



Matemática Multimídia

ANÁLISE DE DADOS
E PROBABILIDADE



GUIA DO PROFESSOR



Software

Medidas do corpo – gráficos univariados

Objetivos da unidade

1. Analisar representação gráfica de dados estatísticos;
2. Familiarizar o aluno com gráficos de barras, setores e histograma;
3. Introduzir o conceito de quantil.

REQUISITOS DE SOFTWARE Navegador moderno (Internet Explorer 7.0+ ou Firefox 3.0+), Adobe Flash Player 9.0+ e máquina Java 1.5+.

RESTRIÇÕES DE ACESSIBILIDADE Este software não possui recurso nativo de alto contraste nem possibilita navegação plena por teclado.

LICENÇA Esta obra está licenciada sob uma licença Creative Commons 



UNICAMP



FUNDO NACIONAL
DE DESENVOLVIMENTO
DA EDUCAÇÃO

Secretaria de
Educação a Distância

Ministério da
Ciência e Tecnologia

Ministério
da Educação



Medidas do corpo – gráficos univariados

GUIA DO PROFESSOR

Sinopse

Neste software, o aluno irá estudar um pouco de estatística descritiva utilizando dados de três variáveis: gênero, altura e número do calçado. Vamos explorar vários tipos diferentes de gráficos de frequências para uma variável e suas características, bem como introduzir a noção de quantil.

Este software faz parte de uma sequência de softwares de estatística que utilizam o mesmo conjunto de dados.

Conteúdos

- Estatística, Interpretação de Gráficos e Dados;
- Gráficos de Barras, Setores e Histograma;
- Quantis.

Objetivos

1. Analisar representação gráfica de dados estatísticos;
2. Familiarizar o aluno com gráficos de barras, setores e histograma;
3. Introduzir o conceito de quantil.

Duração

Uma aula dupla.

Recomendação de uso

Sugerimos que os dados utilizados neste software sejam previamente coletados e registrados em uma tabela pelos alunos.

Material relacionado

- Softwares: Medidas do corpo – Boxplot, Medidas do corpo – gráficos de dispersão;
- Vídeos: Expresso lanches;
- Experimentos: Variáveis antropométricas.



Introdução

A pesquisa científica é um processo de aprendizagem. Neste contexto, os métodos estatísticos oferecem ferramentas matemáticas que permitem otimizar esse processo.

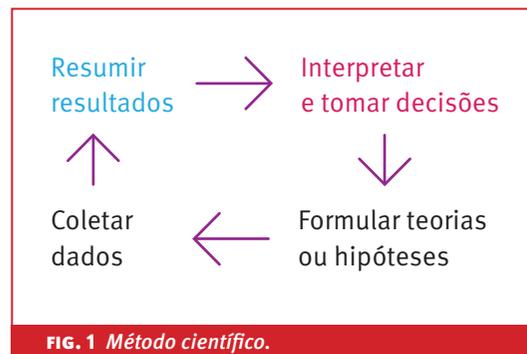
Basicamente, podemos identificar numa pesquisa científica os seguintes estágios:

1. Formulação de uma hipótese, que terá certas consequências;
2. Amostragem ou coleta de dados;
3. Resumo, representação gráfica e comparação dos dados obtidos com o que seria de se esperar de acordo com a hipótese estabelecida;
4. Aceitação ou rejeição da hipótese. No caso de rejeição, segue-se a formulação de uma nova hipótese. No caso de aceitação, a hipótese é mantida até que novas amostras determinem sua rejeição.

Essas etapas formam um ciclo iterativo entre o avanço teórico (hipótese) e os procedimentos de obtenção de dados.

A formulação das hipóteses está relacionada ao levantamento de possíveis respostas para um problema específico.

Ao coletar dados, estamos interessados em obter informação que permita manter a validade de uma hipótese ou que entregue evidências suficientes para rejeitá-la, caso em que serão formuladas novas hipóteses, por sua vez testadas com uma nova coleta de dados e assim por diante.



O software

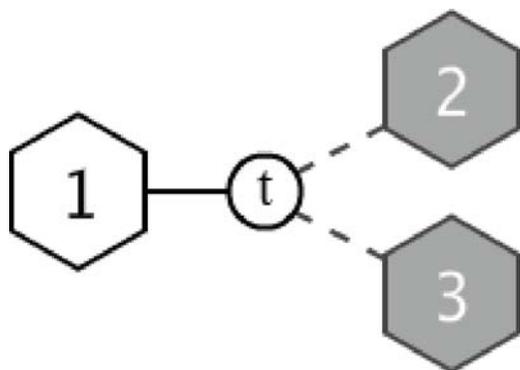
Estrutura do software

Neste software, pretendemos abordar diversas representações gráficas possíveis para um conjunto de dados, como descrito brevemente na Introdução do software e no seu texto de apresentação.

Na ATIVIDADE 1, o aluno deverá fornecer os dados que serão utilizados ao longo do software. Caso outro software da mesma sequência já tenha sido utilizado no mesmo computador, os dados serão carregados automaticamente. O experimento “Variáveis antropométricas” também utiliza um conjunto de dados semelhante e sugerimos que seja usado como uma primeira atividade para que os alunos colem esses dados e depois possam usá-los nos softwares dessa mesma sequência.

Na ATIVIDADE 2, discutimos duas variáveis discretas, em termos de representações gráficas adequadas e medidas-resumo mais comumente utilizadas. Uma das variáveis é Gênero, com duas categorias possíveis, e a outra é Número do Calçado, com categorias que permitem uma ordem natural.

Na ATIVIDADE 3, realizamos o mesmo tipo de análise para uma variável contínua, a saber, Altura de uma pessoa, apresentando gráficos e medidas-resumo adequados para analisar o conjunto de dados.



TELA 1 Mapa do software.

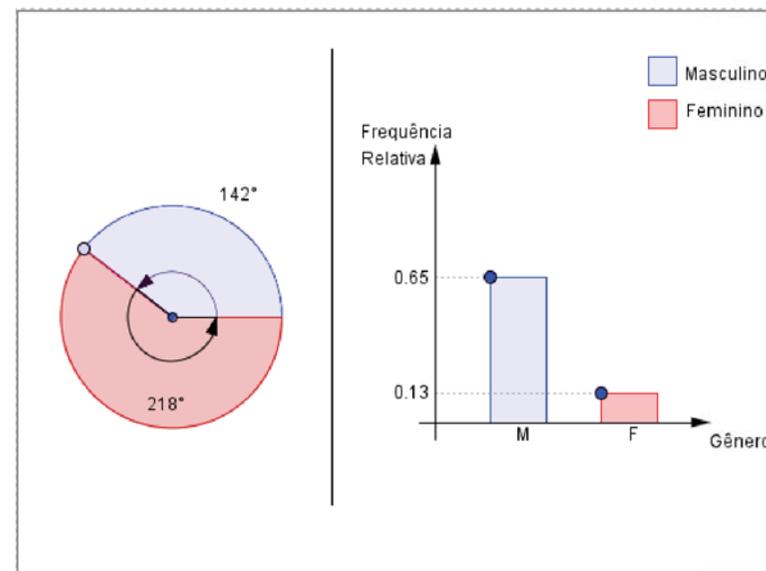
Os alunos devem preencher os dados na ATIVIDADE 1 para ter acesso às atividades 2 e 3.

2 Variável qualitativa e quantitativa discreta

ATIVIDADE

Começamos esta atividade analisando a variável qualitativa Gênero, que tem como possíveis respostas duas categorias não ordenadas: “F”, feminino, e “M”, masculino.

Gráficos típicos para esse tipo de variáveis são o gráfico de setores, ou de pizza, e o gráfico de barras. No primeiro gráfico, a proporção relativa entre as áreas dos setores indica a proporção relativa entre as frequências das categorias, assim como a altura relativa de cada barra, no segundo gráfico.



TELA 2

Assim, se a proporção de mulheres no grupo for de 25%, a relação entre os setores feminino e masculino deve ser de 1 para 3, ou, de outra maneira, a proporção do setor feminino com relação à área total do gráfico deve ser $\frac{1}{4}$. Do mesmo modo, neste exemplo, a altura da barra da categoria “F” deve ser 3 vezes menor que a altura da barra da categoria “M”.

Na segunda parte, analisamos a variável Número do Calçado. Deve-se observar que essa variável tem como possíveis respostas categorias que podem ser ordenadas de acordo com a ordem dos números naturais. Sendo assim, o gráfico de setores, que não entrega informação sobre a ordem das categorias, não é adequado. Já o gráfico de barras permite manter essa informação à vista do leitor.

Novamente, a altura relativa entre as barras representa a frequência na amostra de cada uma das respectivas categorias. Por exemplo, se na amostra o número de pessoas com calçado 37 for 4 vezes maior que o número de pessoas com calçado 34, então a proporção entre as alturas das respectivas barras deve ser de 4 para 1.

Nos dois casos, ambos os gráficos podem representar a *Frequência Absoluta* de cada uma das categorias, que corresponde ao número total de indivíduos da amostra em cada categoria. O mesmo perfil de gráfico é obtido se considerarmos a *Frequência Relativa* de cada categoria, que corresponde à proporção (número entre 0 e 1) de indivíduos na amostra em cada categoria.

A proporção de indivíduos em mais de uma categoria é obtida somando-se as proporções das categorias de interesse. Por exemplo, a proporção de indivíduos com número de calçado maior que 38 é igual à soma das proporções das categorias maiores que 38.

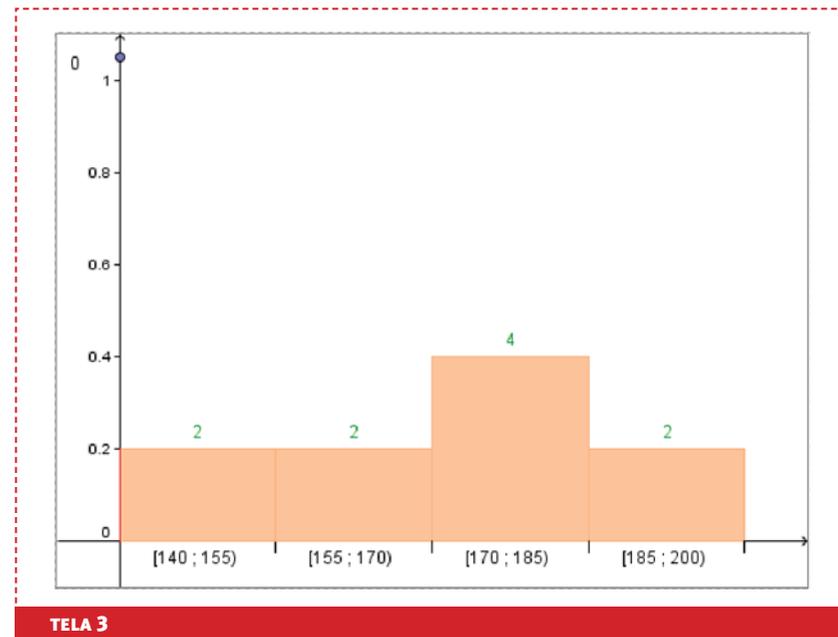
3 Variável quantitativa contínua

ATIVIDADE

Nesta atividade, discutimos representação gráfica e algumas medidas numéricas para uma variável numérica contínua, a Altura de uma pessoa. Uma variável contínua pode, em princípio, assumir qualquer valor real dentro de um intervalo. Na prática, os instrumentos de medição utilizados têm seus limitantes para a precisão da mensuração feita.

Poderíamos tratar a altura como uma variável discreta e definir categorias de 1 em 1 cm. Lembrando que o objetivo da representação gráfica é entregar informação visual clara a respeito das frequências da variável de interesse, percebemos rapidamente que esta não é uma boa estratégia.

Definimos então *Intervalos de Classe*, que correspondem a intervalos disjuntos dentro do conjunto de possíveis valores da variável estudada. A representação gráfica neste caso é o *Histograma*, que entrega a frequência (relativa ou absoluta) de cada intervalo de classe.



Mais precisamente, a cada intervalo de classe associamos um retângulo cuja base é o intervalo, de modo que a área relativa à área total do histograma corresponde à proporção de indivíduos na amostra com valores compreendidos pelo intervalo.

Dessa forma, a frequência de uma dada região é entregue pela proporção da área do histograma nessa região em relação à área total.

Podemos construir um histograma de acordo com uma quantidade fixa de intervalos, o que fornecerá intervalos de classe de mesma amplitude entre os valores mínimo e máximo observados na amostra. Também podemos construí-lo fixando os extremos de cada intervalo, o que permite intervalos de classe de amplitudes diferentes.

Na PARTE 2 desta atividade, o aluno construirá um histograma fixando o número total de intervalos de classe.

Nas PARTES 3 e 4, utilizamos o histograma para determinar algumas medidas numéricas que resumam a informação entregue pela amostra.

Uma medida-resumo bastante utilizada é o *p-quantil*, onde *p* é uma proporção entre 0 e 1. O *p-quantil* de uma amostra é um valor real tal que

1. pelo menos $p \cdot 100\%$ da amostra têm valores menores ou iguais a ele, e
2. pelo menos $(1 - p) \cdot 100\%$ da amostra têm valores maiores ou iguais a ele.

Lembrando que a área do histograma representa proporções amostrais, podemos obter esses quantis a partir da área associada a valores menores ou iguais a um certo valor.

Um quantil importante é a *Mediana*, que representa o quantil 0.5 e “divide” a área do histograma na metade.

Fechamento

O foco deste software é o uso de gráficos para tratar as informações coletadas em uma amostra. Discuta com os alunos a utilidade destas ferramentas, abordando as soluções obtidas por cada grupo.

Na aula seguinte, os alunos podem trazer gráficos de jornais ou revistas para discutir em sala a clareza das informações entregues.

Agora, vamos discutir as questões para o cadernos propostas no software.

Questão 1 para o caderno

Considerando esse conjunto de dados responda: qual é a melhor forma de representação para responder cada uma das questões, gráfico de barras ou de setores? Por quê?

Basicamente, a escolha do gráfico está vinculada ao tipo de variável discreta: ordinal ou nominal, ou seja, trata-se de considerar se as possíveis respostas da variável de interesse são categorias com uma ordem natural ou não.

Como discutido acima, o gráfico de barras permite uma visualização imediata da ordem das categorias, o que não ocorre com o gráfico de setores.

Um segundo critério a ser observado é a quantidade de categorias possíveis. Por exemplo, uma variável pode ter como possíveis respostas 9 categorias não ordenadas. Neste caso, o gráfico de setores pode não permitir uma clara visualização das frequências de cada categoria, sendo mais adequado, então, um gráfico de barras.

Em linhas gerais, o objetivo da representação gráfica é entregar informação visual clara a respeito das frequências da variável de interesse. Quanto mais clara for essa informação, melhor é o gráfico.

O software “Gráficos de barras e setores” pode ser útil na realização de uma atividade que explore essas ideias. Ele permite ao usuário plotar esses dois tipos de gráfico a partir de um conjunto de dados qualquer.

Na segunda parte da ATIVIDADE 3, definimos o histograma a partir de um número fixo de intervalos.

Depois da utilização deste software, no laboratório de informática ou mesmo na sala de aula, o professor pode propor aos alunos (ou grupos de alunos) que definam números diferentes de intervalos; por exemplo, cada grupo escolhe ou 2 ou 5 ou 10 intervalos. O que se percebe ao modificar o número de intervalos? Como é a qualidade da informação para 2, 5 ou 10 intervalos?

Novamente, a discussão pode se basear no objetivo da representação gráfica: entregar informação visual clara a respeito das frequências da variável de interesse. Quanto mais clara for essa informação, melhor é o gráfico.

Tipicamente, poucos intervalos podem esconder informação relevante. Por outro lado, excesso de intervalos pode “sujar” a interpretação pelo excesso de informação pouco relevante.

O software “Histogramas e quantis” pode ser útil na realização de uma atividade que explore essas ideias. Ele permite ao usuário plotar os dois tipos de gráfico a partir de um conjunto de dados qualquer, variando o número de intervalos e visualizando dinamicamente os quantis da amostra.

Existem outros dois softwares que utilizam o mesmo conjunto de dados utilizados neste: “Medidas do corpo – Boxplot” e “Medidas do corpo – Gráficos de Dispersão”. Ambos são boas alternativas para dar continuidade às atividades desenvolvidas a partir deste.

Bibliografia

COSTA NETO, Pedro Luiz de Oliveira. **Estatística**. Editora Edgard Blücher, 2002.

MEYER, Paul. **Probabilidade: Aplicações à Estatística**. Livros Técnicos e Científicos Editora, 2003.

Ficha técnica



AUTOR

Laura Leticia Ramos Rifo

REVISORES

Língua Portuguesa

Ana Cecília Agua de Melo

PROJETO GRÁFICO

Preface Design

ILUSTRADOR

Lucas Ogasawara



UNIVERSIDADE ESTADUAL DE CAMPINAS

Reitor

Fernando Ferreira Costa

Vice-Reitor

Edgar Salvadori de Decca

Pró-Reitor de Pós-Graduação

Euclides de Mesquita Neto

MATEMÁTICA MULTIMÍDIA

Coordenador Geral

Samuel Rocha de Oliveira

Coordenador de Software

Leonardo Barichello

Coordenador de Implementação

Matias Costa

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA (IMECC – UNICAMP)

Diretor

Jayme Vaz Jr.

Vice-Diretor

Edmundo Capelas de Oliveira

LICENÇA Esta obra está licenciada sob uma licença Creative Commons