Fundação
CECIERJ

Consórcio cederj

Centro de Educação Superior a Distância do Estado do Rio de Janeiro

Volume 1

2ª Edição

Giovani Glaucio de Oliveira Costa

Juliana Di Giorgio Giannotti

Estatística Aplicada ao Turismo

GOVERNO DO
Rio de Janeiro

SECRETARIA DE
CIÊNCIA E TECNOLOGIA

UNIVERSIDADE
ABERTA DO BRASIL

Ministério da
Educação

GOVERNO FEDERAL
BRASIL
PAÍS RICO É PAÍS SEM POBREZA

Apoio:

FAPERJ
Fundação Carlos Chagas Filho de Amparo
à Pesquisa do Estado do Rio de Janeiro

Fundação Cecierj / Consórcio Cederj

Rua da Ajuda, 5 – Centro – Rio de Janeiro, RJ – CEP 20040-000
Tel.: (21) 2333-1112 Fax: (21) 2333-1116

Material Didático

ELABORAÇÃO DE CONTEÚDO
Giovani Glaucio de Oliveira Costa
Juliana Di Giorgio Giannotti

COORDENAÇÃO DE DESENVOLVIMENTO INSTRUCIONAL
Cristine Costa Barreto

SUPERVISÃO DE DESENVOLVIMENTO INSTRUCIONAL
Ana Paula Abreu-Fialho

DESENVOLVIMENTO INSTRUCIONAL E REVISÃO
Gustavo de Figueiredo Tarcsay
Luiz Eduardo S.
Marcelo Bastos Matos
Wilson P. de Oliveira Junior

Departamento de Produção

EDITORA
Tereza Queiroz

REVISÃO TIPOGRÁFICA
Cristina Freixinho
Daniela de Souza
Diana Castellani
Elaine Bayma
Patrícia Paula

COORDENAÇÃO DE PRODUÇÃO
Jorge Moura

PROGRAMAÇÃO VISUAL
Márcia Valéria de Almeida

ILUSTRAÇÃO
André Dahmer

CAPA
André Dahmer

PRODUÇÃO GRÁFICA
Verônica Paranhos

C837t

Costa, Giovani Glaucio de Oliveira.
Estatística Aplicada ao Turismo. v. 1 / Giovani Glaucio de Oliveira Costa, Juliana Di Giorgio Giannotti. – 2.ed. – Rio de Janeiro: Fundação CECIERJ, 2013.
192 p.; 19 x 26,5 cm.

ISBN: 978-85-7648-623-7

1. Turismo. 2. Estatística. 3. Métodos estatísticos. 4. Séries estatísticas. 5. Gráficos. I. Giannotti, Juliana Di Giorgio. II. Título.

CDD: 338.4791

Referências Bibliográficas e catalogação na fonte, de acordo com as normas da ABNT e AACR2.

Governo do Estado do Rio de Janeiro

Governador
Sérgio Cabral Filho

Secretário de Estado de Ciência e Tecnologia
Gustavo Reis Ferreira

Universidades Consorciadas

UENF - UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO
Reitor: Silvério de Paiva Freitas

UERJ - UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO
Reitor: Ricardo Vieiralves de Castro

UFF - UNIVERSIDADE FEDERAL FLUMINENSE
Reitor: Roberto de Souza Salles

UFRJ - UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Reitor: Carlos Levi

UFRRJ - UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
Reitor: Ricardo Motta Miranda

UNIRIO - UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
Reitor: Luiz Pedro San Gil Jutuca

Estatística Aplicada ao Turismo | Volume 1

SUMÁRIO

1 Conceitos básicos da Estatística; variáveis e seus tipos

Meta da aula

Apresentar uma visão geral do que é Estatística e os tipos de variáveis com os quais o aluno irá trabalhar.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. definir população e diferenciar amostra, censo e amostragem;
2. compreender a importância da amostragem para a Estatística;
3. distinguir os objetivos e diferenças da Estatística descritiva e inferencial;
4. identificar o que são variáveis e destacar sua importância dentro do contexto da Estatística;
5. distinguir as diversas classificações de variáveis.

Introdução

Desde a antiguidade vários povos faziam registros de dados e informações que consideravam importantes, como o número de habitantes da tribo ou cidade; o número de nascimentos e óbitos; a avaliação de bens e riquezas das pessoas para a posterior cobrança de impostos; a quantificação dos estoques de alimentos e armamentos etc. Assim, por meio de métodos de contagem, surgiram as primeiras ideias da Estatística atual.

O termo Estatística vem do latim *status*, que significa Estado, pois as informações fornecidas e coletadas antigamente estavam ligadas apenas a assuntos relativos ao Estado, como segurança e controle fiscal. Atualmente, a Estatística tornou-se uma ciência cuja aplicação se dá em vários ramos do conhecimento, como, por exemplo, turismo, medicina, engenharias, educação, entre outros.

Assim, para que você encontre maior facilidade na compreensão desta aula, é importante que relacione a Estatística com outras ciências, entenda a amostragem como exercício de comportamento diário que todos nós fazemos no cotidiano de nossas vidas e, por fim, relacione o conceito de variáveis e de suas classificações com exemplos da realidade.

Fenômeno

É tudo que pode ser percebido pelos sentidos ou pela consciência, ou seja, é a definição de qualquer acontecimento observável. É possível listar fenômenos que são relevantes em praticamente qualquer campo de pesquisa. Alguns são ocorrências naturais, como os fenômenos climáticos; outros requerem intervenções do homem como os fenômenos biológicos ou empresariais.

São exemplos de fenômenos: *uma fruta que cai de uma árvore; uma pessoa que nasce; a incidência de uma doença; o comportamento das pessoas numa loja; o consumo de certo produto; o lucro de uma empresa; a porcentagem de ocupação da rede hoteleira num feriado prolongado; distribuição de interesse em* *lugares para serem visitados em um passeio turístico em grupo; faturamento de um hotel.*

Fenômeno coletivo ou de massa

São os fenômenos que não possuem regularidade na observação de casos isolados, mas sim na massa de observações, ou seja, são aqueles que não podem ser definidos por uma simples observação, e sim por um conjunto de observações.

Os exemplos de fenômenos coletivos são: o nível socioeconômico dos hóspedes de um hotel; o preço médio das diárias de hotéis de determinada categoria.

Ciência

É o conjunto orgânico de conhecimentos sobre os fenômenos e suas relações recíprocas. É o processo racional usado pelo homem para se relacionar com a natureza e assim obter resultados que lhe sejam úteis.

A ciência Estatística

A Estatística é uma ciência que envolve coleta, organização, resumo, análise e interpretação dos dados. Assim, a ciência Estatística é utilizada para estudar e medir os fenômenos coletivos ou de massa por meio de técnicas adequadas, algumas das quais você vai estudar no nosso curso.

População

É o conjunto de elementos portadores de pelo menos uma característica comum de interesse para ser estudado pela ciência Estatística.

Exemplos:

- Num estudo sobre satisfação por certo serviço oferecido por um “pacote turístico”, a população é constituída por todos os clientes (aqueles que utilizaram) deste serviço.
- Num estudo sobre hábitos de fumar de clientes de um hotel, a população será formada por todos os clientes deste hotel.

Amostra

É um subconjunto qualquer da população, selecionada para representá-la. Para que as conclusões sobre a população sejam fornecidas adequadamente pela amostra, é necessário que ela seja representativa da população.

Amostras representativas são aquelas que são verdadeiras miniaturas da população, isto é, têm todas as características da população, mas em menor escala.

Para obtermos amostras representativas, existem vários métodos de extração (esses métodos serão estudados nas aulas seguintes), mas os mais eficazes são aqueles em que os elementos que vão compor a amostra são selecionados por sorteio, aleatoriamente.

Exemplos:

- Num estudo sobre satisfação por certo serviço oferecido por um “pacote turístico”, a amostra é constituída <u>por parte</u> dos clientes (aqueles que utilizaram) deste serviço;
- Num estudo sobre hábitos de fumar de clientes de um hotel, a amostra será formada <u>por parte</u> dos clientes deste hotel.

Censo

É o estudo de uma população com base em todos os seus elementos. Quando a amostra é igual à população, isto denomina-se censo.

No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) realiza, de dez em dez anos, a contagem de toda a população brasileira. O objetivo principal da contagem é atualizar os dados estatísticos populacionais no intuito de orientar políticas e ações públicas com informações atualizadas sobre a população.

Exemplos:

- *Quando estudamos o perfil dos clientes de uma pousada, aplicando um questionário de pesquisa a todos os hóspedes.*
- *A contagem de toda a população brasileira realizada pelo IBGE de dez em dez anos.*

Amostragem

É o estudo de uma população com base em uma parte representativa da mesma, isto é, com base numa amostra.

Exemplo:

Quando estudamos o perfil dos clientes de uma pousada, aplicando um questionário de pesquisa a uma parte representativa destes hóspedes, selecionados por sorteio do banco de dados ou do cadastro de clientes do estabelecimento.

Atividade

1. Distinguir a população e a amostra nas duas situações a seguir:

a. Pretende-se estudar os salários dos 500 funcionários de um grupo hoteleiro. Para tanto, selecionam-se ao acaso 36 funcionários, os seus salários são anotados e realiza-se o estudo com base nestes dados.

b. Pretende-se saber a proporção de garotas e rapazes candidatos ao exame vestibular do curso de Licenciatura em Turismo do CEDERJ. Para tanto, anota-se ao acaso o sexo de 120 candidatos e realiza-se o estudo com base nestes dados.

Respostas

a. A população são os salários de todos os 500 funcionários do grupo hoteleiro, e a amostra são os salários dos 36 funcionários selecionados.
b. A população são todos os candidatos a este curso, e a amostra são os 120 candidatos selecionados.

Divisão da Estatística

A Estatística se divide em dois ramos: Estatística descritiva e Estatística inferencial.

Estatística descritiva

É a parte da ciência Estatística que tem o objetivo de descrever os dados observados. Compreende as seguintes etapas:

1. *Obtenção dos dados.*
2. *Organização dos dados.*
3. *Redução dos dados.*
4. *Representação dos dados*.

Os atributos da Estatística descritiva são a obtenção de informações como médias, proporções, dispersões, tendências, índices, taxas, que resumem e representam os fenômenos observados. Isto encerra as atribuições da Estatística descritiva.

No caso de ter se optado pelo estudo estatístico com amostras, a Estatística estabelece técnicas para se conhecerem as informações da população através da informação amostral. Este é o desígnio da Estatística inferencial.

Estatística inferencial

É a parte da Estatística que tem o objetivo de estabelecer técnicas de como generalizar as informações da amostra para a população, através do cálculo das probabilidades.

Exemplo:

Suponha que tivéssemos colhido uma amostra de 50 contracheques de um total de 2.000 funcionários de uma agência de turismo e viagens, e obtivéssemos a porcentagem de pessoas que tiveram descontos por falta ou atrasos num mês considerado. É função da Estatística inferencial generalizar este resultado encontrado em 50 trabalhadores para os 2.000.

Para facilitar o entendimento dos conceitos de população e amostra, assim como os campos de atuação das Estatísticas descritivas e inferenciais, proponho que você observe atentamente a **Figura 1.1**. Nesta figura, temos uma população da qual é retirada uma amostra. Vê-se a Estatística descritiva sendo aplicada para descrever e resumir o que ocorre na população e/ou na amostra. Visualiza-se, também, que a inferência estatística se encarrega de observar o que ocorre na amostra e generalizar, ou extrapolar, as interpretações para a população.

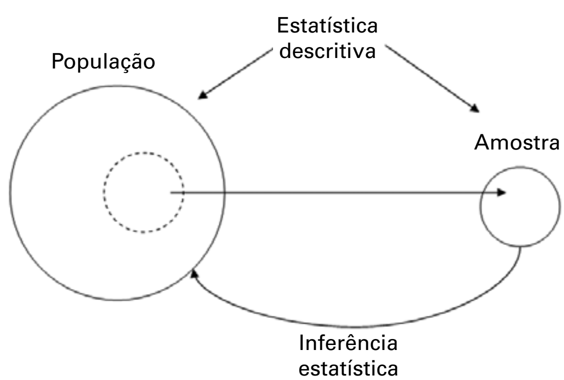


Figura 1.1: Resumo esquemático de população e amostra.

Para você ter uma visão global das etapas envolvidas em um estudo estatístico, observe atentamente o esquema a seguir. Esse esquema mostra que o estudo estatístico se inicia na população, passa pelas etapas explicitadas no esquema que e é finalizado na tomada de decisão a qual foi devidamente embasada pela metodologia estatística envolvida no processo, fato que irá lhe assegurar maior credibilidade.

Esquema lógico de um estudo estatístico

POPULAÇÃO

↓

ESTATÍSTICAS DESCRITIVAS: PARÂMETROS

↓

AMOSTRA ALEATÓRIA

↓

CÁLCULOS DE ESTIMATIVAS DE PARÂMETROS

↓

CÁLCULO DAS PROBABILIDADES

↓

ESTATÍSTICA INFERENCIAL: ***TESTES DE SIGNIFICÂNCIA***

↓

TOMADAS DE DECISÃO

Testes de significância

São aqueles nos quais se utilizam dados amostrais para aceitar ou rejeitar determinada hipótese formulada sobre a população.

Variável

É uma característica qualquer de interesse que associamos à população ou à amostra para ser estudada estatisticamente. São chamadas assim porque apresentam variação de elemento para elemento na população ou amostra em estudo.

Exemplos:

Altura, idade, estado civil e salário dos turistas de uma cidade.

As variáveis podem ser de dois tipos:

Qualitativas e quantitativas

Variáveis qualitativas

Apresentam como possíveis resultados uma qualidade (ou atributo) do elemento pesquisado.

Exemplos:

Sexo (masculino ou feminino), cor (branco, azul, amarelo), boa ou má aparência, status *social (alto, médio, baixo) dos turistas de uma pequena cidade.*

As variáveis qualitativas se classificam em nominais e ordinais.

a. Variável quantitativa nominal

Não existe nenhuma ordenação possível nas realizações das categorias.

Por exemplo, religião é uma variável qualitativa nominal, pois não podemos afirmar que a religião católica é melhor ou pior do que a evangélica, ou que a espírita, ou que a muçulmana seja.

Exemplos:

Sexo, cor, estado civil, nacionalidade dos turistas de uma grande cidade.

b. Variável qualitativa ordinal

Existe certa ordem ou hierarquia entre as realizações de cada variável.

Exemplo:

Nível de instrução é uma variável qualitativa ordinal, pois é possível colocar em uma ordem natural qual é o maior e qual é o menor nível de instrução, ou seja, o nível superior é maior que o nível médio, que é maior que o nível fundamental.

Exemplos:

- *Nível socioeconômico: classe alta > classe média > classe baixa dos turistas de uma pequena cidade.*
- *Cargo ocupado por funcionários em uma empresa de turismo (recepcionista, agente de turismo, gerente).*
- *Posto no serviço militar (coronel, major, capitão, tenente, soldado).*

Variável quantitativa

São aquelas em que as possíveis realizações são valores numéricos. São utilizadas para descrever quantidades e assim são medidas numericamente.

Exemplos:

Idade, estatura e peso corporal dos hóspedes de um SPA.

As variáveis quantitativas se classificam em discretas e contínuas.

a. Variável quantitativa discreta

Seus possíveis valores são números inteiros, ou seja, sem casas decimais, e formam um conjunto enumerável de valores. São aquelas que resultam de contagens.

Exemplos:

Número de filhos de um turista, renda de um turista, números de acidentes numa viagem turística, número de pessoas que chegam ao balcão de um hotel, número de pessoas que utilizam o café da manhã de um hotel.

b. Variável quantitativa contínua

Seus possíveis valores pertencem a um intervalo de valores. São aquelas que resultam de medição. Não são necessariamente inteiras, dependem da precisão adotada e do instrumento de medida.

Exemplos:

Idades dos participantes de um passeio turístico, quilômetros rodados por uma van quando em um passeio turístico, estatura e peso corporal dos participantes de um passeio turístico.

Resumo esquemático da classificação das variáveis:

Esquema

- Variáveis
  - Qualitativas
    - Nominal
    - Ordinal
  - Quantitativas
    - Discreta
    - Contínua

Atividade

Atende aos Objetivos 3 e 5

2. Classifique as seguintes variáveis:

a. Motivo da viagem (turismo, negócios, visita a parentes).

b. Hábito de viajar (sozinho, com família, em excursão).

c. Nacionalidade dos turistas (brasileira, argentina, uruguaia, paraguaia).

d. Número de dias permanecidos no país visitado.

e. Quantia gasta em reais por turista por dia.

f. Nível de escolaridade do turista.

g. Número de filhos do turista.

h. Peso em kg da bagagem por pessoa.

Respostas

a. Variável qualitativa nominal.
b. Variável qualitativa nominal.
c. Variável qualitativa nominal.
d. Variável quantitativa discreta.
e. Variável quantitativa contínua.
f. Variável qualitativa ordinal.
g. Variável quantitativa discreta.
h. Variável quantitativa contínua.

Atividades Finais

Atende aos Objetivos 1, 2, 4 e 5

1. Releia e verifique se você compreendeu os conceitos de: ciência, Estatística, população, amostra, censo, amostragem, estatística descritiva e inferencial.

Comentário

Para desenvolver a Atividade Final 1, você deve ter fixado os conceitos apresentados na aula, mas tente descrevê-los aqui com suas próprias palavras.

2. Relacione as letras com os números no sentido de classificar as variáveis a seguir:

Variável	Classificação
a. Cor dos olhos dos candidatos a modelos de uma agência.	1. Variável qualitativa nominal.
b. Número de filhos de casais visitantes de uma cidade.	2. Variável qualitativa ordinal.
c. Causas de acidentes de trabalho de funcionários de uma fábrica.	3. Variável quantitativa discreta.
d. Salários de funcionários de uma empresa de turismo.	4. Variável quantitativa contínua.
e. Renda de clientes de uma agência de turismo.	
f. Classificação num concurso público.	
g. Nota atribuída por parte de turistas à satisfação com certa viagem.	
h. Peso de hóspedes de um spa.	
i. Estabelecimentos de saúde, públicos e particulares.	
j. Estado civil de entrevistados em uma pesquisa de opinião de satisfação de turistas de um país em período de "alta temporada".	

Resposta

Para classificar corretamente as variáveis do quadro da Atividade Final 2, pense sempre em suas possíveis realizações: se elas são valores alfanuméricos ou numéricos; se forem alfanuméricos, se existe alguma ordenação natural possíveis entre as realizações ou não; sendo valores numéricos, se surgem quando se contam ou quando se medem.

a. 1

b. 3

c. 1

d. 4

e. 4

f. 2

g. 2

h. 4

i. 1

j. 1

3. Nas situações a seguir, identifique a população, a amostra e a variável em estudo e, em seguida, classifique essa variável.

a. Um comprador de determinada rede de restaurantes deve adquirir um lote de 10.000 frangos de 45 dias, produzidos em certa granja. Sabe-se que a avaliação do lote é feita em função do peso dos frangos. Diante da impossibilidade de pesar cada um dos 10.000 frangos, o comprador seleciona ao acaso 50 frangos e realiza a pesagem.

b. Deseja-se estudar a proporção de moradores do município A que são favoráveis à implantação de um complexo turístico próximo desse município. Consultam-se, ao acaso, 200 moradores desse município e a opinião de cada um deles é registrada como sendo contra ou a favor da construção do complexo turístico.

c. Em uma pesquisa de intenção de voto para o governo do estado do Rio de Janeiro, foram entrevistados 1.600 eleitores.

d. Deseja-se investigar se há ou não material contrabandeado em bagagens de passageiros que desembarcam em um aeroporto internacional. Para tanto, selecionam-se, ao acaso, 150 passageiros de determinado voo e suas bagagens são inspecionadas.

Respostas

a. Podemos identificar como sendo a população os 10.000 frangos de 45 dias produzidos na granja, a amostra os 50 frangos selecionados pelo comprador e a variável em estudo é o peso dos frangos. A variável peso é classificada como quantitativa contínua (lembre-se de que ela é proveniente de medição).
b. Podemos identificar como sendo a população todos os moradores do município A, a amostra os 200 moradores consultados e a variável em estudo é a opinião dos moradores (favorável ou não favorável). A variável opinião é classificada como qualitativa nominal.
c. Podemos identificar como sendo a população todos os eleitores do estado do Rio de Janeiro, a amostra os 1.600 eleitores entrevistados e a variável em estudo é a intenção de voto. A variável intenção de voto é classificada como qualitativa nominal.
d. Podemos identificar como sendo a população todos os passageiros do voo em questão, a amostra os 150 passageiros selecionados e a variável em estudo é a presença ou não de material contrabandeado. A variável presença de material contrabandeado é classificada como qualitativa nominal.

2 Noções de amostragem

Meta da aula

Apresentar o conceito de amostragem e suas principais regras de seleção.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. reconhecer o conceito de amostragem e sua importância dentro da Estatística;

2. distinguir amostragem probabilística da não probabilística;

3. identificar as principais regras de amostragem;

4. indicar uma determinada regra de amostragem para uma situação prática que lhe for apresentada;

5. calcular tamanhos de amostras.

Pré-requisito

Para que você encontre maior facilidade na compreensão desta aula, é importante que você tenha sedimentado os conceitos de população, amostra, censo e amostragem, vistos na Aula 1.

Introdução

Quando se deseja informação sobre uma dada situação, o que vem à mente da maioria das pessoas é obter a informação de toda a população de interesse. Porém, nem sempre é possível obter a informação de toda a população, pois muitas vezes não temos esse acesso ou é financeiramente inviável trabalhar com todos os elementos populacionais.

Por esse motivo, um importante ramo de estudo em Estatística é a teoria da amostragem, que consiste no estudo das relações existentes entre uma população e as amostras dela extraída. A extração, ou seleção, de uma amostra a partir de determinada população segue uma regra que visa garantir que tal amostra represente o mais fielmente possível a população que a originou.

Nesta aula, você vai aprender os tipos de amostragens existentes e a forma de fazer o dimensionamento dessas amostras.

Noções de amostragem

Amostragem é o processo (um método de coleta) de retirada de amostras. É o estudo de uma população (ou universo) com base em uma amostra representativa.

Amostra, como vimos na Aula 1, é uma parte da população, retirada segundo uma regra conveniente. Você irá aprender ainda nesta aula as várias regras para seleção de amostras, ou melhor, os vários tipos de amostragem. O pressuposto básico para que se utilize uma determinada regra de amostragem é que ela gere amostras representativas, isto é, com todas as características básicas e importantes do universo: *que seja uma verdadeira miniatura da população*.

As regras de amostragem podem ser classificadas em duas categorias gerais: *probabilísticas e não probabilísticas*.

Amostragens probabilísticas

São amostragens em que a seleção é aleatória, de tal forma que cada elemento da população tenha uma chance real de fazer parte da amostra.

Assim: se uma população tem 100 elementos, ou seja, se N = 100, e se todos os elementos da população possuem igual probabilidade de serem selecionados, teremos que 1/N, ou seja, 1/N = 1/100 = 0,01, é a probabilidade, ou a chance, de cada elemento da população ser selecionado para fazer parte da amostra.

Em amostragem vamos utilizar sempre a letra N para representar o tamanho da população e a letra n para representar o tamanho da amostra.

Tipos de amostragens probabilísticas

a. Amostragem aleatória simples

b. Amostragem sistemática

c. Amostragem estratificada

d. Amostragem por conglomerados

e. Amostragem múltipla

a. Amostragem aleatória simples

Também conhecida por amostragem ocasional, randômica etc., a amostragem aleatória simples se destaca por ser um processo de seleção bastante fácil e muito usado. Nele, todos os elementos da população têm igual probabilidade de serem escolhidos, não só antes de o processo ser iniciado, como até se completar a coleta.

Eis o procedimento para seu uso:

1º. Devemos numerar todos os elementos da população. Se, por exemplo, nossa população tem 5.000 elementos, devemos numerá-los de 1 a 5.000.

2º. Devemos efetuar sucessivos sorteios com ou sem reposição até completar o tamanho da amostra (n).

Reposição – Se cada elemento da população pode ser escolhido mais de uma vez para participar de uma mesma amostra, temos o *sorteio com reposição*. Se cada elemento da população puder ser escolhido apenas uma única vez para participar de uma mesma amostra, temos o *sorteio sem reposição*.

Para realizar o sorteio de forma aleatória, sugere-se o uso de um programa computacional estatístico ou de uma planilha eletrônica. Um exemplo de planilha eletrônica é o Excel, no qual, por meio do comando "aleatório" ou em análise de dados a opção "amostragem", podem-se gerar números aleatórios e amostras aleatórias. O processo termina quando for sorteado o último elemento "n".

b. Amostragem sistemática

Trata-se de uma variação da amostragem aleatória simples, muito conveniente quando a população está naturalmente ordenada, como fichas em um fichário, nomes em listas telefônicas, clientes de uma empresa registrados em um banco de dados etc.

Eis o procedimento para seu uso:

Seja N o tamanho da população e n o tamanho da amostra, então, calcula-se o intervalo de amostragem I = N/n ou o inteiro mais próximo de "I". Sorteia-se, utilizando-se algum dispositivo aleatório qualquer (geração de números aleatórios no excel ou a tabela de números aleatórios), um número x, entre 1 e "I", formando-se a amostra dos elementos correspondentes aos números: x; x + I; x +2I; ... ; x+(n-1)I. Observa-se que a sequência dos elementos sorteados forma uma progressão aritmética de razão igual a "I".

Exemplo:

Seja N = 500 o número de hóspedes de um hotel e n = 50 igual ao tamanho de uma amostra a ser selecionada deste total para uma pesquisa de satisfação de clientes, então, 500/50 = 10 ou I = 10. Este é o intervalo de amostragem.

Sorteia-se um número de 1 a 10. Seja 3 (x = 3) o número sorteado, logo os elementos numerados por 3; 13; 23; 33; . . . serão os componentes da amostra, isto é, os hóspedes do hotel que serão investigados e representarão toda a população de hóspedes deste hotel.

Atividade

Atende aos Objetivos 3 e 4

1. Um hotel tem um fichário com o registo de 5.250 clientes e pretende amostrar 250 fichas. Obtenha, por meio da amostragem sistemática, os números dos registros das 5 primeiras fichas e o número da última ficha. Sabe-se que a primeira ficha sorteada foi a de número 17 (x = 17).

Resposta

$I = 5.250/250$ ou $I = 21$.

Como foi indicado que x=17, então temos que os 5 primeiros registros são:

17; 17+I; 17+2I; 17+3I e 17+4I e o último 17+249I, que resultará nos seguintes números de registros 17; 38; 59; 80; 101 e o último 5.246.

c. Amostragem estratificada

No caso de população heterogênea, na qual podemos distinguir subpopulações mais ou menos homogêneas, denominadas ***estratos***, podemos usar a amostragem estratificada.

Estratos

São grupos homogêneos de elementos de uma população.

Estratificar uma população é dividi-la em L subpopulações, denominadas estratos, tais que $N_1 + N_2 + ... + N_L = N$, onde os estratos são mutuamente exclusivos, ou seja, os elementos do estrato N_1, por exemplo, não podem estar presentes em nenhum outro estrato.

Após a determinação dos estratos, seleciona-se uma amostra aleatória de cada subpopulação.

Se as diversas subamostras (estratos) tiverem tamanhos proporcionais ao respectivos números de elementos no estrato, teremos a estratificação proporcional.

Exemplos:

a. Se o objetivo de uma pesquisa é traçar o perfil dos turistas que chegaram a uma cidade em uma época de alta temporada e se esta população for dividida por nacionalidade, cada nacionalidade pode ser considerada um estrato (subpopulação).
b. Se o objetivo de uma pesquisa é obter a opinião de europeus sobre qual destino turístico seria o mais atrativo para visitação ao Brasil e se esta população for dividida por faixa de renda, cada faixa de renda é um estrato.
c. Se o objetivo de uma pesquisa é obter a opinião de americanos sobre qual cidade brasileira seria a mais atrativa para visitação e se esta população for dividida por sexo, as categorias sexo feminino e sexo masculino são os estratos desta população, ou seja, cada sexo seria considerado um estrato.

Atividade

Atende aos Objetivos 3 e 4

2. Uma agência de turismo tem o cadastro de 500 funcionários, sendo 380 homens e 120 mulheres. Para uma amostra de 10% desses cadastros, retirada obedecendo à amostragem estratificada proporcional, determine o número de cadastros de homens e mulheres que devem estar presentes nessa amostra.

Resposta

N = 500 e 10% de 500 é igual a 50, assim o tamanho da amostra deve ser n = 50. Porém, vamos estratificar esta população; para tanto devemos tomar 10% de cadastros de homens e 10% de cadastros de mulheres. Com o auxílio de um quadro, vamos visualizar como ficará o tamanho dos estratos e da amostra.

Estratos	*População*	*Amostra estratificada*
Homem	*380*	$n_1 = 38$
Mulher	*120*	$n_2 = 12$
Total	*N = 500*	*n = 50*

d. Amostragem por conglomerados

Quando a população apresenta uma subdivisão em pequenos grupos não necessariamente homogêneos, chamamos de conglomerados.

Pode-se fazer a amostragem por conglomerados, que consiste em sortear um número suficiente (será apresentado o cálculo do tamanho de uma amostra mais adinte nesta aula) de conglomerados, cujos elementos constituirão a amostra.

As unidades de amostragem sobre as quais é feito o sorteio passam a ser os conglomerados e não mais os elementos individuais da população.

Esse tipo de amostragem é às vezes adotado por motivos de ordem prática e econômica, ou mesmo por razões de viabilidade: diante da impossibilidade de se obter uma listagem da população numerada para realizar um sorteio.

Exemplos:

1. Quando se pretende avaliar a imagem que os europeus têm quanto ao aspecto turístico do Brasil, cada país da Europa pode ser considerado um conglomerado, pois é complicado se obter uma listagem completa e atualizada dos europeus.

   Em algumas situações, podemos identificar um grupo de elementos que tenha aproximadamente a mesma composição (característica) da população. Neste caso, pode ser interessante realizar a amostragem somente com os elementos desse grupo.
2. Algumas empresas, quando pretendem avaliar a aceitação de um pacote turístico para o eixo Rio-São Paulo, lançam o pacote em Curitiba, cuja população se comporta como uma miniatura do mercado do eixo Rio-São Paulo.

e. Amostragem múltipla

Vamos introduzir o conceito de amostragem múltipla com um exemplo prático para assim facilitar o entendimento. Seja a produção total de uma fábrica chamada de lote. Retira-se uma amostra aleatória de tamanho n deste lote, seguindo uma regra predeterminada, e verifica-se que a proporção de defeitos encontrada na amostra está dentro de um padrão de qualidade. Caso contrário, para confirmação, retira-se, por exemplo, uma segunda amostra e constata-se então o padrão de qualidade estimado para este lote. A este procedimento denominamos amostragem múltipla.

Assim, conceituamente, numa amostragem múltipla, a amostra é retirada em diversas etapas sucessivas. Dependendo dos resultados observados, etapas suplementares podem ser dispensadas. Esse tipo de amostragem é, muitas vezes, empregado na inspeção por amostragem, sendo particularmente importante a amostragem dupla (caso do nosso exemplo descrito anteriormente). Sua finalidade é diminuir o número médio de itens inspecionados a longo prazo, baixando assim o custo de inspeção.

Um caso extremo de amostragem múltipla é a *amostragem sequencial*. A amostra vai sendo acrescida item por item, até se chegar a uma conclusão no sentido de se aceitar ou rejeitar uma dada hipótese. Com a amostragem sequencial, pretende-se tornar mínimo o número médio de itens inspecionados a longo prazo.

Amostragens não probabilísticas

São amostragens em que há uma escolha deliberada dos elementos da amostra.

As amostras não probabilísticas não asseguram naturalmente e necessariamente que a amostra gerada seja representativa da população, uma vez que pode haver elementos da população que não tenham chance real de fazerem parte da amostra. Isso coloca em discussão a sua confiabilidade.

As amostras não probabilísticas são, muitas vezes, empregadas em trabalhos estatísticos, por simplicidade ou por impossibilidade de se obterem amostras probabilísticas como seria desejável. Como em muitos casos, os efeitos da utilização de uma amostragem não probabilística podem ser considerados equivalentes aos de uma amostragem probabilística. Os processos não probabilísticos de amostragem têm também sua importância, principalmente em pesquisas de intenção de voto, quando não se tem acesso a um cadastro da população. Sua utilização, entretanto, deve ser feita com reservas e com a convicção de que não introduzam tendências.

Tipos de amostragens não probabilísticas

a. Inacessibilidade a toda a população
b. Amostragem a esmo ou sem norma
c. População formada por material contínuo
d. Amostragens intencionais
e. Amostragem por voluntários
f. Amostragem por quotas

a. Inacessibilidade a toda a população

Esta situação ocorre com muita frequência na prática. Ocorre quando somos obrigados a colher a amostra na parte da população que nos é acessível. Surge, portanto, uma distinção entre *população objeto e população amostrada*.

População objeto é aquela que temos em mente ao realizar o trabalho estatístico. *População amostrada* é aquela que está acessível na prática para que a amostra seja selecionada.

Se as características da variável de interesse forem as mesmas na população objeto e na população amostrada, então esse tipo de amostragem equivalerá a uma amostragem probabilística. Uma situação muito comum em que ficamos diante da inacessibilidade a toda a população é o caso em que parte da população não tem existência real.

Exemplos:

Seja uma pesquisa em que a população-alvo é constituída por todos os turistas de uma cidade. Mesmo tendo uma quantidade expressiva de turistas na cidade no momento da pesquisa, existe uma parte da população que é formada por turistas que ainda vão chegar à cidade.

Se nos interessa a população de todos os turistas vítimas de violência no Brasil, estaremos diante de um caso semelhante.

Em geral, estudos realizados com base nos elementos da população amostrada terão, na verdade, seu interesse de aplicação voltado para os elementos restantes, ou seja, para a totalidade da população objeto. Esse fato realça a importância de se estar convencido de que as duas populações podem ser consideradas como tendo as mesmas características.

O presente caso de amostragem não probabilística pode ocorrer também quando, embora se tenha a possibilidade de atingir toda a população, retiramos a amostra de uma parte que seja prontamente acessível.

Exemplo:

Se fôssemos investigar o perfil dos turistas de um pacote oferecido por uma agência de viagens, poderíamos fazer a entrevista somente com os turistas que estivessem no exato momento hospedado na pousada do hotel, pois estão mais próximos e acessíveis.

b. Amostragem a esmo ou sem norma

É a amostragem em que o amostrador (pesquisador), para simplificar o processo, procura ser aleatório sem, no entanto, realizar propriamente o sorteio, usando algum dispositivo aleatório confiável.

Exemplo:

Se desejarmos retirar uma amostra de 100 garfos de uma caixa contendo 10.000, evidentemente não faremos uma amostragem aleatória simples, pois seria extremamente trabalhosa. Nesse caso, seria melhor procedermos à retirada simplesmente a esmo.

Os resultados da amostragem a esmo são, em geral, equivalentes aos de uma amostragem probabilística, se a população é homogênea e se não existe a possibilidade de o amostrador ser inconscientemente influenciado por algumas características dos elementos da população.

No caso de haver influência do amostrador para selecionar algum tipo de elemento da amostra, essa influência acarretará tendência nas respostas de um questionário. Por exemplo, se há o interesse em avaliar a intenção de voto da população de um município para determinado candidato, e o amostrador só colher informações dos simpatizantes desse candidato, isso acarretará tendência na resposta, pois todos os informantes indicarão o voto para este candidato.

c. População formada por material contínuo

Nesse caso, é impossível realizar amostragem probabilística devido à impraticabilidade de um sorteio rigoroso: se a população for líquida (por exemplo, o sangue do corpo humano) ou gasosa (por exempo, a poluição presente no ar de determinado local). O que se costuma fazer é ***homogeneizar*** e retirar a amostra a esmo.

> ***Homogeneizar***
>
> É tornar homogêneo, igual, parecido.

Tal procedimento pode, às vezes, ser usado no caso de material sólido, por exemplo, para avaliar a quantidade de microorganismos na areia de uma praia.

Outro procedimento que pode ser empregado nesses casos, especialmente quando a homogeneização não seja praticável, é a *enquartação*, que consiste em subdividir a população em diversas partes (a origem do nome pressupõe a divisão em quatro partes), sorteando-se uma ou mais delas para constituir a amostra ou para delas retirar a amostra a esmo.

d. Amostragens intencionais

Enquadram-se aqui os diversos casos em que o amostrador deliberadamente escolhe certos elementos para pertencer (constituir) a amostra, por julgar tais elementos bem representativos da população. Na amostragem intencional, o pesquisador simplesmente inclui os sujeitos segundo sua conveniência na amostra, dela excluindo os inconvenientes.

O perigo desse tipo de amostragem é obviamente grande, pois o amostrador pode facilmente se equivocar em seu prejulgamento. Apesar disso, o uso de amostragens intencionais, ou parcialmente intencionais, é bastante frequente, ocorrendo em vários tipos de situações reais.

Exemplo:

O editor de uma revista *gay* pede aos seus leitores que respondam a perguntas na revista sobre seu comportamento turístico e opiniões para avaliar seu perfil e estilo de vida, julgando que esta amostra é representativa natural dos *gays* brasileiros.

e. Amostragem por voluntários

Ocorre quando o componente da população se oferece voluntariamente para participar da amostra, independente do julgamento do pesquisador.

Exemplo:

Ocorre, por exemplo, no caso da pesquisa experimental de uma nova droga, em que pacientes são solicitados, havendo concordância, a servirem de "sujeitos" (cobaias) para a verificação da eficácia do novo medicamento. Os que se interessarem se apresentarão voluntariamente ao pesquisador.

f. Amostragem por quotas

É a mais usada e conhecida amostragem não probabilística. Muito praticada no Brasil, sobretudo em pesquisas de mercado e de intenção de voto. Na amostragem por quota, diversas características de uma população, tais como sexo, classe social ou etnia, são amostradas nas mesmas proporções em que figuram na população.

Suponha, por exemplo, que tivéssemos de extrair uma amostra, por quota, da população de turistas de uma dada cidade, em que 42% fossem mulheres e 58%, homens. Usando este método, os entrevistadores recebem a incumbência de localizar uma quota de turistas de tal forma que somente 42% da amostra sejam mulheres e 58%, homens. As mesmas porcentagens que figuram na população são reproduzidas na amostra. Se o tamanho global da amostra fosse de 200 pessoas, então 84 mulheres e 116 homens deveriam ser selecionados.

As amostras por quotas são usadas em certos tipos de pesquisa de mercado e opinião pública, porém as inferências feitas nessas condições não permitem assegurar a confiança do erro (você vai aprender a seguir como esta confiança do erro entrará no cálculo do tamanho da amostra) de amostragem estipulado.

Tamanho de amostras

O tamanho de uma amostra deve alcançar determinadas proporções mínimas, estabelecidas estatisticamente. Ou seja, é o número mínimo de casos a serem amostrados, escolhidos preferencialmente aleatoriamente, para garantir certa segurança estatística em relação à representatividade dos dados.

Às necessidades práticas de tempo, custos, etc. recomenda-se não ultrapassar o tamanho mínimo determinado pela estatística. Portanto, é necessário conhecer a forma de calcular o tamanho da amostra, não só para garantir a possibilidade de generalizar os resultados, mas também pelos aspectos práticos mencionados.

O tamanho da amostra depende dos seguintes fatores:

- Tamanho da população do universo.
- Nível de confiança estabelecido para o erro de amostragem.
- Erro de amostragem permitido.
- Proporção de uma quota característica importante do universo.

Tais fatores serão abordados mais adiante nesta aula.

Classificação da população

Segundo o tamanho da população, o universo divide-se em finito e infinito. Consideram-se universos finitos (limitados) aqueles que não ultrapassam as 100.000 unidades (pessoas, alunos, estabelecimentos educacionais, empresas etc.). Universos infinitos são aqueles que ultrapassam essa quantidade. Tal distinção é importante para determinar o tamanho da amostra, pois as fórmulas são diferentes. No caso do universo infinito, supõe-se que o seu tamanho não influa na fórmula a aplicar. O universo finito depende do número de unidades. Neste curso, apresentam-se as fórmulas mais simples, para amostras aleatórias simples e amostras estratificadas, pois são as mais utilizadas em situações práticas na área do turismo.

Convenciou-se, na literatura científica, o valor superior a 100.000 unidades para se considerar a população infinita.

Fórmula para dimensionar a amostra em uma população infinita:

$$n = \frac{z^2 \times P \times Q}{E^2}$$

onde:

n = tamanho da amostra;

Z = escore da curva normal que é função do nível de confiança escolhido; ou seja, é o nível de confiança dado em número de desvios.

Na tabela a seguir, apresentamos valores de Z, retirados da curva normal, em função do nível de confiança arbitrado:

Tabela 2.1: Tabela baseada nos dados obtidos no **Gráfico 2.1**

Confiança (%)	Escore Z
68	**1,00**
90	**1,65**
95	**2,00**
99	**3,00**

A distribuição normal tem a forma de sino, e os escores Z estão representados na abscissa (eixo X); para determinado nível de confiança tem-se um escore Z diferente.

P = proporção de um resultado de uma característica importante do perfil da população, calculada em porcentagem.
Exemplo: a proporção P de turistas com nacionalidade americana na população é de 40%.
Quando não se dispuser (ou souber) da estimativa de P, pode-se supor que ela seja de 50% na população, o caso mais desfavorável na estimação, pois é aquela em que se trabalha com o tamanho máximo de amostra. É lógico que se a proporção da característica pesquisada fosse outra, seria necessário um menor número de casos.

Q = proporção do universo que não possui a característica pesquisada (Q = 1-P). Em porcentagem: Q = 100 – P.
Exemplo: A proporção P de turistas com nacionalidade americana na população é igual a 40%, tem-se, então, Q = 100-40 = 60%.
E = é o erro amostral expresso na unidade da variável em estudo e estabelecido pelo pesquisador.

Fórmula para dimensionar a amostra em uma população finita:

$$n = \frac{Z^2 \times P \times Q \times N}{E^2 (N-1) + Z^2 \times P \times Q}$$

onde:

N é o tamanho da população; os demais termos foram definidos anteriormente.

Vamos desenvolver exemplos e atividades para o cálculo do tamanho da amostra, para população finita e infinita, em duas situações:

a. quando tivermos uma amostragem aleatória simples;
b. quando tivermos uma amostragem estratificada.

a. Amostragem aleatória simples

Exemplo 1:

Numa pesquisa sobre as atitudes dos turistas que se encontram na cidade de São Paulo em relação às suas expectativas com relação ao país, qual é o tamanho de uma amostra representativa, com um nível de confiança de 99% para um erro permitido de 4%?

Convém utilizar a fórmula para universos infinitos, pois na cidade de São Paulo há mais de 100.000 turistas. Portanto, o tamanho da população não influi no cálculo desta amostra.

$$n = \frac{Z^2 \times P \times Q}{E^2}$$

Para uma confiança de 99%, Z = 3 (obtido da **Tabela 2.1**).

$n = (3^2 \,.\, 50.\, 50) / 4^2 = 1.406{,}25 = 1.406$ turistas.

Assim, o resultado obtido nos permite dimensionar o tamanho da amostra com n=1.406 turistas, para oferecer segurança de probabilidade de 99% de resultados válidos para o universo e de 4% de erro admitido.

Exemplo 2:

Suponha que a pesquisa sobre as atitudes dos turistas seja realizada na Paraíba, onde os turistas não passam de 50.000. Além disso, o pesquisador quer trabalhar apenas com um nível de confiança de 95%, no erro de amostragem de 4%. Qual é o tamanho da amostra com essas exigências? Considerando que o universo é menor que 100.000 turistas no máximo, utiliza-se a fórmula para universos finitos:

$$n = \frac{Z^2 \times P \times Q \times N}{E^2 (N - 1) + Z^2 \times P \times Q}$$

Cálculo:

Para uma confiança de 95% , Z = 2 (obtido na **Tabela 2.1**).

$n = [2^2.50.50.50.000] / [16.(50.000 - 1) + 2^2 .50.50] = 617{,}3 = 617$ turistas.

Atividade

Atende ao Objetivo 5

3. Pretende-se realizar uma amostra aletória para a população de estudantes universitários do Brasil. Dimensione qual seria o tamanho dessa amostra considerando um nível de confiança de 95% e admitindo-se um erro de 5%?

Resposta

Neste caso N>100.000, pois trata-se de todos os estudantes universitários do Brasil. Assim, a fórmula a ser utilizada é:

$$n = \frac{z^2 \times P \times Q}{E^2}$$

Para uma confiança de 95%, Z = 2 (obtido da **Tabela 2.1**)

$n = (2^2 \,.\, 50.\, 50) / 5^2 = 400$ turistas.
Assim, o resultado obtido nos permite dimensionar o tamanho da amostra com n = 400 turistas, para oferecer segurança de probabilidade de 95% de resultados válidos para o universo e de 5% de erro admitido.

Atividade

Atende ao Objetivo 5

4. Pretende-se realizar uma amostra aletória para a população de 20.000 estudantes universitários do estado de Alagoas. Dimensione qual seria o tamanho dessa amostra, considerando um nível de confiança de 95% e admitindo-se um erro de 5%?

Resposta

Neste caso N = 20.000, ou seja, N<100.000. Assim, a fórmula a ser utilizada é:

$$n = \frac{z^2 \times P \times Q \times N}{E^2 (N-1) + Z^2 \times P \times Q}$$

Cálculo:

Para uma confiança de 95% , Z = 2 (obtido na ***Tabela 2.1****).*

$n = [2^2. 50. 50. 20.000] / [25. (20.000 - 1) + 2^2 .50.50] = 392,17 = 392$ *turistas.*

Assim, o resultado obtido nos permite dimensionar o tamanho da amostra com n = 392 turistas, para oferecer segurança de probabilidade de 95% de resultados válidos para o universo e de 5% de erro admitido.

b. Amostragem estratificada

As amostragens estratificadas apresentam um problema especial na determinação de seu tamanho. O número de casos de amostra global pode ser calculado utilizando-se as duas fórmulas apresentadas anteriormente na aula, mas deve-se, também, calcular o tamanho de cada estrato nos grupos. É a condição básica desse tipo de amostra que deve representar, o mais exatamente possível, os estratos, segundo sua proporção na população.

A forma mais simples de calcular o tamanho da amostra estratificada consiste em aplicar ao tamanho global da amostra as respectivas percentagens que cada estrato representa na população. Isso permite determinar o número de casos a ser distribuído em cada um dos estratos.

Exemplo:

Um pesquisador realiza um estudo sobre a imagem que turistas têm das condições políticas e econômicas do país que visitam. De acordo com as informações em poder do pesquisador, esse tipo de comportamento varia muito de acordo com a profissão do turista. Portanto, o pesquisador está interessado em levantar tal imagem estratificando os turistas por profissão. Na época da pesquisa, a cidade tinha 10.000 turistas, o nível de confiança corresponde a 95% e o erro de amostragem permitido é de 4%. De acordo com a informação disponível, os turistas distribuem-se nas seguintes categorias:

• Profissionais de nível superior	1.000
• Profissionais de nível médio	2.000
• Operários	3.000
• Trabalhadores não qualificados	4.000

Determinação do tamanho proporcional dos estratos

- Nível superior: (1.000/10.000) . 100 = 10%
- Nível médio = 20%
- Operários = 30%
- Trabalhadores não qualificados = 40%

Tamanho da amostra

$$n = \frac{2^2 \times 50 \times 50 \times 10.000}{16\,(9999) + 2^2 \times 50 \times 50}$$

Números de casos em cada estrato amostral

• Nível superior	10% de 588 = 58,8 = 59
• Nível médio	20% de 588 = 117,6 = 118
• Operários	30% de 588 = 176,5 = 177
• Trabalhadores não qualificados	40% de 588 = 235,2 = 235
	Total aproximado: **589** turistas

O critério de arredondamento (dado pela ABNT) foi realizado com o objetivo de deixar valores inteiros dos elementos a serem amostrados.

Atividades Finais

Atendem aos Objetivos 1, 2, 3, 4 e 5

1. Um instituto de pesquisa vai realizar uma pesquisa de opinião de intenção de retorno de turistas ao Brasil por amostragem por quotas.

A base da amostra foi obtida junto ao IBGE e está apresentada na tabela a seguir:

Faixas de idade (em anos)				
Sexo	**0 a 20**	**21 a 40**	**41 anos ou mais**	**Total**
Masculino	120	205	125	450
Feminino	180	255	185	620
Total	300	460	310	1.070

Realize o "desenho" (esquema) da amostra para o instituto, desenvolvendo os passos a seguir:

a. Construa uma tabela, baseada na anterior, com as quotas populacionais que devem ser guardadas na amostra;

b. Determine o tamanho da amostra ou o número de entrevistas a serem realizadas para uma margem de erro de 10%, para mais ou para menos, com uma confiança de 95%. Considere P = 50%.

c. Determine a tabela com o número de entrevistas por quotas.

2. Pretende-se estudar as preferências de adolescentes entre lazer de praia ou campo. A população é composta por 68 meninas e 49 meninos. Na impossibilidade de entrevistar todos, opta-se por entrevistar 12% dos adolescentes. Obtenha os componentes proporcionais da amostra estratificada.

3. É dada uma população constituída pelas 12 primeiras letras do alfabeto. Explique como seria possível obter uma amostra sistemática de três elementos distintos, supondo B como sendo a primeira letra sorteada.

4. Indique como seria possível retirar uma amostra sistemática de 35 cadastros de hóspedes de um determinado hotel a partir de uma população ordenada formada por 2.590 cadastros. Suponha que o primeiro cadastro sorteado é o de número 42. Indique os três primeiros cadastros e os três últimos.

5. Numa pesquisa sobre a preferência de destino de viagem para lua de mel dos casais recém-casados de todo o Brasil, determine qual o tamanho de uma amostra que represente esta população, considerando um nível de confiança de 95% e admitindo-se um erro de 6%.

6. Numa pesquisa sobre a preferência de destino de viagem para a lua de mel dos casais recém-casados do estado do Espírito Santo, determine qual o tamanho de uma amostra que repre-

sente esta população, de 50.000 casais, considerando um nível de confiança de 99% e admitindo-se um erro de 3%.

Respostas Comentadas

1.

a. Determinação do tamanho proporcional nos estratos:
Masculino de 0 a 20: (120/1.070) 100 = 11,21 = 11
Masculino de 21 a 40: (205/1.070) 100 = 19,15 = 19
Masculino acima de 41: (125/1.070) 100 = 11,68 = 12
Feminino de 0 a 20: (180/1.070) 100 = 16,82 = 17
Feminino de 21 a 40: (255/1.070) 100 = 23,83 = 24
Feminino acima de 41: (185/1.070) 100 = 17,28 = 17

Porcentagem nos estratos:

Faixas de idade (em anos)				
Sexo	**0 a 20**	**21 a 40**	**41 anos ou mais**	**Total**
Masculino	11%	19%	12%	42%
Feminino	17%	24%	17%	58%
Total	28%	43%	29%	100%

b. Para resolver a atividade proposta, você deve utilizar a fórmula em cada estrato:

$$n = \frac{Z^2 \times P \times Q}{E^2}$$

Para uma confiança de 95%, Z = 2 (obtido da ***Tabela 2.1****)*

Assim, o tamanho total da amostra deverá ser de:
$n = (2^2 . 50. 50) / 10^2 = 100$ turistas.

c. Número de entrevistas por cota.
Como sabemos o tamanho total da amostra e qual a porcentagem de turistas deve estar presente em cada estrato, podemos montar a seguinte tabela:

Número de entrevistas por estrato

Faixas de idade (em anos)				
Sexo	**0 a 20**	**21 a 40**	**41 anos ou mais**	**Total**
Masculino	11	19	12	42
Feminino	17	24	17	58
Total	28	43	29	100

2.
8 meninas; 6 meninos.

3.
Utilizar amostragem sistemática, que resultará nas letras B, F e J.

4.
Primeiros: 42; 116; 190; últimos: 2.410; 2.484; 2.558.

5.
N > 100.000, portanto utilize: $n = \dfrac{Z^2 \times P \times Q}{E^2}$

Para uma confiança de 95%, Z = 2 (obtido da ***Tabela 2.1****).*
$n = (2^2 . 50. 50) / 6^2 = 277{,}78 = 278$ *turistas.*

6.
N = 50.000, N < 100.000, portanto utilize: $n = \dfrac{Z^2 \times P \times Q \times N}{E^2 (N-I) + Z^2 \times P \times Q}$

Para uma confiança de 99% , Z = 1 (obtido na ***Tabela 2.1****).*
$n = [1^2. 50. 50. 50.000] / [9. (50.000 - 1) + 1^2 .50.50] = 276{,}25 = 276$ *turistas.*

3 Fases do método estatístico – do planejamento à coleta de dados

Meta da aula

Apresentar as etapas do planejamento e da coleta de dados para realização de um trabalho estatístico.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. identificar as fases do planejamento e da coleta de dados de um trabalho estatístico;

2. redigir e interpretar resultados de pesquisa estatística.

Pré-requisitos

Para acompanhar esta aula, é importante que você tenha em mente o conceito de fenômeno, fenômeno coletivo ou de massa, ciência, ciência estatística, população, amostra, divisão dos estatísticos, presentes na Aula 1. Métodos de seleção de amostras e dimensionamento do tamanho da amostra, presentes na Aula 2.

Introdução

Antigamente o conhecimento era obtido por acaso ou por necessidades práticas de maneira empírica, ou seja, baseado no que chamamos "tentativa e erro". Como exemplo, podemos citar a cura de determinadas doenças. Os povos antigos observaram que algumas ervas eram úteis para amenizar sintomas de determinadas enfermidades, enquanto outras não provocavam o mesmo efeito. Essa informação foi conseguida por aqueles povos de maneira empírica.

Atualmente, porém, o conhecimento resulta da observação e do estudo de fenômenos, utilizando processos e métodos científicos. O que irá nortear a pesquisa será a curiosidade do pesquisador em constituir o problema da pesquisa ou o objetivo para se realizar o estudo. Para tanto é preciso ficar claro o modo como os resultados da pesquisa poderão ser usados para se tomar decisões.

Assim, todo projeto de pesquisa deve ter um objetivo definido e explícito, mas também, resumidamente, o porquê da realização da pesquisa. Esse "porquê" constitui a necessidade, a motivação ou a relevância da pesquisa.

Fases do método estatístico

Método

É um conjunto de etapas, ordenadamente dispostas, a serem seguidas na investigação da verdade, no estudo de uma ciência ou para alcançar determinado fim.

O método estatístico

É um conjunto de fases, que devem ser seguidas na ordem determinada, para se obter resultados estatísticos.

Quais as fases do método estatístico?

São sete as fases do método estatístico, a saber:

1. planejamento;
2. coleta de dados;
3. crítica de dados;
4. apuração de dados;
5. análise de dados;
6. emissão do relatório final;
7. comunicação dos resultados.

Nesta aula, serão abordados os itens 1 e 2 , na Aula 4, os demais itens serão discutidos.

Descrição das fases do método estatístico

Planejamento

Na fase de planejamento, podemos resumir em uma lista, como a que segue, as principais etapas a serem desenvolvidas:

1. definir as metas da pesquisa;
2. definir a população;
3. identificar cada membro da população;
4. identificar o esquema de amostragem (como escolher a amostra e que tamanho essa amostra deve ter);
5. decidir que método de coleta de dados será utilizado (questionário postal, entrevista etc.);
6. elaborar um questionário (igualmente apropriado para entrevistas pessoais e observação);
7. selecionar e treinar qualquer pessoa envolvida no processo de coleta de dados.

Desse modo, o trabalho estatístico nasce quando um estudioso (pesquisador) sente a curiosidade de conhecer com mais precisão o comportamento de um fenômeno coletivo ou de massa.

Exemplo:

O termo ***turismólogo*** remete-se à categoria de profissionais da área de turismo.

Um ***turismólogo*** deseja conhecer quais as características demográficas, socioeconômicas e as de estilo de vida dos turistas envolvidos em um pacote turístico de sua agência de viagens.

Critérios para determinar se a pesquisa é necessária

Deve-se investir na coleta de informações?

Considerando os volumes relativos de investimentos que o mercado exige para realização adequada de produtos e serviços, supérfluo seria discutir a sua importância.

A pesquisa já há tempos vem sendo uma rotina nas empresas mais avançadas e já se desenvolve em pequenas e médias corporações, inclusive, em agência de turismo ou viagens.

Ninguém quer se arriscar a investir amadoristicamente, com base apenas em suposições. Ninguém quer "queimar" o seu serviço colocando-o inadequadamente no mercado.

O que importa é discutir o momento certo de uma pesquisa.

Para verificar a necessidade e o momento certo, devemos responder às seguintes perguntas:

a. Que informações são necessárias para responder ao meu problema de pesquisa?
b. As informações estão disponíveis e são suficientes para resolver o meu problema de pesquisa no momento?
c. As informações podem ser consultadas em tempo hábil na empresa ou em outras partes?

Se as respostas às ultimas perguntas forem negativas, está estabelecida a necessidade da pesquisa.

Então, podemos concluir que a pesquisa aparece nitidamente como necessidade, quando são trabalhadas e analisadas as informações já disponíveis ao pesquisador. Este observa que ainda resta um vazio, que algumas questões ainda não ficam respondidas, algumas dúvidas não se resolvem.

Exemplo:

No estudo do "Perfil dos turistas envolvidos em um pacote turístico em uma cidade", a motivação para a pesquisa poderá ser a de obter informações para que no futuro se aperfeiçoe o oferecimento do referido pacote turístico cada vez mais moldado ao perfil e à necessidade de quem o demanda.

Quando tudo for estabelecido, resume-se o que se planejou em um pequeno relatório escrito que se chama "Projeto de Pesquisa", com cronogramas, prazos e orçamentos e submete-se às autoridades da empresa ou a outra fonte de aprovação da realização da pesquisa.

Coleta de dados (trabalho de campo)

Para a coleta de dados, são mais utilizados os métodos de questionários, realização de entrevista ou a realização de observação.

A coleta de dados é a parte visível da pesquisa, é a abordagem do entrevistador para os entrevistados em uma rua de grande movimento, ou em suas residências, locais de trabalho, locais de estudo, para colher dados sobre as variáveis da pesquisa.

Em turismo, é comum localizar-se os respondentes (ou entrevistados) de pesquisa em locais ou pontos turísticos de uma cidade, hotéis, pousadas, aeroportos etc.

Fontes de dados

Há quatro diferentes fontes de dados em pesquisa:

a. pesquisado;
b. pessoas que tenham informações sobre o pesquisado;
c. situações similares;
d. dados disponíveis.

a. Pesquisado

Os próprios pesquisados (informantes) são a maior fonte de dados em pesquisa. O dado pode ser obtido do pesquisado pela sua própria declaração, oralmente ou por escrito, ou através de sua observação.

Pesquisado – essa fonte de dados é consultada mais comumente na prática, quando se aplica um questionário, geralmente estruturado, com perguntas objetivas e discursivas ou quando se observa, sem nenhum contato direto, o comportamento do elemento pesquisado, para se obter respostas a perguntas de uma pesquisa. Exemplo: um turista quando vai a uma loja fazer compras, ele geralmente pergunta o preço do produto desejado?

b. Pessoas que tenham informações sobre o pesquisado

Quando ocorre de o pesquisado ser inacessível, possuir pouco conhecimento da informação desejada ou ter dificuldade de se expressar, às vezes é mais fácil conseguir a informação com outras pessoas que convivem com ele.

Exemplos:

⇒ obter informações sobre as crianças de uma casa com a mãe;

⇒ obter informação sobre o marido com a esposa;

⇒ obter informação sobre o chefe com a secretária;
⇒ obter informação sobre o subordinado com o chefe;
⇒ obter informação sobre o cliente com o vendedor.

c. Situações similares

É a busca de conhecimento em situações análogas ou similares.

Exemplos:

⇒ Aprende-se bastante em saber:

1º. como lançar no mercado, para posterior venda, um passeio turístico através do mesmo procedimento já vivido por outra agência de turismo.

2º. Como outras agências do mesmo ramo trabalham com a variável estimulação ao turismo a uma cidade.

3º. Como foram as reações dos turistas quando determinado evento turístico foi lançado em outro país.

d. Dados disponíveis

Existe uma infinidade de dados úteis para negócios turísticos que já foram coletados, tabulados e, às vezes, até analisados que estão catalogados à disposição dos interessados.

Para obter esses dados, exige-se apenas o esforço de dedicar algum tempo a faculdades de turismo, visitas aos órgãos governamentais (por exemplo, Embratur), à leitura de jornais e revistas, à consulta de informações padronizadas de empresas de pesquisa e agências de turismo, e, principalmente no mundo atual, à pesquisa na internet.

Existe uma infinidade de informações sobre turismo, por exemplo, na internet em sites como http://www.world-tourism.org/, http://www.ibge.gov.br/home/, http://www.embratur.gov.br/.

Atividade

Atende ao Objetivo 1

1. Uma agência de viagens irá desenvolver uma pesquisa sobre o panorama de lazer e de cultura de turistas em visita à cidade do Rio de Janeiro. Os turistas seriam localizados em pontos turísticos da cidade, segundo um esquema de amostragem por quotas. De acordo com as informações disponíveis, responda às questões a seguir:

1. Quais as metas da pesquisa?

2. Defina a população.

3. Identifique cada membro da população.

4. Identifique o esquema de amostragem.

5. Decida o método de coleta de dados que poderá ser utilizado (questionário postal, entrevista, etc.).

6. Elabore um questionário (igualmente apropriado para entrevistas pessoais e observação).

7. Sugira um esquema de treinamento para qualquer pessoa envolvida no processo de coleta de dados.

Comentário

A resposta à atividade deve seguir a sugestão das fases do planejamento de uma pesquisa.

Tipos de dados

Dados primários

Os dados primários são aqueles que nunca foram coletados antes, estando ainda em posse dos pesquisados, e que são coletados com o propósito de atender às necessidades específicas da pesquisa.

Ex.: Numa pesquisa de "Satisfação de turistas", os resultados da aplicação do questionário ou da realização da entrevista constituem dados primários.

As fontes básicas de dados primários são:

- pesquisado;
- pessoas que tenham informações sobre o pesquisado;
- situações similares.

Dados secundários

Os dados secundários são aqueles que já foram coletados, tabulados, e ás vezes até analisados e que estão catalogados à disposição dos interessados.

Exemplos:

a. Índice Nielsen Alimentar

Esse índice coleta e processa dados sobre grande número de produtos (bebidas, enlatados, sabonetes, sabões em pó etc.), a cada dois meses, de um painel (conjunto) de lojas de alimentos em todas as regiões metropolitanas dos países.

Os dados coletados e processados incluem compras, vendas, estoques e preços das diversas marcas existentes no mercado, comercializadas pelas lojas do painel.

As informações fornecidas aos clientes compreendem volume de vendas por categoria, por marca, por região, preços praticados etc.

Índice de Audiência – TV (Ibope)

Esse índice faz coleta de audiência já existente em 400 residências em São Paulo, está atualmente sendo expandido, para outros estados brasileiros. Aparelhos instalados nos televisores remetem a cada 30 segundos a audiência dos programas de televisão no estado de São Paulo, via telefone, para uma central de processamento eletrônico, que instantaneamente processa, e os envia aos interessados, via computador e linha telefônica.

Uma variação da pesquisa painel que tem sido frequentemente utilizada por agências de pesquisa é o *Painel Omnibus*, que consiste em uma amostra permanente de uma população que é sempre consultada. Contudo, as informações pesquisadas podem variar conforme as necessidades dos clientes contratantes da pesquisa.

Painel Omnibus é um tipo de pesquisa realizada em intervalos fixos (por exemplo, de quatro em quatro anos) em que se investiga sempre uma mesma amostra de respondentes. Contudo as variáveis do estudo podem ser alteradas de acordo com os interesses dos pesquisadores na época da realização da pesquisa.

Exemplo de pesquisa com dados secundários:

Um empresário do turismo deseja conhecer quais os roteiros mais procurados de um país, para colocar no roteiro de um pacote turístico a ser oferecido por sua agência. Para isso, consulta outros pacotes turísticos oferecidos por outras agências de viagens e seus destinos mais demandados para se basear e organizar o seu pacote turístico.

As fontes básicas de dados secundários são:

⇒ a própria empresa;

⇒ publicações;

⇒ governos;

⇒ instituições não governamentais;

⇒ serviço padronizado de informações em marketing;

⇒ internet.

Atividade

Atende ao Objetivo 1

2. Indique qual a fonte de dados que foi utilizada nos casos a seguir:

a. Para saber os hábitos culturais e de lazer de certos clientes tradicionais de uma agência de turismo, foram consultados os estabelecimentos culturais e de lazer que os mesmos visitaram em destinos passados.

b. Para saber o que os maridos mais apreciaram em uma visita a museus de uma cidade, as esposas deles foram abordadas.

c. Para estabelecer um plano para um grupo de turistas de um país, foram consultados planejamentos turísticos envolvendo pessoas com o perfil do grupo em foco, realizados por outras agências de viagens.

d. Para uma pesquisa de consumo, visitantes de um país foram abordados para entrevistas em pontos turísticos.

e. A internet foi consultada para se obter dados de faturamento de hotéis com turistas holandeses.

Resposta

a. Dados disponíveis.

b. Pessoas que tenham informações sobre o pesquisado.

c. Situações similares.

d. O próprio pesquisado.

e. Dados disponíveis.

Sequência na procura de dados em pesquisas

É comum as pessoas imaginarem que a única forma de se obter dados em pesquisa de negócios seja através de um levantamento de campo. Na verdade, os levantamentos de campo e outras modalidades de coleta de dados primários somente deverão ser usados se outras formas mais rápidas, baratas e eficientes não conseguirem atender às necessidades de dados da pesquisa.

Um grande esforço nos estágios iniciais da pesquisa deverá ser canalizado para procurar tentar descobrir se, ao menos em parte, os dados necessários estejam de alguma forma disponíveis.

Este esforço inicial poderá significar grande economia de tempo, dinheiro e energia na realização da pesquisa.

Na **Tabela 3.1**, é apresentada a sequência dos passos, que normalmente deve ser seguida na definição dos dados e das fontes de dados no processo de pesquisa em negócios de turismo.

Tabela 3.1: Etapas para definição dos dados e das fontes de dados no processo de pesquisa

1.	Definir os objetivos da pesquisa.
2.	Especificar as necessidades de dados.
3.	Planejar as etapas da pesquisa.
4.	Determinar as fontes de dados.
5.	Procurar dados secundários internos.
6.	Procurar dados secundários externos: Publicações: – gerais; – governamentais; – institucionais. Governos: – federal; – estadual; – municipal. Instituições não governamentais: – universidades, faculdades e centros de pesquisas; – associações patronais e de empregados; – sindicatos patronais e de empregados. Serviços padronizados de informações de marketing: – do consumidor; do varejo; do atacado; da indústria.

7.	Determinação das necessidades de dados primários.
8.	Determinação das fontes de dados primários: – pesquisado; – pessoas que tenham informações sobre o pesquisado. – situações similares: • estudo de casos; • experimentos.

Com a presente aula, aprendemos como planejar uma pesquisa estatística para área de turismo, bem como as técnicas disponíveis para a coleta de dados. As fases do planejamento e da coleta são muito importantes para uma pesquisa, porque definem o esquema de pesquisa e geram as informações pelas quais vão se basear todas as análises, interpretação e tomada de decisão administrativa.

Uma vez aprendido os conceitos referidos, podemos agora investir no aprendizado das outras fases do método estatístico que sejam a crítica de dados, apuração de dados, análise de dados, redação do relatório final e comunicação dos resultados. Superadas também essas fases, o pesquisador estará pronto para realizar qualquer inquérito de caráter estatístico.

Atividade Final

Atende ao Objetivo 2

Imagine uma pesquisa cujo objetivo seja identificar características gerais de um pacote turístico capaz de satisfazer às expectativas de visitantes norte-americanos no período de janeiro a março de um determinado ano. Apresente uma sequência de passos no processo de definição dos dados e das fontes de dados necessários para obter as respostas da pesquisa referida.

Resposta Comentada

O pesquisador da agência organizadora do pacote turístico deve definir as metas da pesquisa, sua motivação e procurar fonte de dados internos da empresa. Caso não existam as informações desejadas internamente, elas devem ser procuradas externamente. Não sendo encontradas em fontes externas, configura-se, então, de se obterem os dados em fontes primárias de pesquisa. Qualquer dúvida, consulte a **Tabela 3.1**.

Resumo

As fases do planejamento e coleta de dados são fundamentais no processo de revitalização de um trabalho estatístico.
Na fase do planejamento, são determinados os objetivos da pesquisa, sua finalidade ou motivação e as principais seqüências para se obter a real necessidade de realização da pesquisa.
Já na fase da coleta de dados, as informações são obtidas junto às fontes de dados para crítica e apuração.

Informação sobre a próxima aula

Até este momento estudamos do planejamento à coleta de dados em pesquisa estatística no turismo. Nas próximas aulas, abordaremos as outras fases de um estudo estatístico.

4 Fases do método estatístico – da crítica à comunicação dos resultados

Meta da aula

Apresentar as etapas da crítica à comunicação dos resultados para realização de um trabalho estatístico.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. identificar as fases da crítica, apuração, análise, redação do relatório final e comunicação dos resultados da pesquisa de um trabalho estatístico;

2. redigir e interpretar resultados de pesquisa estatística.

Pré-requisitos

Para acompanhar esta aula, é importante que você tenha em mente os conceitos de fenômeno, fenômeno coletivo ou de massa, ciência, ciência estatística, população, amostra, divisão da estatística, presentes na Aula 1. Os métodos de seleção de amostras e dimensionamento do tamanho da amostra você encontra na Aula 2, além das fases do método estatístico apresentadas na aula anterior (Aula 3).

Introdução

Nesta aula, apresentaremos as próximas fases do método estatístico: crítica de dados, apuração de dados, análise de dados, redação do relatório final e comunicação dos resultados. Portanto, esta aula é a continuação da aula anterior, a Aula 3, em que abordamos as duas primeiras fases do método estatístico: 1ª e 2ª fases. As duas primeiras fases envolveram o planejamento e a coleta de dados. Na Aula 4, abordaremos, então, da 3ª a 7ª fases.
Com o fim desta aula, estaremos aptos a realizar todas as fases de um trabalho estatístico dentro da área do turismo.

3ª – Crítica de dados

A crítica de dados é a preparação dos dados coletados para a apuração, constitui-se na verificação de respostas erradas, verifi-cação de lacunas deixadas pelos *pesquisados* e na quantificação de perdas naturais de informações. As perdas podem ocorrer, por exemplo, pela perda de questionários já preenchidos, pela verificação de um número extenso de "não respostas" às questões. Dependendo do tipo de "erro" apontado no questionário, opta-se pelo aproveitamento, ou não, dele.

4ª – Apuração de dados

Apuração de dados é a contagem ou soma dos resultados observados em cada variável especificada e medida no questionário.

Exemplo de apuração:

Suponha que na pesquisa do "Perfil dos turistas de uma cidade", uma variável observada seja "sexo" e que do resultado da coleta em 10 turistas tivéssemos encontrado:

Indivíduo 1: Masculino,

Indivíduo 2: Feminino,

Indivíduo 3: Masculino,

Indivíduo 4: Masculino,

Indivíduo 5: Masculino,

Indivíduo 6: Feminino,

Indivíduo 7: Masculino,

Indivíduo 8: Feminino,

Indivíduo 9: Feminino,

Indivíduo 10: Masculino.

A apuração de dados, neste caso, é a contagem de quantas pessoas do sexo masculino e do sexo feminino foram observadas.

Tabela 4.1: Quantidade de turistas segundo o sexo

Sexo	**Apuração**
Masculino	IIIIII
Feminino	IIII
Total	IIIIIIIIII

Esse tipo de apuração é a *manual*.

Geralmente o número de resultados observados é bem maior do que dez. Então, a apuração se torna mais trabalhosa ou mesmo operacionalmente inviável. Nesse caso, deve ser realizada eletronicamente através de um pacote específico para apuração. Existem vários programas computacionais voltados para a análise de dados estatísticos, entre eles, o *SPSS*, *Statistica* e o *SAS*.

O SPSS, o Statistica e o SAS são *softwares* estatísticos destinados a pesquisas, principalmente nas áreas das ciências humanas e sociais.

5ª Análise de dados

A análise de dados é a fase em que se estabelece a forma de representação dos dados de acordo com as normas de repre-sentação oficial de dados estatísticos. É a fase, também, em que técnicas estatísticas são usadas para se tirar informações importantes dos dados apurados. Nela são estabelecidas as tabelas para representar os dados, os gráficos necessários, porcentagens, totais, médias, inferências etc. Na análise de dados, os resultados da pesquisa são trabalhados para gerar informações úteis, isto é, os *achados da pesquisa*. Geralmente, com essa fase concluída, pode-se redigir o relatório final da pesquisa.

Exemplo:

Como forma de análise de dados, poderíamos elaborar um gráfico com os dados apurados da variável sexo dos turistas do exemplo anterior:

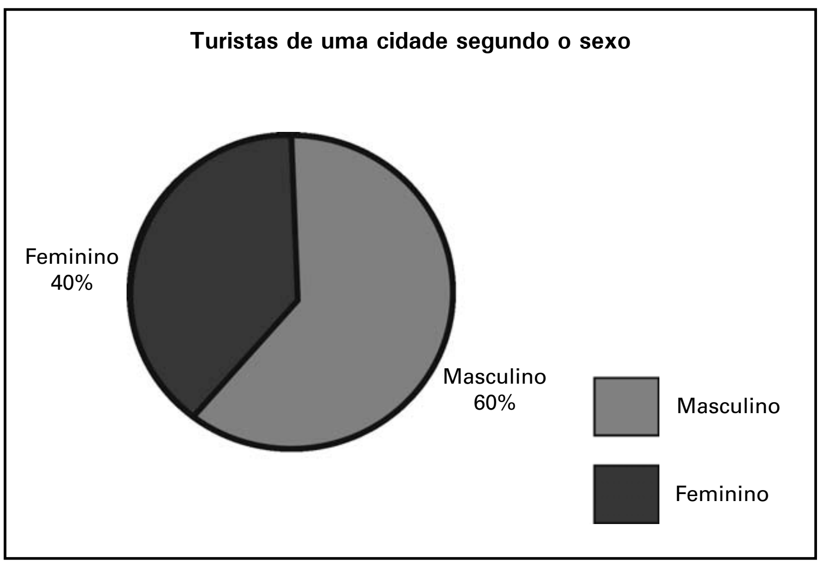


Figura 4.1: Quantidade de turistas segundo o sexo.

Atividade

Atende aos Objetivos 1 e 2

1. Correlacione as colunas de acordo com as fases executadas na pesquisa sobre "Gostos de turistas em um hotel":

(1) Planejamento

(2) Coleta de dados

(3) Crítica de dados

(4) Apuração de dados

(5) Análise de dados

( ) Os 100 questionários aplicados foram numerados e entregues para contagem, isto é, foram introduzidos em um *software* estatístico para gerarem tabelas de frequência das respostas.

( ) Os dados introduzidos no computador foram trabalhados de forma a produzirem gráficos estatísticos, médias e porcentagens, que respondam ao problema da pesquisa de saber os gostos dos turistas.

( ) Nos questionários, foram verificados que alguns turistas não responderam às questões sobre idade e faixa salarial. As não-respostas foram codificadas como dados faltantes.

( ) Uma pesquisa foi estruturada para se conhecer os gostos dos turistas de um hotel em Natal. Nessa fase, foram elaborados os questionários da pesquisa e o esquema de amostragem.

( ) À medida que os turistas iam dando entrada na recepção, os questionários eram entregues a eles para posterior devolução. Foram aplicados os questionários a uma amostra de 100 respondentes.

Resposta Comentada

(4),(5),(3),(1),(2).

O aluno deve recorrer às definições das fases do método estatístico dadas até o momento, inclusive o planejamento e a coleta de dados vistas na Aula 3, para identificar as fases a que se referem o item do caso enunciado na atividade.

6ª – Emissão de relatório final

Depois da análise de dados, redige-se um relatório final de pesquisa, acrescentando a este os resultados da análise, o problema estabelecido no planejamento, a metodologia da pesquisa, as conclusões obtidas e as tomadas de decisão.

O relatório final pode ser comunicado de forma técnica ou em artigo científico, dependendo do interesse dos envolvidos na pesquisa.

7ª – Comunicação dos resultados

Esta etapa do método torna o trabalho do cientista um processo social. As teorias, a metodologia e as conclusões do seu trabalho de pesquisa podem ser relatadas publicamente e sobreviver a um período de debate, avaliação crítica e à repetição dos ensaios e testes por outros profissionais competentes. Só assim os novos conhecimentos são incorporados à ciência universal. Em consequência, tem-se que o conhecimento é um bem público e acervo da humanidade. Além do objetivo de socialização dos resultados, os dados precisam ser comunicados por razões práticas: o homem de negócios precisa dos resultados da pesquisa para tomar decisões administrativas estratégicas com menos riscos.

Atividade

Atende aos Objetivos 1 e 2

2. Identifique nos casos a seguir a que fase se refere cada situação:

a. Um pesquisador de uma pousada redige um relatório contendo informações estatísticas sobre os gastos de turistas com diárias, almoços e telefonemas.

b. O relatório da pesquisa é divulgado para a gerência do estabelecimento que se nutre de informações de tendência de lucro e de estratégias para otimização de resultados.

Resposta Comentada

(a) Emissão de relatório final.
(b) Comunicação do resultados da pesquisa.

No item (a), um documento com os principais achados da pesquisa é redigido. Essa fase é a da emissão de relatório final. No item (b) remete-se a divulgação dos resultados da pesquisa à administração da pousada para as devidas tomadas de decisão. Essa é fase da comunicação dos resultados finais da pesquisa.

Atividade

Atende aos Objetivos 1 e 2

3. Justifique a necessidade de uma pesquisa de campo com questionário para a situação a seguir:

Uma empresa de turismo, para responder ao problema de pesquisa que consiste no traço do perfil econômico dos turistas de um país em uma época de alta temporada, definiu que seriam necessárias informações sobre nacionalidade, profissão, escolaridade, renda familiar, renda *per capita*, bairro de moradias, quantidade de bens móveis e fixos diversos, entre outras informações dos visitantes do país.

Essas características foram investigadas em bancos de dados já disponíveis, e foi constatado que a maioria delas já estava disponível, contudo não poderia ser consultada em tempo hábil pela agência de turismo, inviabilizando a utilização da informação econômica quanto à visita dos turistas ao país.

Resposta Comentada

Uma pesquisa de campo com questionário estruturado para essa situação se faz necessária, uma vez que esta seria, no caso apresentado, mais rápida do que a consulta a dados disponíveis, e as informações poderiam ser utilizadas quando da visita dos turistas ao destino.

Com esta aula, estaremos aptos a desenvolver as principais fases do método estatístico, do planejamento à comunicação dos resultados da pesquisa, com precisão e técnica, dentro do campo do turismo. É importante que internalizemos todas as fases do método estatístico e nessa ordem, pois somente assim estaremos nos utilizando de um método científico legítimo que nos levará aos resultados esperados: um conjunto de informações úteis que serão utilizadas nas tomadas de decisão e que implicarão minimizar o risco de decisões equivocadas. É importante lembrar que o bom desenvolvimento de um trabalho estatístico, que se propõe ser um estudo científico, deve ser baseado na objetividade e ética e também na experiência e vivência de vários anos desenvolvendo o método. Não é da noite para o dia que se faz um bom pesquisador, principalmente na área de estatística.

Atividade Final

Atende aos Objetivos 1 e 2

Leia o texto a seguir e indique à qual fase do método estatístico corresponde cada parágrafo:

(1) Um profissional de turismo de certa agência de viagens foi encarregado de realizar uma pesquisa estatística para determinar a imagem que os turistas tiveram de uma cidade visitada (destino), bem como o seu grau de satisfação.

(2) Os resultados seriam utilizados para tomada de decisões administrativas relativas à otimização futura da visão dos turistas do destino em foco e melhorar o seu grau de satisfação, para uma possível fidelização de visita à cidade.

(3) Em uma primeira reunião com a equipe da agência, foram definidas objetivamente a população-alvo, o esquema de amostragem e redigido o questionário para uma coleta de dados via internet, entre outras coisas.

(4) Os turistas teriam de responder à pesquisa a partir de um questionário que lhes seria enviado por e-mail.

(5) Após a aplicação, cada questionário foi validado para a pesquisa ou não.

(6) Com os questionários válidos, foi montada uma base de dados num *software* estatístico, e foram geradas tabelas de frequências das perguntas do inquérito.

(7) Foram gerados gráficos e estimativas estatísticas, e foram realizadas análises estatísticas.

(8) Os resultados das pesquisas foram apresentados em um relatório final que continha também o problema da pesquisa, a motivação da pesquisa, a metodologia da pesquisa, análises estatísticas, interpretação dos resultados, conclusões, recomendações, bibliografia e anexos.

(9) O relatório da pesquisa foi apresentado aos administradores da empresa que, com este, puderam providenciar a melhoria da imagem dos turistas em relação à cidade visitada e tomar iniciativas de maximizar o grau de satisfação dos visitantes do destino.

Identifique, então, com base no texto anterior, em que parágrafo foram abordados os temas a seguir:

Análise de dados (   )

Apuração de dados (   )

Coleta de dados (   )

Crítica de dados (   )

Divulgação dos resultados (   )

Motivação para a pesquisa (   )

Planejamento da pesquisa (   )

Problema da pesquisa (   )

Relatório final de pesquisa (   )

Respostas

7; 6; 4; 5; 9; 2; 3; 1 e 8.
Você deve recorrer às definições das fases do método estatístico dadas até o momento, inclusive o planejamento e a coleta de dados vistas na Aula 3, para identificar as fases a que se referem o item do caso enunciado na atividade.

Resumo

As fases do método estatístico é um conjunto de etapas dispostas em uma ordem específica para se desenvolver uma pesquisa estatística.

São sete as fases do método estatístico: o planejamento, a coleta de dados, a crítica de dados, a apuração de dados, a análise de dados, a emissão do relatório final e a comunicação dos resultados da pesquisa.

Com as fases desenvolvidas, podemos ter um trabalho estatístico completo e baseado em métodos científicos que permitem inferências válidas.

Informações sobre a próxima aula

Na próxima aula, trataremos do estudo da construção de tabelas estatísticas, normas de representação tabular e classificação de séries estatísticas, tão essenciais na fase da análise de dados.

5 Séries estatísticas

Metas da aula

Apresentar as regras para construção de tabelas estatísticas; classificar as tabelas estatísticas; definir os seus elementos e construir distribuições de frequências.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. identificar as normas para construção de tabelas estatísticas;
2. interpretar resultados de pesquisa estatística envolvendo apresentação de tabelas e gráficos.

Introdução

As séries estatísticas resumem (sumarizam) um conjunto ordenado de observações obtidas através do tempo, espaço e categoria do fenômeno. A organização dos dados provenientes da pesquisa em forma de uma tabela irá proporcionar ao pesquisador maior segurança na explanação dos resultados e conclusões e ao leitor uma compreensão mais rápida e eficiente dos achados da pesquisa. Por isso é tão comum em artigos científicos, relatórios de pesquisa, relatórios de trabalho, monografias etc. a utilização de tabelas para a exposição dos resultados.

Normas de representação tabular

Estas normas têm como objetivo orientar a apresentação racional e uniforme de dados estatísticos, em forma tabular, no Sistema Estatístico, subordinado à orientação normativa e supervisão técnica do IBGE.

Uma tabela estatística compõe-se de elementos essenciais e elementos complementares. Os elementos essenciais de uma tabela são o título, o corpo, o cabeçalho e a coluna indicadora.

- Título é a indicação que precede a tabela e que contém a designação do fato observado, o local e a época em que foi registrado.
- Corpo é o conjunto de colunas e linhas e que contém, respectivamente, em ordem vertical e horizontal, as informações sobre o fato observado. A casa é o cruzamento de uma coluna com uma linha. As casas não deverão ficar em branco, apresentando sempre um número ou sinal convencional.
- Cabeçalho é a parte da tabela que especifica o conteúdo das colunas.
- Coluna indicadora é a parte da tabela que especifica o conteúdo das linhas. Uma tabela pode ter mais de uma coluna indicadora.

Os elementos complementares de uma tabela estatística são a fonte, as notas e as chamadas, e se situam de preferência no rodapé da tabela.

- Fonte é a indicação da entidade responsável pelo fornecimento dos dados ou pela sua elaboração.
- Notas são informações de natureza geral, destinadas a conceituar ou esclarecer o conteúdo das tabelas ou indicar a metodologia adotada no levantamento ou elaboração dos dados.
- Chamadas são informações de natureza específica sobre determinada parte da tabela, destinadas a conceituar ou esclarecer dados. As chamadas são indicadas no corpo da tabela em algarismos arábicos, entre parênteses, à esquerda nas casas e à direita na coluna indicadora. A numeração das chamadas na tabela será sucessiva, de cima para baixo, e da esquerda para direita. A distribuição das chamadas no rodapé da tabela obedecerá à ordem de sua sucessão na tabela, separando-se umas das outras por um parantêse. As chamadas de uma tabela que ocupam mais de uma página devem figurar no rodapé da tabela na última página, de acordo com sua sucessão na mesma.

Exemplo de tabela estatística de acordo com as normas do IBGE:

Tabela 5.1: Número de casos de doenças epidemiológicas e óbitos segundo especificação do estado do Rio de Janeiro, 1970

Especificação	Nº de casos	Nº de óbitos
Sarampo	1) 115	15
Varíola	. . .	10
Cólera	——	——
Meningite	35	5
Total	150	30

Fonte: Departamento Nacional de Endemias Rurais.

Nota: As atividades da campanha de vacinação abrangeram as áreas de maior incidência.

1) Inclusive a área rural.

Séries estatísticas

É a representação das informações em forma de tabelas. Seu objetivo é obter um resumo organizado das informações sobre a variável; fornecer o máximo de informações em um mínimo de espaço.

As séries estatísticas podem ser:

a. Séries temporais;

b. Séries geográficas;

c. Séries especificativas;

d. Séries mistas;

e. Distribuições de frequência.

a. Séries temporais

Nesta série, observa-se a evolução, ao longo do tempo, de determinado acontecimento ou fato. Assim, o que varia é o tempo, ou seja, uma determinada informação é estudada em função do tempo.

Exemplo:

Tabela 5.2: Quantidade de turistas de uma agência de viagens que sofreram furtos nas orlas do Rio de Janeiro, de janeiro a outubro de 2006

Meses	Quantidades
Janeiro	10
Fevereiro	8
Março	9
Abril	3
Maio	1
Junho	1
Julho	1
Agosto	2
Setembro	3
Outubro	2
Total	**40**

Fonte: Dados hipotéticos.

b. Séries geográficas

Uma determinada informação é estudada em função de uma região ou localidade. É feita para apresentar dados de diferentes regiões geográficas.

Exemplo:

Tabela 5.3: Quantidade de turistas de uma agência de viagens que sofreram furtos no Rio de Janeiro, de janeiro a outubro de 2006 segundo as orlas

Orlas ou praias	Quantidade
Flamengo	5
Botafogo	2
Copacabana	10
Ipanema	8
Leblon	6
Barra	9
Total	**40**

Fonte: Dados hipotéticos.

c. Séries especificativas

Neste caso, a informação em estudo é dividida em categorias que a especificam.

Exemplo:

Tabela 5.4: Quantidade de turistas de um pacote turístico segundo o tipo de hospedagem preferida

Tipos de hospedagem	Quantidade
Hotel solteiro	20
Hotel casado	5
Pousada solteiro	10
Pousada casado	2
Apart hotel solteiro	2
Apart hotel casado	1
Total	**40**

Fonte: Dados hipotéticos.

d. Séries mistas

São aquelas séries estatísticas resultantes da combinação das séries estatísticas temporais, geográficas, especificativas ou entre distribuições de frequências.

Exemplo:

Tabela 5.5: Taxas de analfabetismo de pessoas acima de 15 anos, segundo a cor, nos censos demográficos de 1991 e 2000

Cor	*Censos*	
	1991	2000
Branca	11,9	8,3
Preta	31,5	21,5
Amarela	5,4	4,9
Parda	27,8	18,2
Indígena	50,8	26,1
Sem declaração	18,7	16,1
Brasil	**19,4**	**12,9**

Fonte: Dados hipotéticos.

Atividade

Atende ao Objetivo 1

1. Classifique as seguintes séries estatísticas:

Série 1.

Aplicação em R$ em turismo por agente financeiro no Brasil, 2005

Agente financeiro	Aplicação (R$)
BNDES	91.353,00
Banco do Brasil	1.081.238,00
Caixa Econômica Federal	680.821,00
Banco do Nordeste	109.376,00
Banco da Amazônia	15.985,00

Fonte: BRASIL. Ministério do Turismo.

Série 2.

Entrada de turistas argentinos no Brasil.

Ano	Número de turistas
1994	787.117
1995	1.467.922
1996	1.548.571

Fonte: Organização Mundial do Turismo.

Série 3.

Número de agências de turismo por ano segundo região brasileira.

Região	Ano		
	2003	2004	2005
Norte	326	380	415
Nordeste	1.146	1.346	1.449
Sudeste	3.766	3.845	4.223
Sul	2.003	2.080	2.217
Centro-oeste	683	770	826

Fonte: BRASIL. Ministério do Turismo.

Série 4.

Número de agências de turismo por região brasileira, 2005.

Região	Agência de turismo
Norte	415
Nordeste	1.449
Sudeste	4.223
Sul	2.217
Centro-oeste	826

Fonte: BRASIL. Ministério do Turismo

Respostas

Série 1. Série especificativa.
Série 2. Série temporal.
Série 3. Série mista (geográfica e temporal).
Série 4. Série geográfica.

e. Distribuições de frequência

Quando a massa de dados que temos em mãos é grande e se refere a uma única variável (por exemplo: muitas medidas de pesos de bagagens; muitas opiniões a respeito de um determinado pacote turístico etc.), fica difícil para o pesquisador obter alguma informação apenas com a observação dos ***dados brutos***, por isso recomenda-se a construção de uma tabela de distribuição de frequência. Com essa tabela, o pesquisador irá distinguir mais facilmente a distribuição dos dados e realizar interpretações mais rápidas sobre os dados em questão.

Dados brutos

Dados provenientes de pesquisa sem nenhuma manipulação.

Tipos de distribuição de frequência

- Distribuição de frequência simples;
- Distribuição de frequência por classe.

Distribuição de freqüência simples

A cada resultado, ou ocorrência, de uma variável quantitativa é registrada nas células da tabela a frequência com que foi observada na coleta de dados.

Exemplo:

As informações sobre a idade de um grupo de turistas de um pacote turístico em visita ao Rio de Janeiro em julho/2001 foram dispostas na seguinte tabela de distribuição de frequência.

Assim, pela **Tabela 5.6** podemos observar que 15 turistas têm 27 anos, 23 turistas têm 32 anos e assim sucessivamente.

Tabela 5.6: Distribuição da idade dos turistas

Idade	Número de turistas
27	15
32	23
35	32
38	35
42	43
47	28
53	10
Total	**186**

Atividade

Atende aos Objetivos 1 e 2

2. Perguntou-se a turistas, em determinado restaurante, qual o grau de satisfação (A-alto; M-médio; B-baixo) quanto à variedade do cardápio. Tal pesquisa resultou em:

Turista	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Satisfação	A	M	A	M	M	B	A	M	B	M	B	B	M	M	A

Pede-se:

a. Construa uma tabela de frequências simples para a variável qualitativa ordinal grau de satisfação.

Responda:

b. Os turistas estão satisfeitos com o cardápio do restaurante? Justifique.

Respostas

a. Para facilitar a montagem desta tabela, recomenda-se primeiramente realizar a contagem das ocorrências,
Alta: IIII
Média: IIIIIII
Baixa: IIII
e assim colocar esse resultado em uma tabela de frequência simples:

Satisfação	***Número de turistas***
Alta	*4*
Média	*7*
Baixa	*4*
Total	*15*

b. Pela tabela de frequência, observa-se que muitos turistas avaliaram como médio o cardápio oferecido pelo restaurante. Assim, recomenda-se ao restaurante melhorar este serviço.

Distribuição de frequência por classe

Os subconjuntos de resultados de uma variável quantitativa são registrados nas células da tabela, como também, a frequência com que foram observadas na coleta de dados.

A distribuição de frequência por classe é aquela em que a variável observada está dividida em classes, que são subintervalos do intervalo total.

Formas de se representar uma classe:

[[ou |— → inclui à esquerda e exclui à direita

]] ou —| → exclui à esquerda e inclui à direita

[ ] ou |—| → inclui ambos

] [ou — → exclui ambos

Exemplo:

Tempo em segundos dos gastos por turistas para preencher o formulário de entrada em um hotel.

Tabela 5.7: Distribuição do tempo gasto (segundos)

Tempo (em segundos)	Número de turistas
40 |— 45	3
45 |— 50	8
50 |— 55	16
55 |— 60	12
60 |— 65	7
65 |— 70	3
70 |— 75	1
Total	**50**

Elementos de uma distribuição de frequência

a. Classes: são os subintervalos do intervalo total.

Exemplos:

40 $\vdash$ 45, primeira classe.

50 $\vdash$ 55, terceira classe.

70 $\vdash$ 75, última classe.

b. Limites de classe: são os valores que delimitam cada classe, sendo:

- da esquerda (menor valor), chamada limite inferior.
- da direita (maior valor), chamada limite superior.

Exemplo:

55 $\vdash$ 60

55, limite inferior.

60, limite superior.

c. Intervalo de classe (h): é o comprimento de cada classe. É obtido pela diferença entre o limite inferior de uma classe e o limite inferior da classe anterior ou a diferença existente entre o limite superior de uma classe e o limite superior da classe anterior. O intervalo de classe (h) deve ser uma constante na distribuição de frequência por classe.

Exemplos:

45 $\rightarrow$ limite inferior da segunda classe.

40 $\rightarrow$ limite inferior da primeira classe.

Logo, $h = 45 - 40 = 5$

Ou

50 $\rightarrow$ limite superior da segunda classe.

45 $\rightarrow$ limite superior da primeira classe.

$h = 50 - 45 = 5$

d. Ponto médio (x_i): é a média aritmética simples entre o limite superior e inferior de uma mesma classe. Todas as classes têm um ponto médio distinto. O ponto médio é o representante dos valores contidos em uma classe. O ponto médio varia de classe para classe.

Exemplo:

Tabela 5.8: Distribuição do tempo gasto com cálculo do ponto médio

Tempo (em segundos)	Número de turistas	Pontos médios x_i
40 ├─ 45	3	42,5
45 ├─ 50	8	47,5
50 ├─ 55	16	52,5
55 ├─ 60	12	57,5
60 ├─ 65	7	62,5
65 ├─ 70	3	67,5
70 ├─ 75	1	72,5
Total	**50**	

e. Frequência simples (f_i): também chamada de frequência absoluta. É o número de vezes que os valores de uma classe ocorreram na observação, isto é, a frequência com que foram observados.

Exemplo:

Tabela 5.9: Distribuição do tempo gasto com cálculo de frequência simples

Tempo (em segundos)	Frequências simples f_i
40 ├─ 45	3
45 ├─ 50	8
50 ├─ 55	16
55 ├─ 60	12
60 ├─ 65	7
65 ├─ 70	3
70 ├─ 75	1
Total	**50**

f. Frequência relativa (f_{ri}): é o quociente entre a frequência simples da respectiva classe pela frequência total. A soma das frequências relativas é sempre igual a 1.

Exemplo:

Tabela 5.10: Distribuição do tempo gasto com cálculo da frequência relativa

Tempo (em segundos)	Número de turistas	Frequências relativas f_{ri}
40 ├─ 45	3	0,06
45 ├─ 50	8	0,16
50 ├─ 55	16	0,32
55 ├─ 60	12	0,24
60 ├─ 65	7	0,14
65 ├─ 70	3	0,06
70 ├─ 75	1	0,02
Total	**50**	**1,00**

Observação:

Para a **Tabela 5.10**, utilizam-se duas casas decimais nas aproximações.

Se quisermos obter a frequência percentual ($f_{ri}\%$), basta multiplicarmos as frequências relativas por 100.

Exemplo:

Tabela 5.11: Distribuição do tempo gasto com cálculo da frequência percentual

Tempo (em segundos)	Número de turistas	Frequências percentuais $f_{ri}\%$
40 ├─ 45	3	0,06 . 100 = 6
45 ├─ 50	8	0,16 . 100 = 16
50 ├─ 55	16	0,32 . 100 = 32
55 ├─ 60	12	0,24 . 100 = 24
60 ├─ 65	7	0,14 . 100 = 14
65 ├─ 70	3	0,06 . 100 = 6
70 ├─ 75	1	0,02 . 100 = 2
Total	**50**	**100**

g. Frequências acumuladas (FAC_i): são obtidas através da adição sucessiva à frequência simples da primeira classe das frequências simples das classes seguintes. A primeira frequência simples é a primeira frequência acumulada e a última frequência acumulada é o somatório total das frequências absolutas, isto é, a frequência total n).

Exemplo:

Tabela 5.12: Distribuição do tempo gasto com cálculo da FAC

Tempo (em segundos)	Número de turistas	Frequências acumuladas (FACs)
40 ├─ 45	3	3
45 ├─ 50	8	11
50 ├─ 55	16	27
55 ├─ 60	12	39
60 ├─ 65	7	46
65 ├─ 70	3	49
70 ├─ 75	1	50
Total	**50**	—

Atividade

Atende aos Objetivos 1 e 2

3. Construa uma tabela de distribuições de frequência com intervalos de classes para os dados de uma amostra colhida de estaturas em centímetros de crianças nascidas numa maternidade:

44 42 46 58 52 56 50 49 51 60
60 40 48 52 58 50 51 53 56 49
40 45 47 44 50 57 60 58 48 52
50 39 41 48 47 45 54 57 55 43
59 59 51 53 54 49 55 52 43 38
48 46 36 42 48 44 49 51 53 45
51 50 59 43 49 55 48 49 52 47
50 51 54 55 54 43 37 46 50 39
44 53 56 57 47 38 40 41 56 52
53 56 52 54 51 42 53 54 50 55

Resposta Comentada

1º. Apurar os dados:

Medidas (em centímetros)	**Apuração**
36	/
37	/
38	//
39	//
40	///
41	//
42	///
43	////
44	////
45	///
46	///
47	////
48	//////
49	//////
50	////////
51	///////
52	///////
53	//////
54	//////
55	/////
56	/////
57	///
58	///
59	///
60	///
Total	**100**

2º. Obter a distribuição de frequência simples:

Medidas (em centímetros)	*Frequências (f_i)*
36	*1*
37	*1*
38	*2*
39	*2*
40	*3*
41	*2*
42	*3*
43	*4*
44	*4*
45	*3*
46	*3*
47	*4*
48	*6*
49	*6*
50	*8*
51	*7*
52	*7*
53	*6*
54	*6*
55	*5*
56	*5*
57	*3*
58	*3*
59	*3*
60	*3*
Total	***100***

3º. Obter o número de classes e o intervalo de classes:

Para obter o número de classes e o intervalo de classes, deveremos aplicar as fórmulas a seguir:

K = 1 + 3,3 log N, *onde:*

K = número de classes sugeridas, arredondado para inteiro.
N = frequência total.
h = R / K, *onde:*

R = amplitude total, isto é, diferença entre o maior e o menor valor da distribuição.

Cálculos para o exemplo:

K = 1 + 3,3 log 100 = 8 classes

$$h = \frac{60 - 36}{8} = 3$$

Observação: *o valor de K é o número de classes previstas para a distribuição. No momento da montagem da tabela, pode ocorrer de se obter uma classe a mais ou uma classe a menos do que o calculado na fórmula de K.*

4º. Obter a distribuição de frequência com intervalo de classes:

Medidas (em centímetros)	**Frequências (f_i)**
36 ├─ 39	4
39 ├─ 42	7
42 ├─ 45	11
45 ├─ 48	10
48 ├─ 51	20
51 ├─ 54	20
54 ├─ 57	16
57 ├─ 60	9
60 ├─ 63	3
Total	**100**

Conclusão

Conforme visto nesta aula, os dados organizados em tabelas facilitam a exposição e conclusão dos resultados.

Atividades Finais

Atendem aos Objetivos 1 e 2

1. Classifique as séries estatísticas a seguir:

a.

Crianças não vacinadas contra pólio, 1989.

Regiões	Quantidade
Nordeste	512.900
Sudeste	299.585
Norte	148.818
Centro-oeste	124.791
Sul	105.371

Fonte: IBGE.

b.

Avicultura brasileira, 1988.

Espécie	Número (cabeças)
Galinhas	511.834
Patos, marrecos	58.888
Perus	3.823

Fonte: IBGE.

c.

Tempo de licença sem vencimento de funcionários de uma empresa.

Tempo (em dias)	Quantidade
1	10
3	15
4	25
5	31
6	17
7	28
8	5
9	29
10	34

Fonte: Dados hipotéticos.

d.

Número de suicídios ocorridos no Brasil em 1986, segundo a causa atribuída.

Causa atribuída	Nª
Alcoolismo	263
Dificuldade financeira	198
Doença mental	700
Outro tipo de doença	189
Desilusão amorosa	416
Outras	217

Fonte: IBGE.

Respostas

a. série geográfica
b. série especificativa
c. série temporal
d. série especificativa

2. Elabore tabelas estatísticas com os relatórios descritos a seguir:

a. Na Agência Bete Turismo, empresa situada no Rio de Janeiro em 2007, um estatístico realizou um levantamento do número de consultas de turistas do plano de saúde conveniado com a empresa por especialidade médica. Na ortopedia foram 200 turistas, na clínica geral foram 432, na urologia foram 154, na psiquiatria foram 133, na ginecologia foram 220 e na dermatologia foram 90 pacientes.

b. Em 2007, o mesmo profissional do item anterior, resolveu acrescentar a variável sexo ao seu levantamento e constatou: na ortopedia foram 130 indivíduos do sexo masculino e 100 do feminino, na clínica geral foram 153 do sexo masculino e 165 do feminino; na urologia foram 145 turistas, na psiquiatria foram 112 do sexo masculino e 80 do feminino, na ginecologia foram 129 do sexo feminino e na dermatologia foram 29 do sexo masculino e 100 do feminino.

c. O levantamento feito pelos nutricionistas de uma pousada turística revelou a quantidade de refeições fornecidas aos hóspedes segundo os vários turnos de trabalho:

Desjejuns: 74.052;

Almoço: 72.249;

Merenda: 72.178;

Jantar: 72.523;

Ceias: 72.999.

Nas informações, não foram incluídas no total refeições extras, mamadeiras e dietas especiais. Na ceia e no jantar, não são oferecidas sobremesas.

Respostas

a.

Especialidade médica	**Número de consultas**
Ortopedia	*200*
Clínica geral	*432*
Urologia	*154*
Psiquiatria	*133*
Ginecologia	*220*
Dermatologia	*90*

b.

Especialidade médica	**Sexo**	
	Feminino	**Masculino**
Ortopedia	*100*	*130*
Clínica geral	*165*	*153*
Urologia	*0*	*145*
Psiquiatria	*80*	*112*
Ginecologia	*129*	*0*
Dermatologia	*100*	*29*

c.

Refeição	**Quantidade**
Desjejuns	*74.052*
Almoço	*72.249*
Merenda	*72.178*
Jantar	*72.523*
Ceias	*72.999*

3. A tabela a seguir lista os salários e o nível de escolaridade de dez funcionários de determinada agência de turismo.

Funcionário	Salário	Escolaridade
1	1.093,3	fundamental
2	1.159,3	médio
3	1.172,6	médio
4	1.187,9	médio
5	1.312,6	médio
6	1.344,8	médio
7	1.699,2	superior
8	1.729,9	superior
9	1.799,7	superior
10	1.828,7	superior

Pede-se:

a. Classificar as variáveis em estudo.

b. Construir a tabela de distribuição de frequências para cada uma das variáveis, contendo: frequências simples, proporção, porcentagem e frequências acumuladas. (Considere as 3 classes e h = 250 para a variável quantitativa.)

Respostas

a. Salário = Variável quantitativa contínua.

Escolaridade = Variável qualitativa nominal.

b. Tabela de distribuição de frequência para as variáveis em estudo

Classes	***xi***	***ni***	***fi***	***%***	***f_{ac}***
[1093.3; 1343.3)	*1.218,3*	*5*	*0,5*	*50*	*50*
[1343.3; 1593.3)	*1.468,3*	*1*	*0,1*	*10*	*60*
[1593.3; 1843.3)	*1.718,3*	*4*	*0,4*	*40*	*100*
Total		***10***	***1***	***100***	

Escolaridade	***ni***	***fi***	***%***	***f_{ac}***
Fundamental	*1*	*0,1*	*10*	*10*
Médio	*5*	*0,5*	*50*	*60*
Superior	*4*	*0,4*	*40*	*100*
Total	***10,0***	***1,0***	***100***	

4. Os salários dos funcionários de determinado hotel foram divididos nas seguintes classes (em salários mínimos):

Classe de salário	n_i	f_i	%	f_{ac}
1,0 ├─ 2,0	44			
2,0 ├─ 3,0	82			
3,0 ├─ 4,0	52			
4,0 ├─ 5,0	16			
5,0 ├─ 6,0	4			
6,0 ├─ 7,0	2			
Total				

Pede-se:

a. Completar a tabela de frequências.

b. Há uma justa distribuição de salários neste hotel? Por quê?

Respostas

a.

Classe	n_i	f_i	*%*	f_{ac}
1,0 |- 2,0	*44*	*0,22*	*22*	*0,22*
2,0 |- 3,0	*82*	*0,41*	*41*	*0,63*
3,0 |- 4,0	*52*	*0,26*	*26*	*0,89*
4,0 |- 5,0	*16*	*0,08*	*8*	*0,97*
5,0 |- 6,0	*4*	*0,02*	*2*	*0,99*
6,0 |- 7,0	*2*	*0,01*	*1*	*1,00*
Total	***200***	***1,00***	***100***	

b. Não, pois pode-se observar pela tabela de distribuição de frequências, mais precisamente pela frequência acumulada, que 63% dos funcionários, ou seja, mais da metade dos funcionários, recebem entre 1 e 3 salários mínimos e apenas 1% recebe entre 6 e 7 salários mínimos.

6 Gráficos estatísticos para representar a distribuição das variáveis qualitativas

Metas da aula

Apresentar os objetivos dos gráficos estatísticos, as regras para a sua construção, classificação e utilização em diferentes situações práticas, bem como realizar a correta interpretação gráfica.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. classificar os gráficos estatísticos para as variáveis qualitativas;
2. construir gráficos estatísticos utilizando o programa computacional Excel (versão 2003);
3. realizar a interpretação da informação gráfica para diferentes situações práticas.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é necessário que você tenha incorporado os conceitos de variáveis e seus tipos apresentados na Aula 1, séries estatísticas e tabela de distribuição de frequências presentes na Aula 4.

É importante que você procure relacionar o conteúdo desta aula com o que você encontra sobre o assunto na mídia em geral. Seria bom que procurasse reconstruir os exemplos numa planilha eletrônica, como, por exemplo, o Excel, o Open Office etc.

Introdução

A Estatística faz uso de recursos numéricos e visuais para representar os fenômenos. O gráfico é usado para ilustrar ou ressaltar um fenômeno através de uma imagem, o que o torna de mais fácil compreensão.

Em virtude de a leitura e a visualização de um gráfico serem mais fáceis e rápidas do que a leitura de uma tabela, os meios de comunicação (jornais, revistas, televisão, páginas da internet – ver boxe multimídia), artigos científicos, relatórios técnicos, exposições orais etc. utilizam desse dispositivo para melhor ilustrar e apresentar os resultados.

Ressalta-se, porém, que, apesar da abrangência e utilidade dos dispositivos gráficos, neles não estão expostas todas as observações numéricas; assim, apenas com a observação de um gráfico, muitas vezes não é possível recompor o conjunto de dados que o originou.

Nesta aula, você irá aprender a construir os gráficos utilizados para representar a distribuição das variáveis qualitativas.

No exemplo de gráfico presente na mídia internet – no endereço http://www.turismo.gov.br/ na página principal de dados e fatos – você poderá observar a utilização de gráficos estatísticos.

Definição

O gráfico estatístico é um método de representação de dados estatísticos por meio de figuras geométricas.

Requisitos fundamentais de um gráfico

Quando construímos um gráfico, devemos observar que ele tenha:

- Simplicidade: deve-se tomar cuidado para que a quantidade de informações presentes no gráfico não seja excessiva.
- Clareza: os valores numéricos, legendas e títulos devem estar bem descritos e localizados no gráfico.
- Exatidão: utilizar as dimensões adequadas em cada um dos eixos de valores (**Figura 6.1**).

Eixos de valores são dois eixos perpendiculares entre si, sendo o ponto de interseção denominado origem. Os valores das grandezas envolvidas são colocados utilizando uma escala adequada para cada eixo. O eixo na horizontal (por convenção) é denominado eixo das abscissas ou eixo x. O eixo na vertical é denominado eixo das ordenadas ou eixo y.

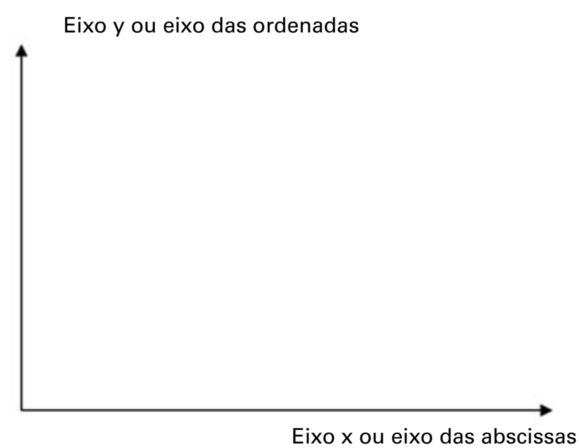


Figura 6.1: Plano cartesiano bidimensional.

Finalidades dos gráficos

- Apresentar os dados de modo agradável e claro.
- Poupar tempo e esforço na análise.
- Dispor os dados de modo a focalizar as comparações de maneira rápida.
- Tornar claros os fatos que possam ser objetos de confusão, por exemplo: duplicidade de dados; dados discrepantes (aqueles cujos valores são muito acima ou muito abaixo dos valores da maioria dos dados do conjunto de dados em estudo) etc.
- Ter uma visão mais geral do que ocorre com a variável objeto do seu estudo/pesquisa, sem precisar ficar observando sequências de números intermináveis.

Construção de gráficos

Antes de iniciarmos a construção de um gráfico, devemos considerar que:

- Todo gráfico deve ter escala e título.
- O título deve ser localizado na parte superior do gráfico.
- As variáveis devem estar identificadas nos dois eixos do gráfico.
- No eixo das abscissas, a escala cresce da esquerda para a direita e no das ordenadas, de baixo para cima.
- Outras informações adicionais, como rodapé, legenda e fonte, podem, ou não, estar explicitadas no gráfico.

Todos os gráficos estudados podem ser facilmente construídos, por exemplo, no programa computacional Excel por meio dos comandos: inserir → gráfico → tipo de gráfico; conforme as **Figuras 6.2** e **6.3**. Existem outros programas computacionais e planilhas eletrônicas nos quais é possível a construção de gráficos, mas no nosso curso iremos utilizar o Excel, devido à sua facilidade, simplicidade e acessibilidade.

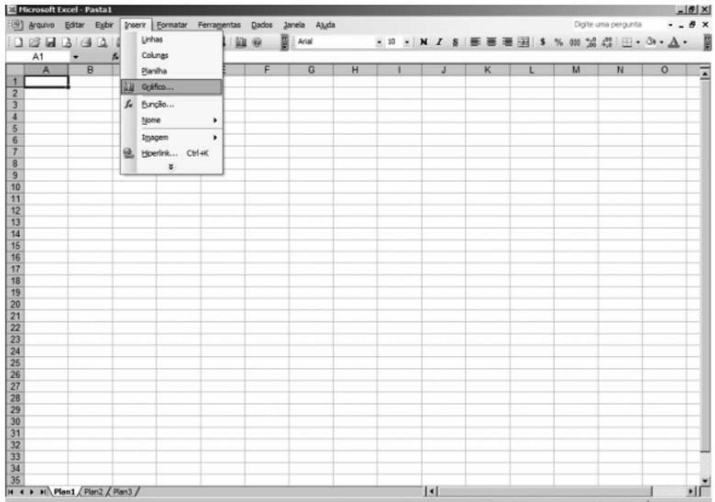


Figura 6.2: Comando inserir gráfico do Excel.

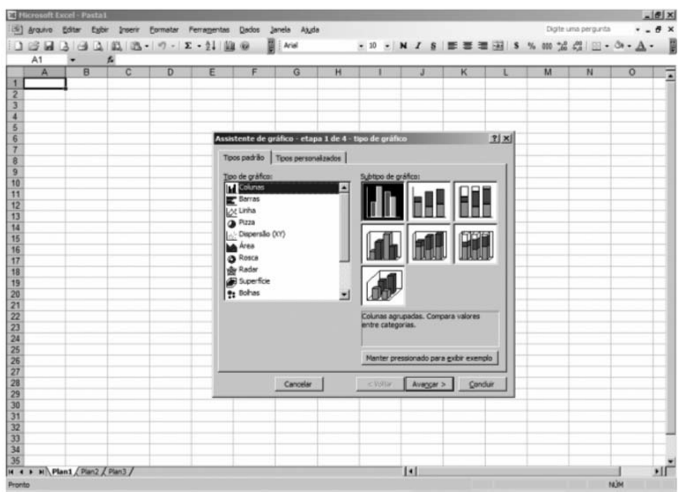


Figura 6.3: Comando tipo de gráfico do Excel.

Tipos de gráficos para representar a distribuição das variáveis qualitativas

Os tipos de gráficos mais utilizados para representar a distribuição das variáveis qualitativas são:

a. Colunas

b. Barras

c. Setograma ou "pizza"

a. Gráfico de colunas

É o tipo mais utilizado para representar as variáveis qualitativas. Pode acomodar um número arbitrário de categorias no eixo x, e seus respectivos valores numéricos no eixo y.

As características do gráfico de colunas são:

- é representado por retângulos dispostos verticalmente;
- a escala de valores numéricos está presente apenas no eixo das ordenadas ou eixo y;
- a largura da coluna não tem significado, ou seja, podemos ter qualquer largura de coluna.

Exemplo:

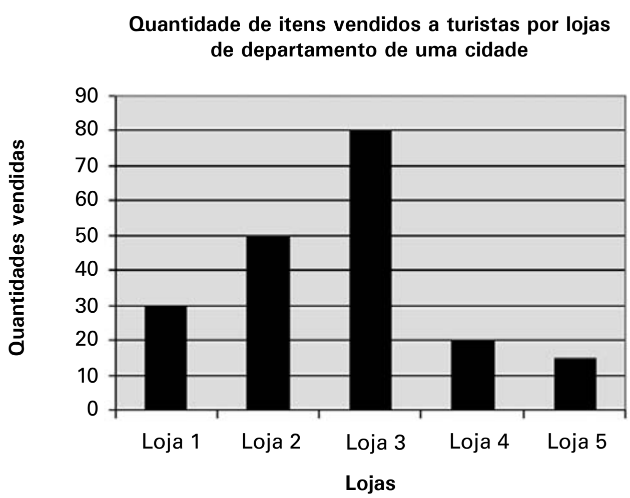


Figura 6.4: Exemplo de um gráfico de colunas.

Interpretação do gráfico

Observando a **Figura 6.4**, podemos concluir que há maiores vendas na loja 3 e as menores vendas estão concentradas nas lojas 4 e 5. As quantidades vendidas nas lojas 1 e 2 podem ser consideradas moderadas na comparação com as lojas 3, 4 e 5.

b. Gráfico de barras

Faz a mesma representação que o gráfico de colunas, porém nesta representação a escala de valores está presente no eixo x. Assim, é utilizado para representar um número arbitrário de categorias no eixo y e seus respectivos valores numéricos no eixo x.

As características do gráfico de barras são:

- é representado por retângulos dispostos horizontalmente;
- a escala de valores numéricos está presente apenas no eixo das abscissas ou eixo x;
- a largura da barra não tem significado, ou seja, podemos ter qualquer largura de barra.

Exemplo:

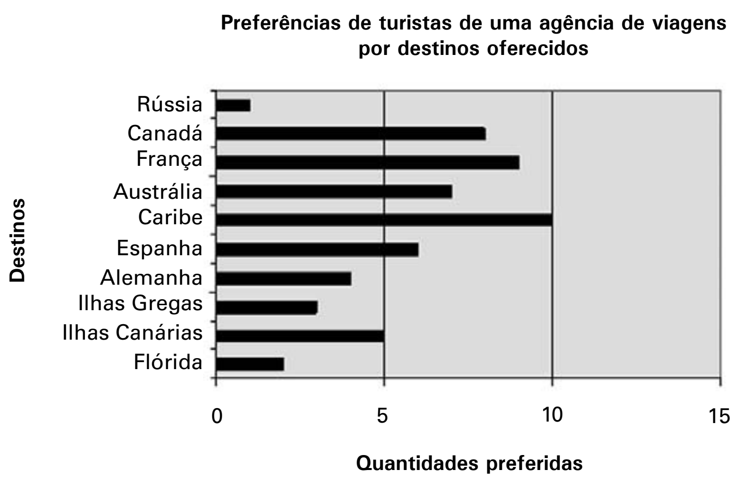


Figura 6.5: Exemplo de um gráfico de barras.

Interpretação do gráfico

Observando a **Figura 6.5**, podemos concluir que o destino preferido pelos turistas é o Caribe, seguido pela França e o Canadá. O destino que apresentou a menor procura foi a Rússia.

c. Setograma ou gráfico de "pizza"

Neste tipo de gráfico, um círculo representa 100% das observações, onde cada resultado da variável em estudo é uma fatia da "pizza".

As características do gráfico de "pizza" são:

- a representação é feita por meio de porcentagem;
- cada fatia da "pizza" tem tamanho proporcional ao valor relativo que o seu resultado representa em relação ao total;
- a variável qualitativa em estudo deve ter poucos resultados, no máximo cinco categorias, para não sobrecarregar a visualização.

Exemplo:

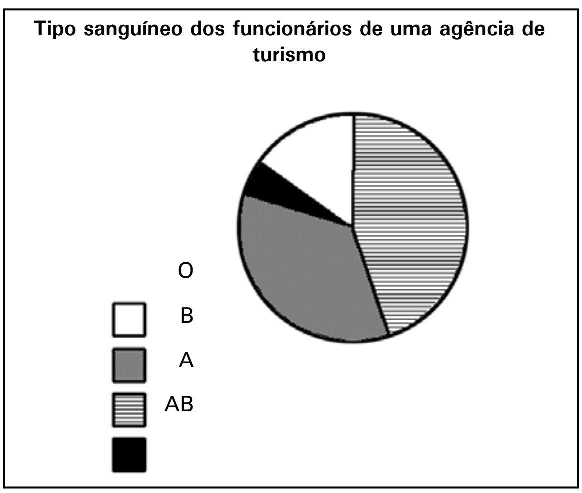


Figura 6.6: Exemplo de um gráfico de setores.

Interpretação do gráfico

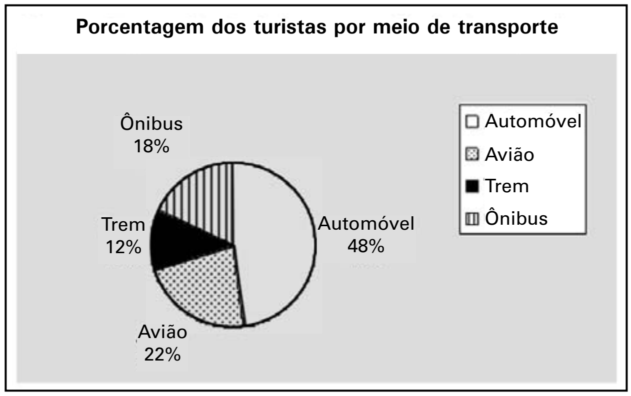
Observando a **Figura 6.6**, podemos concluir que nesta empresa há mais funcionários com o tipo sanguíneo A e poucos funcionários com o tipo sanguíneo AB, enquanto que 50% dos funcionários se dividem entre os tipos sanguíneos B e O.

Atividade

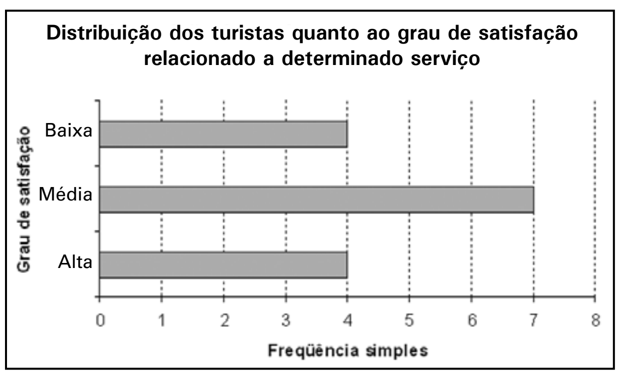
Atende aos Objetivos 1 e 3

1. Identifique e interprete os gráficos apresentados a seguir:

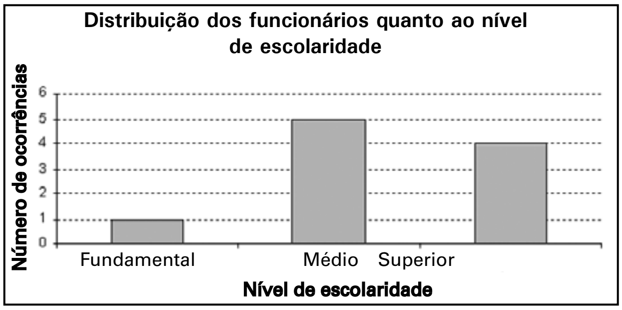
a.



b.



c.



____
____
____
____
____
____

Respostas Comentadas

a. Gráfico de "pizza". Interpretação: a maioria dos turistas utilizou como meio de transporte o automóvel, e uma pequena quantidade de turistas utilizou o trem.

b. Gráfico de barras. Interpretação: a maioria dos turistas classificou como médio o grau de satisfação deste serviço, assim seria interessante que a prestadora deste serviço promovesse melhorias no serviço oferecido.

c. Gráfico de colunas. Interpretação: grande parte dos funcionários desta empresa tem nível de escolaridade médio ou superior, e poucos têm apenas o nível fundamental.

Conclusão

Como conclusão vemos que, com o que foi aprendido nesta aula, temos como compreender a importância que os gráficos têm na apresentação dos resultados de uma pesquisa, de um estudo etc. e como a sua correta interpretação auxilia a ilustração e a interpretação da distribuição de uma variável qualitativa.

Atividades Finais

Atendem aos Objetivos 1, 2 e 3

Na tabela a seguir encontra-se o resultado de uma pesquisa com 10 indivíduos sobre o motivo da viagem, o hábito de viajar, o tempo de permanência (dias) e o gasto por indivíduo por dia (em reais). Resolver, com o auxílio do Excel, as questões 1 e 2.

Tabela 6.1: Resultado de pesquisa sobre o motivo da viagem, o hábito de viajar, a permanência em dias e o gasto por indivíduo por dia em reais (R$) de 10 indivíduos

Indivíduo	Motivo	Hábito	Permanência	Gasto
1	Turismo	Sozinho	16	70,7
2	Turismo	Excursão	11	89,9
3	Visita a parentes	Família	24	108,0
4	Turismo	Família	24	111,1
5	Negócios	Sozinho	22	117,4
6	Visita a parentes	Família	10	120,8
7	Turismo	Excursão	21	152,6
8	Negócios	Sozinho	11	153,6
9	Turismo	Excursão	19	173,9
10	Visita a parentes	Família	17	179,1

Fonte: Dados hipotéticos.

1. Construir um gráfico de barras e um de setores ("pizza") que represente os dados presentes na tabela para a variável motivo da viagem. Interpretar este resultado.

2. Construir um gráfico de colunas e um de setores ("pizza") que represente os dados presentes na tabela para a variável hábito de viajar. Interpretar este resultado.

Respostas Comentadas

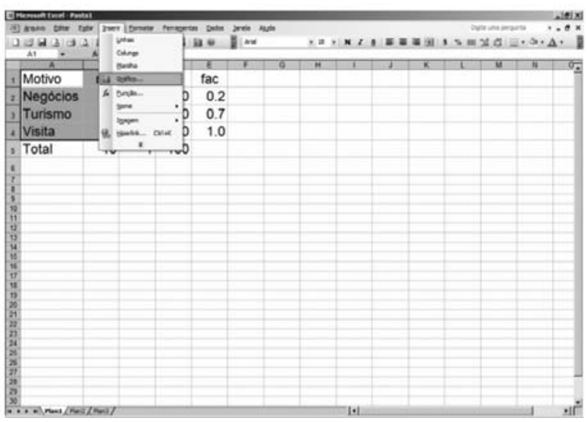
1. Para construir os gráficos pedidos, primeiramente temos de construir a tabela de distribuição de frequências (a construção desta tabela foi assunto da Aula 5) para a variável motivo da viagem, que resultará em:

Motivo	ni	fi	%	fac
Negócios	2	0,2	20,0	0,2
Turismo	5	0,5	50,0	0,7
Visita	3	0,3	30,0	1,0
Total	10	1	100	

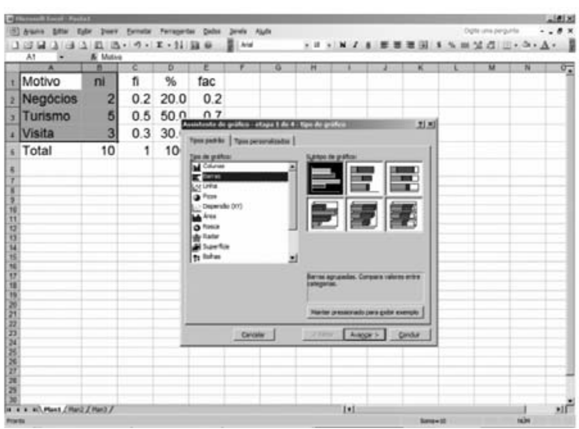
Esta tabela de frequências deve ser digitada no programa Excel e devem ser seguidos os passos descritos na sequência.

Construção do gráfico de barras

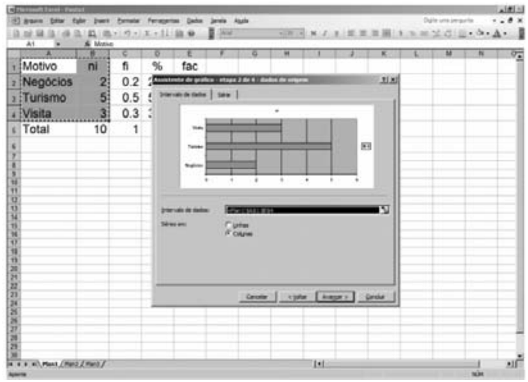
Passo 1*. Selecionar a área que contém os dados (colunas – motivo e ni), clicar no item "Inserir" e selecionar a opção "Gráfico".*



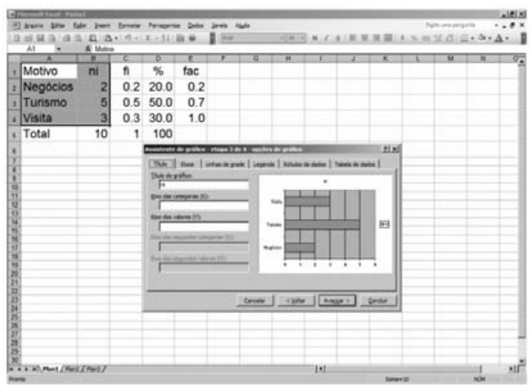
Passo 2*. Dentro do assistente de gráfico, clicar no item "Barras" e "Avançar".*



***Passo 3**. Dentro do assistente de gráfico, clicar nos itens "Avançar" da etapa 2, nomear os eixos e título na etapa 3, e clicar em "Concluir".*

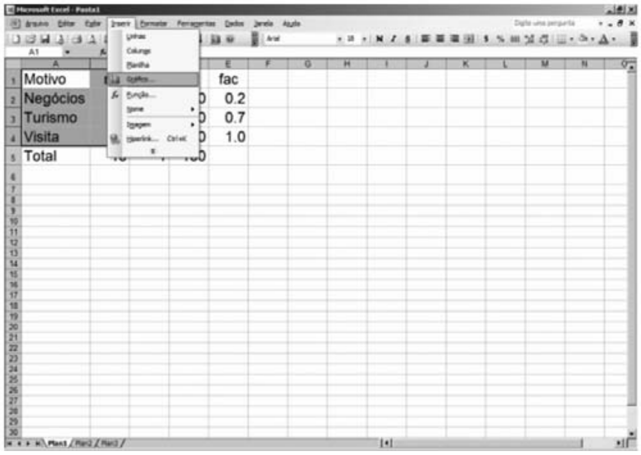


***Passo 4**. Dentro do assistente de gráfico, nomear os eixos na etapa 3 e clicar em "Concluir".*

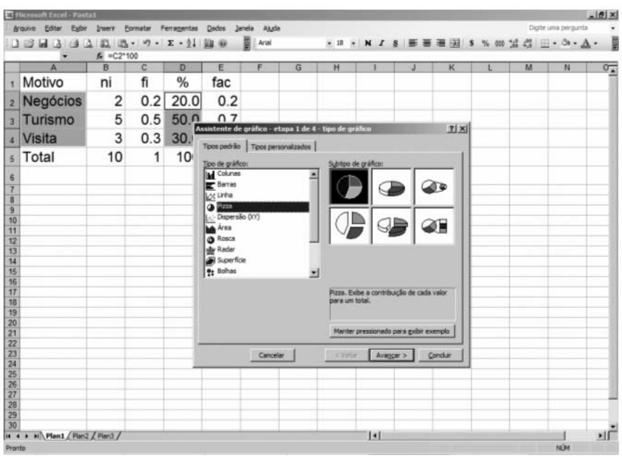


Construção do gráfico de "pizza"

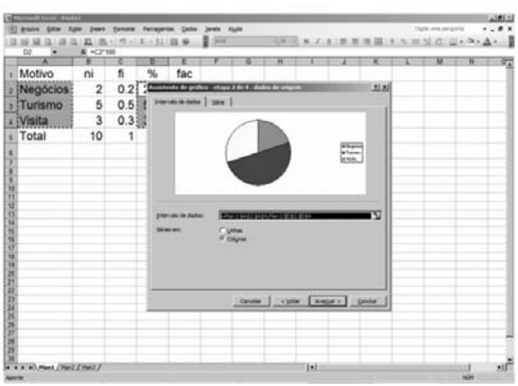
***Passo 1**. Selecionar a área que contém os dados (colunas – motivo e % utilizando a tecla CTRL e o "mouse" para fazer esta seleção), clicar no item "Inserir" e selecionar a opção "Gráfico".*



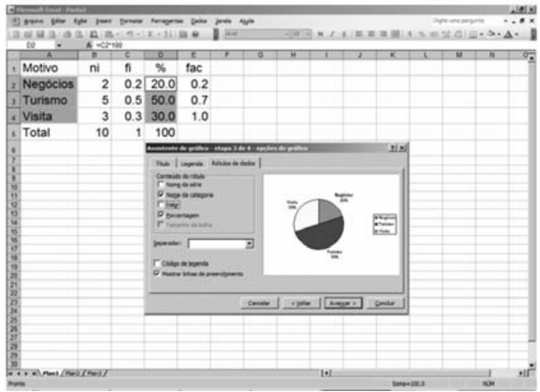
***Passo 2**. Dentro do assistente de gráfico, clicar no item "Pizza" e "Avançar".*



***Passo 3**. Dentro do assistente de gráfico, clicar nos itens "Avançar" da etapa 2, nomear os eixos na etapa 3 e clicar em "Concluir".*



Passo 4*. Dentro do assistente de gráfico, selecionar a pasta "Rótulo de dados" e clicar em "Nome da categoria" e "Porcentagem". Finalizar o gráfico clicando em "Concluir".*



Esse procedimento resultará nos seguintes gráficos:
Gráfico de barras:

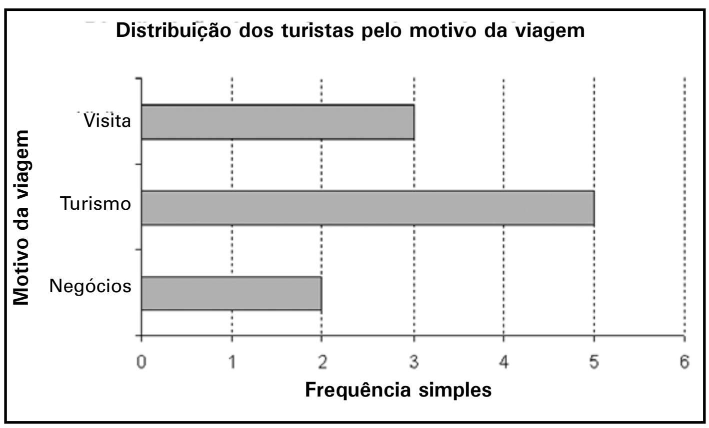
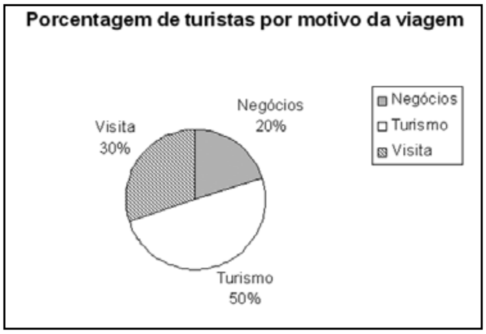


Gráfico de "pizza":

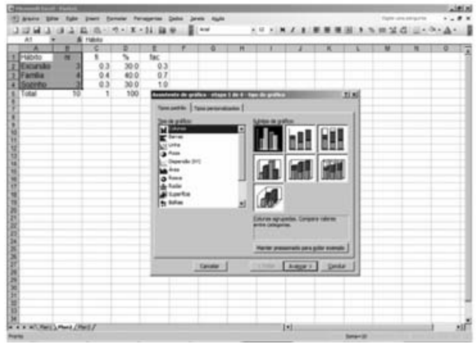


Interpretação: a maioria dos indivíduos desta pesquisa viaja para realizar atividades de turismo, os demais dividem-se em viagens para visitar os parentes e a negócios.

2. Para construir os gráficos pedidos, primeiramente temos de construir a tabela de distribuição de frequências (a construção desta tabela foi assunto da Aula 5) para a variável hábito de viajar.

Hábito	*ni*	*fi*	*%*	*fac*
Excursão	*3*	*0,3*	*30,0*	*0,3*
Família	*4*	*0,4*	*40,0*	*0,7*
Sozinho	*3*	*0,3*	*30,0*	*1,0*
Total	*10*	*1*	*100*	

Para se obterem os gráficos pedidos nesta atividade, deve-se digitar a tabela de distribuição de frequências no Excel e, após, devem ser seguidos os passos explicitados na questão 1 com os dados da questão 2.
A única diferença com relação à questão 1 é no Passo 2. Neste ponto o procedimento é clicar no item "Colunas" e "Avançar".



Com a realização dos passos anteriores, teremos como resultado os seguintes gráficos:
Gráfico de colunas:

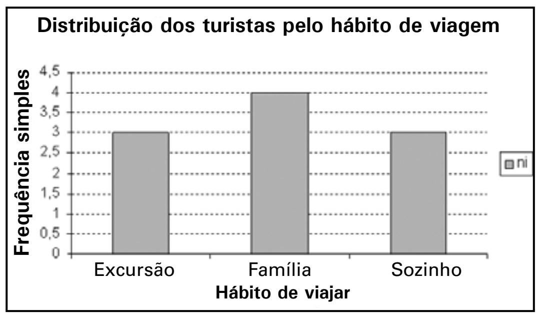
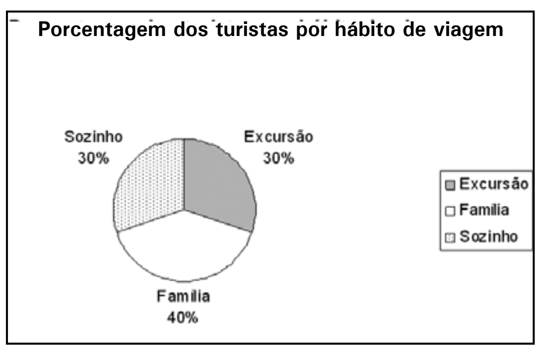


Gráfico de "pizza":



Interpretação: você pode ver por estes dispositivos gráficos que os indivíduos desta pesquisa têm por hábito viajar preferencialmente em família, e as opções de realizar a viagem sozinho ou em excursão foram igualmente citadas.

7 Gráficos estatísticos para representar a distribuição das variáveis quantitativas

Metas da aula

Apresentar e definir os principais gráficos utilizados para representar as variáveis quantitativas.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. classificar os gráficos estatísticos para as variáveis quantitativas;
2. construir gráficos estatísticos utilizando o programa computacional Excel (versão 2003);
3. interpretar gráficos estatísticos apropriados para diferentes situações reais.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é necessário que você tenha incorporado os conceitos de: I) variáveis e seus tipos apresentados na Aula 1; II) séries estatísticas e tabela de distribuição de frequências, ambos presentes na Aula 4; III) gráficos – definição, requisitos, finalidades e construção –, presentes na Aula 5.

Introdução

Conforme você viu na aula anterior, um conjunto de dados é descrito com maior destaque quando se utiliza um gráfico. Desta maneira, verifica-se que os gráficos são amplamente utilizados para a apresentação de dados estatísticos.
No caso das variáveis quantitativas, elas têm como peculiaridade assumir um grande número de valores (pois, como são provenientes de mensuração, dependem da precisão do instrumento de medida), muitas vezes havendo pouca repetição do mesmo valor, em um conjunto de dados. A representação gráfica de uma variável quantitativa difere da apresentada na aula anterior para as variáveis qualitativas, principalmente por apresentar continuidade nos valores numéricos em ao menos um dos eixos do gráfico.

Tipos de gráficos para as variáveis quantitativas

Os tipos de gráficos mais utilizados para representar as variáveis quantitativas são:

a. Gráfico de linhas
b. Histograma
c. Polígono de frequências

a. Gráfico de linhas ou linear

É a representação das informações por meio de uma linha. É usado para representar séries temporais.

As características do gráfico de linhas são:

- é representado por uma linha contínua;
- o tempo é representado no eixo das abscissas (eixo x);
- a série é representada no eixo das ordenadas (eixo y);
- pode-se representar mais de uma série no mesmo gráfico.

Exemplo:

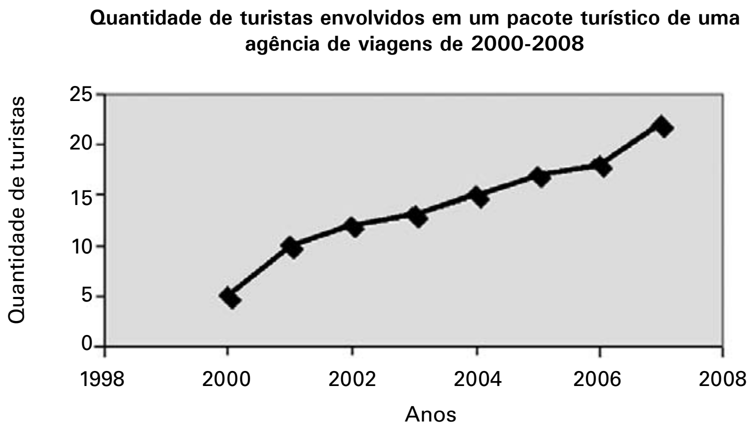


Figura 7.1: Exemplo de gráfico de linhas.

Interpretação do gráfico: Podemos observar, pelo gráfico de linhas (**Figura 7.1**), que houve uma procura crescente por este pacote turístico oferecido por esta agência no decorrer de 10 anos.

Atividade

Atende ao Objetivo 2

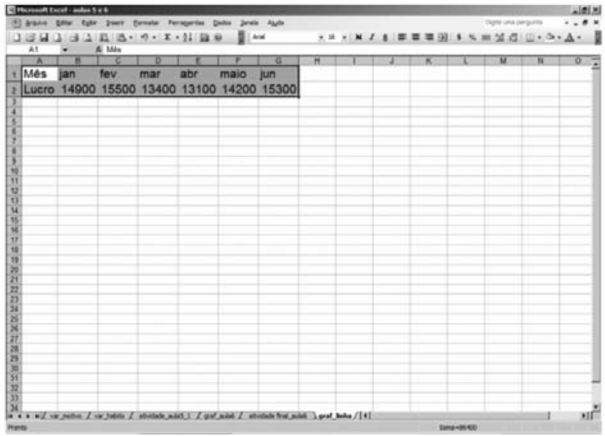
1. Obter e interpretar o gráfico de linhas para os dados dos lucros (em reais) de uma agência de viagens durante o primeiro semestre de determinado ano (**Tabela 7.1**). Utilizar o Excel para a obtenção do gráfico.

Tabela 7.1

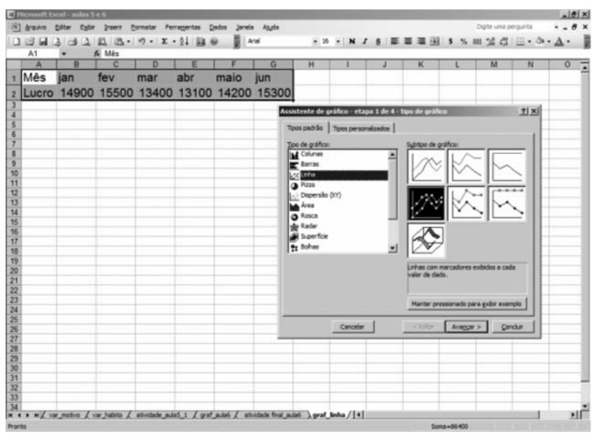
Mês	jan.	fev.	mar.	abr.	maio	jun.
Lucro (R$)	14.900	15.500	13.400	13.100	14.200	15.300

Resposta Comentada

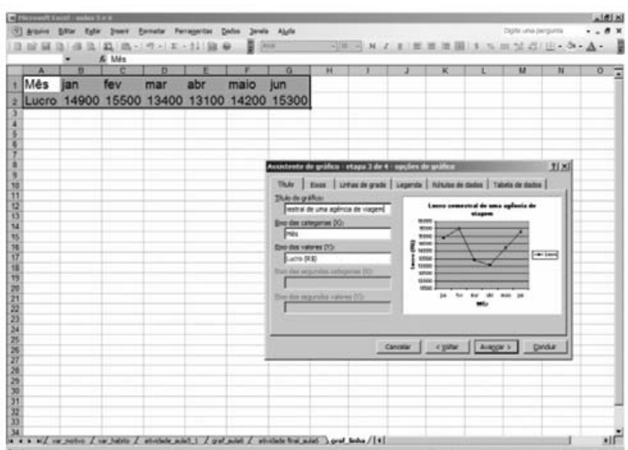
Passo 1*. Digitar a tabela dada no Excel e selecionar toda a área que contém os dados.*

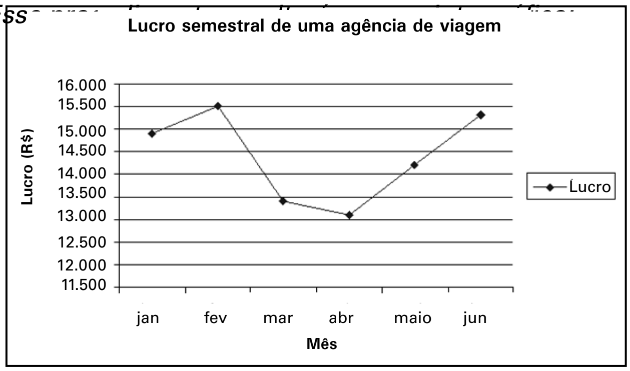


Passo 2*. Clicar no item "Inserir", selecionar a opção "Gráfico", selecionar a opção "Linha".*



Passo 3*. Dentro do assistente de gráfico, clicar nos itens "Avançar" da etapa 2, nomear os eixos e título na etapa 3 e clicar em "Concluir".*





Interpretação do gráfico: Podemos observar que os maiores lucros desta agência de viagens foram obtidos nos meses de janeiro, fevereiro e junho. Isso pode ser explicado pelo fato de nos meses de janeiro e junho haver uma maior concentração de pessoas em férias e em fevereiro pela ocorrência do feriado do carnaval.

b. Histograma

É muito utilizado para representar variáveis quantitativas. É a representação gráfica da distribuição de frequência através de colunas justapostas de maneira contínua, representando cada coluna uma classe. Não há, portanto, espaço entre as colunas, ou seja, onde termina uma classe, imediatamente inicia-se outra.

As características do histograma são:

- é representado por retângulos justapostos;
- a base do retângulo se localiza no eixo das abscissas (eixo x) e é dada pela amplitude das classes;
- a altura do retângulo é representada no eixo ordenadas (eixo y) e é dada pela frequência simples.

Exemplo:

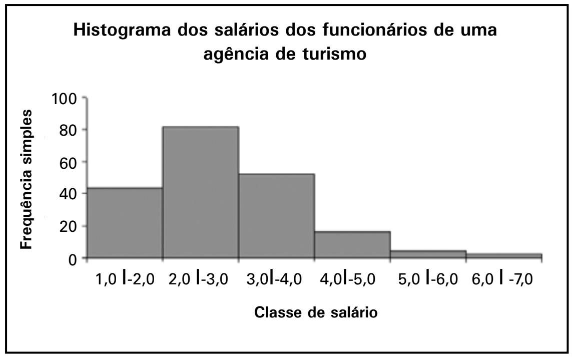


Figura 7.2: Exemplo de histograma.

Interpretação do gráfico: Você pode observar pelo histograma de frequências (**Figura 7.2**) que há uma maior concentração na classe de salários 2 a 3 salários mínimos, e poucos funcionários pertencentes às classes que oferecem maiores salários.

c. Polígono de frequência

É uma linha poligonal que é resultado da interligação de pontos que representam os valores de cada classe, isto é, os pontos médios das classes. Também como o histograma, representa uma distribuição de frequência com intervalo de classe.

As características do polígono de frequência são:

- é um gráfico com configuração linear;
- a linha é formada unindo-se os pontos médios das classes;
- no eixo das abscissas (eixo x) estão representados os pontos médios das classes;
- no eixo das ordenadas (eixo y) estão representadas as frequências simples;
- o histograma pode ou não estar "atrás" do polígono de frequências.

Exemplo:

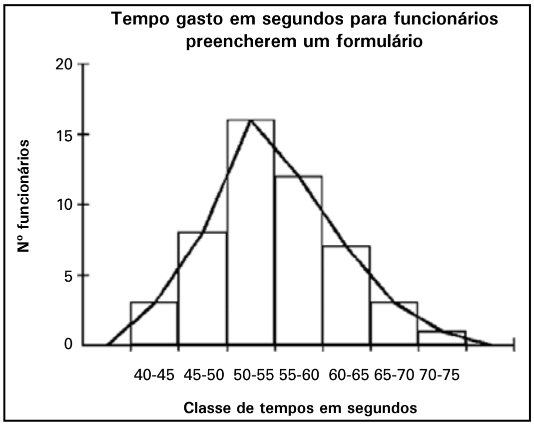


Figura 7.3: Exemplo de polígono de freqüências.

Interpretação do gráfico: Por este gráfico (**Figura 7.3**) você pode observar que a maioria dos funcionários despendeu entre 45 e 65 minutos para preencher o formulário, sendo que poucos gastaram um tempo maior ou menor que este.

Conclusão

Desta aula você pode perceber a importância da representação gráfica das variáveis quantitativas em função de facilitar a visualização da distribuição dos valores, conduzir a conclusões mais eficientes e, assim, auxiliar na interpretação do resultado numérico.

Atividades Finais

Atendem aos Objetivos 1, 2 e 3

Na **Tabela 7.1** encontra-se o resultado de uma pesquisa com 10 indivíduos sobre o motivo da viagem, o hábito de viajar, o tempo de permanência (dias) e o gasto por indivíduo por dia (em reais). Resolver, com o auxílio do Excel, as questões 1 e 2.

Tabela 7.1: Resultado de pesquisa sobre o motivo da viagem, o hábito de viajar, a permanência em dias e o gasto por indivíduo por dia em reais de 10 indivíduos

Indivíduo	Motivo	Hábito	Permanência	Gasto
1	Turismo	Sozinho	16	70,7
2	Turismo	Excursão	11	89,9
3	Visita a parentes	Família	24	108,0
4	Turismo	Família	24	111,1
5	Negócios	Sozinho	22	117,4
6	Visita a parentes	Família	10	120,8
7	Turismo	Excursão	21	152,6
8	Negócios	Sozinho	11	153,6
9	Turismo	Família	19	173,9
10	Visita a parentes	Família	17	179,1

1. Construir o histograma de distribuição de frequências para a variável gasto por turista por dia (em reais) com 3 classes e amplitude entre classes de 75. Interpretar este resultado.

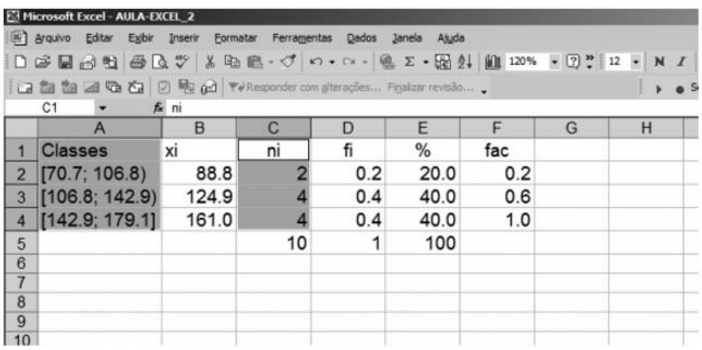
2. Construir o histograma de distribuição de frequências para a variável tempo de permanência.

Respostas Comentadas

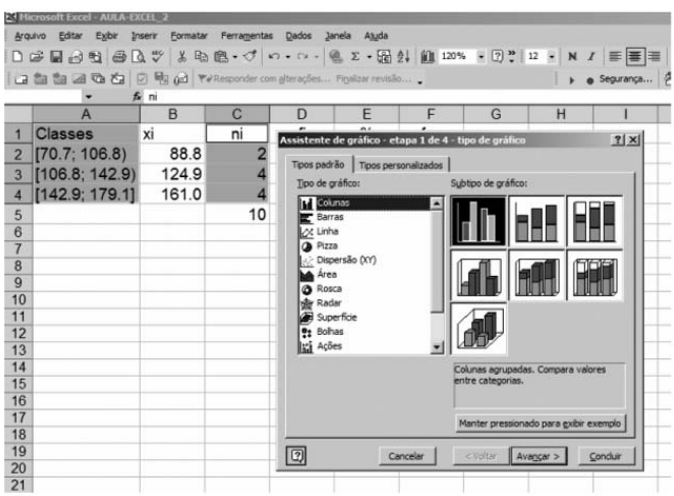
1. Para construir os gráficos pedidos, primeiramente temos de construir a tabela de distribuição de frequências (a construção desta tabela foi assunto da Aula 5) para a variável gasto por turista por dia, que resultará em:

Classes	xi	ni	fi	%	fac
[70,7; 106,8)	88,8	2	0,2	20,0	0,2
[106,8; 142,9)	124,9	4	0,4	40,0	0,6
[142,9; 179,1]	161,0	4	0,4	40,0	1,0
Total		10	1	100	

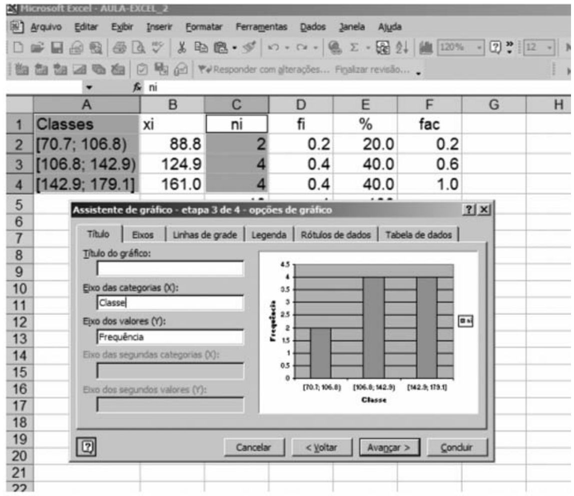
Passo 1*. Digitar a tabela dada no Excel e selecionar a área que contém os dados (colunas – classes e ni).*



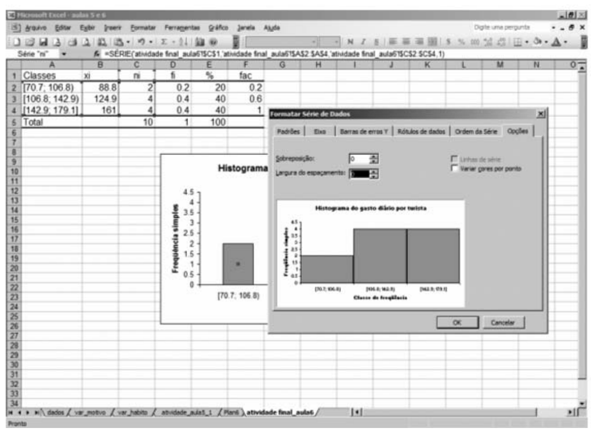
Passo 2*. Clicar no item "Inserir", selecionar a opção "Gráfico", selecionar a opção "Colunas".*



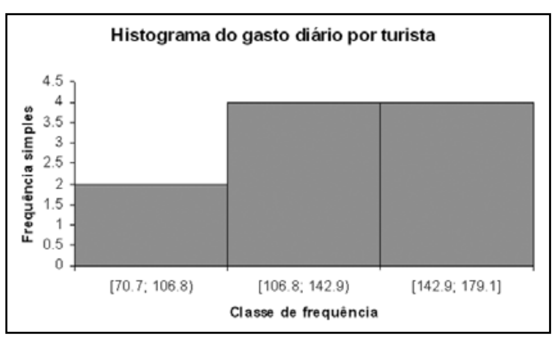
Passo 3*. Dentro do assistente de gráfico, clicar nos itens "Avançar" da etapa 2, nomear os eixos e título na etapa 3, e clicar em "Concluir".*



Passo 4*. Clicar em cima das colunas do gráfico na opção formatar série de dados, entrar em "Opções", em "largura do espaçamento" apagar "150" e escrever "0", e clicar em "Ok".*



Este procedimento resultará no seguinte gráfico:

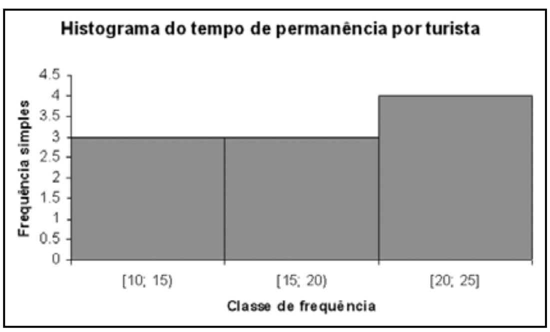


Interpretação do gráfico: As classes de frequência que apresentam a maior concentração dos gastos destes turistas são a segunda e a terceira, sendo que a primeira classe, onde estão concentrados os turistas que gastam menos, apresenta pouca concentração.

2. Para construir o gráfico pedido, primeiramente temos de construir a tabela de distribuição de frequências para a variável tempo de permanência, que resultará em:

Classes	xi	ni	fi	%	fac
[10; 15)	12,5	3	0,3	30,0	0,3
[15; 20)	17,5	3	0,3	30,0	0,6
[20; 25]	22,5	4	0,4	40,0	1,0
Total		10	1	100	

Para se obter o histograma pedido nesta atividade, deve-se digitar a tabela de distribuição de frequências anterior no Excel e, após isso, devem ser seguidos os passos explicitados na questão 1 com os dados da questão 2.
Este procedimento resultará no seguinte gráfico:



8 Separatrizes

Meta da aula

Apresentar o conceito de medidas de posição, com enfoque no seu cálculo e na interpretação do valor resultante.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. reconhecer o conceito de medidas de posição e sua importância dentro da estatística;
2. calcular e identificar, no contexto de um conjunto de dados real, as medidas de posição;
3. interpretar o valor resultante do cálculo da medida de posição dentro do conjunto de dados em estudo.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é imprescindível que os conceitos de população e amostra, distribuição de freqüências e tabelas e gráficos, vistos, respectivamente, nas Aulas 1, 5, 6 e 7, estejam bem claros.

Introdução

Nas aulas anteriores, você aprendeu a organizar e a resumir o conjunto de dados na forma de tabelas de distribuição de frequência e gráficos. Nesta aula, você irá aprender como situar um valor numérico dentro de determinado conjunto de dados.
Essa estatística, a qual chamamos de separatriz, indicará a você quais são os valores que estarão acima ou abaixo de determinado valor do conjunto de dados. Para tanto, primeiramente, você vai aprender os conceitos e formas de calcular: os quartis, os decis e os percentis.

Separatrizes – medidas de posição

São valores que ocupam determinados lugares em um ***ROL***. Assim, cabe salientar que, sempre que formos calcular as separatrizes, os dados do conjunto têm de estar ordenados.

ROL

Sequência ordenada e crescente de dados.

As separatrizes, como o próprio nome já adianta, têm como característica separar o conjunto de dados em partes iguais, e, deste modo, nos indicar quais valores estão acima e quais estão abaixo do valor calculado.

As separatrizes que iremos estudar são:

1. os quartis;
2. os decis;
3. os percentis.

1. Quartis (Qi)

São medidas que dividem um conjunto de dados, uma distribuição, em quatro partes iguais.

Há, portanto, três quartis:

a. $\mathbf{Q_1}$ (1° quartil) – é o valor situado de tal modo na série que 25% dos dados são menores ou iguais a ele, e 75% são maiores que ele.

b. $\mathbf{Q_2}$ (2° quartil) – é o valor situado de tal modo na série que 50% dos dados são menores ou iguais a ele, e 50% são maiores.

c. $\mathbf{Q_3}$ (3° quartil) – é o valor situado de tal forma na série que 75% dos dados são menores ou iguais a ele, e 25% são maiores.

Exemplo:

Faixa salarial, dentro da administração de uma empresa, por exemplo, é a que vai de Q_1 a Q_3, e que concentra a maioria dos valores da distribuição.

2. Decis (D_i)

São medidas que dividem um conjunto de dados, uma distribuição, em dez partes iguais.

Há, portanto, nove decis:

a. $\mathbf{D_1}$ (1° decil) – é o valor situado de tal modo na série que 10% dos dados são menores ou iguais a ele, e 90% são maiores.

b. $\mathbf{D_2}$ (2° decil) – é o valor situado de tal modo na série que 20% dos dados são menores ou iguais a ele, e 80% são maiores.

E, assim, sucessivamente até:

c. $\mathbf{D_9}$ (9° decil) – é o valor situado de tal forma na série que 90% dos dados são menores ou iguais a ele, e 10% são maiores.

3. Percentis (P_i)

São medidas que dividem um conjunto de dados, uma distribuição, em 100 partes iguais.

Há, portanto, 99 percentis:

a. $\mathbf{P_1}$ (1° percentil) – é o valor situado de tal modo na série que 1% dos dados é menor ou igual a ele, e 99% são maiores.

b. $\mathbf{P_2}$ (2° percentil) – é o valor situado de tal modo na série que 2% dos dados são menores ou iguais a ele, e 98% são maiores.

E, assim, sucessivamente até:

c. $\mathbf{P_{99}}$ (99° percentil) – é o valor situado de tal forma na série que 99% dos dados são menores ou iguais a ele e 1% é maior.

Seguem algumas separatrizes e seus nomes particulares.

Aquela que deixa:

- 25% dos dados abaixo dela: 1° quartil = 25° percentil.
- 50% dos dados abaixo dela: 2° quartil = 5° decil = 50° percentil.
- 75% dos dados abaixo dela: 3° quartil = 75° percentil.
- 95% dos dados abaixo dela: 95° percentil.

!

Na coleta de dados, a faixa que concentra a maior parte dos valores em um conjunto de dados é:
P_{25} a $P_{75} = Q_2$ a Q_3.

Cálculo das separatrizes

A seguir, você irá aprender a calcular as separatrizes quando os dados estiverem agrupados em classes e quando não estiverem.

a. Dados agrupados em classe:

Quando os dados estão agrupados em classe, para determinar os quartis, usamos a mesma fórmula da mediana, adaptada, na qual o elemento mediano, EM_e, é substituído por EQ_i, elemento quartílico, dado pela fórmula:

$EQ_i = \dfrac{i \cdot n}{4}$, onde i é o número de ordem dos quartis 1, 2 ou 3.

Atividade

Atende aos Objetivos 1 e 2

1. Determine a faixa salarial dos empregados da agência de turismo X, considerando os dados a seguir:

Salário (R$)	f_i	Fac_i
500 |—— 700	15	15
700 |—— 900	43	58
900 |——1.100	18	76
1.100 |——1.300	5	81
1.300 |——1.500	1	82
1.500 |——1.700	1	83
1.700 |——1.900	1	84
Σ	84	-

Resposta Comentada

A faixa salarial dos empregados da empresa é o intervalo que vai de Q_1 a Q_3, então:

$$EQ_1 = \frac{1 \times 84}{4} = 21$$

$$Q_1 = l_i + h\left[\frac{EQ_1 - 'Fac}{FQ_1}\right]$$

$$Q_1 = 700 + 200 \left[\frac{21 - 15}{43}\right] = R\$ \ 728$$

$$EQ_3 = \frac{3 \times 84}{4} = 63$$

$$Q_3 = l_i + h \left[\frac{EQ_3 - \text{'Fac}}{FQ_3} \right]$$

$$Q_3 = 900 + 200 \left[\frac{63 - 58}{18} \right] = \text{R\$ } 956$$

Faixa salarial: de R$ 728 a R$ 956

b. Dados não agrupados em classe:

Quando os dados não estão agrupados em classe, para determinar as separatrizes, usamos uma regra de três que irá nos indicar a posição do elemento no conjunto de dados da seguinte forma:

100% dos dados ------------------- n

Porcentagem da separatriz ------------------- i

em que:

n = número de elementos do conjunto de dados

i = posição do elemento no ROL

Exemplo: Para o seguinte conjunto de dados encontre os primeiro e terceiro quartis.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x	3	5	10	13	14	19	20	28	28	29	37	45	54	67

Cálculo do 1° quartil:

100% ------------- 14

25% ------------- i

i = (25 . 14)/100 ⇒ i = 3,5. Como não existe a posição 3,5, então aproxima-se para o valor inteiro superior, resultando em i = 4. Sendo assim, o valor do Q_1 nesse conjunto de dados será $Q_1 = 13$.

Cálculo do 3º quartil:

100% ------------- 14

75% ------------- i

$i = (75 \cdot 14)/100 \Rightarrow i = 10{,}5$. Como não existe a posição 10,5, então aproxima-se para o valor inteiro superior, resultando em i = 11. Sendo assim, o valor do Q_1 nesse conjunto de dados será $Q_3 = 37$.

A posição da separatriz é um número inteiro, ou seja, sem casas decimais, dentro do conjunto de dados. Assim, sempre que o valor calculado para a separatriz resultar em casas decimais, esse valor será aproximado para o inteiro superior.

Atividade

Atende aos Objetivos 1 e 2

2. Encontre o 2º decil dos dias de permanência de 11 turistas estrangeiros no Brasil e interprete esse resultado.

i	1	2	3	4	5	6	7	8	9	10	11
x	5	6	7	8	9	9	10	10	11	13	14

Resposta Comentada

Como os dados já se encontram ordenados do menor para o maior, ou seja, em um ROL, podemos iniciar o cálculo da separatriz pedida, 2º decil.

Verificamos que a amostra contém 11 dados, assim,

100% ------------- 11

20% ------------- i

$i = (20 \;.\; 11)/100 \Rightarrow i = 2{,}2$. Como não existe a posição 2,2, então aproxima-se para o valor inteiro superior, resultando em $i = 3$. Sendo assim, o valor do D_2 nesse conjunto de dados será $D_2 = 7$.

Interpretação: Por esse resultado do 2º decil, podemos concluir que ao menos 20% dos turistas estrangeiros permanecem sete dias ou menos, quando em visita ao Brasil.

Com o que foi exposto nesta aula, você pode concluir que, de posse de um conjunto de dados ordenados, há como identificar a posição de determinado elemento e interpretar a respeito do significado desta posição num conjunto de dados reais, ou seja, as separatrizes fornecerão a localização dos dados na amostra.

Atividades Finais

Atendem aos Objetivos 1, 2 e 3

1. O peso médio (em kg) da bagagem de mão de 9 passageiros de uma companhia aérea foi medido e resultou em: 2,59; 2,64; 2,60; 2,62; 2,57; 2,61; 2,50; 2,63; 2,64. Calcule o 2º quartil desses pesos e interprete o resultado.

2. A tabela a seguir mostra as notas em doze disciplinas (em média por disciplina) de dois alunos distintos (X eY), concorrentes a uma bolsa de estudos, do 3º período de determinado curso.

Média por disciplina por aluno

Aluno	**Disciplina**											
	A	B	C	D	E	F	G	H	I	J	K	L
X	6,8	6,4	8,1	5,4	5,3	7,0	6,7	5,0	6,6	7,3	7,4	5,3
Y	8,7	7,9	10,0	7,8	6,5	9,1	8,2	5,7	8,0	9,6	9,5	7,2

Fonte: dados fictícios

a. Calcule, para cada um dos alunos, o 3º quartil. Interprete esse resultado.

b. Baseado nessa informação, qual dos dois alunos tem maiores chances de receber a bolsa de estudos?

Respostas Comentadas

1. Como os dados não se encontram ordenados do menor para o maior, você tem que, primeiramente, ordenar esses dados, o que resultará em:

I	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
X	*2,50*	*2,57*	*2,59*	*2,60*	*2,61*	*2,62*	*2,63*	*2,64*	*2,64*

Agora, com os dados em um ROL, podemos iniciar o cálculo da separatriz pedida, 2º quartil.
Verificamos que a amostra contém nove dados, assim,
100% ------------- 9
50% ------------- i

i = (50 . 9)/100 ⇒ i = 4,5. Como não existe a posição 4,5, então aproxima-se para o valor inteiro superior, resultando em i = 5. Sendo assim, o valor do Q_2 *nesse conjunto de dados será* $Q_2 = 2{,}61$.

Interpretação: Por esse resultado do 2º quartil, podemos concluir que ao menos 50% dos passageiros carregam bagagens de mão com pesos menores ou iguais a 2,61kg.

2.
a. Como os dados não se encontram ordenados do menor para o maior, você tem que, primeiramente, ordenar estes dados, o que resultará em:

X	*5,0*	*5,3*	*5,3*	*5,4*	*6,4*	*6,6*	*6,7*	*6,8*	*7,0*	*7,3*	*7,4*	*8,1*
Y	*5,7*	*6,5*	*7,2*	*7,8*	*7,9*	*8,0*	*8,2*	*8,7*	*9,1*	*9,5*	*9,6*	*10,0*

Agora, com os dados em um ROL, podemos iniciar o cálculo da separatriz pedida, 3º quartil.
Verificamos que a amostra contém doze dados, assim,
100% ------------- 12
75% ------------- i

i = (75 . 12)/100 ⇒ i = 9. Como se encontrou um número inteiro para essa posição, conclui-se que o valor do Q_3 *para as notas do aluno X foi* $Q_3 = 7{,}0$ *e para as notas do aluno Y, foi* $Q_3 = 9{,}1$.

Interpretação: Pelo resultado do 3º quartil, podemos concluir que ao menos 75% das notas do aluno X foram iguais ou menores que 7,0, e ao menos 75% das notas do aluno Y foram iguais ou menores que 9,1.

b. Assim, pelos resultados obtidos no item a, concluímos que o aluno Y tem maiores chances de obter a bolsa de estudos, pois suas notas foram maiores.

Resumo

As separatrizes são medidas estatísticas que servem para posicionar os elementos da série de dados e permitem a classificação e a qualificação desses elementos.
Os quartis dividem a série em quatro partes iguais. São ao todo três quartis.
Já os percentis dividem a série em 100 partes iguais e temos ao todo 99 percentis.

Informação sobre a próxima aula

Na próxima aula, apresentaremos as principais medidas de tendência central que objetivam descrever o nível geral da série e a sua representação.

9 Medidas de tendência central

Meta da aula

Apresentar o conceito de medidas de tendência central, com enfoque no seu cálculo e na interpretação do valor resultante.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. reconhecer o conceito de medidas de tendência central e sua importância dentro da estatística;
2. calcular e identificar no contexto de um conjunto de dados real as medidas de tendência central;
3. interpretar o valor resultante do cálculo da medida de tendência central dentro do conjunto de dados em estudo;
4. identificar qual medida de tendência central melhor irá representar o seu conjunto de dados.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é imprescindível que os conceitos de população e amostra; distribuição de frequências; gráficos e separatrizes, vistos nas Aulas 1, 5, 6, 7 e 8, respectivamente, estejam bem claros.

Introdução

Nas aulas anteriores, você aprendeu a organizar e a resumir o conjunto de dados na forma de tabelas de distribuição de frequência e gráficos. Nesta aula, você aprenderá como resumir ainda mais o conjunto de dados, utilizando apenas um valor numérico.
Esse valor numérico irá indicar a tendência dos dados se agruparem ao redor de um valor central de determinado valor do conjunto de dados. Para tanto, você irá aprender os conceitos e formas de calcular: a média aritmética, a mediana e a moda.

Medidas de tendência central

As medidas de tendência central descrevem o nível geral dos dados coletados, isto é, elas informam a tendência dos dados, dos valores da série. Têm essa denominação em virtude dos dados observados tenderem a se agrupar ao redor dos valores centrais de sua distribuição.

Por descrever o padrão geral dos dados, elas podem ser usadas para resumir e representar os dados a partir das quais foram calculadas.

1ª. Média aritmética ($\bar{x}$)

A média aritmética é definida como a razão entre o somatório dos valores observados e o número deles.

Portanto, se tivermos o seguinte conjunto de dados:

x_1; x_2, x_3; .... x_n, sua média será dada por:

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

Onde *n* é o número total de dados observados (tamanho da amostra).

Média – Somar todos os X_i e dividir pela quantidade deles.

Exemplos:

1. Calcular o tempo médio (em dias) que cinco turistas permaneceram em uma estância, sendo que cada um deles permaneceu: 6, 5, 5, 6 e 8 dias.

Calcula-se a média por:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{6+5+5+6+8}{5} = 6$$

A média dos dias de permanência de turistas em uma estância é de seis dias.

2. Distribuição por intervalo de classe. Dada a seguinte distribuição das notas de satisfação de 80 turistas com uma viagem em grupo, podemos calcular a média das notas de duas maneiras.

Tabela 9.1: Cálculo da média

NOTAS	f_i	x_i	$x_i f_i$	d_i	$d_i f_i$
20|— 29	2	24,5	49,0	-4	-8
30|— 39	9	34,5	310,5	-3	-27
40 |— 49	11	44,5	489,5	-2	-22
50 |— 59	15	54,5	817,5	-1	-15
60 |— 69	17	64,5	1.096,5	0	0
70 |— 79	16	74,5	1.192,0	1	16
80 |— 89	7	84,5	591,5	2	14
90 |— 99	3	94,5	283,5	3	9
$\sum$	80	—	4.830	—	- 33

2.1. Primeira maneira: pelo processo longo.

Utiliza-se a seguinte expressão:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \Rightarrow \bar{x} = \frac{4.830}{80} \cong 60,4$$

2.2. Segunda maneira: pelo processo breve.

Utiliza-se a seguinte expressão:

$$\bar{x} = \bar{x}_0 + h\frac{\sum_{i=1}^{n} d_i f_i}{n} \Rightarrow \bar{x} = 64,5 + 10\frac{(-33)}{80} \cong 60,4$$

Onde,

$\bar{x}_0$ = ponto médio da classe em que o d_i é igual a zero.

Observa-se que pelas duas maneiras o resultado é o mesmo.

Atividade

Atende aos Objetivos 1, 2 e 3

1. Os valores relativos ao tempo (em horas) de espera para o embarque em determinado voo de 31 turistas, em viagens ao Nordeste brasileiro, foram dispostos na seguinte distribuição de frequência:

Tempo de espera (horas)	f_i	$x_i f_i$
7	3	21
6	4	24
5	6	30
4	7	28
3	5	15
2	4	8
1	2	2
$\sum$	31	128

Calcule o tempo médio que esses turistas aguardaram para embarcar em seu voo. Interprete o resultado.

____

____

____

____

____

____

Resposta Comentada

Para você encontrar o tempo médio de espera, deve utilizar a seguinte expressão:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{128}{31} \cong 4{,}1$$

Assim, conclui-se que o tempo médio de espera para embarque foi de 4,1 horas.

2ª. Mediana (M_e):

É o valor do rol (sequência ordenada e crescente de dados) que ocupa o centro da distribuição, ou seja, é o valor que divide a distribuição ao meio.

Exemplos:

1. Encontre a mediana dos dias de permanência de sete turistas estrangeiros no Brasil. Os dados são: 3, 7, 4, 12, 14, 10, 15.

rol: 3, 4, 7, $\boxed{10,}$ 12, 14, 15.

↓

É a mediana M_e. É o nível geral destes números.

$\boxed{M_e = 10}$

2. Encontre a mediana dos dias de permanência de oito turistas estrangeiros no Brasil. Os dados são: 3, 4, 7, 12, 15, 10, 18, 14.

Rol: 3, 4, 7, $\boxed{10,\ 12,}$ 14, 15, 18.

O 10 e o 12 são valores que ocupam o centro da distribuição. A mediana será a média entre esses dois valores.

$$M_e = \frac{(10+12)}{2} = \frac{22}{2} = 11.$$

$\boxed{M_e = 11}$

3. 31 crianças de um país indicaram numa pesquisa o número de irmãos e /ou irmãs que viviam com cada um deles em casa. Os dados resultantes foram dispostos na tabela a seguir. Calcule a mediana:

Tabela 9.2: Cálculo da mediana

Nº de irmãos e /ou irmãs	f_i	Fac_i
5	6	6
4	7	13
3	9	22 16 < 22, Classe da M_e
2	5	27
1	4	31
Σ	31	–

Processo de cálculo:

1°. Calculam-se as Facs como visto na Aula 5.

2°. Calcula-se o elemento mediano (EM_e), da seguinte forma:

$EM_e = \frac{n}{2}$, se n for par.

$EM_e = \frac{n+1}{2}$, se n for ímpar.

3°. Identifica-se M_e de tal forma:

$EM_e < FAC_i$

$$EM_e = \frac{n+1}{2} = \frac{31+1}{2} = \frac{32}{2} = 16$$

$M_e = 3$

4. Encontre a classe mediana da distribuição da estatura de 40 hóspedes de um hotel, dada na tabela a seguir.

Tabela 9.3: Cálculo da mediana e distribuição por classe

Classe das estaturas (cm)	f_i	Fac_i
150 |—— 154	4	4
154 |—— 158	9	13
158 |—— 162	11	24 20 < 24 , Classe da M_e
162 |—— 166	8	32
166 |—— 170	5	37
170 |—— 174	3	40
Σ	40	–

Processo de cálculo:

1°. Calculam-se as Facs

2°. Calcula-se o EM_e

3°. Identifica-se a classe da mediana tal que:

$$\boxed{EM_e \leq Fac_i}$$

4°. Aplica-se a fórmula:

$$M_e = l_i + h = \left[\frac{EM_e - 'Fac}{fmed}\right]$$

Onde:

M_e = mediana.

l_i = limite inferior da classe da mediana.

h = intervalo de classe.

EM_e = elemento mediano.

'Fac = frequência acumulada anterior à classe da mediana.

fmed = frequência absoluta da classe da mediana.

$$EM_e = \frac{40}{2} = 20$$

$$M_e = l_i + h\left[\frac{EM_e - 'Fac}{fmed}\right]$$

$$M_e = 158 + 4\left[\frac{20 - 13}{11}\right] = 160,5$$

Atividade

Atende aos Objetivos 1, 2 e 3

2. Calcule a mediana da distribuição dos dias de permanência de turistas brasileiros em viagens para a Europa. Os dados estão dispostos na tabela a seguir.

x_i	f_i	Fac_i
12	1	1
14	2	3
15	1	4 4 = 4, Classe da M_e
16	2	6
17	1	7
20	1	8
$\sum$	8	–

____
____
____
____
____
____
____
____
____
____

Resposta Comentada

Para proceder ao cálculo da mediana, você precisa primeiramente calcular as Facs para, em seguida, encontrar o elemento mediano EM_e:

$$EM_e = \frac{n}{2} = \frac{2}{2} = 4$$

Assim, você identificará a mediana M_e e encontrará seu valor por

$$M_e = \frac{15+16}{2} = 15,5$$

Interpretação: 50% dos turistas estrangeiros em viagem à Europa permanecem 15,5 dias.

3ª. Moda (M_o):

A moda é definida como o valor que possui a maior frequência em um conjunto de dados ou distribuição. É utilizada para identificar os valores que mais se repetem em uma amostra.

Exemplos:

Para os conjuntos de dados a seguir, que representam o tempo de permanência (em dias) de executivos em viagens de negócios, encontre a moda em cada situação.

1. 7, 8, 9, 10, 10, 10, 11, 12, 13 e 15

$\boxed{M_o}$ = 10 – pois é o valor que mais se repete (tem maior frequência).

2. 3, 5, 8, 10, 12, 13

Não há moda – distribuição amodal, pois não há um valor que mais se repita.

3. 2, 3, 4, 4, 4, 5, 6, 7, 7, 7, 8, 9

$\boxed{M_o}$ = 4 e $\boxed{M_o}$ = 7, a distribuição é bimodal, pois esses são os dois valores que ocorrem com maior frequência.

4. 6, 6, 6, 6, 6, 6

A distribuição é amodal, pois apresenta um único valor que se repete.

5. 3, 3, 3, 8, 6, 6, 6, 10, 15, 15, 15

$\boxed{M_o}$ = 3 $\boxed{M_o}$ = 6 $\boxed{M_o}$ = 15, a distribuição é trimodal, pois esses são os três valores que ocorrem com maior frequência.

6. Dada a distribuição de frequência a seguir, calcule a moda:

Tabela 9.4: Cálculo da moda de distribuição simples

x_i	f_i
3	2
5	7
9	13
12	8
Σ	30

$\boxed{M_o = 9}$, pois é a categoria de valor que apresenta a maior frequência.

Atividade

Atende aos Objetivos 1, 2 e 3

3. Calcule a moda da distribuição do número de faltas de 11 funcionários de uma agência de turismo. Os dados estão dispostos na tabela a seguir.

x_i	f_i
1	3
2	1
3	3
4	1
5	3
∑∑	11

Resposta Comentada

Para você encontrar a moda do número de faltas, basta identificar qual são as categorias de valores que ocorrem com maior frequência. Neste problema, as categorias de valores mais frequentes são: 1 falta, 3 faltas e 5 faltas. Assim, resultará que a distribuição das faltas será trimodal, pois há 3 valores de faltas que ocorreram com maior frequência, que são:

$\boxed{M_o = 1}$ $\boxed{M_o = 3}$ $\boxed{M_o = 5}$

Moda bruta, de King e de Czuber

Para a seguinte distribuição de salários de turistas de uma agência de turismo, você vai aprender a calcular:

- Moda bruta.
- Moda de King.
- Moda de Czuber.

Assim, dada a distribuição, vamos encontrar as modas citadas anteriormente.

Tabela 9.5: Distribuição dos salários

SALÁRIOS (R$)	f_i
500 |—— 700	18
700 |—— 900	31 Classe M_o
900 |—— 1.100	15
1.100 |—— 1.300	3
1.300 |—— 1.500	1
1.500 |—— 1.700	1
1.700 |—— 1.900	1
Σ	70

1ª. Moda bruta

É o ponto médio da classe modal, dado por:

$$M_o = \left(\frac{l_i + l_s}{2}\right)$$

Então, para os dados da tabela, temos:

$$M_o = \left(\frac{700 + 900}{2}\right) = R\$\ 800,00$$

2ª. Moda de King

Para calcular a moda de King, utiliza-se a seguinte expressão:

$$M_o = l_i + h\left[\frac{f_{post}}{f_{ant} + f_{post}}\right]$$

Para os dados da tabela, temos:

$$M_o = 700 + 200\left[\frac{15}{18 + 15}\right] = R\$\ 790,91$$

3ª. Moda de Czuber

Para calcular a moda de Czuber, utiliza-se a seguinte expressão:

$$M_o = li + h\left[\frac{f.máx. - f.ant}{2.f.máx. - (f.ant. + f.post)}\right]$$

Onde:

M_o = moda.
li = limite inferior da classe modal.
f. máx. = frequência máxima.
f. post. = frequência posterior à classe modal.
f. ant. = frequência anterior à classe modal.

Para os dados da tabela, temos:

$$M_o = 700 + 200\left[\frac{31 - 18}{(2 \text{ x } 31) - 18 + 15)}\right]$$

$$M_o = R\$\ 789,7$$

Uso das medidas de tendência central

A medida ideal em cada situação real é aquela que melhor representa a maioria dos dados da distribuição de frequência.

Passos para a escolha da melhor medida

1. Verifique se a distribuição é de variável qualitativa. Se sim, a melhor medida é a moda.
2. Se não for de variável qualitativa, verifique se ela é de alta ou média dispersão. Se for de alta dispersão, a melhor medida é a mediana ou a moda. Se for de baixa dispersão, a melhor medida é a média.

Com o que foi visto nesta aula, podemos concluir que, de posse de um conjunto de dados, podemos calcular estatísticas que irão resumir esses dados em uma única informação numérica. Tais medidas irão também nos auxiliar a verificar qual é o valor central de um conjunto de dados, ao redor do qual os demais dados da amostra se distribuirão.

Atividades Finais

Atendem aos Objetivos 1, 2, 3 e 4

1. O número de quartos de dez hotéis distintos de uma mesma rede hoteleira foi informado como: 127, 136, 148, 156, 163, 175, 189, 192, 200, 213. Calcule a média, a moda e a mediana desses dados. Qual dessas estatísticas representa melhor a tendência central desses dados?

____
____
____
____
____
____

2. A tabela a seguir mostra as notas em doze disciplinas (em média por disciplina) de dois alunos distintos (X eY), concorrentes a uma bolsa de estudos do 3º período de determinado curso.

Média por disciplina por aluno

Aluno	**Disciplina**											
	A	B	C	D	E	F	G	H	I	J	K	L
X	6,8	6,4	8,1	5,4	5,3	7,0	6,7	5,0	6,6	7,3	7,4	5,3
Y	8,7	7,9	10,0	7,8	6,5	9,1	8,2	5,7	8,0	9,6	9,5	7,2

a. Calcule, para cada um dos alunos, a moda, a mediana e a média geral das notas.

b. Baseado nas estatísticas calculadas anteriormente, qual dos dois alunos tem maiores chances de receber a bolsa de estudos?

____
____
____
____
____
____
____
____
____
____
____
____

3. Os salários dos funcionários de determinada agência de turismo foram divididos nas seguintes classes (em salários mínimos). Calcule o salário médio que os funcionários dessa agência recebem.

Classe de salário	x_i	f_i
1,0 |—— 2,0	1,5	44
2,0 |—— 3,0	2,5	82
3,0 |—— 4,0	3,5	52
4,0 |—— 5,0	4,5	16
5,0 |—— 6,0	5,5	4
6,0 |—— 7,0	6,5	2
Total		200

Respostas Comentadas

1.

Para calcular o número médio de quartos desses hotéis, você irá aplicar a expressão:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{127 + 136 + ... + 213}{10} \cong 169,9 \text{ quartos}$$

Assim, esses hotéis têm em média 169,9 quartos, aproximadamente. Como os dados não se encontram ordenados do menor para o maior, você tem que primeiramente ordenar esses dados para encontrar a mediana, o que resultará em:

127	*136*	*148*	*156*	*163*	*175*	*189*	*192*	*200*	*213*

E, para calcular a mediana, você vai utilizar a expressão:

$$M_e = \frac{163 + 175}{2} = 169 \text{ quartos}$$

Como não há nenhum valor que se repete, o conjunto de dados não tem moda. Portanto, é amodal.

Nesse caso, tanto a média quanto a mediana representam muito bem a tendência central dos dados, o que pode ser visto por seus valores estarem bastante próximos.

2.

a. A moda das notas do aluno X é 5,3, pois é o valor que mais se repete. Quanto ao aluno Y, não há nenhuma nota repetida, assim, esse conjunto de dados não tem moda, é amodal.

Os dados não se encontram ordenados, do menor para o maior, assim, você tem que primeiramente ordenar esses dados para encontrar a mediana, o que resultará em:

X	5,0	5,3	5,3	5,4	6,4	6,6	6,7	6,8	7,0	7,3	7,4	8,1
Y	5,7	6,5	7,2	7,8	7,9	8,0	8,2	8,7	9,1	9,5	9,6	10,0

E, para calcular a mediana, você vai utilizar a expressão:

Aluno X: $M_e = \dfrac{6,6+6,7}{2} = 6,65$ pontos.

Aluno Y: $M_e = \dfrac{8,0+8,2}{2} = 8,10$ pontos.

O cálculo da média será feito da seguinte maneira:

Aluno X: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{5,0+5,3+...+8,1}{12} \cong 6,44.$

Aluno Y: $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n} = \dfrac{5,7+6,5+...+10,0}{12} \cong 8,18.$

b. Baseado nestas estatísticas, o aluno Y tem maiores chances de receber a bolsa, pois apresenta tanto média quanto mediana maiores que as do aluno X. Confirmando o que você encontrou na atividade final da Aula 8.

3. Para realizar o cálculo do salário médio desses funcionários, utilizamos a expressão $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i f_i}{n}$, devido aos dados estarem agrupados.

Para facilitar o cálculo, vamos primeiramente efetuar o produto ($x_i\, f_i$) que resultará em:

Classe de salário	x_i	f_i	$(x_i f_i)$
1,0 |—— 2,0	1,5	44	66
2,0 |—— 3,0	2,5	82	205
3,0 |—— 4,0	3,5	52	182
4,0 |—— 5,0	4,5	16	72
5,0 |—— 6,0	5,5	4	22
6,0 |—— 7,0	6,5	2	13
Total		200	560

Assim, aplicando a expressão $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i f_i}{n} = \dfrac{560}{200} = 2,8.$ *Concluímos que, em média, os funcionários desta agência de turismo recebem 2,8 salários mínimos.*

Resumo

As medidas de tendência central são importantes para descrever o nível geral de uma distribuição de frequência.

As medidas de tendência central mais utilizadas são a média, a mediana e a moda. A média é a soma dos maiores divididos pela quantidade deles. A mediana é o valor que divide a distribuição ao meio. Já a moda é o valor dominante da série.

Se os dados são homogêneos, podemos usar a média como um valor que resume da série de dados, caso contrário são sugeridos o uso da mediana ou da moda.

Informação sobre a próxima aula

Na próxima aula, serão apresentadas as medidas que indicam o grau de heterogeneidade dos valores da distribuição de frequência, isto é, as medidas de dispersão ou variabilidade.

10 Medidas de dispersão

Meta da aula

Apresentar o conceito de medidas de dispersão, com enfoque no seu cálculo e na interpretação do valor resultante.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. identificar qual medida de dispersão melhor irá representar a variabilidade de um determinado conjunto de dados, que você estará estudando;

2. calcular e identificar, no contexto de um conjunto de dados real, as medidas de dispersão;

3. analisar o valor resultante do cálculo da medida de dispersão dentro de um determinado conjunto de dados em estudo.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é imprescindível que os conceitos de população e amostra, distribuição de frequências, gráficos, medidas de posição e tendência central, das Aulas 1, 5, 6, 7, 8 e 9, respectivamente, estejam bem claros.

Introdução

Na aula anterior, você aprendeu a calcular medidas que resumem o conjunto de dados em apenas um valor numérico, que são as medidas de tendência central. Porém, muitas vezes, apenas a informação de uma medida de tendência central não irá descrever plenamente o conjunto de dados. Por isso, quando calculamos uma medida de tendência central (a média, por exemplo), associamos a ela uma medida de dispersão, para termos a noção do quão espalhados estão os dados em estudo.
Nesta aula, você irá estudar as seguintes medidas de dispersão: amplitude total, variância, desvio padrão e coeficiente de variação.

Medidas de dispersão

Estas estatísticas irão representar a variabilidade existente dentro de um determinado conjunto de dados, ou seja, irão nos auxiliar a quantificar o quão homogêneos estão esses dados. Assim, as medidas de dispersão são utilizadas para quantificar o grau de espalhamento dos valores presentes em uma distribuição, amostra, população etc.

1ª. Amplitude total (R)

A amplitude total é definida como a diferença entre o maior e o menor valor observado.

Sendo:

X_{max} = maior valor da distribuição

X_{min} = menor valor da distribuição

Então:

$R = X_{max} - X_{min}$.

Quanto maior a amplitude total, mais heterogêneos (mais dispersos) são os dados entre si.

Exemplo 1

Considere os seguintes dados: 15, 12, 10, 17, 16, sobre os dias de permanência de turistas brasileiros em visita à Europa. Qual a amplitude desses dados?

R = 17 – 10 = 7.

Exemplo 2

Considere a série a seguir, que representa a temperatura máxima registrada durante 30 dias numa região visitada por turistas.

Tabela 10.1: Cálculo da amplitude total

Temperatura (ºC)	f_i
10	2
15	6
20	12
30	7
40	3
Σ	30

R = 40 – 10 = 30°C.

Atividade

Atende aos Objetivos 2 e 3

1. Considere a distribuição a seguir, de estaturas de turistas de um evento de esportes radicais. Qual a amplitude total desses dados?

Estaturas (cm)	f_i
150 |—— 154	4
154 |—— 158	9
158 |—— 162	11
162 |—— 166	8
166 |—— 170	5
170 |—— 174	3
Σ	40

Resposta Comentada

Para calcular a amplitude total desses dados, devemos primeiramente identificar quais são os valores máximo e mínimo presentes na distribuição. Feito isso, basta utilizar a expressão $R = X_{max} - X_{min}$ *e obter o resultado:*

$R = 174 - 150 = 24$ *cm.*

Vantagens e desvantagens da amplitude total

Faz-se uso da amplitude total quando se quer determinar a amplitude ou variação da temperatura em um dia ou ano, em controle da qualidade ou quando a compreensão popular é mais importante que a exatidão e a eficácia.

A amplitude total tem o inconveniente de só levar em conta os dois valores extremos da série, descuidando-se do conjunto de valores intermediários, o que quase sempre invalida a precisão e a eficácia do resultado. Ela é apenas uma indicação aproximada da dispersão ou variabilidade.

2ª. Variância (S^2)

É a medida de dispersão mais utilizada na prática, pois considera todos os valores numéricos do conjunto de dados.

É a média aritmética dos quadrados dos desvios em relação à média da distribuição.

Desvio (d_i) é a diferença de cada valor da distribuição de sua média:

$$d_i = (xi - \bar{x}).$$

Temos que a variância de uma amostra é dada por:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i \quad \bar{X})^2}{n}$$

Exemplo 3

Obtenha a variância das notas de uma turma em determinada disciplina. As notas foram: 5,0; 7,0; 8,0; 6,5.

$$\bar{x} = \frac{5,0 + 7,0 + 8,0 + 6,5}{4} = 6,6 \text{ nota}$$

Desenvolvimento dos cálculos dos desvios:

$(5 - 6{,}6)^2 = (-1{,}6)^2 = 2{,}56$

$(7 - 6{,}6)^2 = (0{,}4)^2 = 0{,}16$

$(8 - 6{,}6)^2 = (1{,}4)^2 = 1{,}96$

$(6{,}5 - 6{,}6)^2 = (0{,}1)^2 = 0{,}01$

Somatório dos desvios:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 2{,}56 + 0{,}16 + 1{,}96 + 0{,}01 = 4{,}69$$

$$S^2 = \frac{4{,}69}{4} \cong 1{,}17.$$

Um psicólogo foi encarregado de treinar agentes de turismo para o trabalho de visitas a museus. Após o treinamento, ele deu uma nota a cada agente pelo seu aproveitamento, e os resultados foram:

Tabela 10.2: Cálculo da variância

Notas	f_i	x_i	$x_i f_i$	$x^2_i\ f_i$	d_i	$d_i f^i$	$d^2_i\ f_i$
0 |— 2	5	1	5	5	–2	–10	20
2 |— 4	8	3	24	72	–1	–8	8
4 |— 6	14	5	70	350	0	0	0
6 |— 8	10	7	70	490	1	10	10
8 |— 10	7	9	63	567	2	14	28
$\sum$	44	__	232	1.484	___	6	66

Processo longo:

$$S^2 = \frac{\sum x^2{}_i \cdot f_i - \frac{\sum (x_i \cdot f_i)^2}{n}}{n}$$

Então:

$$S^2 = \frac{1.484 - \frac{(232)^2}{44}}{n} = 5{,}9 \text{ nota}^2.$$

Processo breve:

$$S^2 = [\sum d^2{}_i f_i - \frac{\sum(d_i f_i)^2}{n}] x\ h^2 n).$$

Então:

$$S^2 = [\,66 - \frac{(6)^2}{44}\,] . \frac{2^2}{44}\,) = 5,9 \text{ nota}^2$$

Atividade

Atende aos Objetivos 2 e 3

2. Dada a distribuição do tempo (minutos) de espera de malas por turistas em um aeroporto do Rio de Janeiro, obter a variância.

X_i (tempo em minutos)	f_i	$X_i\, f_i$	$X^2_i\, f_i$
2	1	2	4
3	4	12	36
5	5	25	125
6	3	18	108
7	2	14	98
$\sum$	15	71	371

Resposta Comentada

Como os dados estão agrupados em uma tabela de distribuição de frequência, você irá calcular a variância por:

$$S^2 = \frac{\sum x^2{}_i . f_i - \frac{\sum(x_i . f_i)^2}{n}}{n}$$

Substituindo os valores na expressão anterior, temos:

$$S^2 = \frac{371 - \frac{(71)^2}{15}}{15} = 2,3\ (\text{minutos})^2.$$

3ª. Desvio padrão (S)

É definido como a raiz quadrada da variância, utilizado para que a medida de variabilidade fique na mesma escala da variável original. A vantagem na utilização do desvio padrão é que o resultado é na mesma medida, escala, dos dados originais.

Exemplo 4

Calcule o desvio padrão do exemplo 3:

$S = \sqrt{1,56} \cong 1,25$ nota.

Atividade

Atende aos Objetivos 2 e 3

3. Dados os valores do tempo (minutos) de espera de malas por turistas em um aeroporto do Rio de Janeiro (ver tabela da Atividade 2), encontre o desvio padrão. Em seguida, responda: Qual é a vantagem de utilizar o desvio padrão em vez da variância?

Resposta Comentada

Como já foi calculada a variância na Atividade 2, para o cálculo do desvio padrão basta utilizar a expressão: $S = \sqrt{2,3} \cong 1,5$ minuto.

O desvio padrão é, na prática, mais útil que a variância, pois está na mesma escala dos valores da série original. A variância está numa unidade original ao quadrado.

4ª. Coeficiente de Variação (CV)

Operacionalmente, o coeficiente de variação é a razão percentual entre o desvio padrão e a média. Seu resultado não é da mesma grandeza da escala (exemplo: kg, cm, ton etc.); ao contrário, é o resultado que expressa, em porcentagem, a fração que o desvio padrão é em relação à média.

É definido como:

$$CV = \frac{S}{\bar{x}} 100\%$$

Tem-se que:

CV ≤ 15%, baixa dispersão (dados homogêneos).

15% < CV ≤ 30%, média dispersão.

CV > 30%, alta dispersão (dados muito heterogêneos).

Exemplo:

Fornecidos a média, 51 kg, e o desvio padrão, 6,14 kg dos dados de peso de 20 pessoas, encontre o coeficiente de variação.

Logo, CV = $\frac{6,14}{51}$ x 100% = 12,04%, baixa dispersão.

Utiliza-se o coeficiente de variação para comparar a dispersão:

- de populações com estudos envolvendo variáveis diferentes, isto é, com universos em escalas de medidas distintas (pesos, alturas);
- de populações distintas (crianças, adultos);
- de populações com desvios padrões iguais, mas com médias diferentes.

Exemplos:

1. Numa agência de viagens, o salário médio dos homens é de R$ 400,00 com desvio padrão de R$ 150,00 e o das mulheres é de R$ 300,00 com desvio padrão de R$ 120,00. Qual o grupo mais heterogêneo quanto ao salário?

$CV_H = \frac{150 \times 100\%}{400} =$ 38%, alta dispersão.

$CV_M = \frac{120 \times 100\%}{300} =$ 40%, alta dispersão.

O grupo mais heterogêneo é o das mulheres, pois seus salários tiveram uma maior variação.

2. Sejam duas séries, em que se tenha, respectivamente:

x_1 = 700 mm e S_1 = 20mm;

x_2 = 100 mm e S_2 = 20 mm.

Calcule os CV de cada uma das séries.

$CV_1 = \frac{20 \times 100\%}{700} =$ 2,85%, baixa dispersão.

$CV_2 = \frac{20 \times 100\%}{100} =$ 20%, média dispersão.

O coeficiente de variação é utilizado neste caso para a comparação dos grupos, porque tem médias diferentes e o mesmo desvio padrão.

Atividade

Atende aos Objetivos 2 e 3

4. Tomemos os resultados das medidas das estaturas e dos pesos de uma mesma população de turistas. Qual das variáveis tem menor dispersão?

Variáveis	Média	Desvio padrão
Estaturas	175 cm	5,0 cm
Pesos	68 kg	2,0 kg

____
____
____
____

Resposta Comentada

Como foram dados a média e o desvio padrão, temos:

Estatura: $CV_E = \frac{5,0}{175} \times 100\% = 2,85$, *baixa dispersão.*

Peso: $CV_p = \frac{2,0}{68} \times 100\% = 2,94\%$, *baixa dispersão.*

A distribuição menos dispersa foi das estaturas, por apresentar menor CV.

Usamos o coeficiente de variação na escolha entre a média e a mediana para representar os dados de uma distribuição de frequência. Para efeitos práticos, costuma-se considerar que CV superior a 30% indica alto grau de dispersão e, consequentemente, pequena representatividade da média, enquanto que, para valores inferiores a 30%, a média será tanto mais representativa do fato quanto menor for o valor de seu CV.

Então:

CV > 30, a mediana ou a moda.

CV ≤ 30, a média.

Pelo exposto nesta aula, podemos concluir que as medidas de dispersão têm fundamental importância para nortear a magnitude do espalhamento dos dados de uma amostra ou população. Para quantificar a dispersão de uma distribuição, é, muitas vezes, mais interessante utilizar o CV, pois este agrega duas medidas importantes, que são a média e o desvio padrão. Mas não devemos deduzir que a variância e o desvio padrão careçam de utilidade. Pelo contrário, são medidas muito úteis no tratamento de assuntos relativos à inferência estatística.

Atividades Finais

Atendem aos Objetivos 1, 2 e 3

1. O peso (kg) da bagagem de mão de nove passageiros de uma companhia aérea foi medido e resultou em: 2,59; 2,64; 2,60; 2,62; 2,57; 2,61; 2,50; 2,63; 2,64. Calcule o desvio padrão e o coeficiente de variação desses pesos. Interprete esse resultado sob o aspecto da variabilidade dos dados.

2. O número de quartos de dez hotéis distintos de uma mesma rede hoteleira foi informado como: 127; 136; 148; 156; 163; 175; 189; 192; 200; 213. Calcule a amplitude total, o desvio padrão e o coeficiente de variação desses dados. Qual dessas três medidas melhor representa a variabilidade desses dados?

3. Anotou-se o número de faltas de 50 funcionários de determinada agência de turismo, resultando na seguinte tabela de distribuição de frequências. Calcule o desvio padrão das faltas cometidas.

Faltas	**f_i**
0	25
1	20
2	3
3	1
4	1
Total	50

Respostas Comentadas

1. Para calcular o desvio padrão desses pesos, precisamos primeiramente calcular a média e a variância:

Média: $\overline{x} = \dfrac{2,59 + \ldots + 2,64}{9} = 2,6\,\text{kg}.$

Variância: $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_1 - \overline{X}\right)^2}{n} = \dfrac{(2,59 - 2,60)^2 + \ldots + (2,64 - 2,60)^2}{9} = 0,002\,\text{kg}^2$

Assim, o desvio padrão que é dado pela raiz quadrada da variância resultará em:

Desvio padrão: $S = \sqrt{0,002\,\text{kg}^2} \cong 0,044\,\text{kg}.$

O coeficiente de variação é dado pela razão entre a média e o desvio padrão; assim, seu valor para este conjunto de dados será: $CV = \dfrac{S}{\overline{x}}100\% = \dfrac{0,044}{2,6}100 = 1,69$*, valor que indica baixa dispersão dos pesos das bagagens de mão.*

2. A amplitude é dada pela diferença entre o maior e o menor valor do conjunto de dados; assim, calculamos a amplitude por:

Amplitude: R = 213-127 = 86 quartos.

Para calcular o desvio padrão desses quartos, precisamos primeiramente calcular a média e a variância:

Média: $\overline{x} = \dfrac{127 + \ldots + 213}{10} = 169,9 \cong 170$ quartos.

Variância: $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_1 - \overline{X}\right)^2}{n} = \dfrac{(127 - 169,9)^2 + \ldots + (213 - 169,9)^2}{10} = 735,29\,\textit{quartos}^2.$

Assim, o desvio padrão que é dado pela raiz quadrada da variância resultará em:
Desvio padrão: $S = \sqrt{816,98} \cong 28,58$ *quartos.*

O coeficiente de variação é dado pela razão entre a média e o desvio padrão; assim, seu valor para este conjunto de dados será: $CV = \frac{S}{\bar{x}} 100\% = \frac{28,58}{169,9} 100 \cong 16,82$*, valor que indica média dispersão do número de quartos desses hotéis.*

Das quatro medidas de dispersão, a que melhor representa a variabilidade desses dados é o coeficiente de variação, pois essa estatística quantifica e indica o grau de espalhamento dos dados em questão.

3. Para facilitar o cálculo, vamos primeiramente efetuar os produtos $(x_i\, f_i)$ *e* $(x_i^2\, f_i)$*, que resultarão em:*

Faltas = x_i	**f_i**	**$x_i\, f_i$**	**$x_i^2\, f_i$**
0	25	0	0
1	20	20	20
2	3	6	12
3	1	3	9
4	1	4	16
Total	50	33	57

Assim, para o cálculo da variância, vamos utilizar:

$$S^2 = \frac{\sum x^2 f_i - \frac{\sum (x_i\,.f_i\,)^2}{n}}{n}$$

Substituindo os valores na expressão, teremos:

$$S^2 = \frac{57 - \frac{(33)^2}{50}}{50} = 0,7044 \text{ faltas}^2.$$

Mas foi pedido para calcular o desvio padrão; assim, utilizando a expressão:

$$S = \sqrt{0,7044 \text{ faltas}^2} \cong 0,84 \text{ faltas.}$$

Resumo

As medidas de dispersão ou variabilidade indicam a grande heterogeneidade ou dispersão dos valores da série.
As mais importantes e usadas nas áreas da pesquisa são a variância e o desvio padrão.
A variância é a média dos quadrados do desvio em relação à média. O desvio padrão é a medida de dispersão por excelência, é a raiz quadrada da variância.
As medidas de dispersão completam a informação dada pela média.

Informação sobre a próxima aula

Na próxima aula, apresentaremos as medidas que indicam a forma de uma distribuição. Essa informação é útil para a interferência e modelagem estatística.

11 Medidas da forma de uma distribuição: assimetria

Meta da aula

Apresentar o conceito de medidas da forma de uma distribuição com relação à sua simetria e com enfoque no seu cálculo e na interpretação do valor resultante.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. reconhecer o conceito de assimetria de uma distribuição e sua importância dentro da estatística;
2. calcular a assimetria dentro de um conjunto de dados.

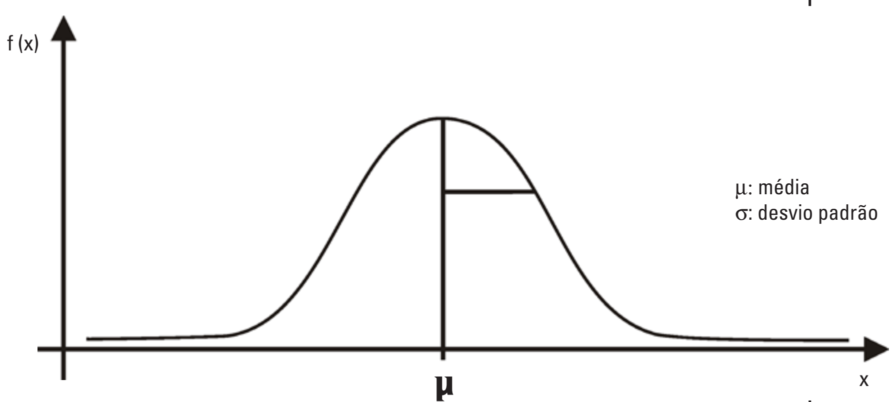
Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é imprescindível que os conceitos de distribuição de frequências, tabelas e gráficos, medidas de posição, tendência central e dispersão, das aulas 2, 5, 6, 7, 8, 9 e 10, respectivamente, estejam bem claros.

Introdução

Você aprendeu, nas aulas anteriores, a representar graficamente uma tabela de distribuição de frequências e, por esta representação gráfica, visualizar a diversidade de formas que uma distribuição de frequências pode assumir. Aprendeu, também, a comparar a distribuição de frequência obtida de uma pesquisa prática com uma distribuição de frequência padrão, teórica, como, por exemplo, uma distribuição de frequência clássica, em forma de sino, denominada curva normal, a qual é de importância vital na fundamentação do Cálculo das Probabilidades e da Teoria da Inferência Estatística.

A curva normal tem forma de sino e é simétrica em relação à média (que tem valor igual à mediana e à moda, quando se trata da curva normal), ou seja, se passarmos uma linha exatamente pelo centro da curva, teremos duas metades perfeitamente iguais.



A importância da curva normal se dá em virtude de ela representar graficamente a distribuição normal, que é uma das distribuições de probabilidade mais utilizadas por se adequar bem a diversas situações práticas; ou seja, são vários os fenômenos do cotidiano que seguem uma distribuição normal, ou aproximadamente normal. O peso da bagagem dos turistas, o preço de diárias em hotéis, o gasto diário de turista etc. são exemplos de variáveis que seguem uma distribuição normal, ou, ao menos, se aproximam dessa distribuição.

Assim, para realizar a comparação da forma das distribuições, são utilizadas, no estudo de estatística, duas medidas, a assimetria e a curtose. Nesta aula, vamos iniciar o nosso estudo da forma de uma distribuição de dados com a assimetria.

Assimetria ou distorção (As)

A assimetria é o estudo do grau de enviesamento ou distorção da curva de frequência. O valor enviesado caracteriza o grau de assimetria de uma distribuição em torno de sua média.

Tipos de assimetria:

a. Assimetria positiva.

b. Assimetria negativa.

c. Curva simétrica.

a. Assimetria positiva

Enviesamento à direita, isto é, cauda mais longa à direita. Um valor enviesado positivo indica uma distribuição com uma ponta assimétrica que se estende em direção a valores mais positivos.

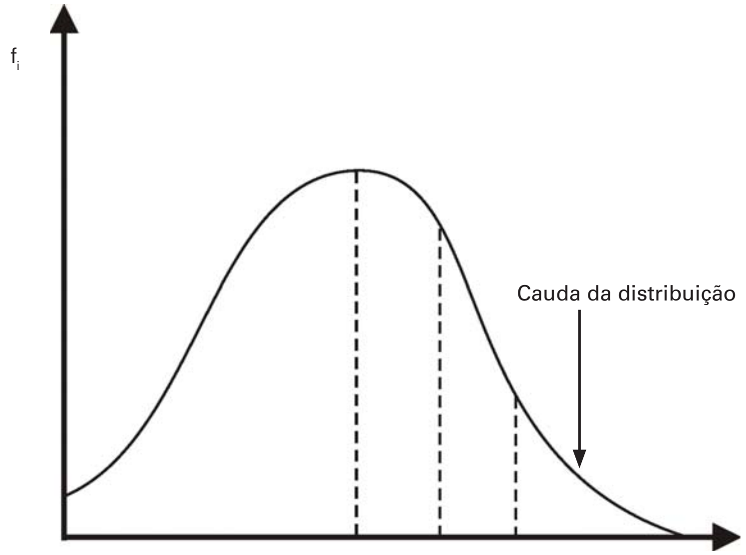


Figura 11.1: Curva normal assimétrica positiva.

Numa distribuição assimétrica positiva, a relação sempre vale:

Moda (M_o) < mediana (M_e) < média ($\overline{X}$).

b. Assimetria negativa

Enviesamento à esquerda, isto é, cauda mais longa à esquerda. Um valor enviesado negativo indica uma distribuição com uma ponta assimétrica que se estende em direção a valores mais negativos.

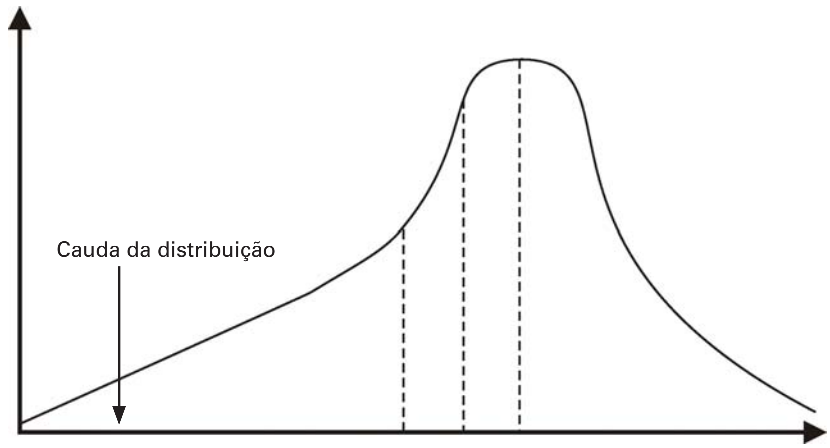


Figura 11.2: Curva normal assimétrica negativa.

Numa distribuição assimétrica negativa, a relação sempre vale:

Média ($\overline{X}$) < mediana (M_e) < moda (M_o).

c. Simétrica

Não há enviesamento. É uma característica da curva normal.

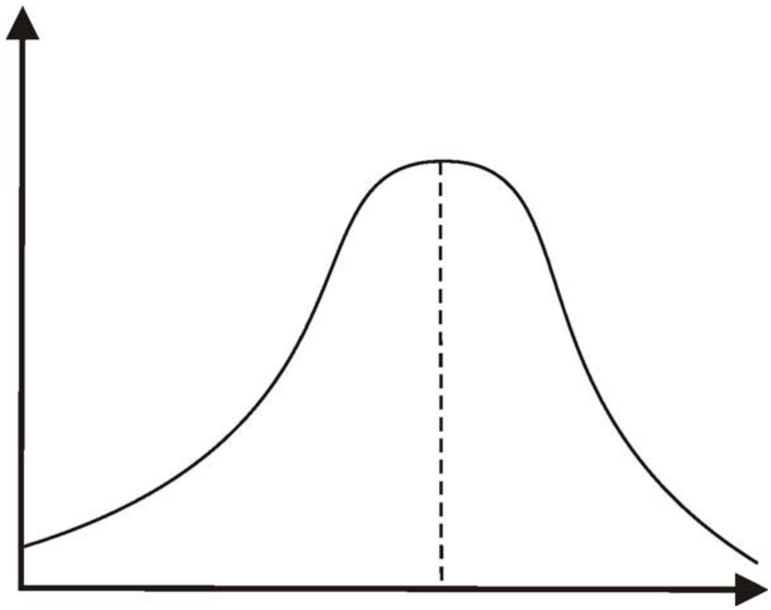


Figura 11.3: Curva normal simétrica.

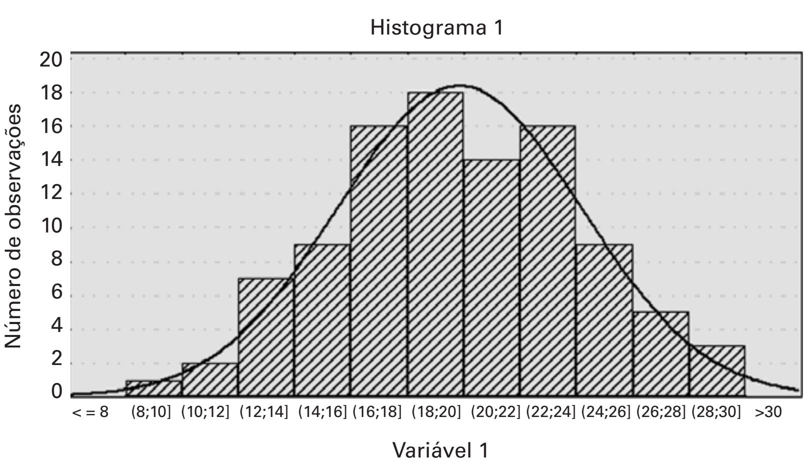
Numa distribuição simétrica, a relação sempre vale:

Média ($\overline{X}$) = mediana (M_e) = moda (M_o).

Atividade

Atende ao Objetivo 1

1. Observe atentamente os histogramas a seguir e classifique as suas distribuições quanto à assimetria.

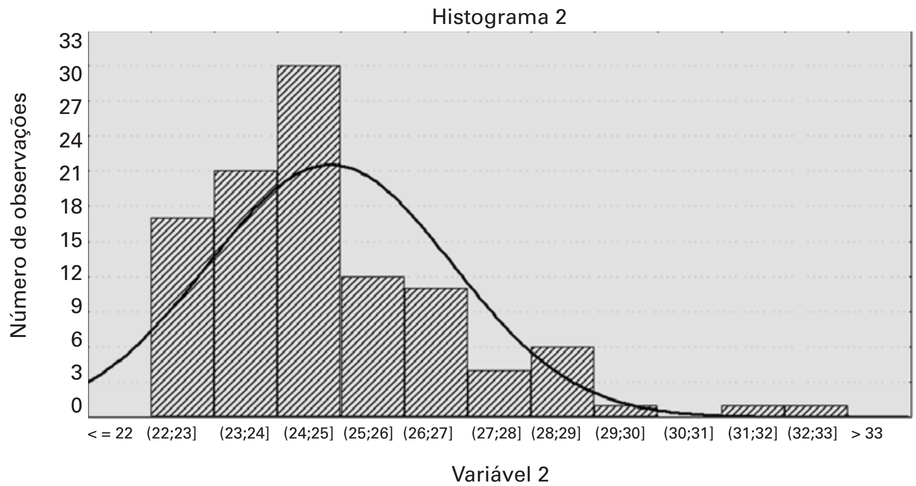


Fonte: Crespo (1999).

____

____

____

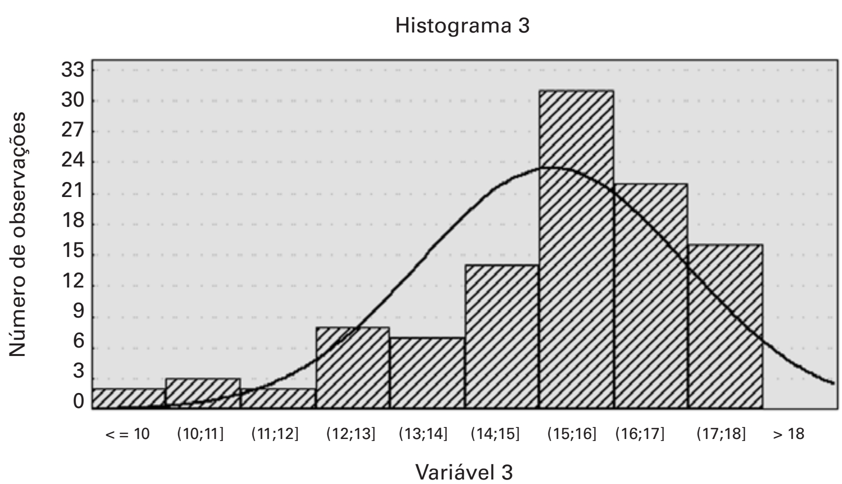


Fonte: Crespo (1999).

____

____

____



Fonte: Crespo (1999).

____

____

____

Resposta Comentada

Histograma 1: curva simétrica. Percebe-se a perfeita simetria existente na curva em questão.
Histograma 2: curva assimétrica à direita, é classificada como assimetria positiva, pois você pode ver que a cauda se prolonga para a direita da curva.
Histograma 3: curva assimétrica à esquerda, é classificada como assimetria negativa, pois você pode ver que a cauda se prolonga para a esquerda da curva.

Coeficientes de assimetria

Existem alguns coeficientes para a determinação e o cálculo da assimetria dos dados de uma distribuição. No nosso curso, iremos aprender dois coeficientes para o cálculo da assimetria, os quais foram propostos por Pearson, denominados coeficientes de assimetria de Pearson, e que são os mais utilizados na prática. São eles:

1) $A_s = \dfrac{\bar{x} - M_o}{S}$.

2) $A_s = 3\,\dfrac{(\bar{x} - M_e)}{S}$.

Em função dos resultados desses coeficientes, é possível determinar o comportamento da curva de freqüência em dois aspectos:

1°) Se:

As = 0, a distribuição é simétrica.

As > 0, a distribuição é assimétrica positiva.

As < 0, a distribuição é assimétrica negativa.

2°) Se:

| As | ≤ 0,15, distribuição praticamente simétrica.

0,15 < | As | ≤ 1, assimetria moderada.

| As | > 1, forte assimetria.

Exemplos:

1. Determinar o coeficiente de assimetria da distribuição de notas de satisfação de turistas com certo passeio turístico e concluir sobre o seu grau de assimetria:

Classes	f_i
50 |—— 60	15
60 |—— 70	20
70 |—— 80	30
80 |—— 90	20
90 |—— 100	15
Σ	100

Calculando, temos:

$\overline{X} = 75$

$M_o = 75$

$S = 12,65$

$$A_s = \frac{x - M_o}{S} = \frac{75 - 75}{12,65} = 0, \quad A_s = 0$$

Assim, devido ao valor zero de assimetria, conclui-se que a distribuição é simétrica.

2. Dizer se a distribuição a seguir pode ser ajustada pela curva normal, analisando o seu grau de assimetria:

Classes	f_i
3 |—— 8	5
8 |——13	15
13|——18	20
18|——23	10

Temos:

$\overline{X} = 14$

$M_o = 15$

$S = 4,5$

$$A_s = \frac{\overline{x} - M_o}{S} = \frac{14 - 15}{4,5} = -\,0,22 \quad A_s = -\,0,22$$

| - 0,22 | – assimetria negativa moderada, em função desse resultado, a distribuição não pode ser ajustada à curva normal.

Atividade

Atende aos Objetivos 1 e 2

2. Calcule e especifique a assimetria dos dias de permanência de 11 turistas estrangeiros em determinada localidade brasileira.

Turista	A	B	C	D	E	F	G	H	I	J	K
Dias	5	6	7	8	9	9	10	10	11	13	14

Resposta Comentada

Para calcular a assimetria, você pode utilizar a expressão:

$$As = \frac{3(\bar{x} - Me)}{s}.$$

Para tanto, são necessários cálculos iniciais, assim:

Cálculo da média: $\bar{x} = \frac{5+6+...+14}{11} = 9,27$ *dias.*

Cálculo da mediana: $Me = x_6 = 9.$

Cálculo do desvio padrão:

$$s = \sqrt{s^2} = \sqrt{\frac{(5-9,27)^2 + ... + (14-9,27)^2}{11-1}} = 2,76 \text{ dias.}$$

Cálculo da assimetria, utilizando os valores encontrados anteriormente:

$$As = \frac{3(9,27-9)}{2,76} = 0,29.$$

Esse valor obtido para o coeficiente de assimetria indica que a distribuição dos dias de permanência de 11 turistas estrangeiros em determinada localidade brasileira apresenta assimetria positiva.

Coeficiente momento de assimetria (MAs)

O coeficiente momento de assimetria é definido como um afastamento do eixo de simetria de uma distribuição de valores, e permite aferir o quanto a distribuição se afasta da normalidade. É calculado pela seguinte expressão:

$$M_{As} = \frac{n}{(n-1)(n-2)} \sum \left[\frac{(x-\bar{x})}{S} \right]^3 fi$$

Exemplo:

Vamos calcular o M_{As} das declarações de despesas feitas pelos executivos de uma agência de viagens (em 100 reais).

Classes	f_i	x_i	$\frac{x_i - x}{S}$	$\frac{x_i - x}{S} f_i$
00|—15	12	7,5	-3,58	-42,96
15|—30	23	22,5	-0,59	-13,57
30|—45	26	37,5	0,00	0,00
45|—60	18	52,5	0,16	2,88
60|—75	13	67,5	1,90	24,70
75|—90	8	82,5	7,20	57,65
Total	100	—	—	28,65

Temos:

$$\overline{X} = 40,65$$

$$S = 21,67$$

$$M_{As} = \frac{100}{(100-1)(100-2)} \,.\, 28,65 = 0,30$$

Assim, pelo valor do coeficiente momento de assimetria, conclui-se que essa distribuição tem assimetria positiva moderada.

Desta forma, pelo que foi exposto nesta aula, você pode concluir que uma distribuição será simétrica quando 50% dos valores observados estiverem acima da observação que ocupa a posição central de distribuição, e 50% dos valores estiverem abaixo. Logo, a condição de assimetria estará presente quando houver deslocamento do centro da distribuição, o que pode ser quantificado por meio dos cálculos dos coeficientes de assimetria.

Atividades Finais

Atendem aos Objetivos 1 e 2

1. Apuraram-se os dados sobre o peso (em kg) das malas de um grupo de turistas em visita a determinado destino europeu, e verificou-se que, em média, as malas pesavam 56,2 kg, apresentaram mediana de 53,2 kg e variância de 11,55 kg^2. A partir dessas informações, calcule e especifique o tipo de assimetria apresentado por esses pesos da bagagem.

2. No quadro a seguir estão as estatísticas referentes à média, mediana e variância do tempo de viagem (em horas) da cidade de origem do turista para a cidade do seu destino turístico, via terrestre e via aérea. Calcule a assimetria pelos dois coeficientes propostos por Pearson e especifique o tipo de assimetria para cada tipo de transporte.

Estatística	**Tipo de transporte**	
	Terrestre	**Aéreo**
Média	9,4	5,3
Mediana	10,4	5,3
Moda	12,4	5,3
Variância	3,7	3,0

Respostas Comentadas

1.

Utilizando a expressão:

$$As = \frac{3(\overline{x} - Me)}{s} \Rightarrow As = \frac{3(56{,}2 - 52{,}9)}{\sqrt{11{,}5}} = 2{,}92$$

O valor obtido para o coeficiente de assimetria indica que a distribuição dos pesos das malas desses turistas apresenta assimetria positiva, ou seja, se traçarmos a curva de distribuição desses dados, veremos que ela será assimétrica à direita.

2.

Utilizando a expressão $As = \frac{3(\bar{x} - Me)}{s}$ *, resultará:*

para o transporte aéreo: $As = \frac{3(5,3 - 5,3)}{\sqrt{3,0}} = 0$ *;*

para o transporte terrestre: $As = \frac{3(9,4 - 10,4)}{\sqrt{3,7}} \cong -1,56$ *.*

Utilizando a expressão $A_s = \frac{\bar{x} - M_o}{s}$ *resultará:*

para o transporte aéreo: $A_s = \frac{(5,3 - 5,3)}{\sqrt{3,0}} = 0$*;*

para o transporte terrestre: $A_s = \frac{(9,4 - 12,4)}{\sqrt{3,7}} \cong -1,56.$

Desse modo, podemos especificar como condição de simetria a distribuição do tempo de viagem via transporte aéreo, e como condição de assimetria negativa (curva assimétrica à esquerda), a distribuição do tempo de viagem via transporte terrestre.

Resumo

As medidas da forma de uma distribuição indicam o modelo de probabilidades da distribuição de frequência em estudo.
As distribuições de frequências são comparadas à distribuição normal quanto à simetria e curtose.
A simetria é o grau de enviesamento da curva de frequência; quanto à simetria, ela pode ser positiva ou negativa.

Informação sobre a próxima aula

Na próxima aula, iremos estudar a próxima medida estatística que indica a forma de uma distribuição de frequência: a curtose.

12 Medidas da forma de uma distribuição: curtose

Meta da aula

Apresentar o conceito de medidas da forma de uma distribuição com relação ao seu achatamento, com enfoque no seu cálculo e na interpretação do valor resultante.

Objetivos

Esperamos que, após o estudo do conteúdo desta aula, você seja capaz de:

1. reconhecer o conceito de curtose de uma distribuição e sua importância dentro da estatística;

2. calcular a curtose dentro de um conjunto de dados.

Pré-requisitos

Para que você encontre maior facilidade na compreensão desta aula, é imprescindível que os conceitos de distribuição de frequências, tabelas e gráficos, medidas de posição, tendência central e dispersão e assimetria, das Aulas 2, 5, 6, 7, 8, 9, 10 e 11, respectivamente, estejam bem claros.

Introdução

Quando é apresentado ao pesquisador um conjunto de dados para estudo, o que ele pretende é extrair o máximo possível de informações desses dados, pois, assim, a compreensão, a análise e as conclusões a respeito dos dados em questão estarão mais embasadas. Por isso, além das medidas de tendência central, de dispersão e de assimetria, outra informação que irá agregar informação estatística ao conjunto de dados é a curtose.

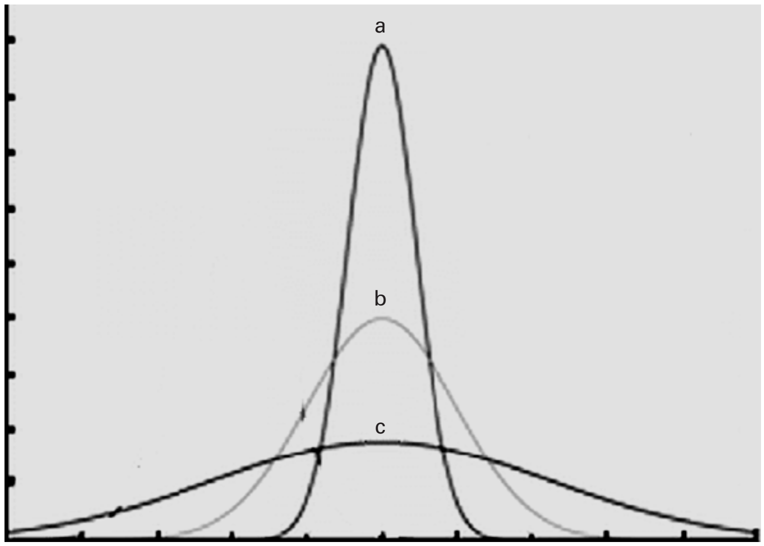
Assim como a assimetria, que estudamos na aula anterior, a curtose é uma medida da forma da distribuição. Nesse caso, estaremos interessados na forma do pico da curva.

Desse modo, vamos estudar o conceito de curtose que será definida formalmente a seguir.

Curtose (K)

É o estudo do grau de achatamento da curva de frequência e irá auxiliar na representação da variabilidade existente dentro de um determinado conjunto de dados. A curtose caracteriza uma distribuição em cume ou plana, se comparada à distribuição normal. De acordo com o grau de curtose, utilizamos três classificações para as curvas de frequência:

- Mesocúrtica: é aquela que denominamos de padrão, não é nem muito achatada nem muito alongada. A curva Normal (ver definição na Aula 11) tem a característica de ser mesocúrtica, ou seja, segue este padrão. Ver **Figura 12.1**, curva b.
- Leptocúrtica: é a curva mais alongada, tem o pico bastante acentuado, quando comparada à curva Normal. Ver **Figura 12.1**, curva a.
- Platicúrtica: é a curva mais achatada, o seu pico é bastante suave, quase imperceptível, quando comparada à curva Normal. Ver **Figura 12.1**, curva c.



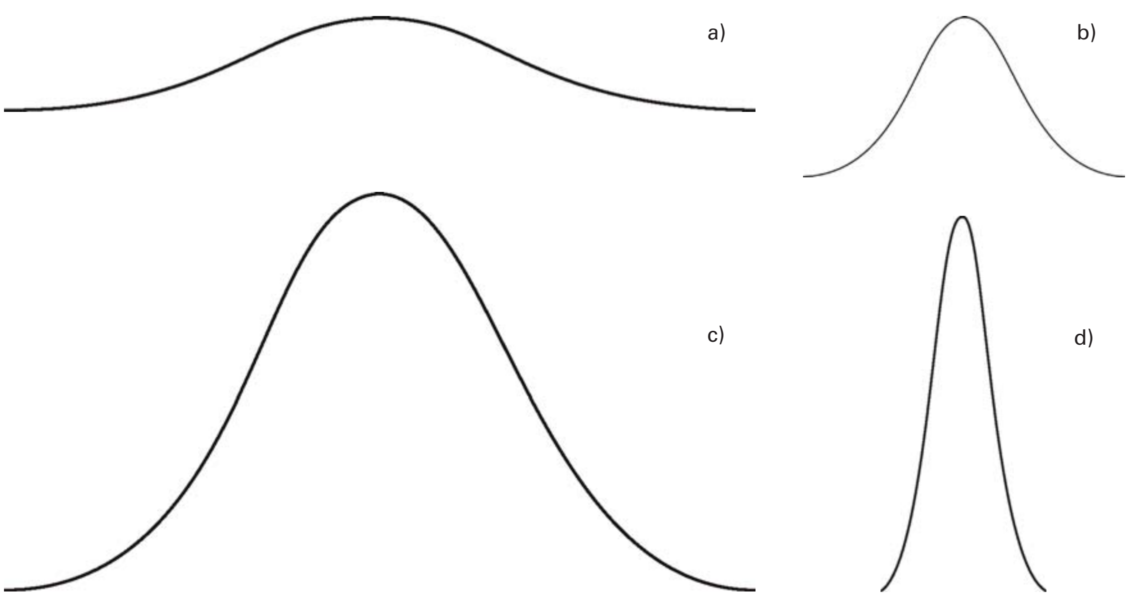
Curva a: leptocúrtiva.
Curva b: mesocúrtica.
Curva c: platicúrtica.

Figura 12.1: Curvas com diferentes graus de achatamento.

Atividade

Atende ao Objetivo 1

1. Observando o formato das seguintes curvas, como você as classifica com relação à curtose?



____
____
____
____
____
____
____
____
____
____

Respostas Comentadas

a. Platicúrtica: você pode ver que essa curva tem um alto grau de achatamento, havendo grande dispersão dos dados.
b. Mesocúrtica: você pode notar que essa curva é bem semelhante à curva Normal, não sendo nem muito achatada nem muito alongada.
c. Mesocúrtica: você pode notar que essa curva é bem semelhante à curva Normal, não sendo nem muito achatada nem muito alongada.
d. Leptocúrtica: você pode ver que essa curva tem um baixo grau de achatamento, havendo menor dispersão dos dados, se compararmos com as demais curvas desta atividade.

Coeficiente de curtose (K)

Para medir o grau de curtose, podemos utilizar o coeficiente:

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

em que:

Q_3 é o terceiro quartil;

Q_1 é o primeiro quartil;

P_{90} é o nonagésimo percentil;

P_{10} é o décimo percentil.

Assim, de acordo com o valor calculado para a curtose (K), iremos compará-lo ao valor 0,263 e verificar qual é o grau de achatamento da distribuição em estudo. Deste modo, se:

$K = 0{,}263$, a distribuição é mesocúrtica;

$K < 0{,}263$, a distribuição é leptocúrtica (distribuição em cume);

$K > 0{,}263$, a distribuição é platicúrtica (distribuição plana).

Quando os dados em questão seguem uma distribuição Normal, o valor calculado para o coeficiente de curtose será sempre 0,263.

Exemplo

Calcular o coeficiente de curtose da distribuição a seguir e classificar a distribuição com relação ao seu grau de achatamento:

Tabela 12.1: Dados para cálculo da curtose

Classes	f_i
3 |— 8	5
8 |— 13	15
13 |— 18	20
18 |— 23	10

Para os dados anteriores, temos que o primeiro quartil, o terceiro quartil, o décimo percentil e o nonagésimo percentil têm os seguintes valores:

$Q_1 = 10{,}5$

$Q_3 = 17{,}38$

$P_{10} = 8{,}00$

$P_{90} = 20{,}5$

Então, podemos colocar esses valores na fórmula e obter o coeficiente de curtose:

$$K = \frac{17{,}38 - 10{,}5}{2\,(20{,}5 - 8{,}0)} = \mathbf{0{,}270} > 0{,}263$$

Por este valor calculado, classificamos a distribuição como platicúrtica ou plana.

Atividade

Atende aos Objetivos 1 e 2

2. No quadro a seguir estão os dias de permanência de 11 turistas estrangeiros em determinada localidade brasileira. Calcule e classifique a curtose dos dias de permanência.

Turista	A	B	C	D	E	F	G	H	I	J	K
Dias	5	6	7	8	9	9	10	11	11	13	14

Resposta Comentada

Para calcular o coeficiente de curtose, utilize a expressão:

$$K = \frac{Q_1 - Q_3}{2(P_{90} - P_{10})}$$

Assim, temos que, primeiramente, obter os valores dos quartis e dos percentis, que irão resultar em:

$Q_1 = 7$
$Q_3 = 11$
$P_{10} = 6$
$P_{90} = 13$

Substituindo os valores dos quartis e dos percentis obtidos para esses dados, podemos calcular o K, que resultará:

$$K = \frac{11 - 7}{2\,(13 - 6)} = 0{,}286 > 0{,}263.$$

Por este valor calculado, classificamos a distribuição como platicúrtica ou plana.

Coeficiente momento de curtose (M_K)

Neste coeficiente, a curtose positiva indica uma distribuição relativamente em cume. A curtose negativa indica uma distribuição relativamente plana. A curtose também pode ser definida da seguinte forma:

$\mathbf{M_K} =$

$$\left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)} \qquad \sum\left[\frac{\overline{X}_i - X}{S}\right]^4 f_i\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Exemplo

Calcular o coeficiente momento de curtose (M_k) para os dados das declarações de despesas feitas pelos executivos de uma agência de viagens (em 100 reais).

Tabela 12.2: Dados para cálculo do momento de curtose

Classes	f_i	x_i	$\left[\frac{X_i - \overline{X}}{S}\right]^4$	$\left[\frac{\overline{X}_i - X}{S}\right]^4 f_i$
00|— 15	12	7,5	5,48	65,72
15|— 30	23	22,5	0,49	11,31
30|— 45	26	37,5	0,00	0,01
45|— 60	18	52,5	0,09	1,61
60|— 75	13	67,5	2,36	30,64
75|— 90	8	82,5	13,91	111,28
Total	100	–	–	220,58

Da tabela, podemos calcular a média e o desvio padrão que resultará nos valores:

$$\overline{X} = 40,65$$
$$S = 21,67$$

Assim, obtemos o coeficiente momento de curtose (M_k), dado por:

$$M_k = \left[\frac{100(100+1)}{(100-1)(100-2)(100-3)}.\ 200,58\right] - \frac{3(100-1)^2}{(100-2)(100-3)} = -0,72$$

Por este valor calculado do coeficiente momento de curtose, concluímos que a distribuição é platicúrtica ou plana.

A curva Normal tem As = 0 e K = 0,263 ou M_{As} = 0 e M_k = 0 simultaneamente. Portanto, a distribuição normal é simétrica e mesocúrtica.

Assim, pelo exposto na aula, você pode concluir que a curtose irá medir o grau de achatamento de uma distribuição de dados em relação à distribuição normal, e tal grau pode ser quantificado por meio dos cálculos dos coeficientes de curtose.

Atividades Finais

Atendem aos Objetivos 1 e 2

1. Na tabela a seguir, encontram-se os resultados (em pontos) de uma avaliação aplicada a funcionários de uma companhia aérea de turnos diurno e noturno. Calcule e interprete a curtose dos pontos dos dois turnos.

	Média	S	Q_1	Q_3	P_{10}	P_{90}
Diurno	19	2,38	17,6	21	16,4	24,6
Noturno	20	4,64	16,5	23,5	13,6	26,4

2. A tabela a seguir mostra as notas em doze disciplinas (em média por disciplina) de dois alunos distintos (X e Y), concorrentes a uma bolsa de estudos do 3º período de determinado curso. Calcule e especifique a distribuição das notas quanto à curtose, para cada um dos alunos.

Aluno	Disciplina											
	A	B	C	D	E	F	G	H	I	J	K	L
X	6,8	6,4	8,1	5,4	5,3	7,0	6,7	5,0	6,6	7,3	7,4	5,3
Y	8,7	7,9	10,0	7,8	6,5	9,1	8,2	5,7	8,0	9,6	9,5	7,2

3. No quadro a seguir estão os pesos (em kg) das bagagens de mão de 11 passageiros de determinado voo. Calcule e interprete a curtose desses pesos de bagagem de mão.

Turista	A	B	C	D	E	F	G	H	I	J	K
Dias	5,5	5,6	5,9	6,3	6,4	6,6	7,3	7,5	9,2	9,4	9,7

Respostas Comentadas

1. Para os pontos dos funcionários do turno diurno, temos:

$$K = \frac{21 - 17,6}{2(26,4 - 16,4)} = \mathbf{0,207} > 0,263.$$

Para os pontos dos funcionários do turno noturno, temos:

$$K = \frac{23,5 - 16,5}{2(26,4 - 13,6)} = \mathbf{0,273} > 0,263.$$

Interpretação: Por esses valores calculados, classificamos a distribuição dos pontos dos funcionários do turno diurno como leptocúrtica (curva alongada) e a distribuição dos pontos dos funcionários do turno diurno, como platicúrtica (curva achatada).

2. Para calcular o coeficiente de curtose, devemos, primeiramente, calcular os quartis e percentis para cada um dos alunos.

Cálculos dos quartis, percentis e curtose para o Aluno X:

Q_1

100% ——— 12

25% ——— $x \Rightarrow Q_1 = x_3 = 5,3$

Q_3

100% ——— 12

75% ——— $x \Rightarrow Q_3 = x_9 = 7,0$

P_{10}

100% ——— 12

10% ——— $x \Rightarrow P_{10} = x_2 = 5,3$

P_{90}

100% ——— 12

90% ——— $x \Rightarrow P_{90} = x_{11} = 7,4$

$$K = \frac{7,0 - 5,3}{2(7,4 - 5,3)} = \mathbf{0,400} > 0,263$$

Interpretação: Em virtude do coeficiente de curtose calculado apresentar maior valor que o padrão, classificamos a distribuição das notas do Aluno X como platicúrtica (curva achatada).

Cálculos dos quartis, percentis e curtose para o Aluno X:

Q_1

100% ——— 12

25% ——— $y \Rightarrow Q_1 = y_3 = 7,2$

Q_3

100% ——— 12

75% ——— $y \Rightarrow Q_3 = y_9 = 9,1$

P_{10}

100% ——— 12

10% ——— $y \Rightarrow P_{10} = y_2 = 6,5$

P_{90}

100% ——— 12

90% ——— $y \Rightarrow P_{90} = y_{11} = 9,6$

$$K = \frac{9,1 - 7,2}{2\,(9,6 - 6,5)} = \mathbf{0,310} > 0,263$$

Interpretação: Em virtude do coeficiente de curtose calculado apresentar maior valor que o padrão, classificamos a distribuição das notas do Aluno Y como platicúrtica (curva achatada).

O fato de os dois coeficientes de curtose acusarem a distribuição das notas dos alunos X e Y como platicúrticas, este não é um bom indicador para determinar qual dos alunos tem maiores chances de receber a bolsa de estudos. Para tanto, devem ser observadas e calculadas outras medidas para esses conjuntos de dados, tais como a média, a mediana, o desvio padrão e o coeficiente de variação, conforme visto nas Aulas 9 e 10.

3. Para calcular o coeficiente de curtose, devemos, primeiramente, calcular os quartis e percentis dos pesos.

Cálculos dos quartis, percentis e curtose para a variável peso (kg):

Q_1

100% ——— 11

25% ——— $x \Rightarrow Q_1 = x_3 = 5,9$

Q_3

100% ——— 11

75% ——— $x \Rightarrow Q_3 = x_9 = 9,2$

P_{10}

100% ——— 11

10% ——— $x \Rightarrow P_{10} = x_2 = 5,6$

P_{90}

100% ——— 11

90% ——— $x \Rightarrow P_{90} = x_{11} = 9,4$

$$K = \frac{9,2 - 5,9}{2\,(9,4 - 5,6)} = 0,434 > 0,263$$

Interpretação: Em virtude do coeficiente de curtose calculado apresentar maior valor que o padrão, classificamos a distribuição dos pesos da bagagem de mão desses 11 passageiros como platicúrtica (curva achatada).

Resumo

A curtose é o grau de achatamento da distribuição de frequência.

A distribuição de frequência é comparada à curva normal, que é mesocúrtica.

Se a distribuição de frequência for mais alta que a normal, ela será leptocúrtica. Se ela for mais baixa que a normal, ela será platicúrdica.

Informação sobre a próxima aula

Na próxima aula, iremos estudar os conceitos fundamentais do cálculo das probabilidades.

Referências

Estatística Aplicada ao Turismo

Aula 1

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. 163p.

HOUAISS, Antonio; VILLAR, Mauro de Salles. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro, 2001. 2922 p.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. São Paulo: Atual, 1981. 321p.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. 160p.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 2003.

TOLED, O. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. 459p.

Aula 2

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3.ed. São Paulo: Alfa-ômega, 1978. 163 p.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 2003.

TOLED, O. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 3

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 1983.

TOLEDO. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

WORLD TOURISM ORGANIZATION. Disponível em: <http://www.world-tourism.org/>. Acesso em: 24 nov. 2008.

Aula 4

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 1983.

TOLEDO. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 5

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 2003.

TOLEDO. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 6

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D. A.; BERENSON, M. L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 7

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D. A.; BERENSON, M. L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. São Paulo: Atual, 1981. p. 321.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. São Paulo: Atlas, 1983.

TOLED, O. G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 8

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D.A.; BERENSON, M.L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. 5. ed. São Paulo: Saraiva, 2003. p. 526.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. 2. ed. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 9

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D.A.; BERENSON, M.L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. 5. ed. São Paulo: Saraiva, 2003. p. 526.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. 2. ed. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

Aula 10

GATTI, B. H.; FERRES, N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D. A.; BERENSON, M. L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. 5. ed. São Paulo: Saraiva, 2003. p. 526.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo*. 2. ed. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas: 1983. p. 459.

Aula 11

CRESPO, Arnot. *Estatística fácil*. São Paulo: Saraiva, 1999.

GATTI, B. H.; FERRES, N. L. *Estatística básica para ciências humanas*. 3. ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D. A.; BERENSON, M. L.; STEPHAN, D. *Estatística*: teoria e aplicações usando Microsoft EXCEL em português. Rio de Janeiro: LCT, 1998. p. 811.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. 5. ed. São Paulo: Saraiva, 2003. p. 526.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

TIBONI, C. G. *Estatística básica para o curso de turismo.* 2. ed. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

TIBONI, C. G. *Estatística básica para o curso de turismo*. 2ª ed. São Paulo: Atlas, 2003.

TOLEDO, G. L.; OVALLE, I. I. *Estatística básica*. São Paulo: Atlas, 1983. p. 459.

NAZARETH, H. *Curso básico de estatística*. São Paulo: Ática, 1996. p. 160.

MORETTIN, P. A.; BUSSAB, W.O. *Estatística básica*. 5ª ed. São Paulo: Saraiva, 2003. p. 526.

GATTI, B. H.; FERRES N. L. *Estatística básica para ciências humanas*. 3ª ed. São Paulo: Alfa-ômega, 1978. p. 163.

LEVINE, D.A.; BERENSON, M.L.; STEPHAN, D. *Estatística: teoria e aplicações usando microsoft EXCEL em português*. Rio de Janeiro: LCT, 1998. p. 811.