

Sistema Computacional para Medidas de Posição - FATEST

Deise Deolindo Silva, Mauricio Duarte, Renata Ueno Sales,
Guilherme Maia da Silva

Faculdade de Tecnologia de Garça – FATEC

deisedeolindo@hotmail.com, maur.duarte@gmail.com,
renataueno@hotmail.com, guilhermemaiasilva@hotmail.com

Abstract.

Resumo. Estatística é um ramo do conhecimento humano utilizado desde a antiguidade e, seu uso intensificou-se nos dias atuais. O que se pode perceber é que as tecnologias computacionais evoluem rapidamente no que se refere a armazenamento e capacidades computacionais que acabam por exigir da estatística, softwares modernos e específicos. O objetivo deste projeto é desenvolver um sistema computacional que seja capaz de auxiliar os cálculos estatísticos mais comumente usados. O projeto se justifica pela grande dificuldade encontrada no ensino de estatística e também como motivação para este aprendizado. O sistema foi implementado em linguagem C e fez uso de estruturas de dados dinâmicas. Estruturas de dados são usadas para a organização e controle dos dados e/ou informações na memória do computador.

1. Introdução

Estatística é um ramo do conhecimento humano utilizado desde a antiguidade e, seu uso intensificou-se nos dias atuais. Esta ciência evolui constantemente e, muitas técnicas tiveram melhor aprimoramento após a década de 90, com o avanço computacional.

O vocábulo – *estatística* – tem a raiz da palavra latina *Status* (Estado), em virtude dos recenseamentos realizados nas antigas civilizações. Existem algumas evidências que as primeiras técnicas foram utilizadas 3.000 anos antes de Cristo, por interesse do Estado que decretava o censo populacional, com o objetivo de registrar os indivíduos, inventariar os bens dessas pessoas para determinar o valor dos impostos a serem cobrados ou para o alistamento militar. (BEARZOTI e BUENO FILHO, p. 7, 2000).

Atualmente, estatística é definida de diversas formas. Lopes (2010) diz que pode ser considerada uma ciência, um método ou uma arte. Ciência quando, utiliza suas próprias teorias a conjuntos de dados, independentemente da natureza, sendo autônoma e universal. É um método quando utilizada como instrumento para outra ciência. E, considerada arte quando é aplicada visando à construção de modelos para representar a realidade.

Magalhães e Lima (p. 1, 2002) definem estatística como a ciência que investiga os processos para a coleta, organização, representação, análise e interpretação de dados, com o objetivo de extrair informações sobre uma população.

Atualmente, as mídias, órgãos públicos ou particulares, instituições escolares e diversas áreas do conhecimento humano utilizam métodos estatísticos cuja estrutura envolve ciência, lógica e tecnologia, estas metodologias auxiliam nas interpretações de dados, previsões ou em tomadas de decisões.

Com a evolução no campo da tecnológica da informação, principalmente o aprimoramento de *softwares*, *hardwares* e o aumento significativo da capacidade de produzir, armazenar e transmitir dados, associados ao crescimento acentuado da demanda por informações em um mundo globalizado, vem exigindo da Estatística avanços paralelos no desenvolvimento de metodologias e novos indicadores cada vez mais complexos que exigem equipamentos modernos, *softwares* estatísticos e técnicos capacitados. (IGNACIO, p. 8, 2010).

Diante deste impacto foi desencadeado semelhante impacto sobre a forma de como a disciplina Estatística está sendo ensinada aos estudantes do curso de Tecnologia em Análise e Desenvolvimento de Sistemas da Fatec de Garça.

O Projeto Pedagógico deste curso afirma que o tecnólogo em Análise e Desenvolvimento de Sistemas deve analisar e manter sistemas computacionais de informação. Além disso, esse profissional trabalha com ferramentas computacionais e equipamentos de informática. Ele deve ter o raciocínio lógico e o emprego de linguagens de programação, deve estar sempre atento às oportunidades que o mercado oferece aproveitando o surgimento de novas tecnologias e estar voltado às tecnologias que surgem quase que diariamente na área de sistemas de informação, procurando soluções adequadas e compatíveis entre as mesmas evitando, por meio de seus projetos, o desperdício de tempo e de recursos financeiros.

O projeto se justifica pela grande dificuldade apresentada pelos alunos no aprendizado dos cálculos essenciais da Estatística e, com o uso de um sistema computacional espera-se diminuir este impacto e dificuldades relacionadas.

A disciplina Estatística Aplicada é ministrada no terceiro termo do curso, em um total de 80 horas, em que o objetivo é conhecer e aplicar conhecimentos de Estatística e desenvolver aplicativos para essa área. Como procedimento metodológico, as professoras desta disciplina, ao planejar os conteúdos específicos, levam os conceitos aos alunos por meio da exposição dos mesmos e resolução de exercícios. Além disso, utilizando os conhecimentos de outras disciplinas do curso (no caso, Estrutura de Dados), fazem com que os alunos programem algum conteúdo trabalhado em Estatística no curso.

O conteúdo trabalhado e que foi elaborado pela turma do 1º semestre de 2011 foi “Medidas de Tendência Central”: após trabalhar a parte conceitual, eles foram levados ao laboratório de informática para programarem a média aritmética, a mediana e a moda.

2. Medidas Resumo

Uma área importante estudada em diversos os cursos é a estatística descritiva. Utilizada na etapa inicial da análise dados, com o objetivo de tirar conclusões de modo informal e direto. Pode ser definida como um conjunto de técnicas destinadas a descrever e resumir os dados a fim de que possamos tirar conclusões a respeito das características de interesse. (MAGALHÃES e LIMA, p. 2, 2002).

2.1 Medidas de Posição ou de Tendência Central

Serão apresentadas as definições de *medidas de tendência central* para um conjunto de dados qualquer. As principais medidas de posição destacadas serão *média*, *mediana* e *moda*.

2.1.1 Média

Magalhães e Lima (p. 94, 2002) apresentam a seguinte definição para a medida aritmética. Considere uma amostra aleatória da variável X , com seus respectivos valores $x_1, x_2, \dots, x_n$, a média aritmética desse conjunto é dada pela soma dos valores dividida pelo número total de observação. Geralmente, denotada por $\bar{x}$.

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

Se os dados estiverem agrupados em tabelas de frequências, a média é obtida ponderando os diferentes valores x_i , pelas suas respectivas frequências relativas. Assim:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_n x_n}{n} = \frac{\sum_{i=1}^n n_i x_i}{n} = \sum_{i=1}^n \frac{n_i}{n} x_i.$$

2.1.2 Mediana

A mediana é o valor que ocupa a posição central dos dados ordenados. Esta pode ser obtida para dados *não agrupados* e para valores *agrupados* em distribuições de frequências.

Magalhães e Lima, (p. 95, 2002) apresenta o procedimento para dados não agrupados. Considere os valores referentes ao número de salários mínimos que 9 funcionários de uma empresa recebe mensalmente: 5, 13, 10, 2, 15, 15, 6, 16, 9. O primeiro passo é ordenar os valores: 2, 5, 6, 9, 10, 13, 15, 15, 16. Em seguida, toma-se aquele valor central que divide a série em duas partes iguais. No exemplo, o valor é 10, já que nessa série, há quatro elementos acima e quatro abaixo de 10. Tem-se, então que a *md* = 10. Note que há um número ímpar de dados e então a mediana é o termo de ordem $(n + 1)/2 = (9 + 1)/2 = 10/2 = 5$, ou seja, o quinto elemento.

Se o número de dados fosse par a *md* será o ponto médio dos dois termos centrais: 2, 6, 7, 10, 12, 13, 15, 16, isto é a média entre o termo de ordem $n/2$ e o termo de ordem $(n/2) + 1$. Então a *md* será a média aritmética entre 10 e 12, ou seja, o valor 11.

Se os dados estiverem em uma distribuição de frequência, o cálculo da mediana se processa de modo muito semelhante àquele dos dados não agrupados, implicando, porém a determinação prévia das frequências acumuladas. Ainda aqui, temos que determinar um valor tal que divida a distribuição em dois grupos que contenham o mesmo número de elementos. Quando a distribuição está subdividida em intervalos de classes, o problema consiste em determinar o ponto do intervalo em que está compreendida a mediana.

Para calculá-la, deve seguir os seguintes passos: 1) Determinar as frequências acumuladas da série; 2) Calcular $n/2$ (50% dos elementos); 3) Marcar a classe correspondente à frequência acumulada (N_i) imediatamente a classe da mediana $n/2$; 4)

Aplicar esta fórmula para obter a mediana: $md = l_i + \frac{[n/2 - N_i(ant)] \times h}{n_i}$. Em que:

l_i : Limite inferior da classe da mediana;

$N_i(ant)$: Frequência Acumulada da classe anterior à classe da mediana;

n_i : Frequência simples da classe mediana;

h : Amplitude do intervalo de classe.

2.1.3 Moda

Moda é o número que ocorre com maior frequência em uma série de valores, ou seja, o valor mais frequente e será denotada por *mo*.

A moda é facilmente reconhecida, bastando para isso procurar o valor que mais se repete. Para dados agrupados o procedimento é encontrar o ponto médio da classe que apresenta a maior frequência, denominada Classe Modal. Ou pode-se obtê-la pela seguinte

fórmula: $mo = \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] \times h$. Em que,

l_i : Limite inferior da classe modal;

$\Delta_1 = n_i - n_i(ant)$: Frequência modal menos a frequência anterior à classe modal;

$\Delta_2 = n_i - n_i(post)$: Frequência modal menos a frequência posterior à classe modal;

h : Amplitude da classe modal.

3. Listas Lineares

Segundo Nicklaus Wirth (1999), um programa pode ser definido como a junção de algoritmos com estruturas de dados. Entende-se por algoritmos como a sequência de passos, logicamente definidos, que visam a solução e/ou execução de uma determinada tarefa e, pode-se dizer que estruturas de dados são formas para armazenar dados na memória do computador a fim de permitir controle e acesso eficientes a estes dados.

São muitas as estruturas de dados definidas: pilhas, filas, listas lineares, árvores binárias, árvores B e B*, grafos, etc. Esta seção dará ênfase especial ao estudo das listas lineares usada no projeto aqui apresentado.

Uma lista linear é um conjunto de n elementos (com $n > 0$) : $x_1, x_2, x_3, \dots, x_n$, cuja propriedade estrutural define forma de armazenamento e política de acesso. Em uma lista linear pode-se realizar várias operações: acesso a elementos, inserções e remoções de elementos, tamanho da lista, etc. Basicamente as listas lineares podem ser representadas em duas formas: listas por contiguidade e listas por encadeamento dos nós e, não existe uma forma geral em que todas as operações implementadas sejam eficientes.

Para uma implementação por contiguidade, tem-se que os elementos da lista são armazenados em um *array* (vetor), tendo os elementos dispostos sequencialmente na memória do computador. Assim, conhecendo-se o endereço do primeiro elemento, tem-se acesso direto aos demais. Para esta representação, a operação de acesso aos elementos é muito eficiente. Porém, têm-se problemas com a predefinição do tamanho do vetor e, caso ele seja desconhecido, poderá ter problemas excesso de memória (sobra de elementos alocados) ou de *overflow* (falta de memória para os elementos desejados).

Em uma implementação por encadeamento dos nós, a lista cresce dinamicamente, isto é, cresce à medida que necessita da inserção de novos elementos. Assim, não se têm os problemas de excesso ou falta de memória característica das listas por contiguidade. Outra vantagem está no fato de não precisar fazer deslocamentos de elementos para manter a lista ordenada. Porém, o acesso aos elementos é prejudicado, pois cada elemento possui além da informação o endereço do elemento seguinte na estrutura. Assim, tem-se a referência ao primeiro elemento e, esta por sua vez a referência do segundo que, tem a referência do terceiro e, assim sucessivamente, o último elemento tem seu próximo NULL, sinalizando que ele não existe. Cada elemento da lista é comumente chamado de nó e, neste caso, cada nó tem dois campos (atributos) a informação e o endereço do próximo nó da lista.

4 - Resultados

O resultado deste trabalho interdisciplinar foi muito satisfatório e pode-se perceber a importância da integração das disciplinas para o aprendizado dos alunos. Porém, em um trabalho interdisciplinar é necessário determinar o valor de cada disciplina, suas estruturas e a intencionalidade de seu papel no currículo para que esses fundamentos possibilitem entender que a interdisciplinaridade é muito mais que uma simples integração de conteúdos.

A interdisciplinaridade não dilui as disciplinas, ao contrário, mantém sua individualidade. Mas integra as disciplinas a partir da compreensão das múltiplas causas ou fatores que intervêm sobre a realidade e trabalha todas as linguagens necessárias para a constituição de conhecimentos, comunicação e negociação de significados e registro sistemático dos resultados. BRASIL (1999, p. 89).

Durante a implementação do software houve algumas dificuldades quanto ao uso das estruturas escolhidas, porém, todas resolvidas em tempo. O sistema foi utilizado por vários alunos que puderam realizar seus testes e interagir com suas funcionalidades. Pode-se perceber durante a aula prática uma maior interação e interesse em aprender os

conteúdos abordados, até então, de forma extremamente teórica. Novos projetos serão realizados neste âmbito.

5. Referências

BRASIL. MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO. SECRETARIA DE EDUCAÇÃO FUNDAMENTAL, **Parâmetros Curriculares Nacionais: terceiro e quarto ciclos - apresentação dos temas transversais**. Brasília: MEC/SEF, 1999.

BEARZOTI, E. e BUENO FILHO, J. S. S. **Introdução à inferência estatística**. Lavras: UFLA/FAEPE, 2000.

IGNACIO, S. A. **Importância da Estatística para o Processo de Conhecimento e Tomada de Decisão**. Curitiba: Iparde, 2010.

LOPES, P. A. **Entendendo a importância da estatística sem ser gênio, matemático ou bruxo**. Disponível em: <<http://www.administradores.com.br/informe-se/artigos/entendendo-a-importancia-da-estatistica-sem-ser-genio-matematico-ou-bruxo/11591/>>. Acesso em: 25 mar. 2012.

MAGALHÃES, M.N. e LIMA, A. C. P. **Noções de Probabilidade e Estatística**. 4 ed. São Paulo: Edusp, 2002.

WIRTH, N. **Algoritmos e Estruturas de Dados**. LTC. 1999.