

TEC
ADM

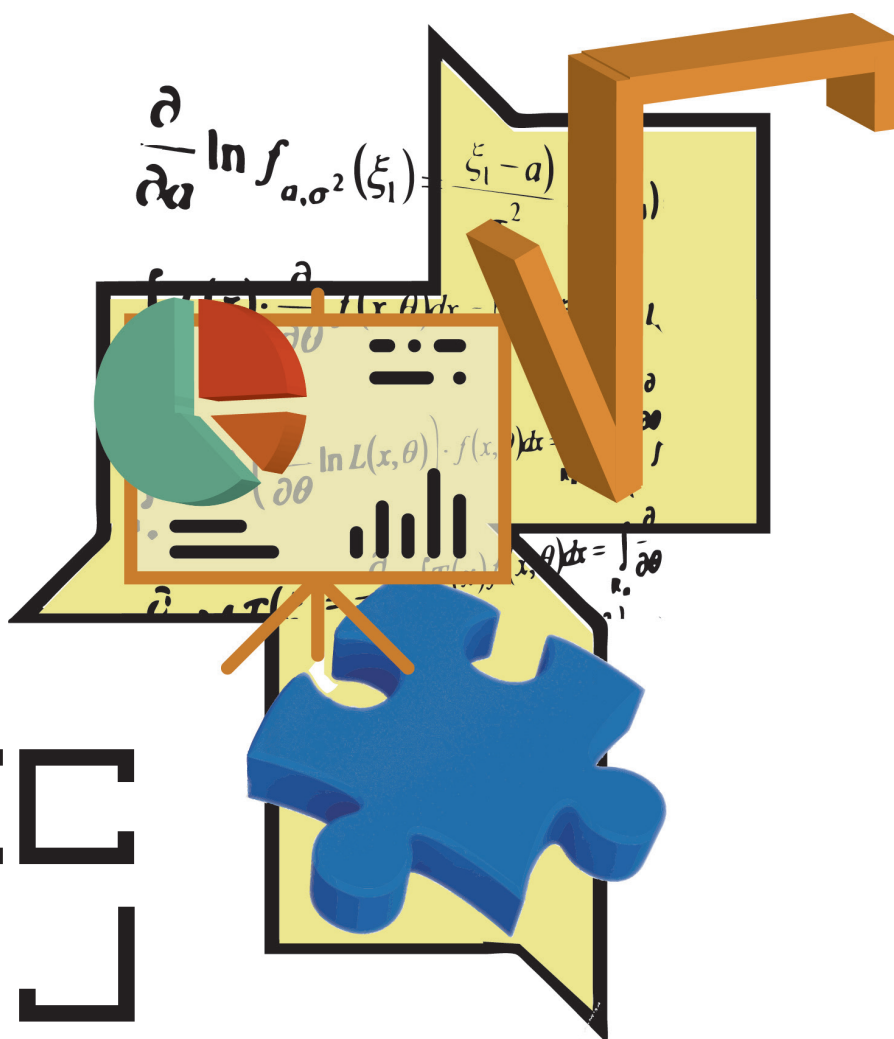
APOSTILA COMPILADA

ORGANIZADOR: GLAUCO CARVALHO CAMPOS

Estatística Aplicada à Contabilidade

VOLUME ÚNICO

TEC
ADM



Governo do Estado do Rio de Janeiro

Governador: Wilson Witzel

Secretário de Estado de Ciência, Tecnologia e Inovação: Leonardo Rodrigues

FAETEC – Fundação de Apoio à Escola Técnica

www.faetec.rj.gov.br

Presidente: Romulo Mello Massacesi

Vice Presidente Educacional: Maicon Luiz Lisboa Felix

Diretora da Diretoria de Desenvolvimento da Educação: Márcia Cristina Pinheiro Farinazo

Coordenadora do Programa Tec RJ: Bárbara Sales Castelhana

Fundação Cecierj

www.cecierj.edu.br

Presidente: Gilson Rodrigues

Vice-Presidente de Educação Superior a Distância: Marilvia Dansa de Alencar

Vice-Presidente Científica: Monica Dahmouche

Organização de Conteúdo
Glauco Carvalho Campos

Coordenação do Programa Tec RJ
Priscila de Souza Costa Couto

Diretoria de Extensão
Michelle Casal Fernandes

Diretoria de Material Didático
Bruno José Peixoto

Coordenação de Design Instrucional
Flávia Busnardo da Cunha

Diretoria de Material Impresso
Ulisses Schnaider

Ilustração
Andre Amaral

Capa
Larissa Averbug e Andre Amaral

Projeto Gráfico
Larissa Averbug

Produção Gráfica
Fábio Rapello Alencar

APOSTILA COMPILADA

Composição do material:

Rede E-TEC Brasil

Título: Estatística Aplicada
(Técnico em Finanças)

Autor: Vilmar dos Santos Alves
UFMT, 2013, 168 p., 11 aulas.

Link: proedu.rnp.br/handle/123456789/1543

Título: Estatística Aplicada
(Técnico em Segurança do Trabalho)

Autor: Paulo Roberto Rufino Pereira et al.
CECIERJ, 2009.





TEC
ADM

ORGANIZADOR: GLAUCO CARVALHO CAMPOS

Estatística Aplicada à Contabilidade

VOLUME ÚNICO

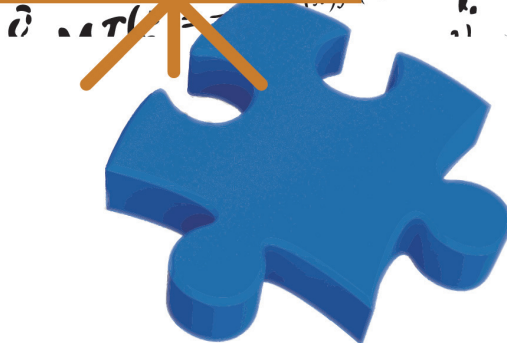


FAETEC

Secretaria de
Ciência, Tecnologia e Inovação



$$\frac{\partial}{\partial a} \ln f_{a, \sigma^2}(\xi_1) = \frac{\xi_1 - a}{\sigma^2}$$





Boas Vindas

Prezado(a) Aluno(a),

Seja bem-vindo ao TEC RJ

Você está recebendo o material didático impresso para acompanhamento de seus estudos, contendo as informações necessárias para seu aprendizado, exercício de desenvolvimento e fixação dos conteúdos. Este material foi selecionado e reeditado à partir da vasta gama de materiais disponibilizados pelo Ministério da Educação na Rede e-Tec Brasil. Assim, um único volume impresso pode apresentar aulas oriundas de materiais produzidos por diferentes instituições atuantes da Rede e-Tec Brasil. Sua ordenação final seguiu a orientação dada pelos coordenadores do TEC RJ.

Além do material didático impresso, disponibilizamos um Ambiente Virtual com conteúdos complementares, atividades individuais de reforço e colaboração com seus colegas, e outros materiais que podem auxiliar na sua aprendizagem. Você também pode enviar perguntas pelos fóruns de dúvida ao corpo docente de seu curso.

Tudo isto foi cuidadosamente planejado para que você tenha uma experiência gratuita de alta qualidade, que resulte em sólida formação técnica.

Bons estudos!



Estatística Aplicada

Vilmar dos Santos Alves



INSTITUTO FEDERAL
RONDÔNIA

Cuiabá-MT
2014

Presidência da República Federativa do Brasil
Ministério da Educação
Secretaria de Educação Profissional e Tecnológica
Diretoria de Integração das Redes de Educação Profissional e Tecnológica

© Este caderno foi elaborado pelo Instituto Federal de Educação, Ciência e Tecnologia / RO, para a Rede e-Tec Brasil, do Ministério da Educação, em parceria com a Universidade Federal do Mato Grosso.

Equipe de Revisão
Universidade Federal de Mato Grosso – UFMT

Coordenação Institucional
Carlos Rinaldi

Coordenação de Produção de Material Didático Impresso
Pedro Roberto Piloni

Designer Educacional
Izabel Solyszko Gomes

Designer Master
Marta Magnusson Solyszko

Ilustração
Tatiane Hirata

Diagramação
Yuri da Silva Peixoto

Revisão de Língua Portuguesa
Marta Maria Covezzi

Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO

Campus Porto Velho Zona Norte

Direção-Geral
Miguel Fabrício Zamberlan

Direção de Administração e Planejamento
Gilberto Laske

Departamento de Produção de EaD
Ariadne Joseane Felix Quintela

Coordenação de Design Visual e Ambientes de Aprendizagem
Rafael Nink de Carvalho

Coordenação da Rede e-Tec
Ruth Aparecida Viana da Silva

Projeto Gráfico
Rede e-Tec Brasil / UFMT

Estatística Aplicada - Finanças

A474e Alves, Vilmar dos Santos.

Estatística aplicada / Vilmar dos Santos Alves. – Cuiabá: Ed.UFMT, 2013.

168 p.
Curso Técnico – Rede E- Tec. (IFRO)

ISBN 978-85-68172-10-0

1. Estatística - Introdução. 2. Estatística Descritiva. I. Título.

CDU 519.2

Ficha Catalográfica Elaborada por Almira de Araújo Medeiros – CRB1 2.327

Apresentação Rede e-Tec Brasil

Prezado(a) estudante,

Bem-vindo(a) à Rede e-Tec Brasil!

Você faz parte de uma rede nacional de ensino que, por sua vez, constitui uma das ações do Pronatec - Programa Nacional de Acesso ao Ensino Técnico e Emprego. O Pronatec, instituído pela Lei nº 12.513/2011, tem como objetivo principal expandir, interiorizar e democratizar a oferta de cursos de Educação Profissional e Tecnológica (EPT) para a população brasileira propiciando caminho de acesso mais rápido ao emprego.

É neste âmbito que as ações da Rede e-Tec Brasil promovem a parceria entre a Secretaria de Educação Profissional e Tecnológica (Setec) e as instâncias promotoras de ensino técnico, como os institutos federais, as secretarias de educação dos estados, as universidades, as escolas e colégios tecnológicos e o Sistema S.

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

A Rede e-Tec Brasil leva diversos cursos técnicos a todas as regiões do país, incentivando os estudantes a concluir o ensino médio e a realizar uma formação e atualização contínuas. Os cursos são ofertados pelas instituições de educação profissional e o atendimento ao estudante é realizado tanto nas sedes das instituições quanto em suas unidades remotas, os polos.

Os parceiros da Rede e-Tec Brasil acreditam em uma educação profissional qualificada – integradora do ensino médio e da educação técnica – capaz de promover o cidadão com capacidades para produzir, mas também com autonomia diante das diferentes dimensões da realidade: cultural, social, familiar, esportiva, política e ética.

Nós acreditamos em você!

Desejamos sucesso na sua formação profissional!

Ministério da Educação
Abril de 2014

Nosso contato
etecbrasil@mec.gov.br



Indicação de ícones

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



Atenção: indica pontos de maior relevância no texto.



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



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



Mídias integradas: remete o tema para outras fontes: livros, filmes, músicas, *sites*, programas de TV.



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



Reflita: momento de uma pausa na leitura para refletir/escrever sobre pontos importantes e/ou questionamentos.



Palavra do Professor-autor

Caro (a) estudante,

Seja bem vindo(a) ao componente curricular Estatística Aplicada!

É com imenso prazer que apresento a você o conteúdo desta disciplina, com carga horária de 60 horas.

Esperamos que este estudo venha contribuir significativamente para o seu processo de aprendizagem e que possa trazer as informações necessárias para um bom desempenho profissional.

No decorrer das aulas, você terá a oportunidade de interagir através da plataforma Moodle e participar de *chats*, fóruns, realizar atividades individuais e em grupo, o que lhe proporcionará o crescimento e o aperfeiçoamento do aprendizado.

Vamos aprender mais? Somos parceiros nesta etapa da sua formação, conte comigo.

Contamos com sua participação!

Dedique-se aos estudos e faça a diferença!





Apresentação da Disciplina

Olá! Seja bem-vindo (a) ao Curso de Habilitação Profissional Técnico de Nível Médio em Finanças!

A disciplina Estatística Aplicada, que você está iniciando, faz parte deste curso. No decorrer de nossas aulas, mostraremos a você que a Estatística é um instrumento que fornece informações úteis para a tomada de decisões dentro e fora da empresa.

A Estatística é uma ciência com técnicas específicas que tem aplicação em todas as áreas do conhecimento e segmentos profissionais. Por meio da utilização da estatística, é possível realizar previsões no sentido de ajudar a planejar a obtenção de dados, interpretar e analisar os dados obtidos e apresentar os resultados de maneira a facilitar a tomada de decisões razoáveis.

Desta forma, conhecer bem os conceitos, técnicas e aplicações da estatística constitui um aspecto muito importante para a sua formação e atuação no mercado de trabalho como técnico em finanças, bem como para aplicação em outras atividades de sua vida profissional.

Neste material, temos informações preciosas que podem colaborar com seu processo de construção do conhecimento. Bons estudos!



Sumário

Aula 1 - Introdução à Estatística	13
1.1. Um pouco de história	14
1.2. Conhecendo a estatística	16
1.3. Os Dados	23
1.4. Elementos e variáveis	24
Aula 2 - Conceitos Básicos e Elementares de Matemática	31
2.1. Porcentagem	31
2.2. Dados absolutos e dados relativos	35
Aula 3 - Método Estatístico, População e Amostra	41
3.1. O método estatístico e suas fases	42
3.2. População e amostra	47
3.3. Técnicas de amostragem	51
3.4. Amostragem probabilística simples	53
3.5. Amostragem probabilística estratificada	53
3.6. Amostragem probabilística sistemática	54
Aula 4 - Estatística Descritiva	57
4.1. Ramos da estatística	57
4.2. Distribuição de frequências	61
Aula 5 - Medidas de Posição	71
5.1. Média aritmética simples	72
5.2. Mediana	76
5.3. Moda	79
5.4. Média ponderada	81
5.5. Média de dados agrupados	83
Aula 6 - Medidas de Dispersão	87
6.1. Amplitude total	87
6.2. Desvio médio	89
6.3. Variância (S^2)	92
6.4. Desvio padrão (S)	93
6.5. Desvio padrão em uma distribuição de frequência agrupada	97



Aula 7 - Tabelas e Gráficos	101
7.1 Tabelas	101
7.2. Gráficos	106
Aula 8 - Correlação	115
8.1. Correlação positiva	116
8.2. Correlação negativa	117
8.3. Correlação não linear	118
8.4 Coeficiente de correlação	119
Aula 9 - Regressão	127
9.1 Regressão linear	127
9.2 Equação de uma reta de regressão	129
9.3 Coeficiente de determinação	132
9.4 Aplicações das linhas de regressão	133
Aula 10 - Análise de Dados e Indicativos	137
10.1 Análise de dados	138
10.2 Coeficiente de correlação	144
10.3 Regressão linear	145
10.4 Coeficiente de determinação	146
Aula 11 - Probabilidade	149
11.1 Conceito de probabilidade	149
11.2 Representação e interpretação de valores de probabilidade	151
Palavras finais	159
Guia de Soluções	160
Referências	165
Obras Consultadas	167
Bibliografia Básica	167
Currículo do Professor-autor	168

Aula 1 - Introdução à Estatística

Objetivos:

- Conceituar a estatística.
- Reconhecer as aplicações da estatística.
- Apontar um conjunto de dados estatísticos.
- Identificar e exemplificar os tipos de variáveis.

Prezado(a) estudante, esta é a primeira aula de nossa disciplina. Nesta aula, veremos um pouco da história da estatística, o conceito de estatística e de dados, bem como a identificação de variáveis.

Esperamos que, você compreenda bem estes conceitos e que os mesmos sirvam como base para o desenvolvimento de nossas próximas aulas desta disciplina.

Para começo de conversa, é relevante observar que a estatística está presente em nosso dia a dia, nas mais variadas formas e significações. Você certamente já se deparou com informações como a da figura 1 em jornais, livros, internet ou em outros materiais que trazem informações provenientes de dados ou pesquisas estatísticas e que talvez você não tenha conseguido compreender bem.



Veja o quanto o Brasil cresceu (em %)

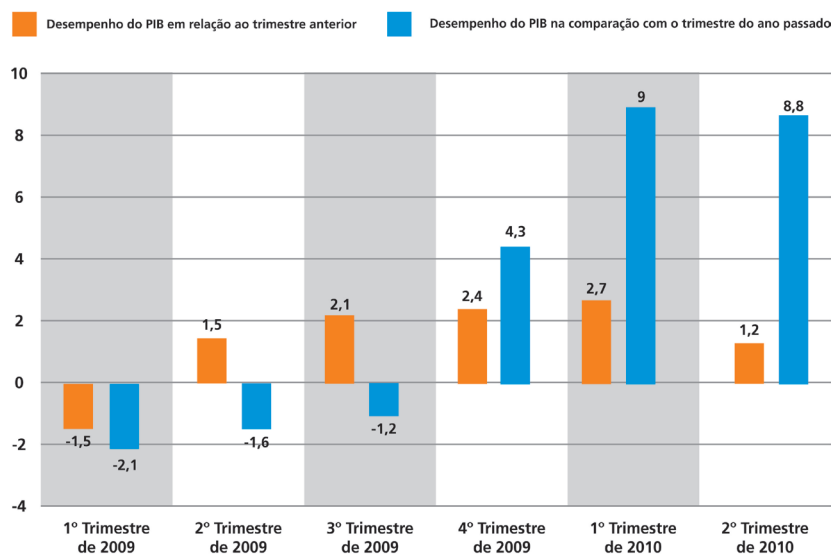


Figura 1 - Crescimento do PIB Brasileiro - 2009/2010

Fonte: IBGE (2010)

Ao final desta disciplina, com certeza, você terá maior clareza e segurança para ler dados estatísticos expressos em gráficos e tabelas.

Em nosso estudo sobre o conteúdo de Estatística, parece-nos importante conhecer um pouco da origem desta técnica tão relevante e tão presente em nossas vidas. Neste sentido, o texto a seguir traz alguns questionamentos:

- Quando a humanidade começou a registrar fatos e fazer uso dessas informações?
- De que forma eram feitos estes registros?
- O que é efetivamente estatística?

Pois bem, estes e outros questionamentos serão respondidos mais adiante!

1.1 Um pouco de história

Desde o início da formação das sociedades, a humanidade tem se preocupado em criar representações para registrar os fatos. É óbvio que as pessoas na pré-história não possuíam o domínio da escrita, tampouco faziam uso de planilhas como usualmente fazemos hoje. Contudo, a falta de domínio da escrita não significava necessariamente a ausência de formas de registros.



Nos primeiros registros, utilizavam-se apenas de símbolos e ilustrações para se comunicar e para registrar sua memória, deixando-nos como herança daquele período as artes rupestres.

Na vida da humanidade pré-histórica, tinha lugar a Arte e o espírito de conservação daquilo de que se necessitava.

Estudos arqueológicos demonstram que o Homem da Pré-História (a fase da História que precede a escrita) já conservava, além de cerâmicas, armas e utensílios trabalhados na pedra, nos ossos dos animais que abatiam e no metal. Arqueólogos e antropólogos datando e estudando peças extraídas em escavações conferem a estes vestígios seu real valor como “documentos históricos”, verdadeiros testemunhos da vida do Homem em tempos remotos e de culturas extintas. Disponível em: <http://pt.wikipedia.org/wiki/Arte_rupestre> Acesso em :20 fev. 2013.

A gravura abaixo é uma arte rupestre encontrada no sítio arqueológico de Vila Nova de Foz Côa – Portugal.



Figura 2 - Gravura rupestre Vila Nova de Foz Côa - Portugal

Fonte: Wikimedia Commons (2013)

Com o passar do tempo, a humanidade aprimorou-se criando a escrita, passando posteriormente a registrar não somente os fatos, mas também a organizar dados, processar análises e tomar decisões com base nos dados analisados. Marca-se assim o início da significação dos registros do fatos e dados, pois agora tais registros tornam-se muito úteis nas tomadas de decisões dos povos antigos.



A Arte rupestre, pintura rupestre ou, ainda, **gravura rupestre**, são termos dados às mais antigas representações artísticas conhecidas, as mais antigas, datadas do período Paleolítico Superior (40.000 a.C.), gravadas em abrigos ou cavernas, em suas paredes e tetos rochosos, ou também em superfícies rochosas ao ar livre, mas em lugares protegidos, normalmente datando de épocas pré-históricas. Você encontrará mais informações no site http://pt.wikipedia.org/wiki/Arte_rupestre



1.2 Conhecendo a estatística

1.2.1 A estatística milenar

As necessidades humanas que exigem o conhecimento numérico dos recursos disponíveis começaram a surgir a partir do nascimento das organizações nas sociedades primitivas.

Diversas evidências indicam que vários povos, desde a Antiguidade, já registravam o número de seus habitantes, número de nascimentos, número de óbitos, estimavam riquezas, poderio militar, dentre outros, por meio de procedimentos que se assemelhavam ao que hoje chamamos de estatística.

Entre as diversas evidências, conforme diversos autores, dentre os quais destacamos Memória (2004) e Vieira (2008), podemos citar os seguintes acontecimentos:

- Há indícios de que 3000 anos a.C. já se faziam censos na Babilônia, na China, onde usualmente estas informações eram utilizadas para a taxa-ção de impostos ou para o alistamento militar.
- No antigo Egito, os faraós fizeram uso sistemático de informações de caráter estatístico. Acredita-se que, aproximadamente em 3050 a.C., Heródoto realizou um estudo das riquezas da população do Egito com a finalidade de averiguar quais os recursos humanos e econômicos disponíveis para a construção das pirâmides.
- O recenseamento de todos os judeus pelo Império Romano, na época de Cesar Augusto.
- Os balancetes do império romano, o inventário das posses de Carlos Magno, o Doomsday Book, registro que Guilherme, o Conquistador, invasor normando da Inglaterra, no século XI, mandou realizar, das propriedades rurais dos conquistados anglo-saxões, para inteirar-se de suas riquezas, são alguns exemplos anteriores à emergência da estatística descritiva no século XVI, na Itália.
- Guilherme, "O Conquistador", que reinou entre 1066 e 1087, ordenou a realização de um levantamento estatístico da Inglaterra para levantar informações sobre terras, proprietários, uso da terra, animais, entre outros, a fim de servir de base para o cálculo de impostos.

Contudo, mesmo que a prática de coletar dados sobre colheitas, composição da população humana ou de animais, impostos, entre outros, fosse conhecida pelos egípcios, hebreus, romanos e diversos outros povos, apenas no século XVII, a Estatística passou a ser considerada disciplina autônoma, tendo como objetivo básico a descrição dos BENS do Estado.



O desenvolvimento da estatística teve origem nas aplicações práticas ou empíricas, pois nenhuma disciplina tem interagido tanto com as demais disciplinas em suas atividades do que ela, dado que é, por sua natureza, a ciência do significado e do uso dos dados. Contudo, apesar de seu uso se configurar como uma prática milenar, só começou realmente a existir como ciência autônoma a partir do início do século 20, com o surgimento da estatística moderna.

1.2.2 A Estatística a partir do século XX

A estatística é uma técnica milenar, inclusive com aplicações na agricultura, nas estratégias de governos e no desenvolvimento de novos produtos. O surgimento da **inferência estatística**, entretanto, caracteriza o avanço da estatística moderna, e “tem funcionado como uma grande caixa de ferramentas dos métodos estatísticos empregados”. (WALPOLE, 2009, p. 1).

Convém destacar que tais métodos foram desenvolvidos para contribuir com o processo de realizar julgamentos científicos diante de incerteza e variação.

Para Walpole (2009, p. 1), “os métodos estatísticos são usados para analisar os dados de um processo, e identificar, dentro de um processo, que mudanças podem ser feitas para melhorar a qualidade”.

Há várias razões pelas quais a estatística e a necessidade de estudar a estatística têm crescido nos últimos anos, especialmente nas últimas cinco décadas. Uma delas consiste no crescimento da abordagem quantitativa aplicada em todas as ciências, como na Administração e em diversas atividades em nossas vidas. Isso inclui o planejamento, a análise de situações recorrentes, tais como acidentes de trabalho em determinadas áreas de serviço, ou o volume de chuvas em um período específico, entre diversas outras.

O avanço tecnológico e o surgimento da informática, razões estas que merecem especial destaque, aumentaram enormemente a nossa capacidade de coleta e armazenamento de dados, possibilitando a realização de análises estatísticas sofisticadas em busca de melhorias, até então muito difíceis de serem realizadas. (FREUND, 2006).

Tomando um exemplo das possibilidades e usos da informática como instrumento de coleta e de armazenamento, podemos citar a medição do índice pluviométrico de determinada região (índice de chuvas). Tal medida não é nova e, além disso, não é necessariamente preciso um computador para



A palavra Estatística surgiu somente em 1749, atribuída pelo alemão Gottfried Achenwall (1719-1772), professor da Universidade de Göttingen.



realizá-la, uma vez que a referida medida é dada em milímetros, ou seja, em litros de chuva por metro quadrado.

Isso significa que só precisaríamos de um recipiente de 1m^2 disposto ao ar livre durante a chuva e já teríamos a coleta da água.

Tomando como exemplo, dizer que em uma região choveu 100 mm, significa dizer que em uma área de 1 m^2 , a lâmina de água formada pela chuva que caiu apresenta uma altura de 100 milímetros. Esse volume pode ser obtido calculando o volume do paralelepípedo de 1 m^2 de área da base e altura de 100 milímetros = 0,1 metros.

Assim, o volume da chuva será dado por:

$$V = (\text{área da base}) \times \text{altura}$$

$$V = 1 \times 0,1 = 0,1\text{ m}^3$$

Esse volume pode ser determinado em litros, lembrando que $1\text{ m}^3 = 1000$ litros.

Assim, uma chuva de 100 milímetros equivale a um volume, em litros, de:

$$V = 0,1 \times 1000 = 100\text{ litros}$$

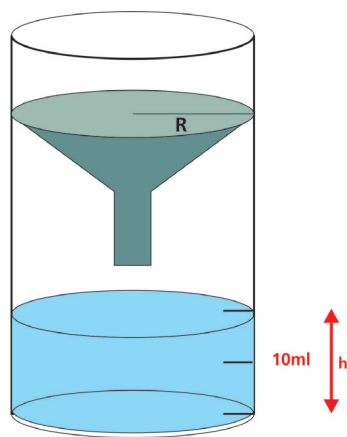
Isso implica dizer que, para cada metro quadrado da região, houve uma precipitação de 100 litros.



Figura 3

Fonte: Ilustradora

Mas agora, digamos que quiséssemos saber qual foi o índice de chuva por hora durante o intervalo de um ano. Neste caso, teríamos que conferir a cada hora, pontualmente, o que seria muito trabalhoso.



A-Z

O **pluviômetro** é um aparelho meteorológico destinado a medir, em milímetros, a altura da lâmina de água gerada pela chuva que caiu numa área de 1m^2 .

Figura 4 - Pluviômetro artesanal

Fonte: Ilustradora

Com o avanço da informática, a novidade é que só precisamos de um pequeno pluviômetro com sensor ligado a um dispositivo computadorizado. Neste aparelho, seria feita a programação para capturar o resultado a cada intervalo de tempo de uma hora, quinze, dez ou até um minuto, e o próprio computador se encarregaria de armazenar toda a série de dados nos horários estabelecidos.

Percebeu a diferença? Pois bem, mas, se por um lado, a informática facilitou a coleta e o armazenamento de dados, também exigiu o desenvolvimento de modelos estatísticos mais elaborados, pois de nada adiantam grandes volumes de dados, se a análise destes for confusa e sujeita a erros.

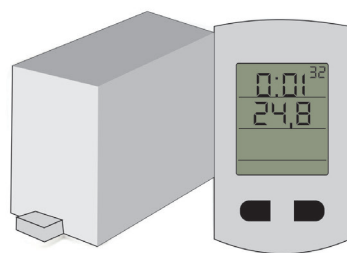


Figura 5 - Modelo de um pluviômetro eletrônico

Fonte: Ilustradora

Vale lembrar que a medição do índice de chuvas é apenas um exemplo, dentre as mais variadas utilizações da estatística no mundo moderno.

1.2.3 O que é, afinal, Estatística?

A estatística significa, para a maioria das pessoas, apenas uma descrição numérica muitas vezes utilizada em nosso cotidiano como sinônimo de apresentar tabelas, de números, em colunas esportivas ou econômicas em jornais e revistas, ilustrações com gráficos, ou quando muito, associa-se a estatística à previsão de resultados eleitorais.

A figura 5 traz alguns dos exemplos reais de informações a que temos acesso em nosso cotidiano.

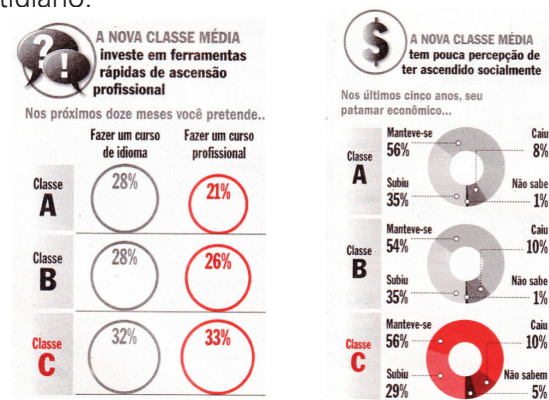


Figura 6 - Exemplos de estatística em notícias

Fonte: Revista Veja (14 dez 2011)

É correto afirmar que as gravuras trazem informações estatísticas, no entanto, estas, quando observadas de forma isolada do banco de dados (observações) que lhe deu origem, não constituem a estatística propriamente dita, e sim um texto de natureza gráfica que a representa.



Figura 7

Fonte: Ilustradora

Como veremos mais adiante, apesar de as situações da figura 4 serem confundidas com a estatística, esta envolve métodos e técnicas específicas, sem



os quais não passa de um conjunto de dados.

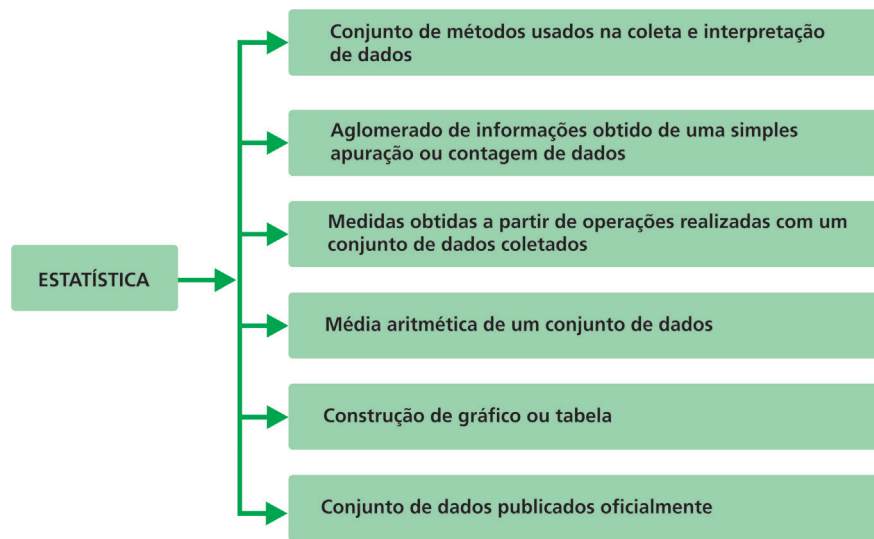


Figura 8 - O que não é estatística

Fonte: Adaptado de FREUND (2006)

Caro(a) estudante, é importante compreender a utilização da palavra estatística. Situações semelhantes às acima exemplificadas são equivocadas, pois, muito embora, tudo o que trate de coleta, processamento, interpretação e apresentação de dados pertença ao domínio da Estatística, a existência de uma destas atividades isoladamente não se caracteriza como **Estatística**. (FREUND, 2006).

Convém observar que a estatística também não é um ramo da matemática que investiga os processos de obtenção, organização e análise de dados sobre uma determinada população. A estatística também não se limita a um conjunto de elementos numéricos relativos a um fato social, nem a números, tabelas e gráficos usados para o resumo, a organização e a apresentação dos dados de uma pesquisa, embora este seja um aspecto da estatística que pode ser facilmente percebido no cotidiano.

A estatística é uma ciência multidisciplinar, que permite a análise estatística de dados de físicos, economistas, agrônomos, químicos, geólogos, matemáticos, biólogos, sociólogos, psicólogos ou cientistas políticos.

Em síntese, a Estatística (Larson, 2010) consiste em um conjunto de técnicas e métodos de pesquisa que envolve etapas, como, por exemplo, o planejamento do experimento a ser realizado, a coleta qualificada dos dados, a inferência, o processamento e a análise na resposta para um determinado problema.



Assistir aos jornais da TV e ler nas revistas dados, porcentagens, projeções, a bolsa de valores que sobe e desce, reflete apenas o que pesquisas de opinião e previsões dizem. Mas a Estatística vai além, é preciso que esses números sejam confiáveis e tenham garantia de qualidade. É importante que pessoas comuns saibam avaliar dados a fim de que não sejam manipulados por números mascarados e não tomem decisões equivocadas que vão de encontro aos seus próprios interesses. Portanto, conhecer e saber utilizar a estatística significa ter controle de suas decisões.



Saiba o que significa inferência! A **inferência estatística** consiste na generalização sobre a variável aleatória feita a partir de estimativas obtidas por meio de amostras. (MYNBAEV, 2004).



Você também pode conhecer mais sobre inferência estatística no endereço <http://www.inf.ufsc.br/~marcelo/Cap9.pdf>



1.2.4 Aplicações da estatística

A Estatística possui aplicações em diversas áreas.

Você consegue identificá-las? Que tal pensar em algumas de suas aplicações?

Pois bem, são muitas as aplicações do conhecimento de estatística nas mais variadas atividades humanas, das quais relacionamos algumas no quadro abaixo.

Quadro 1 - Exemplos de usos da estatística

Áreas do conhecimento	Aplicações	Exemplo de pesquisas
Indústria	Controle de qualidade	O fabricante deseja verificar o percentual de itens com defeitos em relação ao total fabricado.
Educação	Controle da evasão escolar	Estudo para verificar o percentual de alunos desistentes em relação ao total de matrículas.
Políticas de Governo	Auxiliar nas definições de investimentos	Quando se deseja identificar o quantitativo/percentual de famílias sem moradia e que tenham renda familiar mensal de até 5 salários mínimos, por exemplo.
Gestão empresarial	Qualidade do atendimento	O empresário/investidor deseja identificar o grau de satisfação dos clientes.

Fonte: Elaboração própria.

O quadro acima traz apenas alguns exemplos do imenso campo de utilização da estatística. Convém reforçar que a aplicação da Estatística se dá tanto na pesquisa científica, para a melhoria de recursos econômicos, para o aumento da qualidade e produtividade, como nas questões judiciais, nas previsões e em muitas outras áreas do conhecimento humano.

O estatístico planeja e coordena o levantamento de informações por meio de questionários, entrevistas e medições. Organiza, analisa e interpreta os resultados para explicar fenômenos sociais, econômicos ou naturais. Cabe a ele montar bancos de dados para os mais diversos usos, como controle de qualidade da produção de uma indústria, recenseamento populacional, pesquisa eleitoral ou lançamento de produtos no mercado de consumo. Na indústria, acompanha os testes de qualidade, ajuda a fazer previsão de vendas e desenvolve modelos matemáticos para ajustá-los a situações práticas. Em laboratório, cria tabelas para sistematizar os resultados de experimentos e pesquisas.

Para você que está cursando esta disciplina no Curso Técnico em Finanças, é importante saber que todos nós devemos conhecer e saber fazer uso da estatística, entretanto, é importante lembrar que existe uma profissão específica, a do estatístico.



Figura 9

Fonte: Ilustradora



Caro(a) estudante, que tal refletir um pouco sobre o que vimos até agora?



Vamos à atividade de aprendizagem!

Atividade de aprendizagem



1. Em um texto dissertativo, baseando-se no conteúdo desta aula, descreva o que é estatística e que aplicabilidade tem no ambiente empresarial.

1.3 Os dados

Os dados estatísticos são obtidos por meio de processos que envolvem a observação ou outra mensuração de características de uma determinada população ou amostra, tais como total de filhos por família em uma cidade, gênero dos membros de uma comunidade, estatura dos habitantes de uma cidade, dentre outros.

Os dados são fatos e números coletados, analisados e sintetizados para a compreensão e interpretação de um determinado cenário, causas ou efeitos. Os dados coletados em um estudo são denominados conjunto de dados. A tabela 1 mostra um conjunto de dados contendo informações referentes à educação básica, extraídas do site do IDEB.



Tabela 1 – IDEB 2005, 2007, 2009, 2011 e Projeções do Ensino Médio para o BRASIL

	IDEB Observado				Metas				
	2005	2007	2009	2011	2007	2009	2011	2013	2021
Total	3.4	3.5	3.6	3.7	3.4	3.5	3.7	3.9	5.2
Dependência Administrativa									
Pública	3.1	3.2	3.4	3.4	3.1	3.2	3.4	3.6	4.9
Estadual	3.0	3.2	3.4	3.4	3.1	3.2	3.3	3.6	4.9
Privada	5.6	5.6	5.6	5.7	5.6	5.7	5.8	6.0	7.0

*Os resultados marcados em verde referem-se ao IDEB que atingiu a meta.

Fonte: Saeb e Censo Escolar.

Fonte: INEP (2012)

Há dois tipos de conjuntos de dados utilizados na estatística: a **população** e a **amostra**.

Caracteriza-se população como uma coleção de todos os resultados, de medições ou contagens de uma determinada área de interesse. Por exemplo, habitantes de um país, alunos de uma escola, produção de uma fábrica, entre outros.

Enquanto que **amostra** consiste no subgrupo de uma população.

Na aula 3, trataremos os conceitos de população e amostra mais detalhadamente.

1.4 Elementos e variáveis

Os elementos são entidades ou períodos dos quais os dados são coletados. Para o conjunto de dados na tabela 2, cada mês é um elemento. Como são 4 meses na tabela, logo, existem 4 elementos no conjunto de dados, pois cada intervalo na série temporal é classificado como um elemento.



Não são apenas os intervalos de datas que são elementos; também o são: características como idade, escolaridade; delimitações geográficas, como cidades, estados, regiões.

Na aula 7, no conteúdo de tabelas, mostraremos a classificação e a aplicabilidade de cada um desses tipos de elementos.

Tabela 2 - Contas de microcrédito no Brasil

Data	Contas correntes simplificadas		Contas de poupança simplificadas	
	Quantidade total	Contas ativas	Quantidade total	Contas ativas
jul/12	12.431.587	7.720.594	12.262	11.388
ago/12	12.578.085	7.877.702	12.354	11.374
set/12	12.646.188	7.945.213	12.369	11.271
out/12	12.713.003	8.034.503	12.393	11.220

Observação: série histórica atualizada após retificação dos dados enviados pelas instituições financeiras.
Fonte: Banco Central do Brasil

Uma variável é uma característica de interesse para os elementos. O conjunto de dados na tabela 2 apresenta as 4 seguintes variáveis:

- (i) quantidade total de contas correntes simplificadas;
- (ii) quantidade de contas correntes simplificadas ativas;
- (iii) quantidade total de contas poupança simplificadas; e
- (iv) quantidade de contas poupança simplificadas ativas.

É importante destacar que cada uma dessas características como, quantidade de contas ativas, quantidade de contas total, ou ainda outras características, tais como, gênero, total de filhos por família, estatura, dentre outras, tendem a apresentar certo grau de variabilidade quando se fazem mensurações sucessivas.

Essas características, devido ao seu grau de variabilidade, são denominadas de variáveis. Variável é a palavra usada pela ciência para definir conceitos, propriedades, dimensões e funções de um objeto que está sendo estudado. As variáveis são sempre quantificáveis mesmo que seja em nível dicotômico, como, por exemplo, sim ou não; masculino ou feminino.

As variáveis são, na realidade, valores, fatos ou fenômenos que, numa hipótese, são considerados em sua dimensão de inter-relação causal, de modo que um ou mais deles sejam determinados como causa, e outros, como efeitos. (GRESSLER, 2004).

A-Z

Dicotômico

quando a variável resposta assume apenas dois valores possíveis. Ex. sim/não.



Figura 10

Fonte: Ilustradora

Claramente tais variáveis têm naturezas diferentes no que tange aos possíveis valores que podem assumir. Tal fato deve ser levado em conta nas análises dos dados, pois para cada tipo de variável existe um tratamento diferente.

Observe este exemplo: Suponha que um questionário foi aplicado aos alunos do Curso de Finanças do IFRO fornecendo as seguintes informações:

- 1) Identificação do aluno
- 2) Gênero: F se feminino, M se masculino.
- 4) Altura: Altura em metros
- 6) Filhos: Número de filhos na família

Quadro 2 - Tipos de variáveis e resultados possíveis

Variável	Resultados possíveis
Gênero	Masculino, feminino
Altura	..., 160cm, 161cm, 162cm,...
Nº de filhos na família	0, 1, 2, 3,...

Fonte: Elaboração própria.

Vamos analisar o quadro a seguir?

No caso da variável gênero, temos apenas duas opções de respostas, masculino ou feminino, e trata-se de uma resposta não numérica, de natureza dicotômica (que possui duas opções mutuamente excludentes) sobre uma característica do indivíduo.

A variável filhos, por se tratar de resultados de um processo de contagem, tem como opções de respostas números inteiros positivos. Já a variável estatura, apesar de também ter opções de respostas numéricas, estas não necessariamente serão números inteiros como respostas, apresentando-se assim uma continuidade.

Mas como estas variáveis são, então, classificadas?

Vamos ver!

As variáveis são classificadas em variáveis quantitativas (contínuas ou discretas) e variáveis qualitativas (nominais ou ordinais), conforme exposto na figura 11.

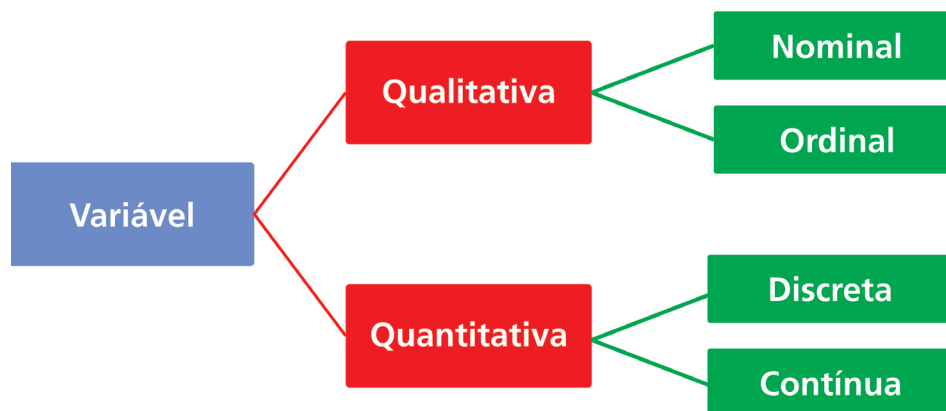


Figura 11 - Classificação das variáveis

Fonte: Elaboração própria

As variáveis quantitativas são do tipo numéricas, ou seja, expressas em número e podem ser subdivididas em discretas e contínuas. As variáveis quantitativas discretas são resultantes de contagens e, portanto, assumem valores de números inteiros. Já as variáveis contínuas são resultantes de mensuração e podem assumir qualquer valor entre dois números reais.

As variáveis qualitativas não são numéricas, uma vez que seus valores representam atributos ou quantidades. Tais variáveis também são subdivididas em duas subcategorias, chamadas de nominais e ordinais. As variáveis classificadas como qualitativas ordinais apresentam uma ordenação natural, indicando intensidades crescentes de realização. Ao passo que as variáveis qualitativas nominais definem apenas categorias, entretanto, não é possível



estabelecer uma ordem natural entre seus valores.

O quadro 3 sintetiza as variáveis em seus quatro níveis e traz alguns exemplos de sua aplicação.

Quadro 3 - Resumo dos quatro níveis de mensuração de uma variável		
Variável	Exemplo de conjunto de dados	Cálculos significativos
Nível nominal (dados qualitativos)	Preferência musical dos ouvintes de uma rádio Rock MPB Sertaneja Pagode	Coloque em uma categoria. Por exemplo, a preferência musical dos ouvintes de uma rádio poderia ser colocada em quatro categorias.
Nível ordinal (dados qualitativos ou quantitativos)	Ano Escolar de alunos do nível médio 1º Ano 2º Ano 3º Ano	Coloque em uma categoria que ordene. Por exemplo, para cursar o 3º ano é necessário que tenha sido aprovado no 2º. Para cursar o 2º ano, deve ter sido aprovado no 1º ano.
Nível discreto (dados quantitativos)	Número de alunos matriculados Sala A – 40 Sala B – 38 Sala C – 39	Coloque em uma categoria, ordene e encontre a diferença entre os valores. Por exemplo, $40 - 38 = 2$, logo a sala A tem 2 alunos a mais que a sala B
Nível contínuo (dados quantitativos)	Consumo de combustível por km. Veículo litros/km Carro A – $1/10 = 0,1$ Carro B – $1/8 = 0,13$ Carro C – $1/5 = 0,2$	Coloque em uma categoria, ordene, encontre a diferença e a razão entre os valores. Por exemplo, $0,2 / 0,1 = 2$, então o carro C consome duas vezes mais combustível que o carro A.

Fonte: Elaboração própria.

Resumo

Você estudou nesta aula, que a Estatística faz parte do grupo das ciências “cujos primeiros passos remontam aos primórdios da história da humanidade e cujo desenvolvimento formal tende a estar em sintonia com a evolução do conhecimento humano”

A estatística pode ser ainda caracterizada como a área da Matemática que coleta, analisa e interpreta dados numéricos para o estudo de fenômenos naturais, econômicos e sociais, e tal como as demais áreas do conhecimento científico, consiste em uma ciência que está sempre absorvendo novas técnicas e contribuições de outras ciências, como novas descobertas e novas teorias.

A estatística tem aplicabilidade nas diversas atividades humanas, fenômenos da natureza e áreas do conhecimento, como, por exemplo, na indústria, na saúde, nas políticas de governo, na mensuração de chuvas e temperatura, e

na pesquisa científica.

Vimos que dados estatísticos são obtidos por meio de processos que envolvem a observação ou outra mensuração de características de uma determinada população ou amostra, tais como total de filhos por família em cidade, gênero dos membros de uma comunidade, estatura dos habitantes de uma cidade, entre outras. São características de um fenômeno ou evento, e que são coletadas e processadas por meios de elementos estatísticos, ou seja, das características dos dados observados e de variáveis, que consistem no aspecto de variação dos elementos observados.

Vamos praticar o que acabamos de apresentar?

Atividade de aprendizagem

2. Dada a tabela abaixo, relacione as colunas com os conceitos adequados.



Tabela 3 - Perfil dos Alunos

Turma	Estatura	Nº de Irmãos	Escolaridade
Turma A	1,65	3	1º Ensino Médio
Turma B	1,70	2	2º Ensino Médio
Turma C	1,72	2	3º Ensino Médio

Fonte: Elaborada pelo autor (dados fictícios)

- | | |
|-------------------|------------------------------------|
| a) Turma A, B e C | () Variável quantitativa discreta |
| b) Estatura | () Variável quantitativa contínua |
| c) Nº de irmãos | () Variável qualitativa ordinária |
| d) Escolaridade | () Elementos estatísticos |

Caro(a) estudante

Chegamos ao final da primeira aula desta disciplina. Nela, abordamos o tema introdução à estatística, um pouco da sua história, em que consistem a estatística, os dados e as variáveis.

Na próxima aula, veremos alguns conceitos básicos e elementares da matemática utilizados na estatística! Continue atento(a). Ainda há muito para aprender.



Aula 2 - Conceitos Básicos e Elementares de Matemática

Objetivos:

- Reconhecer e exemplificar operações que envolvam porcentagem.
- Identificar a diferença entre dados absolutos e relativos
- Reconhecer e utilizar as representações de dados relativos

Prezado(a) estudante,

Esta é a segunda aula de Estatística Aplicada e o assunto a ser tratado são os conceitos básicos de matemática, tais como porcentagem, dados absolutos e relativos, necessários para o bom desenvolvimento desta disciplina. O aprendizado do conteúdo desta aula auxiliará em muito na compreensão e no desenvolvimento das próximas aulas.

Esperamos que você compreenda bem estes conceitos e que os mesmos sirvam como base para o seu desenvolvimento profissional. Boa aula!!!

2.1 Porcentagem

A porcentagem ou percentagem é uma medida de razão com base 100 (cem). É um modo de expressar uma proporção ou uma relação entre 2 (dois) valores (um é a parte e o outro é o inteiro) a partir de uma fração cujo denominador é 100 (cem), ou seja, é dividir um número por 100 (cem).

Dizer que algo, por exemplo, o número de faltas de um aluno x é de “20%” na disciplina de estatística aplicada (lê-se: “o aluno x faltou em vinte por cento das aulas”), significa dizer que as faltas equivalem a 20 elementos em um conjunto universo de 100 elementos, ou seja, que a razão é a divisão.

$$\frac{70}{100} = 0,7$$



Vamos imaginar que um aluno tivesse acertado 9 de 12 questões de uma prova.

Temos que a razão entre o número de questões acertadas e o número total de questões é de:

$$\frac{9}{12} = \frac{3}{4} = 0,75$$

Portanto, o aluno acertou 0,75 da prova. Mas quanto isso representa em termos percentuais?

É simples descobrir. Um número percentual significa que tem denominador igual a 100. E para incluí-lo, basta que multipliquemos 0,75 por $\frac{100}{100}$. A multiplicação com numerador e denominador iguais não altera o resultado, uma vez que $\frac{100}{100} = 1$.

$$0,75 * \frac{100}{100} = \frac{75}{100}$$

Uma nomenclatura muito utilizada para representar os números percentuais é substituir o 100 pelo % (por cento).

$$\frac{75}{100} = 75\%$$

Agora, vamos aos elementos dos cálculos percentuais.

$$\frac{\text{porcentagem}}{\text{principal}} = \frac{\text{taxa}}{100}$$

Porcentagem, denotada por p (p minúsculo), é o valor que representa a quantidade de unidades tomadas de outra, proporcionalmente a uma taxa.

Taxa, representada por r, é o valor que representa a quantidade de unidades tomadas de outra, em cada 100.

Principal, denotada por P (P maiúsculo), é o valor da grandeza em que se calcula a porcentagem.

Assim, podemos representá-la por:

$$\frac{p}{P} = \frac{r}{100}$$



Vamos aplicar ao nosso exemplo anterior?

$$\frac{p}{P} = \frac{r}{100}$$

$$\frac{9}{12} = \frac{75}{100} \text{ ou } 75\%$$

No exemplo dado, o número de acertos do aluno representa a percentagem (9), o total de questões (12) é o principal e 75 é a taxa.

Taxa unitária

Como vimos, a taxa percentual (% por cento) refere-se a uma taxa por 100, ou dividida por 100, $\frac{x}{100}$.

Na resolução de muitos problemas em nosso dia a dia, é muito mais prático utilizarmos a taxa unitária. Como o próprio nome já diz, isso significa que a taxa é proporcional a unidade ou a 1.

Pelo exemplo anterior, temos que,

$$\frac{75}{100} = 0,75$$

Sendo que a 0,75 chamamos de taxa unitária.

Como fica na fórmula?

Se a taxa unitária (i) é dada por

$$\frac{r}{100} = i$$

Então, podemos afirmar que,

$$\frac{p}{P} = i$$

Ou seja,

$$\frac{p}{P} = i$$

Vamos exemplificar!



A turma em que estudo é de 40 alunos. Sabendo que 22 são do gênero masculino, vamos verificar em termos percentuais e em taxa unitária o quanto isso representa.

Pela taxa percentual:

Os alunos do gênero masculino representam a percentagem. O total de alunos, o principal, e queremos descobrir a taxa.

$$\frac{p}{P} = \frac{r}{100}$$

$$\frac{22}{40} = \frac{r}{100}$$

Fazendo uma regra de três simples, nós temos,

$$r = \frac{22}{40} * 100$$

$$r = 0,55 * 100$$

$$\mathbf{r = 55\%}$$

Pela taxa unitária:

$$i = \frac{p}{P}$$

$$i = \frac{22}{40}$$

$$\mathbf{i = 0,55}$$

Como você observou, a taxa percentual é de 55% e a taxa unitária é de 0,55. Caso queira representar a taxa unitária em termos percentuais, basta multiplicá-la por 100, conforme mostraremos a seguir.

$$i = 0,55 * 100$$

$$\mathbf{i = 55\%}$$

Lembre-se de que, ao transformar uma taxa em percentual, devemos inserir o símbolo %, para indicar que a taxa é percentual.



2.2 Dados absolutos e dados relativos

Na seção anterior, conhecemos e trabalhamos com operações que envolvem porcentagem. Nesta seção, vamos aprender a identificar e resolver operações com dados relativos.

Um conjunto de dados pode ser classificado como absoluto ou relativo.

Os **dados absolutos** são aqueles que ainda não sofreram alterações, apresentando os dados reais, tais como foram coletados.

Já os **dados relativos** são aqueles que sofreram alterações e são geralmente apresentados em percentuais, índices, coeficientes e taxas.

Vamos detalhar cada um destes conceitos aplicados aos dados relativos.

Porcentagens

Como vimos na seção anterior, a porcentagem ou porcentagem é uma medida de razão com base 100 (cem). É um modo de expressar uma proporção ou uma relação entre 2 (dois) valores (um é a parte e o outro é o inteiro) a partir de uma fração cujo denominador é 100 (cem), ou seja, é dividir um número por 100 (cem).

Índices

Os índices são razões entre duas grandezas tais que uma não inclui a outra.

São exemplos de índices:

$$\text{renda per capita} = \frac{\text{renda}}{\text{população}}$$

$$\text{receita per capita} = \frac{\text{receita}}{\text{população}}$$

$$\text{consumo per capita} = \frac{\text{consumo}}{\text{população}}$$

$$\text{renda per capita} = \frac{\text{renda}}{\text{população}}$$

$$\text{densidade geográfica} = \frac{\text{população}}{\text{superfície}}$$

$$\text{quociente eleitoral} = \frac{\text{número de votos válidos}}{\text{número de cadeiras preenchidas}}$$



Como calcular a Renda Familiar Per Capita

Observe este exemplo prático de utilização de índice na definição de benefícios sociais do Governo Federal.

Para verificar se a família do idoso ou da pessoa com deficiência recebe menos de $\frac{1}{4}$ de salário mínimo por pessoa, ou seja, se a renda mensal familiar per capita é inferior a $\frac{1}{4}$ de salário mínimo, devem ser somados todos os rendimentos recebidos no mês por aqueles que compõem a família, compreendendo o (a) requerente (idoso ou pessoa com deficiência); o (a) cônjuge ou companheiro (a); os pais e, na ausência deles, a madrasta ou o padrasto; irmãos(as) solteiros(as); filhos (as) e enteados (as) solteiros(as) e os(as) menores tutelados (as).

O valor total dos rendimentos, chamado de renda bruta familiar, deve ser dividido pelo número dos integrantes da família. Se o valor final for menor do que $\frac{1}{4}$ do salário mínimo, o(a) requerente poderá receber o BPC, desde que cumpridos todos os demais critérios.

Os rendimentos que entram no cálculo da renda bruta mensal são aqueles provenientes de: salários; proventos; pensões; pensões alimentícias; benefícios de previdência pública ou privada; seguro desemprego; comissões; pró-labore; outros rendimentos do trabalho não assalariado; rendimentos do mercado informal ou autônomo; rendimentos auferidos do patrimônio; Renda Mensal Vitalícia – RMV, e o Benefício de Prestação Continuada da Assistência Social - BPC.
<http://www.mds.gov.br/assistenciasocial/beneficiosassistenciais/bpc/como-calculer-a-renda-familiar-per-capita>



Coeficientes

Os coeficientes são razões entre o número de ocorrências e o número total, que compreende as ocorrências e não ocorrências, tais como:

$$\text{coeficiente de mortalidade} = \frac{\text{número de óbitos}}{\text{população local}}$$

$$\text{coeficiente de evasão escolar} = \frac{\text{número de alunos evadidos}}{\text{número inicial de matrículas}}$$

$$\text{coeficiente de natalidade} = \frac{\text{número de nascimentos}}{\text{população local}}$$

Taxas

As taxas são dadas pelo próprio coeficiente, entretanto, o que a diferencia é que esta é multiplicada por 10, 100, 1000, 10.000 ou 100.000, para tornar o resultado mais compreensivo.

Tomando como exemplo os coeficientes mostrados acima, temos:

$$\text{Taxa de mortalidade} = \text{coeficiente de mortalidade} \times 100.000$$

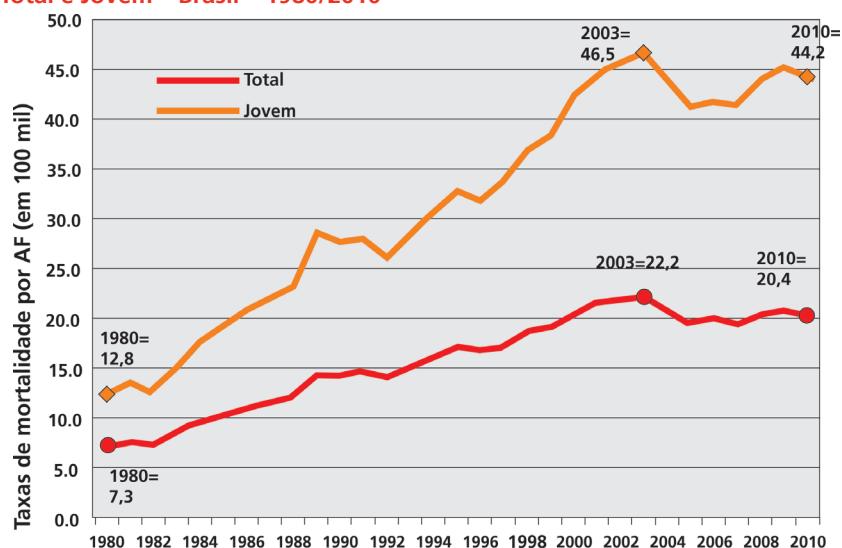
$$\text{Taxa de evasão escolar} = \text{coeficiente de evasão escolar} \times 100$$

$$\text{Taxa de natalidade} = \text{coeficiente de natalidade} \times 1000$$

As taxas são comumente utilizadas na representação de resultados, devido à forma simplificada de representação. Por exemplo, o gráfico 1 apresenta, de acordo com Waiselfisz (2013), as taxas de mortalidade (em cada 100 mil habitantes) por arma de fogo no período de 1980 a 2010.

Observe que, entre os jovens, a taxa em 1980 era de 12,8, o que significa que 12,8 a cada 100.000 jovens no Brasil, morreram vítimas de armas de fogo no ano de 1980. Este número subiu para 44,2 em 2010, o que é muito alarmante. Um grande problema que depende do empenho tanto das autoridades como da população para a conscientização da chaga da violência brasileira, principalmente entre jovens.

Gráfico 1 – Taxas de mortalidade (em 100 mil habitantes) por armas de fogo. População Total e Jovem – Brasil – 1980/2010



Fonte: WAISELFISZ (2013)

Com certeza, você está se perguntando como calcular a taxa. Vamos exemplificar a partir dos dados da tabela 4.

Tabela 4 - Número de vítimas letais por armas de fogo na população total e na população jovem segundo causa básica. Brasil. 2010

POPULAÇÃO TOTAL					15 A 29 ANOS				
Acidente	Suicídio	Homicídio	Indeterminado	Total arma de fogo	Acidente	Suicídio	Homicídio	Indeterminado	Total arma de fogo
352	969	36792	779	38892	152	299	21843	400	22694

Fonte: WAISELFISZ (2013)

Para calcular a taxa de mortalidade da população brasileira por arma de fogo no ano de 2010, consideramos a população total de 190.755.799, conforme dados extraídos do Censo 2010. (IBGE, 2010).

Quadro 4 – Cálculo da taxa de mortalidade da população brasileira, por arma de fogo em 2010

$$\begin{aligned}
 \text{Taxa de mortalidade da população brasileira por arma de fogo em 2010} &= \frac{\text{número de óbitos}}{\text{população local}} \times 1000 \\
 &= \frac{38892}{190.755.799} \times 1000 \\
 &= 0,000204 \times 1000 \\
 &= 20,4
 \end{aligned}$$

Fonte: Elaboração própria

A Escola Novo Amanhã, no ano de 2012, teve 300 alunos matriculados. No final do ano letivo, verificou-se que 45 alunos haviam desistido. Quais foram



o coeficiente e a taxa de evasão escolar observados?

Coeficiente:

$$\text{coeficiente de evasão escolar} = \frac{\text{número de alunos evadidos}}{\text{número inicial de matrículas}}$$

$$\text{coeficiente de evasão escolar} = \frac{45}{300} = 0,15$$

Taxa:

$$\text{Taxa de evasão escolar} = \text{coeficiente de evasão escolar} \times 100$$

$$\text{Taxa de evasão escolar} = 0,15 \times 100$$

$$\text{Taxa de evasão escolar} = 15\%$$

Portanto, o coeficiente e a taxa de evasão escolar na Escola Novo Amanhã são de 0,15 e 15%, respectivamente.

Resumo

Nesta aula, você pôde verificar que a porcentagem é uma medida de razão com base 100. É um modo de expressar uma proporção ou uma relação entre 2 (dois) valores (um é a parte e o outro é o inteiro) a partir de uma fração cujo denominador é 100 (cem), ou seja, é dividir um número por 100 (cem).

Pôde também aprender que um conjunto de dados pode ser classificado como absoluto ou relativo. Os dados absolutos são aqueles que ainda não sofreram alterações, apresentando os dados reais tais como foram coletados. Já os dados relativos são aqueles que sofreram alterações e são geralmente apresentados em percentuais, índices, coeficientes e taxas.

Demonstramos que coeficientes são razões entre o número de ocorrências e o número total, que compreende as ocorrências e não ocorrências.

Por último, mostramos que as taxas são dadas pelo próprio coeficiente, entretanto, o que as diferencia é que estas são multiplicadas por 10, 100 ou 1000, para tornar o resultado mais compreensivo.



Atividade de aprendizagem

1. Brasil Carinhoso pode baixar pobreza extrema infantil a 0,6%

Trecho de matéria extraída do site do IPEA:



O Programa Brasil Carinhoso tem a capacidade de reduzir a pobreza extrema entre crianças de 0 a 15 anos a um patamar residual, segundo estudo lançado hoje pelo Instituto de Pesquisa Econômica Aplicada (**IPEA**). A *Nota Técnica nº 14* da Diretoria de Estudos e Políticas Sociais (DISOC), intitulada *O Bolsa Família depois do Brasil Carinhoso: uma análise do potencial de redução da pobreza extrema* revela que, se o desenho atual do programa tivesse sido implementado em 2011, a taxa de pobreza extrema entre a população de 0 a 15 poderia ter caído para apenas 0,6%. (IPEA, 2012).

Com base no texto acima, responda o seguinte:

- a)** Indique a informação em que o autor utilizou o conceito de porcentagem para apresentá-la no texto.
- b)** Construa um modelo em que seja possível calcular o coeficiente de pobreza de um país. (Observação: não é preciso calcular).

Caro(a) estudante

Chegamos ao final da segunda aula sobre Estatística Aplicada. Nela aula, os temas estudados foram porcentagem, dados absolutos e relativos.

Na próxima aula, abordaremos os temas: método estatístico, população e amostra. Continue disciplinado(a) em seus estudos e não deixe de realizar as atividades de aprendizagem!



Aula 3 - Método Estatístico, População e Amostra

Objetivos:

- Reconhecer a diferença entre a estatística descritiva e estatística inferencial.
- Distinguir e exemplificar as fases do método estatístico.
- Expressar e exemplificar a diferença entre população e amostra.
- Identificar e utilizar a amostragem estatística.

Prezado(a) estudante.

Esta é a terceira aula da disciplina Estatística Aplicada, e os assuntos a serem tratados são o método estatístico e suas fases, população e amostra e o procedimento de amostragem. Esperamos que você compreenda bem este conteúdo, pois o mesmo é importante para o entendimento dos assuntos a serem expostos nas próximas aulas. Leia com atenção e não deixe de realizar as atividades de aprendizagem. Boa aula!!!

Acredito que você se recorda que, nas primeiras aulas, nós vimos:

- Que a Estatística é uma técnica de uso milenar;
- A diferenciação entre a estatística e a apresentação de resultados; e
- A classificação de dados e suas variáveis.

Pois bem, nesta aula, veremos:

- Que a estatística é subdividida em dois ramos: a estatística descritiva e a estatística inferencial;
- As fases de aplicação do método estatístico;
- Conceito de população e amostra;
- A utilização de amostragem em pesquisas estatísticas.



Preparado(a) para recomeçar? Vamos iniciar pelo método estatístico e suas fases.

3.1 O método estatístico e suas fases

Conforme Martins (2007), o método consiste no conjunto de meios dispostos convenientemente para observar fenômenos e inferir conclusões a partir de tais observações. Os métodos podem ser classificados em experimentais ou estatísticos: o método experimental, que consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que se possa descobrir seus efeitos. Já o método estatístico admite todas as causas presentes, variando-as, dada a impossibilidade de manter as causas constantes, registrando estas variações e procurando determinar que influências cabem a cada uma delas.

Neste estudo, nós nos concentraremos apenas no método estatístico.

O objetivo de todo método estatístico é coletar dados e, então, usá-los para tomar uma decisão. Larson (2010, p. 15) observa que “qualquer decisão que seja tomada usando os resultados de um estudo estatístico será tão boa quanto o processo utilizado para a obtenção desses dados.” Daí a importância de conhecer bem as técnicas estatísticas, bem como realizar o adequado planejamento de todas as etapas, pois “se o processo tiver falhas, então a decisão resultante será questionável”.

A pesquisa estatística, como qualquer outro estudo, requer um adequado planejamento de todas as suas etapas.

O método estatístico abrange as seguintes fases, conforme ilustra a próxima figura:



Conheça dois Institutos que produzem estatísticas no Brasil: IBGE e IBOPE.

O Instituto Brasileiro de Geografia e Estatística – IBGE – constitui-se no principal provedor de dados e informações do país, que atende às necessidades dos mais diversos segmentos da sociedade civil, bem como dos órgãos das esferas governamentais federal, estadual e municipal. O IBGE oferece uma visão completa e atual do País, através do desempenho de suas principais funções, vinculadas à produção, análise e difusão de informações estatísticas.

O IBOPE é a maior empresa privada de pesquisa da América Latina e a 12ª maior do mundo.

No Brasil, adquiriu grande notoriedade pela medição de audiência de TV e pelas pesquisas eleitorais e de opinião pública, contribuindo com o amadurecimento da democracia e dos mercados que atende.

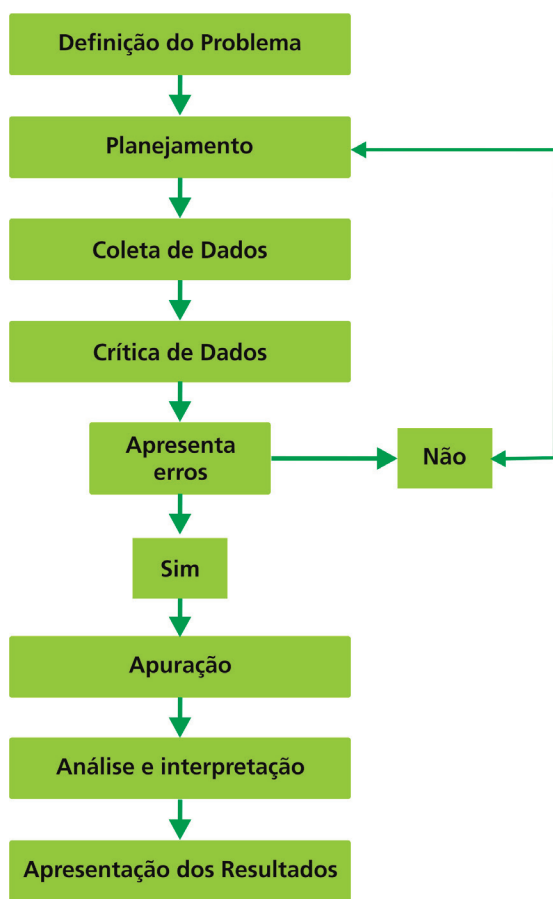


Figura 12 - Fases do método estatístico

Fonte: Elaborado pelo autor

a) Definição do Problema

A fase de definição do problema consiste no levantamento do que se quer pesquisar, identificando exatamente o que se pretende pesquisar, definindo o problema corretamente (variáveis, população, hipóteses etc.). Faz parte também desta fase a definição dos objetivos de realização do estudo.

Durante a definição do problema, é importante o(a) pesquisador(a) examinar outros levantamentos realizados no mesmo campo ou semelhantes, a fim de facilitar o trabalho e diminuir os possíveis erros estatísticos, provenientes de vícios na coleta de dados, é recomendável examinar outros levantamentos realizados no mesmo campo, por meio de uma revisão da literatura.

b) Planejamento

Nesta fase, deve-se determinar o procedimento necessário para resolver o problema, em especial, como levantar informações sobre o assunto ou ob-



No campo da estatística, um **pré-teste** é a aplicação de um questionário, na sua versão preliminar, a uma amostra de indivíduos, com o objetivo de identificar perguntas-problema que justifiquem uma modificação da redação, alteração do formato ou mesmo sua eliminação da versão final.



jeto de estudo. O(a) pesquisador(a) deve ainda escolher e formular as perguntas, e realizar um pré-teste do questionário a fim de certificar-se de que quem irá responder será capaz de compreender as questões e de que as respostas atenderão ao problema proposto.

A realização de um pré-teste é muito importante, pois este possibilitará verificar se as perguntas estão claras, se o vocabulário utilizado está acessível e se não gera ambiguidade (dupla interpretação).

Este procedimento, quando realizado durante o planejamento de uma pesquisa, evita a possível perda parcial ou total de dados, ocasionada pela descoberta de erros durante ou após a realização da pesquisa de campo.

Ainda nesta fase, é importante a definição de qual tipo de levantamento será realizado (censo ou amostragem), levando em consideração tempo e recursos disponíveis.

c) Coleta ou levantamento dos dados

A fase de coleta de dados é a etapa em que a pesquisa efetivamente é realizada, a fase de campo. Antes de iniciar esta etapa é imprescindível que o(a) pesquisador(a) tenha certeza de que todas as etapas anteriores já foram cumpridas, e entre elas, merece especial atenção a certificação de que existe coerência entre o problema de estudo proposto e o questionário.

A coleta pode ser: direta - diretamente da fonte de dados, ou indireta - feita através de outras fontes.

Os dados podem ser obtidos pela própria pessoa (primários) ou baseiam-se no registro de terceiros (secundários).

d) Crítica dos dados

Nesta fase, é realizada a observação dos dados à procura de falhas e imperfeições, visando eliminar erros grosseiros. Cabe observar que existem basicamente dois grupos de falhas na pesquisa amostral: as de natureza externa ou interna.

São classificadas como falhas:



- **De natureza externa:** oriundas de erros por parte do informante, do respondente da pesquisa;
- **De natureza interna:** ocorrem devido a falhas no instrumento de coleta ou ainda da abordagem do entrevistador.

Observa-se que, com um bom planejamento, elaboração e revisão cuidadosa do instrumento de pesquisa (questionário) e correta definição da amostra, o pesquisador reduz, e muito, os riscos de falhas durante a realização da pesquisa. Caso nesta fase sejam verificadas falhas, é recomendável que se retorne à etapa de planejamento.

e) Apuração dos Dados

A fase de apuração consiste em resumir os dados, através de contagem e agrupamento. A apuração pode ser manual ou mecânica, entretanto, os grandes institutos de pesquisa como o IBGE e o IBOPE, entre outros, utilizam-se de recursos tecnológicos (Palmtops, notebook, tablet etc.) que permitem a apuração de forma automatizada, para que haja pouca interferência humana na consolidação dos resultados. Há aqueles casos ainda, em que respondemos a questionários *on line*, cuja apuração de resultados se dá de forma automática.

f) Análise e interpretação dos dados

É a fase mais importante e também a mais delicada, pois é o momento em que se encontram conclusões que auxiliam o(a) pesquisador(a) a resolver seu problema, o que nem sempre é tão simples de fazer. Porém, um bom planejamento e conhecimento das técnicas da estatística, conforme veremos nos capítulos seguintes, o auxiliarão muito na realização desta fase.

g) Apresentação dos resultados

É a fase em que vamos mostrar os resultados obtidos na coleta e na organização dos dados. A apresentação dos resultados é realizada tanto com a utilização de tabelas, consistindo em uma apresentação mais numérica, como por meio de gráficos. Nas aulas seguintes, detalharemos este assunto.

Vamos observar a aplicação do método estatístico a uma situação real?

Imagine que queiramos vender algum produto alimentício para a turma de seu curso. Mas antes de realizar o investimento, desejamos saber o que a



turma gostaria de comer. Então vamos ao método!

Veja o exemplo das ações em cada uma das fases.

Passo I - Definição do Problema

O que queremos saber?	Que tipo de alimentos a maioria das pessoas deseja consumir.
De quais variáveis precisarei?	Doces, tipos de doces, salgados assados, salgados fritos, sucos, sabor dos sucos etc.
Qual a minha população?	A turma de meu curso. Como exemplo, vamos considerar que serão 50 alunos.
Verificar se há outros estudos.	Caso haja outros estudos anteriores sobre o assunto: qual foi o resultado? Posso comparar com o estudo que realizarei?

Passo II - Planejamento

Como conseguirei o resultado que desejo?	Por meio de entrevista? Por aplicação de questionário? Caso exista outro estudo idêntico, posso retirar os dados deste para serem analisados conforme a minha necessidade?
Na fase anterior, definimos as variáveis. Mas como perguntarei?	Exemplos de perguntas: Na escola você prefere comer doces ou salgados? Você prefere salgados assados ou fritos?
Pré-teste	Depois de elaborar todas as perguntas é imprescindível verificar se as perguntas não são tendenciosas ou difíceis de serem compreendidas. Por isso, nesta fase, você deve pedir para algumas pessoas responderem ao questionário e comentar o que entenderam. Verifique se o que elas entenderam e responderam é justamente o que você gostaria de perguntar. Caso a resposta a esta pergunta seja negativa, é recomendável que o questionário seja reelaborado. No caso de nosso exemplo, durante o pré-teste, podemos aplicar o questionário com 5 alunos.
Definição do número de pessoas.	Você ouvirá todos os alunos, ou somente parte deles. Se somente parte deles, quantos serão e qual será o critério de escolha?
Quando será aplicada a pesquisa?	É hora de definir o cronograma de trabalho.
Se você precisar de material, recursos ou ajuda de outras pessoas.	Nesta fase é importante definir quem poderá lhe ajudar, de que materiais e de quanto recurso você precisará.

Passo III - Coleta ou levantamento dos dados

Definidas as fases anteriores, nesta fase, faremos a coleta de dados.	Aqui faremos a coleta de dados, seja por meio de entrevista, questionário, observação ou ainda pela análise de estudos anteriores.
---	--



Passo IV - Crítica dos dados

Já temos todas as respostas. E agora?

É importante conferir se todos os questionários estão devidamente respondidos. É recomendável que os questionários incompletos sejam retirados do estudo.

As pessoas entenderam o que você gostaria de saber?

Caso as perguntas não apresentem respostas significativas, é bom reavaliar as perguntas e, se for o caso, fazer nova coleta de dados.

Passo V - Apuração dos Dados

Até agora você terá um conjunto de questionários respondidos. Mas, nesta fase, você deve organizá-los estatisticamente.

Nesta fase, é feita a tabulação dos resultados em distribuições de frequências.

Além disso, nesta fase, aplicam-se também os procedimentos de análise estatística, tais como as medidas de posição e dispersão.

Conheceremos estas medidas nas próximas aulas.

Passo VI - Análise e interpretação dos dados

Análise e interpretação

Após a apuração dos resultados, é preciso saber o que eles significam.

Ilustrativamente, vamos imaginar que entrevistamos 10 pessoas. Dentre estas, 4 responderam que desejam comer doces, e 6 desejam salgados. Dos que responderam que desejam salgados, 4 preferem salgados assados.

Isso significa que 40% dos estudantes preferem doces, 40% salgados assados e 20% salgados fritos. Com base nestas perguntas, você poderá definir quais tipos de alimentos irá vender. Percebeu isso?

Passo VII - Apresentação dos dados

Apresentação dos resultados

Na fase anterior, realizou-se a análise dos dados. Nesta fase, são geralmente construídos gráficos e tabelas que sejam capazes de demonstrar os resultados encontrados.

3.2 População e amostra

Um dos conhecimentos mais relevantes no universo da estatística é a distinção entre os dois tipos de conjuntos de dados: população e amostra.

Uma população, na estatística, pode ser classificada como finita ou infinita conforme o seu tamanho. Convém salientar que, quando a observação abrange uma população muito grande, isso implica dificuldades ou impossibilidade de observar as características em todos os elementos. Como alternativa, nesses casos, é possível realizar o estudo a partir de uma parte representativa da população. Esta parte representativa e finita da população é denominada de amostra.

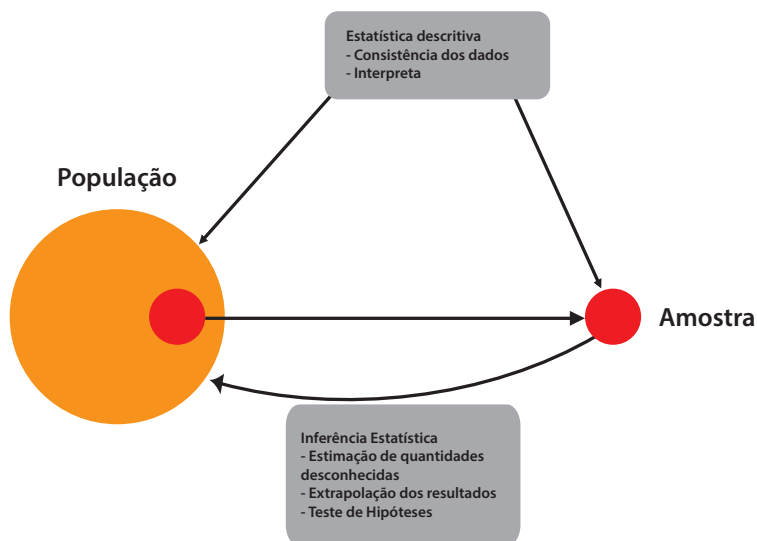


Figura 13 – Etapas da análise estatística

Fonte: ilustradora

População – uma cidade



Amostra – parte dos moradores

É importante ressaltar que o termo população, utilizado em estatística, não possui o mesmo significado aplicado à geografia.



População é o conjunto de todos os resultados de interesse, enquanto amostra é um subgrupo desse conjunto maior. Amostra é um subconjunto da população.

Observe como nos explica um estudioso do assunto, “Se um conjunto de dados consiste em todas as observações concebíveis (ou hipoteticamente) possíveis de um dado fenômeno, dizemos que é uma **população**; se um conjunto de dados consiste em apenas uma parte de uma população, dizemos que é uma **amostra**” (FREUND, 2006, p. 57).

Originalmente, a Estatística tratava apenas da descrição de populações humanas, resultados de censos e semelhantes. Mas, à medida que seus objetivos se ampliaram, o termo “população” passou a ter a conotação muito mais ampla dada a ele na distinção precedente entre população e amostras. (FREUND, 2006). A palavra “população” na estatística constitui, portanto, um termo técnico com significação própria.

Vale, porém, destacar que um conjunto de dados pode ser considerado como amostra ou população, dependendo da ênfase do estudo. Uma turma do Curso Técnico em Finanças, por exemplo, pode ser considerada como população, da qual se podem tirar diversas amostras. Por outro lado, se quisermos observar todos os alunos do Instituto Federal, esta mesma turma seria então apenas uma amostra da escola.





A contagem de uma população inteira é chamada de **censo**. A realização de um censo fornece informações completas, mas ela é, frequentemente, cara e difícil de realizar.

Uma amostragem é uma contagem ou medição de parte de uma população e é mais comumente usada nos estudos estatísticos. É importante observar que, para coletar dados imparciais, o pesquisador deve ter a certeza de que a amostra representa a população. Outro aspecto importante é o de que, mesmo com os melhores métodos, pode ocorrer um erro de amostragem. (LARSON, 2010).

O erro de amostragem consiste na diferença entre os resultados da amostra e os da população. Um exemplo bastante conhecido deste tipo de erro de amostragem são as frequentes divergências entre as pesquisas de intenções de votos e os resultados das eleições, nas quais em muitos casos, candidatos estatisticamente considerados eleitos, perdem a eleição.

Porque fazer amostragem ao invés de censo?

Economia.

Menor tempo.

Maior qualidade nos dados levantados.

População infinita.

Mais fácil, com resultados satisfatórios.

Quando fazer censo?

População pequena (tamanho da amostra grande em relação ao da população).

Quando se exige o resultado exato.

Quando já se dispõe dos dados da População.

3.2.1 Tabela determinante do tamanho da amostra

Uma das etapas mais importantes da pesquisa estatística consiste na deter-



minação do tamanho da amostra. Uma amostra muito pequena pode não ser representativa o suficiente para se tirar conclusões. Por outro lado, se o tamanho da amostra for maior do que o necessário, você estará desperdiçando dinheiro e tempo, não é verdade?

Pois bem, a tabela abaixo indica três níveis de erro amostral: 3%, 5% e 10%. Cada um deles está subdividido em dois níveis de split diferentes.

A-Z

Split

O split na tabela de amostragem demonstra o nível de variação das respostas na pesquisa, isto é, o grau de homogeneidade da população



Cabe observar que uma população mais homogênea corresponde a uma população que possua características semelhantes, tais como: mesmo nível de renda, idade, sexo etc. Para uma população mais homogênea (com respostas similares), aplica-se um split de 80/20, que indica uma menor variação nas respostas (população mais homogênea).

Por outro lado, em populações em que as respostas têm maior variação, utiliza-se o split de 50/50, que indica muita variação entre as respostas dos entrevistados (população mais heterogênea). (SEBRAE, 2005).

Tabela 5 – Determinação do tamanho da amostra para pesquisa estatística com nível de confiança de 95%

População	Erro amostral = +/- 3%		Erro amostral = +/- 5%		Erro amostral = +/- 10%	
	Split 50/50	Split 80/20	Split 50/50	Split 80/20	Split 50/50	Split 80/20
100	92	87	80	71	49	38
250	203	183	152	124	70	49
500	341	289	217	165	81	55
750	441	358	254	185	85	57
1000	516	406	278	198	88	58
2500	748	537	333	224	93	60
5000	880	601	357	234	94	61
10000	964	665	378	243	96	61
25000	1023	665	378	243	96	61
50000	1045	674	381	245	96	61
100000	1056	678	383	245	96	61
1000000	1066	678	383	245	96	61
100000000	1067	683	383	245	96	61

Fonte: SEBRAE (2005)

Quando não se tem noção do grau de homogeneidade da população, deve-se considerar um split de 50/50 (população mais heterogênea) para se alcançar um nível maior de confiança nas respostas.

Observe este exemplo: Um empreendedor deseja realizar uma pesquisa de opinião sobre a satisfação de seus clientes em relação à qualidade do atendimento em sua empresa. Verificando-se o cadastro de clientes, definiu-se que a população é de 3000 pessoas. Como se admite um erro amostral de 10%, com nível de confiança de 95%, e considerando que este empreendedor não tem a mínima noção de homogeneidade da população, determine o tamanho da amostra.

Solução:

Tabela 6 – Exemplo de determinação da amostra para pesquisa estatística com nível de confiança de 95%

População	Erro amostral = +/- 3%		Erro amostral = +/- 5%		Erro amostral = +/- 10%	
	Split 50/50	Split 80/20	Split 50/50	Split 80/20	Split 50/50	Split 80/20
5000	880	601	357	234	94	61

Fonte: SEBRAE (2005)

Com base na tabela, verifica-se que o tamanho da amostra ideal é de 94 pessoas.

3.3 Técnicas de amostragem

Para que a informação seja relevante, a amostra deve ser representativa, ter uma dimensão adequada e ser selecionada aleatoriamente. Caso estes pressupostos não se verifiquem, não se pode fazer a extrapolação dos resultados obtidos na amostra para a população. Assim, o conhecimento da estatística permite que se avaliem os métodos de recolha de dados, os resultados e as conclusões definidos num dado estudo, permitindo assim que se detectem falsas conclusões.

Para que uma amostra seja válida, é necessário que seja extraída com base em critérios rígidos, sendo estes bem respeitados. Veremos a seguir quais as formas de seleção da amostra.

Amostragem Probabilística: São amostragens em que a seleção é aleatória, de tal forma que cada elemento da população tem uma probabilidade conhecida de fazer parte da amostra. São métodos rigorosamente científicos.

Observe, aleatório é o significado de resultados não previsíveis.

Nos experimentos aleatórios, mesmos que as condições iniciais sejam sem-



pre as mesmas, os resultados finais de cada tentativa do experimento serão diferentes e previsíveis.

Exemplos:

- a) O lançamento de uma moeda;
- b) Lançamento de um dado;
- c) Determinação da vida útil de um componente eletrônico.

Todas as situações acima estão associadas ao resultado que não é previsível, por isso são chamadas de eventos aleatórios. (MORETTIN, 2010).

Amostragem Não Probabilística: são amostragens em que há uma escolha deliberada dos elementos da amostra. Depende dos critérios e julgamento do pesquisador.

Conforme mostra a figura abaixo, a amostragem probabilística pode ser classificada como: aleatória simples; aleatória estratificada; aleatória por conglomerado; sistemática, enquanto que a não probabilística pode ser classificada como: por conveniência ou acidental; intenção ou por julgamento; por cotas ou proporcional; por tráfego.

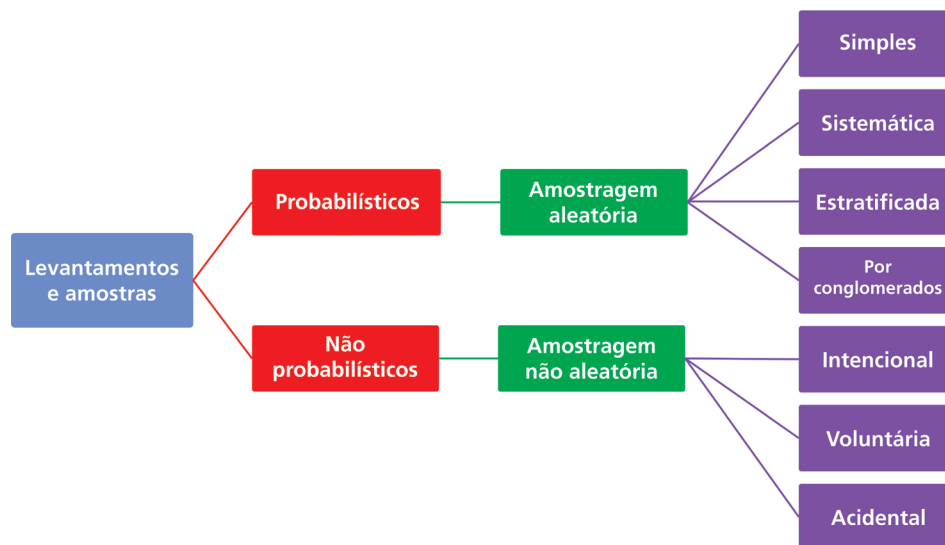


Figura 14 – Classificação dos levantamentos amostrais

Fonte: Elaboração própria

Neste estudo, nós nos concentraremos apenas na amostragem probabilística aleatória simples, sistemática e estratificada.



3.4 Amostragem probabilística simples

Na Amostragem Probabilística Aleatória Simples, os elementos da amostra são escolhidos aleatoriamente. A amostra é escolhida por meio de sorteios, de sistemas ou de qualquer outra forma aleatória.

Como é equivalente a um sorteio, pode ser realizada numerando-se a população e, em seguida, retirando um dos números ou mais, dependendo do tipo de sorteio que será utilizado.

Aprenda com este exemplo: para obter uma amostra representativa do nível de escolaridade de uma comunidade de 70 pessoas, pode-se proceder da seguinte forma:

Solução

- Enumere as 70 pessoas de 1 a 70.
- Em seguida, utilize um globo de bingo para sortear 14 números, um a um, no intervalo de 1 a 70.
- Esses 14 números representam a amostra equivalente a 20% da população, para os quais o entrevistador fará a seguinte pergunta: Qual é o seu nível de escolaridade?

Viu como é bastante simples?

3.5 Amostragem probabilística estratificada

Na Amostragem Probabilística Estratificada, a amostra é definida proporcionalmente ao tamanho de cada estrato da população. Subdivide-se a população em subgrupos (estratos). Selecionam-se amostras aleatórias simples de cada estrato. Por fim, combina-se o resultado de cada estrato para estimar parâmetros da população.

Analise este exemplo:

Imagine agora que você queira saber o nível de escolaridade dos membros da comunidade, do exemplo anterior, porém classificados por gênero. Digamos que, entre os membros, existam 40 pessoas do gênero feminino e 30 do gênero masculino.

Solução:



Neste caso, você deve:

- Primeiro enumerar os homens de 1 a 30 e as mulheres de 31 a 70.
- Depois, utilize dois globos de bingo, sendo que um possui a numeração de 1 a 30 e o outro, de 31 a 70.
- Em seguida, sorteie 6 números do primeiro e 8, do segundo globo.
- Agora que você fez isso, possui uma amostra representativa que considere a proporcionalidade entre homens e mulheres.

3.6 Amostragem probabilística sistemática

Na Amostragem Probabilística Sistemática, os elementos da amostra são escolhidos segundo um fator de repetição (intervalo fixo). Cada elemento pode ser identificado pela sua posição (uma lista ou fila).

Observe este exemplo: Vamos supor que uma empresa decida que deseja saber a opinião de 10% de seus clientes sobre a qualidade do atendimento. Caso ele adote a amostragem probabilística sistemática, basta que ele pergunte a um de cada 10 clientes.

Solução:

Portanto, perguntaria ao 10º, depois ao 20º, ao 30º e assim sucessivamente.

Caro (a) estudante, estamos concluindo a nossa aula. Vamos lembrar o que vimos até agora?

Resumo

Nesta aula mostramos que o método consiste no conjunto de meios dispostos convenientemente para se chegar ao fim que se deseja. Os métodos podem ser classificados em experimentais ou estatísticos. O objetivo de todo método estatístico é coletar dados e, então, usá-los para tomar uma decisão. As fases de um método estatístico compreendem: definição do problema; planejamento; coleta de dados; crítica dos dados; apuração dos resultados; análise e interpretação; e apresentação dos resultados.

Apontamos que População é o conjunto de todos os dados que estamos interessados em conhecer estatisticamente, enquanto amostra é uma parte representativa, selecionada para a realização do estudo (pesquisa ou observação). É importante ressaltar que o termo população, utilizado em estatística, não possui o mesmo significado aplicado à geografia.



Demonstramos que as técnicas de amostragem são utilizadas para selecionar uma amostra de uma dada população. A amostragem pode ser probabilística ou não probabilística. Na amostragem probabilística a escolha é realizada de forma aleatória, enquanto que na não probabilística a amostra é de livre escolha do pesquisador.

As técnicas de amostragem probabilísticas ou aleatórias mais utilizadas são: amostragem aleatória simples; estratificada; e sistemática.

A amostragem aleatória simples é escolhida aleatoriamente entre todos os indivíduos da população. Na amostragem aleatória estratificada, a escolha se dá de forma aleatória, porém, dentro de cada estratificação ou subconjunto adotados, como, por exemplo, gênero, idade, localização, dentre outros.

Por fim, explicamos que a amostragem sistemática se dá em intervalos ou ordem sistematicamente definidos, como, por exemplo, uma a cada 10 peças fabricadas, sendo selecionadas obrigatoriamente a décima, a vigésima e assim sucessivamente.

Atividades de aprendizagem



1. O que você entende por método estatístico? E que aspectos são importantes serem observados em cada fase?

2. Numere a 2ª coluna de acordo com a 1ª

- | | |
|---|--------------------------|
| 1. Contempla o tipo de levantamento, cronograma, custos envolvidos etc. | () Apresentação |
| 2. Definir o que se deseja pesquisar. | () Problema de pesquisa |
| 3. Estão associadas ao cálculo de medidas. | () Apuração |
| 4. Estão dispostos em forma de gráficos e tabelas. | () Planejamento |
| 5. Registro sistemático de dados com um fim específico | () Coleta dos dados |



Caro(a) estudante

Chegamos ao final da terceira aula da disciplina Estatística Aplicada. Nela, você teve oportunidade de aprender sobre o método estatístico e suas fases, população e amostra e as técnicas de amostragem.

Na próxima aula, abordaremos os temas: estatística descritiva, estatística inferencial e distribuição de frequência. Serão informações importantes dentro da área em que você está se qualificando. Sua dedicação fará grande diferença em seu processo de aprendizagem.



Aula 4 - Estatística Descritiva

Objetivos:

- Identificar a estatística descritiva.
- Distinguir e exemplificar a estatística inferencial.
- Apontar e exemplificar a distribuição de frequências.

Prezado (a) estudante,

Esta é a quarta aula da disciplina Estatística Aplicada e os assuntos a serem tratados são: a estatística descritiva, a estatística inferencial e a distribuição de frequências. Esperamos que, nesta aula, você compreenda bem os conceitos de estatística descritiva e inferencial, bem como aprenda a utilizar as distribuições de frequências.

No desenvolvimento de nossa aula, observaremos a importância da distribuição de frequência para a organização de conjunto de dados. A distribuição de frequência não é uma medida de análise estatística propriamente dita, no entanto, serve para agrupar os dados, facilitando assim a aplicação dos modelos estatísticos.

Ao estudar sobre estatística descritiva e inferencial, você perceberá que, enquanto a estatística descritiva, como indica o nome, tem por objetivo descrever o comportamento de determinado fenômeno ou evento por meio dos dados observados, a estatística inferencial, por outro lado, objetiva a generalização dos resultados. Em outras palavras, consiste considerar os resultados de amostra como válidos para toda a população. Prepare-se para este novo conteúdo.

4.1 Ramos da estatística

Como vimos na aula 2, nos últimos anos, especialmente a partir de 1940,



com o advento da informática, o volume de dados cresceu significativamente. Isso fez com que a análise de um conjunto completo de dados, devido ao seu tamanho, se tornasse em muitos casos trabalhosa. Como alternativa, tem-se a opção de, ao invés de analisar todo o conjunto de dados, utilizar apenas uma parte representativa para avaliar e, da escolhida, fazer-se a generalização para todo o conjunto.

Vale lembrar que, apesar de a estatística descritiva ser um ramo importante da Estatística e continuar sendo amplamente utilizada, as informações estatísticas quase sempre são obtidas de amostras (observações feitas apenas em parte de um conjunto grande de itens).

A pesquisa eleitoral é um exemplo que ilustra, e muito bem, esta situação. Pois imagine se, para descobrir a preferência de votos para a Presidência da República, tivéssemos que ouvir todos os eleitores brasileiros? Logicamente, isso seria impraticável, pois é altíssimo o custo de uma pesquisa como esta. Daí a importância de se utilizar da estatística inferencial.



Figura 15

Fonte: Ilustradora

Outro exemplo do uso de generalização de resultados se dá quando é necessário prever o fluxo de tráfego numa rodovia que ainda não foi construída, com base no tráfego observado no passado, em rodovias alternativas. Como vimos nestes dois casos, a análise realizada exige generalizações que vão além dos dados e, conseqüentemente, a característica mais importante do recente crescimento da estatística é uma mudança de ênfase da estatística descritiva para os métodos da inferência estatística. (FREUND, 2006).

O estudo da estatística, didaticamente, divide-se em dois ramos principais: a **estatística descritiva** e a **estatística inferencial**.

A **estatística descritiva** é o ramo que trata da organização, do resumo e da



apresentação dos resultados, e é responsável pelo estudo das características de uma dada população.

Enquanto que a **estatística inferencial**, também conhecida como estatística indutiva, é o ramo que trata de tirar conclusões sobre uma população a partir de uma amostra de uma dada população ou universo.

Em última análise, ao fazer previsões a partir da parte para o todo, ou seja, com base na análise de um conjunto limitado de alguns dados (amostra) recolhidos junto de um conjunto total de indivíduos (população), pretendemos caracterizar a população.

Vamos analisar alguns casos procurando identificar qual parte do estudo representa o ramo descritivo da estatística. Quais conclusões podem ser tiradas do estudo usando a estatística inferencial em cada um dos casos.

Vamos analisar alguns exemplos?

Caso 1:

“Uma amostra entrevistada pelo Instituto Data Popular (2012) mostrou que os consumidores com renda mensal per capita a partir de R\$ 1.020 (classe alta) vão gastar, em média, R\$ 1.062 com presentes neste Natal, ante o gasto de R\$ 484 dos consumidores considerados da classe média, com renda per capita mensal entre R\$ 292 e R\$ 1.019.”

Fonte: Terra

Disponível em: <http://economia.terra.com.br/noticias/noticia.aspx?idNoticia=201212211326_TRR_81863147>

Solução:

A estatística descritiva inclui afirmações, tais como: “consumidores com renda mensal per capita a partir de R\$ 1.020 (classe alta) vão gastar, em média, R\$ 1.062 com presentes neste Natal” e “ante o gasto de R\$ 484 dos consumidores considerados da classe média, com renda per capita mensal entre R\$ 292 e R\$ 1.019”.

Uma possível inferência tirada deste estudo é a de que a classe alta tenderá a gastar mais do que a classe média nas compras de natal.



Caso 2:

Um Instituto de pesquisa independente ouviu uma amostra de 500 pessoas na cidade XY para saber qual a sua preferência musical. Como resultado, observou-se que 27% preferem música sertaneja, 23%, música religiosa, 10%, pagode, 10%, MPB e 30% preferem outros ritmos musicais.

Solução:

A estatística descritiva inclui afirmações tais como “uma amostra de 500 pessoas na cidade XY para saber qual a sua preferência musical. Como resultado, observou-se que 27% preferem música sertaneja, 23%, música religiosa, 10%, pagode, 10%, MPB e 30% preferem outros ritmos musicais”.

Uma possível inferência tirada desse estudo é a de que, dentre a população da cidade XY, 27% preferem música sertaneja, 23%, música religiosa, 10%, pagode, 10%, MPB e 30% preferem outros ritmos musicais.

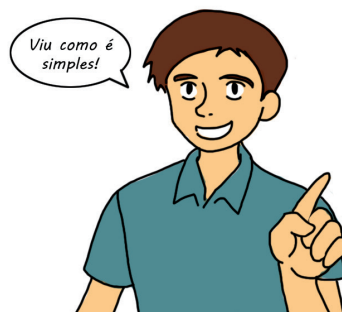


Figura 16

Fonte: Ilustradora

Você deve ter observado que a estatística descritiva apenas descreve os resultados encontrados, enquanto que a estatística inferencial generaliza os resultados encontrados para toda a população da qual a amostra foi retirada. Nas próximas aulas, veremos mais sobre este assunto.

Vamos seguir com as atividades para você praticar o que vimos até agora!



Atividade de aprendizagem

1. Para uma pesquisa, foram entrevistados 200 mulheres e 200 homens da Cidade X. Foi aplicada a seguinte pergunta para todos os participantes da pesquisa: “quanto livros você leu no último ano?”





Como resultado, verificou-se que, entre os homens, a média de leitura foi de 2 livros e, entre as mulheres, de 3 livros.

Diante do exposto, indique a parte que se refere à estatística descritiva e qual a inferência estatística possível para o exemplo dado.

4.2 Distribuição de frequências

Como já mostramos anteriormente, muitas vezes é necessário trabalhar com dados amostrais a fim de analisar populações e dados por meio da inferência estatística. Mas, ocasionalmente, é possível e necessário trabalhar com os dados da população inteira. Como exemplos, podemos citar o censo realizado pelo IBGE a cada 10 anos, por meio do qual é contada toda a população brasileira. Ou a nota de 40 alunos de uma turma matriculados na escola X.

É importante observar que, quando um banco de dados tem muitas entradas, pode ser difícil visualizar padrões. Muitas vezes, é impraticável a realização de uma análise sem antes realizar o devido agrupamento dos dados, pois nestes casos temos grande volume de dados e, para tanto, precisamos organizá-los adequadamente por meio de agrupamentos, a fim de facilitar a análise dos resultados.

Algumas características mais importantes devem ser consideradas quando organizamos e descrevemos um conjunto de dados; são elas: **o centro, a variabilidade** (ou amplitude) e **a forma**. Porém, nesta aula, nós nos concentraremos especificamente na forma com que os dados são descritos.

Os conteúdos sobre medidas centrais e variabilidade serão tratados nas próximas aulas.



Larson (2010, p. 32) define que a “**distribuição de frequências** é uma tabela que mostra classes ou intervalos das entradas e de dados com uma contagem do número de entradas de cada classe. A frequência f de uma classe é o número de entrada de dados de uma classe”.

A distribuição de frequências demonstrada abaixo está distribuída em quatro classes. Na coluna das classes (à esquerda) verifica-se que cada classe tem um limite inferior de classe, que é o menor número de cada classe, e um limite superior, representado pelo maior número pertencente à classe.

Na distribuição de classe na tabela abaixo, os limites inferiores são representados por 1, 16, 31 e 46, e limites superiores pelos algarismos 15, 30, 45 e 60. Outro aspecto que pode ser observado, ainda nesta coluna, é a largura das classes, a distância entre os limites inferiores. Por exemplo, a largura da classe de distribuição de frequência mostrada é $16 - 1 = 15$.

Classe	Frequência f
1 - 15	3
16 - 30	7
31 - 45	4
46 - 60	2

Na coluna de frequência, verifica-se o número de observações para cada uma das classes. Suponhamos que os dados da tabela acima refiram-se ao número de integrantes de uma família e os intervalos de classe representam a idade das pessoas, neste caso, observamos que a família hipotética possui 3 pessoas entre 1 e 15 anos, 7 com idades entre 16 e 30 anos, e assim por diante.

A **amplitude** é outro conceito utilizado na distribuição de frequências e consiste na diferença entre as entradas máxima e mínima. No exemplo acima, seria a pessoa com maior e a com menor idade. No entanto, no intuito de facilitar a sua compreensão, mostraremos este conceito detalhadamente mais adiante.

Quando se coletam os dados para uma pesquisa, tais informações são classificadas como dados brutos.

Vejamos o rol de notas de uma turma de 40 alunos. Observa-se que se trata de um conjunto de dados brutos, ou seja, não permite em sua forma que se faça uma análise dos dados apresentados.

O conjunto de dados amostrais a seguir lista a nota de 40 estudantes.

Vamos construir uma distribuição de classes?

80	50	60	40	90	85	83	88	71	64
32	12	92	91	23	35	25	45	48	79
36	55	65	72	78	86	73	67	74	87
58	76	92	69	30	62	76	84	90	76

Mas você deve estar se perguntando em quantas classes dividir e que critérios utilizar.



Figura 17
Fonte: ilustradora

Para facilitar a análise e a visualização dos resultados, é recomendável que se utilize de 5 a 20 classes. Apesar de o número de classes poder ser definido de forma arbitrária, conforme a necessidade ou gosto do pesquisador, existe um modelo universalmente aceito.

Vamos conferir?

O número de classes (k) a ser utilizado pode ser calculado com base no número de observações (n) realizadas.

Por exemplo, em uma pesquisa, foram entrevistados $n = 40$ indivíduos. Então, o número de classes será dado por:

$$k = \sqrt{n}$$



Ou seja, o número de classes será igual ao resultado da raiz quadrada do número de observações, logo,

$$k = \sqrt{40} = 6,32$$

Como o número de classes deve ser sempre um número inteiro, arredondaremos para 6 classes.

Já definimos o número de classes (k) em que iremos organizar os dados, agora precisamos agrupá-los, e para isso é necessário saber a amplitude do intervalo de classe (c).

Cabe destacar que o intervalo de classe só pode ser definido após a definição da amplitude total.



O cálculo de intervalo de classe $k = \sqrt{n}$ é válido para $n \leq 100$. Para $n > 100$, utiliza-se $k = 5 \log n$



Figura 18

Fonte: ilustradora

A **amplitude total** (A) corresponde à diferença entre o maior e o menor valor observado.

Isso quer dizer que, no rol de notas dos alunos, conforme acima exposto, teremos:

$$A = 92 - 12.$$

Agora que temos a amplitude total calculada, vamos obter a amplitude do intervalo de classes (c)

$$C = \frac{A}{k - 1}$$

$$C = \frac{92 - 12}{6 - 1}$$

$$C = \frac{80}{5} = 16$$

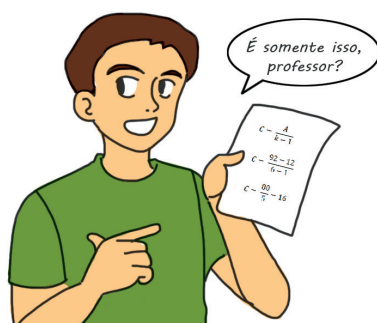


Figura 19

Fonte: ilustradora

Nós descobrimos a amplitude total e a amplitude do intervalo de classes, entretanto, ainda falta definir o intervalo de classes.

E como definir o intervalo de classes?

Também é simples definir o intervalo de classes. A entrada mínima de dados é considerada como o **ponto médio** do limite inferior, assim, para encontrar o limite inferior da primeira classe, basta dividir a largura do intervalo de classes por 2 e, em seguida, subtraí-lo do menor valor observado.

O ponto médio de uma classe é a soma dos limites inferiores e superiores da classe dividida por dois. O **ponto médio** é também denotado como marca da classe. (LARSON, 2010).

$$\text{Ponto médio} = \frac{(\text{limite inferior da classe}) + (\text{limite superior da classe})}{2}$$

Vamos continuar resolvendo o nosso exemplo?

$$\frac{C}{2}$$

Limite inferior da primeira classe = menor valor observado -

$$\text{Limite inferior da primeira classe} = 12 - 16/2 = 12 - 8 = 4.$$

Reexplicando, no exemplo dado, a menor nota foi 12 pontos, este número será considerado o ponto médio da classe inferior. Se tivermos o ponto médio do intervalo de classe, que no exemplo dado foi de 16 pontos, é simples descobrir o limite inferior deste intervalo; basta subtrair a metade do intervalo de classes ($16/2 = 8$), ou seja, $12 - 8 = 4$, logo, o limite inferior para o primeiro intervalo de classes é 4 pontos.



Para encontrar os limites inferiores das cinco classes restantes, adicione a largura de classe 16 ao limite superior de cada classe anterior.

Dessa forma, temos:

4 20	primeira classe
20 36	segunda classe
36 52	terceira classe
52 68	quarta classe
68 84	quinta classe
84 100	sexta classe

Você deve ter percebido que o limite superior da classe anterior e o limite inferior da classe seguinte são iguais. Por exemplo, 4 | 20 e 20 | 36. Mas também deve ter observado o símbolo |.

Pois bem, o símbolo | significa que o intervalo é fechado no limite inferior e aberto no superior. Em outras palavras, o limite superior não está incluso no intervalo. Assim, tomando o exemplo anterior, o aluno que obtiver nota igual a 20, pertencerá ao intervalo 20 | 36, e não ao intervalo 4 | 20.

A distribuição de frequências é mostrada a seguir. A primeira classe, 4 | 20, tem 1 frequência, a segunda classe, 20 | 36, tem frequência 5 e assim por diante. Observe que o número de frequência é igual à quantidade de alunos dentro de cada um dos intervalos. Observe ainda que a soma das frequências é 40, que é o número de alunos observados em nosso conjunto de dados.

A soma é denotada por Σf , em que Σ é a letra maiúscula sigma.

Classe	Frequência f
4 20	1
20 36	5
36 52	5
52 68	7
68 84	12
84 100	10
	$\Sigma f = 40$

A **frequência relativa** de uma classe é a porção ou porcentagem de dados que está em determinada classe. Para a frequência relativa de uma classe,



é necessário dividir a frequência f pelo tamanho da amostra. A frequência pode ser descrita tanto em número decimal como em percentual. (LARSON, 2010).

$$\text{Frequência relativa} = \frac{\text{Frequência de classe}}{\text{Tamanho da amostra}} = \frac{f}{n}$$

A frequência acumulada de uma classe é a soma da frequência para aquela classe e todas as classes inferiores. A frequência acumulada da última classe é, portanto, igual ao tamanho n da amostra.

Encontrando pontos médios, frequência relativa e acumulada

Voltando ao exemplo anterior, vamos encontrar os pontos médios, frequências relativas e acumuladas.

Solução:

O ponto médio, a frequência relativa e a acumulada para as três primeiras classes são:

Classe	f	Ponto Médio	Frequência relativa	Frequência acumulada
4 20	1	$\frac{4+20}{2} = 12$	$\frac{1}{40} = 0,025$	1
20 36	5	$\frac{20+36}{2} = 28$	$\frac{5}{40} = 0,125$	6
36 52	5	$\frac{36+52}{2} = 44$	$\frac{5}{40} = 0,125$	11

Vejam os pontos médios, as frequências relativas e acumuladas de todas as classes na tabela a seguir:

Classe	Frequência f	Ponto Médio	Frequência relativa	Frequência acumulada
4 20	1	12	0,025	1
20 36	5	28	0,125	6
36 52	5	44	0,125	11
52 68	7	60	0,175	18
68 84	12	76	0,3	30
84 100	10	92	0,25	40
$\Sigma f = 40$			$\Sigma \frac{f}{n} = 1$	



Interpretação

Observe que os dados organizados pela distribuição de frequências facilitaram a leitura e a interpretação, sua principal vantagem, especialmente quando trabalhamos com grande quantidade de dados. Mas a distribuição de frequência não constitui uma ferramenta de análise estatística, propriamente dita. Por enquanto, podemos analisar, na distribuição de frequências, que a nota mais comum entre os estudantes foi de 69 a 84 pontos. Entretanto, existem diversos padrões de interpretação dos resultados, dos quais trataremos nas aulas a seguir.

A-Z

Histogramas

são constituídos por um conjunto de retângulos, com as bases assentadas sobre um eixo horizontal, tendo o centro da mesma no ponto médio da classe que representa, e cuja altura é proporcional à frequência da classe.

4.2.1 Representação gráfica de uma distribuição

O **histograma** é formado por um conjunto de retângulos justapostos cujas bases se localizam sobre o eixo horizontal, de tal modo que os seus pontos médios coincidam com os pontos médios dos intervalos de classe e seus limites coincidam com os limites da classe.

Na figura 9, você pode verificar o histograma para a frequência das notas dos alunos, construído a partir do exemplo anterior. Observe, no eixo horizontal, estão representados os pontos médios, sendo que, para o ponto, há um retângulo representativo ao número de alunos com notas dentro do intervalo de classe. Por exemplo, o intervalo 4 | 20 é representado pelo ponto médio 12, e para este intervalo observa-se um único aluno.

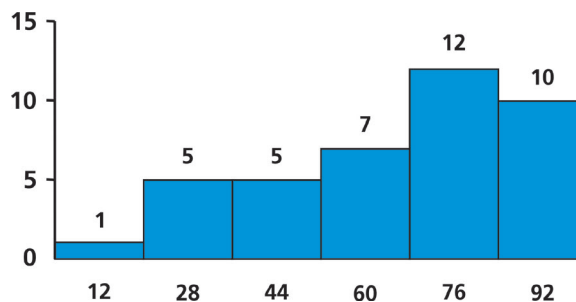


Figura 20 - Histograma das notas dos alunos

Fonte: ilustradora

Vamos relembrar o conteúdo de nossa aula.

Resumo

A estatística divide-se em descritiva e inferencial. A estatística descritiva descreve os fenômenos, utilizando técnicas estatísticas a partir dos dados observados. Já a estatística inferencial consiste em fazer generalizações, verdades estatísticas, para uma população inteira.



A distribuição de frequências consiste na organização dos dados em intervalos de classe, com o objetivo de facilitar a leitura e a utilização dos dados para a realização de análises e inferências.

Atividade de aprendizagem



2. Dado o rol dos salários de 25 funcionários da empresa Alfa, conforme abaixo, faça o que se pede:

670	1050	700	720	900
620	1020	920	910	1200
930	855	1060	730	780
580	760	920	690	1030
580	760	920	860	830

Calcule o número de classes, o ponto médio de cada intervalo de classes, a frequência relativa e acumulada e construa o histograma.

Caro(a) estudante

Chegamos ao final da nossa quarta aula. Nesta aula, abordamos os temas estatística descritiva e inferencial, e distribuição de frequência.

Na próxima aula trataremos de medidas de posição, ou medidas de tendência central, das quais se destacam a média, a mediana e a moda. Não fique com dúvidas. Volte sempre ao que já foi exposto e leia com calma até conseguir entender o que já foi apresentado até aqui. Até a nossa próxima aula!



Aula 5 - Medidas de Posição

Objetivos:

- Identificar as medidas de posição ou de uma tendência central, conceituando-as e exemplificando-as.
- Determinar a média aritmética, a mediana e a moda de uma amostra.

Prezado (a) estudante.

Esta é a quinta aula da disciplina Estatística Aplicada e os assuntos a serem tratados são as medidas de posição ou medidas de tendência central. Mais especificamente, você irá conhecer a média, a mediana e a moda. Esperamos que, nesta aula, você compreenda bem estes conceitos e que os mesmos lhe auxiliem nas tarefas que você tiver que executar na função que for exercer após o término deste curso.

Boa aula!

Nas aulas anteriores você teve oportunidade de estudar a distribuição de frequências, tabelas e gráficos, instrumentos que nos permitem descrever o padrão de variação de um determinado fenômeno estatístico observado.

Nesta aula, vamos apresentar as medidas de posição, também chamadas de medidas de tendência central, as quais descrevem os valores médios de um conjunto de dados observados.

Entre as medidas de posição utilizadas estão a média aritmética, a mediana e a moda. Porém existem outras, tais como a média geométrica, a média harmônica, a quadrática, a cúbica e a biquadrática.

Como você verá, um grande número de métodos numéricos está disponível para descrever conjuntos de dados quantitativos. A maior parte destes



métodos mede ou a tendência central ou a variabilidade de um conjunto de dados ou medições.

- a)** A tendência central do conjunto de medições – ou seja, a tendência dos dados de se agrupar, ou centralizar, entre certos valores.
- b)** A variabilidade do conjunto de medições – isto é, a dispersão dos dados.

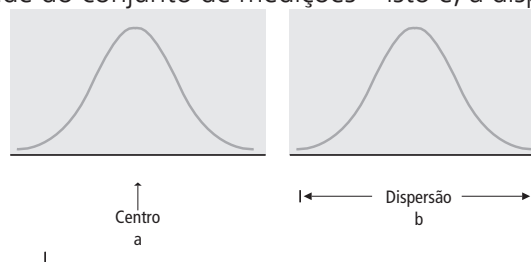


Figura 21 - Medidas numéricas descritivas

Fonte: McClave (2009).

Nesta aula, nós nos concentraremos nas medições de tendência central, também conhecidas como medidas de posição.

5.1 Média aritmética simples

A medida de tendência central mais popular é a que o leigo chama de “média”. De acordo com Freund (2006, p. 58), “está certo usar o termo ‘média’ para a média aritmética, mas como existem outras médias em Estatística, não podemos permitir um emprego vago da expressão quando há risco de ambiguidade”.

De acordo com Freund (2006), a popularidade da média como uma medida do “meio” ou “centro” de um conjunto de dados não é acidental, pois toda vez que utilizamos um único número para descrever algum aspecto de um conjunto de dados, há certas exigências ou características desejáveis que devemos considerar. Para o autor acima citado, a popularidade da média se dá pelas seguintes propriedades:

- 1.** A média pode ser calculada para qualquer conjunto de dados numéricos e, portanto, sempre existe.
- 2.** Qualquer conjunto de dados numéricos tem uma, e uma só média e, portanto, é sempre única.
- 3.** A média se presta a outros tratamentos estatísticos; por exemplo, como



veremos, as médias de vários conjuntos de dados podem ser sempre combinadas em uma média global de todos os dados.

4. A média é relativamente confiável, no sentido de que as médias de amostras repetidas extraídas da mesma população geralmente não flutuam, ou variam tanto quanto outras medidas estatísticas usadas para estimar a média de uma população.

5. A média leva em conta todos os elementos de um conjunto de dados.

A média consiste em uma medida de posição central. Além de ser a mais conhecida e utilizada, a média talvez seja também a medida de posição mais importante. A média, quando obtida a partir dos dados de uma amostra, é representada por \bar{x} , porém quando obtida a partir de dados de uma população, a média é denotada pela letra grega μ .

A média de n números é a sua soma dividida por n .

Como a média é bastante comum mesmo no dia a dia, é conveniente ter uma fórmula simples, de fácil aplicação. Isso exige que os dados da média que calcularemos sejam denotados por símbolos, tais como x , y ou z e que o número de dados de uma amostra, isto é, o tamanho da amostra, seja denotado pela letra n , a fim de que assim possamos generalizar uma única fórmula para todos os casos. Vamos considerar que, escolhendo a letra x para denotar os dados observados de uma determinada amostra de n valores de uma amostra, nós os escreveremos da seguinte forma x_1, x_2, \dots, x_n (onde lemos "xis um", "xis dois", ..., "xis n ") e escrevemos como

$$\text{média de amostra} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

A partir da demonstração acima, temos no numerador da média da amostra dado pela soma das n observações. Ou seja

$$\sum x_i = x_1 + x_2 + \dots + x_n$$

Donde podemos concluir que a média da amostra é representada pela seguinte expressão

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



A letra grega Σ é o símbolo do somatório. Um **somatório** é um operador matemático que nos permite representar facilmente somas de um grande número de termos até infinitos e é definido por:

$$\sum_{i=m}^n f(i) = f(m) + f(m+1) + \dots + f(n-1) + f(n)$$

A variável i é o **índice do somatório** que designa um valor inicial chamado **limite inferior**, m .

A variável i percorre os valores inteiros até alcançar o **limite superior**, n .

Por exemplo, se queremos expressar a soma dos cinco primeiros números naturais podemos representá-la assim:

$$\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15$$

Vamos a um exemplo, para melhor explicar o conceito de média.

Para ilustrar o cálculo da média da amostra, consideraremos os seguintes dados referentes à idade de cinco estudantes de um curso técnico.

18 16 20 19 22

Solução:

Utilizando-se da notação x_1 , x_2 , x_3 , x_4 e x_5 para a idade de cada um dos estudantes, teremos:

$$x_1 = 18 \quad x_2 = 16 \quad x_3 = 20 \quad x_4 = 19 \quad x_5 = 22$$

Para calcular a média da idade dos alunos na amostra dada, podemos utilizar a fórmula do cálculo da média, conforme explicada acima.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{18 + 16 + 20 + 19 + 22}{5} = 19$$

A idade média dos estudantes, portanto, é de 19 anos.

A média de uma população N é definida da mesma forma que a média de uma amostra, trocando o tamanho da amostra n pelo tamanho da população N . Denotando a média de uma população por μ , escrevemos pela



seguinte notação.

A média da População é denotada pela fórmula

$$\mu = \frac{\sum x_i}{N}$$

É importante lembrar que o $\sum x_i$ no cálculo da média de uma população, consiste na soma de todos os N valores de x que constitui a população.

Você certamente deve estar se perguntando qual a diferença entre usar a descrição de amostra ou de população; se esta diferença é apenas uma diferença na utilização de símbolos.



a letra grega minúscula μ que lemos "um", corresponde a m.

A descrição de amostras constitui uma estatística, ao passo que a descrição de uma população é um parâmetro.

Ainda está com dúvidas? Não se preocupe! Pois bem, então vamos exemplificar.

Imagine que estejamos interessados em verificar a eficiência de um lote de 20.000 pneus de uma determinada marca. É óbvio que não poderemos testar todos os pneus, pois neste caso, não restaria nenhum para vender ou usar. A única alternativa viável é, então, calcularmos \bar{x} e usarmos o valor deste como uma estimativa de μ .

Verifique se com este exemplo vai ficar mais claro. Se os pneus da amostra duram 32.000, 30.000, 35.500, 33.000, 40.000 quilômetros rodados, o que poderemos concluir sobre a duração média dos 20.000 pneus do lote? Vamos considerar uma amostra de 5 pneus.

Solução:

A média da amostra é dada por:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{32000 + 31000 + 36500 + 33000 + 35000}{5} = 33500 \text{ quilômetros}$$

Supondo que os dados analisados constituem uma amostra no sentido técnico, ou seja, um conjunto de dados a partir dos quais podemos tirar generalizações válidas, podemos estimar que a duração média dos 20.000 pneus é de $\mu = 33.500$ km.



Para dados não negativos, a média não só descreve o meio de um conjunto de dados, mas também impõe certa limitação ao seu tamanho, pois se multiplicarmos por n ambos os lados da equação $\bar{x} = \frac{\sum x_i}{n}$, veremos que $\sum x = n \times \bar{x}$, e, portanto, que nenhuma parte, ou subconjunto dos dados, pode exceder $n \times \bar{x}$.

5.2 Mediana

A mediana consiste em uma medida de posição central de uma variável. “A mediana é o valor que fica no meio da sequência quando os dados são arranjados na ordem ascendente (classificação do menor para o maior). Com um conjunto ímpar de observações, a mediana é o valor do meio. Um número par de observações não tem um valor único no meio. Neste caso, seguimos a convenção de definir a mediana como sendo a média dos valores das duas observações do meio.” (ANDERSON, 2002, p. 82).

A mediana é representada por Md.

Em síntese, a definição da mediana é dada nos seguintes termos:

Com os valores dos dados arranjados na ordem ascendente:

- a) Para um número ímpar de observações, a mediana é o valor do meio;
- b) Para um número par de observações, a mediana é a média dos dois valores centrais.



A mediana é o valor do elemento do meio, se n é ímpar, e a média de dois valores do meio, se n é par.

Vale ressaltar que, apesar de a média ser muito mais utilizada, há casos específicos em que a sua utilização de forma isolada pode levar o leitor ou analista desavisado a tirar conclusões errôneas.

Vejamos o exemplo do rol abaixo que traz o valor do salário de cinco funcionários de uma determinada empresa: caso utilizássemos a média, teríamos como medida central o valor de R\$ 2920,00, conforme segue:

700 1000 1300 1600 8000

$$\bar{x} = \frac{\sum x_i}{n} = \frac{700 + 1000 + 1300 + 1600 + 8000}{5} = 2920$$

Como você pode observar, a média de 2920 é bem superior à maioria dos salários. Isso ocorreu devido ao alto valor do salário de um único funcionário.

Neste caso, o empregador poderia muito bem utilizar esta média para afirmar que, em média, paga excelentes salários, apesar de não estar mentindo, uma vez que o cálculo da média está correto.

O valor da média, neste caso, não representa, portanto, a realidade da maioria dos indivíduos observados. Em outras palavras, o valor da média encontrado é bem superior ao salário de 4 dos 5 funcionários.

Mas com certeza você está se perguntando. O que fazer nestes casos em que a média não constitui uma medida representativa da amostra?

Bom, caro estudante, para este caso, a mediana é capaz de melhor representar uma medida de tendência central, conforme vamos aqui calcular.

É bastante fácil, vejamos!

Solução:

Basta arranjarmos os valores em ordem crescente, e como temos cinco indivíduos, portanto, um número ímpar. Assim, a mediana será a observação que estiver no meio, ou seja, a terceira observação.

Tabela 7 - Salários de Empregados	
Funcionário	Valor do salário
x1	700
x2	1000
x3	1300
x4	1600
x5	8000

Fonte: Elaboração própria com dados fictícios

Logo, pode-se afirmar que a mediana é igual a R\$ 1300,00. Mas qual medida melhor representa a realidade dos salários dos 5 empregados, a média de 2920 ou a mediana de 1300?



Logicamente, a mediana, neste caso, parece melhor representar a realidade dos salários dos funcionários da amostra observada.

A mediana é de muito valor quando se está descrevendo grandes conjuntos de dados. Se o conjunto é caracterizado por um histograma de frequência relativa (figura 11), a mediana é o ponto sobre o eixo x em que metade da área sob o histograma encontra-se acima da mediana e metade abaixo dela. A notação da mediana de uma amostra é m . (MCCLAVE, 2009).

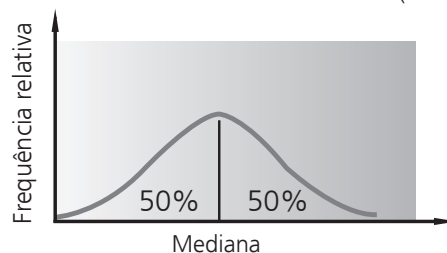


Figura 22 - Localização da mediana

Fonte: McClave (2009).

Uma comparação da média e da mediana nos dá um método geral para detectar assimetria em conjuntos de dados, como mostra o quadro a seguir. Um conjunto de dados é dito não simétrico quando uma curva de distribuição tem mais observações extremas do que a outra.

Vamos resolver mais um exercício que serve como exemplo?

Em uma prova, 6 alunos obtiveram as seguintes notas: 7; 6; 9; 4; 10; e 3. Calcule a mediana.

Solução:

Passo I – Vamos ordenar todas as notas em ordem crescente.

3 4 6 7 9 10

Passo II – Como você deve ter observado, o tamanho da amostra é par, portanto, não há um valor único na posição central. Neste caso, a mediana é o somatório dos dois termos centrais dividido por dois. Vejamos:

$$Md = \frac{6 + 7}{2} = \frac{13}{2} = 6,5$$

Portanto, a mediana das notas dos alunos é 6,5. Isso significa que a metade dos alunos obtiveram notas menores do que 6,5 e a outra metade, maior



do que 6,5.

Que tal praticar um pouco?

Atividade de aprendizagem



1. Em uma pesquisa em 5 supermercados, o kg de arroz apresentava os seguintes preços: R\$ 2,00; R\$1,90; R\$2,30; R\$2,20; R\$2,50; e R\$2,40.

Com base nos dados propostos, calcule o valor da média e da mediana para o kg de arroz.

5.3 Moda

A moda é uma medida utilizada para descrever o meio ou o centro de um conjunto de dados, é a moda, definida simplesmente como o valor ou categoria que ocorre com a maior frequência e mais de uma vez. Para Freund (2006), as duas vantagens principais da moda são: não exige cálculo algum, apenas uma contagem, e pode ser determinada tanto para dados numéricos quanto para categóricos.

Suponha que uma empresa decidiu fazer um levantamento geral das atividades. O quadro abaixo mostra os valores das vendas em milhares de R\$, nos dois últimos 24 meses.

16	29	16	19	24	17	18	20	19	22	17	24
34	20	19	22	11	14	13	19	20	26	19	21

Dado o conjunto de informações, vamos calcular a moda.



Solução:

Passo I – Vamos colocar os dados em ordem crescente:

11 – 13 – 14 – 16 – 16 – 17 – 17 – 18 – 19 – 19 – 19 – 19 – 19 – 20 – 20 –
20 – 21 – 22 – 22 – 24 – 24 – 26 – 29 – 34

Passo II – Para saber a moda da amostra, basta contar o resultado com maior frequência. Como é fácil perceber, a moda é 19.

Passo III – Vejamos a frequência no histograma.

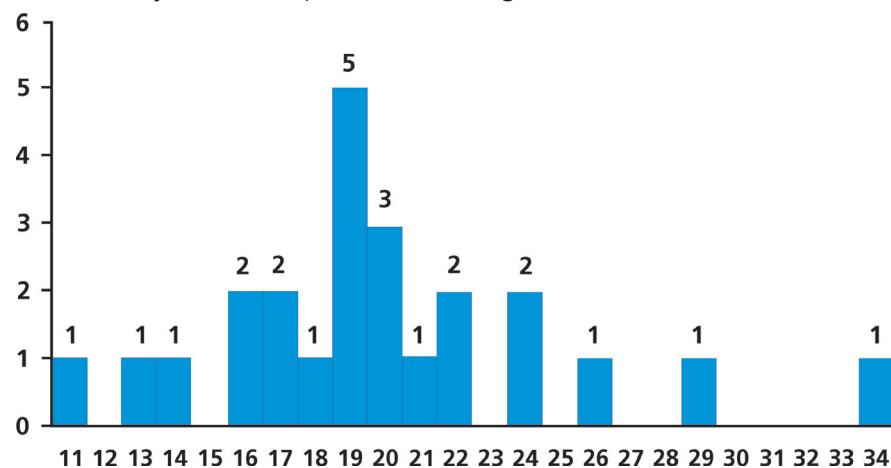
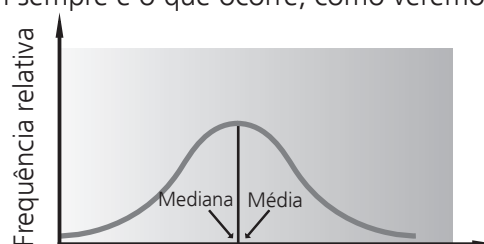


Gráfico 2 - Histograma de frequência

Fonte: Elaboração própria

Detectando assimetria com a comparação entre a média e a mediana.

Um conjunto de dados é considerado simétrico quando a média e a mediana são iguais, conforme ilustra a figura 23. Nestes casos, qualquer uma das medidas de tendência central é capaz de descrever bem o conjunto de dados, entretanto, nem sempre é o que ocorre, como veremos nas figuras 24 e 25.



Se o conjunto de dados é simétrico,
a mediana é igual à média

Figura 23 – Simetria

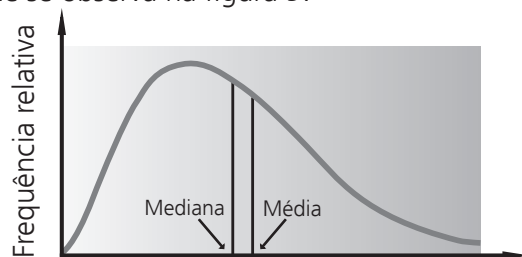
Fonte: ilustradora



Há casos em que a mediana é menor do que a média, e nestes casos, o conjunto é considerado assimétrico à direita. O exemplo anterior encaixa-se perfeitamente nesta situação. Observe que a mediana foi de R\$ 1300,00 e a média, de R\$ 2920,00.

Mas por que isso acontece?

Bom, você deve ter percebido que a maioria dos empregados, no caso dado, recebe salários abaixo da média e como a mediana representa o ponto médio de um conjunto de dados, logicamente a mediana será menor do que a média, conforme se observa na figura 9.

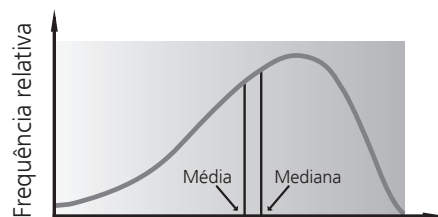


Se o conjunto de dados é simétrico,
a mediana é igual à média

Figura 24 - Simetria à direita

Fonte: ilustradora

De modo análogo ao caso anterior, são considerados assimétricos à esquerda os conjuntos que possuem a mediana maior do que a média. Isso ocorre sempre que a maioria dos dados observados em um conjunto (amostra) são maiores do que a média.



Se o conjunto de dados tem simetria à esquerda,
a média é menor (ou à esquerda) do que a mediana

Figura 25 - Simetria à esquerda

Fonte: ilustradora

5.4 Média ponderada



Quando calculamos uma média e as grandezas em jogo não têm todas a mesma importância ou a mesma significância, podemos muito bem não obter uma medida estatística que nos diga o que estamos esperando descrever. Vamos considerar o seguinte caso.

Às vezes, os dados contêm entradas que têm um maior efeito na média do que outras. Nestes casos, a média aritmética não é mais adequada, uma vez que esta considera que todas as variáveis possuem o mesmo peso ou grau de importância. Para encontrar a média de tais conjuntos, é preciso calcular a média ponderada.

Para obter essa média significativa, é necessário que cada elemento tenha um peso de importância relativa para, então, calcular uma média ponderada. Em geral, a média ponderada \bar{x}_w de um conjunto de números x_1, x_2, x_3, \dots e x_n cuja importância relativa é expressa numericamente por um conjunto correspondente de números w_1, w_2, w_3, \dots e x_n é dada por

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum w \cdot x}{\sum w}$$

Aqui, $\sum w \cdot x$ é a soma dos produtos obtidos multiplicando cada x pelo peso correspondente, e $\sum w$ é simplesmente a soma dos pesos. Note que, quando os pesos são todos iguais, a fórmula da média ponderada reduz-se à da média aritmética.

Verifique se você entendeu. Um consumidor deseja saber o preço médio que pagará por kg de carne. Sabendo que este comprará 2 kg de carne tipo patinho, 5 kg de alcatra e 3 kg de picanha. Dados os preços de R\$ 9,00, R\$ 16,00 e R\$ 23,00, respectivamente, calcule o preço da média ponderada.

Como queremos saber o preço médio a ser pago pela carne, então temos que o valor por kg é representado por x_i e o peso por kg será representado por w_i .

Portanto,

$$x_1 = 2, x_2 = 5 \text{ e } x_3 = 3$$

$$w_1 = 9, w_2 = 16 \text{ e } w_3 = 23$$



Agora que já definimos os valores de cada variável, basta aplicarmos a fórmula da média ponderada,

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum w \cdot x}{\sum w}$$

$$\bar{x}_w = \frac{2 * 9 + 5 * 16 + 3 * 23}{2 + 5 + 3} = \frac{\sum 167}{\sum 10}$$

$$\bar{x}_w = 16,70$$

Portanto, o preço médio que o consumidor irá pagar pelo kg de carne será de R\$ 16,70.

Devemos adotar como x_i a variável que desejamos saber, a média e, por w_i , a variável que indica o peso ou o índice de cada elemento. A inversão destas variáveis levará a um resultado incorreto.



5.5 Média de dados agrupados

A média de uma distribuição de frequência para uma amostra é aproximada por:

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

Em que x e f são pontos médios e as frequências de uma classe, respectivamente.

Dicas:	
Em palavras	Em símbolos
Encontre o ponto médio de cada classe	$x = \frac{(\text{limite inferior}) + (\text{limite superior})}{2}$
Encontre a soma dos produtos dos pontos médios e as frequências	$\sum (x.f)$
Encontre a soma das frequências	$n = \sum f$
Encontre a média da distribuição de frequências	$\bar{x} = \frac{\sum (x.f)}{n}$



Observe como encontrar a média da distribuição de frequência:

Use a distribuição de frequência abaixo para aproximar a nota média dos alunos em uma amostra.

Ponto médio das classes x	Frequência f	(x.f)
12	1	12
28	5	140
44	5	220
60	7	420
76	12	912
92	10	920
n = 40		$\sum(x.f) = 2624$

Solução:

$$\bar{x} = \frac{\sum (x.f)}{n} = \frac{2624}{40} = 65,6$$

Então, a média das notas dos alunos foi de aproximadamente 65,6 pontos.

Chegamos ao fim de mais uma aula. Vamos relembrar o que aprendemos!

Resumo

Nesta aula, você pôde verificar que as medidas de posição consistem em resultados estatísticos que representam o ponto central de uma amostra. Entre as medidas de posição, estão a média, a mediana e a moda.

A média consiste em uma medida de posição central. Além de ser a mais conhecida e utilizada, a média talvez seja também a medida de posição mais importante. A média, quando obtida a partir dos dados de uma amostra, é representada por \bar{x} , porém quando obtida a partir de dados de uma população, a média é denotada pela letra grega μ .

A mediana consiste em uma medida de posição central de uma variável. A mediana é o valor que fica no meio da sequência quando os dados são arranjados na ordem ascendente (classificação do menor para o maior). Com um conjunto ímpar de observações, a mediana é o valor do meio. Um número par de observações não tem um valor único no meio. Neste caso, se-



guimos a convenção de definir a mediana como sendo a média dos valores das duas observações do meio.

A moda é uma medida utilizada para descrever o meio ou o centro de um conjunto de dados; é a moda, definida simplesmente como o valor ou categoria que ocorre com a maior frequência e mais de uma vez. A moda não exige cálculo algum, apenas uma contagem, e pode ser determinada tanto para dados numéricos quanto para categóricos.

Atividade de aprendizagem



2. O rol abaixo demonstra o consumo de energia elétrica, em reais, de uma empresa durante 30 meses.

290	321	308	213	318	302	358	286	393	396
352	235	329	333	409	351	325	458	314	181
396	340	356	369	281	386	334	331	348	339

Com base nos dados, calcule a média, a mediana e a moda.

Prezado(a) estudante.

Chegamos ao final da quinta aula de Estatística Aplicada. Nesta aula, o tratamos de medidas de posição, também chamadas de medidas de tendência central e, dentre elas, demonstramos as mais importantes, ou seja, a média, mediana e moda. Acreditamos que o conteúdo exposto trouxe uma oportunidade para você conhecer aspectos da estatística que podem ser úteis tanto na sua profissão como nas tarefas do seu cotidiano.

Na próxima aula, abordaremos as medidas de dispersão. Até lá!



Aula 6 - Medidas de Dispersão

Objetivos:

- Identificar a amplitude de um conjunto de dados
- Reconhecer a variância e o desvio padrão da população e da amostra

Prezado(a) estudante, esta é a sexta aula da disciplina que você está cursando, e o assunto a ser tratado são as medidas de dispersão que consistem na amplitude total, no desvio padrão e na variância.

É importante você ter bem claro que a média, a mediana e a moda podem ser usadas para sintetizar em um único número o que é médio ou aquilo que é típico de distribuição. Quando empregada individualmente, entretanto, qualquer medida de tendência central produz somente um quadro incompleto de um conjunto de dados e, portanto, pode enganar ou distorcer tanto quanto esclarecer. (LEVIN, 2012).

As medidas de dispersão têm por finalidade esclarecer possíveis discrepâncias de um conjunto de dados, o que permite maior confiabilidade nos resultados. Complementarmente, vale observar que, na maioria dos casos, as medidas de dispersão e posição são utilizadas conjuntamente.

Esperamos que, nesta aula, você compreenda bem estes conceitos e que os mesmos sirvam como base para o seu desempenho profissional. Boa aula!!!

6.1 Amplitude total

A medida mais simples é a amplitude do conjunto. A amplitude de um conjunto é a diferença entre as entradas máximas e mínimas no conjunto, ou seja, o maior e menor valor numérico de uma série de dados. Observe a expressão “**valor numérico**”, é isso mesmo, só é possível encontrar a amplitude se os dados forem quantitativos.



Amplitude = (Entrada máxima de dados) – (Entrada mínima de dados)

Existem basicamente duas formas de calcular a amplitude total de uma base de dados agrupados.

- a)** Pelo Ponto Médio das Classes – consiste na diferença entre o ponto médio da última classe e o ponto médio da primeira classe;
- b)** Pelo Limite das Classes – é dado pela diferença entre o limite superior da última classe e o limite inferior da primeira classe.

Exemplo:

Em uma pesquisa sobre a altura dos alunos de uma determinada turma, verificou-se que os alunos possuem as seguintes medidas.

Altura (cm)	Frequência f
146 ┤ 154	4
154 ┤ 162	6
162 ┤ 170	7
170 ┤ 178	12
178 ┤ 186	8
186 ┤ 194	3

Podemos encontrar duas diferentes amplitudes totais para a tabela acima, de acordo com o critério adotado.

- a)** Se adotarmos o critério de amplitude pelo ponto médio das classes, teremos o seguinte:

$$A = 190 - 150 = 40 \text{ cm}$$

- b)** Se adotarmos o critério de amplitude pelos limites das classes, teremos o seguinte:

$$A = 194 - 146 = 48 \text{ cm}$$

A **Amplitude Total** (ou intervalo total), embora seja a medida de dispersão mais simples, apresenta algumas restrições quanto ao seu uso, porque é



muito instável. Essa instabilidade ocorre em virtude de a amplitude total levar em consideração apenas os valores extremos da série, não sendo afetada pela dispersão dos seus valores internos. (CASTANHEIRA, 2010).

6.2 Desvio médio

O Desvio Médio Simples é uma medida da dispersão dos dados em relação à média de uma sequência, o “afastamento” em relação a essa média.

Veja como chegamos à ideia de “dispersão”...

Em 2 grupos de amigos, pesquisou-se as notas que tinham obtido em um trabalho escolar. Cada grupo tinha 5 pessoas e identificamos cada grupo com as letras A e B.

As notas que obtiveram foram:

$$A = \{4; 3; 8; 3; 2\}$$

$$B = \{4; 4; 4; 4; 4\}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{4 + 3 + 8 + 3 + 2}{5} = 4$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{4 + 4 + 4 + 4 + 4}{5} = 4$$

Você observou que as médias em ambos os grupos é 4, ou seja, são iguais!

Pois bem, a média no caso acima não foi suficiente para definir qual dos grupos de alunos teve o melhor desempenho. Portanto, precisamos utilizar outro parâmetro de comparação.

Vamos trabalhar com a ideia de DISPERSÃO em torno da média. A ideia aqui é identificar o quanto cada nota se “desviou” da média.

Como se resolve?

Para comparar o desempenho desses grupos de alunos, podemos começar calculando a média de cada um deles. Como já sabemos, a média em ambos



os casos é 4.

Passaremos para o próximo passo que consiste em calcular o desvio do desempenho de cada aluno em relação à média. Observe que, no grupo A, os alunos 2, 4 e 5 possuem desvios negativos. Como o que desejamos saber é o desvio da média, portanto, o quanto a nota se distancia da média do grupo, devemos sempre considerar os valores como positivos, uma vez que não existem distâncias negativas.



Módulo ou valor absoluto

Calculando o módulo

Considere a reta real:



Chamamos a distância de um ponto da reta à origem (distância do ponto até o zero) de módulo ou valor absoluto.

Assim, a distância do ponto 4 à origem é 4. Dizemos que o módulo de 4 é igual a 4. E representamos $|4| = 4$

Da mesma forma, a distância do ponto -2 à origem é 2, ou seja, o módulo de -2 é 2, pois não há muito sentido em considerarmos distâncias negativas. Assim: $|-2| = 2$, $|3| = 3$

Para resolver o problema dos desvios negativos, podemos utilizar o conceito de módulos, que considera o valor absoluto do número.

Finalmente, somamos os desvios das notas de todos os alunos e dividimos pelo número de alunos.

Grupo A	Aluno 1	Aluno 2	Aluno 3	Aluno 4	Aluno 5
Nota	6	3	8	3	2
Comparação com a média	$6 - 4 = 2$	$3 - 4 = -1$	$8 - 4 = 4$	$3 - 4 = -1$	$2 - 4 = -2$
Desvio (com módulo)	$ 2 = 2$	$ -1 = 1$	$ 4 = 4$	$ -1 = 1$	$ -2 = 2$
Desvio médio	$2 + 1 + 4 + 1 + 2 = \frac{10}{5} = 2$				



Grupo B	Aluno 1	Aluno 2	Aluno 3	Aluno 4	Aluno 5
Nota	4	4	4	4	4
Comparação com a média	$4 - 4 = 2$	$4 - 4 = 0$	$4 - 4 = 0$	$4 - 4 = 0$	$4 - 4 = 0$
Desvio (com módulo)	$4 - 4 = 0$	$4 - 4 = 0$	$4 - 4 = 0$	$4 - 4 = 0$	$4 - 4 = 0$
Desvio médio	$0 + 0 + 0 + 0 + 0 = \frac{0}{5} = 0$				

Como se pode perceber, pelo exemplo acima, o desvio médio nada mais é do que a média aritmética dos desvios dos valores em relação à média, considerados em valor absoluto.

Quando desejamos analisar a dispersão ou o afastamento dos valores de uma série em relação à média, é conveniente analisar essa dispersão de cada um dos valores, sem exceção.

O desvio médio é representado D_m e pode ser calculado pela fórmula:

$$D_m = \frac{\sum |x_i - \bar{x}| \cdot f_i}{n}$$

Onde $|x_i - \bar{x}|$ é o módulo de cada desvio em relação à média, i varia de 1 a n observações e N é igual $\sum f_i$ (somatório das observações).

É muito importante ressaltar que a diferença entre o valor de uma observação x_i pela média do grupo é representada dentro módulo $|x_i - \bar{x}|$, conforme se observa. Portanto, módulo do desvio de cada elemento em relação à média, porque de acordo com a propriedade de média aritmética, a soma de todos os desvios em relação à média é sempre igual a zero. Entretanto, trabalhando com o módulo (que retorna o valor sempre positivo), este problema é resolvido.



x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $
2	$2 - 5 = -3$	3
3	$3 - 5 = -2$	2
5	$5 - 5 = 0$	0
6	$6 - 5 = 1$	1
9	$9 - 5 = 4$	4
		$\sum x_i - \bar{x} = 10$

Dada a fórmula do desvio médio,

$$D_m = \frac{\sum |x_i - \bar{x}| \cdot f_i}{n}$$



Como cada valor de x_i ocorre uma única vez, então f_i como 1 é igual a 1. Logo, teremos que:

$$Dm = \frac{10}{5} = 2$$

É importante ressaltar, entretanto, que para dados agrupados em classes ou intervalos, substitui-se x_i na fórmula, pelo ponto médio de cada classe.

6.3 Variância (S^2)

A variância é outra medida de identificação da dispersão ou variabilidade dos dados de um determinado conjunto que leva em consideração cada valor em uma distribuição. Podemos ficar tentados a somar todos os desvios, entretanto a soma dos desvios reais, $\sum(x_i - \bar{x})$, é sempre zero, uma vez que os desvios maiores e menores se cancelam.

A exemplo do conceito de desvio médio, no qual se utiliza da função modular para tornar os desvios positivos, na variância, este problema é resolvido elevando-se os desvios reais da média ao quadrado e, em seguida, somando-os, o que é denotado pela seguinte fórmula $\sum(x_i - \bar{x})^2$.

Da mesma forma que no caso dos desvios médios, tendo somado os quadrados dos desvios da média, divide-se essa soma por N, sendo N o número de dados observados. O resultado é a média do quadrado dos desvios, que é conhecida como variância e simbolizada por S^2 .

Vejamos a denotação de variância,

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

Onde

s^2 = variância

$\sum(x_i - \bar{x})^2$ = soma dos quadrados em relação à média

n = número de observações (tamanho da amostra)

Entenda melhor com este exemplo: a empresa Tradicional Ltda deseja avaliar a variância em seu volume vendas com base no faturamento dos últimos



6 meses. Dados os valores mensais abaixo em milhares de reais, calcule a variância.

15 16 17 18 19 20 23

Solução:

Passo I -

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$

$$S^2 = \frac{(15 - 18,29)^2 + (16 - 18,29)^2 + (17 - 18,29)^2 + (18 - 18,29)^2 + (19 - 18,29)^2 + (20 - 18,29)^2 + (23 - 18,29)^2}{7}$$

$$S^2 = \frac{(-3,286)^2 + (-2,286)^2 + (1,286)^2 + (0,286)^2 + (0,714)^2 + (1,714)^2 + (4,714)^2}{7}$$

$$S^2 = \frac{10,796 + 5,224 + 1,653 + 0,082 + 0,51 + 2,939 + 22,224}{7}$$

$$S^2 = \frac{43,418}{7} = 6,204$$

Portanto, o valor da variância no volume de vendas é de 6,204 mil reais.

6.4 Desvio padrão (S)

Como você deve ter observado, para calcular a variância, o desvio da média foi elevado ao quadrado $\sum (x_i - \bar{x})^2$. Isso significa dizer que a unidade de medida foi alterada, o que torna a variância um tanto quanto difícil de ser interpretada. (LEVIN, 2012).

Para colocar a medida na perspectiva certa, é necessário retornar à sua unidade original de medida, que é dada pela raiz quadrada da variância.

Vejamos.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N}$$



$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Isso nos dá o desvio padrão, uma medida de variabilidade obtida somando-se os quadrados dos desvios em relação à média, dividindo-se por N e tomando-se a raiz quadrada.

Para melhor compreensão, observe o exemplo: vamos tomar por base o conjunto de dados abaixo, representado pela idade dos jovens matriculados em um curso profissionalizante.

15 16 17 18 19 20 23

Solução:

Passo 1 – Calcule a média para a distribuição

x	
15	$x = \frac{\sum x}{n}$ $= \frac{126}{7}$ $= 18$
16	
17	
18	
19	
20	
21	
$\sum X = 126$	

Passo 2 – Subtraia o dado bruto da média

X	\bar{x}	$x - \bar{x}$
15	18	$15 - 18 = -3$
16	18	$16 - 18 = -2$
17	18	$17 - 18 = -1$
18	18	$18 - 18 = 0$
19	18	$19 - 18 = 1$
20	18	$20 - 18 = 2$
21	18	$21 - 18 = 3$

Passo 3 – Eleve ao quadrado cada desvio para, então, obter a soma dos



quadrados dos desvios.

X	\bar{x}	$x - \bar{x}$	$(x - \bar{x})^2$
15	18	$15 - 18 = -3$	$(15 - 18)^2 = 9$
16	18	$16 - 18 = -2$	$(16 - 18)^2 = 4$
17	18	$17 - 18 = -1$	$(17 - 18)^2 = 1$
18	18	$18 - 18 = 0$	$(18 - 18)^2 = 0$
19	18	$19 - 18 = 1$	$(19 - 18)^2 = 1$
20	18	$20 - 18 = 2$	$(20 - 18)^2 = 4$
21	18	$21 - 18 = 3$	$(21 - 18)^2 = 9$
			$\sum (x_i - \bar{x})^2 = 28$

Passo 4 – Utilizando-se da fórmula de desvio padrão, dada por $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$, aplique o resultado do somatório de $\sum (x_i - \bar{x})^2 = 28$ encontrado, dividindo-o por n que é igual a sete. Em seguida, obtenha a raiz quadrada, como segue:

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

$$s = \sqrt{\frac{28}{7}}$$

$$s = \sqrt{4}$$

$$s = 2$$

Agora podemos dizer que o desvio padrão é de 2 anos para os 7 alunos matriculados no curso profissionalizante. Isso significa que os dados nessa distribuição desviam da média por 2 anos. Por exemplo, na distribuição, apenas um aluno de 15 e outro de 21 anos ficaram fora do desvio padrão, ou seja, abaixo de 16 e acima de 20 anos.

6.4.1 Desvio padrão a partir de dados brutos

Há um método mais simplificado de calcular o desvio padrão. Esse método funciona diretamente com os dados brutos. Conforme veremos:

$$s^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

$$s = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

onde



$\sum x^2$ = soma dos quadrados dos dados brutos. É importante observar que cada dado bruto é elevado ao quadrado, para então serem somados.

n = número de observações (tamanho da amostra)

\bar{x}^2 = Quadrado da média

Continue seguindo o exemplo, utilizando-se do mesmo banco de dados de alunos matriculados do exemplo anterior, agora calcularemos o desvio padrão a partir dos dados brutos.

15 16 17 18 19 20 23

Solução:

Passo 1 – Eleve cada dado bruto ao quadrado e some os resultados.

X	x^2
15	225
16	256
17	289
18	324
19	361
20	400
21	441
	$\sum x^2 = 2296$

Passo 2 – Obtenha a média e eleve ao quadrado

X	
15	$\bar{x} = \frac{\sum x}{n}$ $= \frac{126}{7}$ $= 18$ $\bar{x}^2 = 324$
16	
17	
18	
19	
20	
21	
$\sum X = 126$	

Passo 3 – Utilizando a formula de desvio padrão com base em dados brutos.



$$s = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$s = \sqrt{\frac{\sum 2296}{7} - 324}$$

$$s = \sqrt{328 - 324}$$

$$s = \sqrt{4}$$

$$s = 2$$

Você observou que a fórmula do desvio padrão com base em dados brutos teve exatamente o mesmo resultado da fórmula original?

6.5 Desvio padrão em uma distribuição de frequência agrupada

Na aula anterior, vimos que as medidas de posição podem ser calculadas a partir de um conjunto de dados agrupados em forma de uma distribuição simples. De forma similar, veremos como calcular a variância e o desvio padrão.

Assim como é possível encontrar uma medida de tendência central para dados de uma distribuição de frequência agrupada, o caso do desvio padrão é bastante similar, como vamos ver a seguir.

É importante ressaltar que, independentemente da abordagem utilizada, seja esta de frequência simples ou agrupada, os valores para a média e s serão os mesmos.

O desvio padrão pode ser aproximado a partir de uma distribuição de frequência agrupada por meio da adaptação das fórmulas de uma distribuição de frequências simples que substitua os valores dos dados x pelos pontos médios dos intervalos de classe f, especificamente agrupados. (LEVIN, 2012).

$$\bar{x} = \frac{\sum x.f}{n}$$

$$s = \sqrt{\frac{\sum f.x^2}{n} - \bar{x}^2}$$



Dada a distribuição de frequência abaixo, vamos calcular o desvio padrão.

Altura do aluno	x	f	$f \cdot x$	$f \cdot x^2$
1,46 1,54	1,5	4	6	9
1,54 1,62	1,58	6	9,48	14,9784
1,62 1,70	1,66	7	11,62	19,2892
1,70 1,78	1,74	12	20,88	36,3312
1,78 1,86	1,82	8	14,56	26,4992
1,86 1,94	1,9	3	5,7	10,83
		$n = 40$	$\sum f \cdot x = 68,24$	$\sum f \cdot x^2 = 116,928$

Usando os conceitos de média e desvio de padrão de frequência agrupada, temos:

Cálculo da média:

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

$$\bar{x} = \frac{68,24}{40}$$

$$\bar{x} = 1,706$$

Cálculo do desvio padrão:

$$s = \sqrt{\frac{\sum f \cdot x^2}{n} - \bar{x}^2}$$

$$s = \sqrt{\frac{116,928}{40} - 1,706^2}$$

$$s = \sqrt{2,9232 - 2,91044}$$

$$s = 0,112978$$

Por arredondamento, temos que $s = 0,11$, ou seja, o desvio padrão é igual a 0,11.

Como a altura dos alunos está expressa em metros, podemos convertê-la em centímetros, bastando para isso multiplicá-la por 100. Assim, pode-se afirmar que o desvio padrão da altura dos alunos observado é de 11 centímetros.



De acordo com Levin (2012), é importante lembrar que essas são apenas aproximações, pois os pontos médios servem como representantes para todos os valores observados nos respectivos intervalos de classe. Ele sugere que, quando os dados observados estiverem disponíveis, é recomendável que não seja utilizada a fórmula de frequência agrupada.

Chegamos ao final de nossa aula. Vamos rever o que aprendemos?

Resumo

Nesta aula, mostramos que a amplitude de um conjunto é a diferença entre as entradas máximas e mínimas no conjunto, ou seja, o maior e o menor valor numérico de uma série de dados. Observou-se que só é possível encontrar a amplitude se os dados forem quantitativos.

Você pôde verificar que o Desvio Médio Simples é uma medida da dispersão dos dados em relação à média de uma sequência, o “afastamento” em relação a essa média.

Foi informado também que variância é outra medida de identificação da dispersão ou variabilidade dos dados de um determinado conjunto que leva em consideração cada valor em uma distribuição.

Por último, vimos que o desvio padrão consiste na medida de verificação da dispersão da média. Este pode ser aproximado a partir de uma distribuição de frequência agrupada por meio da adaptação das fórmulas de uma distribuição de frequências simples que substitua os valores dos dados x pelos pontos médios dos intervalos de classe f .

Atividade de aprendizagem

1. Dado o conjunto de dados a seguir, calcule a variância e o desvio padrão.



290 321 308 213 318 302 358 286 393 396





Prezado(a) estudante,

Você finalizou a 6ª aula e já venceu a metade do conteúdo desta disciplina. Mas ainda faltam informações que você precisa juntar às já obtidas para, futuramente, atuar como técnico em finanças. Na próxima aula, trataremos de tabelas e gráficos estatísticos. Preparado(a) para prosseguir?



Aula 7 - Tabelas e Gráficos

Objetivos:

- Identificar os termos de uma tabela.
- Reconhecer e exemplificar os diferentes tipos de tabelas.
- Distinguir e construir os principais tipos de gráficos.

Prezado (a) estudante

Esta é a sétima aula de Estatística Aplicada, trataremos de tabelas e gráficos que são utilizados na apresentação de resultados da pesquisa estatística. Esperamos que esta aula lhe proporcione o entendimento necessário para utilizá-los, bem como a habilidade para ler gráficos e tabelas em jornais, revistas e relatórios. Boa aula!!!

A assimilação das informações geradas pelos dados de experimentos é mais fácil quando as mesmas estão dispostas em tabelas ou em gráficos. Uma tabela é um arranjo sistemático de dados numéricos dispostos de forma (colunas e linhas) a favorecer a comparação. A apresentação em forma de tabela deve expor os dados de modo fácil e que deixe a leitura mais rápida.

Os gráficos, por sua vez, são recursos mais utilizados para facilitar a leitura de dados estatísticos. Tem grande finalidade em relatórios, em apresentações e noticiários.

7.1 Tabelas

Nesta seção, iremos identificar os termos de uma tabela.

Bons estudos!



A tabela é uma forma não discursiva de apresentar informações, das quais o dado numérico se destaca como informação central, na sua forma, identificam-se espaços e elementos.

As tabelas consistem em um tipo de texto técnico, por meio do qual quem pesquisa ou elabora o texto é capaz de apresentar um conjunto de informações de um determinado assunto ou evento, anteriormente coletadas.

E, como todo texto, a boa compreensão por quem lê será possível e até mesmo facilitada, quando o(a) autor(a) observar alguns cuidados básicos, porém necessários para a elaboração de uma boa tabela. Além de uma boa organização dos dados, é necessário escolher o tipo de tabela mais adequado para a informação a ser comunicada.

No Brasil, a definição das normas para a elaboração de uma tabela é dada pela Associação Brasileira de Normas Técnicas (ABNT) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Descrevemos as mais essenciais.

Uma boa tabela, portanto, atende aos seguintes requisitos:

- Conter todas as informações para uma completa compreensão do texto, dispensando outras consultas, apresentadas de maneira simples e objetiva e, de preferência, em uma página;
- Quando são apresentadas intercaladas no texto, devem estar próximas e logo após o trecho em que são citadas;
- Podem ser apresentadas anexas ao texto, quando o volume de tabelas for grande;
- Devem ser alinhadas preferencialmente às margens laterais do texto e, quando pequenas, devem ser centralizadas;
- Devem ser dispostas de maneira a evitar que sua leitura tenha sentido diferente do normal;
- Deve apresentar, a maior parte das células, a informação de qual dado se trata, evitando-se as informações desnecessárias.

Em síntese, as **tabelas** trazem dados estatísticos numéricos; os lados, esquerdo e direito, são abertos, as partes superior e inferior são fechadas e não se colocam traços horizontais e verticais para separar os números, diferentemente dos quadros que possuem os quatro lados fechados com traços horizontais e verticais para separar as informações. Devem apresentar a fonte, acrescidos de nota, se necessário. Esta nota deve registrar informações ou comentários elucidativos.

O título deve ser digitado acima da tabela, obedecendo a margem esquerda; somente as iniciais da frase e dos nomes próprios são escritas em letras maiúsculas. O título é digitado após a palavra Tabela, separado por hífen.

Portanto, na tabela estatística, são considerados elementos essenciais os seguintes itens:

- Título: informa o conteúdo do corpo da tabela, de maneira completa, concisa e indicando a natureza do fato estudado;
- Corpo: é o conjunto de linhas e colunas que contêm as séries verticais e horizontais de informação;
- Fonte: é o indicativo, no rodapé da tabela, da entidade responsável pela informação.

Vamos ver um exemplo?

Coluna indicadora Cabeçalho Título

Tabela 1 - Micro e pequenas empresas de comércio e serviços, pessoal ocupado Brasil - 2001

Faixas de pessoal ocupado	Número de micro e pequenas empresas	Pessoal ocupado em 31.12.2001
Até 5 pessoas ocupadas	1 536 272	2 958 944
De 6 a 19 pessoas ocupadas	439 719	2 668 873
20 ou mais pessoas ocupadas	68 574	1 662 853
Total de pessoas ocupadas	2 044 565	7 290 670

Linhas

Fonte: IBGE (2001).
Rodapé

Está claro até aqui?

Conseguiu identificar os termos de uma tabela?

Caso não tenha ficado claro, releia os itens anteriores e visualize a figura da tabela para entendê-la melhor.

Muito bem!



Agora que conseguimos identificar os termos de uma tabela, vamos conhecer os diferentes tipos de tabelas. Existem diversos tipos de tabelas, dentre as quais se destacam: as tabelas de série histórica, série geográfica e série específica ou categórica.

7.1.1 Série histórica

A série histórica consiste em que o elemento variável é o fator cronológico, como no exemplo abaixo, em que este é representado em anos. Essas séries também são denominadas temporais ou cronológicas.

Tabela 8 - Crédito concedido ao microempreendedor no Brasil	
Mês	Contratos no mês
mai/12	239.621
jun/12	249.488
jul/12	236.938
ago/12	267.643
set/12	260.950
out/12	429.234

Fonte: Adaptado pelo autor a partir de dados do BACEN (2012)

Observa-se, nesta tabela, que se deseja demonstrar a quantidade de contratos realizados em cada mês, logo, pode-se afirmar que o elemento variável são os meses, daí a denominação de série histórica ou cronológica.

7.1.2 Série específica ou categórica

A série específica ou categórica descreve os valores da variável em determinado tempo e local, discriminados segundo especificações ou categorias.

Tabela 9 - Atividade dos dirigentes das Cooperativas de Rondônia - 2009	
Atividade	Dirigente
Empresário	57
membros de outros conselhos ou cooperativas	14
Empregado de outras empresas ou entidades	5
Diretor de entidade de classe	6
Diretor de sindicato	1
Outros	15

Fonte: Elaborado pelo autor a partir de dados do SESCOOP/RO (2009)

Observa-se, nesta tabela, que permanece fixo o indicador Dirigentes de cooperativas no Estado de Rondônia. No entanto, o fenômeno estudado, que



são as atividades dos dirigentes de cooperativas, está sofrendo variação em diversas categorias, daí o nome da série categórica.

7.1.3 Série geográfica

Descrevem os valores da variável, em determinado tempo, discriminados segundo a localização geográfica da variável.

Tabela 10 - População da Região Norte - 2010

Atividade	Número de habitantes
Acre	733.559
Amapá	669.526
Amazonas	3.483.985
Pará	7.581.051
Tocantins	1.383.445
Rondônia	1.562.409
Roraima	450.479

Fonte: IBGE (2010)

Verifica-se, facilmente, nesta tabela, que são fixos o fenômeno estudado (número de habitantes) e a época da pesquisa (2010), embora o elemento local sofra variação, caracterizando, por isso, esta série estatística como série geográfica. Este tipo de série é também conhecido como: espacial; territorial ou de localização.

7.1.4 Tabelas de dupla entrada

Existem também as de dupla entrada. Este tipo de tabela é utilizado quando há necessidade de apresentar diferentes situações ao mesmo tempo, ou quando se pretende comparar dados. As tabelas de dupla entrada são compostas de séries estatísticas resultantes da combinação das séries estatísticas temporais, geográficas ou específicas.

Tabela 11 - Quantidade, perfil e participação dos associados das Cooperativas em Rondônia - 2009

Perfil dos associados	Cooperativas não filiadas na OCB	Cooperativas filiadas na OCB
Associados ativos	78%	72%
Associadas (sexo feminino)	22%	18%
Associados presentes na última assembléia geral	50%	17%

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

Você observou que na tabela acima é possível comparar o perfil do associado



com o tipo de cooperativa? Essa é uma das grandes aplicabilidades deste tipo de série.

Está tudo bem até aqui?

Percebeu como é fácil reconhecer e exemplificar os diferentes tipos de tabela?

Pois bem, passaremos agora para a próxima sessão, na qual conheceremos os gráficos estatísticos.

7.2 Gráficos

Em estatística, o gráfico é a tentativa de expressar visualmente estatísticas simplificadas, matemáticas ou não, de algum dado ou valor obtido, facilitando a compreensão. Os gráficos são recursos visuais muito utilizados para facilitar a leitura e a compreensão de informações sobre fenômenos e processos naturais, sociais e econômicos.

No cotidiano, jornais, revistas e livros, além de telejornais e programas educativos mostram o quanto esse recurso é explorado pelos meios de comunicação.

O gráfico é uma representação com forma geométrica construída de maneira exata e precisa a partir de informações numéricas obtidas através de pesquisas e organizadas em uma tabela.



A elaboração de gráficos é algo bastante simples, quando utilizamos um software. Caso deseje conhecer como elaborar gráficos no Excel, acesse dois vídeos disponíveis em: <<http://youtu.be/Szbxu3BPojY>> e <<http://youtu.be/xexGAIE5snI>>.

Existem vários tipos de gráficos e os mais utilizados são os de colunas, os de linhas e os circulares.

7.2.1 Gráficos de colunas

O gráfico de colunas é composto por duas linhas ou eixos, um vertical e outro horizontal. No eixo horizontal, são construídas as colunas que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade. Essa intensidade é indicada pelo eixo vertical.

As colunas devem sempre possuir a mesma largura e a distância entre elas deve ser constante.

Você já conhece este tipo de tabela, agora procure identificar os dados da

planilha a partir da representação dada no gráfico de coluna abaixo.

Tabela 12 - Atividade dos dirigentes das Cooperativas de Rondônia - 2009

Atividade	Dirigente
Empresário	57
membros de outros conselhos ou cooperativas	14
Empregado de outras empresas ou entidades	5
Diretor de entidade de classe	6
Diretor de sindicato	1
Outros	15

Fonte: Elaborado pelo autor a partir de dados do SESCOOP/RO (2009)

Você deve ter percebido que cada uma das atividades (elementos) da planilha e seus respectivos valores estão representados no gráfico. É importante salientar que neste tipo de gráfico, podem ser usados tanto os dados absolutos da pesquisa, bem como sua conversão em percentuais.

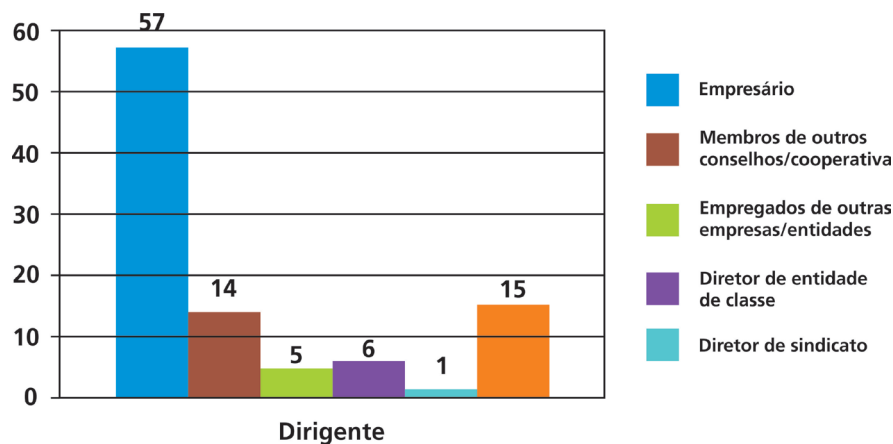


Gráfico 3 - Atividade dos dirigentes das Cooperativas Rondonienses

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

7.2.2 Gráficos de barras

O gráfico de barras é composto por duas linhas ou eixos, um vertical e outro horizontal. No eixo vertical são construídas as barras que representam a variação de um fenômeno ou de um processo de acordo com sua intensidade. Essa intensidade é indicada pelo eixo horizontal. As barras devem sempre possuir a mesma largura e a distância entre elas deve ser constante.



Tabela 13 - Critérios das Cooperativas Rondonienses para a realização de investimento

Critérios utilizados	%
Experiência em Métodos Gerenciais	27%
Pesquisa de Mercado	23%
Orientação da Central	17%
Conhecimento da Concorrência	16%
Contatos com Fornecedores	11%
Nenhum/Não Investe	38%
Outros	2%

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

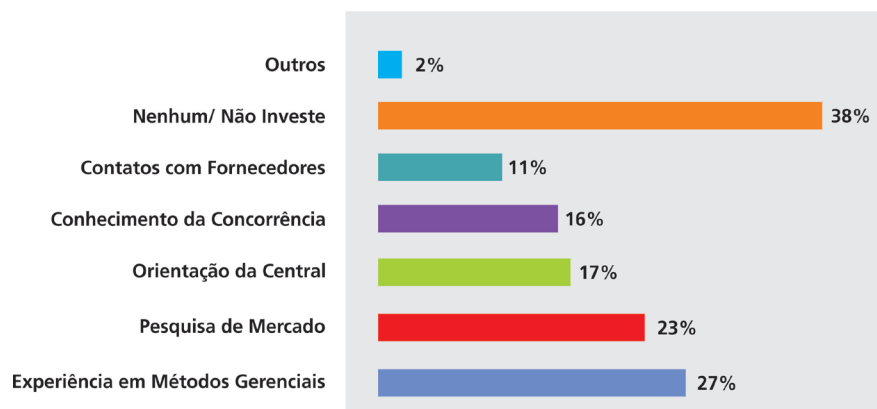


Gráfico 4 - Critérios das Cooperativas Rondonienses para a realização de investimento

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

A representação dos dados no gráfico de barras é muito similar ao gráfico de colunas, como você deve ter percebido.

7.2.3 Gráficos de setores

Os gráficos de setor ou pizza são representados por círculos divididos proporcionalmente de acordo com os dados do fenômeno ou do processo a ser representado. Os valores são expressos em números ou em percentuais (%).

Tabela 14 - Participação por gênero na direção das cooperativas de Rondônia

Gênero do diretor(a)	%
Masculino	87,50
Feminino	12,50

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

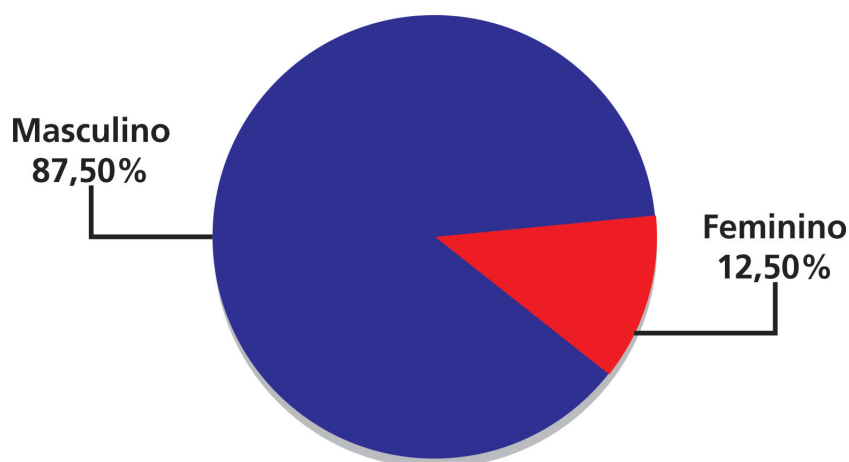


Gráfico 5 - Participação por gênero na direção das cooperativas de Rondônia

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)



É muito importante, que antes de optar pela utilização deste gráfico, se verifique que o somatório seja igual a 100%, caso contrário, os resultados serão distorcidos.



Figura 26

Fonte: ilustradora

Observe, na tabela, que o somatório dos grupos A, B e C é maior do que 100%.

Tabela 15 - Tabela de exemplo	
Perfil dos associados	Exemplo
Grupo A	78%
Grupo B	22%
Grupo C	50%
Total	150%

Fonte: Elaboração própria

Vamos analisar como seria a representação no gráfico de setores?

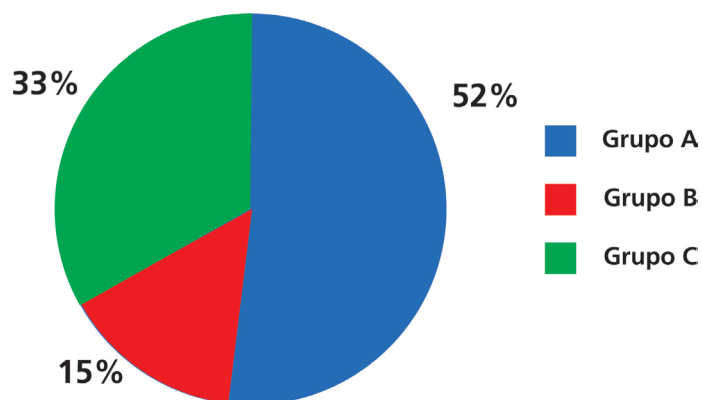


Gráfico 6 - Gráfico de exemplo

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

Os dados são os mesmos, mas você observou que os valores percentuais no gráfico estão menores do que os da tabela?

Apesar de serem os mesmos dados, ao incluir os dados no gráfico de setores, este considerou 150 como 100%.

De forma similar, dividiu-se também $78/150 = 0,52$ ou 52%, $50/150 = 0,33$ ou 33% e $22/150 = 0,15$ ou 15%.

Portanto, sempre se lembre de verificar o seu conjunto de dados, antes de escolher o gráfico para melhor representá-lo.

7.2.4 Gráfico de linha

O gráfico de linha é composto por dois eixos, um vertical e outro horizontal, e por uma linha que mostra a evolução de um fenômeno ou processo, isto é, o seu crescimento ou diminuição no decorrer de determinado período.

Tabela 16 - Formação dos colaboradores de cooperativas de Rondônia

Formação	Colaboradores
Não terminou o ensino fundamental	49
Ensino fundamental	101
Ensino médio	390
Superior completo	209
Pós-graduação	72

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

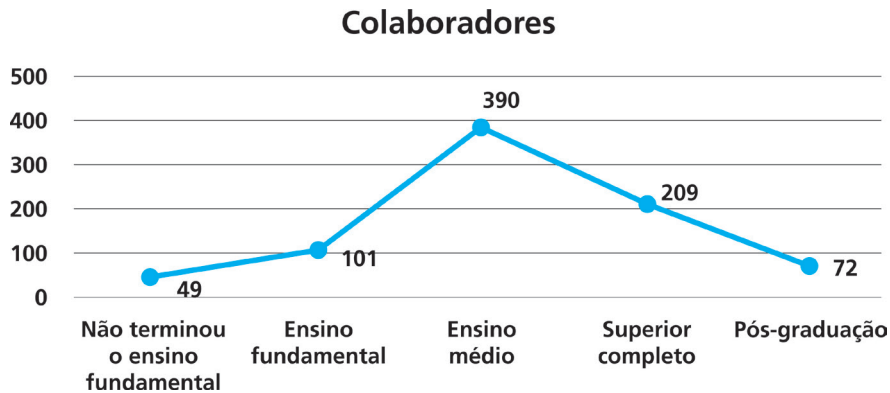


Gráfico 7 - Formação dos colaboradores de cooperativas de Rondônia

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

7.2.5 Gráficos de múltiplas entradas

De forma análoga às tabelas, também é possível representar, ao mesmo tempo, dois ou mais fenômenos, bem como comparar valores. Nestes casos, é necessário utilizar um gráfico múltiplo de colunas ou de barras.

No exemplo a seguir, mostraremos somente o uso do gráfico de barras, mas pode ser aplicado o mesmo conhecimento para a montagem de um gráfico de colunas.

Tabela 17 - Quantidade, perfil e participação dos associados das Cooperativas Rondônia - 2009

Perfil dos associados	Cooperativas não filiadas na OCB	Cooperativas filiadas na OCB
Associados ativos	78%	72%
Associadas (sexo feminino)	22%	18%
Associados presentes na última assembleia geral	50%	17%

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

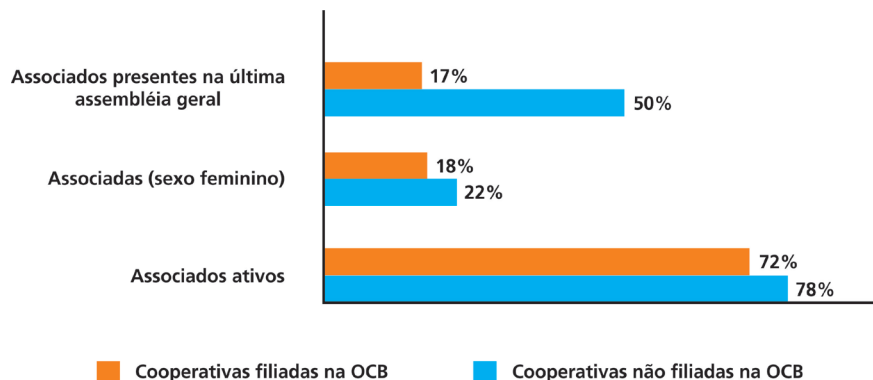


Gráfico 8 - Perfil dos associados

Fonte: Elaborado pelo autor (SESCOOP/RO, 2009)

Resumo

Nesta aula, você pôde constatar que as **tabelas** consistem em um tipo de texto técnico, por meio do qual quem pesquisa ou elabora o texto é capaz de mostrar a quem lê um conjunto de informações de um determinado assunto ou evento, anteriormente coletadas. As tabelas são classificadas em: geográfica, histórica, categórica e de dupla entrada.

Pôde perceber também que, para possibilitar uma boa compreensão, é importante escolher o tipo de tabela que seja mais adequada para o tipo de informação a ser comunicada. A boa organização dos dados também é imprescindível.

Por último, teve a oportunidade de verificar que os gráficos são utilizados para expressar visualmente estatísticas simplificadas, matemáticas ou não, de algum dado ou valor obtido, a fim de facilitar a compreensão. São muito utilizados para facilitar a leitura e a compreensão de informações sobre fenômenos e processos naturais, sociais e econômicos.



Atividade de aprendizagem

1. Dada a tabela abaixo, retirada da atividade da página 69, construa um gráfico representativo da frequência de renda.



Nível salário	Frequência f	Ponto Médio	Frequência relativa	Frequência acumulada
502,5 ┤ 657,5	3	580	0,12	0,12
657,5 ┤ 812,5	8	735	0,32	0,44
812,5 ┤ 967,5	9	890	0,36	0,8
967,5 ┤ 1122,5	4	1045	0,16	0,96
1122,5 ┤ 1277,5	1	1200	0,04	1
	25		1	

Prezado(a) estudante.

Chegamos ao final de mais uma aula de Estatística Aplicada. Nesta aula, você pôde estudar gráficos e tabelas, que consistem em recursos representativos utilizados para facilitar a leitura de dados e resultados estatísticos.

Na próxima aula, abordaremos o tema correlação. Não deixe de realizar as atividades de aprendizagem e de acessar os sites indicados. Essas ações, certamente, contribuem para a sua aprendizagem.



Guia de Soluções

Aula 1

1. Padrão de resposta. Espera-se que o estudante seja capaz de precisar os conceitos fundamentais da estatística e destacar exemplos da aplicabilidade em uma ou mais situações reais do ambiente empresarial.

2. Resposta: A sequência correta é C, B, D e A.

Aula 2

Atividade 1

a) O autor utilizou o conceito de taxa conforme se verifica na expressão “a taxa de pobreza extrema entre a população de 0 a 15 poderia ter caído 0,6%”.

b) Como as taxas são dadas por um coeficiente multiplicado por 10, 100 ou 1000, para tornar o resultado mais compreensivo, pode-se afirmar que a taxa de pobreza extrema entre a população de 0 a 15 está relacionada ao coeficiente de pobreza extrema em relação à população brasileira total com idade entre 0 e 15 anos. Este coeficiente pode ser denotado por:

$$\text{Coeficiente de pobreza extrema} = \frac{\text{nº de brasileiros de 0 a 15 na pobreza extrema}}{\text{nº total de brasileiros de 0 a 15}}$$

Aula 3

1. Padrão de resposta: Espera-se que o aluno seja capaz de descrever em que consiste o método estatístico, sua finalidade, bem como o que é realizado em cada uma das etapas.

2. Resposta: 4, 2, 3, 1 e 5

Aula 4

1. Resposta:

Estatística Descritiva: Em uma pesquisa foram entrevistadas 200 mulheres e 200 homens da Cidade X; verificou-se que, entre os homens, a média de leitura foi de 2 livros e, entre as mulheres, foi de 3 livros.



Estatística Inferencial: As mulheres leem em média 1 livro a mais, por ano, do que os homens.

2. Resposta:

Número de classe: $k = 5$

Nível salarial	Frequência f	Ponto Médio	Frequência relativa	Frequência acumulada
502,5 657,5	3	580	0,12	0,12
657,5 812,5	8	735	0,32	0,44
812,5 967,5	9	890	0,36	0,8
967,5 1122,5	4	1045	0,16	0,96
1122,5 1277,5	1	1200	0,04	1
	25		1	

Aula 5

1. Resposta:

Média = 2,22

Mediana = 2,25

2. Resposta:

Média = 331,73

Mediana = 333,5

Moda = 396

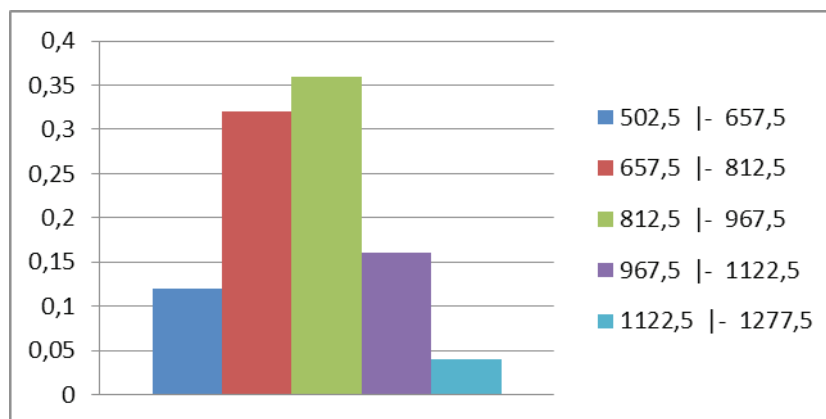
Aula 6

1. Resposta: O desvio padrão é igual a 54,27.



Aula 7

1. Resposta: O gráfico abaixo demonstra a classificação das pessoas por nível salarial.





Referências

BONAFINI, Fernanda Cesar. **Estatística**. São Paulo: Pearson Education do Brasil, 2012.

BISQUERRA, Rafael. **Introdução à estatística**: enfoque informático com o pacote estatístico SPSS. Porto Alegre: Artmed, 2004.

CABRAL, Álvaro. **Dicionário técnico de psicologia**. 14. ed. – São Paulo: Cultrix, 2006.

CASTANHEIRA, Nelson Pereira. **Estatística aplicada a todos os níveis**. 5. ed. ver. e atual. – Curitiba: Ibpx, 2010.

FREUND, John E. **Estatística aplicada**: economia, administração e contabilidade. Tradução Claus Ivo Doering. 11. Ed. – Porto Alegre: Bookman, 2006.

IPEA. **Brasil Carinhoso pode baixar pobreza extrema infantil a 0,6%**. Disponível em: <http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=16545&catid=10&Itemid=9> Acesso em: jan.2013.

LARSON, Ron; FARBER, Betsy. **Estatística aplicada**. Tradução Luciane Ferreira Pauleti Vianna. 4. ed. São Paulo: Pearson Prentice Hall, 2010.

LEVIN, Jack; FOX, James Alan. FORDE, David R. **Estatística para ciências humanas**. Tradução Jorge Ritter; revisão técnica Fernanda Bonafini. 11. ed. – São Paulo: Pearson Education do Brasil, 2012.

MARTINS, Gilberto de Andrade. **Metodologia da investigação científica para ciências sociais aplicadas**. São Paulo: atlas, 2007.

MCCLAVE, James T.; BENSON, P. George; SINCICH, Terry. **Estatística para administração e economia**. Tradução Fabrício Pereira Soares e Fernando Sampaio Filho; revisão técnica Galo Carlos Lopez Noriega. – São Paulo: Pearson Prentice Hall, 2009.

MEMORIA, José Maria Pompeu. **Breve história da estatística**. Brasília, DF: Embrapa Informação Tecnológica, 2004. Disponível em: <http://www.im.ufrj.br/~lpbraga/prob1/historia_estadistica.pdf> Acesso em: mar 2013.

MINISTÉRIO DA EDUCAÇÃO. **Caderno de questões do exame nacional do ensino médio**. INEP, 2011.

MORETTIN, Luiz Gonzaga. **Estatística básica**: probabilidade e inferência, volume único. São Paulo. Pearson Prentice Hall, 2010.

SEBRAE. **Curso Iniciando um Pequeno Grande Negócio**. 2005. Disponível em: <<http://www.sebraemg.com.br/arquivos/parasuaempresa/planodemercado/mercado.pdf>> Acesso em: jan.13.

VIEIRA, Maria Teresa Ferreira de Amorim da Silva. **Amostragem**. Portugal: Universidade



de Aveiro, 2008. Disponível em: <<http://ria.ua.pt/bitstream/10773/2909/1/2009000495.pdf>>. Acesso: abr. 13.

WAISELFISZ, Julio Jacobo. **Mapa da violência**. 2013. Disponível em: <http://mapadaviolencia.org.br/pdf2013/MapaViolencia2013_armas.pdf>. Acesso: abr. 13.

WALPOLE, Ronald E.; MYERS, Raymond H.; MYERS; Keying Ye. **Probabilidade e estatística para engenharia e ciências**. São Paulo: Pearson Prentice Hall, 2009.

WIKIPÉDIA. **Arte rupestre**. <http://pt.wikipedia.org/wiki/Arte_rupestre>. Acesso em: mar 2013.



Obras Consultadas

BONAFINI, Fernanda Cesar. **Matemática**. São Paulo: Pearson Prentice Hall, 2012.

CASTANHEIRA, Nelson Pereira. **Métodos quantitativos**. Curitiba: Ibpx, 2008.

JACQUES, Ian. **Matemática para economia e administração**. São Paulo: Pearson Prentice Hall, 2010.

NEUFELD, John L. **Estatística aplicada à administração usando Excel**. São Paulo: Prentice Hall, 2003.

Bibliografia Básica

BONAFINI, Fernanda Cesar. **Estatística**. São Paulo: Pearson Education do Brasil, 2012.

FONSECA, Jairo Simon da. **Curso de Estatística**. 6. ed. São Paulo: Atlas, 2006.

FREUND, John E. **Estatística aplicada: economia, administração e contabilidade**. Tradução Claus Ivo Doering. 11. Ed. — Porto Alegre: Bookman, 2006.

MARTINS, Gilberto de Andrade. **Estatística geral e aplicada**. São Paulo: Atlas, 2008.



Currículo do Professor-autor



Vilmar dos Santos Alves, graduado em Licenciatura Plena em Matemática e Mestre em Administração pela Universidade Federal de Rondônia. Pós-graduado em MBA em Finanças, Controladoria e Auditoria pela Faculdade São Lucas. Atualmente, ocupa a função de coordenador de filial na Gerência de Desenvolvimento Urbano e Rural Porto Velho da Caixa Econômica Federal. Professor e diretor operacional da Tutores Reforço Es-

colar Multidisciplinar em Porto Velho. Professor pela Faculdade de Ciências Administrativas e de Tecnologia FATEC-RO e professor conteudista do Curso Técnico em Finanças na Modalidade à Distância, do Instituto Federal de Rondônia - IFRO.