



Matemática Multimídia

NÚMEROS
E FUNÇÕES



GUIA DO PROFESSOR



Software

Medidas do corpo – Box plot

Objetivos da unidade

1. Analisar representação gráfica de dados estatísticos;
2. Familiarizar o aluno com gráfico de Box Plot e análise estatística bivariada;
3. Utilizar 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 – box plot

GUIA DO PROFESSOR

Sinopse

Neste software, o aluno irá estudar um exemplo de estatística descritiva utilizando medidas de três variáveis: gênero, altura e número do calçado. Vamos explorar a análise da relação entre duas variáveis, uma qualitativa e outra quantitativa, através do chamado gráfico de Box Plot.

Conteúdos

- Estatística, Interpretação de Gráficos e Dados;
- Gráficos bivariados, Box Plot;
- Quantis de uma variável quantitativa.

Objetivos

1. Analisar representação gráfica de dados estatísticos;
2. Familiarizar o aluno com gráfico de Box Plot e análise estatística bivariada;
3. Utilizar o conceito de quantil.

Duração

Uma aula dupla, com possibilidade de extensão por mais aulas caso o professor deseje resolver todos os problemas propostos.

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 – Gráficos Univariados, Medidas do Corpo – Gráficos de Dispersão;
- Experimento: Variáveis Antropométricas;
- Áudio: História da Estatística.



Introdução

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

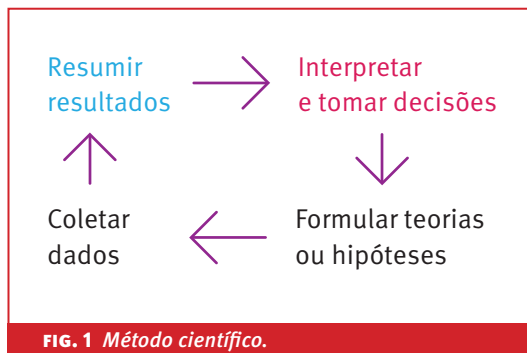
Basicamente, podemos identificar numa pesquisa científica os seguintes estágios:

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 uma possível representação gráfica para a análise da relação entre duas variáveis, uma qualitativa e outra quantitativa, como descrito brevemente na Introdução do software e na introdução da ATIVIDADE 1, PARTE 2.

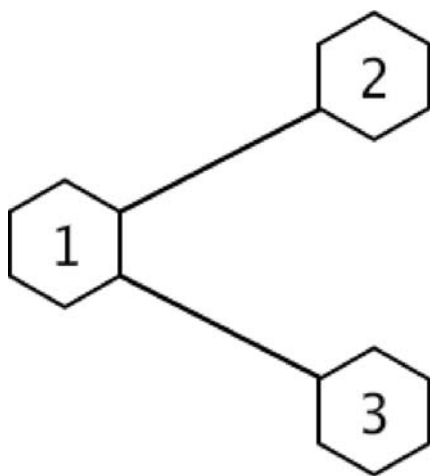
Na ATIVIDADE 1, o aluno deverá fornecer os dados que serão utilizados ao longo do software. Caso outro software da mesma sequência já tenha sido utilizado no mesmo computador, os dados serão carregados automaticamente. O experimento Variáveis Antropométricas também utiliza um conjunto de dados semelhante.

Na ATIVIDADE 2, o aluno deverá construir o gráfico de Box Plot para a variável “altura” para as duas categorias, “meninas” e “meninos”, com os dados coletados na classe, comparando os valores obtidos em ambas as categorias.

Na ATIVIDADE 3, realizamos o mesmo tipo de análise para a variável “número do calçado”, comparando os valores obtidos para as categorias da variável “gênero”.

Como as ATIVIDADES 2 e 3 possuem a mesma estrutura, sugerimos ao professor que defina previamente quais atividades devem ser resolvidas pelos alunos e em que ordem.

Na construção dos gráficos, os alunos deverão definir e compreender as medidas de posição: máximos e mínimos amostrais, quartis e mediana de uma amostra.



TELA 1 Mapa do software.

1 Tabela de dados

ATIVIDADE

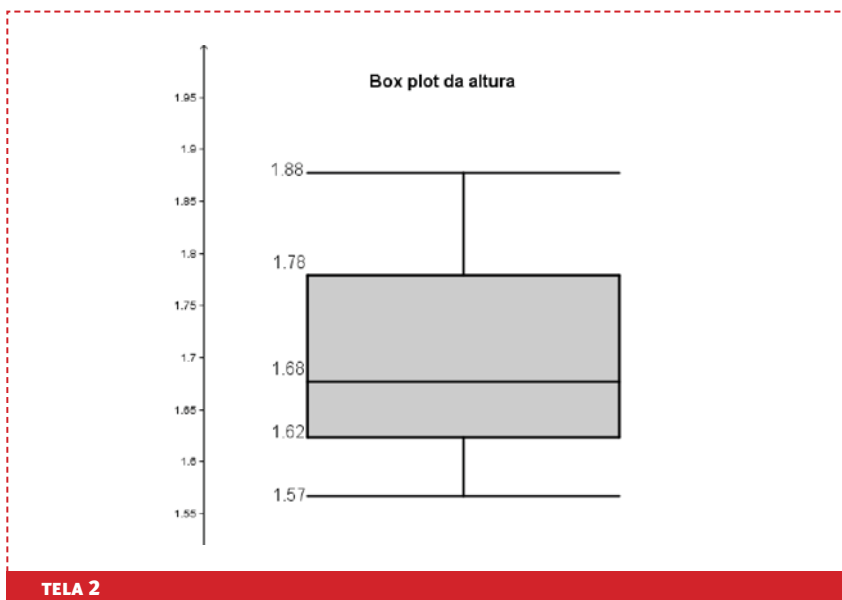
Começamos esta atividade registrando os dados referentes às variáveis “gênero”, “altura” e “número de calçado”, obtidos na classe. O aluno deve preencher a tabela com os dados obtidos.

A variável qualitativa “gênero” tem como possível resposta duas categorias não ordenadas: “F”, feminino, e “M”, masculino. A variável quantitativa “altura” deve ser registrada em cm, e a variável quantitativa “número de calçado” será um número inteiro não negativo.

Nesta atividade, o aluno também aprenderá a construir o gráfico de Box Plot, entendendo visualmente os conceitos de mediana e quartis.

No gráfico de Box Plot, a variável analisada é representada no eixo vertical, e o gráfico é formado por um retângulo, como na TELA 2.





A base do retângulo representa o primeiro quartil ou 0,25-quantil. A parte superior do retângulo representa o terceiro quartil ou 0,75-quantil. A linha dividindo o retângulo representa a mediana, ou segundo quartil ou 0,5-quantil.

Estes valores são medidas de posição da variável analisada e entregam informação sobre a localização dos valores observados na reta real.

Dado p em $(0, 1)$, definimos o p -quantil como o valor real q_p tal que pelo menos $p \times 100\%$ da amostra sejam valores menores ou iguais a q_p , e pelo menos $(1 - p) \times 100\%$ da amostra sejam valores maiores ou iguais a q_p .

Por exemplo, suponhamos que os valores observados na amostra, já ordenados, sejam

1, 2, 2, 2, 2, 2, 3, 4, 4, 4

Então, a mediana ou o 0,5-quantil corresponde ao valor 2, já que 60% da amostra tem valores menores ou iguais a 2, e 90% da amostra tem valores maiores ou iguais a 2. Assim, o valor 2, e nenhum outro valor real, satisfaz a definição de mediana.

Observe que, com esta definição, pode existir mais de um p-quantil. Por exemplo, suponhamos que os valores observados na amostra, já ordenados, sejam

1, 2, 2, 2, 2, 3, 3, 4, 4, 4

Então, a mediana ou o 0,5-quantil corresponde a qualquer valor entre 2 e 3, já que 50% da amostra tem valores menores ou iguais a 2, e 50% da amostra tem valores maiores ou iguais a 2, satisfazendo a definição. Do mesmo modo, 70% da amostra tem valores menores ou iguais a 3 e 50% da amostra tem valores maiores ou iguais a 3. Por fim, para qualquer valor estritamente entre 2 e 3, temos que 50% dos valores amostrados são menores ou iguais a ele, e 50% da amostra tem valores maiores ou iguais a ele.

A altura do retângulo, igual à diferença entre o terceiro e o primeiro quartil, é chamada distância interquartil e será denotada por d .

Os segmentos verticais acima e abaixo do retângulo se estendem até os valores extremos observados na amostra, que distam menos que $1,5d$ (por convenção, apenas) dos quartis. Assim, se a diferença entre o primeiro quartil e o menor valor da amostra for menor que $1,5d$, então o menor valor será representado pela base do segmento inferior. Se não, a base do segmento inferior será o menor valor da amostra a uma distância de até $1,5d$ do primeiro quartil. Os valores a uma distância maior serão representados por asteriscos e denominados valores atípicos, já que se encontram isolados dos valores mais centrais da amostra.

A mesma definição se aplica ao segmento superior para valores acima do terceiro quartil.

A distância entre os valores máximo e mínimo observados é chamada amplitude da amostra.

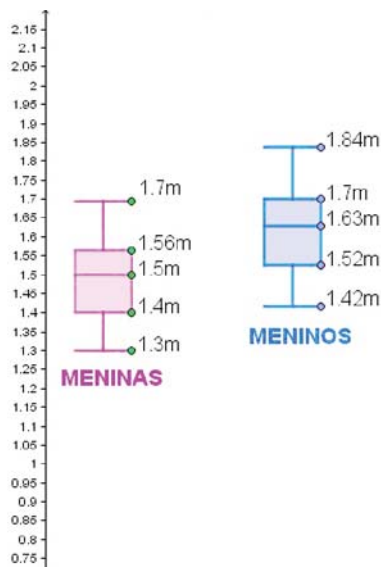


2 Box Plot para “altura”, por categorias

Nesta atividade, o aluno deverá construir o gráfico Box Plot para a variável “altura” para cada um dos subgrupos da amostra formados pelas categorias da variável “gênero”: meninas e meninos.

Neste exemplo, a variável quantitativa é uma variável contínua, já que, em princípio, pode assumir qualquer valor dentro de um intervalo da reta, dependendo apenas da precisão do instrumento de medição. Ou seja, em um mundo ideal, com um instrumento de medição perfeito, qualquer valor real dentro de um intervalo poderia ser observado.

Tipicamente, para uma variável contínua representando uma medição biométrica dentro de um grupo homogêneo, a mediana assume um único valor.



TELA 3

Podemos perceber que esta representação gráfica nos permite explorar rapidamente as diferenças entre ambas as categorias.

As comparações que devem ser feitas em uma primeira análise descritiva são basicamente as referentes a:

1. mediana
2. quartil
3. valor mínimo
4. valor máximo
5. distância interquartil
6. amplitude

para os valores amostrados em cada uma das categorias.

As perguntas da atividade permitem que o aluno adquira sensibilidade para este tipo de análise gráfica.

As comparações 1 a 4 dizem respeito a medidas de posição da variável estudada, indicando a região da reta real em que os valores amostrados se encontram.

As comparações 5 e 6 dizem respeito à variabilidade da variável estudada dentro de cada categoria. Menor variabilidade indica comportamento mais homogêneo da variável estudada dentro da categoria.

3 Box Plot para “número de calçado”, por categorias

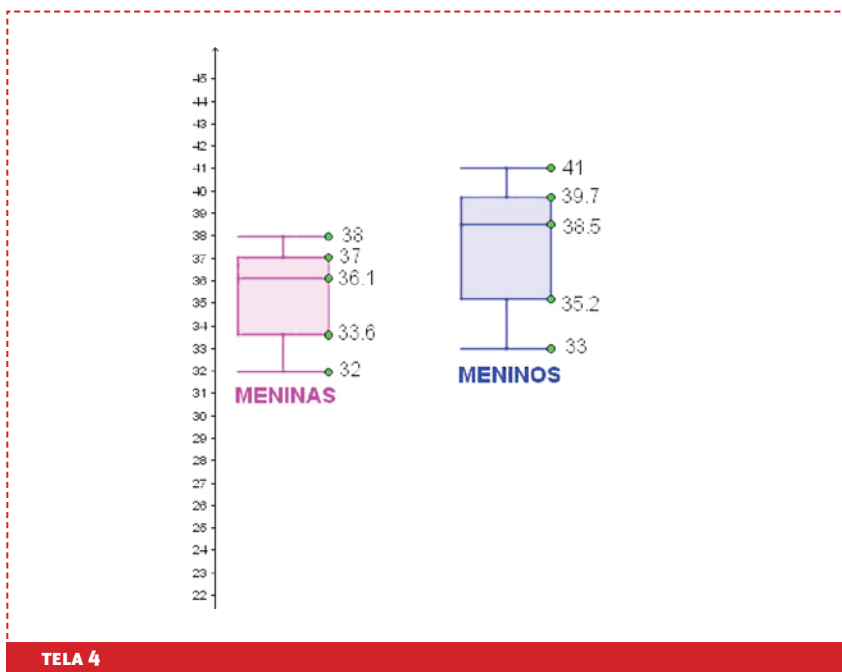
ATIVIDADE

Nesta atividade, discutimos a relação entre a variável qualitativa “gênero” e a quantitativa “número de calçado”, realizando uma análise gráfica similar à feita na ATIVIDADE 2.

Neste exemplo, a variável quantitativa é discreta, ou seja, o conjunto de possíveis valores é um conjunto discreto, já que é um subconjunto dos números naturais.

Tipicamente, para uma variável deste tipo, a mediana será um intervalo de possíveis valores. Podemos então adotar como mediana o ponto médio deste intervalo.





TELA 4

Assim como no exemplo anterior, o objetivo desta atividade é explorar as diferenças entre as categorias meninas e meninos para a variável quantitativa estudada.

Novamente, as perguntas da atividade permitem que o aluno adquira sensibilidade para este tipo de análise gráfica.

Fechamento

O foco deste software é o uso de gráficos para tratar as informações coletadas em uma amostra, em particular o uso de gráficos para comparação de variáveis quantitativas em diferentes categorias.

Esta análise pode ser estendida para mais de duas categorias. Por exemplo, poderíamos comparar a variável altura por “gênero” e por turmas diferentes de uma mesma série (A, B, C, D), ou por períodos diferentes de uma mesma série (matutino, vespertino e noturno).

Discuta com os alunos a utilidade destas ferramentas, abordando as soluções obtidas por cada grupo.

Os alunos podem trazer para a aula seguinte um conjunto de dados com informação sobre variáveis qualitativas e quantitativas.

Um ponto de partida para o fechamento desta atividade são as questões para o caderno propostas no final das ATIVIDADES 2 e 3:

Questão 1 para o caderno

Considerando esse conjunto de dados, analise graficamente as diferenças das variáveis quantitativas para as diversas categorias amostradas.

Como discutido acima, o gráfico de Box Plot permite uma visualização imediata das diferenças entre categorias.

Outra análise gráfica possível, mais complexa, é a comparação entre histogramas da variável quantitativa em cada uma das categorias. Mas isto foge ao escopo desta atividade.

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

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



O software “Histogramas e Quantis”, por sua vez, pode ser útil na realização de uma atividade que explore variáveis quantitativas.

Existem outros dois softwares, semelhantes ao discutido neste Guia do professor, que utilizam o mesmo conjunto de dados: Medidas do Corpo – Gráficos Univariados e Medidas do Corpo – Gráficos de Dispersão. Ambos são boas alternativas para complementar as 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

E ILUSTRAÇÕES TÉCNICAS

Preface Design



**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 



FUNDO NACIONAL
DE DESENVOLVIMENTO
DA EDUCAÇÃO

Secretaria de
Educação a Distância

Ministério da
Ciência e Tecnologia

Ministério
da Educação

