

Ministério da Educação
Escola Técnica Aberta do Brasil
Universidade Tecnológica Federal do Paraná

Estatística - I

Thelma Pretel Brandão Vecchi
Flávia Aparecida Reitz Cardoso



Cuiabá, 2009

Comissão Editorial Prof^ª Dr^a Maria Lucia Cavalli Neder - UFMT
Prof^ª Dr^a Ana Arlinda de Oliveira - UFMT
Prof^ª Dr^a Lucia Helena Vendrusculo Possari - UFMT
Prof^ª Dr^a Gleyva Maria Simões de Oliveira - UFMT
Prof. Dr. Henrique Oliveira da Silva - UTFPR
Prof. M. Sc. Oreste Preti - UAB/UFMT

Designer Educacional Prof. M. Sc. Oreste Preti - UAB/UFMT

FICHA CATALOGRÁFICA

V397e Vecchi, Thelma Pretel Brandão
Estatística I / Thelma Pretel Brandão Vecchi, Flávia
Aparecida Reitz Cardoso. – Cuiabá : EdUFMT :
Curitiba : UFPR, 2009.
89p.: il.; color.

Inclui bibliografia.
ISBN 978-85-61819-37-8

1. Estatística – Textos didáticos. 2. Estatística descri-
tiva – Textos didáticos. I. Cardoso, Flávia Aparecida
Reitz. II. Título.

CDU – 311(075)

Revisão Germano Aleixo Filho
Capa Marcelo Velasco
Ilustração Marcelo Velasco e Hugo Leonardo Leão Oliveira
Diagramação Elizabeth Kock Carvalho Netto

PROGRAMA e-TEC BRASIL



Amigo(a) estudante:

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

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

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

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

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

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

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

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

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

Brasília, Ministério da Educação – setembro de 2008.

CURSO TÉCNICO EM MEIO AMBIENTE

Caro estudante:

Os avanços tecnológicos na área de informática e comunicação, associados a modelos pedagógicos apoiados no uso de tecnologia, deram origem à modalidade de ensino chamada Ensino a Distância (EaD). A característica desse modelo é a separação física entre aluno e professor. Para suprir a distância, a interação entre o aluno e o professor é mediada tanto por recursos tecnológicos quanto pelo material impresso. Nessa modalidade de ensino, o material impresso, juntamente com recursos de vídeo, videoconferência e um Ambiente Virtual de Aprendizagem são as bases tecnológicas, às quais você terá acesso durante sua formação.

Todos esses recursos são meios de comunicação entre professor e aluno. Cada recurso possui característica própria e necessita de canal específico de comunicação. Para assistir aos vídeos, participar de videoconferência ou realizar as atividades do Ambiente Virtual de Aprendizagem você precisará ter acesso a computadores e a Internet. Porém, tais recursos tecnológicos nem sempre estão disponíveis em tempo integral, por isso, a importância do material impresso, que permitirá a você ter acesso ao conhecimento independente de possuir a sua disposição as tecnologias de informática e de comunicação.

Aliado às atividades presenciais e às atividades a distância, o material impresso irá também apoiá-lo na realização das atividades de estudos, estimulando-o a participar de forma mais ativa no seu processo de ensino-aprendizagem, construindo, progressivamente, conhecimento de maneira interativa. Assim, o professor deixa de ser a única fonte de informação. O distanciamento físico não será impedimento para o processo de cooperação e interação entre você e o professor da sua instituição, responsável pela oferta da disciplina no curso. Esse professor criará oportunidades para que você participe de forma ativa durante seu processo de aprendizagem.

O material foi elaborado visando a formação de Técnicos em Meio Ambiente, segundo os parâmetros do Catálogo Nacional de Cursos Técnicos. O profissional formado deverá ter qualificação para atender à demanda regional em consonância com as tendências tecnológicas. Deverá ser capaz de coletar, armazenar e interpretar informações, dados e documentações ambientais, colaborar na elaboração de laudos, relatórios e estudos ambientais, auxiliar na elaboração, acompanhamento e execução de sistemas de gestão ambiental, atuar na organização de programas de educação ambiental, de conservação e preservação de recursos naturais, de redução, reuso e reciclagem, bem como identificar as intervenções ambientais, analisar

suas conseqüências e operacionalizar a execução de ações para preservação, conservação, otimização, minimização e remediação dos seus efeitos.

Além disso, deve estar ancorado em uma base de relacionamento interpessoal e comunicação oral. Deve também ter pensamento crítico e racional, capacidade para resolver problemas de ordem técnica, capacidade criativa e inovadora, capacidade de gestão e visão estratégica. Essa base o tornará competitivo no mercado de trabalho. Mas, isso somente não é suficiente, você também deve demonstrar: honestidade, responsabilidade, adaptabilidade, capacidade de planejamento, ser ágil e ter capacidade de decisão. Além de ser possuidor de um espírito crítico, uma formação tecnológica generalista e uma cultura geral sólida e consistente.

Foi pensando nessa formação que equipes de professores da rede pública federal de educação elaboraram seu material. Professores que atuam no ensino médio e no ensino superior. Todos profissionais conceituados em suas respectivas áreas de atuação. O objetivo desses profissionais é auxiliar você na sua formação profissional.

Os recursos didáticos pedagógicos e os profissionais envolvidos fazem parte do projeto Escola Técnica Aberta do Brasil, e-TecBrasil. Um projeto que estabelece parceria entre Instituições de Ensino Público Federal, no papel de formadores, e município, ou estado, que disponibilizam os pólos que receberão os cursos oferecidos na modalidade de EaD.

Mas, lembre-se, simplesmente ter acesso aos recursos didáticos e tecnológicos, além de ter a disposição uma equipe especializada de profissionais não é suficiente. É necessário que esse material seja utilizado intensamente, de forma a tornar-se fonte de conhecimento que o auxiliará em todos os momentos de sua formação.

Cientes de que esse também é o seu desejo, a equipe do e-TecBrasil deseja a todos um ótimo processo de aprendizagem.

Atenciosamente,

Equipe de formadores da Escola Técnica Aberta do Brasil
da Universidade Tecnológica Federal do Paraná.

SUMÁRIO

Estatística I - Thelma Pretel Brandão Vecchi e Flávia Aparecida Reitz Cardoso

CONVERSA INICIAL	9
UNIDADE I - INTRODUÇÃO À ESTATÍSTICA	11
UNIDADE II - ESTATÍSTICA DESCRITIVA	21
UNIDADE III - AMOSTRAGEM E DISTRIBUIÇÃO	37
RETOMANDO A CONVERSA INICIAL	55
REFERÊNCIAS	81

CONVERSA INICIAL

Prezado estudante:

Bem-vindo ao mundo mágico da Estatística, com seus números, gráficos, quadros, tabelas. Um mundo capaz de nos abrir os olhos para enxergar e tentar entender o intermediário que talvez exista entre o certo e o errado, ou possibilitar a discussão para entender cada um desses extremos.

Um mundo em que a aleatoriedade dos acontecimentos não permite que se tomem decisões ou se façam afirmações sem antes analisar e avaliar dados numéricos colhidos em experimentos científicos, no caso de ramos das ciências como a medicina, a biologia e a química, ou em pesquisas sobre opiniões de pessoas ou de grupos sobre os mais diversos temas ou assuntos, dentre os quais têm um forte apelo cultural aqueles associados a tópicos ambientais.

Tal crescimento é evidenciado pelo expressivo número de pesquisas e de publicações, bem como pela crescente atuação técnica de profissionais estatísticos e ambientalistas em atividades públicas e privadas que visam à elaboração de diagnósticos e análises de temas geoambientais e socioambientais.

Você já pode perceber, então, a importância da Estatística em seu curso e em sua formação profissional, pois ela tornará possível sua interação com o meio ambiente, ao estudá-lo e ao tirar conclusões a partir de dados coletados e processados.

Diante da importância que a Estatística tem em sua formação, ela corresponde a uma carga expressiva no curso (120 horas). Porém, para facilitar seu estudo, dividimos o conteúdo da disciplina em dois cadernos. Neste primeiro, abordaremos temas relacionados com a Estatística Descritiva (40 horas) e, no próximo, com a Estatística Inferencial (80 horas).

Escrevemos os cadernos com o objetivo de apresentar a Estatística de forma clara e objetiva. Esperamos que, por meio da leitura, você conheça os métodos e as técnicas da Estatística e que isso lhe possibilite desenvolver melhor sua capacidade de tomar decisões, no campo profissional e pessoal, a partir dos cálculos realizados com os dados coletados.

Assim, nesse primeiro caderno, abordaremos os seguintes temas:

- Introdução à Estatística
- Estatística descritiva
- Amostragem e distribuição.

Na Unidade I, você terá contato inicial com o campo da Estatística, sua importância, sua finalidade, sua definição, seus principais elementos e suas variáveis.

Na Unidade II, estudará a Estatística descritiva. Realizará os primeiros cálculos estatísticos, tais como média e desvio-padrão, bastante usados na prática. Você aprenderá a trabalhar com dados estatísticos e distribuição de frequências.

Na Unidade III, você irá estudar as tabelas e os gráficos estatísticos. Estes são instrumentos muito utilizados para apresentar os dados estatísticos de forma clara e objetiva.

Então, vamos dar início à leitura deste caderno com vontade, garra e persistência. Leia com atenção cada tópico, refaça os exemplos verificando se realmente houve aprendizado e, em seguida, faça os exercícios propostos para verificar seu nível de compreensão do tema em estudo.

Bom trabalho e sucesso em seus estudos...



UNIDADE I

INTRODUÇÃO À ESTATÍSTICA

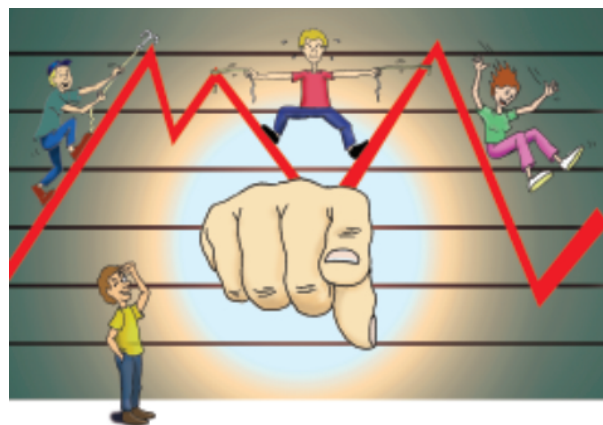
O que você pensou ao ter em suas mãos este caderno de Estatística? Sentiu algum tipo de resistência, imaginando se tratar de conteúdo difícil ou que nada tem a ver com sua formação como técnico em meio ambiente?

Não se preocupe. Isso é comum acontecer, pois, na escola, infelizmente criamos certo “medo” de disciplinas que trabalham com números, como se fossem uma espécie de bicho-papão, não é?

Ou ficou curioso para saber de que trata a Estatística e em que ela pode contribuir para sua formação?

Nesta unidade, você terá contato com as primeiras definições que são necessárias para conhecer, um pouco, o trabalho estatístico, quais elementos são envolvidos neste trabalho, sua finalidade e aplicação.

O ensino da estatística possui importante papel na educação porque representa uma maneira de pensar, não desenvolvida em outros ramos da matemática, capaz de torná-la mais humana pela ligação a problemas do cotidiano. Pois, ela se relaciona com as ciências experimentais, naturais, econômicas e sociais de todos os tipos, como ferramentas de trabalho ligadas à matemática.



Em meados de 1997, os Parâmetros Curriculares Nacionais incluíram, na disciplina de matemática para o ensino fundamental e médio, o conteúdo de estatística.



Mas por quê? Qual a “utilidade” ou finalidade da Estatística na formação do profissional e do cidadão?

De acordo com o Ministério da Educação e do Desporto (MEC), a finalidade da inclusão da Estatística no ensino fundamental, médio e pós-médio é fazer com que o aluno venha a construir procedimentos para coletar, organizar, comunicar e interpretar dados, utilizando tabelas, gráficos e representações que aparecem frequentemente em seu dia a dia,

[...] uma vez que os alunos precisam adquirir entendimento sobre o propósito e a lógica das investigações estatísticas, bem como sobre o processo de investigação. Deve-se possibilitar aos estudantes o entendimento intuitivo e formal das principais idéias matemáticas implícitas em representações estatísticas, procedimentos ou conceitos (Conhecimentos de Matemática, MEC, 2006, p. 79).

Por isso, nesta primeira unidade, você terá contato com os seguintes temas:

- primeiras definições necessárias para conhecer, um pouco, o trabalho estatístico;
- os elementos envolvidos neste trabalho,
- sua finalidade e aplicação.

1 Estatística: conceituação e aplicações

Leia com atenção a reportagem publicada pela revista Veja.

As notícias sobre a Amazônia que chegam aos olhos e ouvidos dos brasileiros são, por natureza, fragmentadas e muitas vezes contraditórias. Ora se dá conta de que a selva tropical brasileira nunca esteve tão protegida. Ora soam os clarins do apocalipse e anuncia-se a morte iminente da maior reserva de água doce, plantas e animais do planeta.



O SENSO COMUM: a Floresta Amazônica é o pulmão do planeta. Se ela desaparecer, o aquecimento global vai se acelerar de forma dramática.

A VERDADE: as pesquisas recentes mostram que o efeito mais visível do desaparecimento da Amazônia seria o desequilíbrio das chuvas no mundo.



O SENSO COMUM: o governo garantiu que o desmatamento está sob controle.

A VERDADE: o desmatamento aumentou seu ritmo em 30% nos últimos meses.

Fonte: Revista Veja on-line, março de 2008



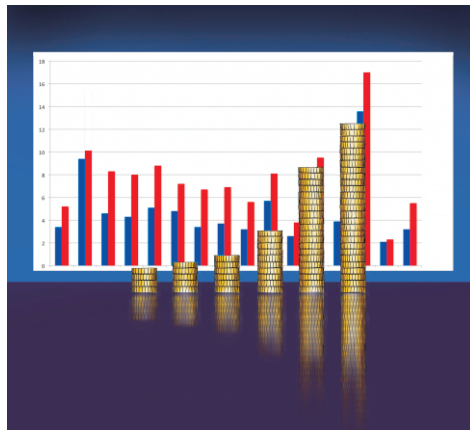
Mas, afinal, o que é a Estatística?



Pegue um dicionário, localize a palavra estatística e, em seguida, transcreva abaixo a explicação encontrada:

Cada dicionário ou livro que trata de Estatística dá sua compreensão desse campo do conhecimento, mas todos eles têm algo em comum. Apontam para o fato de a Estatística ser uma reunião de métodos que coletam, organizam e analisam dados. Pois bem.

A Estatística é um conjunto de métodos para coletar, organizar, resumir, analisar, planejar experimentos, interpretar um conjunto de observações, para que se possa realizar a tomada de decisões.



O objeto de estudo da Estatística são os **fenômenos aleatórios** que representam as variabilidades dos dados dos **fenômenos da natureza**. Caracterizam-se por se repetirem, por não serem os mesmos resultados a cada repetição e por serem feitas várias repetições apresentando acentuada regularidade em seus resultados, o que caracteriza determinado comportamento.

Espero que você tenha uma idéia inicial mais clara sobre o que é a Estatística. Mas, deve estar se perguntando:



Mas, para que serve a Estatística?
Qual sua aplicação e utilidade prática?
Como fazer a leitura de tabelas, gráficos?

Fenômeno aleatório

Para entender, vamos recorrer ao seguinte exemplo: temos o peso e a altura das pessoas, pois se repetem, não podem ser previstos e apresentam certa regularidade, sendo, portanto, objeto de estudo estatístico.

Pesquisadores, engenheiros, cientistas e demais profissionais que estão constantemente em contato com conjuntos de fatos ou dados, utilizam a Estatística, com seus métodos, para resumir e organizar os dados, para tirar conclusões a respeito das informações neles contidas.

Sendo assim, a Estatística tem basicamente duas finalidades:

- descrever os fenômenos e suas características;
- fazer previsões sobre as ocorrências futuras de certo fenômeno em condições semelhantes às aquelas em que ele ocorreu no passado.

Por exemplo, ambientalistas estudam os efeitos do aquecimento global e, a partir de estatísticas, fazem previsões sobre seus efeitos futuros. Leia atentamente a reportagem que encontramos no site www.ibge.gov.br/ibgeteen/mudancas, em 29-10-2008.



O Efeito Estufa e a Vida na Terra Degradação do Meio Ambiente gera mudanças climáticas

A Terra nunca esteve tão quente como nos últimos 25 anos e não temos como negar que o principal responsável por essa mudança climática somos nós mesmos: os seus habitantes.

Nos dias atuais, convivemos com as consequências do aumento de temperatura através de fenômenos como El Niño ou como o desaparecimento do gelo na região ártica do Alasca.

Nos últimos 40 anos, o gelo ártico diminuiu 40%. Nos anos 90, as emissões de poluentes no ar aumentaram 6%. Em 2000, só nos Estados Unidos, as emissões estavam 13% mais altas do que em 1990. Até 2100, a temperatura na Ásia, geralmente acima de 40°C, terá aumentado 7°C.

Como os ambientalistas chegaram a essas conclusões?

Coletando dados ao longo de determinado período, comparando com dados anteriores, com outras pesquisas.

Mas, como fazer isso, como proceder para realizar tal coleta de dados?

Para desenvolver um trabalho estatístico, podemos seguir os seguintes passos:

- a) Definir o problema que será abordado;
- b) Planejar sua solução;
- c) Executar o trabalho planejado a partir da coleta de dados e processamento dos mesmos;
- d) Obter, analisar e divulgar os resultados conseguidos.

À medida que formos avançando nossos conhecimentos no decorrer dos capítulos deste Fascículo, essas fases do trabalho estatístico serão trabalhadas e você terá condições de desenvolvê-las.

Vamos agora verificar se você compreendeu bem esse tópico? Procure, então, realizar as seguintes atividades.



ATIVIDADE

Atividade de aprendizagem – 1 a 2

1) Pesquise, em três diferentes livros, a definição de Estatística e transcreva-a. Identifique aspectos comuns e diferentes entre elas.

2) Dê exemplos de duas situações práticas em que a Estatística pode auxiliar o trabalho do profissional em meio ambiente.

Agora que percebeu a importância da Estatística em sua vida e em sua formação técnica, vamos estudar um aspecto importante desse campo do conhecimento: o sistema de amostragem.

2. População e amostra

Você saberia diferenciar os termos população e amostra no campo da Estatística?

Certamente, deve ter uma idéia inicial, pois são palavras bastante usadas nos meios de comunicação, sobretudo em época de eleições. Mas, vamos juntos aprofundar esse tópico, pois é fundamental no estudo da Estatística.

2.1 População Estatística

Sempre que trabalharmos com algum tipo de pesquisa - seja de indivíduos, de objetos, de animais, enfim de elementos que apresentam a mesma característica -, dizemos que se trata de **população estatística**. A população estatística pode ser dividida em:

- **Finita:** quando possui um número finito de elementos.
Exemplo: número de alunos da classe.
- **Infinita:** quando possui um grande número de elementos.
Exemplo: a população da cidade de São Paulo.

O universo estatístico é infinito, mas, devido à impossibilidade ou à viabilidade econômica, não é viável fazer uma observação que abranja todos os seus elementos. Isso significa que as informações produzidas por uma pesquisa são limitadas a apenas uma parte da população.

Essa parte é denominada de **amostra**, e técnica de coleta é chamada de amostragem. Tem como objetivo garantir que cada elemento da população possa ter a mesma chance de ser escolhido.

2.2 Variáveis

Ao se observar uma população estatística, tomamos como base o estudo de certas características dos elementos dessa população. Por exemplo: tempo de duração de uma lâmpada, período de reprodução de acordo com a espécie.

Essas características são chamadas de variáveis, ou seja, são as medidas controladas ou manipuladas em uma pesquisa. Diferem em muitos aspectos, principalmente no papel que a elas é dado em uma pesquisa e na forma como podem ser medidas.

Há, inicialmente, uma divisão principal para as variáveis estatísticas, que consiste em considerá-las como:

- **variável quantitativa** - são atribuídos valores numéricos.
Por exemplo: altura, peso, massa, número de parentes, etc.
- **variável qualitativa** - não são atribuídos valores numéricos porque se apresentam em formato de texto.
Por exemplo: sexo, esporte, cor, meio de comunicação, etc.

A variável quantitativa pode ser classificada como **discreta** ou **contínua**.

As variáveis **discretas** não podem assumir nenhum valor, dentro de um intervalo de valores de resultados possíveis.

Exemplo

Você está fazendo uma pesquisa em determinada região sobre quantos televisores há em uma casa. O entrevistado não pode responder que tem dois televisores e meio, a resposta não pode assumir todos os valores de um intervalo. Ou ainda podemos dizer que terá que ser uma resposta em número inteiro.

Variáveis discretas

São aquelas resultantes de contagem, constituem um conjunto finito de valores: número de reprovações em matemática, idade em anos completos, etc.

Ao contrário da variável discreta, as variáveis **contínuas** podem assumir qualquer valor dentro de um intervalo de resultado possível.

Exemplo

Se está fazendo uma pesquisa em determinada região onde se quer saber qual é a massa das pessoas, a resposta dos entrevistados pode ser 65,50 Kg, 71,84 Kg.
Massa: medida da quantidade de matéria em um objeto.

Vamos verificar se compreendeu bem esse tópico?



Atividade de aprendizagem - 3

3) Cite duas situações para cada tipo de variável estatística estudada neste tópico: a variável discreta e a variável contínua. Dê preferência para exemplos relacionados com o meio ambiente.

2.3 Amostragem

Você deve ter tido a sorte, indo a uma loja ou a um médico, de ganhar uma “amostra” de alguma coisa, não é? Qual o “sentido” dessa amostra? É uma espécie de “aperitivo” do que representa o produto.

Mas, como posso ter certeza de que aquilo que estou experimentando é representativo daquele produto?

Em época de eleição, são divulgadas pesquisas de opinião dos eleitores em relação aos candidatos. Como posso ter confiança de que aquelas poucas pessoas ouvidas representam o pensamento dos demais eleitores?

É aqui que entram as técnicas estatísticas que buscam dar confiabilidade e credibilidade aos dados coletados em uma “amostra” da população.

A amostragem, portanto, está presente em nossa vida no dia a dia. O simples fato de provar um alimento representa uma *amostragem*. Nas pesquisas científicas, normalmente se utiliza a amostragem para obter valores aproximados ou estimativas (parâmetros desconhecidos) por meio de uma amostra que é uma pequena parte de um todo.

Esse tipo de pesquisa é denominado levantamento por amostragem. Contudo, para que se possa aplicar essa técnica é necessário que se utilize uma metodologia adequada para que os dados a serem obtidos sejam suficientes e possam gerar as informações esperadas.

Vamos apresentar, a seguir, os principais tipos de amostragem e os procedimentos para definir a amostra.

Amostragem probabilística

Essa técnica é especial para recolher amostras, pois garante, quanto possível, o acaso na escolha. Ela exige que cada elemento da população possua determinada probabilidade de ser selecionado. Assim, se N for o tamanho da população, a probabilidade de cada elemento ser selecionado será $\frac{1}{N}$.

Em outras palavras, se você tem doze lápis de cores distintas dentro de um estojo, a probabilidade de você retirar o lápis da cor branca, por exemplo, será $\frac{1}{12}$.

Trata-se do método que garante cientificamente a aplicação das técnicas estatísticas de inferências. Somente com base em amostragens probabilísticas é que se pode realizar dedução sobre a população a partir do conhecimento da amostra.

Amostragem casual ou aleatória simples

É uma técnica simples e frequentemente utilizada. Basta seguir os seguintes procedimentos:

- numerar a população de 1 a n ;
- sortear, a seguir, por meio de um dispositivo aleatório qualquer, x números dessa sequência, os quais corresponderão aos elementos pertencentes à amostra.

Exemplo

Para obter uma amostra de 10%, representativa para a pesquisa da estatura de 90 alunos de uma escola:

- 1º. pegue a lista de alunos e os numere de 1 a 90.
- 2º. escreva os números dos alunos, de 1 a 90, em pedaços iguais de papel, coloque na urna e, depois de misturar bem, retire, um a um, nove números que formarão a amostra.

Quando o número de elementos da amostra é muito grande, esse tipo de sorteio se torna muito trabalhoso. Neste caso, utiliza-se uma tabela de números aleatórios, construída de modo que os algarismos de 0 a 9 são distribuídos ao acaso nas linhas e colunas

Amostragem proporcional estratificada

Quando a população se divide em estratos (subpopulações), convém que o sorteio dos elementos da amostra leve em consideração tais estratos, como sexo, cor, raça, nível de escolaridade, renda, etc.

Assim, podemos obter os elementos da amostra proporcional ao número de elementos desses estratos.

Exemplo

Para obter uma amostra proporcional estratificada de 10%, supondo que, de 90 alunos, 54 sejam meninos e 36 sejam meninas. São, portanto dois estratos (sexo masculino e sexo feminino). Logo, temos:

Sexo	População	10%	Amostra
MASC.	54	5,4	5
FEMIN.	36	3,6	4
Total	90	9,0	9

Foram sorteados mais meninos do que meninas porque há uma proporção maior de sujeitos do sexo masculino na população a ser pesquisada.

Amostragem sistemática

Quando os elementos da população já se acham ordenados, não há necessidade de construir o sistema de referência. É o caso, por exemplo, de lotes numa área rural, de casas numa determinada rua ou quadra de um bairro, de levantamentos ou diagnósticos já disponíveis, etc.

Exemplos

Em uma amostra de 20% dos doentes de gonorréia diagnosticados numa clínica de DST, sorteia-se um valor de 1 a 5. Se o sorteado for o 2, incluem-se na amostra o paciente 2, o 7, o 12 e assim por diante, de cinco em cinco.

Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador.

Você quer realizar uma pesquisa de opinião em determinado bairro. Como selecionar a amostra?

Numa rua com 900 casas, quer-se obter uma amostra formada por 50 casas para uma pesquisa de opinião. Pode-se, neste caso, usar o seguinte procedimento: como $900/50 = 18$, escolhe-se por sorteio casual um número de 1 a 18, o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, supondo que o número sorteado fosse 4, a amostra seria: 4ª casa, 22ª casa, 40ª casa, 58ª casa, 76ª casa, etc.

Amostragem por conglomerados (ou agrupamentos)

Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos. Não obstante isso, pode ser relativamente fácil identificar alguns subgrupos da população. Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode se colhida, e uma contagem completa deve ser feita para o conglomerado sorteado. Agrupamentos típicos são bairros, famílias, organizações, agências, edifícios, etc.

Exemplo

Num levantamento da população de determinada cidade, pode-se dispor do mapa indicando cada bairro e não dispor de uma relação atualizada de seus moradores. Pode-se, então, colher uma amostra dos bairros e fazer a contagem completa de todos os que residem naqueles bairros sorteados.

Amostragem acidental

Trata-se de amostra formada por aqueles elementos que vão aparecendo, cujo número é possível obter até completar a amostra. Geralmente utilizada em pesquisas de opinião, em que os entrevistados são acidentalmente escolhidos.

Exemplo

Pesquisas de opinião em praças públicas, ruas de grandes cidades.

Amostragem por quotas

Um dos métodos de amostragem mais comumente usados em levantamentos de mercado e em prévias eleitorais. Ele abrange três fases:

1ª - classificação da população quanto às propriedades que se sabe, ou presume, serem relevantes para a característica a ser estudada;

2ª - determinação da proporção da população para cada característica, com base na constituição conhecida, presumida ou estimada da população;

3ª - fixação de quotas para cada entrevistador, a quem tocará a responsabilidade de selecionar entrevistados, de modo que a amostra total, observada ou entrevistada, contenha a proporção e cada classe, tal como determinado na 2ª fase.

Exemplo

Caso se queira fazer uma pesquisa sobre o "trabalho das mulheres na atualidade", provavelmente haverá interesse em considerar: a divisão cidade e campo, a habitação, o número de filhos, a idade dos filhos, a renda média, as faixas etárias, etc.

A primeira preocupação é descobrir as proporções (porcentagens) dessas características na população. Imagina-se que haja 47% de homens e 53% de

mulheres na população. Logo, uma amostra de 50 pessoas deverá ter 23 homens e 27 mulheres. Então o pesquisador receberá uma "quota" para entrevistar 27 mulheres. A consideração de várias categorias exigirá uma composição amostral que atenda ao n determinado e às proporções populacionais estipuladas.

Atividade de aprendizagem - 4



- 4) Leia novamente os métodos de amostragem que foram descritos nessa unidade e dê três exemplos para quaisquer três desses métodos, tentando dar preferência a exemplos relacionados com o meio ambiente.



- Ao longo desta unidade, pudemos aprender alguns conceitos de extrema importância no estudo da estatística. Dentre eles, vimos que toda pesquisa pode ser feita com base na observação de determinado grupo que tenha alguma característica comum, denominada população.
- Também estudamos os tipos de amostragem que podem ser empregados para que a técnica de coleta garanta que cada elemento da população possa ter a mesma chance de ser escolhido.
- Finalizando a unidade, estudamos as variáveis como medidas controladoras ou manipuladoras em uma pesquisa, diferindo em muitos aspectos, principalmente no papel que a elas é dado em uma pesquisa e na forma como podem ser medidas. É muito importante que os estudos sejam continuados para que, quando formos interpretar dados obtidos, eles expressem a realidade estudada.

Vamos verificar agora se o conteúdo dessa unidade foi compreendido?



Atividade de aprendizagem - 5 a 8

- 5) Uma agência de turismo cadastrou 3.000 clientes. Para garantir bom atendimento, pesquisou a preferência deles em relação aos preços, ao número de passeios, ao número de acompanhantes e à qualidade com

relação aos serviços prestados pela agência. De modo aleatório, foram consultados 800 clientes. Com base nessas informações, responda:

a) Qual é a quantidade de pessoas da população estatística envolvida na pesquisa?

b) Qual a quantidade de pessoas presentes na amostra?

c) Quais foram as variáveis qualitativas pesquisadas?

d) Quais variáveis quantitativas pesquisadas? Determine as discretas e as contínuas.

6) Obtenha uma amostra representativa proporcional estratificada de 60% dos 50 tigres de uma reserva florestal, em que 20 são fêmeas e 30 são machos.

7) Leia atentamente as frases a seguir e as complete, conforme o que você estudou nesta unidade:

a) As pesquisas são feitas com base em observações de um grupo que apresenta determinada característica comum, chamada de
..... Esse grupo recebe o nome de

b) é uma parte da população que será estudada.

c) Se determinada característica observada, recebe o nome de variável qualitativa e, quando ela, recebe o nome de variável quantitativa.

8) Em janeiro de 2003, o Instituto Brasileiro de Opinião Pública e Estatística (IBOPE) divulgou o resultado de uma pesquisa de opinião sobre a aceitação de alimentos transgênicos pela população brasileira. Uma das perguntas dessa pesquisa foi: Você já ouviu falar em alimentos transgênicos? _____

A tabela a seguir descreve o percentual de resposta a essa pergunta, com o que o Ibope concluiu que a maioria dos brasileiros não sabe o que é um alimento transgênico.

Resposta	Percentual
Já ouviu	37%
Nunca ouviu	61%
Não sabe ou não opinou	2%

Com base nesses dados responda:

a) Como são feitas pesquisas como essa?

b) O IBOPE teria entrevistado toda a população brasileira para chegar a essa conclusão?



UNIDADE II

ESTATÍSTICA DESCRITIVA

Nesta unidade, você terá contato com a Estatística descritiva. Como o próprio nome já diz, a Estatística descritiva se preocupa em descrever os dados de uma pesquisa estatística, pois a análise e a interpretação desses dados são realizadas pela Estatística inferencial, objeto da disciplina Estatística II, que você irá estudar no decorrer deste curso técnico.

Começaremos estudando a distribuição de frequências e, a partir daí, avançaremos para os cálculos das medidas de tendência central (média, moda e mediana) e das medidas de dispersão (variância e desvio-padrão).

O objetivo desta unidade é que você conheça e compreenda as técnicas que a Estatística descritiva utiliza para sintetizar uma série de valores de mesma natureza, permitindo dessa forma que se tenha visão global da variação desses valores, que serão organizados em forma de tabelas.

Leia com atenção o texto e desenvolva os exemplos e exercícios propostos para que você possa compreender os princípios básicos da Estatística descritiva.



1. Distribuição de frequências

Para que as observações de certo fenômeno possam nos fornecer informações de interesse, é necessário que os dados coletados estejam bem organizados. Para isto, utilizamos tabelas e gráficos que tornam a visualização dos dados clara e objetiva.

Vamos supor que foram coletados dados relativos à estatura de trinta alunos da UTFPR:

Tabela 1-Estatura (em cm) de 30 alunos da UTFPR

175	157	159	157	178	160	170	170	175	167
165	178	173	167	165	159	160	157	168	165
175	163	160	175	157	175	170	163	168	170

Esta tabela é denominada **tabela primitiva**, cujos elementos não foram numericamente organizados.

A maneira mais simples de organizar os dados é por meio de certa ordenação (crescente ou decrescente). Esta tabela já ordenada recebe o nome de **rol**.

A partir de agora, para que você compreenda mais facilmente as operações no campo da Estatística, iremos lhe propor exercícios graduais. É importante que

você os faça, pois contribuirão no processo de aprendizagem, tornarão a leitura **mais agradável e facilitarão o estudo do tema seguinte.**



Atividade de aprendizagem - 1 e 2

- 1) Complete a tabela abaixo com os dados da tabela 1, em ordem crescente dos dados numéricos

Tabela 2 - Estatura (em cm) de 30 alunos da UTFPR

157	157	157	157	159	159	160	160	163

O que você percebeu ao completar a tabela? É fácil fazer a leitura?

Você não acha que a variável "estatura" seria observada e estudada muito mais facilmente se dispuséssemos os valores ordenados em uma coluna e colocássemos, numa coluna ao lado, o número de vezes que aparece repetido?

- 2) Procure completar a tabela com os dados sobre estatura

Tabela 3 - Estatura (em cm) de 30 alunos da UTFPR

Estatura (cm) x_i	Freqüências f_i
157	4
159	2
Total	30

Verifique agora, no gabarito ao final do caderno, se organizou de maneira correta os dados.

O que achou? Olhe novamente para a tabela. Não é mais fácil a leitura dos dados coletados por meio dessa organização?

Porém, a tabela ao lado, dependendo da quantidade de dados, pode ser inconveniente, necessitando de muito espaço para sua construção.

Então, pode-se resolver este problema agrupando os valores em vários intervalos, denominados **classes** ou **categorias**.

Um critério utilizado na determinação do número de classes (k) é por meio da fórmula:

$$K = 1 + 3,3.\log n$$

Em que n representa o total de observações, k representa o número de classes, e $\log n$ representa o logaritmo decimal de n .

A amplitude (h) de cada classe, ou seja, o 'tamanho' que cada classe terá, será dada por $h = \frac{a}{k}$ em que a representa a amplitude total das observações, definida como a diferença entre o maior e o menor valor observado, ou seja,

Logaritmo decimal é o logaritmo cuja base é igual a dez.

$$a = \text{maior valor observado} - \text{menor valor observado}$$

No exemplo da estatura, vamos calcular os valores de k e h :

$$\begin{aligned} k &= 1 + 3,32 \cdot \log n & h &= \frac{a}{k} \\ k &= 1 + 3,32 \cdot \log 30 & h &= \frac{178 - 157}{6} \\ k &\cong 5,90 & h &= 3,5 \\ k &\cong 6 & h &= 3,5 \end{aligned}$$

Logo, $k = 6$ significa que teremos 6 classes

$h = 3,5$ representa a amplitude de cada classe, ou seja, a 1ª classe incluirá as estaturas que estiverem no intervalo de 157 cm (incluído) a 160,5 cm (excluído).

$$157 + 3,5 = 160,5$$

A 2ª classe representará as estaturas que estiverem no intervalo de 160,5 cm (incluído) a 164 cm (excluído), e assim por diante, até chegar à 6ª classe.

$$160,5 + 3,5 = 164$$

O segmento vertical (|—) simboliza que o número está incluído no intervalo

A frequência de cada classe abrangerá a frequência de todas as estaturas pertencentes à classe em questão.

Por exemplo, a 1ª classe (representada por 157 |— 160,5) terá frequência igual a 9, que corresponde à soma das frequências das estaturas 157 cm, 159 cm e 160 cm, que são 4, 2 e 3, respectivamente.

Desta forma, como será construída a tabela, no caso da estatura dos alunos da UTFPR? Observe como organizei os dados na tabela a seguir:

Tabela 4 - Estatura (em cm) de 30 alunos da UTFPR

Estatura (cm) x_i	Frequência f_i
157,0 — 160,5	9
160,5 — 164,0	2
164,0 — 167,5	5
167,5 — 171,0	
171,0 — 174,5	1
174,5 — 178,0	7
Total	30

Você observou que

- a 6ª classe tem um símbolo diferente das anteriores (|—|)? O símbolo é necessário, pois o valor 1,78 cm deve estar incluído na última classe;
- o total da frequência continua sendo igual a 30, pois o número de alunos pesquisados é o mesmo.
- o que difere as tabelas 3 e 4 é o número de linhas, sendo este menor na tabela 4.

Além dessa maneira mais simples de organizar as frequências, podemos trabalhar também com as frequências relativas e acumuladas.

A **frequência relativa**, que será denotada por f_r , é definida pelo quociente entre a frequência de classe (f_i) e o número total de observações, ou frequência total (n), e multiplicado por 100 para que o resultado represente o percentual da frequência para cada classe.

Em nosso exemplo, vamos calcular a frequência relativa correspondente à 1ª classe (157,0 — 160,5), cuja frequência $f_i = 9$:

$$f_r = \frac{9}{30} \cdot 100 = 30\%$$

Isto significa que, dos 30 alunos da UTFPR que foram pesquisados, 30% deles possuem estaturas (em cm) que pertencem ao intervalo 157,0 — 160,5.

A **frequência acumulada**, que será denotada por f_a , representa a frequência acumulada das classes imediatamente inferiores à classe em questão.

Pode-se ter também a frequência acumulada relativa, denotada por f_{ar} que será determinada a partir da frequência relativa acumulada das classes imediatamente inferiores à classe em questão.

Queremos desafiá-lo a preencher corretamente a tabela 5. Nós iniciamos o exercício, para ajudá-lo, mas você continua, ok?



Atividade de aprendizagem - 3

3) Calcule a frequência acumulada, a frequência relativa e a frequência relativa acumulada.

Tabela 5 - Estatura (em cm) de 30 alunos da UTFPR

Estatura (cm) x_i	Frequência f_i	Frequência acumulada f_a	Frequência relativa f_r	Frequência relativa acumulada f_{ar}
157,0 — 160,5	9	9	30 %	30 %
160,5 — 164,0	2	11	6,67 %	36,67 %
164,0 — 167,5	5	16	16,67 %	53,34 %
167,5 — 171,0	6	-	-	-
171,0 — 174,5	1	-	-	-
174,5 — 178,0	7	-	-	-
Total	30	-	100 %	-

Conseguiu completar a tabela? Muito bem! Caso tenha tido dificuldades, verifique o gabarito ao final desta unidade e identifique o erro cometido e o porquê.

Observe que o valor correspondente à frequência f_a da última classe é igual ao total da frequência f_i . Note ainda que o total da frequência f_r é igual a 100% e o mesmo ocorre com o valor da frequência f_{ar} da última classe.

Bem, até aqui você pôde aprender como organizar os dados de uma pesquisa estatística utilizando tabelas e distribuição de frequências, mas ainda é preciso saber o que fazer com esses dados para poder conhecer melhor sua pesquisa e poder tirar conclusões sobre ela. Para isso, leia com atenção o tema que segue sobre as medidas de tendência central.

2. Medidas de tendência central

Para representar um conjunto de dados, podemos utilizar um valor único, que representa em termo “médio” todo conjunto. Esse valor tende a se localizar no centro do conjunto de dados, sendo conhecido como **medida de tendência central**.

As medidas de tendência central mais conhecidas e que vamos estudar são: média aritmética, mediana e moda.

Média aritmética



A média aritmética, que denotaremos por $\bar{x}$, de um conjunto de n valores $x_1, x_2, \dots, x_n$ é definida pela soma de todos os valores $x_1, x_2, \dots, x_n$ dividida pelo total de valores n , ou seja:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{ou, ainda,} \quad \bar{x} = \frac{\sum x}{n}$$

Em que o símbolo $\sum$ significa somatório e $\sum_{i=1}^n x_i$ significa somatório de todos os valores de x , que pode ser escrito mais simplificado por $\sum x$.

Resumidamente, a média aritmética consiste em somar todos os valores pertencentes a um conjunto de termos, e dividir o resultado pelo total de elementos.



Vamos supor que as notas de dez alunos na disciplina de Estatística I foram:

6	5	8	7	9	6	4	7	8	4
---	---	---	---	---	---	---	---	---	---

Procure determinar a nota média: _____

Como determinou a nota média da turma? _____

Utilizando a definição anterior, você deve ter realizado o seguinte processo:

$$\bar{x} = \frac{6+5+8+7+9+6+4+7+8+4}{10}$$
$$\bar{x} = 6.4$$

A nota média é 6,4. Bem simples, não acha? Logo, este valor 6,4 representa a média da nota dos dez alunos na disciplina de Estatística I.

Vamos, agora, estudar outro tipo de medida.

Média ponderada



Se, em um conjunto de dados, alguns valores se repetirem, ou seja, se $x_1, x_2, \dots, x_n$ ocorrerem com as frequências $f_1, f_2, \dots, f_n$ respectivamente, a média aritmética será dada pela expressão:

$$\bar{x} = \frac{\sum_{i=1}^n (x_i \cdot f_i)}{\sum_{i=1}^n f_i}$$

ou ainda,

$$\bar{x} = \frac{\sum x \cdot f}{\sum f}$$

Isso significa dizer que será feita a soma dos produtos de cada valor por sua correspondente frequência e dividido pelo somatório das frequências, que representa o total de termos.

Vamos retomar o exemplo anterior das notas dos alunos na disciplina de Estatística I. Temos a seguinte distribuição de frequências:

Tabela 6

Notas x_i	Frequência f_i	$x_i f_i$
4	2	8
5	1	5
6	2	12
7	2	14
8	2	16
9	1	9
Total Σ	10	64

Então, utilizando a definição da média ponderada, temos:

$$\bar{x} = \frac{4.2 + 5.1 + 6.2 + 7.2 + 8.2 + 9.1}{10} = \frac{8 + 5 + 12 + 14 + 16 + 9}{10} = \frac{64}{10} = 6.4$$

Mais complicado? Nem tanto, somente uma operação a mais, ao multiplicar a nota pela frequência (x_i, f_i).

Ficou surpreendido com o resultado? É o mesmo obtido no exemplo anterior. Isso era esperado, pois os dados são os mesmos. A única diferença está em considerar o número de repetições de cada elemento.

Caso os dados sejam distribuídos em classes, os valores x_i , são determinados fazendo a média entre o limite inferior e o superior de cada classe, ou seja,

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

Este valor x_i irá representar os valores pertencentes à classe em questão, já que esta abrange um intervalo de valores.

Deu para entender? Ainda não? Então, me acompanhe neste outro exercício. Vamos considerar a seguinte distribuição de frequências, relativa às notas de quarenta alunos na disciplina de Estatística II:

Tabela 7

Notas x_i	Frequência f_i
0 — 1	2
1 — 2	4
2 — 3	3
3 — 4	5
4 — 5	6
5 — 6	2
6 — 7	7
7 — 8	5
8 — 9	4
9 — 10	2
Σ	40

Como vou determinar a nota média?

Antes de tudo, e sobre isso quero chamar sua atenção, você precisa primeiro encontrar, nesse caso, um valor de x que represente cada intervalo. Este valor vai ser determinado como a média entre os valores inferiores e superiores de cada intervalo.

Para o primeiro intervalo, 0 — 1, temos:

$$x_1 = \frac{0+1}{2} = 0.5$$

Para o segundo, 1 — 2, temos:

$$x_2 = \frac{1+2}{2} = 1.5$$

e assim por diante, até o último intervalo, 9 — 10:

$$x_{10} = \frac{9+10}{2} = 9.5$$



Atividade de aprendizagem - 4

4) Acrescente esses valores na tabela abaixo (3ª coluna) e os multiplique por suas correspondentes frequências (4ª coluna):

Tabela 8

Notas x_i	Frequência f_i	x_i	$x_i f_i$
0 — 1	2	0,5	1
1 — 2	4	1,5	6
2 — 3	3		
3 — 4	5		
4 — 5	6		
5 — 6	2		
6 — 7	7		
7 — 8	5		
8 — 9	4		
9 — 10	2	9,5	19
Σ	40		206

Conseguiu acompanhar nossos raciocínio e nossos passos? Se tiver dificuldade, releia o tópico com mais atenção e, caso necessário, entre em contato com o professor da disciplina e/ou com seu tutor. A tabela anterior está completa no gabarito ao final do caderno. Verifique se você acertou...

Com o procedimento que acabamos de descrever, você obterá a nota média dos quarenta alunos, a partir do quociente:

$$\bar{x} = \frac{206}{40} = 5,15$$

Então, 5,15 representa a nota média dos quarenta alunos na disciplina de Estatística II.

Mediana



A mediana M_e de um conjunto de n valores ordenados $x_1, x_2, x_3, \dots, x_n$, é representada pelo valor central do conjunto, ou seja, pelo elemento que se encontra:

- no caso de n ímpar, na posição

$$\frac{n+1}{2}$$

- no caso de n par, pela média aritmética dos dois termos que se encontram nas posições $\frac{n}{2}$ e $\frac{n}{2} + 1$.

Acompanhe nossa explicação. Para que fique mais claro, vamos recorrer a um exemplo.

Vamos supor que você deva determinar a mediana do conjunto de valores:

5 9 7 19 20 13 17

Como proceder?

a. Em primeiro lugar, coloque os valores em ordem crescente ou decrescente, ou seja:

5 7 9 13 17 19 20

b. Em seguida, observe o número de termos. São 7, um número ímpar. Portanto, terá que utilizar a definição de mediana para um número ímpar de termos:

$$\frac{n+1}{2}$$

c. Aplique agora a fórmula:

$$\frac{n+1}{2} = \frac{7+1}{2} = 4$$

O que significa?

Significa que o 4º termo da sequência representa a mediana. Então, observando a sequência de dados ordenada de forma crescente, percebemos que o 4º termo é o número 13, o que significa dizer que metade dos termos do conjunto de valores é inferior ou igual a 13, e a outra metade é superior ou igual a 13.

E se o número de valores for par? O mesmo procedimento, só muda a definição da mediana. Vamos ver?



Experimente, agora, fazer o exercício.

Experimente determinar a mediana do conjunto de valores:

10 7 3 12 15 8

1º passo: ordene os dados em ordem crescente, ou decrescente:

número de termos é par ou ímpar? _____

São seis não é? Então, é número par.

2º passo: Como o número de termos é igual a 6, um número par, utilize a definição para determinar o valor da mediana, para verificar que elementos se encontram nas posições $\frac{n}{2}$ e $\frac{n}{2}+1$

$$\frac{n}{2} =$$

$$\frac{n}{2} + 1 =$$

Verifique abaixo os cálculos indicados:

$$\frac{n}{2} = \frac{6}{2} = 3 \Rightarrow \text{Elemento da 3ª posição} = 8$$

$$\frac{n}{2} + 1 = \frac{6}{2} + 1 = 3 + 1 = 4 \Rightarrow \text{Elemento da 4ª posição} = 10$$

Logo, a média entre os valores 8 e 10, que é dada por $\frac{8+10}{2}=9$, representa a mediana deste conjunto de termos. Isto significa dizer que metade dos termos deste conjunto é inferior a 9, e a outra metade superior a 9, já que 9 não é um elemento do conjunto de valores.

A mediana é útil principalmente quando o conjunto de dados é muito influenciado pelos extremos, refletindo com mais fidelidade que a média aritmética

Para reforçar sua aprendizagem, pare um pouco a leitura do texto e vamos juntos realizar o seguinte exercício:

Vamos utilizar os dados da tabela anterior em que havíamos organizado a frequência das notas dos 40 estudantes da UTFPR (lembra?) para determinarmos a mediana. Para tal, precisamos, antes de tudo, acrescentar à tabela a coluna da frequência acumulada:

Tabela 9

Notas	Frequência	Frequência acumulada
x_i	f_i	$x_i f_i$
4	2	2
5	1	3
6	2	5
7	2	7
8	2	9
9	1	10
Total Σ	10	

Pronto, está calculada a frequência acumulada. Agora, para determinar a mediana, precisamos primeiramente determinar a posição em que ela se encontra. O número de termos da distribuição é igual a 10, um número par. Logo, temos:

$$P = \frac{n}{2} = \frac{10}{2} = 5$$

$$P = \frac{n}{2} + 1 = \frac{10}{2} + 1 = 5 + 1 = 6$$

Assim, a mediana é a média aritmética entre os termos que se encontram na 5ª e 6ª posições da distribuição. Para a 5ª posição (olhar na coluna da frequência acumulada, onde se encontra o número 5) temos o valor de x igual a 6, e para a 6ª posição (olhar na frequência acumulada, onde se encontra o número 6, 4ª linha da tabela) e o valor correspondente de x é 7.

Portanto, a mediana é dada por: $M_e = \frac{6+7}{2} = 6.5$

Isto significa dizer que 50% dos dados da distribuição são inferiores a 6.5 e os demais 50% superiores. Em outras palavras, 50% das notas dos estudantes são inferiores a 6.5, e as demais 50% são superiores a 6.5.

E se a distribuição de frequência é dada por intervalos de classes, como calcular a mediana?

A mediana pode ser calculada pela expressão:

em que:

$$M_e = L_i + \frac{P - f'_a}{f_{Me}} \cdot h$$

L_i → é o limite inferior da classe mediana (classe que contém a mediana);

$P = \frac{n}{2}$ → é a posição da classe mediana;

f' → é a frequência acumulada da classe anterior à mediana;

h → é a amplitude do intervalo da classe ($h = L_s - L_i$);

f_{Me} → é a frequência da classe mediana.

Será que ficou claro?

Então, sugiro que nos acompanhe no exercício a seguir.

Vamos retomar a outra tabela anterior, referente às notas de quarenta alunos na disciplina de Estatística II:

Notas x_i	Frequência f_i	Frequência Acumulada f_a
0 — 1	2	2
1 — 2	4	6
2 — 3	3	9
3 — 4	5	14
4 — 5	6	20
5 — 6	2	22
6 — 7	7	29
7 — 8	5	34
8 — 9	4	38
9 — 10	2	40
Σ	40	

Veja que é necessário que a tabela apresente as colunas referentes às frequências f_i (frequência de cada classe) e f_a (frequência acumulada).

Agora, para calcular a mediana precisaremos utilizar a fórmula que acabamos de apresentar.

Iniciamos determinando a posição da classe a que pertence a mediana:

$$P = \frac{n}{2} = \frac{40}{2} = 20$$

Logo, observando na coluna da frequência acumulada, a classe que se encontra na vigésima posição é 4 |— 5.

Assim, temos:

- $L_i \rightarrow$ é o limite inferior da classe mediana (classe que contém a mediana): $L_i = 4$
- $P = \frac{n}{2} \rightarrow$ é a posição da classe mediana: $P = 20$

● $f_a \rightarrow$ é a frequência acumulada da classe anterior à mediana: $f_a = 14$

● $h \rightarrow$ é a amplitude do intervalo da classe ($h = L_s - L_i$): $h = 5 - 4 = 1$

● $f_{Me} \rightarrow$ é a frequência da classe mediana: $f_{Me} = 6$

Substituindo esses valores na fórmula dada, temos:

$$M_e = L_i + \frac{P - f_a}{f_{Me}} \cdot h \rightarrow M_e = 4 + \frac{20 - 14}{6} \cdot 1 \rightarrow M_e = 5$$

Logo, a mediana é igual a 5, ou seja, 50% das notas são inferiores ou iguais a 5, e os demais 50% superiores ou iguais a 5.

Moda

A moda M_o de um conjunto de n valores, $x_1, x_2, \dots, x_n$ é representada pelo valor que ocorre o maior número de vezes. Um conjunto de valores pode não apresentar moda (quando nenhum valor se repete), como também a moda poderá não ser única. Vamos ver como isso ocorre?

Por exemplo, no conjunto 3, 5, 7, 7, 7, 8, 10, o elemento que mais se repete é o número 7, com 3 repetições. Logo, a moda deste conjunto é 7.

Já no conjunto 1, 2, 2, 3, 4, 5, 5, 6, os elementos 2 e 5 possuem o mesmo número de repetições, duas cada. Neste exemplo temos um conjunto com duas modas, os elementos 2 e 5.

Bimodal

O nome dado a um conjunto com duas modas

Amodal

O nome dado a um conjunto que não possui modas

Observe, agora, o conjunto 2, 4, 5, 7, 8, 9. O que você percebe? Não temos elementos que se repetem. Aqui não temos moda. Simples, não acha?

Vamos, agora, analisar o que ocorre no caso de os dados serem agrupados em classes de frequências.

A moda pode ser calculada pela expressão:

$$M_o = L_i + \frac{f'}{f + f'} \cdot h$$

- $L_i \rightarrow$ é o limite inferior da classe modal (a classe de maior frequência);
- $f' \rightarrow$ é a frequência de classe imediatamente inferior à classe modal;
- $f'' \rightarrow$ é a frequência de classe imediatamente posterior à classe modal;
- $h \rightarrow$ é a amplitude do intervalo da classe modal.

Voltando à tabela em que há uma distribuição de frequências relativas às notas dos 40 alunos na disciplina de Estatística II, temos:

Notas x_i	Frequência f_i
0 — 1	2
1 — 2	4
2 — 3	3
3 — 4	5
4 — 5	6
5 — 6	2
6 — 7	7
7 — 8	5
8 — 9	4
9 — 10	2
Σ	40

Como encontrar a moda, neste caso?

Como vimos, a moda representa o elemento que obteve o maior número de repetições. Neste caso, o elemento é o intervalo de classe e o que obteve maior frequência foi o intervalo 6 |— 7.

Sendo assim, temos a fórmula: $M_o = L_i + \frac{f'}{f + f'} h$ em que:

- $L_i = 6 \rightarrow$ é o limite inferior da classe modal;
- $f' = 2 \rightarrow$ é a frequência de classe imediatamente inferior à classe modal;
- $f'' = 5 \rightarrow$ é a frequência de classe imediatamente posterior à classe modal;
- $h = 1 \rightarrow$ é a amplitude do intervalo da classe modal ($7 - 6 = 1$).

Então, substituindo os valores na fórmula, temos:

$$M_o = 6 + \frac{5}{2+5} \cdot 1 \rightarrow M_o = 6 + \frac{5}{7} \rightarrow M_o = 6,71$$

Assim, podemos concluir que o valor que mais se repetiu no conjunto de dados foi o número 6,71.

O que você achou dessas medidas de tendência central? Ficou claro como são calculadas e sua finalidade?

Medidas de tendência central

Possibilitam a idéia de todo o conjunto por meio de um valor único, que pode ser a média, a moda ou a mediana.

Talvez, nesse momento, ainda lhe parecem distantes, muito abstratas, sem utilidade para o dia a dia e para sua atuação como técnico em meio ambiente. Mas não é bem assim. Esperamos que, ao final da leitura deste fascículo, você mude de ideia.

Medidas de dispersão

As medidas de tendência central, porém, são insuficientes para descrever, mais detalhadamente, o comportamento de todo o conjunto, ou seja, como os dados se comportam em torno destas medidas.

Exemplo

Considere os tempos de três máquinas semelhantes, para executar certa operação industrial. Foram tomados os tempos (em segundos) de operações para cada máquina, fornecendo os resultados:

Máquina A: 10, 10, 10, 10, 10

Máquina B: 11, 10, 9, 11, 9

Máquina C: 3, 4, 5, 20, 18.

Calculando a média aritmética para cada máquina, temos:

$$\bar{x}_A = \frac{10+10+10+10+10}{5} = 10$$

$$\bar{x}_B = \frac{11+10+9+11+9}{5} = 10$$

$$\bar{x}_C = \frac{3+4+5+20+18}{5} = 10$$

Percebeu que as médias são iguais, ou seja, todas possuem média de 10 segundos? As três máquinas, mesmo trabalhando em ritmo diferente, apresentaram a mesma média.

Interessante, não é? Mas, você nos dirá, certamente: “E daí? Será que o desempenho das três foi o mesmo?”

Você está nos perguntando, em outras palavras, se as três máquinas apresentaram a mesma “variabilidade” em torno da média, isto é, se a dispersão dos tempos de cada máquina em torno da média de 10 segundos é a mesma.

Pois bem, para uma análise quantitativa dessa maior ou menor variação (ou dispersão) do conjunto de valores em torno do valor médio, você precisa estudar as medidas de dispersão.

As que iremos apresentar a seguir são as principais: variância, desvio-padrão e coeficiente de variação.

4.1 Variância



A variância, que será representada pelo símbolo s^2 , de um conjunto de n valores, $x_1, x_2, \dots, x_n$ é a média aritmética dos quadrados dos desvios desses valores em relação à sua média aritmética, ou seja,

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

Vamos determinar a variância para os dados das máquinas A, B e C citadas em momento anterior:

$$s_A^2 = \frac{(10-10)^2 \cdot 5}{5} = 0$$

$$s_B^2 = \frac{(9-10)^2 \cdot 2 + (10-10)^2 \cdot 1 + (11-10)^2 \cdot 2}{5} = 0,8$$

$$s_C^2 = \frac{(3-10)^2 \cdot 1 + (4-10)^2 \cdot 1 + (5-10)^2 \cdot 1 + (18-10)^2 \cdot 1 + (20-10)^2 \cdot 1}{5} = 42,8$$

Analisando os resultados obtidos, percebe-se que:

- → a máquina A não possui variação dos dados em torno da média, pois o valor obtido para a variância foi zero;
- → a máquina B possui uma variação pequena, pois o valor da variância é bem próximo de zero;
- → a máquina C possui uma variação bem maior dos dados em torno da média, pois o valor obtido da variância foi de 42,8.

Preste atenção!

Caso você não considere os dados de toda a população, mas somente de uma amostra, a fórmula anterior é alterada, substituindo o denominador por

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

Tem-se, então:

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

Você notou que a unidade da variância é expressa pelo quadrado da unidade da variável em estudo. Sendo assim podemos contornar este problema definindo-se o que chamamos de desvio-padrão.

Desvio-padrão



O desvio-padrão, que será representado pelo símbolo s , é definido como a raiz quadrada positiva da variância, ou seja:

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

Vamos determinar o desvio-padrão para os dados do exemplo das máquinas A, B e C, utilizando os resultados obtidos para sua respectiva variância:

$$s_A = \sqrt{0} = 0 \qquad s_B = \sqrt{0,8} \cong 0,89 \qquad s_C = \sqrt{42,8} \cong 6,54$$

O que esses dados expressam?

Expressam uma medida na mesma unidade de medida da variável em estudo, pois se extrai a raiz quadrada da variância que representa uma unidade ao quadrado, porque, para determinarmos a variância, utilizamos a soma dos quadrados dos desvios em relação à média.

Agora, observe atentamente os resultados acima. O que você pode concluir?

Em primeiro lugar, que a máquina C é a que possui maior variação dos dados em torno da média, pois seu desvio-padrão é maior do que das demais máquinas.

Em segundo lugar, que a máquina A não apresenta variação, já que seu desvio-padrão é zero.

Muito bem, e o que isso significa?

Significa que a máquina C é mais irregular que todas as outras máquinas, ou seja, seu tempo de produção varia muito. Já a máquina B é mais regular e a máquina A é totalmente regular, pois seu tempo de produção não apresenta variação.

Coeficiente de variação



Outra medida de dispersão é o coeficiente de variação, que é representado por cv .

É uma medida de dispersão, definida por:

$$cv = \frac{s}{\bar{x}}$$

O coeficiente de variação pode ser expresso em porcentagem. Assim, quando se deseja comparar a dispersão de conjuntos com médias bem diferentes, deve-se utilizar o coeficiente de variação, pois este leva em consideração a ordem de grandeza deles.

Exemplo

Suponha que um conjunto de dados tem média $\bar{x}_1 = 30 \text{ cm}$ e desvio-padrão $s_1 = 3 \text{ cm}$, e outro conjunto tem média $\bar{x}_2 = 80 \text{ cm}$ e desvio-padrão $s_2 = 4 \text{ cm}$.

Com base nestes dados, vamos determinar o coeficiente de variação de cada conjunto e fazer uma análise dos resultados obtidos.

- primeiro conjunto (média $\bar{x}_1 = 30 \text{ cm}$ e desvio-padrão $s_1 = 3 \text{ cm}$) → o coeficiente de variação é dado por :

$$cv = \frac{s}{\bar{x}} = \frac{3}{30} = 0,1 \text{ ou } 10\% .$$

- o segundo conjunto de dados (média $\bar{x}_2 = 80 \text{ cm}$ e desvio-padrão $s_2 = 4 \text{ cm}$) → o coeficiente de variação é dado por:

$$cv = \frac{s}{\bar{x}} = \frac{4}{80} = 0,05 \text{ ou } 5\% .$$

Podemos concluir então que, se as médias dos conjuntos são diferentes, o coeficiente de variação fornece um resultado mais coerente, pois mesmo o desvio-padrão sendo menor no conjunto de dados 1, o coeficiente de variação foi maior do que o do conjunto 2. Isto mostra que os dados do conjunto 1 são mais dispersos do que os do conjunto 2, isto é, os dados variam mais em torno da média no conjunto 1 do que no conjunto 2.

Conseguiu acompanhar nosso raciocínio?

Vamos, então, propor-lhe um exercício para que você mesmo verifique se está compreendendo bem as medidas de dispersão.

Vimos que a média das notas dos 40 alunos na disciplina de Estatística II é de **5,15**.

Vamos, agora, completar a 4ª coluna da tabela abaixo, pois será utilizada para o cálculo da variância.

Vamos mostrar o cálculo que será feito para a primeira linha e para a quarta coluna da tabela:

$$(x_i - \bar{x})^2 \cdot f = (0,5 - 5,15)^2 \cdot 2 = 43,245$$



Atividade de Aprendizagem 5

a. Procure fazer este cálculo para as demais linhas da tabela:

Tabela 10

Notas x_i	f_i	x_i	$(x_i - \bar{x})^2 \cdot f$
0 — 1	2	0.5	43,2450
1 — 2	4	1.5	
2 — 3	3	2.5	
3 — 4	5	3.5	
4 — 5	6	4.5	
5 — 6	2	5.5	
6 — 7	7	6.5	
7 — 8	5	7.5	
8 — 9	4	8.5	
9 — 10	2	9.5	
Σ	40		

Calcule, em seguida, a variância, o desvio padrão e o coeficiente de variação.

b. Cálculo da variância:

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

c. Cálculo do desvio padrão:

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

d. Cálculo do coeficiente de variação:

$$cv = \frac{s}{\bar{x}} =$$

Foi difícil?

Verifique, ao final do fascículo, o gabarito. Caso tenha errado, releia esse tópico e procure identificar o procedimento que não ficou claro. É importante que você estude também com colegas, mantenha contato com seu tutor e converse com ele sobre suas dificuldades.



Nesta unidade, apresentamos alguns conceitos modernos utilizados amplamente em situações que envolvem a Estatística.



Nosso mundo é rodeado por números. Por isso, é tão importante que saibamos relacioná-los com os fatos, ainda mais que, se não o fizermos, correremos o risco de não acompanhar as rápidas transformações do dia a dia ou de sermos enganados por resultados manipuladores.



De modo geral, podemos dizer que a análise dos números, ou de um grupo de dados (ou de valores) se caracteriza pelo apontamento das tendências de cada amostra e, conforme foi estudado, isso se concretizou com o estudo das medidas estatísticas e suas representações gráficas.



Com base nisso, na próxima unidade, iremos pôr em prática os conceitos estudados ao longo das duas primeiras unidades, de modo que você tenha uma visualização das relações existentes entre dados numéricos e possíveis acontecimentos futuros, pois isso faz com que se estabeleçam processos de análise e tomadas de decisão alicerçados em métodos estatísticos.

Para que você mesmo possa avaliar sua compreensão dos temas estudados nessa unidade, queremos lhe propor atividades de aprendizagem. As respostas corretas e as explicações, você as encontrará ao final do fascículo. Bom trabalho!



Atividade de aprendizagem - 6 a 13

6) Uma indústria de produtos químicos produz certo resíduo prejudicial ao meio ambiente e este é eliminado em um córrego próximo à empresa. Uma equipe da prefeitura fez uma pesquisa, durante 30 dias consecutivos, para verificar a quantidade deste resíduo despejado na natureza. Com base nessa pesquisa serão tomadas providências para evitar danos ao meio ambiente. Os dados obtidos, em litros, foram:

Tabela 1

21	24	23	25	25	29	31	35	28	20
27	38	35	32	26	28	41	39	22	20
27	30	31	29	28	38	28	21	27	34

Antes de resolver os itens a seguir, volte a ler o tópico que trata da distribuição de frequências, e compare o que se pede com os exemplos resolvidos neste capítulo.

a) Construa o rol;

b) Construa a tabela de distribuição de frequências (f_p, f_a)

c) Determine o número de classes e a amplitude de cada classe para a tabela de distribuição de frequências com intervalos de classes;

d) Construa a tabela de distribuição de frequências (f_p, f_a, f_v), utilizando intervalos de classes.

7) Determine a média aritmética, a mediana e a moda dos dados referentes à tabela do item b do exercício anterior (exercício 6), analisando os resultados obtidos;

--	--	--

8) Determine a média aritmética, a mediana e a moda dos dados referentes à tabela do item d do exercício 6, comparando com os resultados obtidos no exercício 6;

--	--	--

9) Determine a variância, o desvio-padrão e o coeficiente de variação dos dados referentes à tabela do item b do exercício 6, e faça uma análise dos resultados obtidos;

--	--	--

10. Determine a variância, o desvio-padrão e o coeficiente de variação dos dados referentes à tabela do item d do exercício 6, comparando com os resultados obtidos no exercício 4.

--	--	--

11) Uma empresa de produtos agrícolas produz certo tipo de fertilizante que contém, em sua composição, um componente que causa problemas ao meio ambiente. Devido a isto, o limite máximo deste componente para que não haja impacto ambiental é de 15 gramas por embalagem. Mas, devido a pequenas variações no processo de produção, as embalagens variam com relação a esta medida. Uma pesquisa foi realizada para verificar esta variação. Foi feita uma amostra de 50 embalagens deste fertilizante produzidas em um dia, e os seguintes valores do componente em questão foram medidos em gramas (dados fictícios):

Tabela 2

14	16	15	14,5	13	17	15,5	16,3	17,1	13,2
14,5	15,1	13,6	16,2	15,7	13	12,4	16,9	17	15,8
16,9	17,1	15,7	12,8	18	18,3	14,7	13	16	14
15	15,2	16,1	17	16,3	17,2	14,3	13,7	13,8	12,4
12	15,2	17,2	14	15,2	16	17,2	17,3	14,6	15,2

a) Construa o rol;

b) Determine o número de classes e a amplitude de cada classe para a tabela de distribuição de frequência com intervalos de classes;

--

c) Construa a tabela de distribuição de frequência utilizando intervalos de classes;

12) Determine a média aritmética a mediana e a moda dos dados referentes à tabela do item c do exercício anterior (exercício 11), analisando os resultados obtidos;

13) Determine a variância, o desvio-padrão e o coeficiente de variação dos dados referentes à tabela do item c do exercício 11, e faça uma análise dos resultados obtidos.

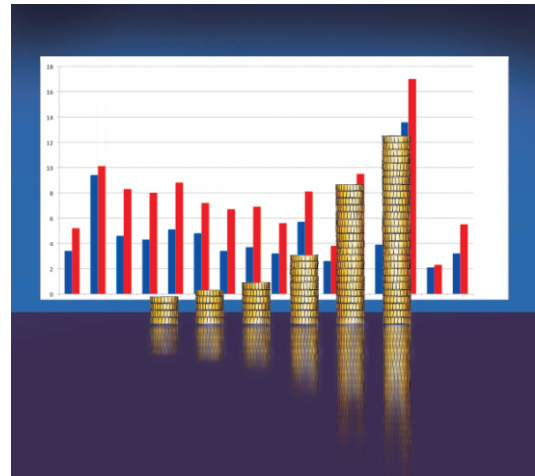


UNIDADE III

CONSTRUÇÃO DE TABELAS E DE GRÁFICOS

Nesta última unidade, iremos estudar formas diferentes de apresentar os dados estatísticos que dão ao leitor informações claras e objetivas. As formas mais utilizadas são as tabelas e os gráficos. Esperamos que, ao final dessa unidade, você saiba analisar e interpretar uma tabela e um gráfico estatístico e perceber o quanto são importantes para apresentação dos dados de uma pesquisa.

Você já deve ter tido contato, inúmeras vezes, com tabelas e gráficos estatísticos, ao ler jornais e revistas, assistir jornais na TV, navegar na internet, em propagandas de marketing de produtos, principalmente em época de propagandas eleitorais.



PARA REFLETIR

Qual tem sido sua postura?
Você aceita as explicações dadas ou procura fazer uma leitura dos dados apresentados?

Temos certeza de que, a partir de agora, com os conhecimentos que essa disciplina está lhe proporcionando, você desenvolverá atitude crítica e de questionamento diante dos dados estatísticos.

Por isso, é importante que nos acompanhe no que iremos apresentar nesta unidade. Iremos tratar das tabelas, dos gráficos e de seus elementos fundamentais, pois, para construir tabelas e gráficos, é necessário compreender seus elementos fundamentais, o que será trabalhado a seguir.

No decorrer desta unidade, serão utilizados exemplos de tabelas e gráficos que consideram situações aplicadas ao tema meio ambiente, já que este é o tema deste curso técnico.

1. Tabelas

No seu dia a dia, você já deve ter utilizado a palavra tabela: acompanhar a tabela dos jogos, verificar a tabela de preços num supermercado, jogar por tabela, etc.

Mas, o que é uma tabela?



Tabela é uma palavra latina que significa “pequena tábua”. Na Antiguidade, era sobre tabuinhas (de terracota, de cera...) que altos funcionários (eram poucas as pessoas que sabiam ler e, menos ainda, as que sabiam fazer contas) lançavam dados, informações sobre produção agrícola, compra, venda, bens, dívidas, empréstimos, lista de preços, censo da população, etc.)



No campo da Estatística,

Tabelas são quadros que contêm um conjunto de dados, com o objetivo de uniformizar e racionalizar os dados, de forma mais simples e fácil para leitura.

1.1 Elementos fundamentais de uma tabela estatística

Uma tabela deve ser feita de modo que forneça o máximo de esclarecimentos, com o mínimo de espaço. Por isso, ela é organizada, contendo pelo menos os seguintes elementos:

Título: colocado na parte superior da tabela, onde deve estar definido o fato observado, com a especificação de local e

época;

Corpo: constituído por linhas e colunas, que fornecem o conteúdo das informações prestadas;

Cabeçalho: a parte da tabela que apresenta a natureza do que contém cada coluna. Apresenta o conteúdo das colunas.

Fonte: aparece logo abaixo da tabela, contendo a informação de onde os dados foram retirados.

Veja o exemplo de uma tabela, retirada do sítio do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA):

Tabela 1- Danos identificados nos acidentes de 2006 e 2007, e número de vezes em que apareceram **Título**

Dano	2006	2007	Cabeçalho
Águas subterrâneas	0	1	Corpo
Atmosfera	5	9	
Danos patrimoniais	3	6	
Fauna	15	31	
Flora	13	10	
Habitat frágil/raro	0	22	
Lago	3	7	
Mar	17	35	
Óbitos/ feridos	13	25	
População afetada/evacuada	7	23	
Praia	9	5	
Rio / córrego	39	58	
Solo	18	26	
Suspensão de abastecimento de água	1	6	
Sem dano	17	9	
Sem informação	0	2	
Total	160	275	
Fonte: IBAMA			Fonte

Vamos identificar cada elemento?

O **título da tabela**: Danos identificados nos acidentes de 2006 e 2007, e número de vezes em que apareceram. O título dá a informação sobre o assunto que será tratado na tabela.

O **cabeçalho**: Dano – 2006 – 2007. Indica que, na primeira coluna estão os tipos de danos, na segunda coluna a frequência que cada dano ocorreu no ano de 2006 e na terceira coluna, a frequência em que se deu no ano de 2007.

O **corpo da tabela** traz as informações de todos os danos verificados e sua respectiva frequência, nos anos de 2006 e 2007. Ao final do corpo da tabela está o total, que indica que, no ano de 2006, ocorreram 160 danos ao todo e, em 2007, 275 danos.

A **fonte** indica que os dados da tabela foram informados pelo IBAMA.

Às vezes, encontramos alguns sinais em tabelas que é importante conhecermos, utilizados em situações específicas:

- (...) -quando o dado existe, mas não o conhecemos, ou não dispomos dele, utiliza-se os três pontinhos.
- (—) -quando o dado não existe, utiliza-se o traço horizontal;
- (?) -quando há dúvida quanto à exatidão de determinado dado, utiliza-se o ponto de interrogação;
- (Z) -quando o dado for rigorosamente zero, utiliza-se a letra Z;
- (0) -quando o valor é muito pequeno para ser expresso pela unidade adotada, utiliza-se o número zero.

1.2 Tipos de tabela

As tabelas se diferenciam dependendo do número de variáveis que elas apresentam. Neste fascículo vamos estudar dois tipos de tabela, a simples e a de dupla entrada.

A **tabela simples**, também chamada de **unidimensional**, é a tabela que apresenta dados ou informações relativas a uma única variável.

Observe a tabela unidimensional retirada do Anuário Estatístico do Estado de São Paulo:

Tabela 2 - Síntese da Caracterização Geográfica do Território - Estado de São Paulo – 2003

Extensão dos principais limites naturais (km)	Medidas
Rio Grande, ao norte	590
Rio Canoas, ao norte e a nordeste	40
Serra da Mantiqueira, a leste	70
Serra do Mar, a leste	50
Oceano Atlântico, a sudeste e ao sul	622
Serra do Taquari, ao sul	40
Rio Pardo, ao sul	80
Rio da Ribeira, ao sul	80
Rio Itapirapuã, ao sul	40
Rio Itararé, a sudoeste	210
Rio Paranapanema, a oeste	440
Rio Paraná, a Oeste e a noroeste	410

Fonte: Instituto Geográfico e Cartográfico – IGC

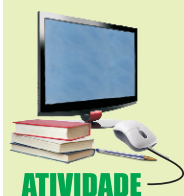
A **tabela de dupla entrada**, também chamada de **bidimensional**, apresenta dados ou informações relativas a mais de uma variável.

Observe esta outra tabela bidimensional:

Tabela 3 - Concentrações de amônia Cubatão
(Vila Parisi e Vale do Mogi) em 1998

Mês	Vila Parisi Media $\frac{\mu g}{m^3}$	Vale do Mogi média $\frac{\mu g}{m^3}$
Julho	-	-
Agosto	32	26
Setembro	33	30
Outubro	32	-

Fonte: CETESB



Agora que você já tomou conhecimento dos elementos de uma tabela e dos tipos de tabelas existentes, procure, em jornais, revistas, internet ou propagandas, diferentes exemplos de tabelas, e procure identificar o título, o cabeçalho, o corpo, a fonte, e também o tipo de tabela (unidimensional ou bidimensional) de cada tabela que você pesquisou.

Se você encontrar dificuldades nessa atividade, procure seu tutor para tirar suas dúvidas.

Outra maneira de apresentar os dados de uma pesquisa estatística de forma clara, de fácil visualização e objetiva, é por meio de gráficos estatísticos. Estes são muito utilizados, pois chamam a atenção devido à sua forma e ao colorido.

2. Gráficos



Os gráficos informam as mesmas ideias das tabelas, porém causam impressão e compreensão mais rápida, mais viva, pois eliminam os detalhes desnecessários, visualizando somente as características mais importantes dos dados.

Os gráficos apresentam as seguintes características: simplicidade, clareza e objetividade. Os gráficos também têm regras para sua construção, assim como as tabelas, que já estudamos. Veja as regras necessárias para a construção de gráficos:

- ➡ Título e fonte, assim como as tabelas;
- ➡ Legendas, quando necessário, todas iguais e colocadas abaixo ou ao lado do gráfico;

- ➡ A leitura é feita sempre da esquerda para a direita.
- ➡ São incluídas somente no gráfico as coordenadas indispensáveis à leitura.
- ➡ Escala horizontal cresce para a direita e a escala vertical para cima.
- ➡ As linhas do gráfico devem ser individualizadas, usando cores diferentes ou espessuras diferentes.

2.1 Representação gráfica dos dados estatísticos

O gráfico estatístico é uma das formas mais utilizadas para a apresentação dos dados estatísticos. Tem como objetivo produzir, em quem o analisa, uma informação direta e objetiva do fenômeno em estudo. Por ser colorido, utilizar formas geométricas, algumas vezes figuras, chama a atenção do leitor para a pesquisa estatística, devido às suas características visuais.

Existem vários tipos de gráficos estatísticos. Vamos estudar aqueles que são mais utilizados: gráficos de colunas, gráfico em barra, gráfico em setores, gráfico de linhas e histograma.

Gráfico de colunas

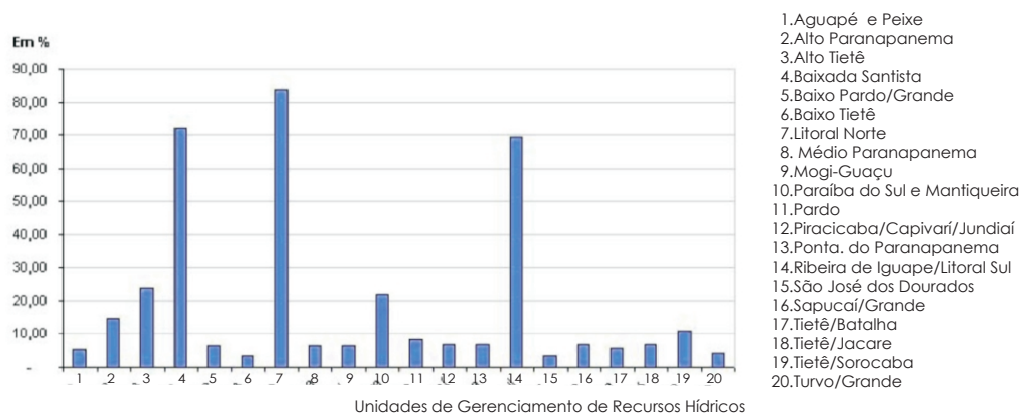
É a representação de uma série estatística por meio de retângulos, dispostos verticalmente. Os retângulos possuem a mesma base, e sua altura é proporcional aos respectivos dados.

É o mais usado para :

- séries temporais (o elemento variável é um fator cronológico, ou seja, depende do tempo);
- séries específicas (o elemento variável é um fato); ou
- séries geográficas (o elemento variável é um fator geográfico).

Observe dois gráficos de colunas retirados do Anuário Estatístico do Estado de São Paulo:

Proporção de área coberta por vegetação natural, segundo unidades de gerenciamento de recursos hídricos –Estado de São Paulo - 2001

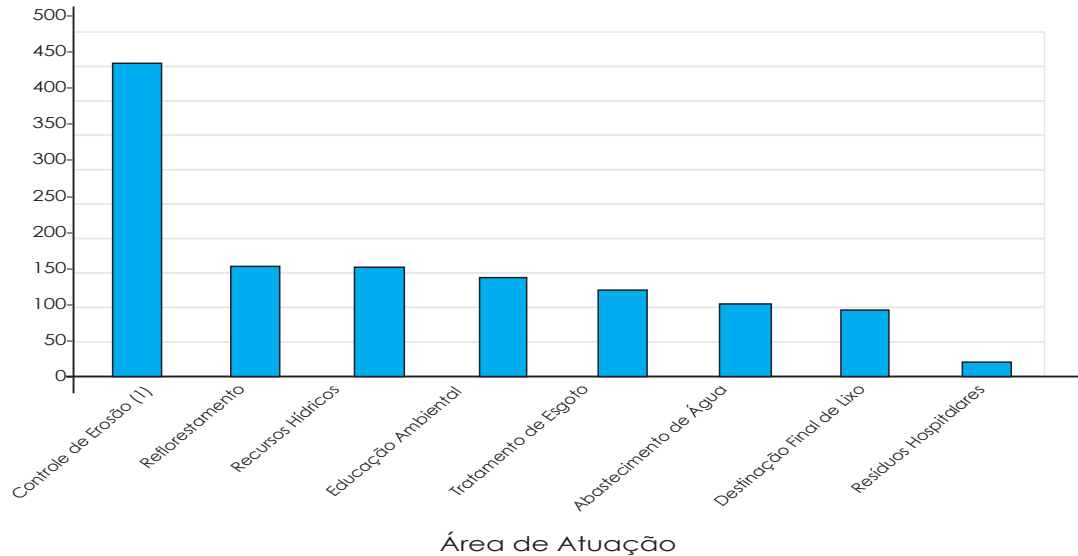


Fonte: Secretaria de Estado do Meio Ambiente, Relatório de Qualidade Ambiental do Estado de São Paulo 2004-2005, Instituto Florestal- IF, Inventário Florestal da Vegetação Natural do Estado de São Paulo.

Veja que o gráfico apresenta título, para que fique claro o assunto que o gráfico contempla, e a fonte, para que se tenha conhecimento da origem dos dados.

Segue mais um exemplo:

Número de municípios participantes de consórcios intermunicipais em Meio Ambiente, segundo áreas de atuação, Estado de São Paulo, 2003.



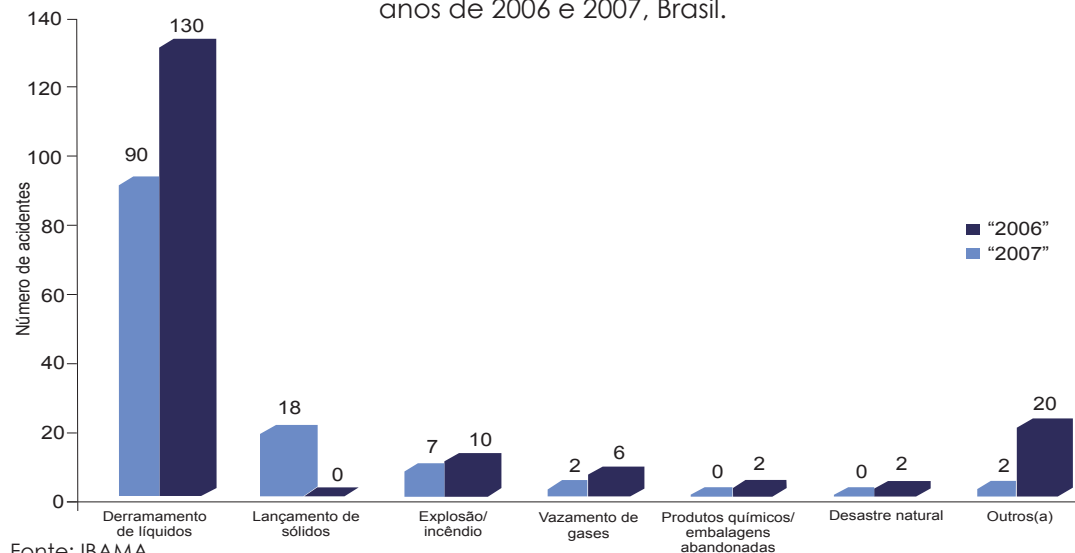
Fonte Companhia de Desenvolvimento Agrícola do Estado de São Paulo: Codasp; Fundação Seade. Pesquisa Municipal Unificada - PMU

(1) Estão incluídos 76 consórcio intermunicipais, envolvendo 427 municípios, constituídos com finalidade de integrar o Programa Pró- Estrada, desenvolvido pela Companhia de Desenvolvimento Agrícola do Estado de São Paulo - Codasp, cujo objetivo principal é a conservação de estradas rurais.

Nota: As áreas de atuação dizem respeito à Classificação temática das atividades desenvolvidas pelos consórcios intermunicipais com vistas à solução de problemas ambientais, relacionados tanto à infra-estrutura como à ação institucional.

Outro tipo de gráfico de colunas muito utilizado é aquele que faz uma comparação entre duas ou mais variáveis. A seguir alguns exemplos:

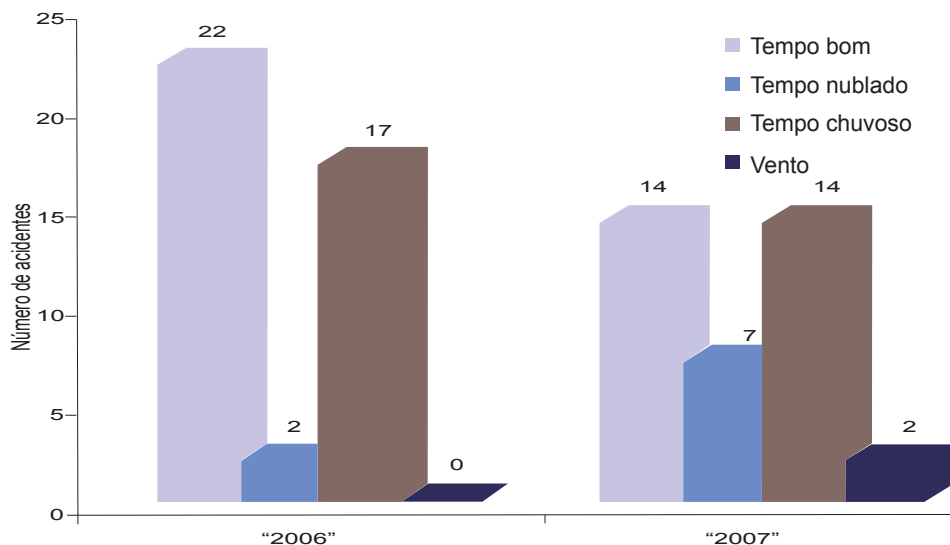
Número de acidentes ambientais por tipo de evento registrado nos anos de 2006 e 2007, Brasil.



Fonte: IBAMA

Outro exemplo:

Número de acidentes nas 4 condições meteorológicas disponibilizadas no formulário, nos anos de 2006 e 2007, Brasil.



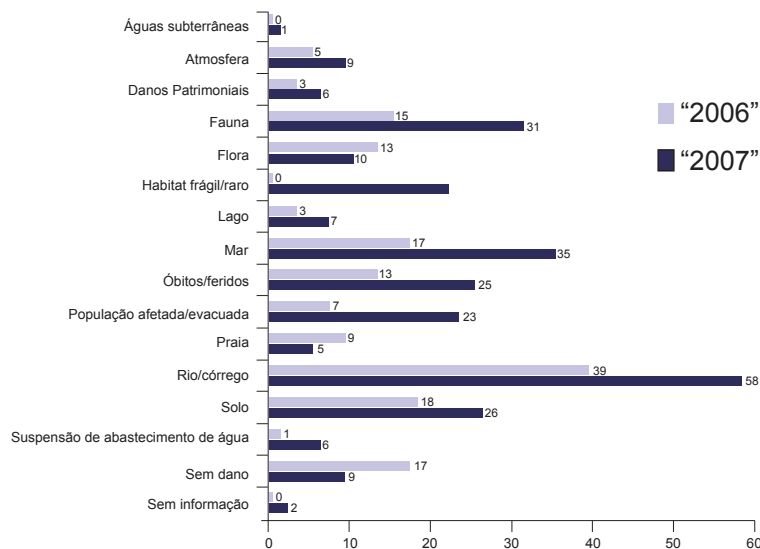
Fonte: IBAMA

Gráfico em barra

É a representação de uma série estatística por meio de retângulos dispostos horizontalmente. Os retângulos possuem a mesma altura e o seu comprimento é proporcional aos respectivos dados.

Segue um exemplo de um gráfico de barras:

Número de vezes em que cada dano foi observado nos acidentes ambientais de 2006 e 2007, Brasil.



Fonte: IBAMA

Você pode observar que este é um gráfico de comparação entre o número de vezes que cada dano foi observado nos acidentes ambientais ocorridos nos anos de 2006 e 2007.

Com base em um gráfico de comparação, é possível tirar importantes conclusões sobre a variável em questão, de forma rápida e clara.

Segue mais um exemplo de gráfico de barras:.

Número de vezes em que cada instituição atuou no atendimento dos acidentes registrados em 2006 e 2007, Brasil.

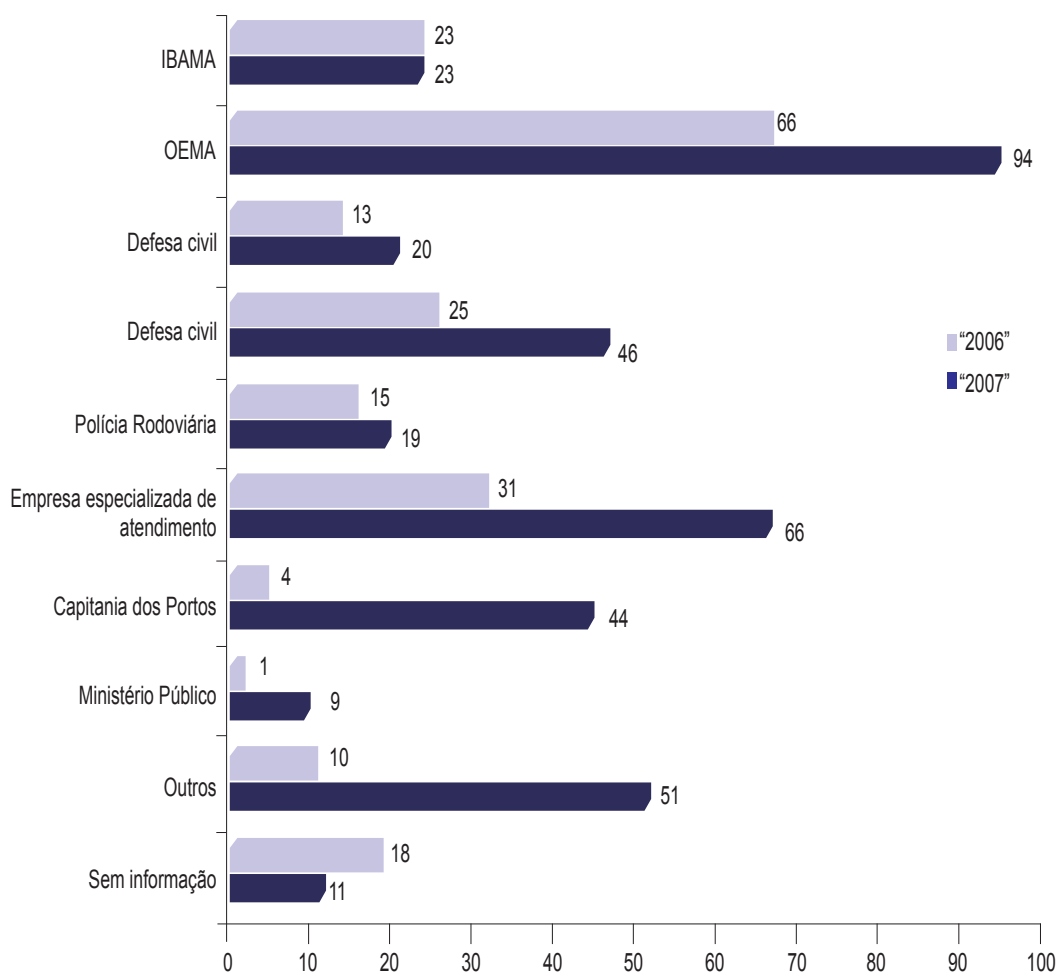
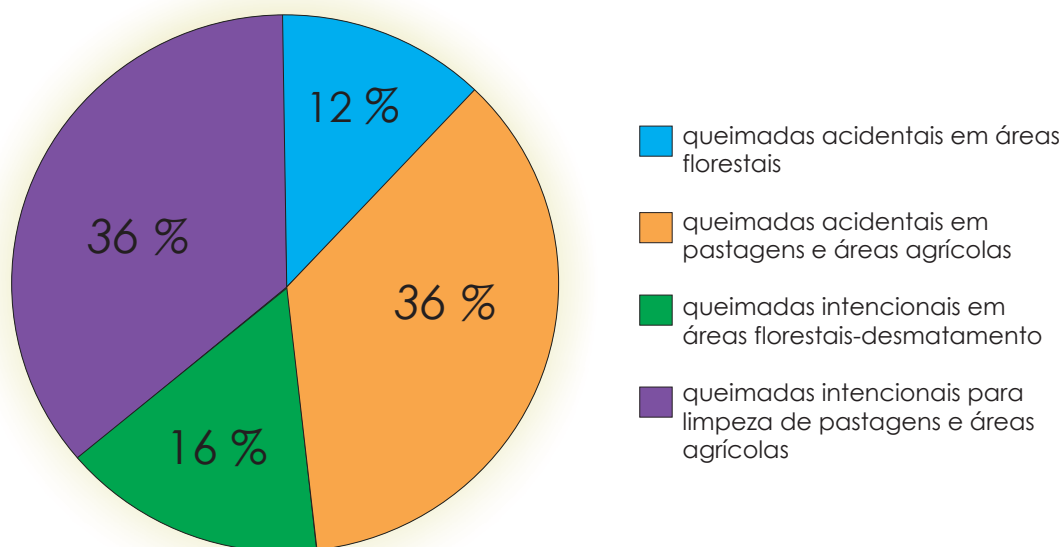


Gráfico de setores

É designado por meio de um círculo, em que cada classe é representada por uma fatia do círculo, sendo dividido proporcionalmente ao tamanho da amostra.

É utilizado quando se deseja mostrar as partes de um todo, quando se deseja comparar proporções.

Queimadas ocorridas no Brasil - 1995

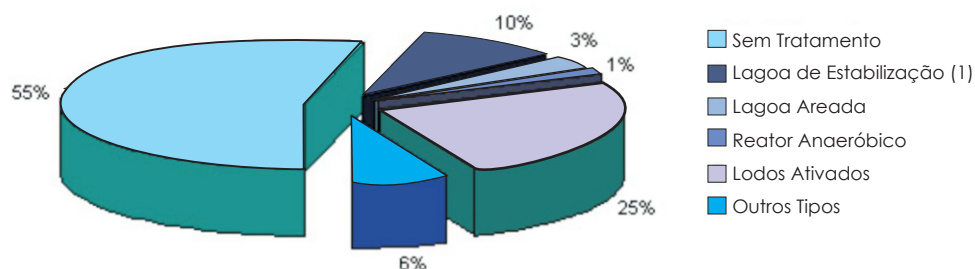


Fonte: IBAMA

Este gráfico proporciona rápida comparação entre os tipos de queimadas ocorridas no ano de 1995 e, ainda, apresenta a respectiva percentagem, o que garante ótima relação com o todo.

Segue mais um exemplo de gráfico de setores:

Volume de esgoto coletado,
por tipo de tratamento dispensado, Estado de São Paulo, 2003



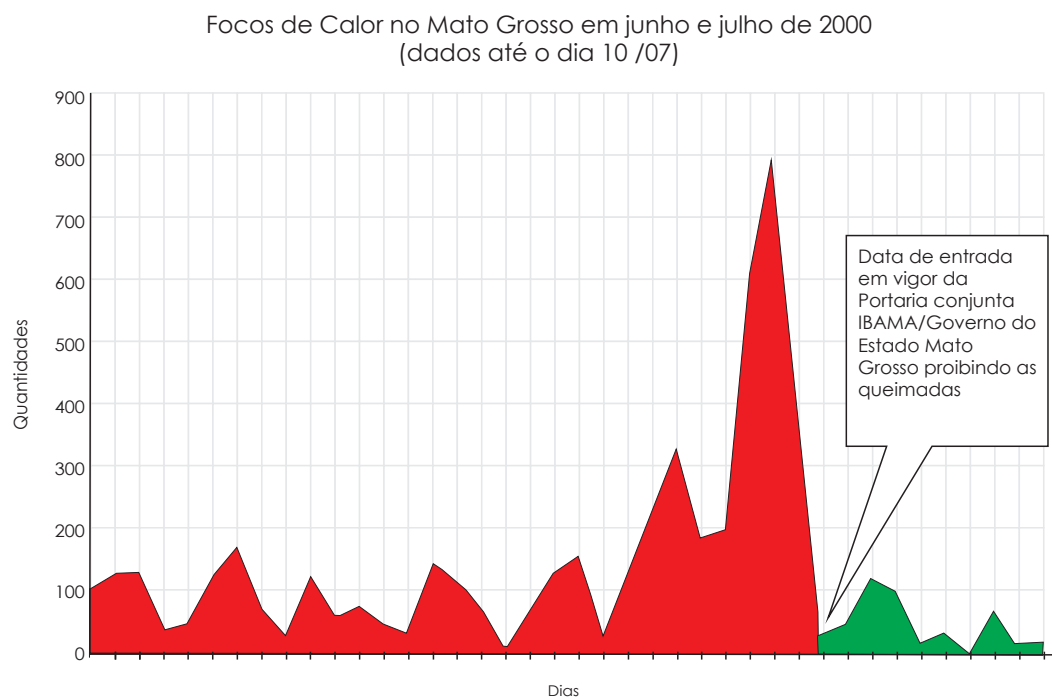
Fonte: Companhia de Saneamento Básico do Estado de São Paulo – Sabesp; Fundação Seade.
Pesquisa Municipal Unificada – PMU.
(1) Inclui lagoa australiana e facultativa.
(2) Incluem, nos municípios litorâneos, emissário submarino.

Observe que este gráfico apresentou, logo após a fonte, duas observações referentes à legenda. São as notas que devem aparecer no rodapé do gráfico, quando forem necessárias.

Gráfico de linhas

É mais utilizado para a representação de séries temporais.

A seguir, um gráfico de linhas, retirado do sítio do IBAMA, que mostra o número de focos de calor em Mato Grosso no período de 1º de junho de 2000 a 10 de julho do mesmo ano.



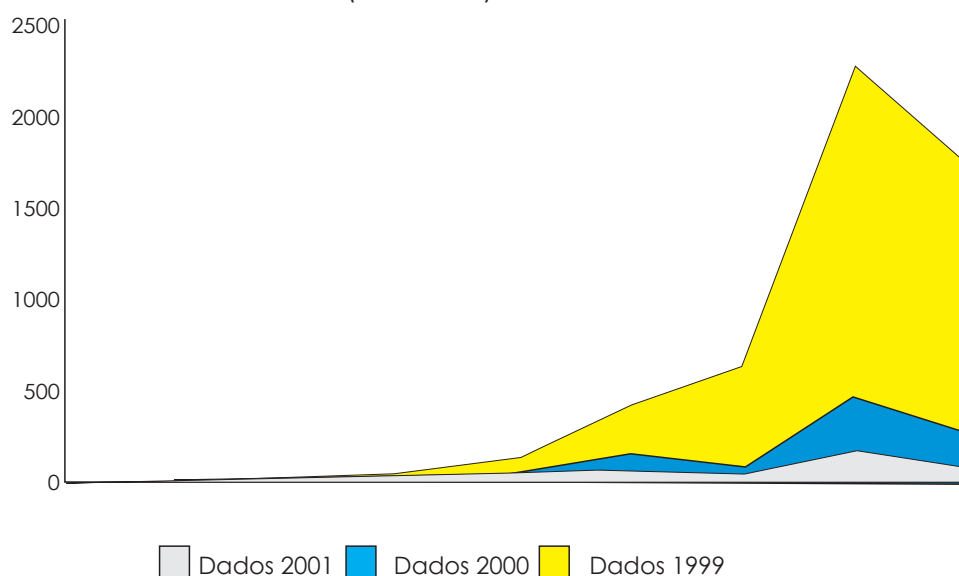
Fonte: IBAMA

Ao analisar este gráfico, podem-se perceber os intervalos de tempo em que houve queda do número de focos de calor, ou, ainda, quando houve aumento do número de focos. Percebe-se também que houve uma queda brusca exatamente na data de entrada em vigor da portaria conjunta do IBAMA com o Governo de Estado do Mato Grosso que proíbe as queimadas.

Com certeza, este era o objetivo maior deste gráfico: informar ao leitor que a Portaria conjunta entre o IBAMA e o Estado de Mato Grosso, que proíbe queimadas, realmente foi eficaz para o decréscimo do número de queimadas no Estado.

Mais um exemplo de gráfico comparativo de linhas:

Número de Alertas Verdes na Amazônia
(1999-2001)



Fonte: Ministério do Meio Ambiente

	Jan.	Fev.	Mar	Abril	Mai	Junho	Julho	Agosto	set.
1999	0	0	7	17	83	263	570	1.837	1.482
2000	5	12	17	1	17	86	29	294	193
2001	5	11	6	12	20	50	33	173	73

Os alertas verdes fazem parte da metodologia usada no sistema de monitoramento de fogo do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis – IBAMA, e são emitidos por este instituto.



O IBAMA, em conjunto com Instituto de Pesquisas Espaciais - INPE, utiliza imagens de alguns satélites para a geração de pares de coordenadas que possam indicar possível incêndio florestal.

Analisando o gráfico, o que você pode concluir sobre o comparativo que foi feito entre o número de alertas verdes na Amazônia, nos anos 1999, 2000 e 2001?

Dentre outras conclusões que você pode tirar com base nesta análise, a que está bem evidente é que o maior número de alertas verdes ocorreu sempre no mês de agosto. Por que será que isto aconteceu? Se você ficar curioso, pesquise e procure responder a esta questão.

Histograma

É destinado à representação de distribuição de frequência. Os intervalos de classes são representados por retângulos, um para cada classe, vertical e unidos (sem espaço entre eles). As alturas são determinadas pela frequência de cada classe.

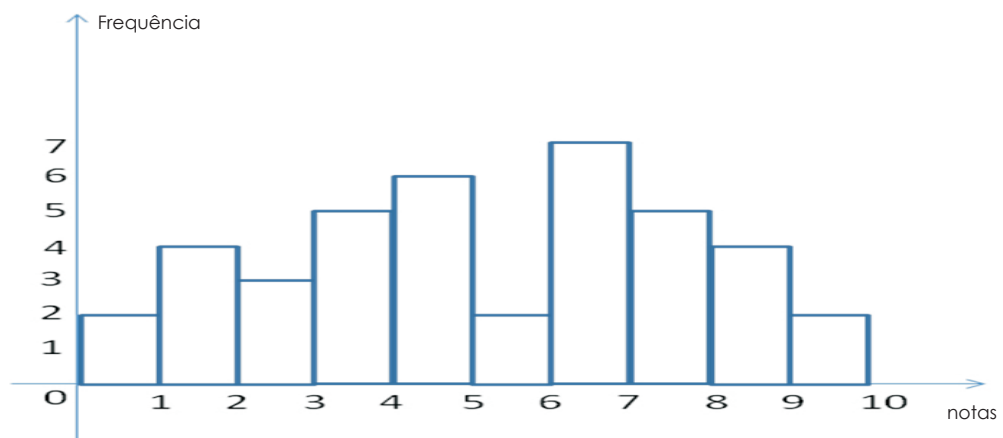
Vamos considerar como exemplo os dados do exemplo 4 da unidade II deste caderno:

Tabela 7 - Notas de 40 alunos na disciplina de Estatística II – UTFPR, 2007

Notas x_i	Frequência f_i
0 —— 1	2
1 —— 2	4
2 —— 3	3
3 —— 4	5
4 —— 5	6
5 —— 6	2
6 —— 7	7
7 —— 8	5
8 —— 9	4
9 —— 10	2
Σ	40

Esses dados podem ser apresentados de outra maneira, por meio do histograma:

Estatística II – UTFPR, 2007



Comparando as duas maneiras de apresentar os mesmos dados, qual você acha que oferece leitura mais fácil e imediata?

Você percebeu como os gráficos são ótima ferramenta utilizada pela estatística para mostrar os dados de uma pesquisa? Então, a partir de agora, procure prestar mais atenção quando tiver contato com estes gráficos, analisando-os e tirando conclusões sobre o assunto que este desvenda.



Nesta unidade, vimos a importância das tabelas e dos gráficos que nos dão uma ótima apresentação dos dados estatísticos.

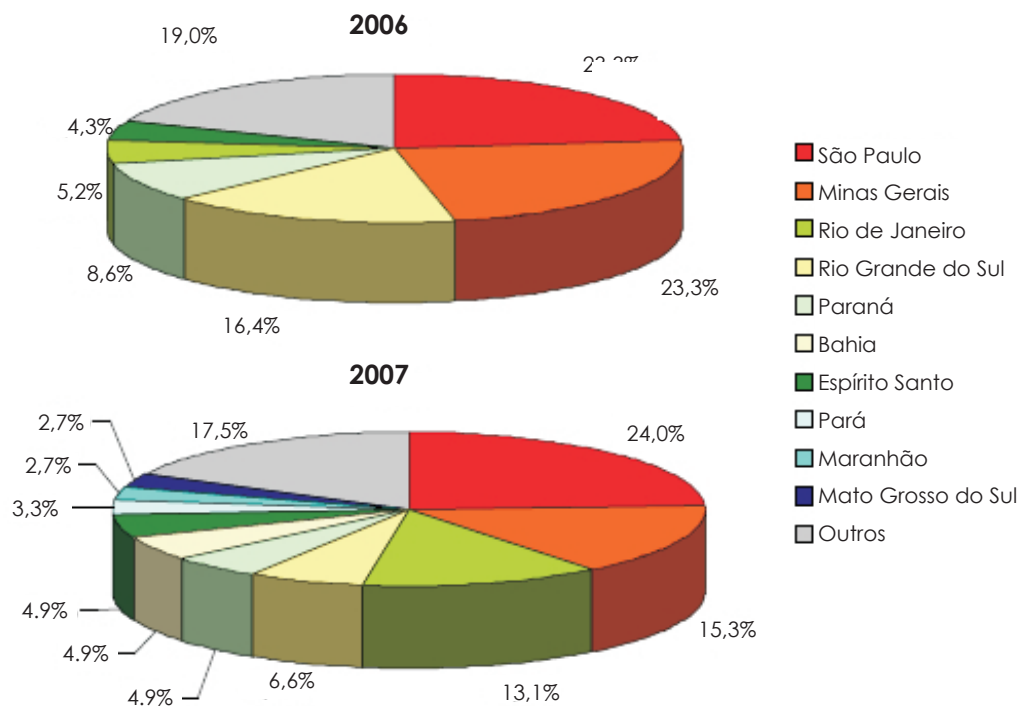
De agora em diante, procure prestar mais atenção quando tiver contato com tabelas e gráficos em revistas, jornais escritos e falados, livros, internet, propagandas. Faça uma análise dos dados, verifique se este está sendo um bom recurso para sua apresentação e se você tem uma boa compreensão dos fatos com base neste recurso.



Propomos, a seguir, alguns exercícios para que você possa avaliar sua compreensão dos temas estudados nesta unidade. Como as questões são abertas, você não encontrará as respostas ao final da unidade, pois é sua opinião que estará sendo dada. Portanto, procure discutir com colegas, e até mesmo com seu tutor, as respostas que você obteve. Com certeza, irá enriquecer seu conhecimento esta troca de experiências. Bom trabalho!

- 1) Pesquise, em jornais e revistas, gráficos estatísticos referentes a dados relativos ao tema meio ambiente e faça uma análise destes gráficos:
- 2) Pesquise, em jornais e revistas, tabelas estatísticas referentes a dados relativos ao tema meio ambiente e faça a construção do correspondente gráfico utilizando a ferramenta de construção de gráficos do Microsoft Excel:
- 3) Construa o histograma referente aos dados das tabelas dos exercícios 1 e 6 da unidade 2, tabelas de distribuição de frequência com intervalos de classes (reveja o exemplo de histograma dado no texto antes de fazer esta atividade):
- 4) Analise os dados que constam na tabela e nos gráficos abaixo, retirados do sítio do IBAMA, e faça um comentário sobre o que você pode concluir:

Estados	2006	2007
Acre	0	2
Alagoas	1	1
Amapá	0	0
Amazônia	2	4
Bahia	1	9
Ceará	1	1
Distrito Federal	1	1
Espírito Santo	5	9
Goiais	1	3
Maranhão	3	5
Mato Grosso do Sul	0	5
Minas Gerais	27	28
Pará	1	6
Paraíba	2	2
Paraná	10	9
Rio de Janeiro	6	24
Rio Grande do Norte	2	2
Rio Grande do Sul	19	12
Rondônia	0	0
Roraima	0	0
Santa Catarina	1	4
São Paulo	27	44
Sergipe	0	4
Tocantins	1	1
TOTAL	116	183



RETOMANDO A CONVERSA INICIAL

Caro estudante:

Durante o século XX, os métodos estatísticos foram desenvolvidos como uma mistura de ciência, tecnologia e lógica para a solução e investigação de problemas em várias áreas do conhecimento humano. Ela foi reconhecida como um campo da ciência neste período, mas sua história teve início bem anterior a 1900.



Como você pode perceber, a Estatística é considerada uma tecnologia que permite avaliar e estudar as incertezas e seus efeitos no planejamento e na interpretação de experiências e de observações de fenômenos da natureza e da sociedade. E, como já vimos, ela não pode ser considerada como uma caixa-preta, nem bola de cristal, nem mágica. Nem mesmo um conjunto de técnicas úteis para algumas áreas isoladas ou restritas da ciência. em análise de decisões, em questões judiciais, previsões e em muitas outras áreas.

REFERÊNCIAS BIBLIOGRÁFICAS

- BIEMBENGUT, M. S. *Modelagem matemática & implicações no ensino-aprendizagem de matemática*. Blumenau: FURB, 1998.
- CARVALHO FILHO, Sérgio de. *Estatística básica para concurso*. Rio de Janeiro: Impetus, 2004.
- COSTA NETO, P. L. *Estatística*. 2 ed. São Paulo: Edgard Blucher, 2002.
- MARQUES, Jair Mendes e MARQUES, Marcos Augusto Mendes. *Estatística básica para os cursos de engenharia*. Domínio do saber, 2005.
- MARTINS e DONAIRE. *Princípios de estatística*. 4 ed. São Paulo: Atlas, 1990.
- MILONE, Giuseppe. *Estatística geral e aplicada*. São Paulo: Thomson Pionera, 2003.
- Ministério da Educação. *Parâmetros curriculares nacionais – ensino médio*. Secretaria de Educação Média e Tecnológica. Brasília: MEC, 1999.
- MORETTIN, Luiz Gonzaga. *Estatística básica: inferência*. 4 ed. São Paulo: Macron Books, 2000.
- SILVA, J.D. e FERNANDES, V.S. *Matemática*. São Paulo, IBEP, 1998.
- SOUZA, M.H. e SPINELLI, W. *Matemática 2º grau*. Volume 1. São Paulo: Scipione, 1996.
- SPIEGEL, M. *Estatística: Chaum*. 5 ed. São Paulo: Macgraw Hill, 1993.
- TOLEDO, Geraldo Luciano; OVALLE, Ivo Izidoro. *Estatística básica*. São Paulo: Atlas, 1985.

GABARITO

UNIDADE I

5)

- a) A população estatística é representada pelos 3.000 clientes cadastrados na agência.
- b) A amostra pesquisada é representada pelos 800 clientes escolhidos aleatoriamente.
- c) Como as variáveis qualitativas não são expressas por números, a qualidade com relação aos serviços prestados é uma variável qualitativa.
- d) As variáveis quantitativas discretas relacionam contagem, ou seja, o número de acompanhantes e o número de passeios. Já as quantitativas contínuas, relacionam o tempo de duração e o preço da viagem.

6)

Estratos	População	60% dos tigres	Amostra
fêmeas	20	60% de 20 = 12	12
machos	30	60% de 30 = 18	18
total	50	60% de 50 = 30	30

Assim podemos concluir que a amostra estratificada é formada por 30 tigres, sendo 12 fêmeas e 18 machos.

7) a) variável e população estatística

b) amostra

c) não é mensurável e é mensurável

8) a) Primeiramente precisa saber qual é a população estatística que se vai realizar a pesquisa para depois, então, estudar os comportamentos da amostra.

b) Não. Da população estatística, no caso a população brasileira, é retirada uma amostra e a pesquisa é feita na amostra.

UNIDADE II

1)

Tabela 2

Estaturas (em cm) de 30 alunos da UTFPR

157	157	157	157	159	159	160	160	160	163
163	165	165	165	167	167	168	168	170	170
170	170	173	175	175	175	175	175	178	178

Estaturas(cm) x_i	Frequência f_i
157	4
159	2
160	3
163	2
165	3
167	2
168	2
170	4
173	1
175	5
178	2
Total	30

3) Tabela 5
Estaturas (em cm) de 30 alunos da UTFPR

Estaturas(cm) x_i	Frequência f_i	Frequência acumulada f_a	Frequência relativa f_r	Frequência relativa acumulada f_{ar}
157,0 — 160,5	9	9	30 %	30 %
160,5 — 164,0	2	11	6,67 %	36,67 %
164,0 — 167,5	5	16	16,67 %	53,34 %
167,5 — 171,0	6	22	20%	73,34 %
171,0 — 174,5	1	23	3,33 %	76,67 %
174,5 — 178,0	7	30	23,33 %	100 %
Total	30		100 %	

5) a) Tabela 10

Notas	Frequência x_i	f_i	$(x_i - \bar{x})^2 \cdot f$
0 — 1	2	0.5	1
1 — 2	4	1.5	6
2 — 3	3	2.5	7.5
3 — 4	5	3.5	17.5
4 — 5	6	4.5	27
5 — 6	2	5.5	11
6 — 7	7	6.5	45.5
7 — 8	5	7.5	37.5
8 — 9	4	8.5	34
9 — 10	2	9.5	19
Σ	40		206

b) - variância

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i} = \frac{257.1}{40} \cong 6.43$$

c) - desvio padrão:

$$s = \sqrt{s^2} = \sqrt{6.43} \cong 2.54$$

d) - coeficiente de variação:

$$cv = \frac{s}{\bar{x}} = \frac{2.54}{5.15} \cong 0.49 \text{ ou } 49\%$$

Exercícios propostos ao final da Unidade II

6) a) Rol:

20	20	21	21	22	23	24	25	25	26
27	27	27	28	28	28	28	29	29	30
31	31	32	34	35	35	38	38	39	41

b)

Resíduos (L) x_i	f_i	$f_{\%}$
20	2	2
21	2	4
22	1	5
23	1	6
24	1	7
25	2	9
26	1	10
27	3	13
28	4	17
29	2	19
30	1	20
31	2	22
32	1	23
34	1	24
35	2	26
38	2	28
39	1	29
41	1	30
Total	30	

Percebeu como essa tabela é muito grande?

Por isso utilizamos a distribuição de frequências com intervalos de classes, que será feita no item d deste exercício. Compare as duas tabelas (item b e item d) e analise, qual das duas é mais viável?

- a) Utilizando a fórmula $k = 1 + 3,32 \cdot \log n$ que foi apresentada à você neste texto, vamos determinar o número de classes para esta distribuição, sendo $n = 30$, pois n representa o total de termos:

$$\begin{aligned} k &= 1 + 3,32 \cdot \log n \\ k &= 1 + 3,32 \cdot \log 30 \\ k &\approx 6 \end{aligned}$$

Isto quer dizer que a distribuição terá 6 classes.

Agora vamos calcular a amplitude das classes utilizando a fórmula que também foi apresentada no texto, onde $a = 41 - 20 = 21$, representa a diferença entre o maior e menor valores observados.

$$h = \frac{a}{k}$$

Então:

que também foi apresentada no texto, onde $a = 41 - 20 = 21$, representa a diferença entre o maior e menor valores observados. Então:

$$\begin{aligned} h &= \frac{a}{k} \\ h &= \frac{21}{6} \end{aligned}$$

$$h = 3,5$$

Isto significa que cada classe terá amplitude de 3,5.

- d) considerando os resultados obtidos no item anterior, temos que a primeira classe será 20|—23,5; a segunda de 23,5|—27; e assim por diante até a última classe que será 37,5|—41. Lembre-se que o símbolo da última classe é diferente, pois o número 41 está incluído nela.

Resíduos	f_c	f_{sc}	f_s
20 — 23,5	6	6	20%
23,5 — 27	4	10	13,33%
27 — 30,5	10	20	33,33%
30,5 — 34	3	23	10%
34 — 37,5	3	26	10%
37,5 — 41	4	30	13,34%
Total	30		100%

7) Para calcular a média, vamos acrescentar à tabela referente ao exercício 1, item b, a coluna do produto entre os valores de x_i e as suas respectivas frequências f_i :

Resíduos (L) x_i	f_i	f_{ac}	$x_i \cdot f_i$
20	2	2	40
21	2	4	42
22	1	5	22
23	1	6	23
24	1	7	24
25	2	9	50
26	1	10	26
27	3	13	81
28	4	17	112
29	2	19	58
30	1	20	30
31	2	22	62
32	1	23	32
34	1	24	34
35	2	26	70
38	2	28	76
39	1	29	39
41	1	30	41
Total	30		862

Então a média $\bar{x}$ será calculada utilizando a fórmula:

$$\bar{x} = \frac{\sum(x_i \cdot f_i)}{\sum f_i}$$

Substituindo os valores da tabela na fórmula, temos:

$$\bar{x} = \frac{862}{30} \cong 28,73$$

o que significa que a quantidade média de resíduos é de 28,73 litros. Para calcularmos a moda, vamos verificar qual o valor da maior frequência na tabela anterior.

Percebemos que a maior frequência é 4, e é referente ao valor de 28 litros. Isto quer dizer que, a quantidade de resíduos despejados na natureza que mais ocorreu foi de 28 litros.

Percebemos que a maior frequência é 4, e é referente ao valor de 28 litros. Isto quer dizer que, a quantidade de resíduos despejados na natureza que mais ocorreu foi de 28 litros.

Para o cálculo da mediana, precisamos primeiramente determinar o valor da posição em que se encontra a mediana. Como o valor de $n = 30$ é par, fazemos a média aritmética dos termos que se encontram nas posições:

$$\frac{n}{2} e \frac{n}{2} + 1$$

Então:

$$\frac{n}{2} = \frac{30}{2} = 15 \quad (15^{\text{a}} \text{ posição}) \quad e \quad \frac{n}{2} + 1 = \frac{30}{2} + 1 = 15 + 1 = 16 \quad (16^{\text{a}} \text{ posição})$$

Observando na coluna da frequência acumulada, o elemento que se encontra nas posições citadas à cima é 28. Logo, a mediana é 28 litros, o que significa que, em 50% dos dias a quantidade de resíduos despejada na natureza é inferior ou igual a 28 litros, e os demais 50% é superior ou igual a 28 litros.

8) Para o cálculo da média, vamos acrescentar à tabela referente ao item d do exercício 1, a coluna de x_i (que é a média aritmética entre os valores extremos dos intervalos de classe) e a coluna do produto entre os valores de x_i e as suas respectivas frequências f_i :

Por exemplo, para o valor de x_i $20+23,5 = 43,5 : 2 = 21,75$, e assim por diante....
E para o valor de $x_i \cdot f_i$ $21,75 \times 6 = 130,5$, e assim por diante...

Resíduos (L)	f_i	f_a	$x_i \cdot f_i$	$x_i \cdot f_i$
20 — 23,5	6	6	21,75	130,5
23,5 — 27	4	10	25,25	101
27 — 30,5	10	20	28,75	287,5
30,5 — 34	3	23	32,25	96,75
34 — 37,5	3	26	35,75	107,25
37,5 — 41	4	30	39,25	157
Total	30			88

Então a média $\bar{x}$ será calculada utilizando a fórmula:

$$\bar{x} = \frac{\sum(x_i \cdot y_i)}{\sum f_i}$$

Substituindo os valores da tabela na fórmula, temos:

$$\bar{x} = \frac{880}{30} \cong 29,33 ,$$

o que significa que a quantidade média de resíduos é de 29,33 litros.

Para calcularmos a moda, vamos verificar qual o valor da maior frequência na tabela anterior.

Percebemos que a maior frequência é 10, e é referente ao intervalo 27 — 30,5. Utilizando a fórmula estudada, temos:

$$M_o = L_i + \frac{f'}{f + f'} \cdot h$$

$$M_o = 27 + \frac{3}{4 + 3} \cdot 3,5$$

$$M_o = 28,5$$

Isto quer dizer que, a quantidade de resíduos despejados na natureza que mais ocorreu foi de 28,5 litros.

Para o cálculo da mediana, precisamos primeiramente determinar o valor da posição em que se encontra a mediana.

$$P = \frac{n}{2} , \text{ então } P = \frac{30}{2} = 15 \quad (15^{\text{a}} \text{ posição})$$

Observando na coluna da frequência acumulada, a classe que se encontra na 15ª posição é 27 — 30,5.

Utilizando a fórmula estudada da mediana, temos:

$$M_e = L_i + \frac{P - f_a}{f_{Me}} \cdot h$$

$$M_e = 27 + \frac{15 - 10}{10} \cdot 3,5$$

$$M_e = 28,75$$

Logo, a mediana é 28,75 litros, o que significa que, em 50% dos dias a quantidade de resíduos despejada na natureza é inferior ou igual a 28,75 litros, e os demais 50% é superior ou igual a 28,75 litros.

9) Para determinarmos a variância, vamos acrescentar à tabela referente ao exercício 2 a coluna dos cálculos $(x_i - \bar{x})^2 \cdot f_i$ para podermos utilizar a fórmula da variância estudada:

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

Sabemos que a média para esses dados já foi calculada no exercício 2, e o seu valor é $\bar{x} = 28,73$. Então, para o primeiro valor da tabela, temos:
 $(20 - 28,73)^2 \cdot 2 = 152,43$ (usaremos 2 casas decimais, e se a terceira casa decimal for 5 ou superior a 5, arredonda-se o valor da segunda casa decimal, aumentando em uma unidade). Procedendo da mesma forma, teremos a tabela a seguir:

Resíduos (L) x_i	f_i	f_a	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
20	2	2	40	152,43
21	2	4	42	119,51
22	1	5	22	45,29
23	1	6	23	32,83
24	1	7	24	22,37
25	2	9	50	27,83
26	1	10	26	7,45
27	3	13	81	8,98
28	4	17	112	2,13
29	2	19	58	0,15
30	1	20	30	1,61
31	2	22	62	10,31
32	1	23	32	10,69
34	1	24	34	27,77
35	2	26	70	78,63
38	2	28	76	171,87
39	1	29	39	105,47
41	1	30	41	150,55
Total	30		862	975,87

Logo, substituindo na fórmula

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i}, \text{ temos: } s^2 = \frac{975,87}{30} = 32,53.$$

O desvio padrão é a raiz quadrada da variância, ou seja $s = \sqrt{s^2} = \sqrt{32,53} = 5,70$.

Isto quer dizer que a média dos desvios dos termos em relação à média de 28,73 é de 5,70, para mais ou para menos.

Para o cálculo do coeficiente de variação, vamos usar a fórmula já estudada, sendo que $s = 5,70$ e $\bar{x} = 28,73$ já foram calculados anteriormente. Então:

$$cv = \frac{s}{\bar{x}} = \frac{5,70}{28,73} = 0,1984$$

ou 19,84%, o que significa dizer que existe uma variação de aproximadamente 19,84% dos dados em torno da média.

10) Para determinarmos a variância, vamos acrescentar à tabela referente ao exercício 3 a coluna dos cálculos $(x_i - \bar{x})^2 \cdot f_i$, para podermos utilizar a fórmula da variância estudada:

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

Sabemos que a média para esses dados já foi calculada no exercício 3, e o seu valor é $\bar{x} = 29,33$. Então, para o primeiro valor da tabela, temos:

$$(21,75 - 29,33)^2 \cdot 6 = 344,74$$

(usaremos 2 casas decimais, e se a terceira casa decimal for 5 ou superior a 5, arredonda-se o valor da segunda casa decimal, aumentando em uma unidade). Procedendo da mesma forma, teremos a tabela a seguir:

Resíduos (L)	x_i	f_i	f_a	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
20 23,5	6	6	21,75	130,5	344,74
23,5 27	4	10	25,25	101	66,59
27 30,5	10	20	28,75	287,5	3,36
30,5 34	3	23	32,25	96,75	25,58
34 37,5	3	26	35,75	107,25	123,65
37,5 41	4	30	39,25	157	393,63
Total	30			880	957,55

Logo, substituindo na fórmula $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i}$ temos $s^2 = \frac{957,55}{30} = 31,92$.

O desvio padrão é a raiz quadrada da variância, ou seja, $s = \sqrt{s^2} = \sqrt{31,92} = 5,65$.

Isto quer dizer que a média dos desvios dos termos em relação à média de 29,33 é de 5,65, para mais ou para menos.

Para o cálculo do coeficiente de variação, vamos usar a fórmula já estudada,

sendo que

$cv = \frac{s}{\bar{x}} = \frac{5,65}{29,33} = 0,1926$ ou 19,26%, o que significa dizer que existe uma variação de aproximadamente 19,26% dos dados em torno da média.

11) a) Rol:

12	12,4	12,4	12,8	13	13	13	13,2	13,6	13,7
13,8	14	14	14	14,3	14,5	14,5	14,6	14,7	15
15	15,1	15,2	15,2	15,2	15,2	15,5	15,7	15,7	15,8
16	16	16	16,1	16,2	16,3	16,3	16,9	16,9	17
17	17	17,1	17,1	17,2	17,2	17,2	17,3	18	18,3

b) Utilizando a fórmula $k = 1 + 3,32 \cdot \log n$ que foi apresentada à você neste texto, vamos determinar o número de classes para esta distribuição, sendo $n = 50$, pois n representa o total de termos:

$$k = 1 + 3,32 \cdot \log n$$

$$k = 1 + 3,32 \cdot \log 50$$

$$k = 6,64$$

$$k = 7 \quad \text{Isto quer dizer que a distribuição terá 7 classes.}$$

Agora vamos calcular a amplitude das classes utilizando a fórmula $h = \frac{a}{k}$, que também foi apresentada no texto, onde $a = 18,3 - 12 = 6,3$ representa a diferença entre o maior e menor valores observados. Então: $h = \frac{a}{k} \rightarrow h = \frac{6,3}{7} \rightarrow h = 0,9$

Isto significa que cada classe terá amplitude de 0,9.

c) considerando os resultados obtidos no item anterior, temos que a primeira classe será $12 \vdash 12,9$; a segunda de $12,9 \vdash 13,8$; e assim por diante até a última classe que será $17,4 \vdash 18,3$. Lembre-se que o símbolo da última classe é diferente, pois o número 18,3 está incluído nela.

Componente (g)	f_i	f_a	f_r
12,0 $\vdash$ 12,9	4	4	8%
12,9 $\vdash$ 13,8	6	10	12%
13,8 $\vdash$ 14,7	8	18	16%
14,7 $\vdash$ 15,6	9	27	18%
15,6 $\vdash$ 16,5	10	37	20%
16,5 $\vdash$ 17,4	11	48	22%
17,4 $\vdash$ 18,3	2	50	4%
Total	50		100%

- 12) Para calcular a média, vamos acrescentar à tabela referente ao exercício 6, a coluna da coluna de x_i e a coluna do produto entre os valores de x_i e as suas respectivas frequências f_i :

Componente (g)	f_i	f_a	f_r	x_i	$x_i f_i$
12,0 —— 12,9	4	4	8%	12,45	49,8
12,9 —— 13,8	6	10	12%	13,35	80,1
13,8 —— 14,7	8	18	16%	14,25	114
14,7 —— 15,6	9	27	18%	15,15	136,35
15,6 —— 16,5	10	37	20%	16,09	160,9
16,5 —— 17,4	11	48	22%	16,95	186,45
17,4 —— 18,3	2	50	4%	17,85	35,7
Total	50		100%		763,3

Então a média $\bar{x}$ será calculada utilizando a fórmula

$$\bar{x} = \frac{\sum (x_i \cdot y_i)}{\sum f_i} \quad \text{Substituindo os valores da tabela na fórmula, temos:} \quad \bar{x} = \frac{763,3}{50} \cong 15,27$$

O que significa que a quantidade média do componente é de 15,27 gramas.

Para calcularmos a moda, vamos verificar qual o valor da maior frequência na tabela anterior.

Percebemos que a maior frequência é 11, e é referente ao intervalo 16,5 — 17,4. Utilizando a fórmula estudada, temos:

$$M_o = L_i + \left(\frac{f'}{f + f'} \right) h$$

$$M_o = 16,5 + \frac{2}{10 + 2} \cdot 0,9$$

$$M_o = 16,65$$

Isto quer dizer que, a quantidade do componente encontrado no fertilizante que mais ocorreu foi de 16,65 gramas.

Para o cálculo da mediana, precisamos primeiramente determinar o valor da posição em que se encontra a mediana $P = \frac{n}{2}$, então $P = \frac{50}{2} = 25$ (25ª posição)

Observando na coluna da frequência acumulada, a classe que se encontra na 25ª posição é 14,7 — 15,6.

Utilizando a fórmula estudada da mediana, temos:

$$M_e = L_i + \frac{P - f_a}{f_{Me}} \cdot h$$

$$M_e = 14,7 + \frac{25 - 18}{9} \cdot 0,9$$

$$M_e = 15,4$$

Logo, a mediana é 15,4 gramas, o que significa que, em 50% dos fertilizantes a quantidade do componente químico encontrada é inferior ou igual a 15,4 gramas, e nos demais 50% é superior ou igual a 15,4 gramas.

13) Para determinarmos a variância, vamos acrescentar à tabela referente ao exercício 7 a coluna dos cálculos $(x_i - \bar{x})^2 \cdot f_i$, para podermos utilizar a fórmula da variância estudada:

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

Sabemos que a média para esses dados já foi calculada no exercício 7, e o seu valor é $\bar{x} = 15,27$. Então, para o primeiro valor da tabela, temos:

$$(12,45 - 15,27)^2 \cdot 4 = 31,81$$

(usaremos 2 casas decimais, e se a terceira casa decimal for 5 ou superior a 5, arredonda-se o valor da segunda casa decimal, aumentando em uma unidade).

$$s = \sqrt{s^2} = \sqrt{2,32} = 1,52$$

Procedendo da mesma forma, teremos a tabela a seguir:

Componente (g)	f_i	f_a	f_r	x_i	$x_i f_i$	$(x_i - \bar{x})^2 \cdot f_i$
12,0 ── 12,9	4	4	8%	12,45	49,8	31,81
12,9 ── 13,8	6	10	12%	13,35	80,1	22,12
13,8 ── 14,7	8	18	16%	14,25	114	8,32
14,7 ── 15,6	9	27	18%	15,15	136,35	0,13
15,6 ── 16,5	10	37	20%	16,09	160,9	6,72
16,5 ── 17,4	11	48	22%	16,95	186,45	31,05
17,4 ── 18,3	2	50	4%	17,85	35,7	13,31
Total	50		100%		763,3	113,46

Como nesta situação temos uma amostra e não toda população, a fórmula da variância será dada por:

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

,temos: $s^2 = \frac{113,46}{50 - 1} = 2,32$

O desvio padrão é a raiz quadrada da variância, ou seja, $s = \sqrt{s^2} = \sqrt{2,32} = 1,52$

Isto quer dizer que a média dos desvios dos termos em relação à média de 15,27 é de 1,52, para mais ou para menos.

Para o cálculo do coeficiente de variação, vamos usar a fórmula já estudada, sendo que

$$cv = \frac{s}{\bar{x}} = \frac{1,52}{15,27} = 0,0995$$

ou 9,95%, o que significa dizer que existe uma variação de aproximadamente 9,95% dos dados em torno da média.

AS AUTORAS

Prof.^a Thelma Pretel Brandão Vecchi

É natural de Campo Mourão, Paraná. Licenciada em Matemática pela Universidade Estadual de Maringá – PR em 1993 e especialista em Educação Matemática pela Faculdade Estadual de Ciências e Letras de Campo Mourão – PR em 1998. Concluiu mestrado em Métodos Numéricos em Engenharia, pela Universidade Federal do Paraná, no ano de 2004. Iniciou sua carreira de docente em 1993, ministrando aulas de matemática para o ensino fundamental e médio. Trabalhou também no ensino superior, ministrando aulas de cálculo, estatística, geometria analítica, cálculo numérico, entre outras disciplinas, em instituição pública e privada. Atualmente é professora efetiva da Universidade Tecnológica Federal do Paraná, *campus* Campo Mourão, desde julho de 2006.

Prof.^a Flávia Aparecida Reitz Cardoso

É natural de Enéas Marques, Paraná. Licenciada em matemática pela Faculdades Reunidas de Administração, Ciências Contábeis e Econômicas de Palmas – PR em 1992 e especialista em Modelagem Matemática pela Faculdade Estadual de Ciências e Letras de Campo Mourão – PR em 1999. Concluiu mestrado em Métodos Numéricos em Engenharia, pela Universidade Federal do Paraná, no ano de 2004. Iniciou sua carreira de docente em 1993, ministrando aulas de matemática e física para o ensino médio. Trabalhou também no ensino superior, ministrando aulas de cálculo, estatística, geometria analítica, cálculo numérico, entre outras disciplinas, em instituição pública. Atualmente é professora efetiva da Universidade Tecnológica Federal do Paraná, *campus* Campo Mourão, desde setembro de 2004.