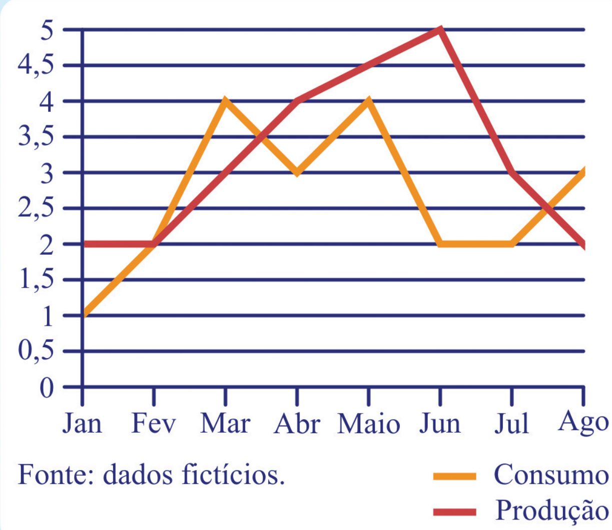
5.4.1 Gráfico de pontos

5.4.2 Gráfico em curvas ou em linhas

Neste tipo de gráfico utiliza-se uma linha poligonal para representar séries temporais, principalmente quando a série cobrir um grande número de períodos de tempo.

Exemplo:

Relação entre consumo e produção de óleo combustível no RS no período compreendido entre Jan /Ago de 2006 (em milhões de barris):



Fonte: dados fictícios.

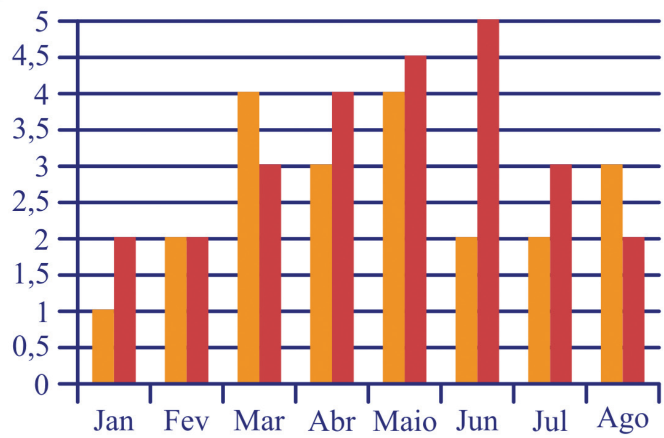
5.4.3 Gráfico de superfície

5.4.4 Gráfico em colunas (simples sobrepostas e justapostas)

É a representação de uma série estatística através de retângulos, dispostos em colunas (na vertical) ou em retângulos (na horizontal) e pode ainda conter barras múltiplas. Este tipo de gráfico representa praticamente qualquer série estatística.

Exemplo:

Relação entre consumo e produção de óleo combustível no RS no período compreendido entre Jan/Ago de 2006 (em milhões de barris):



Fonte: dados fictícios.

Observação:

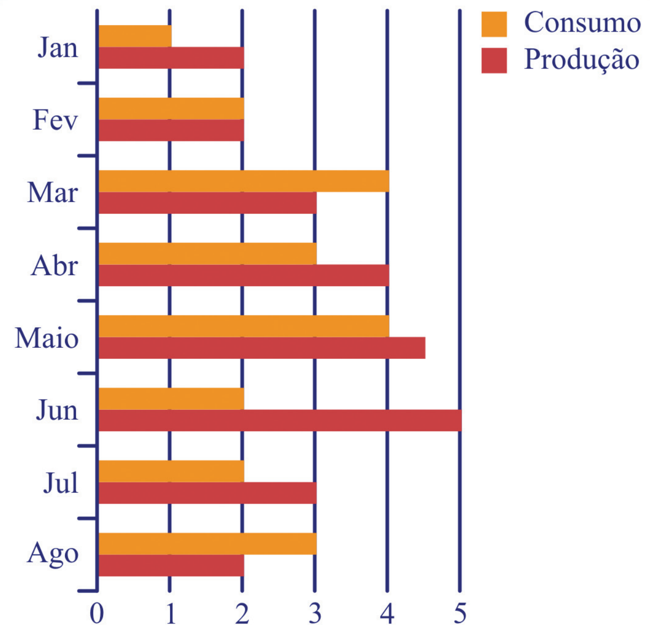
As regras para a construção são as mesmas do gráfico em curvas. As bases das colunas são iguais e as alturas são proporcionais aos respectivos dados. O espaço entre as colunas pode variar de 1/3 a 2/3 do tamanho da base da coluna.

5.4.5 Gráfico em barras (simples, sobrepostas e justapostas)

É representado por retângulos dispostos horizontalmente, prevalecendo os mesmos critérios adotados na elaboração de gráfico em colunas.

Exemplo:

Relação entre consumo e produção de óleo combustível no RS no período compreendido entre Jan /Ago de 2008 (em milhões de barris):



Fonte: dados fictícios.

5.4.6 Gráfico em setores

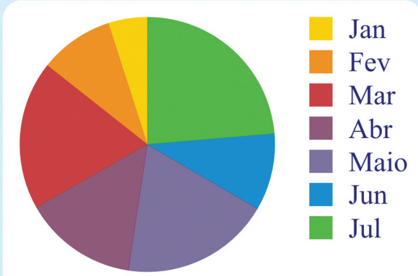
É a representação gráfica de uma série estatística em um círculo de raio qualquer, por meio de setores com ângulos centrais proporcionais às ocorrências. É utilizado quando se pretende comparar cada valor da série com o total.

O total da série corresponde a 360° (total de graus de um arco de circunferência). O gráfico em setores representa valores absolutos ou porcentagens complementares.

As séries geográficas, específicas e as categorias em nível nominal são mais representadas em gráficos de setores, desde que não apresentem muitas parcelas (no máximo sete).

Exemplo:

Consumo de óleo combustível no RS no período compreendido entre Jan /Jul de 2007 (em milhões de barris):



Fonte: dados fictícios.

Observação:

Para construir este gráfico, cada setor será expresso graficamente em graus (ângulo do setor) e a porcentagem calculada através de uma regra de três:

$$\text{Total } (\%) \longrightarrow 360^{o}$$
$$\text{Parte } (\%) \longrightarrow x^{o}$$

Exemplo:

No ano passado, certo produtor rural, após anos de pouca lucratividade, resolveu diversificar sua produção. Atualmente, 25% da sua renda vem da criação de suínos, 18% da psicultura, 31% do leite, 17% do cultivo de hortaliças e 9% da fruticultura. Represente sua produção em um gráfico de setores.

25% suínos: $100 . x = 360 . 25 \rightarrow x = 9000 \div 100 \rightarrow x = 90^{o}$

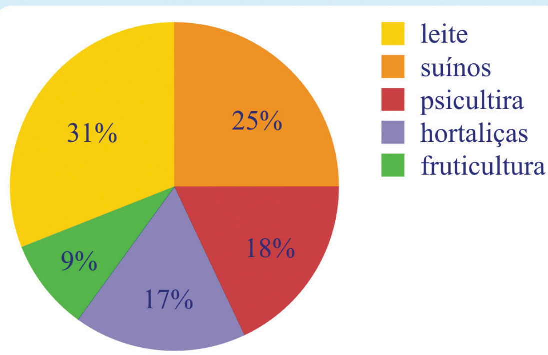
18% psicultura: $100 . x = 360 . 18 \rightarrow x = 6480 \div 100 \rightarrow x = 64.8^{o}$

31% leite: $100 . x = 360 . 31 \rightarrow x = 11160 \div 100 \rightarrow x = 111.6^{o}$

17% hortaliças: $100 . x = 360 . 17 \rightarrow x = 6120 \div 100 \rightarrow x = 61.2^{o}$

9% fruticultura: $100 . x = 360 . 9 \rightarrow x = 3240 \div 100 \rightarrow x = 32.4^{o}$

Assim temos:



5.5 Atividades de aprendizagem e avaliação

01. Explique os tipos de gráficos e dê um exemplo de cada um deles.
02. Qual a principal vantagem de um gráfico em relação a uma tabela?
03. A representação gráfica deve obedecer a certos requisitos. Quais são eles?

5.6 Síntese

Nesta unidade foram estudados os gráficos, os quais 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.

UNIDADE 6 – DISTRIBUIÇÃO DE FREQUÊNCIAS

6.1 Objetivos de aprendizagem

- Compreender a representação dos dados (amostrais e populacionais);
- Compreender e reconhecer os elementos de uma distribuição de frequência;
- Conhecer os tipos de frequências;
- Compreender os gráficos representativos de uma distribuição de frequência.

6.2 Introdução

Um conjunto de observações de certo fenômeno, não estando adequadamente organizado, fornece pouca informação de interesse ao pesquisador e ao leitor. Para uma visão rápida e global do fenômeno em estudo é preciso que os dados estejam organizados. Um dos primeiros passos em uma análise de dados é organizar, condensar, resumir e comunicar a informação obtida.

A fim de facilitar o entendimento do conteúdo, no decorrer desta unidade usaremos o mesmo exemplo em todos os tópicos.

6.3 Representação dos dados (amostrais ou populacionais)

As tabelas, chamadas primitiva e rol, são utilizadas na representação de dados que não estão organizados numericamente (dados brutos).

Dados brutos: são aqueles que não foram numericamente organizados, ou seja, estão na forma com que foram coletados.

Exemplo:

Considere o levantamento de dados da idade de 50 trabalhadores da indústria A (variável x), cujos resultados, em anos, mostrados na tabela a seguir, estão colocados na sequência como foram obtidos.

Idade de 50 trabalhadores da indústria A

29	40	21	20	42	20	25	37	24	20
22	21	39	23	26	33	30	45	34	48
45	22	33	40	45	45	49	41	30	24
34	41	46	22	33	37	26	28	28	41
25	31	35	23	38	44	22	23	46	47

O primeiro passo para a organização dos dados é ordená-los de forma crescente ou decrescente. A tabela assim organizada recebe o nome de **rol**.

20	21	23	25	29	33	37	40	44	46
20	22	23	26	30	33	37	41	45	46
20	22	24	26	30	34	38	41	45	47
20	22	24	28	31	34	39	41	45	48
21	23	25	28	33	35	40	42	45	49

A simples organização dos dados em um rol de ordem crescente já permite determinar diretamente o menor valor (x = 20 anos), o maior valor (x = 49 anos), e a amplitude da variação (a distância entre o maior e o menos, $\Delta x = 49 - 20 = 29$ anos).

Assim temos:

20, 20, 20, 20, 21, 21, 22, 22, 22, 23, 23, 23, 24, 24, 25, 25, 26, 26, 28, 28, 29, 30, 30, 31, 33, 33, 33, 34, 34, 35, 37, 37, 38, 39, 40, 40, 41, 41, 41, 42, 44, 45, 45, 45, 46, 46, 47, 48, 49.

Exemplo:

Idade	Freq.	Idade	Freq.	Idade	Freq.
20	4	30	2	42	1
21	2	33	3	44	1
22	3	34	2	45	4
23	3	35	1	46	2
24	2	37	2	47	1
25	2	38	1	48	1
26	2	39	1	49	1
28	2	40	2		
29	1	41	3		

6.4 Distribuições de frequências

Uma maneira mais concisa de mostrar os dados do rol é apresentar cada dado seguido pelo número de vezes que ocorre, ao invés de repeti-los. O número de ocorrências de um determinado valor recebe o nome de **frequência**.

Exemplo:
A idade de 20 anos ocorre 4 vezes, assim temos $f(20) = 4$. Já a idade de 46 anos ocorre 2 vezes, que escreve-se $f(46) = 2$.

A tabela que contém todos os valores, cada um com sua respectiva frequência, recebe o nome de **distribuição de frequência**.

Ainda assim, o processo exige muito espaço, em especial quando o número de valores da variável (n) aumenta. O mais razoável nestes casos, quando a variável é contínua, é agrupar os valores por intervalos. Deste modo, ao invés de listar cada um dos valores que ocorrem, listam-se os intervalos de valores e a frequência correspondente.

Exemplo:
Ao invés de colocar 1 trabalhador com 20 anos, 1 trabalhador com 21 anos, etc, coloca-se, por exemplo, 4 trabalhadores entre 20 e 24 anos. Este intervalo é escrito como 20 ⊢ 24 que corresponde a $20 \leq x < 24$ (a variável pode estar desde 20 inclusive até 24 exclusive), portanto valores 20 e 23, mas 24 não. Definindo o rol de acordo com intervalos, tem-se a seguinte tabela:

Estatura de 40 alunos da escola A.

Idades	Freqüência
20 ⊢ 24	12
24 ⊢ 28	6
28 ⊢ 32	6
32 ⊢ 36	6
36 ⊢ 40	4
40 ⊢ 44	6
44 ⊢ 48	8
48 ⊢ 52	2
Total	50

Fonte: dados fictícios.

Procedendo desta forma perde-se a informação detalhada das idades, mas se ganha em praticidade, pois a análise dos dados fica simplificada. Examinando a tabela acima, podemos facilmente verificar que a maioria dos trabalhadores tem idade entre 20 e 23 anos e que uma minoria é maior que 48 anos.

" — ":
Não inclui os valores da direita nem os da esquerda.

" ⊣ ":
Inclui o valor da direita mas não o da esquerda.

" ⊢ ":
Inclui o valor da esquerda mas não o da direita.

" ⊢⊣ ":
Inclui tanto os valores da direita quanto os da esquerda.

Frequentemente procedemos desta forma numa análise estatística, pois o objetivo da Estatística é justamente fazer o apanhado geral das características de um conjunto de dados, desinteressando-se por casos particulares.

6.5 Elementos de uma distribuição de frequência

6.5.1 Classe

As **classes** são intervalos de variação de uma variável. As classes são representadas simbolicamente por i, sendo i = 1; 2; .... ; k, onde k é o número total de classes. O número total de valores é simbolizado por n.

Exemplo:
Assim, o intervalo 28 ⊢ 32 define a terceira classe ($i = 3$), o intervalo 36 ⊢ 40 define a quinta classe (i = 5) e assim por diante. Como a distribuição tem oito classes, logo k = 8, a variável x assume 50 valores, logo $n = 50$.

6.5.2 Limites de classe

Os **limites de classe** são os extremos de cada classe. Para uma determinada classe i, o **limite inferior** é simbolizado por $\boldsymbol{l_i}$ e o **limite superior** por $\boldsymbol{L_i}$.

Exemplo:
O limite inferior da segunda classe do exemplo é escrito como $l_2 = 24$, enquanto o limite superior da segunda classe é escrito como $L_2 = 28$.

De acordo com o IBGE as classes devem ser escritas como desta quantidade até menor que aquela, usando para isso o símbolo "⊢". Assim, $l_i \vdash L_i$ significa inclusão de $\boldsymbol{l_i}$ e exclusão de $\boldsymbol{L_i}$.

Exemplo:
O trabalhador com idade de 28 anos, estaria na quarta classe ($i = 3$) e não na segunda.

6.5.3 Intervalo de classe

A **amplitude de um intervalo de classe** ou simplesmente **intervalo de classe** é o tamanho do intervalo que define a classe. O intervalo da classe i é simbolizado por h_i e é obtido pela diferença entre os seus limites:

$$h_i = L_i - l_i$$

Exemplo:

O tamanho do intervalo da terceira classe (h_2) vale:

$$h_2 = L_2 - l_2 = 28 - 24 = 4 \text{ anos.}$$

Todas as classes do exemplo também têm intervalo de 4 anos, pois este é o intervalo entre cada um dos limites inferiores e os limites superiores correspondentes.

6.5.4 Amplitude total da distribuição

A **amplitude total** da distribuição (AT) é o intervalo total compreendido por todas as classes da distribuição, isto é, desde o limite inferior da primeira classe (l_1) até o limite superior da última classe (L_k). Matematicamente, escrevemos isso como:

$$AT = L_k - l_1$$

Exemplo:

Ainda no nosso exemplo, temos oito classes ($k = 8$). O limite superior da última classe ($i = 8$) vale $L_8 = 52$, enquanto o limite inferior da primeira classe ($i = 1$) vale $l_1 = 20$. Portanto, substituindo os valores na fórmula, temos:

$$AT = L_6 - l_1 = 52 - 20 = 32 \text{ anos.}$$

Numa distribuição em que as classes possuem o mesmo intervalo, a amplitude total pode ser escrita como o intervalo de classe multiplicado pelo número de classes:

$$AT = h_i \,.\, k$$

6.5.5 Amplitude amostral

A **amplitude amostral** (AA) é o intervalo entre o maior valor $(\text{máx}(x))$ e o menor valor $(\text{mín}(x))$ dos dados da amostra:

$$AA = \text{máx}(x) - \text{mín}(x)$$

Exemplo:
A maior idade é de 49 e a menor 20, logo:

$$AA = 49 - 20 = 29 \text{ anos.}$$

6.5.6 Ponto médio de uma classe

O **ponto médio de uma classe** é o ponto que divide a classe ao meio. O ponto médio da classe i é simbolizado por x_i e calculado efetuando-se a média entre os limites da classe:

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

Exemplo:
O ponto médio da segunda classe é:

$$x_i = \frac{l_i + L_i}{2} = \frac{24 + 28}{2} = 26 \text{ anos}$$

O ponto médio de uma classe é o valor representativo da classe.

6.6 Tipos de frequências

6.6.1 Frequência simples ou absoluta

A frequência, simples ou absoluta (f_i), de uma classe ou de um valor individual é o número de vezes que o valor ocorre numa amostra.

Exemplo:
Idade de 50 trabalhadores da indústria A:

$$f_1 = 4;\ f_2 = 9;\ f_3 = 11;\ f_4 = 8;\ f_5 = 5;\ f_6 = 3;\ f_7 = 3;\ f_8 = 3.$$

A soma de todas as frequências é representada pelo símbolo de somatório (Σ). Onde $\sum_{i=1}^{k} f_i$ significa a soma dos f_i, sendo que i vai desde 1 até k. Pode-se entender que a soma de todas as frequências é igual ao número total de valores na amostra:

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

Quando não há dúvidas, podemos escrever simplesmente:

$$\sum f_i = n$$

Exemplo:

Assim, escrever $\sum_{i=1}^{8} f_i$ é como escrever $f_1+f_2+f_3+f_4+f_5+f_6+f_7+f_8$, ou seja:

$$\sum_{i=1}^{8} f_i = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8$$

$$12 + 6 + 6 + 6 + 4 + 6 + 8 + 2 = 50$$

Neste ponto podemos reescrever a distribuição de frequência com a seguinte representação técnica da tabela:

Estatura de 40 alunos da escola A.

i	Estaturas (cm)	f_i
1	20 ⊢ 24	12
2	24 ⊢ 28	6
3	28 ⊢ 32	6
4	32 ⊢ 36	6
5	36 ⊢ 40	4
6	40 ⊢ 44	6
7	44 ⊢ 48	8
8	48 ⊢ 52	2
		$\sum f_i = 50$

Fonte: dados fictícios.

6.6.2 Frequência relativa

Frequências relativas (fr_i) são as razões entre as frequências simples (fi) e a frequência total (n), e são geralmente expressas em percentuais:

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

A frequência relativa de uma classe mostra a parcela que aquela classe representa na amostra.

Exemplo:

A frequência relativa da terceira classe é:

$$fr_3 = \frac{f_3}{\sum f_i} = \frac{6}{50} = 0.12$$

Então a terceira classe corresponde a uma fração de 0.12 do total ou 12 %.

6.6.3 Frequência acumulada

A frequência acumulada (F_i) é a soma das frequências simples de todas as classes com intervalos inferiores a uma determinada classe. Isto é, corresponde ao total (acumulado) das frequências absolutas observadas até o nível em questão (inclusive) que é representado pela letra " j ":

$$F_i = \sum_{i=1}^{j} f_i = f_1 + f_2 + \ldots + f_j$$

Exemplo:

A frequência acumulada correspondente à terceira classe é:

$$F_3 = \sum_{i=1}^{3} f_i = f_1 + f_2 + f_3 = 12 + 6 + 6 = 24$$

Significa que existem 24 alunos com idade inferior a 32 anos (limite superior da terceira classe).

6.6.4 Frequência relativa acumulada

Frequência acumulada relativa (FR_i) é a frequência acumulada da classe dividida pela frequência total da distribuição:

$$FR_i = \frac{F_i}{\sum f_i} = \frac{F_i}{n}$$

Exemplo:

Para a terceira classe, temos:

$$FR_3 = \frac{F_3}{n} = \frac{24}{50} = 0.4$$

Significa que a fração de 0.4 alunos (ou 40%) têm idades inferiores a 32 anos (limite superior da terceira classe) .

Exemplo:
A tabela completa fica assim:

Idade de 50 trabalhadores da indústria A.

i	Idades (anos)	x_i	f_i	fr_i	F_i	FR_i
1	20 ⊢ 24	22	12	0.24	12	0.24
2	24 ⊢ 28	26	6	0.12	18	0.36
3	28 ⊢ 32	30	6	0.12	24	0.48
4	32 ⊢ 36	34	6	0.12	30	0.60
5	36 ⊢ 40	38	4	0,08	34	0.68
6	40 ⊢ 44	42	6	0.12	40	0.80
7	44 ⊢ 48	46	8	0.16	48	0.96
8	48 ⊢ 52	50	2	0.04	50	1.00
			$\sum f_i = 50$	$\sum fr_i = 1$		

x_i: Ponto médio de uma classe;
f_i: Frequência simples;
fr_i: Frequenência relativa;
F_i: Frequência acumulada;
FR_i: Frequência relativa acumulada.

Fonte: dados fictícios.

Examinando a tabela, vemos por exemplo que a terceira classe corresponde a maior fração de trabalhadores ($fr_1 = 12$), isto é, a maioria dos trabalhadores têm idades entre 20 anos (inclusive) e 24 anos (excluindo). Também é possível ver que 80% dos alunos têm idade inferior a 44 anos pois a freqüência acumulada até a quarta classe ($F_6 = 40$) é 0.80 que corresponde a 80 %.

6.7 Determinação do número de classes e intervalos de classe

Quando dispomos de uma tabela primitiva ou de um rol, precisamos estabelecer a quantidade e o intervalo das classes que vamos criar, isto é, para que a distribuição de frequência seja útil para a nossa análise.

Uma das maneiras de determinar o número de classes é pelo uso da **Regra de Sturges,** que determina k em função de n:

$$k \cong 1 + 3{,}3 \log(n)$$

Onde k é o número de classes e n o número de dados. Da mesma forma, podemos usar outra regra que associa k e n de outra forma:

$$k \cong \sqrt{n}$$

Exemplo:

Usando a Regra de Sturges temos $n = 50$, logo $k = 1 + 3{,}3 \log (50) = 6.606 \cong 7$. Desta maneira utilizamos 7 classes. Com a outra regra, temos $k = \sqrt{50} = 7.07 \cong 7$, cujo resultado para o número de classes é o mesmo.

Para facilitar o cálculo do número de classes pela Regra de Sturges, utilize a tabela abaixo:

n = nº total de observações	k = nº de classes a usar
1	1
2	2
3 |—| 5	3
6 |—| 11	4
12 |—| 22	5
23 |—| 46	6
47 |—| 90	7
91 |—| 181	8
182 |—| 362	9
363 |—| 724	10
725 |—| 1448	11
1449 |—| 2896	12

Observação:

Qualquer regra para determinação do número de classes da tabela não nos leva a uma decisão final. Esta vai depender, na realidade, de um julgamento pessoal, que deve estar ligado a natureza dos dados.

Sabendo o número de classes *(k)* que vamos usar, podemos determinar o intervalo de classes *(H)* através da amplitude total da distribuição *(AT)*:

$$H \cong \frac{AT}{k}$$

Nas equações acima foi usado o símbolo de aproximadamente (≅)

ao invés de igualdade (=), porque estas fórmulas representam valores típicos a serem usados, mas que podem ser alterados, ligeiramente, de acordo com o objetivo da distribuição ou para evitar classes com frequências nulas enquanto outras têm valores muito altos. Com relação ao intervalo de classe, lembre-se que a amplitude total *(AT)* deve ser ligeiramente maior que a amplitude amostral *(AA)* para que a distribuição tenha intervalos para incluir todos os valores da amostra.

6.8 Gráficos representativos de uma distribuição de frequências em classes

Feita a coleta de dados, seja através de censo, de levantamento por amostragem, ou de experimento, geralmente esses dados apresentam-se de maneira desorganizada, devendo ser organizadas e resumidas para possibilitar a obtenção de informações úteis. Uma maneira de organizar os dados são as distribuições de frequências, que podem ser representadas além da forma de tabela, também através de gráficos.

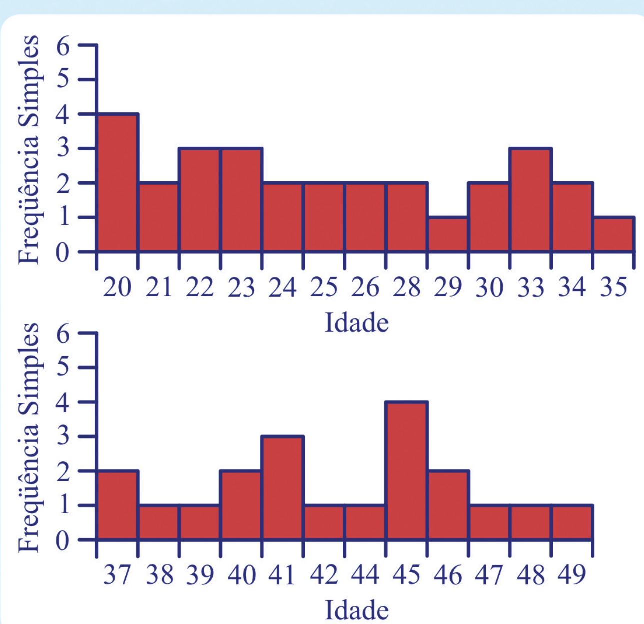
6.8.1 Histograma

O **histograma** é formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal (eixo *x*). Onde são representados os intervalos de classe numa escala contínua, não sendo necessário que a escala inicie do zero, de tal modo que os seus pontos médios coincidam com os pontos médios dos intervalos de classe e seus limites coincidam com os limites da classe.

Um histograma para a frequência simples é mostrado a seguir.

Exemplo:

Histograma das idades dos trabalhadores
da indústria A entre 20 e 49 anos



Observação:
No histograma, a base do retângulo é proporcional a amplitude do intervalo classe, a altura proporcional a frequência da classe e a sua área é proporcional à soma das frequências.

6.8.2 Polígono de frequências

É um gráfico, em linha, de uma distribuição de frequências. Esse gráfico é obtido unindo-se, por segmentos de reta, os pontos médios das bases superiores dos retângulos de um histograma. Podendo também ser feito da mesma forma para frequências acumuladas.

6.9 Atividades de aprendizagem e avaliação

01. O que são dados brutos?
02. Diferencie os elementos de uma distribuição de frequência e dê exemplos de cada um deles.
03. Cite e explique os tipos de frequência.
04. Explique o significado da regra de Sturges e dê um exemplo.
05. Através de um exemplo de histograma, faça sobre ele um polígo-
de frequência .

6.10 Síntese

Ao longo desta unidade estudamos que uma das formas de organizar os dados coletados em uma pesquisa pode ser através das tabelas de distribuições de frequências ou de gráficos, as quais permitem uma síntese adequada e uma melhor compreensão dos resultados.

UNIDADE 7 – MEDIDAS DESCRITIVAS

7.1 Objetivos de aprendizagem

- Reconhecer as medidas descritivas básicas;
- Aplicar as medidas descritivas, de acordo com a utilização exigida;
- Analisar e interpretar os procedimentos e fórmulas utilizadas nas medidas descritivas, diferenciando-as.

7.2 Introdução

Na maior parte das vezes em que os dados estatísticos são analisados, procuramos obter um valor para representar um conjunto de dados. Este valor deve sintetizar, da melhor maneira possível, o comportamento do conjunto do qual ele é originário.

Com isto, a Estatística Descritiva visa descrever os dados disponíveis da forma mais completa possível sem, no entanto, se preocupar em tirar conclusões sobre um conjunto maior de dados.

As medidas descritivas fazem parte da Estatística Descritiva e elas indicam os valores em torno dos quais ocorre a maior concentração do fenômeno em estudo.

As medidas descritivas básicas mais importantes são as de:

> Posição, dispersão ou variabilidade, momentos, assimetria e curtose.

As medidas de **posição** se dividem em medidas de tendência central ou promédias (verifica-se uma tendência dos dados observados a se agruparem em torno dos valores centrais) e separatrizes.

As **medidas de tendência central** mais utilizadas são: média aritmética, moda e mediana. Outras promédias menos usadas são as médias: geométrica, harmônica, quadrática, cúbica e bi quadrática.

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

7.3 Medidas de posição

7.3.1 Medidas de tendência central

Na utilização de dados numéricos observa-se uma tendência destes, se agruparem, em torno de um valor central, indicando que algum valor central é característica dos dados e que o mesmo pode ser usado para descrevê-los e representá-la.

As medidas de tendência central são insuficientes para descrever adequadamente uma amostra, pois é necessário também descrever em que medida os dados de observações estão agrupados ao redor da média.

7.3.1.1 Média aritmética

Simbolizada por $\overline{x}$ (x barra) no caso de amostra e μ (mi) para população, é a medida de tendência central mais utilizada para descrever um conjunto de dados.

a) Média aritmética para dados não tabelados

Consiste na soma de todas as observações (x_i) dividida pelo número (n) de observações de um determinado grupo:

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

b) Média aritmética para dados tabelados

Quando os dados estiverem agrupados em uma tabela de frequências, pode-se obter a média aritmética da distribuição, calculando-se:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum_{i=1}^{n} f_i}$$

onde:

x_i = ponto médio da classe i
f_i = frequência absoluta da classe i

Exemplos:

1. A média do conjunto $x_i = \{1; 2; 3; 4; 5\}$ é $\overline{x} = 3$, enquanto a média do conjunto $y_i = x_i + 10 = \{11; 12; 13; 14; 15\}$ é $\overline{y} = \overline{x} + 10 = 13$
2. A média do conjunto $x_i = \{1; 2; 3; 4; 5\}$ é $\overline{x} = 3$, enquanto a média do conjunto $y_i = 10 . x_i = \{10; 20; 30; 40; 50\}$ é $\overline{y} = 10 . \overline{x} + 10 = 30$
3. A média simples de produção de uma vaca ao longo de 7 dias é 10; 14; 13; 15; 16; 18; 12 litros de leite. Cada dia vale:

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{10+14+13+15+16+18+12}{7} = \frac{98}{7} = 14$$

Assim, uma outra vaca que produzisse 14 litros de leite em todos os 7 dias teria produzido, no final, o mesmo que esta vaca cuja produção tenha sido 10; 14; 13; 15; 16; 18; 12.

4. Os desvios da média dos valores 10; 14; 13; 15; 16; 18; 12 da produção de leite de uma determinada vaca são:

$d_1 = x_1 - \overline{x} = 10 - 14 = -4$ $\quad d_5 = x_5 - \overline{x} = 16 - 14 = 2$
$d_2 = x_2 - \overline{x} = 14 - 14 = 0$ $\quad d_6 = x_6 - \overline{x} = 18 - 14 = 4$
$d_3 = x_3 - \overline{x} = 13 - 14 = -1$ $\quad d_7 = x_7 - \overline{x} = 12 - 14 = -2$
$d_4 = x_4 - \overline{x} = 15 - 14 = 1$

Somando todos os desvios:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \overline{x}) = -4 + 0 - 1 + 1 + 2 + 4 - 2 = 0$$

conforme esperado pela propriedade.

7.3.1.2 Moda

A **moda** *(Mo)* ou norma é o valor que ocorre com maior frequência em um conjunto de valores, sendo uma medida de dominância, não sendo afetada por valores extremos.

Quando os dados estão agrupados em classes, a moda corresponde à frequência simples mais alta e o valor da moda é tomado como o ponto médio do intervalo da classe. Se os limites inferior e superior da classe mais frequente são (l_*) e (L_*), moda será calculada por: $Mo = (l_* + L_*) / 2$.

Exemplo:
Considere os seguintes salários: 100, 90, 110, 100, 100, 2500 e 3000. A moda é o valor que mais ocorre, portanto $Mo = 100$. Neste caso a média é $\overline{x} = 500$.

a) Distribuição modal: é aquela que possui uma só moda.

Exemplo:

$$x = \{100, 90, 110, 100, 100, 2500\} \quad Mo = 100$$

b) Distribuição bimodal ou plurimodal: a série possui dois ou mais valores modais.

Exemplo:

$$x_i = \{100, 200, 100, 100, 150, 210, 200, 120, 200\}$$
$$Mo = 100 \text{ e } Mo = 200$$

c) Distribuição amodal: ocorre quando nenhum valor é repetido, isto é, a distribuição não possui moda.

Exemplo:

$$x_i = \{1, 2, 3, 6, 7, 22, 300\} \ \nexists \ Mo$$

7.3.1.3 Mediana

A **mediana** ou **valor mediano** *(Md)* é o valor que ocupa a posição central, quando todos os itens do grupo estão dispostos em termos de valor, em ordem crescente ou decrescente de magnitude. Ela divide a série ordenada em dois conjuntos com o mesmo número de valores. Se a série tem um número ímpar de valores, a mediana é o valor que está no meio (ponto mediano) da série.

Exemplo:

Na série ordenada $\{2; 5; 6; 8; 10; 13; 15; 16; 18\}$, temos que $Md = 10$, pois abaixo de 10 temos 4 números (2, 5, 6, 8) e acima de 10 também 4 (13, 15, 16, 18).

Se a série tem um número par de valores, então, utiliza-se como mediana, o valor médio entre os dois valores que estão no meio da série. Desta forma, quando o número de observações for par, deve-se somar os dois números centrais e dividir por dois:

$$\text{Posição mediana} = \frac{(n+1)}{2}$$

Exemplos:

1. Na série ordenada $\{1; 3; 6; 8; 9; 10\}$, temos que $Md = (6+8) \div 2 = 7$, pois não há um só número no centro da série, assim utilizamos o valor médio dos dois números centrais.
2. Uma empresa deseja comparar o desempenho de dois empregados, com base na produção diária de uma determinada peça durante cinco dias:

   Empregado A : 70, 71, 69, 70, 70, sendo $\bar{x} = 70$

   Empregado B : 60, 80, 70, 62, 83, sendo $\bar{x} = 71$

   O desempenho médio do empregado A é de 70 peças produzidas diariamente, enquanto que o do empregado B é de 71 peças. Com base na média aritmética, verifica-se que o desempenho de B é melhor do que o de A. Porém, observando bem os dados, percebe-se que a produção de A varia apenas de 69 a 71 peças, ao passo que a de B varia de 60 a 83 peças, que revela que o desempenho de A é bem mais uniforme do que de B.

	Definição	Variação	Desvantagens
Média	Centro de distribuição de frequências.	Reflete cada valor.	É afetada por valores extremos.
Moda	Valor mais frequente.	Valor "típoco". Maior quantidade de valores concentrados nesse ponto.	Não se presta para análise matemática. Pode não ter moda para certos conjuntos de dados.
Mediana	Metade dos valores são maiores, metade são menores.	Menos sensível a valores extremos do que a média.	Difíceis de determinar para grandes quantidades de dados.

7.4 Separatrizes

Além das medidas de posição que estudamos, há outras que, consideradas individualmente, não são medidas de tendência central, mas estão ligadas à mediana relativamente à sua característica de separar a série em duas partes que apresentam o mesmo número de valores.

As medidas (**quartis, decis e percentis)** são, juntamente com a **mediana**, conhecidas pelo nome genérico de **separatrizes**.

Enquanto a mediana separa a distribuição em duas partes iguais, as outras separatrizes dividem a distribuição da seguinte maneira:

Quartis: dividem a distribuição em quatro partes iguais;
Decis: dividem em dez partes iguais;
Percentis: dividem em cem partes iguais.

Notações:
Q_i : quartil de ordem i;
D_i : decil de ordem i;
P_i : percentil de ordem i.

7.4.1 Quartis

São os valores de uma série que subdividem uma distribuição de medidas quando dispostas em termos de valores em ordem crescente ou decrescente, em quatro partes iguais.

Precisamos, portanto de 3 quartis (Q_1, Q_2 e Q_3) para dividir a série em quatro partes iguais.

Observação:
O segundo quartil (Q_2) será sempre igual a mediana da série.

a) Quartis para dados não agrupados em classes

Os quartis dividem um conjunto de dados em 4 partes iguais. Assim, para o cálculo das posições usaremos:

Primeiro Quartil (Q_1 ou P_{25}) é o número da série tal que um quarto (25%) dos dados está abaixo dele e as três quartas partes restantes (75%) estão acima dele. Para encontrar a posição do Q_1 emprega-se:

$$Q_1 \rightarrow p_1 = \frac{n+1}{4}$$

Segundo Quartil (Q_2) é com evidência, coincidente com a mediana $(Q_2 = Md)$. A posição do Q_2 é obtida por:

$$Q_2 \rightarrow p_2 = \frac{2(n+1)}{4} = \frac{(n+1)}{2}$$

Terceiro Quartil (Q_3 ou P_{75}) é o número da série tal que três quartos (75%) dos dados estão abaixo dele e uma quarta parte (25%), está acima dele.

Calcula-se a posição do Q_3 como:

$$Q_3 \rightarrow p_3 = \frac{3(n+1)}{4}$$

Onde n = número de dados (valores)

Exemplos:

1. Considere as seguintes notas de uma turma e calcule os quartis dessa amostra:

52,0	55,9	56,7	59,4	60,2	54,4	55,9	56,8	59,4	60,3
54,5	56,2	57,2	59,5	60,5	55,7	56,4	57,6	59,8	60,6
55,8	56,4	58,9	60,0	60,8					

Utilizando as relações para a posição dos quartis, temos:

$$Q_1 \rightarrow p_1 = \frac{25+1}{4} = p_1 = 6{,}5 \cong 7$$

$$Q_2 \rightarrow p_2 = \frac{2(25+1)}{4} = p_2 = 13$$

$$Q_3 \rightarrow p_3 = \frac{3(25+1)}{4} = p_3 = 19{,}5 \cong 20$$

O primeiro quartil (Q_1) abrange 25% dos valores da série, o segundo quartil (Q_2) 50%, e o terceiro (Q_3) 75%.

Logo em $\{2, 5, 6\}$ a mediana é 5. Ou seja: será o quartil 1(Q_1).
Em $\{10, 13, 15\}$ a mediana é 13. Ou seja: será o quartil 3 (Q_3).

2. Calcule os quartis da série: $\{1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13\}$.
Como a série já está ordenada, calcularemos o quartil 2:

$$Q_2 = Md = (5 + 6) \div 2 = 5.5$$

O quartil 1 será a mediana da série à esquerda de Md: $\{1, 1, 2, 3, 5, 5\}$.

$$Q_1 = (2 + 3) \div 2 = 2.5$$

O quartil 3 será a mediana da série à direita de Md: $\{6, 7, 9, 9, 10, 13\}$.

$$Q_3 = (9 + 9) \div 2 = 9$$

7.4.2 Decis

São valores que dividem a série em 10 partes iguais. Assim, para dados não agrupados o cálculo das posições (decis) usaremos:

$$D_1 \rightarrow p_1 = \frac{n+1}{10}$$

$$D_2 \rightarrow p_2 = \frac{2(n+1)}{10}$$

$$\vdots$$

$$D_5 \rightarrow p_5 = \frac{5(n+1)}{10}$$

$$\vdots$$

$$D_9 \rightarrow p_9 = \frac{9(n+1)}{10}$$

Onde n = número de dados (valores)

Exemplo:
Considere as seguintes notas de uma turma e calcule os decis dessa amostra:

52,0	55,9	56,7	59,4	60,2	54,4	55,9	56,8	59,4	60,3
54,5	56,2	57,2	59,5	60,5	55,7	56,4	57,6	59,8	60,6
55,8	56,4	58,9	60,0	60,8					

Utilizando as relações para a posição dos decis, temos:

$$p_1 = 3 \quad p_2 = 5 \quad p_5 = 13 \quad p_9 = 23$$

O primeiro decil (D_1) abrange 10% dos termos da série, o segundo decil (D_2) 20% e assim por diante, o nono (D_9) 90%.

7.4.3 Percentil ou centil

São os valores que delimitam proporções dentro de uma série segundo o percentil escolhido. Assim, para dados não agrupados, o cálculo das posições (percentis) usaremos:

$$P_1, P_2, \ldots, P_{42}, \ldots, P_{99}$$

E é evidente que:

$$P_{50} = md \qquad P_{25} = Q_1 \text{ e } P_{75} = Q_3$$

Logo:

$$P_1 \rightarrow p_1 = \frac{n+1}{100}$$

$$P_2 \rightarrow p_2 = \frac{2(n+1)}{100}$$

$$\vdots$$

$$P_{50} \rightarrow p_{50} = \frac{50(n+1)}{100}$$

$$\vdots$$

$$P_{99} \rightarrow p_{99} = \frac{99(n+1)}{100}$$

Onde n = número de dados (valores).

Por exemplo, o percentil 97,5 é um valor que se apresenta acima de 97,5 % das observações e tem acima dele os 2,5% dos valores restantes.

Exemplo:

Considere as seguintes notas de uma turma e calcule os percentis dessa amostra:

52,0	55,9	56,7	59,4	60,2	54,4	55,9	56,8	59,4	60,3
54,5	56,2	57,2	59,5	60,5	55,7	56,4	57,6	59,8	60,6
55,8	56,4	58,9	60,0	60,8					

Utilizando as relações para a posição dos percentis, temos:

$$p_{10} = 3 \quad p_{20} = 5 \quad p_{90} = 23$$

O décimo percentil (P_{10}) abrange 10% dos termos da série, o vigésimo percentil (P_{20}) 20% e assim por diante, o nonagésimo (P_{20}) 90%.

7.4.4 Amplitude interquartil

A amplitude interquartil (AIQ) é a diferença entre o valor do terceiro quartil (Q_3) e o valor do primeiro quartil (Q_1) e compreende os 50% dos dados centrais da série. É menos afetada pelos valores extremos do que a amplitude total, tornando-se uma medida de grande utilidade, principalmente em séries assimétricas. A AIQ é válida para dados ordinais, intervalares ou de razão, é calculada por:

$$AIQ = Q_3 - Q_1$$

7.5 Medidas de dispersão

São medidas utilizadas para medir o grau de variabilidade ou dispersão dos valores observados em torno da média aritmética. Servem para medir a representatividade da média e proporcionar o conhecimento do nível de homogeneidade ou heterogeneidade dentro de cada grupo analisado.

A dispersão ou variabilidade representa um dos mais importantes grupos de medidas da Estatística. Para o conhecimento pleno e adequado de uma série ou de uma distribuição de frequências é necessário determinar não apenas a posição central dos valores, através das medidas de posição, mas também é preciso conhecer o real grau de afastamento de um conjunto de números em relação à sua média.

As medidas de dispersão se dividem em:

Absoluta	Amplitude, desvio médio, desvio padrão e variância.
Relativa	Coeficiente de variação.

7.5.1 Medida de dispersão absoluta

7.5.1.1 Variância

As medidas de tendência central são insuficientes para descrever adequadamente uma amostra. É necessário também descrever em que medida os dados de observações estão agrupados ao redor da media. A variância (s^2) mede a dispersão dos dados de observações de uma amostra em relação à respectiva média.

Deste modo, define-se a **variância** como:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{ou} \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

7.5.1.2 Desvio padrão

Desvio padrão *(s)* é a mais importante medida de dispersão dos valores individuais ao redor da média. Apresenta vantagem sobre a variância, pois utiliza a mesma unidade de medida de dados (kg, cm, etc.) empregadas na tomada das observações.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad \text{ou} \quad s = \sqrt{\frac{1}{n-1}\sum(x_i - \overline{x})^2}$$

O desvio padrão *(s)* é a raiz quadrada da variância, então é calculado por:

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

Exemplos:

1. Considere a série a seguir: $x_i = \{8; 10; 12; 9; 11; 7; 13\}$, para a qual $\overline{x} = 10$. Para calcular a variância e o desvio padrão é útil construirmos uma tabela com os desvios:

x_i	$d_i = x_i - \overline{x}$	d_i^2
8	−2	4
10	0	0
12	2	4
9	−1	1
11	1	1
7	−3	9
13	3	9
	$\sum d_i = 0$	$\sum d_i^2 = 28$

Deste modo, o desvio padrão pode ser calculado:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}} \qquad n = 7$$

$$s = \sqrt{\frac{(8-10)^2+(10-10)^2+(12-10)^2+(9-10)^2+(11-10)^2+(7-10)^2+(13-10)^2}{7-1}}$$

$$s = \sqrt{\frac{(-2)^2+0^2+2^2+(-1)^2+1^2+(-3)^2+3^2}{6}}$$

$$s = \sqrt{4.67} = 2.16$$

Média: $\overline{x} = 10$ Desvio padrão: s = 2.16

2. Do mesmo modo, considere a série $y_i = \{10; 11; 9; 10; 10; 9; 11\}$, que tem a mesma média $\overline{y} = 10$ e para cuja tabela de desvios tem-se:

x_i	$d_i = x_i - \overline{x}$	d_i^2
10	0	0
11	1	1
9	−1	1
10	0	0
10	0	0
9	−1	1
11	1	1
	$\sum d_i = 0$	$\sum d_i^2 = 28$

Procedendo da mesma forma, calcula-se o desvio padrão:

$$s = \sqrt{\frac{1}{6} \cdot 4} = 0.81$$

Média: $\overline{x} = 10$ Desvio padrão: s = 0.81

7.5.1.3 Desvio médio

É a média aritmética dos valores absolutos dos desvios tomados em relação à média ou à mediana. Considera-se o módulo de cada desvio:

$$|\, d_i \,| = |\, X_i - x \,|$$

É dado por:

$$Dm = \sum \frac{|\, X_i - x \,|}{n} \cdot |\, X_i - x \,| = \sum \frac{d_i}{n}$$

7.5.1.4 Amplitude de variação

É a mais simples e precária medida de variabilidade. É a diferença entre o valor mais alto *(X máx.)* e o valor mais baixo *(X mín.)* de uma série, representada pela letra H:

$$H = X\,máx. - X\,mín.$$

7.5.1.5 Soma dos quadrados

A soma dos quadrados refere-se à soma dos quadrados dos desvios em relação à média:

$$SQ = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}$$

7.5.2 Medida de dispersão relativa

Empregada quando se deseja comparar a variabilidade de duas ou mais distribuições, mesmo quando essas se referem a diferentes fenômenos e sejam expressas em unidades de medida distintas. Isto é, para comparar a variação do desvio padrão com a média, usa-se a razão entre o desvio padrão e a média, chamado de **coeficiente de variação de Pearson.** Desta forma, este coeficiente é uma medida de dispersão relativa.

> *"O coeficiente de variação para um conjunto de n observações é definido como o quociente entre o desvio padrão e a média aritmética da distribuição."*

O **coeficiente de variação** independe da unidade de medição empregada, permitindo com isso, a comparação de vários tipos de dados, tais como: pressão arterial e temperatura. O coeficiente de variação (CV) é, portanto, a magnitude relativa do desvio-padrão expresso em porcentagem da média:

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

Exemplos:

1. Considerando uma média valendo $\bar{x} = 980$ e um desvio padrão $s = 56$, temos:

$$CV = \frac{56}{980}.100 = 5.7\%$$

Que indica a dispersão da amostra.

Observação:

O coeficiente de variação de Pearson ou apenas coeficiente de variação (CV), geralmente é expresso em porcetagem. Alguns analistas consideram:

- Baixa dispersão: $CV < 15\%$
- Média dispersão: $15\% < CV < 30\%$
- Alta dispersão: $CV > 30\%$

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

2. Considere uma amostra que apresenta uma distribuição de frequência com $x = 161$ e $s = 4.22$; logo temos coeficiente de variação valendo:

$$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{4.22}{161} \cdot 100 = 2.62 \cong 2.6\%$$

7.6 Momentos, assimetria e curtose

As medidas de assimetria e curtose complementam as medidas descritivas de posição e de dispersão, no sentido de proporcionar uma descrição e compreensão mais completa das distribuições de frequências. Estas distribuições não diferem apenas quanto ao valor médio e à variabilidade, mas também quanto a sua forma (assimetria e curtose).

Para conhecer as medidas de assimetria e curtose é necessário compreender certas quantidades, conhecidas como **momentos**.

7.6.1 Momentos

São medidas descritivas de caráter mais geral e dão origem às demais medidas descritivas, tais como as: de posição, dispersão, assimetria e curtose. De acordo com a potência considerada tem-se a ordem ou o grau do momento calculado.

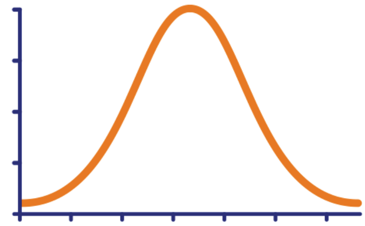
Os momentos podem ser simples ou centrados na origem, centrados na média e momentos abstratos.

7.6.2 Assimetria e simetria

Assimetria é o grau de desvio, afastamento da simetria ou grau de deformação de uma distribuição de frequências. Os coeficientes de assimetria servem para medir o grau de deformação da distribuição. Se não houver desvio então a distribuição é simétrica.

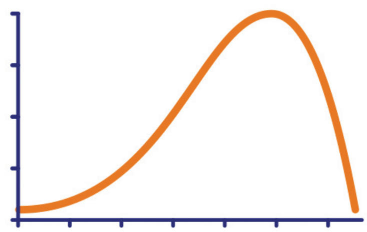
a) Assimétrica nula ou simétrica: uma distribuição de frequências com classes é simétrica quando:

Média = Mediana = Moda



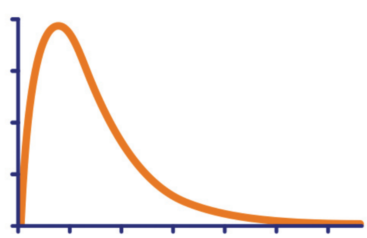
b) Assimétrica à esquerda ou negativa: uma distribuição de frequências com classes é assimétrica à esquerda ou negativa quando:

Média < Mediana < Moda



c) Assimétrica à direita ou positiva: uma distribuição de frequências com classes é assimétrica à direita ou positiva quando:

Média > Mediana > Moda



Para fazer a classificação do tipo de assimetria através de cálculo, considerando a média *(x)* e a moda *(Mo)*, podemos utilizar a seguinte relação:

$x - Mo$

Se:

$x - Mo = 0$ assimetria nula ou distribuição simétrica.

$x - Mo < 0$ assimetria negativa ou à esquerda.

$x - Mo > 0$ assimetria positiva ou à direita.

Exemplo:

Determinando os tipos de assimetria das distribuições abaixo:

Distribuição A

Classe	f_i
2 → 6	6
6 → 10	12
10 → 14	24
14 → 18	30
18 → 22	6
	$\sum f_i = 78$

Distribuição B

Classe	f_i
2 → 6	6
6 → 10	30
10 → 14	24
14 → 18	12
18 → 22	6
	$\sum f_i = 78$

Distribuição C

Classe	f_i
2 → 6	6
6 → 10	12
10 → 14	24
14 → 18	12
18 → 22	6
	$\sum f_i = 60$

Calculando a média, mediana, moda e o desvio padrão para cada distribuição temos:

Distribuição A:
$x = 12.9$
$Md = 13.5$
$Mo = 16$ kg
$s = 4.20$

Distribuição B:
$x = 11.1$
$Md = 10.5$
$Mo = 8$ kg
$s = 4.20$

Distribuição C:
$x = 12$
$Md = 12$
$Mo = 12$ kg
$s = 4.42$

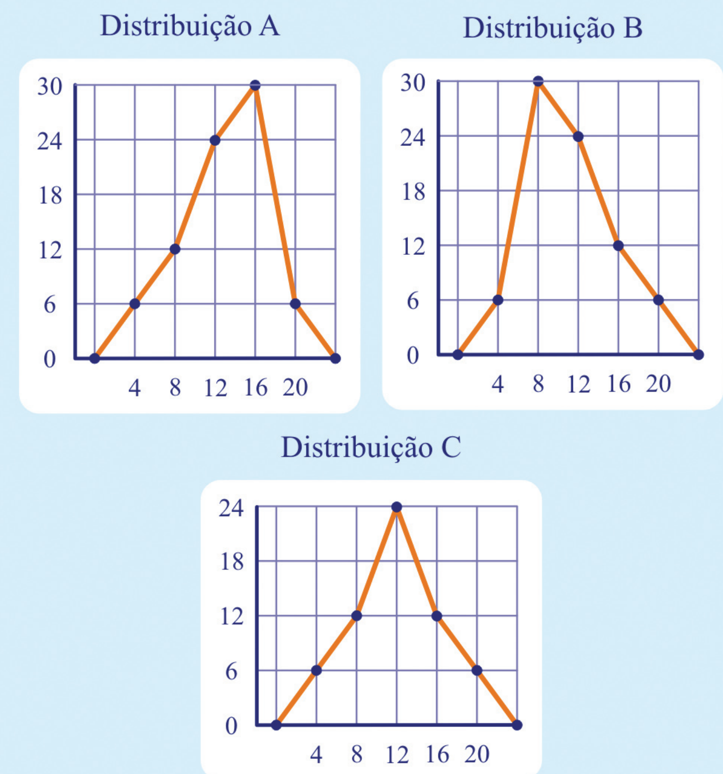
Logo, calculando o tipo de assimetria:

Distribuição A:
$x - Mo$
$12.9 - 16 = -3.1$
a distribuição é assimétrica negativa

Distribuição B:
$x - Mo$
$11.1 - 8 = 3.1$
a distribuição é assimétrica positiva

Distribuição C:
$x - Mo$
$12 - 12 = 0$
a distribuição é simétrica

Construindo os gráficos das distribuições anteriores, temos:



7.6.2.1 Coeficiente de assimetria

A medida anterior *(x-Mo)*, por ser absoluta, apresenta a mesma deficiência do desvio padrão, isto é, não permite a comparação entre as medidas de duas distribuições. Por este motivo, daremos preferência ao **coeficiente de assimetria de Pearson** *(AS)*, definido por:

$$AS = \frac{3(x - Md)}{S}$$

Escalas de assimetria:

$0.15 < |AS| < 1$ assimetria moderada

$|AS| \geq 1$ assimetria elevada

Observação:

Coeficiente de assimetria de Pearson é considerado em módulo para se fazer a classificação acima.

Exemplo:

Considerando as distribuições do exemplo anterior (A, B e C), calcule os coeficientes de assimetria e faça a análise dos resultados obtidos:

$$ASA = \frac{3(x - Md)}{S} = \frac{3(12 - 12)}{4.42} = 0 \text{ (simetria)}$$

$$ASB = -0.429 \text{ (assimetria moderada e negativa)}$$

$$ASC = 0.429 \text{ (assimetria moderada e positiva)}$$

7.6.3 Curtose

É o grau de achatamento (afilamento) de uma curva em relação à curva Normal (Guassiana), tomada como padrão.

Pode ser classificada em:

Platicúrtica: a curva é mais achatada do que a normal.
Mesocúrtica: a curva é normal.
Leptocúrtica: a curva é mais alta do que a normal.

7.7 Atividades de aprendizagem e avaliação

01. Explique o significado de variância e desvio padrão.

02. Dê um exemplo e explique o coeficiente de variação de Pearson.

03. Faça um paralelo entre assimetria e simetria.

04. Qual a classificação da curtose?

7.8 Síntese

Nesta unidade estudamos as medidas de dispersão, aprendemos que a descrição de um conjunto de dados é mais completa quando se considera além de uma medida de tendência central, uma medida de variação. Isto ocorre porque é comum encontrar séries que, apesar de apresentarem a mesma média, são compostas de maneiras diferentes. Desta forma, mostramos que as medidas de tendência central são insuficientes para descrever adequadamente uma série estatística.

UNIDADE 8 – PROBABILIDADE

8.1 Objetivos de aprendizagem

- Conhecer e entender o trabalho estatístico;
- Diferenciar e aplicar os conceitos de espaço amostral, experimento aleatório e evento;
- Aplicar e conhecer o cálculo da probabilidade de ocorrência de um evento.

8.2 Introdução

No ano de 1654, um jogador da sociedade parisiense chamado Chevalier de Mére, propôs ao matemático Blaise Pascal algumas questões sobre possibilidades de vencer em jogos. Uma destas questões foi: "Um jogo de dados entre dois adversários chega ao fim quando um dos jogadores vence três partidas em primeiro lugar. Se esse jogo for interrompido antes do final, de que maneira cada um dos jogadores deverá ser indenizado?". As estas reflexões em torno destes problemas, levaram Pascal a corresponder-se com Pierre de Fermat, desencadeando discussões a respeito dos princípios de uma nova teoria, que veio a ser chamada de Teoria de Probabilidade.

Já sabemos que para obter informações sobre alguma característica da população, o tamanho amostral é de fundamental importância. Estudaremos agora Probabilidade, que é uma ferramenta usada e necessária para fazer ligações entre a amostra e a população, de modo que a partir de informações da amostra se possam fazer afirmações sobre características da população.

Assim, pode-se dizer que a Probabilidade é a ferramenta básica da Estatística Inferencial.

8.3 Experimento aleatório

É aquele experimento que quando repetido em iguais condições, podem fornecer resultados diferentes, ou seja, são resultados explicados ao acaso, isto é, não podem ser previamente determinados, dependem exclusivamente do acaso. Se o fenômeno seguir um modelo não-determinístico, tem-se um experimento aleatório, que possui as seguintes características:

- O experimento pode ser repetido;

- Embora não seja possível afirmar que resultado em particular ocorrerá, é possível descrever o conjunto de todos os resultados possíveis do experimento;

- Á medida que aumenta o número de repetições aparece uma certa regularidade que torna possível a construção de um modelo matemático.

Exemplo:
Lançamento de dois dados, lançamento de uma moeda, sorteio de um cupom dentre cem mil cupons, etc.

8.4 Espaço amostral

É o conjunto universo ou o conjunto de todos os resultados possíveis de um experimento aleatório. A letra que representa o espaço amostral é S.

Exemplos:

1. No experimento "lançamento de um dado" ao observar a face superior temos como espaço amostral o conjunto $S = \{1, 2, 3, 4, 5, 6\}$. Portanto: $n(S) = 6$.
2. Num jogo de futebol, entre duas equipes, uma das equipes pode obter resultados tais como: vitória (v), empate (e) ou derrota (d).
   Tem-se então:
   $S = \{v, e, d\}$

8.5 Evento

É um conjunto qualquer de resultados de um experimento aleatório. Pode-se dizer que um evento é qualquer subconjunto de um espaço amostral, sendo representado pela letra E.

Exemplo:
No lançamento de duas moedas, o espaço amostral é $E = \{(C,C), (C,K), (K,C), (K,K)\}$. Se aparecerem faces iguais, o conjunto $C = \{(K, K), (C, C)\}$, subconjunto de E, é um evento de E.

8.5.1 Tipos de eventos

a) Eventos certos: são eventos que possuem todos os elementos do espaço amostral. Logo, se $E = S$ então o evento é o próprio espaço amostral.

Exemplo:
Lançamento de um dado e ocorrência de um número menor do que 6 na face superior.

b) Eventos impossíveis: são eventos que não possuem elementos no espaço amostral. Logo, se $E = \varnothing$ então o evento E é chamado de evento impossível.

Exemplo:
Lançamento de um dado e ocorrência de um número maior do que 6 na face superior.

c) Eventos elementares: se $E \subset S$ (E está contido em S) e E é um conjunto unitário, então E é chamado evento elementar.

Exemplo:
Lançamento de um dado e ocorrência de um número ímpar maior do que 4 na face superior.

d) **Evento complementar**: em relação a determinado evento A podemos definir seu evento complementar $(\bar{A})$ que é caracterizado pela não ocorrência daquele evento. Por exemplo: considerando o lançamento de um dado, onde S = {1, 2, 3, 4, 5, 6}, e definido o evento A ={números pares} então o **evento complementar de** A é o conjunto dos números ímpares, $\bar{A}$ = {1,3,5}.

8.6 Cálculo da probabilidade de um evento ocorrer

Considerando que todos os elementos do espaço amostral (S) têm a mesma chance de ocorrer, podemos definir o cálculo da probabilidade de um evento ocorrer como a razão (divisão) entre o número de elementos do evento e o número de elementos do espaço amostral:

$$P(S) = \frac{n(E)}{n(S)}$$

Onde:
$n(E)$ = o número de elementos do evento
$P(S)$ = a probabilidade de ocorrer o evento
$n(S)$ = o número de elementos do espaço amostral

8.7 Probabilidade de um evento complementar $P(\bar{A})$

A probabilidade de ocorrer o evento complementar $\bar{A}$ é igual a um menos a probabilidade de ocorrer A, que pode ser representada por:

$$P(\bar{A}) = 1 - P(A)$$

Exemplo:
Fazendo-se inspeção em um lote de 240 peças de motor, o departamento de controle de qualidade constatou que 20 peças estavam com defeito. Retirando-se ao acaso uma das 240 peças, a probabilidade de esta peça NÃO ser defeituosa é:

- Sendo S o conjunto dos elementos do espaço amostral, casos possíveis, e $n(S)$ o número de elementos deste conjunto.
- Sendo E o conjunto dos elementos das peças defeituosas, e $n(E)$ o número de elementos deste conjunto.
- Sendo E' o conjunto dos elementos das peças não defeituosas, e $n(E')$ o número de elementos deste conjunto. Neste caso, é o conjunto dos casos favoráveis.

$$n(S) = 240 \quad n(E) = 20 \quad n(E') = 220$$

Para calcular a probabilidade de retirada uma peça que seja não defeituosa, faça assim:

$$P(E') = \frac{n(E)}{n(S)} = \frac{220}{240} = \frac{11}{12}$$

8.8 Probabilidade da união de eventos

8.8.1 Eventos mutuamente exclusivos

Dois eventos A e B são mutuamente exclusivos se $A \cap B = \emptyset$, neste caso a probabilidade da união desses eventos é calculada por:

$$P(A \cup B) = P(A) + P(B)$$

8.8.2 Eventos não mutuamente exclusivos

Dois eventos A e B não são mutuamente exclusivos se $A \cap B \neq \varnothing$, neste caso a probabilidade da união desses eventos é calculada por:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Exemplo:

No lançamento de um dado, determine a probabilidade de se obter um número ímpar ou menor que 3.

- Espaço amostral: $S = \{1, 2, 3, 4, 5, 6\}, \; n(S) = 6$
- Evento A (números ímpares): $A = \{1, 3, 5\}, \; n(A) = 3$
- Evento B (números menores que 3): $B = \{1, 2, 3\}, \; n(B) = 3$
- Evento de $A \cap B$: $A \cap B = \{1, 3\} \; n(A \cap B) = 2$

Calculando a probabilidade, temos:

$$P(A \cap B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} + \frac{2}{6} + \frac{4}{6} + \frac{2}{3}$$

8.9 Probabilidade condicional

Sejam dois eventos A e B associados a um espaço amostral S. Denota-se por $P\ (A|B)$, a probabilidade do evento A ocorrer dado que o evento B ocorreu é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ onde } P(B) \neq 0, \text{ pois } B \text{ já ocorreu}$$

Portanto, quando calculamos $P(A|B)$, tudo se passa como se o evento B fosse um novo espaço amostral reduzido dentro do qual queremos calcular a probabilidade do evento A.

Pode-se escrever, também, através do Teorema do Produto:

$$P(A \cap B) = P(A/B)\,.\,P(B)$$

$$P(A \cap B) = P(B/A)\,.\,P(A)$$

Assim, encontramos uma alternativa para o cálculo da probabilidade da interseção de dois eventos A e B, denotada por $P(A \cap B)$.

Exemplo:

Considere o conjunto de números inteiros $S = \{1, 2, 3, 4, 5, ..., 21, 22, 23, 24, 25\}$ e por meio de um sorteio ao acaso, retire um número. Se o número sorteado for ímpar, qual a probabilidade de número ser 13?

- Espaço amostral: $S = \{1, 2, 3, ..., 23, 24, 25\}$, $n(S) = 25$
- Evento A: $A = \{13\}$, $n(A) = 1$
- Evento B: Condição para ocorrência do evento A: $\{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25\}$, $n(B) = 13$

$$(A \cap B) = \{13\}, n(A \cap B) = 1$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{13}$$

8.10 Eventos independentes

Para que dois ou mais eventos sejam independentes é necessário que a ocorrência de um não dependa (ou não é condicionada, ou não se vincula) da ocorrência do outro evento. Isto é, a informação adicional de que um dos eventos já ocorreu em nada altera a probabilidade de ocorrência do outro evento.

Portanto, sejam dois eventos independentes A e B, a probabilidade de que ocorram os eventos A e B simultaneamente é dada por:

$$P(A \cap B) = P(A).\, P(B)$$

Exemplo:

Em um grupo de alunos de uma determinada escola, a probabilidade de que um estudante escolhido ao acaso, tenha média final acima de 7.0 é de 1/5. Nesse mesmo grupo, a probabilidade de que um estudante saiba jogar tênis é de 5/6.

Qual a probabilidade de escolhermos um estudante (ao acaso), que tenha média final maior que 7.0 e que saiba jogar tênis?

- Evento A: ter média final acima de 7.0;
- Evento B: saber jogar tênis;
- Eventos A e B: ter média final acima de 7.0 e saber jogar tênis.

Como queremos calcular $P(A \text{ e } B)$, admita que: de todos os estudantes, 1/5 têm média final acima de 7.0 e 5/6 sabem jogar tênis.

Dessa forma, 5/6 de 1/5, ou seja, 5/6 . 1/5 = 1/6, sabem jogar tênis e têm média final acima de 7.0.

Portanto:

$$P(A \cap B) = P(A).P(B) = \frac{5}{6} \cdot \frac{1}{5} = \frac{1}{6}$$

8.11 Atividades de aprendizagem e avaliação

01. Defina probabilidade.

02. Faça um paralelo entre evento e espaço amostral, exemplificando-os.

03. Diferencie eventos mutuamente exclusivos e não-exclusivos.

8.12 Síntese

Nesta unidade estudamos os princípios de Probabilidade. Verificamos que o trabalho estatístico se desenvolve a partir da observação de determinados fenômenos e emprega dados numéricos relacionados aos mesmos a fim de tirar conclusões que permitam conhecê-los e explicá-los. Neste caso, com determinado grau de crença, para obter o desenvolvimento teórico do fenômeno é necessário a formulação de um modelo.

No campo da Estatística, os modelos matemáticos utilizados são denominados modelos não-determinísticos ou probabilísticos, ou seja, permitem avaliar a probabilidade de ocorrência de resultados.

Com esta unidade finalizamos o estudo de Estatística, adquirindo conhecimentos básicos e aplicados, os quais servirão de base para o estudo de outras disciplinas, tais como Gestão da Qualidade, entre outras.

REFERÊNCIAS

COSTA NETO, P. L. de O. **Probabilidades**. São Paulo: Editora Edgard Blucher Ltda, 2 ed., 2005.

COSTA NETO, P. L. de O. **Estatística.** São Paulo: Editora Edgard Blucher Ltda, 2002.

CRESPO, A. A. **Estatística Fácil.** São Paulo: Editora Saraiva, 2002.

DAL MOLIN, Beatriz Helena, et al. **Mapa Referencial para Construção de Material Didático** - Programa e-Tec Brasil. 2. ed. revisada. Florianópolis: Universidade Federal de Santa Catarina – UFSC, 2008.

DANTE, L. R. **Matemática: Contexto e Aplicações**. São Paulo: Atica, 3v., 2003.

DAVID, R. A.; Dennis, J. S.; Thomas, A. W. **Estatística Aplicada a Administração e Economia.** São Paulo: Thomson, 2 ed., 2007.

DOWNING, D.; CLARK, J. **Estatística Aplicada.** São Paulo: Editora Saraiva, 2000.

LAPPONI, J. C. **Estatística usando Excel.** São Paulo: Editora Lapponi, 2000.

LEVIN, J. **Estatística Aplicada a Ciências Humanas.** São Paulo: Harbra, 2 ed., 1987.

NICK, E.; KELLNER, S. R. O. **Fundamentos de Estatística para as Ciências do Comportamento.** Rio de Janeiro: Editora Renes, 1971.

SIEGEL, S.; CASTELLAN, N. J. **Estatística não-paramétrica para Ciências do Comportamento.** Porto Alegre: Artmed, 2 ed., 2006.

STEVENSON, W. J. **Estatística Aplicada à Administração.** São Paulo: Harbra, 1997.

TRIOLA, M. F. **Introdução à Estatística.** Rio de Janeiro: LTC (Livros Técnicos e Científicos), Editora S.A., 7 ed., 1999.

CURRÍCULO SINTÉTICO DO PROFESSOR-AUTOR

Paulo Roberto da Costa é natural de Santa Maria – RS, professor do Colégio Técnico Industrial (CTISM) da Universidade Federal de Santa Maria (UFSM). Engenheiro Eletricista, graduado pela Universidade Federal de Santa Maria. Licenciado em Matemática com Habilitação em Física pela Faculdade de Filosofia Ciências e Letras Imaculada Conceição (FIC), em Santa Maria/RS. Licenciado no Esquema I (Formação de professores) com habilitações em desenho técnico, eletrônica e eletricidade pela Universidade Luterana do Brasil (ULBRA), em Canoas/RS. Especialista em Engenharia Clínica pela Universidade Federal do Rio Grande do Sul (UFRGS), em Porto Alegre/RS. Especialista em Engenharia de Segurança do Trabalho pela UFSM. Mestre em Engenharia de Produção (Qualidade e Produtividade) pela UFSM. Doutor em Engenharia Agrícola pela UFSM.

Atualmente no CTISM/UFSM, é Coordenador do Curso Técnico de Segurança do Trabalho. Ministra aulas em diversos cursos técnicos, como Mecânica e Segurança do Trabalho, em disciplinas como: Gerenciamento de Riscos, Gestão da Qualidade e Administração e Organização para o Trabalho, Segurança do Trabalho, entre outras, conforme o semestre letivo.

Ministrou disciplinas no Colégio Politécnico da UFSM, Colégio Agrícola de Frederico Westphalen e no Curso de Especialização de Engenharia de Segurança do Trabalho, também na UFSM. Foi Coordenador do Curso Técnico em Segurança do Trabalho e Diretor do Departamento de Relações Empresariais e Comunitárias, todos no CTISM. Participou de diversos conselhos administrativos (CEPE, CPPD, Colegiado, Comissões) da UFSM. É membro do Núcleo de Ensino a Distância, e participa também em projetos de extensão, ministrando cursos de capacitação para eletricistas da geração, transmissão e distribuição de energia elétrica nas áreas de Eletricidade e de Segurança do Trabalho.

Caro Estudante!

O Curso Técnico de Automação Industrial na modalidade de Educação a Distância - Programa Escola Aberta do Brasil (e-Tec Brasil), oferecido pelo Colégio Técnico Industrial de Santa Maria (CTISM), Universidade Federal de Santa Maria (UFSM), é resultado da participação do CTISM no Edital de Seleção de Projetos de Cursos de Educação Profissional Técnica de Nível Médio, na modalidade Educação a Distância, publicado dia 27 de abril de 2007 no D.O.U., seção III. pág. 60, através do Edital de Seleção n° 1/2007/SEED/SETEC/MEC.

O Curso é oferecido em etapas, totalizando oito bimestres, com duração de dois anos. Tem como objetivo capacitar os alunos para, a partir de conhecimentos específicos nas áreas de Mecânica, Eletrônica, Eletrotécnica, Informática e Gestão, que sejam formados profissionais capazes de atuar no planejamento, implantação, otimização e manutenção de linhas de produção automatizadas nas indústrias, visando valorizar e qualificar empresas e os sistemas locais de produção. A carga horária do curso é de 1200 horas, complementando-se com estágio (200 horas) ou Trabalho de Conclusão de Curso (TCC).

O conteúdo programático é composto por unidades relacionadas à utilização de micro-controladores voltados para a Automação Industrial, funcionamento de equipamentos e de e instalações industriais, manutenção em equipamentos com comando numérico computadorizado e controladores lógicos programáveis, aplicação de técnicas de gerenciamento de qualidade, montagem de sistemas integrados eletrônicos, eletros-pneumáticos, eletros-hidráulicos e mecânicos, integrar equipes multiprofissionais e controlar a qualidade de produtos em processos. As disciplinas são de natureza teórica e prática na medida em que são necessários conhecimentos de base para compreensão dos diversos processos e habilidades para a execução de tarefas.

O responsável pelo Curso é o Professor/Coordenados Paulo Roberto Colusso, do Colégio Técnico Industrial de Santa Maria, da Universidade Federal de Santa Maria-RS. Esperamos que o trabalho tenha pleno êxito e empenho de todos, em especial de professores, tutores e alunos.

Uma mensagem de otimismo:

Vai, acelera as estradas do Pensamento
Em direção aos seus alvos brilhantes;
Ao longe o lavrador semeia suas sementes,
Para que o trigo que colheis sejam almas.
Ralph Waldo Emerson

Boa sorte a todos.

Paulo Roberto Colusso - coordenador do curso

e-Tec Brasil
Escola Técnica Aberta do Brasil

ISBN