rede e-Tec Brasil

Estatística

Paulo Roberto da Costa

Colégio Técnico Industrial UFSM

Santa Maria - RS
2011

Presidência da República Federativa do Brasil

Ministério da Educação

Secretaria de Educação a Distância

Este material didático 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.

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

Coordenação Institucional
Araci Hack Catapan/UFSC

Coordenação do Projeto
Silvia Modesto Nassar/UFSC

Cordenação de Design Instrucional
Beatriz Helena Dal Molin/UNIOESTE

Designers Intrucionais
Helena Maria Maullmann/UFSC
Jorge Luiz Silva Hermenegildo/CEFET-SC

WEB Designers
Beatriz Helena Dal Molin/UNIOESTE
Mércia Freire Rocha Cordeiro Machado/ETUFPR

Supervisão de Projeto Gráfico
Ana Carine García Montero/UFSC

Diagramação
João Ricardo Zattar/UFSC
Luís Henrique Lindler/UFSC

Revisão
Lúcia Locatelli Flôres/UFSC

Comissão de Acompanhamento e Validação
Colégio Técnico Industrial de Santa Maria/CTISM

Coordenador Institucional
Paulo Roberto Colusso/CTISM

Professor-autor
Paulo Roberto da Costa/CTISM

Coordenação Técnica
Iza Neuza Teixeira Bohrer/CTISM

Coordenação de Design
Erika Goellner/CTISM

Revisão Pedagógica
Andressa Rosemárie de Menezes Costa/CTISM
Francine Netto Martins Tadielo/CTISM
Marcia Migliore Freo/CTISM

Revisão Técnica
Carmen Vieira Mathias/UFSM
Eduardo Lehnhart Vargas/CTISM

Revisão Textual
Lourdes Maria Grotto de Moura/CTISM
Vera da Silva Oliveira/CTISM

Diagramação e Ilustração
Gustavo Schwendler/CTISM
Leandro Felipe Aguilar Freitas/CTISM
Marcel Santos Jacques/CTISM
Máuren Fernandes Massia/CTISM
Maíra Rodrigues/CTISM
Rafael Cavalli Viapiana/CTISM
Ricardo Antunes Machado/CTISM

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

C837e Costa, Paulo Roberto da
Estatística / Paulo Roberto da Costa. – 3. ed. – Santa Maria : Universidade Federal de Santa Maria, Colégio Técnico Industrial de Santa Maria, Curso Técnico em Automação Industrial, 2011.
95 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

Apresentação e-Tec Brasil

Prezado estudante,

Bem-vindo ao e-Tec Brasil!

Você faz parte de uma rede nacional pública de ensino, a Escola Técnica Aberta do Brasil, instituída pelo Decreto nº 6.301, de 12 de dezembro 2007, com o objetivo de democratizar o acesso ao ensino técnico público, na modalidade a distância. O programa é resultado de uma parceria entre o Ministério da Educação, por meio das Secretarias de Educação a Distância (SEED) e de Educação Profissional e Tecnológica (SETEC), as universidades e escolas técnicas estaduais e federais.

A educação a distância no nosso país, de dimensões continentais e grande diversidade regional e cultural, longe de distanciar, aproxima as pessoas ao garantir acesso à educação de qualidade, e promover o fortalecimento da formação de jovens moradores de regiões distantes dos grandes centros geograficamente ou economicamente.

O e-Tec Brasil leva os cursos técnicos a locais distantes das instituições de ensino e para a periferia das grandes cidades, incentivando os jovens a concluir o ensino médio. Os cursos são ofertados pelas instituições públicas de ensino e o atendimento ao estudante é realizado em escolas-polo integrantes das redes públicas municipais e estaduais.

O Ministério da Educação, as instituições públicas de ensino técnico, seus servidores técnicos e professores acreditam que uma educação profissional qualificada – integradora do ensino médio e educação técnica, – é capaz de promover o cidadão com capacidades para produzir, mas também com autonomia diante das diferentes dimensões da realidade: cultural, social, familiar, esportiva, política e ética.

Nós acreditamos em você!

Desejamos sucesso na sua formação profissional!

Ministério da Educação
Janeiro de 2010

Nosso contato
etecbrasil@mec.gov.br

Indicação de ícones

Os ícones são elementos gráficos utilizados para ampliar as formas de linguagem e facilitar a organização e a leitura hipertextual.

Atenção: indica pontos de maior relevância no texto.

Saiba mais: oferece novas informações que enriquecem o assunto ou "curiosidades" e notícias recentes relacionadas ao tema estudado.

A-Z

Glossário: indica a definição de um termo, palavra ou expressão utilizada no texto.

Mídias integradas: sempre que se desejar que os estudantes desenvolvam atividades empregando diferentes mídias: vídeos, filmes, jornais, ambiente AVEA e outras.

Atividades de aprendizagem: apresenta atividades em diferentes níveis de aprendizagem para que o estudante possa realizá-las e conferir o seu domínio do tema estudado.

Sumário

Palavra do professor-autor

"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

"Uma probabilidade razoável é a única certeza."

Samuel Howe

"A experiência não permite nunca atingir a certeza absoluta. Não devemos procurar obter mais que uma probabilidade."

Bertrand Russel

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

Esperamos que esta apostila facilite sua aprendizagem e desenvolva a sua forma de pensar matematicamente. Queremos que ela seja bastante útil na busca de soluções para uma série de problemas do dia a dia.

Faz-se necessário muito estudo para obter sucesso no processo de aprendizagem; mãos à obra e que o aproveitamento seja significativo.

Paulo Roberto da Costa

Apresentação da disciplina

Investigação sugere pesquisa, busca de informações e análise de dados. Tudo isso 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 socioeconômicas; é de aplicação intensa na Engenharia Industrial, no controle de qualidade e na comparação de fabricações, é também muito aplicada na engenharia agrícola, entre outras.

A aplicação da Estatística intensificou-se nas últimas décadas, tornando-se o foco do estudo de especialistas para as áreas econômicas, sociais, culturais, políticas, educacionais, de saúde, meio ambiente, etc.

Enfim, explícita ou implicitamente, a Estatística está presente em todos os aspectos da vida moderna, e essa presença só tende a crescer, pois o estudo estatístico colabora como indicardor para trabalhar com diversos produtos, possibilitando maiores estratégias na busca e no planejamento de soluções.

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, é utilizado o termo bioestatístico para distingui-la 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 um dos objetivos da Estatística extrair informação dos dados para obter uma melhor compreensão das situações que representam.

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, procura-se agrupá-los e reduzi-los, sob forma de amostra, deixando de lado a aleatoriedade presente.

Seguidamente o objetivo do estudo estatístico pode ser o de estimar uma quantidade ou testar uma hipótese, utilizando-se técnicas estatísticas convenientes as quais 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 na nossa vida. As empresas utilizam 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 necessária a utilização dessas ferramentas. Por isso, é necessário saber ler gráficos, interpretá-los, prever situações, analisar dados, etc. A estatística nos ajudará nesta tarefa.

Panorama histórico da Estatística

Embora a Estatística seja recente na área de pesquisa, ela já era observada de forma rudimentar e imprecisa na antiguidade, quando então os governantes coletavam e faziam registros de dados que consideravam importantes, tais como as informações sobre suas populações e suas riquezas, tendo como objetivo fins militares ou tributários.

Dados têm sido coletados ao longo da história. Vejamos:

a) **Antiguidade** – vários povos já registravam o número de habitantes, de nascimentos e óbitos. Já se faziam estimativas das riquezas sociais, distribuíam-se equitativamente terras 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) **Século XVI** – surgem as primeiras análises sistemáticas, as primeiras tabelas, os números relativos e o cálculo de probabilidade.

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

Até o início do século 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) **Século XVII** – iniciou-se na Inglaterra uma nova fase de desenvolvimento da Estatística, voltada para a análise dos fenômenos observados. Era a fase da **Estatística Analítica**.

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

e) **Século XVIII** – a origem do termo **Estatística** surgiu no século XVIII. Quanto à origem da estatística, a data de seu aparecimento não parece ser encarada com unanimidade. Há quem diga que o seu autor foi Godofredo Achenwall (1719-1772), que usou pela primeira vez o termo estatística (*statistik*, do grego *statizein*). Outros também afirmam 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 século XVIII e XIX** – a Estatística desenvolveu-se muito, associada ao cálculo das probabilidades que haviam se desenvolvido e à 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 a serviço dos diferentes ramos do saber. Surge então a fase da **Estatística Aplicada**. É com essas características que a Estatística é hoje reconhecida, pois informações numéricas são necessárias para cidadãos e organizações de qualquer natureza e de qualquer parte do mundo globalizado. Portanto, é uma ciência moderna, imprescindível para entender aspectos e problemas em todas as áreas do conhecimento.

Projeto instrucional

Disciplina: Estatística (carga horária: 30h).

Ementa: Introdução à estatística, divisão da estatística, fases do método estatístico, tabelas e gráficos, distribuição de frequência, medidas de posição, medidas de dispersão, medidas de assimentria, probabilidade e controle estatístico de processos.

AULA	OBJETIVOS DE APRENDIZAGEM	MATERIAIS	CARGA HORÁRIA (horas)
1. Conceitos	Definir estatística e seus tipos. Reconhecer os processos relacionados ao estudo da população na estatística. Diferenciar e aplicar os métodos científicos. Diferenciar e determinar as variáveis quantitativas e qualitativas.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	02
2. Fases do método estatístico	Conhecer e analisar as fases do método estatístico. Empregar as fases do método estatístico.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	02
3. Tabelas	Conhecer e aplicar a representação tabular. Entender os elementos constituintes de uma tabela.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	02
4. Séries estatísticas	Reconhecer, através da série estatística, a existência de seus fatores. Empregar a nomenclatura técnica no estudo e na interpretação da série estatística.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	02
5. Distribuição de frequências	Compreender a representação dos dados (amostrais e populacionais). Compreender e reconhecer os elementos de uma distribuição de frequência. Conhecer os tipos de frequência.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	07
6.Gráficos estatísticos	Conhecer e identificar os tipos de gráficos existentes. Interpretar e comparar os valores colocados nos gráficos.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	04

AULA	OBJETIVOS DE APRENDIZAGEM	MATERIAIS	CARGA HORÁRIA (horas)
7. Medidas descritivas	Reconhecer as medidas descritivas básicas. Aplicar as medidas descritivas, de acordo com a utilização exigida. Analisar e interpretar os procedimentos e fórmulas utilizados nas medidas descritivas, diferenciando-os.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	07
8. Probabilidade	Conhecer e entender a ideia do trabalho estatístico. Diferenciar e aplicar o espaço amostral, experimento aleatório e evento. Aplicar e conhecer o cálculo da probabilidade da ocorrência de um evento.	Ambiente virtual: plataforma *moodle*; Apostila didática; Recursos de apoio: *links*, exercícios.	04

Aula 1 – Conceitos

Objetivos

Definir estatística e seus tipos.

Reconhecer os processos relacionados ao estudo da população na estatística.

Diferenciar e aplicar os métodos científicos.

Diferenciar e determinar as variáveis quantitativas e qualitativas.

1.1 Apresentação dos conceitos de estatística

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, devem ser expressas em termos estatísticos.

Existem muitas definições propostas por autores que objetivam 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.

1.2 Método científico

1.2.1 Método

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

1.2.2 Método experimental

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

1.2.3 Método estatístico

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

1.3 O que é estatística?

Saiba mais sobre a história da estatística em:
http://www.im.ufrj.br/~lpbraga/prob1/historia_estatistica.pdf

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

Figura 1.1: Para que serve a Estatística?
Fonte: CTISM

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

Esse 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: Estatística Descritiva e Estatística Indutiva ou Inferencial.

1.3.1 Estatística descritiva – coleta, organização e descrição de dados

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.

Assista a uma aula sobre estatística descritiva
http://www.youtube.com/watch?v=bWG2AcdXDY4

Essa área apenas descreve e analisa um conjunto de dados, **sem tirar conclusões**.

Exemplo

Os 150 trabalhadores de uma fábrica ganham em média R$ 1.200,00 por mês.

1.3.2 Estatística indutiva ou inferencial e/ou inferência estatística – análise e interpretação de dados

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

Saiba mais sobre método experimental em:
http://www.prof2000.pt/users/isis/psique/unidade1/metodos/experimental.html

É a que trata das inferências e conclusões, isto é, a partir da análise de dados **são tiradas conclusões**. A inferência refere-se ao processo de generalização a partir de resultados particulares.

Exemplo

Um teste à opinião revelou que 65% da população brasileira apoia um determinado candidato para presidente da república. Se este candidato for realmente inscrito para as eleições, é de se esperar que ele se eleja.

O processo de generalização do método indutivo está associado a uma margem de incerteza. Isso 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.

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

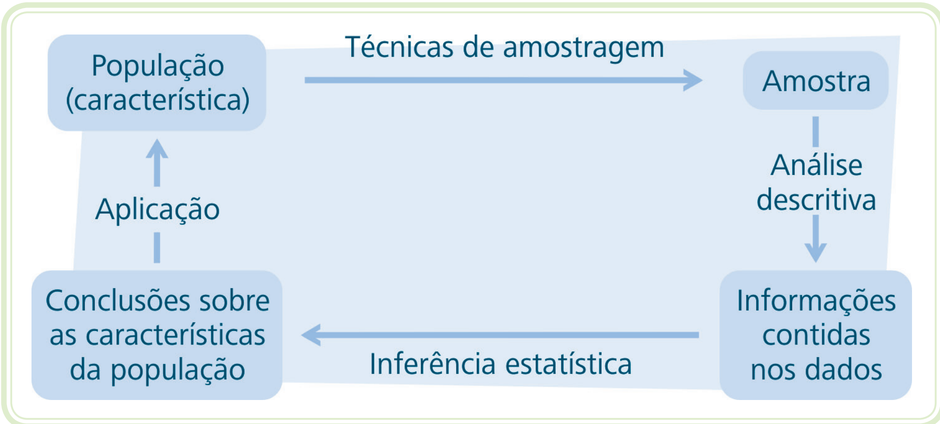


Figura1.2: Análise estatística de dados de uma amostra
Fonte: CTISM

1.4 População

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

1.4.1 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 (inferir). Em outras palavras, é o conjunto de todas as medidas e observações relativas ao estudo de determinado fenômeno que formam o universo de nosso estudo.

Exemplo
Os estudantes constituem uma população com uma característica em comum: matriculados em um determinado curso de uma universidade.

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 os que não pertencem.

A população pode ser constituída por pessoas, animais, minerais, vegetais, etc.

Exemplos

1. Deseja-se saber se nas indústrias situadas no Estado do Rio Grande do Sul, em 2010, existia algum tipo de controle ambiental.

   População ou universo – indústrias situadas no Estado do Rio Grande do Sul em 2010.

   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 Santa Maria – RS, no ano de 2010.

   População ou universo – todas as residências que estavam ligadas à rede elétrica em Santa Maria - RS, no ano de 2010.

   Características – consumo anual de energia elétrica em Mwh.

a)

1.4.1.1 Divisão da população

a) **População finita** – apresenta um número limitado de elementos. Dessa forma, é possível enumerar todos os elementos componentes.

Exemplo
Idade dos alunos do Curso Técnico em Automação Industrial do CTISM em 2010.

População – todos os alunos do Curso Técnico em Automação Industrial do CTISM, em 2010.

b) **População infinita** – apresenta um número ilimitado de elementos, o que torna impossível a enumeração de todos os elementos componentes. Entretanto, tal definição existe apenas no campo teórico, porque na prática, nunca encontraremos populações com infinitos elementos, mas somente, populações com grande número de componentes. Nessas circunstâncias, tais populações são tratadas como se fossem infinitas.

Exemplo
Febre aviária.

População – aves.

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

Observação
Em grandes populações, torna-se interessante a realização de uma amostragem, que ocorre na impossibilidade de colher informações sobre a população total.

1.5 Amostra

É uma parte ou um subconjunto representativo de uma população, isto é, é um conjunto de elementos extraídos da população. Os dados de observação registrados na amostra fornecem informações sobre a população. O processo pelo qual são tiradas conclusões sobre a população, com base nos resultados obtidos na amostra, refere-se à inferência estatística vista na definição. 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 sejam, em geral, as mesmas que as do todo (população). Enfim, a amostra é um subconjunto finito e representativo de uma população.

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

Exemplo
Para se verificar a concentração de poluentes no ar de algumas casas noturnas, é feita a verificação por um aparelho que testa uma pequena quantidade de ar.

1.6 Amostragem

É a coleta das informações de parte da população chamada amostra mediante métodos adequados de seleção dessas unidades. Amostragem é considerada uma técnica especial de escolher amostras, de forma a garantir 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.

1.6.1 Técnicas de amostragem

Existem técnicas adequadas para recolher amostras, de forma a garantir (tanto quanto possível) o sucesso da pesquisa e dos resultados.

Definidos os objetivos e a população a ser estudada, deve-se pensar em como será constituída a amostra dos dados e quais as características ou variáveis a serem estudadas.

a) **Amostragem casual ou aleatória simples** – este tipo de amostragem é baseado no sorteio da amostra. Numera-se a população de 1 a *n*, e, utilizando um dispositivo aleatório qualquer, por exemplo, sorteio, escolhem-se *k* números dessa sequência, os quais corresponderão aos elementos da amostra.

b) **Amostragem proporcional estratificada** – quando as populações se dividem em subpopulações (estratos), pode ser razoável supor que a variável de interesse apresente comportamento distinto nos diferentes estratos. Assim, para que uma amostra seja representativa, é necessário utilizar-se uma amostragem proporcional estratificada, que considera os estratos (subgrupos) e obtém a amostragem proporcional a estes.

Exemplo
Suponhamos um grupo de 90 alunos de uma escola, onde 54 sejam meninos e 36 sejam meninas. Teremos dois estratos (sexo masculino e sexo feminino) e queremos uma amostra de 10% da população, assim:

- Definimos a amostra em estratos:

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

- Numeramos os alunos de 01 a 90, sendo que 01 a 54 correspondem aos meninos e de 55 a 90, meninas.

  Para determinar a amostra, efetuamos os sorteios até atingirmos 5 meninos (por exemplo, os de números 05, 17, 31, 46 e 53) e quatro meninas (por exemplo, as de números 63, 74, 75 e 90).

- Nesse caso, serão obtidas as características dos seguintes alunos:

  Meninos: 05, 17, 31, 46 e 53.

  Meninas: 63, 74, 75 e 90.

c) **Amostragem estratificada uniforme** – não utiliza o critério de proporcionalidade, pois se seleciona a mesma quantidade de elementos de cada estrato, devendo ser usada para comparar os estratos ou obter estimativas separadas para cada estrato.

Exemplo
Uma empresa de automação conta com 480 funcionários, dos quais 288 são do sexo feminino e os 192 restantes do sexo masculino. Considerando a variável "sexo" para estratificar essa população, vamos obter uma amostra estratificada uniforme de 50 funcionários.

Supondo que haja homogeneidade dentro de cada categoria, pode-se obter amostra estratificada uniforme de 50 funcionários com a seleção de 25 elementos de cada estrato.

Estrato (por sexo)	População	Amostra estratificada uniforme
Feminino	288	$n_1 = 25$
Masculino	192	$n_2 = 25$
Total	480	50

d) Amostragem sistemática – é um procedimento para a amostragem aleatória aplicada quando os elementos da população já estão ordenados. Assim, não é necessário construir um sistema de referência ou de amostragem.

Exemplos

Casas e prédios de uma rua, os funcionários de uma empresa, as linhas de produção, lista de alunos, etc.

Num estoque de 100 peças, para obtermos 10 amostras sistemáticas podemos retirar as peças de número 10, 20, 30, e assim por diante, até completarmos 10 amostras sistematicamente colhidas.

Para encontrarmos os pontos onde faremos as coletas sistemáticas das amostras, podemos seguir os seguintes passos:

1º) Define-se tamanho da população: $N = 1.600$.

2º) Define-se o tamanho da amostragem total: $n = 100$.

$$I = \frac{N}{n} = \frac{1600}{100} = 16$$

Onde: I é igual ao intervalo de seleção

3º) Sorteia-se um número de 1 a 16, que será o primeiro número da amostra, logo, as próximas amostras serão retiradas de 16 em 16.

Observação

O intervalo da seleção corresponde ao número de vezes que a amostra cabe na população.

e) **Amostragem por estágios múltiplos** – essa estratégia de amostragem pode ser vista como uma combinação de dois ou mais planos amostrais.

Exemplo
Uma população estratificada onde o número de estratos é muito grande, ao invés de sortear uma amostra de cada estrato, o que poderia ser inviável devido à quantidade de estratos, o pesquisador poderia optar por sortear alguns estratos e, em seguida, selecionar uma amostra de cada estrato sorteado. Neste caso, teríamos uma amostragem em dois estágios usando, nas duas vezes, a amostragem aleatória simples, sendo que no primeiro estágio as unidades amostrais são os estratos e no segundo são as componentes da população.

f) **Amostragem de conveniência** – para casos onde é viável realizar um sorteio entre todos os componentes da população alvo.

Exemplo
Pesquisas clínicas.

g) **Amostragem por meio de conglomerados** – é aplicada quando a população apresenta uma subdivisão em pequenos grupos, chamados conglomerados. É possível e, muitas vezes, conveniente fazer 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 conglomerados e não mais elementos individuais da população. A amostragem por meio de conglomerados é adotada por motivos de ordem prática e econômica.

Exemplo
Quarteirões de um bairro.

1.7 Censo

Saiba mais sobre recenseamento em: http://www.somatematica.com.br/estat/basica/pagina2.php

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 pelo censo das favelas do PAC (Programa de Aceleração do Crescimento) na comunidade de Manguinhos - RJ
Fonte: http://www.rj.gov.br/web/imprensa

1.8 Variáveis estatísticas

Representam o atributo ou característica que se pretendem estudar em uma população ou amostra e são divididas em dois tipos: qualitativas e quantitativas.

1.8.1 Variáveis qualitativas

Variável que assume como possíveis valores qualidade ou atributos. Dividem-se em:

a) **Variáveis nominais** – quando não existe ordenação nos atributos.

Exemplo
A cor dos olhos, cor da pele, estado civil, cidade natal, marcas de carro, sexo, etc.

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 bom exemplo, assim como a faixa etária dos indivíduos.

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).

Escolaridade (ensino fundamental, médio e superior), mês de observação (janeiro, fevereiro e março...), grau de satisfação (escala de 0 a 5).

Observação
A resposta é expressa através de "palavras".

1.8.2 Variáveis quantitativas

São características que podem ser medidas em escala qualitativa, ou seja, apresentam valores numéricos que fazem sentido.

a) **Variáveis contínuas** – são aquelas que podem assumir qualquer valor num certo intervalo (contínuo) da reta real. Essas variáveis, geralmente, provêm de medições.

Exemplo
A altura dos alunos é uma variável contínua, pois teoricamente, um aluno poderá possuir altura igual a 1,70 m, 1,71 m, 1,711 m, 1,712 m (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.

Observação
A resposta é expressa em "valores numéricos".

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

Resumo

Nesta aula iniciamos o estudo relacionado à estatística de uma forma geral. Foi possível perceber como ela pode ser aplicada no estudo de população. Estudamos também as relações entre população, amostra e amostragem e, por fim, estudamos as características relacionadas à população, conhecidas também como variáveis.

Atividades de aprendizagem

1. Explique o que se entende por estatística.

2. Como a estatística se divide?

3. Quais as diferenças entre população, amostra e amostragem?

4. Defina os métodos de amostragem.

5. Qual o significado do censo?

Aula 2 – Fases do método estatístico

Objetivos

Conhecer e analisar as fases do método estatístico.

Empregar as fases do método estatístico.

2.1 Apresentação das fases do método estatístico

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

2.2 Divisão das fases

As principais fases do método estatístico são:

- Definição do problema.
- Planejamento.
- Coleta de dados.
- Apuração dos dados.
- Apresentação dos 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.

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

A seguir listamos algumas perguntas presentes nessa fase:

Como levantar informações? Que dados deverão ser obtidos? Qual levantamento a ser utilizado? Censitário? Por amostragem? E o cronograma de atividades? E 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, o passo seguinte é a coleta de dados, que pode ser de dois tipos:

- **Coleta direta**

A coleta direta de dados ocorre quando os dados são obtidos pelo próprio pesquisador, através de levantamento de registros (nascimentos, óbitos, notas fiscal, impostos, entre outros), ou coletados diretamente através de inquéritos, questionários, etc. Esses dados são chamados de **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. Exemplo: registros.

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

Ocasional – quando feita em determinada situação para atender a um objetivo, como pesquisa de mortalidade de um rebanho, pesquisa de um produto no mercado, etc. Exemplo: casos emergenciais.

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.

- **Coleta indireta**

A **coleta indireta** é inferida 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 de dados secundários, os dados são obtidos de algo já disponível (dados primários).

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

Observação
É mais seguro trabalhar com fontes primárias. O uso da fonte secundária traz 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 a tabulação de dados. É a etapa de soma e processamento dos dados obtidos mediante critérios de classificação. Pode ser manual ou eletrônica.

e) **5ª fase – Apresentação dos dados** – há duas formas de apresentação que não se excluem mutuamente. Primeiro, 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, segundo, a apresentação gráfica dos dados numéricos, que constitui uma apresentação geométrica que permite uma visão rápida e clara do fenômeno.

f) **6ª fase – Análise e interpretação dos dados** – a última fase do trabalho estatístico é a mais importante e a mais delicada. Está ligada essencialmente ao cálculo de medidas e coeficientes cuja finalidade principal é descrever o fenômeno. Porém esse processo de generalização, está além do escopo do nosso trabalho nessa disciplina, constituindo um campo mais avançado da Estatística que, como já foi comentado, é da estatística indutiva.

Resumo

Nessa aula 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.

Atividades de aprendizagem

1. Cite as fases do método estatístico.

2. Explique como ocorre a fase de coleta de dados.

3. Como se faz um planejamento através do método estatístico?

Aula 3 – Tabelas

Objetivos

Conhecer e aplicar a representação tabular.

Entender os elementos constituintes de uma tabela.

3.1 Apresentação de tabelas

As tabelas são recursos utilizados pela estatística, com o objetivo de organizar e facilitar a visualização e comparação dos dados.

As tabelas permitem uma visão geral dos valores assumidos pelas variáveis dentro de certos parâmetros.

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.

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

3.2 Elementos de uma tabela

Idade e sexo dos alunos matriculados no curso de Automação

Idade	Sexo	Idade	Sexo
18	F	20	M
18	M	20	M
19	F	21	F
19	F	21	M
19	F	21	M
19	M	22	F
19	M	23	F
19	M	26	M
20	F	32	F
20	F	41	M

Fonte: coordenação do curso. (rodapé)

Título

Cabeçalho

Célula

Corpo

Fonte

Figura 3.1: Elementos de uma tabela
Fonte: CTISM

Corpo	Conjunto de linhas e colunas que contêm informações sobre a variável.
Título	Conjunto das informações, (mais completas possíveis) que responde 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 à 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 à exatidão de determinado valor.

4. Zero (0) quando o valor é muito pequeno para ser expresso pela unidade utilizada. Se os valores são expressos em numerais decimais, precisamos acrescentar à 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.

Resumo

Nessa aula, estudamos as tabelas que têm por finalidade expor sinteticamente e num só local os resultados sobre determinado assunto, de modo que se obtenha uma visão global mais rápida daquilo que está em análise.

Atividades de aprendizagem

1. Cite os elementos que fazem parte de uma tabela.

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

Aula 4 – Séries estatísticas

Objetivos

Reconhecer, através da série estatística, a existência de seus fatores.

Empregar a nomenclatura técnica no estudo e na interpretação da série estatística.

4.1 Apresentação de séries estatísticas

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

Leia o texto sobre séries estatísticas em: http://www.oderson.com/educacao/estatistica/series-revisao.doc

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 eles podem ser divididos, conforme o que se apresenta a seguir.

4.2 Série temporal, histórica, cronológica ou evolutiva

É a série cujos dados estão em correspondência com o tempo, ou seja, variam com o tempo, enquanto o fato e o local permanecem constantes.

Exemplo

GAÚCHA VEÍCULOS LTDA
Vendas no primeiro bimestre de 2010

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

* Em mil unidades.
Fonte: dados fictícios.

4.3 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), enquanto o tempo e o fato permanecem constantes.

Exemplo

GAÚCHA VEÍCULOS LTDA
Vendas no primeiro bimestre de 2010

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

* Em mil unidades.
Fonte: dados fictícios.

4.4 Série específica ou categórica

É a série cujos dados estão em correspondência com a espécie, ou seja, variam com o fenômeno. O local e o tempo permanecem constantes, enquanto o fato varia.

Exemplo

GAÚCHA VEÍCULOS LTDA
Vendas no primeiro bimestre de 2010

Marca	Unidades vendidas*
Fiat	18
Chevrolet	12
Total	30

* Em mil unidades.
Fonte: dados fictícios.

4.5 Séries mistas, conjugadas ou tabela de dupla entrada

As combinações entre as séries anteriores constituem novas séries que são denominadas séries compostas ou mistas e são apresentadas em tabelas de dupla entrada, e permitem variar simultaneamente o tempo, o lugar e o fato, havendo duas ordens de classificação: uma horizontal e outra vertical.

Para obter mais informações sobre história da estatística, acesse:
http://www.im.ufrj.br/~lpbraga/prob1/historia_estatistica.pdf

Exemplo

GAÚCHA VEÍCULOS LTDA
Vendas no primeiro bimestre de 2010

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

* Em mil unidades.
Fonte: dados fictícios.

GAÚCHA VEÍCULOS LTDA
Vendas no primeiro bimestre de 2010

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

* Em mil unidades.
Fonte: dados fictícios.

Resumo

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

Atividades de aprendizagem

1. Qual a definição de série estatística e quais os fatores que a compõem?

2. Elabore um exemplo de série geográfica.

Aula 5 – Distribuição de frequência

Objetivos

Compreender a representação dos dados (amostrais e populacionais).

Compreender e reconhecer os elementos de uma distribuição de frequência.

Conhecer os tipos de frequência.

5.1 Apresentação da distribuição de frequência

Após a realização de uma pesquisa em que os dados foram coletados, é necessário organizá-los e classificá-los. Isso pode ser feito mediante tabelas e gráficos. Em geral, construímos inicialmente uma tabela que contemple as informações coletadas em função dos respectivos parâmetros.

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 aula, usaremos o mesmo exemplo em todos os tópicos.

5.2 Definições básicas

Dados primitivos ou brutos – são os dados coletados durante a pesquisa e que ainda não foram organizados.

Rol – é a ordenação dos valores obtidos (dados brutos) em ordem crescente ou decrescente de grandeza numérica ou qualitativa.

Classe – ao organizar os dados coletados, estes são subdivididos convenientemente em categorias. Cada uma dessas subdivisões recebe o nome de classe.

5.3 Representação dos dados (amostrais ou populacionais)

As tabelas, chamadas primitivas e rol, são utilizadas nas representações de dados que não estão organizados numericamente (dados brutos).

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	22	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
21	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 menor valor, $\Delta x = 49 - 20 = 29$ anos). Assim, temos:

20, 20, 20, 21, 21, 22, 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, 45, 46, 46, 47, 48, 49.

5.4 Distribuição de frequência

Uma maneira mais concisa de mostrar os dados do rol é apresentar cada dado seguido pelo número de vezes em 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 3 vezes, assim temos $f(20) = 3$. Já a idade de 46 anos ocorre 2 vezes, que se escreve $f(46) = 2$.

A tabela que contém todos os valores com a sua frequência recebe o nome de **distribuição de frequência**.

Exemplo

Idade de 50 trabalhadores da indústria A

Idade	Freq.	Idade	Freq.	Idade	Freq.
20	3	30	2	41	3
21	2	31	1	42	1
22	4	33	3	44	1
23	3	34	2	45	4
24	2	35	1	46	2
25	2	37	2	47	1
26	2	38	1	48	1
28	2	39	1	49	1
29	1	40	2		

Ainda assim, o processo exige muito espaço, em especial, quando o número de valores da variável (n) aumenta. O mais razoável nesses casos, quando a variável é contínua, é agrupar os valores por intervalos. Desse 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 3 trabalhadores com 20 anos, 2 trabalhadores com 21 anos, etc., coloca-se, por exemplo, 12 trabalhadores entre 20 e 24 anos. Este intervalo é escrito como 20 ⊢ 24 que corresponde a 20 (símbolo de menor ou igual) $x < 24$ (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:

Idade de 50 trabalhadores da indústria A.

Idades	Frequência
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.

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.

Observação
" – ": 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.

5.5 Distribuição de frequência de dados numéricos agrupados em intervalos de classe

Quando os dados numéricos coletados assumem grande quantidade de valores diversificados, fica inviável que cada valor isolado represente uma categoria. Sendo assim, convém agruparmos os valores coletados em intervalos de classe.

Passos necessários

- Determinação do número de classes (*K*).
- Amplitude amostral (*AA*).
- Cálculo de amplitude do intervalo de classe (*h*).
- Limite inferior e superior do intervalo de classe.
- Determinação dos intervalos de classe.
- Determinação das frequências dos intervalos de classe.
- Amplitude total (*AT*).

5.5.1 Determinação do número de classes (*K*)

O número de classes que irá compor a tabela pode ser estabelecido pelo estatístico que elabora a pesquisa, pois ele é soberano para decidir o número de classes. Para tornar o processo mais uniforme, existem duas maneiras para estabelecer o número de classes em função do número de dados da tabela.

a) **Primeira maneira** – é a fórmula desenvolvida pelo matemático Sturges para o cálculo do número de classes.

$$K = 1 + 3{,}322 \,.\, \log n$$

Sendo: K = número de classes

n = número de elementos coletados na pesquisa

b) **Segunda maneira** – recomenda-se a utilização quando o número de dados coletados for menor ou igual a 50.

$$K = \sqrt{n} \text{ para } n \leq 50$$

Exemplos

O número de classes pela fórmula de Sturges, quando $n = 30$ é:

$$K = 1 + 3{,}322 \,.\, \log 30 = 5{,}90 \cong 6 \text{ classes}$$

O número de classes utilizando a fórmula $K = \sqrt{n}$, quando $n = 30$ é:

$$K = \sqrt{n} = 5{,}48 = 5 \text{ classes}$$

O número de classes pela fórmula de Sturges, quando n = 120 é:

$$K = 1 + 3{,}322 \,.\, \log 120 = 7{,}91 \cong 8 \text{ classes}$$

O número de classes utilizando a fórmula K = $\sqrt{n}$, quando n = 120 é:

$$K = \sqrt{n} = 10{,}95 = 11 \text{ classes}$$

5.5.2 Amplitude amostral (*AA*)

É a diferença entre o maior e o menor valor observado nos valores coletados.

$$AA = X_{máx.} - X_{mín.} \text{ (maior valor observado - menor valor observado)}$$

5.5.3 Cálculo da amplitude do intervalo de classe (*h*)

Definido com quantas classes vamos construir a tabela de distribuição de frequência, pode-se passar ao cálculo de amplitude do intervalo de classe.

h = amplitude de classe ou amplitude do intervalo de classe

$$\text{Amplitude de classe } (h) = \frac{\text{maior valor observado - menor valor observado}}{\text{número de classes}}$$

Ou seja: É a razão entre:

$$h = \frac{AA}{K}$$

Observação
O valor obtido no cálculo da amplitude de classe nem sempre é um valor exato. Sendo assim, para preservar o número de classes estabelecido, faz-se o arredondamento da amplitude de classe para valores acima do valor obtido.

Exemplo
Se todos os dados coletados forem múltiplos de 5, a amplitude de intervalo de classe deverá ser arredondada para o múltiplo de 5 imediatamente superior ao valor obtido.

5.5.4 Limites inferior e superior do intervalo de classe

Os valores do conjunto que compõe o intervalo de classe estão limitados entre dois números que representam os extremos inferior e superior do intervalo.

Apresentação da convenção matemática para os símbolos utilizados na representação de intervalos abertos e fechados.

Tipo de intervalo	Símbolos de representação		
	1º tipo	2º tipo	3º tipo
Intervalo fechado à esquerda e fechado à direita	⊢⊣	[a, b]	[a, b]
Intervalo aberto à esquerda e aberto à direita	—	]a, b[	(a, b)
Intervalo fechado à esquerda e aberto à direita	⊢	[a, b[	[a, b)
Intervalo aberto à esquerda e fechado à direita	⊣	]a, b]	(a, b]

- Limite inferior (l_i) é o menor valor numérico do intervalo de classe.

- Limite superior (L_i) é o maior valor numérico do intervalo de classe.

O símbolo " ⊢" estabelece "Inclusão" ou "exclusão" para os valores limites num intervalo de classe.

Exemplo

Dado o intervalo 15 ⊢ 22

- O valor do limite inferior é 15
- O valor do limite superior é 22.
- O valor 15 está dentro do intervalo, pois o valor pertence ao intervalo, estando fechado à esquerda.
- O valor 22 está fora do intervalo, não pertencendo ao intervalo, pois o intervalo é aberto à direita.
- A amplitude do intervalo de classe é $h = 7$, pois $h = 22 - 15 = 7$.

5.5.5 Amplitude total (*AT*)

Calcula-se em relação aos valores dos intervalos de classe da tabela de distribuição de frequência.

$$\text{Amplitude de classe } (h) = \frac{\text{maior valor observado - menor valor observado}}{\text{número de classes}}$$

5.5.6 Ponto médio de uma classe (X_i)

É o ponto que, por situar-se numa posição média da distribuição de valores do intervalo de classe, divide o intervalo em duas partes iguais.

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

Exemplo

Se considerarmos um intervalo de classe de valores: 30 ⊢ 44, o ponto médio desse intervalo é calculado pela expressão:

$$X_i = \frac{l_i + L_i}{2} = \frac{30 + 44}{2} = \frac{74}{2} = 37$$

5.6 Tipos de frequência

5.6.1 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:

$f_1 = 12;\ f_2 = 6;\ f_3 = 6;\ f_4 = 6;\ f_5 = 4;\ f_6 = 6;\ f_7 = 8;\ f_8 = 2.$

A soma de todas as frequências é representada pelo símbolo de somatório ($\sum$), onde $\sum_{i=1} f_i$ significa a soma dos f_i, sendo que i vai de 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:

Idade de 50 trabalhadores de indústria A

i	Idade	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.

5.6.2 Frequência relativa (fr_i)

Frequência relativa de uma classe é o quociente entre a frequência absoluta (f_i) da classe considerada e o número total de dados (n) coletados na pesquisa.

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

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

Frequência relativa percentual ($fr_i\%$) de uma classe é o produto da frequência relativa por 100.

$$fr_i\,\% = fr_i \times 100$$

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%.

5.6.3 Frequência acumulada (fac_i ou f_i)

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).

5.6.4 Frequência relativa acumulada ($fRAC_i$ ou fR_i)

Frequência relativa acumulada (fR_i) é a frequência acumulada da classe dividida pela frequência total da distribuição, ou a razão entre a frequência acumulada da classe considerada e o número total de dados (n) coletados na pesquisa:

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

Frequência relativa acumulada percentual ($fRAC_i\%$ ou $fR_i\%$) de uma classe é o produto da frequência relativa acumulada por 100%.

$$fRAC_i\ \% = fRAC_i \times 100$$

Exemplo

Para a terceira classe, temos:

$$fR_3 = \frac{f_3}{n} = \frac{24}{50} = 0{,}4$$

Significa que a fração de 0,4 trabalhadores (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	$fr_i\%$	f_i	fR_i	$fR_i\%$
1	$20 \vdash 24$	22	12	0.24	24	12	0.24	24
2	$24 \vdash 28$	26	6	0.12	12	18	0.36	36
3	$28 \vdash 32$	30	6	0.12	12	24	0.48	48
4	$32 \vdash 36$	34	6	0.12	12	30	0.60	60
5	$36 \vdash 40$	38	4	0.08	8	34	0.68	68
6	$40 \vdash 44$	42	6	0.12	12	40	0.80	80
7	$44 \vdash 48$	46	8	0.16	16	48	0.96	96
8	$48 \vdash 52$	50	2	0.04	4	50	1.00	100
			$\sum f_i = 50$	$\sum fr_i = 1$				

Fonte: dados fictícios.

x_i: Ponto médio de uma classe;
f_i: Frequência simples;
fr_i : Frequência relativa;
fr_i : Frequência relativa percentual;
f_i: Frequência acumulada;
fR_i : Frequência relativa acumulada;
$fR_i\%$: Frequência relativa acumulada percentual.

5.7 Gráficos representativos de uma distribuição de frequência

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 organizados e resumidos para possibilitar a obtenção de informações úteis. Uma maneira de organizar os dados é a distribuição de frequência, que pode ser representada além da forma de tabela, através de gráficos.

5.7.1 Histograma

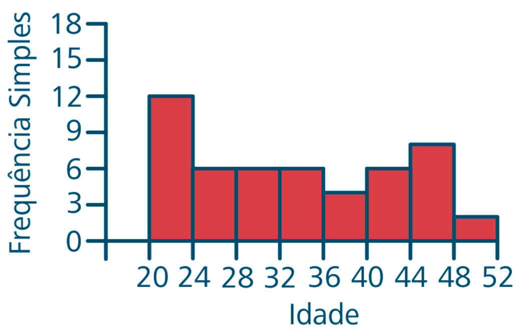
O histograma é formado por um conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal (eixo x). Sobre esse eixo são representados os intervalos de classe numa escala contínua, não sendo necessário que a escala inicie no zero, de tal modo que os seus pontos médios coincidam com os pontos médios dos intervalos de classe e seus limites coincidam com os limites da classe.

O número de retângulos encontrados em um histograma é igual ao número de intervalos de classe, sendo a largura de cada retângulo igual à amplitude do intervalo de classe, enquanto sua altura representa a frequência do intervalo de classe, e a área é proporcional à soma das frequências.

Exemplo

O histograma para a idade de 50 trabalhadores da indústria A, fica assim:

Idade (anos)	Frequência simples
20 ⊢⊣ 24	12
24 ⊢⊣ 28	6
28 ⊢⊣ 32	6
32 ⊢⊣ 36	6
36 ⊢⊣ 40	4
40 ⊢⊣ 44	6
44 ⊢⊣ 48	8
48 ⊢⊣ 52	2



Observação

A diferença entre o gráfico de coluna e o histograma é o distanciamento entre as colunas. No histograma, não há separação entre os retângulos adjacentes.

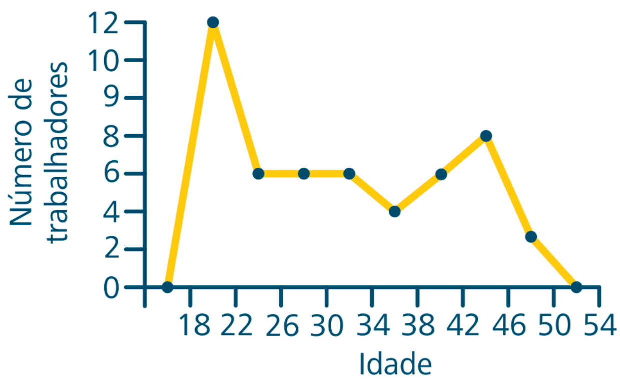
5.7.2 Polígono de frequências

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

É um gráfico de linha em que cada ponto é obtido considerando-se como valor de X, o ponto médio do intervalo de classe e, como valor de Y, a respectiva frequência do intervalo. Consideramos também uma classe anterior à primeira e outra posterior à última. Ligando todos os pontos, temos o polígono de frequência.

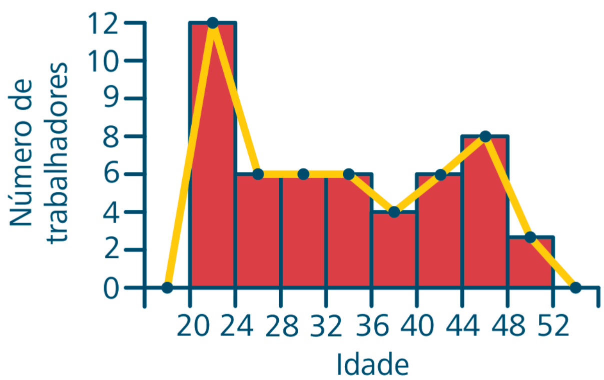
Exemplo

Observando o histograma do exemplo anterior, o polígono de frequências para esta distribuição, onde os pontos médios dos intervalos são dados por 22, 26, 30, 34 e 38, 42, 46, 50 e 54; acrescentamos à primeira e posterior à última, com frequência nula teremos:



Observação

Da mesma forma que o histograma, o polígono de frequências também apresenta área proporcional à soma das frequências. Também podemos construí-lo por meio dos pontos médios das bases superiores dos retângulos do histograma, unidos ao ponto anterior à primeira classe e ao posterior à última.



Resumo

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

Atividades de aprendizagem

1. O que são dados brutos?

2. Diferencie os elementos de uma distribuição de frequência e dê exemplos de cada um deles.

3. Cite e explique os tipos de frequência.

4. Explique o significado da regra de Sturges e dê um exemplo.

5. Utilizando um exemplo de histograma, elabore um polígono de frequência.

Aula 6 – Gráficos estatísticos

Objetivos

Conhecer e identificar os tipos de gráficos existentes.

Interpretar e comparar os valores colocados nos gráficos.

6.1 Apresentação de gráficos estatísticos

Os gráficos são desenhos que envolvem formas e cores cuja construção utiliza técnicas de desenho.

Os gráficos são de extrema importância na visualização e interpretação de informações e dados acerca de temas de aspectos naturais, sociais e econômicos. 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 permitem a representação de uma relação entre variáveis e facilitam a compreensão de dados.

Saiba mais sobre construção de gráficos em: http://www.netknow.mat.br/atividades/Graficos.pdf

Além disso, os gráficos devem ser correspondentes às tabelas estatísticas, mas não devem substituí-las. Desse modo, a representação gráfica é um complemento importante da apresentação tabular.

A vantagem de um gráfico sobre a tabela está na possibilidade de 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, tais como:

Simplicidade – deve ser destituído de detalhes e traços desnecessários.

Clareza – deve possuir uma correta interpretação dos valores representativos do fenômeno em estudo.

Veracidade – deve expressar a verdade sobre o fenômeno em estudo.

6.2 Formas de apresentação dos gráficos

Quando for levado em conta o visual de acordo com a composição de formas, os gráficos podem ser:

6.2.1 Gráficos de informação

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

6.2.2 Gráficos de análise

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

6.3 Principais tipos de gráficos

6.3.1 Gráfico em curvas ou em linhas

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

Os dados de uma tabela geralmente são colocados num sistema cartesiano ortogonal, tendo pontos ligados por segmentos de reta.

Exemplo

Relação entre consumo e produção de óleo combustível no RS no período compreendido entre Jan/Ago de 2010 (em milhões de barris).

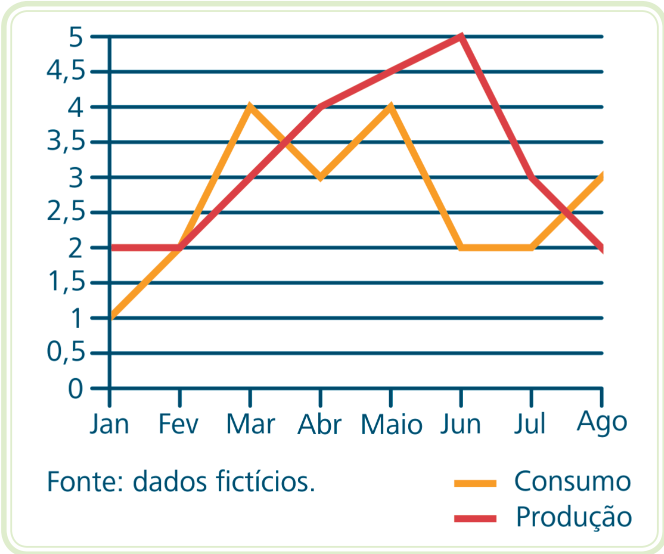


Figura 6.1: Gráfico em linhas
Fonte: CTISM

6.3.2 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). Pode também conter barras múltiplas. Esse 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 2010 (em milhões de barris).

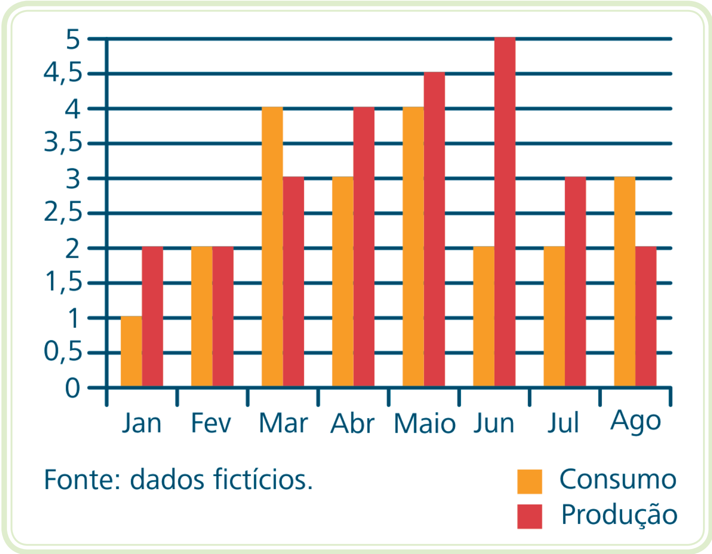


Figura 6.2: Gráfico em colunas
Fonte: CTISM

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.

6.3.3 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 2010 (em milhões de barris).

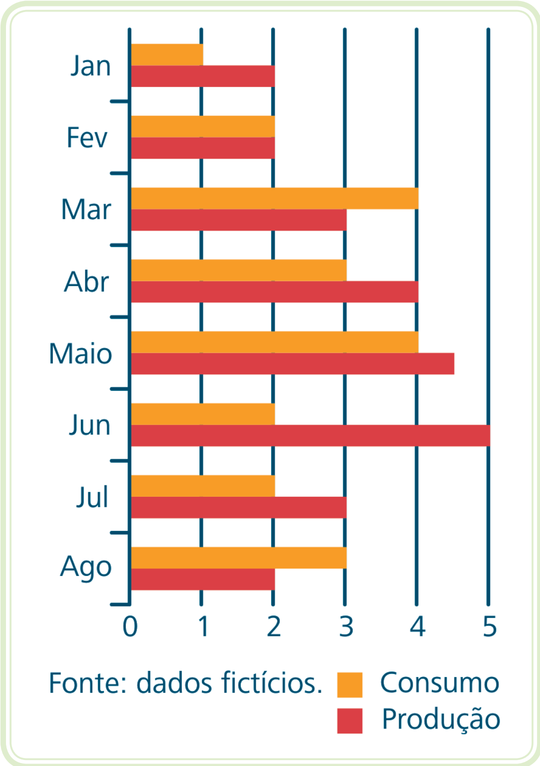


Figura 6.3: Gráfico em barras
Fonte: CTISM

6.3.4 Gráfico em setores

Os dados são apresentados em setores circulares proporcionais aos valores a eles atribuídos. 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 360º (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/Ago de 2010 (em milhões de barris).

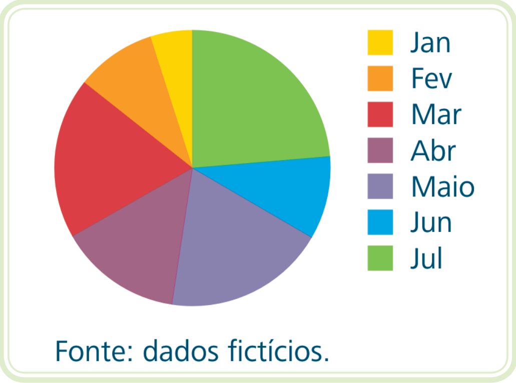


Figura 6.4: Gráfico em setores
Fonte: CTISM

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:

$$\begin{aligned} \text{Total (\%)} &\longrightarrow 360° \\ \text{Parte (\%)} &\longrightarrow x° \end{aligned}$$

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: $100 \times x = 360 \times 25 \rightarrow x = 9000 \div 100 \rightarrow x = 90°$
18% piscicultura: $100 \times x = 360 \times 18 \rightarrow x = 6480 \div 100 \rightarrow x = 64{,}8°$
31% leite: $100 \times x = 360 \times 31 \rightarrow x = 11160 \div 100 \rightarrow x = 111{,}6°$
17% hortaliças: $100 \times x = 360 \times 17 \rightarrow x = 6120 \div 100 \rightarrow x = 61{,}2°$
9% fruticultura: $100 \times x = 360 \times 9 \rightarrow x = 3240 \div 100 \rightarrow x = 32{,}4°$

Assim temos:

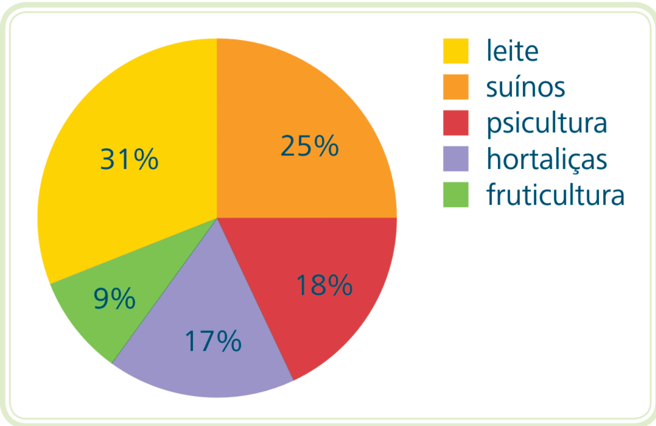


Figura 6.5: Produção rural
Fonte: CTISM

Resumo

Nesta aula foram estudados os gráficos, os quais propiciam uma ideia inicial mais satisfatória da concentração e da dispersão dos valores, uma vez que, através deles, os dados estatísticos se apresentam em termos de grandezas visualmente interpretáveis.

Atividades de aprendizagem

1. Explique os tipos de gráficos e dê um exemplo de cada um deles.

2. Qual a principal vantagem de um gráfico em relação a uma tabela?

3. A representação gráfica deve obedecer a certos requisitos. Quais são eles?

Para obter mais informações, acesse: http://intervox.nce.ufrj.br/~diniz/d/direito/ou-estatística.doc

Aula 7 – Medidas descritivas

Objetivos

Reconhecer as medidas descritivas básicas.

Aplicar as medidas descritivas, de acordo com a utilização exigida.

Analisar e interpretar os procedimentos e fórmulas utilizados nas medidas descritivas, diferenciando-os.

7.1 Apresentação das medidas descritivas

Na maior parte das vezes em que os dados estatísticos são analisados, procuramos obter um valor para representar um conjunto de dados. Esse valor deve sintetizar, da melhor maneira possível, o comportamento do conjunto do qual ele é originário. Com isso, a Estatística Descritiva visa descrever os dados disponíveis da forma mais completa possível sem, no entanto, preocupar-se em tirar conclusões sobre um conjunto maior de dados.

saiba mais sobre estatística em: http://www.intervox.nce.ufrj.br/~diniz/d/direito/ou-estatistica.doc

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

As medidas descritivas básicas mais importantes são as de:

Posição, **dispersão** ou **variabilidade**, **momentos** e **assimetria**.

As medidas de **posição** dividem-se 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 biquadrática.

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

7.2 Medidas de posição

7.2.1 Medidas de tendência central (promédios)

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

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.

7.2.1.1 Média aritmética

Simbolizada por $\bar{x}$ (*x* barra), no caso de amostra, e μ (mi) para população, esta 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, isto é, é o quociente da soma dos valores de um rol pelo número de elementos. Sendo x_1, x_2, ... x_n os elementos, temos:

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

b) **Média aritmética para dados tabelados ou ponderada** – quando os dados estiverem agrupados em uma tabela de frequências, pode-se obter a média aritmética da distribuição, se os elementos x_1, x_2, ..., x_n apresentarem, respectivamente, frequências f_1, f_2, ..., f_n, ou pesos da variável, então:

$$\bar{x} = \frac{x_1 . f_1 + x_2 . f_2 + x_3 . f_3 \ldots + x_n . f_n}{f_1 + f_2 + f_3 + \ldots + f_n} = \frac{\sum_{i=1}^{n} x_i . f_i}{\sum_{i=1}^{n} f_i}$$

Exemplo

notas (x_i)	frequência (f_i)
3	3
4	5
5	6
6	7
7	6

$$\bar{x} = \frac{3.3 + 4.5 + 5.6 + 6.7 + 7.6}{3 + 5 + 6 + 7 + 6} = \frac{143}{27} = 5{,}296$$

7.2.1.2 Moda

A **moda** (*Mo*) ou norma é o valor que ocorre com maior frequência ou repetições, em um conjunto de valores. É uma medida de dominância, não 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), a moda será calculada por $Mo = (l + L) / 2$.

Exemplo

Considere as seguintes distâncias em quilômetros: 100, 90, 110, 100, 100, 2500 e 3000. A moda é o valor que mais ocorre, portanto $Mo = 100$. Neste caso a média é $\bar{x} = 500$.

a) **Distribuição unimodal** – é aquela que possui uma só moda.

Exemplo

$x_i = \{100, 90, 110, 100, 100, 2500\}$ $Mo = 100$

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 e *Mo* = 200

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

Exemplo

x_i = {1, 2, 3, 6, 7, 22, 300} Não existe *Mo*

7.2.1.3 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

{2; 4; 6; 8; 10; 12; 15; 17; 19; 21; 23}, temos que *Md* = 12, pois abaixo de 12 temos 5 números (2, 4, 6, 8, 10) e acima de 12 temos também 5 números (15, 17, 19, 21, 23).

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. Dessa forma, quando o número de observações for par, deve-se somar os dois números centrais e após dividir por dois.

$$\text{Posição mediana} = \frac{n}{2}$$

Exemplo

Para a série: {3; 5; 7; 9; 11; 13}, temos que *Md* = (7+9)/2 = 8

Uma fábrica deseja comparar o desempenho de duas máquinas, com base na produção diária de uma determinada peça durante cinco dias:

$$\text{Máquina A} : 70, 71, 69, 70, 70, \text{ sendo } \bar{x} = 70$$
$$\text{Máquina B} : 60, 80, 70, 62, 83, \text{ sendo } \bar{x} = 71$$

O desempenho médio da máquina A é de 70 peças produzidas diariamente, enquanto que a da máquina B é de 71 peças. Com base na média aritmética, verifica-se que o desempenho da B é melhor do que o da 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, o que revela que o desempenho da A é bem mais uniforme do que o da B.

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 após dividir por dois.

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

	Definição	Variação	Desvantagens
Média	Centro de distribuição de frequências.	Reflete cada valor.	É afetada por valores extremos.
Moda	Valor mais frequente	Valor "típico". 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.

7.3 Separatrizes

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

As medidas (**quartis**, os **decis** e os **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.

7.3.1 Quartis

Assista a um vídeo sobre quartis
http://www.youtube.com/watch?v=4BmgvVuj3bs

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

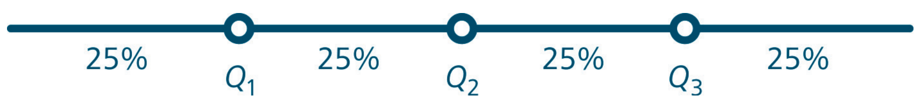
Precisamos, portanto, de 3 quartis (Q_1, Q_2 e Q_3) para dividir a série em quatro partes iguais.

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

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 um quarto (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)
p_1 e p_3 refere-se à posição quartil

Exemplos

1. Em relação às notas em Estatística dos alunos do Curso Técnico em Automação Industrial – EaD, calcule os quartis desta 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. Em relação à série: {2, 3, 3, 5, 6, 6, 8, 9, 10, 11, 11, 12}. Como a série já está ordenada, calculando o quartil 2 teremos.

$$Q_2 = Md = (6 + 8) \div 2 = 7$$

O quartil 1 será a mediana da série à esquerda de *Md*: {2, 3, 3, 5, 6, 6}.

$$Q_1 = (3 + 5) \div 2 = 4$$

O quartil 3 será a mediana da série à direita de *Md*: {8, 9, 10, 11, 11, 12}.

$$Q_3 = (10 + 11) \div 2 = 10{,}5$$

7.3.2 Decis

Decis para dados não agrupados – são valores que dividem a série em 10 partes iguais de tal forma que cada intervalo do decil contenha 10% dos elementos coletados. Assim, para dados não agrupados o cálculo das posições (decis) usaremos:

Onde: n = número de dado (valores)

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

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

$$\vdots$$

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

$$\vdots$$

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

Exemplo

Considerando o exemplo das notas de Estatística dos alunos do Curso Técnico em Automação do exercício anterior, calculamos os decis desta 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%.

7.3.3 Percentil ou centil

Percentis para dados não agrupados – são os valores que delimitam proporções dentro de uma série segundo o percentil escolhido, de tal forma que cada intervalo do percentil contenha 1% dos elementos coletados. Assim, para o cálculo das posições (percentis) usaremos:

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

E é evidente que:

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

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)

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

Exemplo

Da amostra dos exemplos anteriores, calcula-se os percentis:

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_{90}) 90%.

7.3.4 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. A *AIQ* é válida para dados ordinais, intervalares ou de razão. É calculada por:

$$AIQ = Q_3 - Q_1$$

7.4 Medidas de dispersão ou de variabilidade

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 é preciso conhecer o real grau de afastamento de um conjunto de números em relação a sua média.

As medidas de dispersão se dividem em:

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

7.4.1 Medida de dispersão absoluta

7.4.1.1 Variância (S^2)

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 média. A variância (s^2) mede a dispersão dos dados de observações de uma amostra em relação à respectiva média. Ela relaciona os desvios em torno da média, isto é, é a média aritmética dos quadrados dos desvios.

Assista a um vídeo sobre variância
http://www.youtube.com/watch?v=yVP8kLI_blU

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

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

7.4.1.2 Desvio padrão (σ ou S)

Desvio padrão (σ) é a raiz quadrada da variância, então é calculado por:

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

Por consequência, a variância é o quadrado do desvio padrão.

Exemplo

Com os dados sobre a produção diária de 3 trabalhadores de uma indústria, através do desvio padrão, podemos identificar qual deles A, B, C, apresenta menor variabilidade na produção diária de peças para carro.

Trabalhador	Dia					Média Diária	Amplitude Total
	1°	2°	3°	4°	5°		
A	82	70	65	60	73	70	22
B	60	78	68	62	82	70	22
C	53	72	75	75	75	70	22

Para a A, utilizando a definição, temos:

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

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

$$\sigma = \sqrt{\frac{(82-70)^2+(70-70)^2+(65-70)^2+(60-70)^2+(73-70)^2}{5 - 1}}$$

$$\sigma = \sqrt{69{,}55} = 8{,}34, \text{ a variância é igual a } (8{,}34)^2 = 69{,}55$$

E calculando o desvio padrão para B e para C, encontraremos os valores respectivamente iguais a 9,69 e 9,59.

Portanto, podemos observar que o trabalhador A apresenta a menor dispersão na produção diária de peças.

7.4.2 Medida de dispersão relativa

7.4.2.1 Coeficiente de variação/coeficiente de variação de *Pearson*

Quando se deseja comparar a variabilidade de duas ou mais distribuições, mesmo quando elas se referem a diferentes fenômenos e são 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, chamada de **coeficiente de variação de *Pearson***.

> "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 com temperatura. O coeficiente de variação (*CV*) é, portanto, a magnitude relativa do desvio padrão expresso em porcentagem da média:

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

Exemplos

1. Considerando uma média valendo $x = 980$ e um desvio padrão $\sigma = 58$, temos:

$$CV = \frac{58}{980} \times 100 = 5{,}7\%$$

Que indica a dispersão da amostra.

Observação

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

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

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

2. Considere uma amostra que apresenta uma distribuição de frenquência com $x = 161$ e $\sigma = 4{,}22$, logo temos coeficiente de variação valendo:

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{4{,}22}{161} \times 100 = 2{,}62 \cong 2{,}6\%$$

7.5 Momentos e assimetria

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

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

7.5.1 Momentos

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

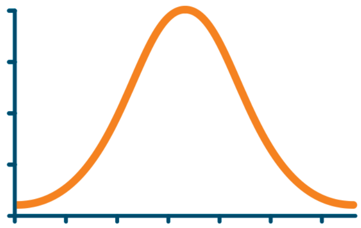
Os momentos podem ser simples ou centrados na origem, centrados na média e abstratos.

7.5.2 Assimetria e simetria

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

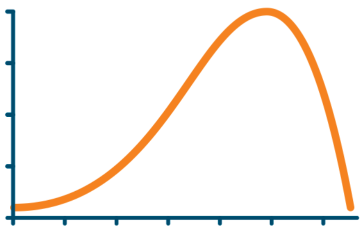
a) **Assimétrica nula ou simétrica** – uma distribuição de frequências com classes é simétrica quando:

Média = Mediana = Moda



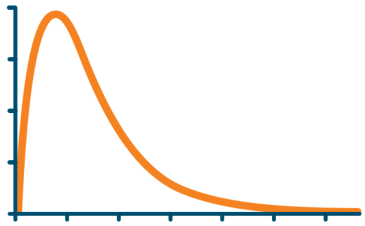
b) **Assimétrica à esquerda ou negativa** – uma distribuição de frequências com classes é assimétrica à esquerda ou negativa quando:

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, temos 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 distibuições 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
	$\sum f_i = 60$

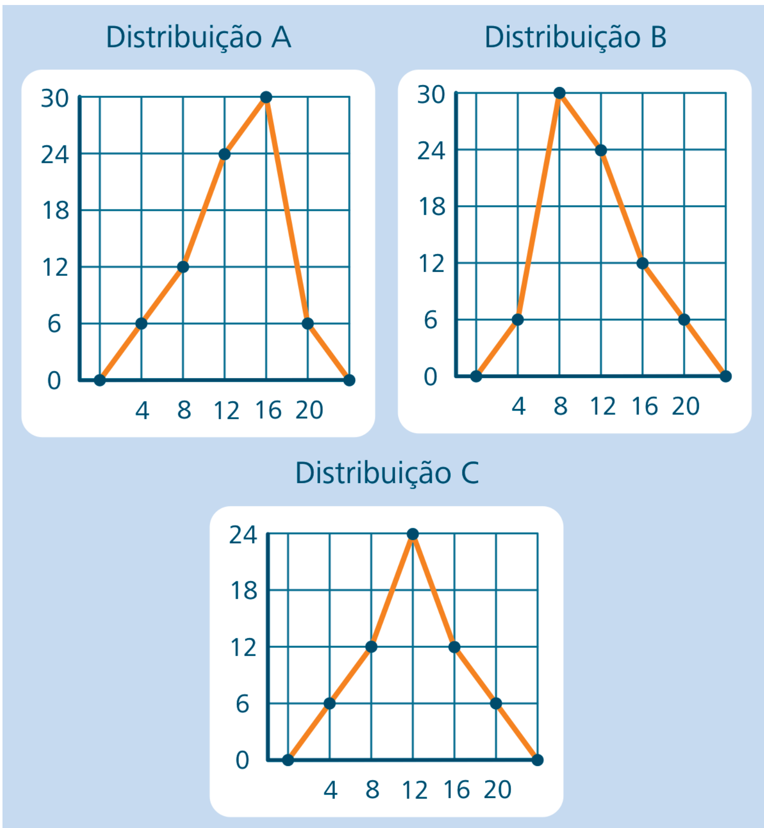
Calculando a média, a mediana, a moda e o desvio padrão para cada distribuição temos:

Distribuição A:	Distribuição B:	Distribuição C:
$\bar{x} = 12{,}9$	$\bar{x} = 11{,}1$	$\bar{x} = 12$
$Md = 13{,}5$	$Md = 10{,}5$	$Md = 12$
$Mo = 16$ kg	$Mo = 8$ kg	$Mo = 12$ kg
$\sigma = 4{,}20$	$\sigma = 4{,}20$	$\sigma = 4{,}42$

Logo, calculando o tipo de assimetria:

Distribuição A:	Distribuição B:	Distribuição C:
$\bar{x} - Mo$	$\bar{x} - Mo$	$\bar{x} - Mo$
$12{,}9 - 16 = -3{,}1$	$11{,}1 - 8 = 3{,}1$	$12 - 12 = 0$
a distribuição é assimétrica negativa	a distribuição é assimétrica positiva	a distribuição é simétrica

Construindo os gráficos das distribuições anteriores, temos:



7.5.2.1 Coeficiente de assimetria

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

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

Escalas de assimetria: $0{,}15 < |AS| < 1$ assimetria moderada
$|AS| \leq 1$ 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:

$$ASA = \left|\frac{3(12{,}9 - 13{,}5)}{4{,}20}\right| = 0{,}428 \text{ (assimetria moderada)}$$

$$ASB = \left|\frac{3(11{,}1 - 10{,}5)}{4{,}2}\right| = 0{,}428 \text{ (assimetria moderada)}$$

$$ASC = \left|\frac{3(12 - 12)}{4{,}2}\right| = 0 \text{ (simetria)}$$

Resumo

Nesta aula 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. Isso ocorre porque é comum encontrar séries que, apesar de apresentarem a mesma média, são compostas de maneiras diferentes. Dessa forma, mostramos que as medidas de tendência central são insuficientes para descrever adequadamente uma série estatística.

Atividades de aprendizagem

1. Explique o significado de variância e de desvio padrão.

2. Dê um exemplo e explique o coeficiente de variação de *Pearson*.

3. Faça um paralelo entre assimetria e simetria.

Aula 8 – Probabilidade

Objetivos

Conhecer e entender a ideia do trabalho estatístico.

Diferenciar e aplicar o espaço amostral, experimento aleatório e evento.

Aplicar e conhecer o cálculo da probabilidade da ocorrência de um evento.

8.1 Apresentação da probabilidade

No ano de 1654, um jogador da sociedade parisiense chamado Chevalier de Mére propôs ao matemático Blaise Pascal algumas questões sobre possibilidades de vencer jogos. Uma dessas 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 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 Teoria da Probabilidade.

Assista a vídeos sobre probabilidade
http://www.youtube.com/watch?v=uHSpupVbcmE

http://www.youtube.com/watch?v=KEampel5L5U

http://www.youtube.com/watch?v=A0HggMjT9JI

http://www.youtube.com/watch?v=05m5BRVw7P8

http://www.youtube.com/watch?v=mPE8zh64SL0

Em condições normais podemos prever a que temperatura o leite ferve, por exemplo. Esse tipo de experimento, cujo resultado é previsível, recebe o nome de **determinístico**. Mas, ao lançar um dado, uma ou mais vezes, não podemos saber com antecedência o número que se vai obter; sabemos apenas que os possíveis resultados são 1, 2, 3, 4, 5 ou 6. Esse tipo de experimento, cujo resultado não pode ser previsto, é chamado **aleatório**.

Já sabemos que, para se 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, possam ser feitas afirmações sobre características da população.

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

8.2 Experimento aleatório

É aquele experimento que, quando repetido em iguais condições, pode 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, temos 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.
- À 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.

Observação

Na teoria das probabilidades, estudamos os experimentos aleatórios equiprováveis, ou seja, aqueles em que qualquer resultado pode ocorrer com a mesma chance. É o caso do lançamento de uma moeda: a possibilidade de ocorrer cara ou coroa é a mesma.

8.3 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" e 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}. Portanto: n (S) = 3

8.4 Evento

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

Exemplo

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

8.4.1 Tipos de eventos

a) **Eventos certos** – são eventos que possuem todos os elementos do espaço amostral. Logo, se $E = S$, ele 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 $E = \emptyset$, 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 ($\overline{A}$) que é caracterizado pela não ocorrência daquele evento.

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, $\overline{A} = \{1, 3, 5\}$.

A probabilidade de um evento ocorrer num experimento aleatório equiprovável é dada pelo quociente da divisão do número de casos favoráveis pelo número de casos possíveis.

8.5 Conceito de probabilidade

Se em um fenômeno aleatório as possibilidades são igualmente prováveis, então a probabilidade de ocorrer o evento A é:

$$P(A) = \frac{\text{número de casos favoráveis}}{\text{número de casos possíveis}}$$

Exemplo

No lançamento de um dado, um número para pode ocorrer de 3 maneiras diferentes, dentre 6 igualmente prováveis, portanto, P= 3/6 = ½ = 50%

Dizemos que um espaço amostral S (finito) é equiprovável quando seus eventos elementares têm probabilidades iguais de ocorrência.

Num espaço amostral equiprovável S (finitos), a probabilidade de ocorrência de um evento A é sempre:

$$P(A) = \frac{\text{número de elemtos } A}{\text{número de elementos } S} = \frac{n(A)}{n(S)}$$

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 $\overline{A}$ o conjunto dos elementos das peças defeituosas, e $n\ (A)$ 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.

$$n(S)=240 \quad n(\overline{A})=20 \quad n(A)=220$$

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

$$P\ (E') = \frac{n\ (A\ )}{n\ (\ S\ )} = \frac{220}{240} = \frac{11}{12} = 0{,}916\ ...$$

Isso significa que a probabilidade de retirar uma peça não defeituosa é de 91,6% aproximadamente.

8.6 Eventos complementares

Se A e $\overline{A}$ são eventos complementares, então:

$$P(A) + P(\overline{A}) = 1$$

8.7 Probabilidade condicional

Antes da realização de um experimento, é necessário que já se tenha alguma informação sobre o evento que se deseja observar. Nesse caso, o espaço amostral se modifica e o evento tem a sua probabilidade de ocorrência alterada.

Fórmula da probabilidade condicional:

$$P(E_1 \text{ e } E_2 \text{ e } E_3 \text{ e } .... \text{ e } E_{n-1} \text{ e } E_n\ ) \text{ é igual a } P(E_1)$$
$$P(E_2\ /E_1)\ .\ P(E_3/E_1 \text{ e } E_{12})\ ...\ P\ (E_n/E_1 \text{ e } E_2 \text{ e } ...\ E_{n-1})$$

Onde:

P (E_2/E_1) é a probabilidade de ocorrer E_2 condicionada pelo fato de já ter ocorrido E_1;

P (E_3/E_1 e E_2) é a probabilidade de ocorrer E_3, condicionada pelo fato de já terem ocorrido E_1 e E_2;

P (E_n/E_1 e E_2 e... E_{n-1}) é a probabilidade de ocorrer E_n, condicionada ao fato de já ter ocorrido E_1 e E_2... E_{n-1}.

$$P\ (A \text{ e } B) = P(A)\ .\ (B/A)$$

Exemplo
Uma urna tem 30 bolas, sendo 10 vermelhas e 20 azuis. Se ocorrer um sorteio de 2 bolas, uma de cada vez e sem reposição, qual será a probabilidade de a primeira ser vermelha e a segunda ser azul?

Seja o espaço amostral S = 30 bolas e consideramos os seguintes eventos:

A: vermelha na primeira retirada e P(A) = 10/30

B: azul na segunda retirada e P(B) = 20/29

Assim:

$$P(A \text{ e } B) - P(A)\ .\ (B/A) = 10/30\ .\ 20/29 = 20/87$$

8.8 Eventos independentes

Dizemos que E_1 e E_2 e... E_{n-1} são eventos independentes quando a probabilidade de ocorrer um deles não depende do fato de os outros terem ou não terem ocorrido.

$$P(E_1 \text{ e } E_2 \text{ e } E_3 \text{ e } \ldots \text{ e } E_{n-1}) = P(E_1)\ .\ P(E_2)\ .\ P(E_3) \ldots P(E_n)$$

Exemplo
Uma urna tem 30 bolas, sendo 10 vermelhas e 20 azuis. Se sortearmos 2 bolas uma de cada vez e repondo a sorteada na urna, qual será a probabilidade de a primeira ser vermelha e a segunda ser azul?

Como os eventos são independentes, a probabilidade de sair vermelha na primeira retirada e azul na segunda retirada é igual ao produto das probabilidades de cada condição, ou seja, P(A e B) = P(A) . P(B).

Ora, a probabilidade de sair vermelha na primeira retirada é 10/30 e a de sair azul na segunda retirada 20/30. Daí, usando a regra do produto temos:

$$10/30 \;.\; 20/30 = 2/9$$

Observe que na segunda retirada foram consideradas todas as bolas, pois houve reposição. Assim, P(B/A) = P(B), porque o fato de sair bola vermelha na primeira retirada não influenciou a segunda retirada, já que ela foi reposta na urna.

8.9 Probabilidade de ocorrer a união de eventos

$$P(E_1 \text{ ou } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ e } E_2)$$

De fato, se existirem elementos comuns a E_1 e E_2, estes eventos estarão computados no cálculo de $P(E_1)$ e $P(E_2)$. Para que sejam considerados uma vez só, subtraímos $P(E_1 \text{ e } E_2)$.

Probabilidade de ocorrer união de eventos mutuamente exclusivos:

$$P(E_1 \text{ ou } E_2 \text{ ou } E_3 \text{ ou } \dots E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$$

Exemplos

1. Se dois dados, azul e branco, forem lançados, qual a probabilidade de sair 5 no azul e 3 no branco?

Considerando os eventos:

A: Tirar 5 no dado azul e P(A) = 1/6

B: Tirar 3 no dado branco e P(B) = 1/6

Sendo *S* o espaço amostras de todos os possíveis resultados, temos:

$$n(S) = 6{,}6 = 36 \text{ possibilidades}$$

Daí temos:

$$P(A \text{ ou } B) = 1/6 + 1/6 = 1/36 = 11/36$$

2. Se retirarmos aleatoriamente uma carta de um baralho com 52 cartas, qual a probabilidade de ser um 8 ou um rei?

Sendo *S* o espaço amostral de todos os resultados possíveis, temos: $n(S) = 52$ cartas.

Considere os eventos:

A: Sair 8 e P(A) = 4/52

B: Sair um rei e P(B) = 4/52

Assim, P(A ou B) = 4/52 + 4/52 - 0 = 8/52 = 2/13

Note que P(A e B) = 0, pois uma carta não pode ser 8 e rei ao mesmo tempo.

Observação
Quando isso ocorre, $A \cap B = \varnothing$, dizemos que os eventos A e B são **mutuamente exclusivos**.

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

Resumo

Nesta aula estudamos os princípios da 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. Nesse caso, com determinado grau de crença, para se obter o desenvolvimento teórico do fenômeno, é necessária 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 essa aula finalizamos uma introdução à Estatística, adquirindo conhecimentos básicos os quais servirão de base para o estudo de outras disciplinas como Gestão da Qualidade, entre outras.

Atividades de aprendizagem

1. Defina probabilidade.

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

3. Diferencie eventos mutuamente exclusivos e não exclusivos.

4. Determine a probabilidade de sair o número 5 em dois lançamentos sucessivos de um dado.

5. Em uma urna há cinco bolas azuis e 9 bolas vermelhas. Retiramos uma bola da urna e, em seguida, sem repor a bola retirada, retiramos uma segunda bola. Determine a probabilidade de:

a) Ambas serem bolas vermelhas.

b) Ambas serem azuis.

c) A primeira bola ser azul e a segunda vermelha..

6. Numa determinada cidade gaúcha, 52% dos habitantes são mulheres e destas, 2,4% são canhotas. Dos homens, 2,5% são canhotos. Calcule a probabilidade de que um indivíduo escolhido ao acaso seja canhoto.

Referências

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

DALMOLIN, 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. 3 v. São Paulo: Ática, 2003.

DAVID, R. A.; Dennis, J. S.; Thomas, A. W. **Estatística Aplicada a Administração e Economia**. 2. ed. São Paulo: Thomson, 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**. 2. ed. São Paulo: Harbra, 1987.

MOORE, David. **A Estatística Básica e sua Prática**. 5. ed. Editora LTC, 2011.

MOROTTIN, Pedro Alberto; BUSSAB, Wilson de Oliveira. **Estatística Básica**. 6. ed. Editora Saraiva, 2010.

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**. 2. ed. Porto Alegre: Artmed, 2006.

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

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

Currículo 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-RS. Mestre em Engenharia de Produção (Qualidade e Produtividade) pela UFSM. Doutor em Engenharia Agrícola pela UFSM.

Atualmente no CTISM/UFSM, é professor do Curso Técnico de Segurança do Trabalho e Tecnólogo em Redes de Computadores, ministra aulas em diversos cursos técnicos, como Mecânica e Eletromecânica, em disciplinas como: Gerenciamento de Riscos, Gestão da Qualidade, Administração e Organização para o Trabalho, Segurança do Trabalho, Higiene do Trabalho, Normalização e Legislação Aplicada, Eletricidade e Magnrtismo, Estatística, 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, ambos no CTISM. Participou de diversos conselhos administrativos (CEPE, CPPD, Colegiado, Comissões e Bancas de Tese de Doutorado) 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.