

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ALINE BRUM LORETO

**Análise da Complexidade Computacional
de Problemas de Estatística Descritiva com
Entradas Intervalares**

Tese apresentada como requisito parcial
para a obtenção do grau de
Doutor em Ciência da Computação

Profa. Dra. Leila Ribeiro
Orientadora

Profa. Dra. Laira Vieira Toscani
Co-orientadora

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Loreto, Aline Brum

Análise da Complexidade Computacional de Problemas de Estatística Descritiva com Entradas Intervalares / Aline Brum Loreto. – Porto Alegre: PPGC da UFRGS, 2006.

93 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2006. Orientadora: Leila Ribeiro; Co-orientadora: Laira Vieira Toscani.

1. Medidas de tendências central. 2. Medidas de dispersão. 3. Separatrizes. 4. Complexidade computacional. 5. Extensão intervalar. 6. Problema NP-Difícil. I. Ribeiro, Leila. II. Toscani, Laira Vieira. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Pró-Reitor de Coordenação Acadêmica: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora de Pós-Graduação: Prof^a. Valquíria Linck Bassani

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Flávio Rech Wagner

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“Tenha sempre em mente que a pele enruga,
o cabelo embranquece e os dias convertem-se em anos...
Mas o que é importante não muda.
A sua força e convicção não tem idade
O seu espírito é como qualquer teia de aranha.
Atrás de cada linha de chegada há sempre uma de partida.
Atrás de cada conquista, vem um novo desafio.
Enquanto estiver vivo, sinta-se vivo...
Se sentir saudades do que fazia volte a fazer.
Não viva de fotografias amareladas...
Continue quando todos esperam que desista.
Não deixe que enferruje o ferro que existe em você.
Faça com que em vez de pena tenham respeito por você.
Quando não conseguir correr através dos anos, caminhe!
Quando não conseguir caminhar use uma bengala...
Mas nunca se deixe deter...”*

— MADRE TERESA DE CALCUTÁ

SUMÁRIO

LISTA DE SÍMBOLOS	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	9
RESUMO	11
ABSTRACT	12
1 INTRODUÇÃO	13
1.1 Problema dos valores de amostra em Estatística Descritiva	13
1.2 Aritmética Intervalar e Estatística Descritiva Intervalar	15
1.3 Complexidade dos problemas da Computação Intervalar	17
1.4 Objetivos do Trabalho	19
1.5 Estrutura da tese	20
1.6 Comentários Finais	21
2 COMPLEXIDADE DE PROBLEMAS INTERVALARES	22
2.1 Aritmética Intervalar	22
2.2 Complexidade Computacional	26
2.2.1 Algoritmos	26
2.2.2 Problema Computacional	28
2.2.3 Complexidade	29
2.2.4 Classes de Complexidade	30
2.3 Complexidade de Problemas Intervalares	31
2.3.1 O PBCI com restrições	33
2.4 Comentários Finais	37
3 ESTATÍSTICA DESCRITIVA INTERVALAR	38
3.1 Representação dos reais em intervalos	38
3.2 Relação de ordem em Intervalos	39
3.3 Nível Intervalar dos Indicadores Estatísticos	40
3.3.1 Média aritmética Intervalar	40
3.3.2 Mediana Intervalar	40
3.3.3 Moda Intervalar	41
3.3.4 Amplitude total Intervalar	41
3.3.5 Variância Intervalar	41
3.3.6 Desvio padrão Intervalar	41

3.3.7	Coefficiente de variação Intervalar	42
3.3.8	Covariância Intervalar	42
3.3.9	Coefficiente de correlação Intervalar	42
3.3.10	Separatrizes Intervalares	42
3.3.11	Estado da Arte dos problemas da Estatística Descritiva Intervalar	43
3.4	Comentários Finais	44
4	ANÁLISE DA COMPLEXIDADE DOS PROBLEMAS DE ESTATÍSTICA DESCRITIVA INTERVALAR	46
4.1	Redefinindo os Problemas	46
4.2	Complexidade dos Problemas de computar Indicadores Estatísticos Intervalares	47
4.2.1	Problema: Média Intervalar	47
4.2.2	Problema: Mediana Intervalar	48
4.2.3	Problema: Moda Intervalar	48
4.2.4	Problema: Separatrizes - Quartil, Decil e Percentil Intervalares	49
4.2.5	Problema: Amplitude total Intervalar	49
4.2.6	Problema: Variância Intervalar	49
4.2.7	Problema: Desvio Padrão Intervalar	50
4.2.8	Problema: Coeficiente de Variação Intervalar	50
4.2.9	Problema: Covariância Intervalar	51
4.2.10	Problema: Coeficiente de Correlação Intervalar	51
4.3	Resultados de Complexidade	52
4.4	Comentários Finais	53
5	VERIFICAÇÃO DO INTERVALO SOLUÇÃO	54
5.1	Qualidade da solução intervalar	54
5.2	Medida de correção de uma resposta intervalar	54
5.3	Estudo de Caso	56
5.4	Verificação da Qualidade de Aproximação	57
5.4.1	IMC dos alunos da 5a. série	57
5.4.2	IMC dos alunos da 6a. série	60
5.4.3	IMC dos alunos da 7a. série	62
5.4.4	IMC dos alunos da 8a. série	66
5.5	Comentários Finais	68
6	CONCLUSÃO	70
6.1	Resultados Obtidos	70
6.2	Prosseguimento do Trabalho	72
6.2.1	Ordenação de dados intervalares	72
6.2.2	Intervalos de confiança intervalar	72
	REFERÊNCIAS	74
	ANEXO A INDICADORES ESTATÍSTICOS	78
A.1	Média Aritmética	78
A.2	Mediana	78
A.3	Moda	78
A.4	Amplitude Total	78
A.5	Variância	78

A.6	Desvio padrão	79
A.7	Coefficiente de Variação	79
A.8	Covariância	79
A.9	Coefficiente de Correlação	79
A.10	Separatrizes	79
A.10.1	Quartis	79
A.10.2	Decis	79
A.10.3	Percentil ou Centil	80
ANEXO B ALGORITMOS E ANÁLISE DA COMPLEXIDADE		81
B.1	Algoritmo da Média Intervalar	81
B.2	Algoritmo Ordenação de Intervalos	82
B.3	Algoritmo da Moda Intervalar	83
B.4	Algoritmo Separatrizes Intervalares	83
B.5	Algoritmo Amplitude Intervalar	83
B.6	Algoritmo da Variância Intervalar	83
B.7	Algoritmo do Desvio Padrão Intervalar	84
B.8	Algoritmo do Coeficiente de Variação Intervalar	84
B.9	Algoritmo da Covariância Intervalar	84
B.10	Algoritmo do Coeficiente de Correlação Intervalar	84
ANEXO C ALTURA, PESO E IDADE DOS ALUNOS DA ESCOLA IMACULADA CONCEIÇÃO		89
C.1	Altura, peso e idade da 5a. série	89
C.2	Altura, peso e idade da 6a. série	89
C.3	Altura, peso e idade da 7a. série	89
C.4	Altura, peso e idade da 8a. série	89
ANEXO D INTERVALO DE CONFIANÇA		93

LISTA DE SÍMBOLOS

X	Variável aleatória
$+$	Operação de soma
$-$	Operação de subtração
$*$	Operação de multiplicação
$/$	Operação de divisão
$^$	Operação de exponenciação
$=$	Igual
\neq	Diferente
\subset	Inclusão estrita
\subseteq	Inclusão
Σ	Somatório
$<$	Menor
$>$	Maior
\leq	Menor ou igual
\geq	Maior ou igual
\in	Relação de pertinência
\forall	Quantificador universal
\exists	Quantificador existencial
$\mathbf{x} = [\underline{x}, \bar{x}]$	Intervalo cujos extremos são \underline{x} e \bar{x}
$\inf(\mathbf{x})$	Extremo inferior \underline{x} do intervalo \mathbf{x}
$\sup(\mathbf{x})$	Extremo superior \bar{x} do intervalo \mathbf{x}
\mathbb{Z}	Conjunto dos número inteiros
\mathbb{R}	Conjuntos dos números reais
\mathbb{IR}	Conjunto dos intervalos reais
$f(x)$	Função real
$F(\mathbf{x})$	Função intervalar
$Dom(f)$	Domínio da função f
$Cod(f)$	Contradomínio da função f

LISTA DE FIGURAS

Figura 3.1: Formulário dos Indicadores Estatísticos Intervalares	43
--	----

LISTA DE TABELAS

Tabela 2.1:	Tipos de Funções, Complexidade do PBCI com n Variáveis e com Polinômio de Ordem Fixa	35
Tabela 2.2:	Função Polinomial de Ordem Fixa e Complexidade do PBCI com Polinômio de Ordem Fixa	36
Tabela 2.3:	Tipos de Funções Polinomiais e Complexidade do PBCI com Coeficientes Restritos	36
Tabela 2.4:	Tipos de Funções Polinomiais com x_i arbitrários e Complexidade do PBCI com Polinômios Fixos e Intervalos Estreitos	36
Tabela 2.5:	Tipos de Funções Polinomiais com x_i fixos e Complexidade do PBCI com Polinômios Arbitrários e Intervalos Fixos	37
Tabela 4.1:	Computação Intervalar e Complexidade do Problema Variância Intervalar	50
Tabela 4.2:	Computação Intervalar e Complexidade do Problema Covariância Intervalar	51
Tabela 4.3:	Computação Intervalar e Complexidade do Problema Coeficiente de Correlação Intervalar	52
Tabela 4.4:	Indicadores Estatísticos Intervalares e Complexidade dos Problemas dos Indicadores Estatísticos Intervalares	52
Tabela 4.5:	Problemas da Estatística Descritiva Intervalar, complexidade dos Problemas com extensão intervalar e com imagem intervalar	53
Tabela 5.1:	Idade, IMC Meninas e IMC Meninos.	56
Tabela 5.2:	Indicadores Estatísticos, dados com IMC reais (meninas da 5a. série) e dados com IMC intervalares.	58
Tabela 5.3:	Indicadores Estatísticos, erros absolutos e erro relativo (meninas 5a. série)	58
Tabela 5.4:	Indicadores Estatísticos, dados com IMC reais (meninos da 5a. série) e dados com IMC intervalares.	59
Tabela 5.5:	Indicadores Estatísticos, erros absolutos e erro relativo (meninos 5a.série).	59
Tabela 5.6:	Indicadores Estatísticos, dados com IMC reais (meninas da 6a. série) e dados com IMC intervalares.	60
Tabela 5.7:	Indicadores Estatísticos, erros absolutos e erro relativo (meninas 6a. série).	61
Tabela 5.8:	Indicadores Estatísticos, dados com IMC reais (meninos da 6a. série) e dados com IMC intervalares.	62

Tabela 5.9: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 6a. série).	62
Tabela 5.10: Indicadores Estatísticos, dados com IMC reais (meninas da 7a. série) e dados com IMC intervalares.	63
Tabela 5.11: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 7a. série).	64
Tabela 5.12: Indicadores Estatísticos, dados com IMC reais (meninos da 7a. série) e dados com IMC intervalares.	65
Tabela 5.13: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 7a. série).	65
Tabela 5.14: Indicadores Estatísticos, dados com IMC reais (meninas da 8a. série) e dados com IMC intervalares.	66
Tabela 5.15: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 8a. série).	67
Tabela 5.16: Indicadores Estatísticos, dados com IMC reais (meninos da 8a. série) e dados com IMC intervalares.	68
Tabela 5.17: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 8a. série).	68

RESUMO

A Estatística é uma ferramenta indispensável em todos os campos científicos. A Estatística descritiva é usada para sintetizar dados. O principal problema desta área está relacionado aos valores de uma amostra, os quais geralmente possuem erros que ocorrem durante a obtenção dos dados. Um dos objetivos deste trabalho é apresentar uma forma de representação para os valores amostrais que considera os erros contidos nestes valores. Esta representação é realizada através de intervalos. A literatura mostra que foram realizadas pesquisas somente em problemas de calcular os valores intervalares das medidas de dispersão variância, covariância e coeficiente de correlação, que a utilização da computação intervalar na solução de problemas de medidas de dispersão intervalar sempre fornece solução com intervalos superestimados (intervalos com amplitude grande), e que ao procurar uma solução com intervalos de amplitude pequena (através da computação da imagem intervalar), o problema passa a pertencer a classe de problemas NP-Difícil. Com o objetivo principal de analisar a complexidade computacional dos problemas de computar os valores dos indicadores estatísticos descritivos com entradas intervalares, e realizar uma classificação quanto a classe de complexidade, a presente tese apresenta: i) definições intervalares de medidas de tendência central, medidas de dispersão e separatrizes; ii) investigação da complexidade de problemas das medidas de tendência central média, mediana e moda, das medidas de dispersão amplitude, variância, desvio padrão, coeficiente de variação, covariância, coeficiente de correlação e das separatrizes e iii) representação intervalar dos valores reais, de tal modo que garante a qualidade de aproximação nos intervalos solução calculado através da extensão intervalar. Primeiramente, apresentamos uma abordagem intervalar para os indicadores estatísticos e propomos algoritmos para a solução dos problemas de computar os intervalos de medidas de tendência central intervalar, dispersão intervalar e separatrizes intervalares. Tais algoritmos utilizam a aritmética intervalar definida por Moore, a extensão intervalar e foram projetados para serem executados em ambientes intervalares como IntLab e Maple Intervalar. Por meio da análise da complexidade computacional verificamos que os problemas de medidas de tendência central, dispersão e separatrizes, com entradas intervalares, pertencem à classe de problemas P. Este trabalho apresenta, portanto, algoritmos de tempo polinomial que calculam os intervalos dos indicadores estatísticos com entradas intervalares, e que retornam como solução intervalos com qualidade de aproximação. Os resultados obtidos no desenvolvimento do trabalho tornaram viável a computação da Estatística Descritiva Intervalar.

Palavras-chave: Medidas de tendências central, medidas de dispersão, separatrizes, complexidade computacional, extensão intervalar, problema NP-Difícil.

Analysis of the Computational Complexity of Descriptive Statistic Problems with intervals input

ABSTRACT

Statistics is a fundamental tool for all scientific areas. Descriptive Statistics are used to analyze data. The main problem in this area is related to sample values, which are typically error prone due to the experimental processes used to obtain these values. One of the aims of this thesis is to suggest a representation form to sample values that takes these errors into account by using intervals. There are already approaches to calculate values of dispersion, variance, covariance and correlation coefficient using intervals, but the interval computation of dispersion measures gives always overestimated intervals (with wide amplitude) as results. Moreover, if better solutions are needed, by using the interval image representation, calculating such measures turns out to be an NP-hard problem. The main aim of this thesis is to analyze the computational complexity of the descriptive statistics measures using intervals as input values. More concretely this work presents: i) definitions of central trend, dispersion and separatriz using intervals; ii) investigation of the complexity of problems related to the calculation of central trend measures mean, median and moda, of the dispersion measures amplitude, variance, standard deviation, variance coefficient, covariance, correlation coefficient and separatriz; and iii) interval representation of real values such that guarantee the quality of the approximation in the result intervals calculated using interval extension does not occur. First, an interval approach to statistics measures is presented, and then algorithms to compute the above described measures are proposed. Such algorithms are based on the interval arithmetics defined by Moore, on interval extension, and were designed to execute in ambients like IntLab and Interval Maple. By analyzing the computational complexity of these algorithms, we conclude that all defined measures are in the P class of complexity. This means that there are polynomial time algorithms that calculate the resulting intervals of the main descriptive statistical measures with interval input values with quality of the approximation. The impact of this result is to make Interval Descriptive Statistics feasible.

Keywords: measures of central trend, measure of dispersion, computational complexity, interval extension.

1 INTRODUÇÃO

Este trabalho está inserido na área de Matemática da Computação, tendo como tema a complexidade dos problemas de cálculo de indicadores estatísticos com entradas intervalares.

O trabalho é composto por 6 capítulos, sendo este primeiro destinado a apresentação do tema da tese e estrutura do volume. As próximas seções apresentam a motivação, justificativa e o estado da arte do tema proposto. Na seção seguinte é considerada a formulação da tese e os objetivos propostos. Para finalizar, as seções 1.5 e 1.6 apresentam a estrutura da tese e comentários finais.

1.1 Problema dos valores de amostra em Estatística Descritiva

A palavra estatística origina-se da palavra “Status” (Estado em latim), por ser utilizada no levantamento de dados, com a finalidade de orientar decisões governamentais. Considera-se a estatística como “a ciência que se preocupa com a organização, descrição, análise e interpretação de dados experimentais” ou “matemática aplicada a dados em observação” (LIPSCHUTZ, 1972). Segundo Meyer (1968) considera-se a estatística como “um conjunto de métodos e processos quantitativos que serve para estudar e medir os fenômenos coletivos” ou ainda como “um conjunto de métodos destinados à coleta, organização, resumo, apresentação e análise de dados e observações, bem como a tomada de decisões baseadas em tais análises”.

Popularmente o uso da estatística está ligado a representação de dados experimentais através de tabelas e gráficos (CHRISTMANN, 1978).

Christmann (1978), afirma que com o atual desenvolvimento, a Estatística se tornou uma ferramenta indispensável em todos os campos científicos. Na engenharia e na indústria, por exemplo, ela é indispensável no controle de qualidade, na análise de custos, no dimensionamento de estoques, no estudo de tempos e, de uma maneira geral, a estatística é pré-requisito em projetos de interesse coletivo, na análise de experiências e no estabelecimento de testes, em pesquisas científicas, no estabelecimento de fórmulas empíricas, etc.

Em função dos dados na análise estatística, a estatística se divide em descritiva e inferencial. A estatística descritiva é constituída de métodos usados para sintetizar os dados. A inferência estatística é o processo de utilizar os dados obtidos a partir de uma amostra para fazer estimativas ou testar hipóteses sobre as características de uma população. A estatística descritiva desenvolve o trabalho até a descrição dos dados; a análise destes fica para ser desenvolvida pela estatística inferencial (ANDERSON; SWEENEY; WILLIAMS, 2003).

A estatística descritiva envolve inicialmente a distribuição de frequências, medidas

de tendência central, medidas de dispersão e separatrizes. As medidas de localização ou de tendência central mostram o valor representativo em torno do qual os dados tendem a agrupar-se, com maior ou menor frequência. São usadas para sintetizar em um único número o conjunto de dados observados. As medidas de variabilidade ou de dispersão mostram o grau de afastamento dos valores observados em relação àquele valor representativo.

O principal interesse do presente trabalho, dentro da teoria da estatística descritiva, está relacionado ao problema dos valores $\{x_1, \dots, x_n\}$ de uma amostra. Geralmente os valores de uma amostra possuem erros que ocorrem durante a obtenção dos dados.

Jain (1991) comenta que utilizar quaisquer dados que aconteçam estar disponíveis ou utilizar dados que foram obtidos com pouco cuidado pode levar a uma informação pobre e enganosa. Assim, tomar as medidas necessárias para a obtenção de dados exatos pode assegurar informação confiável e valiosa para a tomada de decisão. A medição de uma grandeza deve ter como objetivo, entre outros, obter um valor com melhor estimativa do verdadeiro valor da grandeza e ter alguma indicação sobre a imprecisão da medida. Os erros de observação, normalmente, não obedecem a regras simples e podem ter várias causas. Geralmente, são enquadrados em dois grupos: erros acidentais e erros sistemáticos. Por vezes os dois tipos de erro ocorrem combinados, sendo difícil distinguir entre uns e outros.

Os erros sistemáticos podem ser devidos ao observador ou ao instrumento utilizado e são inerentes ao método escolhido, ao processo de medida ou ao próprio instrumento. Este tipo de erro afeta sempre o resultado no mesmo sentido e tem sempre a mesma magnitude sendo, por isso, muito difícil de detectar. Não existem princípios gerais para tratar este tipo de erro e a sua detecção e correção exige grande experiência da parte do observador (por exemplo, a utilização de uma régua mal graduada introduz um erro sistemático na medição) (VUOLO, 1996).

Os erros acidentais ou aleatórios podem ter diversas causas, mas são na sua grande maioria erros de leitura, têm uma incidência desordenada, influenciando os resultados ora num sentido ora noutro. Este tipo de erro é inevitável, pode no entanto ser detectado aumentando o número de experiências efetuadas, ou seja, diminuindo a probabilidade de ocorrência de um "mesmo acidente". São usualmente tratados de forma estatística (VUOLO, 1996).

Um erro sistemático não impede que se obtenha uma medida de grande precisão (os valores obtidos podem ser próximos entre si) no entanto pode conduzir a uma medida de baixo rigor (o valor obtido pode ser bastante diferente do valor real).

A precisão de cada uma das medidas efetuadas depende sempre das limitações técnicas do instrumento utilizado. Na maioria das vezes a precisão dos aparelhos é indicada, pelos fabricantes, no próprio aparelho ou no seu manual. O valor da precisão de uma medida nunca pode ser menor que a precisão do instrumento utilizado nessa medida. Por mais cuidadosa que seja uma medição e por mais preciso que seja o instrumento, não é possível realizar uma medida perfeita. Ou seja, sempre existe uma incerteza ao se comparar uma quantidade de uma dada grandeza física com sua unidade (VUOLO, 1996).

No desenvolvimento do presente trabalho a incerteza está relacionada ao instrumento de medida, o qual não é abordado neste trabalho. Supõe-se que as medidas foram realizadas nas mesmas condições e com o mesmo instrumento. A aleatoriedade está relacionada com a amostra de uma população, e os valores de uma amostra podem conter erros aleatórios ou sistemáticos.

O presente trabalho não identifica o tipo de erro presente nos valores de uma amostra,

e sim sugere uma forma de representação que considera os erros contidos nestes valores. Esta representação é realizada através do uso de intervalos, onde representam-se os valores $\{x_1, \dots, x_n\}$, com uma margem de precisão δ , em intervalos $\mathbf{x}_1 = [x_1 - \delta, x_1 + \delta], \dots, \mathbf{x}_n = [x_n - \delta, x_n + \delta]$.

1.2 Aritmética Intervalar e Estatística Descritiva Intervalar

O uso de intervalos remonta dos primórdios da humanidade, sendo seu conceito utilizado por muitos matemáticos, mesmo antes da era cristã (DENNIS; KREINOVICH; RUMP, 1998). No entanto, os primeiros relatos de desenvolvimento de uma aritmética entre intervalos são referidos a partir do trabalho de Burkill, em 1924, e Young, em 1931, e mais efetivamente a partir dos trabalhos de Sunaga (1958) e de Moore (1966). A Análise Intervalar surge com o objetivo inicial de controlar a propagação de erros de arredondamento em procedimentos numéricos computacionais (VACCARO, 2001).

A aritmética intervalar utiliza intervalos reais para representar valores infinitos, valores desconhecidos ou para representar valores contínuos que podem ser conhecidos ou não. Os intervalos servem para representar dados inexatos, aproximações e erros de truncamento de procedimentos, como consistência de programas, critério de parada de processos iterativos.

Os intervalos foram definidos com o objetivo inicial de automatizar a análise do erro computacional. Através da utilização de intervalos, tem-se um controle automático de erros com limites confiáveis, além de provas de existência e não existência de solução de diversos problemas.

Na matemática intervalar, em vez de se aproximar um valor real x para um número de máquina, ele é aproximado por um intervalo \mathbf{x} , que possui como limite inferior e superior números de máquina de forma que o intervalo contenha x . O tamanho deste intervalo pode ser usado como medida para avaliar a qualidade de aproximação (RATSCHKE; ROKNE, 1988). Os cálculos reais são substituídos por cálculos que utilizam a aritmética intervalar.

É importante ressaltar que existem questões do tipo: “Por que utilizar técnicas intervalares se existem na maioria dos sistemas de computadores bibliotecas matemáticas avançadas e eficientes, que resolvem a maioria dos problemas?”.

A justificativa do uso de técnicas intervalares, segundo Rastchek *et al* (1988), inicia pelo fato de que os computadores empregam aritméticas chamadas de ponto flutuante ou ponto fixo. Nestas aritméticas, números reais são aproximados para um subconjunto finito de números reais chamados números de máquina representáveis. Devido esta representação são gerados alguns tipos de erros. O primeiro tipo de erro ocorre quando um valor real de entrada é aproximado para um número de máquina. Exemplificando, considera-se o caso do número $\pi = 3, 1415\dots$ que pode ser aproximado, como por exemplo, para o número de máquina 3, 14. O segundo tipo de erro é causado pelos resultados intermediários gerados na execução de cada operação e que vão se acumulando. Existe ainda um outro tipo de erro que está relacionado com a incerteza dos dados de entrada, o que acontece muito no caso de experimentos físicos e químicos onde os dados de entrada são incertos.

No processo de resolução de problemas podem ser constatadas algumas fontes de erros: erros de propagação nos dados iniciais, erros de arredondamento e erros de truncamento, causados ao se truncar seqüências infinitas de operações aritméticas, após um número finito de etapas. Neste contexto percebe-se a importância de técnicas intervalares. Ressalta-se que uma resposta intervalar carrega com ela a garantia de sua incerteza. Um valor pontual não carrega medidas de sua incerteza. Mesmo quando uma análise de

sondagem do erro é executada, o número resultante é somente uma estimativa do erro que pode estar presente.

Segundo Keiffott *et al* (1996), são muitas as aplicações de intervalos e nas mais diversas áreas, tais como: programação matemática, manipulação de equações, análise e projeto de circuitos elétricos, psicologia matemática, estatística, equações diferenciais, física e muitos outros.

Trabalhos desenvolvidos na área de estatística com dados intervalares são muito recentes, destaque-se a tese desenvolvida por Campos (1997). A tese propõe uma metodologia para estender a probabilidade real de tal forma a possibilitar seu cálculo automático através do uso de intervalos, limitando os erros computacionais e mantendo uma semântica consistente entre a probabilidade e os resultados intervalares obtidos. Campos obteve a probabilidade intervalar com a utilização de conceitos de extensões intervalares de funções reais, propostos por Moore (1966) e Acioy (1991). A partir da definição do que seja um espaço de probabilidade intervalar, provou-se que a extensão proposta tinha propriedades similares às da probabilidade real (probabilidade usual). Nesse sentido, a probabilidade intervalar estende a probabilidade real sendo esta uma restrição da primeira.

Dos diversos problemas existentes na computação intervalar, o principal é a existência de métodos que freqüentemente retornam um intervalo z que superestima a imagem procurada y tal que $y \subseteq z$ e $y \neq z$ (KREINOVICH *et al*, 1998).

Segundo Kreinovich *et al* (1998), existem casos onde é extremamente importante conhecer como diferentes valores reais de y podem ser estimados como $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$. Por exemplo, se estimamos que a quantidade de óleo em um poço é de 100 milhões de toneladas, então as ações dependerão de quanto exata é esta estimativa: se a estimativa é razoavelmente exata, ou seja, se o valor real é 100 ± 10 milhões ton., provavelmente poderíamos começar a exploração comercial deste poço de óleo; agora, se a estimativa não é exata, e o erro $\Delta y = \tilde{y} - y$ pode ser tão grande quanto ± 100 , logo é possível que a quantidade atual de óleo y esteja próximo de 0, ou seja, que não existe uma quantidade de óleo suficiente para exploração comercial. Isso significa que as medições realizadas foram insuficientes para dar informações adequadas para decidir sobre o que fazer, e em consequência, se deve realizar medições mais precisas.

Agora, cita-se um exemplo prático do problema de computar o valor de uma função polinomial intervalar: o custo de uma viagem de férias - neste caso a soma correta do total da despesa da viagem pode ser obtida adicionando, subtraindo, e multiplicando diversos números, tais como o custo da passagem (aérea ou terrestre), o valor das diárias em um hotel, refeições, compras de lembranças do local de passeio. Os únicos valores certos que se conhece são os custos da passagem e hotel, os valores referentes as refeições e compras em geral não são precisos, neste caso pode-se considerar que estes valores estão contidos em um intervalo x_i , onde tem-se a idéia de gastar um valor mínimo e máximo em relação àquelas despesas que pode-se chamar de despesas variáveis. Como resultado do total de despesas da viagem, a soma pode ser um polinômio de várias variáveis $f(x_1, \dots, x_n)$, onde $x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n$ e cada intervalo x_i corresponde uma despesa variável.

A literatura ((FERSON *et al.*, 2002), (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002), (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2004), (KREINOVICH, 2003)) mostra que foram realizadas pesquisas somente em problemas das medidas de dispersão variância intervalar, covariância intervalar e coeficiente de correlação intervalar, e que a utilização da computação intervalar na solução de problemas de medidas de dispersão intervalar sempre fornece solução com intervalos superestimados (ou intervalos com amplitude grande), e que ao procurar uma solução

com intervalos de amplitude pequena (através da computação da imagem intervalar), o problema passa a pertencer a classe de problemas NP-Difícil.

Com o objetivo de tornar viável a computação de medidas de dispersão com entradas intervalares, procura-se desenvolver algoritmos eficientes (TOSCANI; VELOSO, 2001), ou seja, algoritmos com tempo de processamento razoável. A certificação da eficiência de tais algoritmos, e a busca de uma possível solução para os problemas de computar os valores intervalares de medidas de dispersão, dá-se através da utilização da complexidade computacional.

1.3 Complexidade dos problemas da Computação Intervalar

O termo complexidade, no contexto de algoritmos, refere-se aos recursos necessários para que um algoritmo possa resolver um problema sob o ponto de vista computacional, ou seja, à quantidade de trabalho despendido pelo algoritmo (TOSCANI; VELOSO, 2001). Quando o recurso é o tempo, são escolhidas uma ou mais operações fundamentais e então são contados os números de execuções desta operação fundamental na execução do algoritmo. Segundo Toscani (2001) a escolha de uma operação como operação fundamental é aceitável se o número de operações executadas pelo algoritmo é proporcional ao número de execuções da operação fundamental.

A complexidade também pode ser vista como uma propriedade do problema, o que significa dar uma medida independente do tratamento dado ao problema, independente do caminho percorrido na busca da solução, portanto independente de algoritmos. Alguns problemas são bem comportados, isto é, permitem chegar a limites de complexidades bem definidos, outros estão em classes com contornos não bem claros (TOSCANI; VELOSO, 2001).

No decorrer das pesquisas sobre problemas que envolvem computação intervalar, verificou-se que a maioria dos problemas pesquisados pertenciam à classe de problemas NP-Difícil (KREINOVICH et al, 1998) (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002) (KREINOVICH, 2003). Diante de tais resultados, passou-se a pesquisar sobre a computação intervalar e trabalhos relacionados.

A análise da complexidade em problemas que envolvem computação intervalar teve como ponto de partida o problema básico da computação intervalar (PBCI) (KREINOVICH et al, 1998), ou seja, o problema de computar o intervalo imagem de uma função contínua e computável f .

Uma aplicação clássica deste problema é, por exemplo, o das medições de quantidade de óleo em um poço (KREINOVICH et al, 1998). Geralmente se conhecem os valores aproximados \tilde{x}_i de uma quantidade física x_i , a exatidão Δ_i de cada medição e o algoritmo $f(x_1, \dots, x_n)$ que transforma os valores x_i no valor da quantidade desejada y . Como resultado, sabe-se que x_i pertence ao intervalo $\mathbf{x}_i = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$ e se deseja conhecer o conjunto de possíveis valores de y . Para uma função contínua $f(x_1, \dots, x_n)$, este conjunto é um intervalo denotado por $\mathbf{y} = [y, \bar{y}]$.

Considerando o exemplo acima, surge a questão: “Pode-se calcular os extremos y e \bar{y} do intervalo \mathbf{y} em tempo razoável?”. Kreinovich *et al* (1998), procuraram responder esta questão analisando a complexidade do PBCI, verificando que o mesmo pertence a classe de problemas NP-Difícil, ou seja, é pelo menos tão difícil de resolver quanto qualquer problema NP. Em outras palavras, é um problema que não tem algoritmo de tempo de processamento polinomial conhecido.

Mesmo o PBCI sendo NP-Difícil, Ferson *et al* (2002) consideraram valores interva-

lares $[\underline{x}_i, \bar{x}_i]$, $i = 1, \dots, n$, para cada um dos n valores de amostra $\{x_1, \dots, x_n\}$ como dados de entrada de problemas da estatística, mais precisamente para os indicadores estatísticos variância, covariância e coeficiente de correlação. Da análise da complexidade computacional do problema de computar o valor da variância va contido em um intervalo $VA = [\underline{va}, \bar{va}]$, foi verificado que o mesmo é NP-Difícil (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002)(FERSON et al., 2002), e considerando algumas restrições sobre os dados de entrada, Ferson *et al* (2002) e Wu 2003, apresentaram algoritmos razoáveis (TOSCANI; VELOSO, 2001) que computavam o extremo inferior \underline{va} do intervalo VA em tempo polinomial. Os problemas de computar o intervalo da covariância intervalar CO e o intervalo da correlação intervalar CC também pertencem a classe de problemas NP-Difícil (FERSON et al., 2002) (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2004), porém para estes problemas não foram consideradas restrições sobre os dados de entrada para melhorar o resultado da complexidade.

A análise da complexidade do problema de computar medidas de dispersão com valores intervalares justifica-se devido a afirmação de Traylor (1995) de que nenhum algoritmo razoável (TOSCANI; VELOSO, 2001) é possível para computar a estimativa de um intervalo ótimo, isto é, a menos que o tempo de processamento seja exponencial. Esta afirmação foi constatada nos resultados encontrados na bibliografia pesquisada sobre a complexidade de problemas com entradas intervalares, porém não com a utilização da computação intervalar (operações aritméticas intervalares e aritmética de exatidão máxima) devido a superestimação no intervalo obtido, e sim com processamento de dados utilizando as operações de aritmética real (computação da imagem intervalar). A estimativa na computação intervalar não fornece a imagem exata, e sim retorna um intervalo que contém o valor da imagem. Esta imagem é chamada de estimativa do intervalo ótimo (“*optimal interval*”) (TRAYLOR; KREINOVICH, 1995).

Analisando a bibliografia relacionada ao tema “complexidade de problemas que envolvem computação intervalar”, verifica-se que não foi abordado o problema de estimar valores esperados (ou esperança) com entradas intervalares. Em decorrência, desenvolveram-se os trabalhos (LORETO et al., 2004) e (LORETO; SILVA; CLAUDIO; TOSCANI; LEAL; RIBEIRO, 2005), os quais mostram que o problema de estimar o valor esperado de uma função com entradas intervalares é NP-Difícil. Em (LORETO; SILVA; CLAUDIO; TOSCANI; LEAL; RIBEIRO, 2005) mostra-se a possibilidade de aproximar um problema NP-Difícil da computação intervalar, através do uso de heurísticas com ênfase em algoritmos baseados no Método de Monte Carlo. A fim de apresentar heurísticas para o problema genérico de estimar numericamente, dada uma precisão ε , o valor esperado de uma função racional $f(x_1, \dots, x_n)$ com $(x_1, \dots, x_n) \in \prod_{k=1}^n I_k$ escolhidos de acordo com uma função de densidade de probabilidade $P(x_1, \dots, x_n)$ conjunta, com I_k sendo intervalos, mostra-se para o particular caso de um grafo regular de grau 2, cujos vértices podem assumir somente dois valores -1 e $+1$, o que restringe os intervalos a serem todos discretos $I_k = [-1, 1] \cap \mathbb{Z}^*$, $k = 1, \dots, n$. Verifica-se que o problema do valor esperado entre os pares de vértices adjacentes, tivera seu custo reduzido de $O(2^n)$ operações, através do cálculo exato, para $O(N_{sample})$ que seria a complexidade encontrada quando aproxima-se o resultado obtido pela Heurística, onde N_{sample} é o número de amostras de configurações escolhidas através do processo de amostragem por importância no contexto do algoritmo de Metrópolis.

O trabalho (LORETO et al., 2005), desenvolvido sobre o tema da tese, apresenta a complexidade computacional dos problemas de computar, com entradas intervalares, as medidas de tendência central média, mediana e moda, e as medidas de dispersão am-

plitude total, variância, desvio padrão, coeficiente de variação, covariância e coeficiente de correlação. Para a investigação da complexidade elabora-se uma abordagem intervalar para os indicadores estatísticos e uma forma de representação dos valores reais em valores intervalares, de tal modo que não ocorre superestimação nos intervalos solução. Por meio da análise da complexidade computacional verifica-se que os problemas de computar as medidas de tendência central e dispersão, com entradas intervalares, pertencem à classe de problemas P, quando resolvidos através da extensão intervalar (MOORE, 1979).

Ao analisar a complexidade de um problema pode-se informar qual a classe de complexidade que o mesmo pertence. Conhecendo a classe em que o problema pertence, os projetistas de algoritmos, e até mesmo profissionais da área de estatística, poderiam se concentrar naqueles problemas para os quais existem algoritmos razoáveis, ou seja, que possam ser resolvidos através do computador em tempo polinomial de processamento.

Segundo Toscani (2001), identificar a tratabilidade e a intratabilidade dos problemas, mesmo aqueles que possuem algoritmos imediatos (algoritmos de fácil construção), é de extrema importância para os projetistas de algoritmos, pois conhecendo as classes de complexidade em que os problemas pertencem, os projetistas poderiam ter uma medida real quanto às soluções disponíveis e a expectativa de melhorar esses resultados.

Conforme comentado anteriormente, a estatística descritiva é amplamente utilizada nas mais diversas áreas. Salienta-se que um dos problemas mais preocupantes referem-se aos valores de amostras que geralmente contém erros de medições. A confiabilidade dos valores das medidas de tendência central média, mediana, moda, das medidas de dispersão amplitude, variância, desvio padrão, coeficiente de variação, covariância, coeficiente de correlação e das separatrizes quartil, decil e percentil dependem exclusivamente dos valores de amostra.

Uma alternativa proposta por este trabalho é o uso de intervalos para considerar erros de medidas, inerentes aos valores de amostra, a fim de que se obtenham valores mais confiáveis para os indicadores estatísticos. Além disso, com a abordagem intervalar, com a utilização da extensão intervalar e com algoritmos eficientes (verificados através da análise da complexidade computacional) os indicadores estatísticos com entradas intervalares tornam-se possíveis de serem calculados através de ambientes intervalares como Maple Intervalar (MAPLE, 2005) e IntLab (MATLAB, 2005).

1.4 Objetivos do Trabalho

O objetivo de tese proposta é seguinte:

“Analisar a complexidade computacional dos problemas de computar os valores de indicadores estatísticos com entradas intervalares, a fim de obter uma classificação quanto a classe de complexidade destes problemas”.

O trabalho desenvolvido neste volume visa a realização dos seguintes objetivos específicos:

- Determinar uma abordagem intervalar para os indicadores estatísticos descritivos média, mediana, moda, amplitude total, variância, coeficiente de variação, covariância, coeficiente de correlação e separatrizes quartil, decil e percentil;
- Apresentar uma forma de representação dos valores reais em valores intervalares, de tal modo que garanta a qualidade de aproximação nos intervalos solução em relação a resposta real exata;

- Definir um novo problema, para cada indicador estatístico, através da definição do domínio intervalar de funções intervalares com variáveis intervalares, considerando a extensão intervalar como método de resolução. A definição de um novo problema se faz necessária devido ao PBCI, aplicado a estatística, possuir como domínio da função f valores reais x_1, \dots, x_n contidos em intervalos x_i , ou seja, $Dom(f) = \mathbb{R}^n$ (onde n é o número de argumentos da função) e o mesmo pertencer a classe de problemas NP-Difícil (FERSON et al., 2002)(FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002)(FERSON et al., 2002)(KREINOVICH, 2003).
- Propor algoritmos para a solução dos problemas de computar os intervalos de medidas de tendência central, dispersão e separatrizes com entradas intervalares. Tais algoritmos utilizam a aritmética intervalar definida por Moore (MOORE, 1966) e a extensão intervalar (MOORE, 1979);
- Reclassificar, quanto a classe de complexidade, os problemas de calcular os indicadores estatísticos variância, covariância e coeficiente de correlação com entradas intervalares;
- Mostrar, através de exemplos de cálculos intervalares usando sistema de ponto flutuante, arredondamento direcionado (KULISCH; MIRANKER, 1981) e extensão intervalar, a qualidade dos intervalos solução.

Por fim, é objetivo tornar viável a computação dos valores dos indicadores estatísticos intervalares variância, covariância e coeficiente de correlação, os quais não eram passíveis de serem calculados devido os resultados de NP-dificuldade apresentados na bibliografia pesquisada.

1.5 Estrutura da tese

Os demais capítulos deste trabalho observam a seguinte estrutura:

- O Capítulo 2 apresenta conceitos básicos sobre os quais está formalizada a teoria da complexidade computacional. Descreve-se, ainda, um apanhado geral do estado da arte do tema complexidade de problemas intervalares (seção 2.3), o PBCI com restrições (seção 2.3.1);
- O Capítulo 3 apresenta uma forma de representação dos valores reais em valores intervalares, a ordenação intervalar quanto todos os intervalos são disjuntos ou quanto são todos encaixados e definições, a nível intervalar, de medidas de tendência central, dispersão e separatrizes, e o estado da arte da complexidade dos problemas de estatística descritiva com entradas intervalares;
- O Capítulo 4 apresenta a investigação da complexidade computacional de problemas de medidas de tendência central, dispersão e separatrizes com entradas intervalares e os resultados de complexidade dos problemas dos indicadores estatísticos considerando a imagem intervalar (estado da arte) e extensão intervalar (nossos resultados);
- O Capítulo 5 apresenta medidas de correção de uma resposta intervalar, exemplos numéricos dos indicadores estatísticos intervalares realizados com o estudo de caso

do IMC (Índice de Massa Corporal) de alunos do ensino fundamental e averiguação da qualidade dos intervalos solução;

- O Capítulo 6 apresenta as conclusões do trabalho, juntamente com indicações de desenvolvimentos futuros.

O volume encerra-se com a apresentação das referências bibliográficas e os anexos utilizados para o desenvolvimento deste trabalho.

1.6 Comentários Finais

O presente capítulo apresentou o tema sobre o qual é formulada a presente tese, ressaltando a motivação e a justificativa para sua escolha, bem como aspectos relacionados à estruturação e desenvolvimento deste trabalho.

O próximo capítulo apresentará os fundamentos teóricos de complexidade computacional, sobre o qual o presente trabalho está direcionado.

2 COMPLEXIDADE DE PROBLEMAS INTERVALARES

Este capítulo apresenta, de forma sucinta, conceitos de Complexidade Computacional (problema computacional, classes de complexidade) e aritmética intervalar. Tais conceitos são de fundamental importância para a compreensão e o desenvolvimento do trabalho.

Nas seções seguintes apresentamos o desenvolvimento de pesquisa na área da complexidade computacional de problemas que envolvem computação intervalar.

2.1 Aritmética Intervalar

Nesta seção são apresentadas as definições básicas da aritmética intervalar, necessárias para o desenvolvimento do presente trabalho. Tais definições estão baseadas nos trabalhos de Moore (1966) (1979) e Oliveira (1997).

A aritmética intervalar é uma ferramenta matemática para a solução de problemas relacionados com erros numéricos na computação científica. Por aritmética intervalar entende-se um sistema algébrico consistindo do conjunto formado de todos os intervalos fechados da reta real, com algumas operações definidas sobre ele. Ao invés de algoritmos usuais para resolver certos problemas matemáticos, usam-se algoritmos com operações intervalares desenvolvidos para produzir intervalos como resultados, em lugar de números reais, os quais: (i) devem ser confiáveis no sentido que incluem a solução exata do problema original; (ii) devem ter como resposta um intervalo com amplitude tão pequena quanto possível (CAMPOS, 1997).

Definição 2.1. (Intervalo de reais) *Seja \mathbb{R} o conjunto dos números reais, e sejam $x_1, x_2 \in \mathbb{R}$, tais que $x_1 \leq x_2$. Então, o conjunto $\{x \in \mathbb{R} \mid x_1 \leq x \leq x_2\}$ é um intervalo de reais ou simplesmente um **intervalo**, e será denotado por $\mathbf{x} = [x_1, x_2]$.*

Definição 2.2. (Conjunto \mathbb{IR}) *Define-se e denota-se por \mathbb{IR} o conjunto de todos os intervalos de reais, ou seja, $\mathbb{IR} = \{[x_1, x_2] \mid x_1, x_2 \in \mathbb{R}, x_1 \leq x_2\}$.*

Observa-se que todo número real $x \in \mathbb{R}$ pode ser visto como um intervalo de \mathbb{IR} . Para tanto, basta identificar os pontos $x \in \mathbb{R}$ com os intervalos pontuais $\mathbf{x} = [x, x] \in \mathbb{IR}$. Estes intervalos também são chamados de intervalos degenerados ou intervalo pontual (OLIVEIRA; DIVERIO; CLAUDIO, 1997).

Definição 2.3. (Igualdade entre intervalos) *Sejam $\mathbf{x} = [x_1, x_2]$ e $\mathbf{y} = [y_1, y_2]$ dois intervalos de \mathbb{IR} . Dizemos que $\mathbf{x} = \mathbf{y}$ se e somente se $x_1 = y_1$ e $x_2 = y_2$.*

Relação de Ordem: Há na literatura (MOORE, 1966), (MOORE, 1979), (SUNAGA, 1958), vários tipos de relações de ordem definidas sobre \mathbb{IR} . Entre essas, tem-se uma ex-

tensão da relação de ordem $<$ definida sobre o conjunto dos números reais. Esta extensão foi definida por Moore (1966) (1979) e é como segue:

$$\mathbf{x} < \mathbf{y} \Leftrightarrow x_2 < y_1.$$

Desde que intervalos reais são também conjuntos, outra relação de ordem que pode ser definida (MOORE, 1966) (MOORE, 1979) (SUNAGA, 1958) é \subseteq dada a seguir:

$$\mathbf{x} \subseteq \mathbf{y} \Leftrightarrow y_1 \leq x_1, x_2 \leq y_2.$$

A relação de ordem definida por Kulish e Miranker (1981) é a ordem \leq , dada por

$$\mathbf{x} \leq \mathbf{y} \Leftrightarrow x_1 \leq y_1, x_2 \leq y_2.$$

Os símbolos $\langle \mathbb{IR}, < \rangle$, $\langle \mathbb{IR}, \subseteq \rangle$, $\langle \mathbb{IR}, \leq \rangle$ indicam o espaço dos intervalos com as relações de ordem $<$, \subseteq e \leq , respectivamente.

Na definição da aritmética intervalar as operações sobre \mathbb{IR} devem manter a propriedade do fechamento, ou seja, o resultado ainda sejam um intervalo.

Definição 2.4. (Operações aritméticas em \mathbb{IR}) Sejam $\mathbf{x}, \mathbf{y} \in \mathbb{IR}$ dois intervalos de reais. As operações de soma, subtração, multiplicação e divisão em \mathbb{IR} são definidas por $\mathbf{x} \star \mathbf{y} = \{x \star y \mid x \in \mathbf{x}, y \in \mathbf{y}\}$ onde $\star \in \{+, -, *, /\}$ é quaisquer uma das quatro operações aritméticas. Se ω é uma operação unária, então $\omega \mathbf{x}$ é definida por $\omega \mathbf{x} = \omega(\mathbf{x}) = \{\omega x \mid x \in \mathbf{x}\} = [\min\{\omega(x) \mid x \in \mathbf{x}\}; \max\{\omega(x) \mid x \in \mathbf{x}\}]$.

Observa-se que \mathbf{x}/\mathbf{y} somente está definido se $0 \notin \mathbf{y}$.

Sejam os intervalos reais \mathbf{x} e \mathbf{y} de forma $\mathbf{x} = [x_1, x_2]$ e $\mathbf{y} = [y_1, y_2]$. Tem-se as operações aritmética definidas como segue:

$$\mathbf{x} + \mathbf{y} = [x_1 + y_1, x_2 + y_2]$$

$$\mathbf{x} - \mathbf{y} = [x_1 - y_2, x_2 - y_1]$$

$$\mathbf{x} * \mathbf{y} = [\min\{x_1 y_1, x_1 y_2, x_2 y_1, x_2 y_2\}, \max\{x_1 y_1, x_1 y_2, x_2 y_1, x_2 y_2\}]$$

$$\mathbf{x}/\mathbf{y} = \mathbf{x} * 1/\mathbf{y} = [x_1, x_2] * [1/y_2, 1/y_1], 0 \notin \mathbf{y}.$$

Definição 2.5. (Diâmetro de um intervalo) Seja $\mathbf{x} = [x_1, x_2] \in \mathbb{IR}$ um intervalo. Defina-se o diâmetro do intervalo \mathbf{x} , denotado por $\text{diam}(\mathbf{x})$ ou $w(\mathbf{x})$, como sendo o número real não negativo $\text{diam}(\mathbf{x}) = w(\mathbf{x}) = x_2 - x_1$.

Definição 2.6. (Ínfimo e Supremo de um intervalo) Defina-se o ínfimo do intervalo \mathbf{x} como o limite inferior do intervalo. O supremo do intervalo é definido como o limite superior do intervalo. São denotados por: $\text{inf}(\mathbf{x}) = x_1$ e $\text{sup}(\mathbf{x}) = x_2$.

Definição 2.7. (Ponto Médio) Defina-se o ponto médio do intervalo \mathbf{x} como sendo o número real $m(\mathbf{x}) = \frac{x_1 + x_2}{2}$.

Avaliação de expressões e funções: A aritmética intervalar é baseada no uso de intervalos finitos, fechados $[x_1, x_2]$ de números reais como elementos básicos e sua idéia básica do ponto de vista computacional, é: dada uma função $f(x)$ de variável real x pertencente a um intervalo $\mathbf{x} = [x_1, x_2]$ onde $x_1, x_2 \in \mathbb{R}$. O domínio da f é dado por:

$$f(x) = \{y \mid y = f(x), x_1 \leq x \leq x_2\}.$$

Este domínio não é, em geral, dado exatamente, mas é sempre possível determinar um intervalo $\mathbf{y} = [y_1, y_2]$ tal que $f(\mathbf{x}) \subseteq \mathbf{y}$, isto é $y_1 \leq f(x) \leq y_2$. Pode-se então definir uma função intervalar F associada a f pela transformação do intervalo $[x_1, x_2]$ em $[y_1, y_2]$, isto é:

$$f(x) \subseteq F(\mathbf{x}) = \mathbf{y}.$$

Esta função F , chamada extensão intervalar de f , deve ser aquela que se afasta o mínimo possível de $f(\mathbf{x})$. O erro obtido no cálculo de $f(x)$ a partir do intervalo \mathbf{x} é obtido através do diâmetro $w(F(\mathbf{x})) = y_2 - y_1$.

O cálculo de expressões na aritmética intervalar consiste na extensão das operações aritméticas, o qual junto com um conjunto de funções *standards*, forma a aritmética intervalar.

A aritmética intervalar utiliza um arredondamento especial, chamado arredondamento direcionado, o que significa que os resultados são arredondados para o menor e para o maior número de máquina que contém o resultado das operações, obtendo-se com isso um intervalo de máquina, com diâmetro mínimo, no qual a solução se situa.

Uma distinção importante na aritmética intervalar é da imagem intervalar de uma função da avaliação intervalar da função. A Imagem Intervalar de uma função f , contínua no intervalo \mathbf{x} , é definida como o intervalo limitado pelo mínimo da imagem de $f(x)$ e pelo máximo da imagem de $f(x)$, sendo x um elemento do intervalo \mathbf{x} .

$$I(f(\mathbf{x})) = [\min\{f(x)\}, \max\{f(x)\}].$$

Na computação intervalar, pode-se calcular o intervalo solução $\mathbf{y} = [\underline{y}, \bar{y}] = f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ através de métodos de aproximação, técnicas de otimização, extensão intervalar e, ainda, por métodos considerados mais sofisticados (KREINOVICH et al, 1998) como forma centrada.

Dentre os métodos de aproximação existentes, descreve-se o Método de Linearização de uma função. Quando é suficiente obter os valores *aproximados* dos extremos \underline{y} e \bar{y} do intervalo \mathbf{y} , pode-se *linearizar* a função $f(x_1, \dots, x_n)$, isto é, representar x_i como $x_i = \tilde{x}_i - \Delta x_i$, e expandir a expressão resultante $f(x_1, \dots, x_n) = f(\tilde{x}_1 - \Delta x_1, \dots, \tilde{x}_n - \Delta x_n)$ em série de Taylor em relação a Δx_i , omitindo termos quadráticos e de ordem superior nesta expansão. Como resultado obtém-se uma fórmula aproximada

$$f(x_1, \dots, x_n) \approx a_0 + a_1 \cdot \Delta x_1 + \dots + a_n \cdot \Delta x_n,$$

onde $a_0 = f(\tilde{x}_1, \dots, \tilde{x}_n) = \tilde{y}$ e $a_1 = -\frac{\partial f}{\partial x_1}(\tilde{x}_1, \dots, \tilde{x}_n)$. Para a função linear resultante (aproximada), calcula-se o intervalo \mathbf{y} da seguinte maneira:

- $\underline{y} = \tilde{y} - |a_1| \cdot \Delta_1 - \dots - |a_n| \cdot \Delta_n;$
- $\bar{y} = \tilde{y} + |a_1| \cdot \Delta_1 + \dots + |a_n| \cdot \Delta_n.$

Segundo Kreinovich *et al* (1998), para muitos problemas práticos, métodos de aproximação não são suficientes, pois deve-se ter uma estimativa garantida para a imagem y , e devido a uma série de fatores como:

- Medições de erros são, às vezes, relativamente grandes, de modo que seus quadrados não podem ser omitidos seguramente;
- A função $f(x_1, \dots, x_n)$ que descreve a relação entre quantidades medidas diretamente x_i e a quantidade desejada y pode não ser linear;
- Muitos algoritmos de processamento de dados processam valores em grande quantidade, ou seja, quando eles processam os valores medidos em diferentes momentos de tempo;
- Existem problemas nos quais necessita-se de uma estimativa garantida, pois uma superestimação poder ser desastrosa,

métodos de aproximação não são muito empregados.

Uma das técnicas de otimização é o cálculo do ponto de máximo e de mínimo da uma função (MOORE, 1979), também conhecido como Imagem Intervalar (OLIVEIRA; DIVERIO; CLAUDIO, 1997). O problema de encontrar os extremos do intervalo solução é um problema típico de otimização: o extremo inferior \underline{y} é a solução do problema de minimização $f(x_1, \dots, x_n) \rightarrow \min$ sobre as condições $\underline{x}_i \leq x_i \leq \bar{x}_i$, $1 \leq i \leq n$ (onde $\underline{x}_i = \tilde{x}_i - \Delta_i$ e $\bar{x}_i = \tilde{x}_i + \Delta_i$), e o extremo superior \bar{y} é a solução do problema de maximização $f(x_1, \dots, x_n) \rightarrow \max$ sobre as condições $\underline{x}_i \leq x_i \leq \bar{x}_i$, $1 \leq i \leq n$. Para encontrar o máximo, é suficiente encontrar o ponto cuja derivada é igual a 0 (ou todos os pontos existentes sobre o intervalo \mathbf{x}); calcular o valor $f(x)$ de todos os pontos “candidatos” e de todos os extremos significa encontrar todos os valores de $f(x)$ de todos estes pontos, o maior valor de $f(x)$ é o máximo desejado (correspondentemente, o menor dos valores de $f(x)$ é o mínimo desejado). Para uma aplicação proveitosa desta técnica deve-se considerar o tipo de função, se a função $f(x)$ é muito complicada, então a equação $df/dx = 0$ também será muito complicada e, por essa razão, difícil de resolver. Porém para funções razoavelmente simples este método é muito eficiente.

Ratschek *et al* (1988) afirmam que a maioria dos métodos de otimização têm, no mínimo, dois defeitos (falhas). O primeiro defeito é que o método não garante que os pontos de mínimo, ou de máximo, possam ser encontrados para uma dada tolerância. O segundo defeito é que o método, dependendo das condições da função, permite encontrar somente o mínimo local ao invés do global. Estes defeitos dificultam a solução de problemas de otimização global. A otimização global é considerada, por essa razão, um assunto intratável.

Segundo Ferson *et al* (2002), historicamente, o primeiro método para computar o intervalo solução é a extensão intervalar (MOORE, 1979), ou avaliação intervalar (OLIVEIRA; DIVERIO; CLAUDIO, 1997). Este método está baseado no fato que em um computador, todo algoritmo consiste de operações elementares (aritméticas e lógicas). Para cada operação elementar $f(a, b)$, se são conhecidos os intervalos a e b para a e b , pode-se computar a imagem exata $f(\mathbf{a}, \mathbf{b})$, através da aritmética intervalar definida por Moore em (MOORE, 1966). Na extensão intervalar, repetindo a computação formando o programa f passo-a-passo, substituindo cada operação elementar de números reais pela correspondente operação da aritmética intervalar.

Outro método, considerado sofisticado, é a forma centrada (RATSCHEK; ROKNE, 1988). Estima-se o intervalo solução de $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ de uma função como

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) \subseteq f(\tilde{x}_1, \dots, \tilde{x}_n) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}_1, \dots, \mathbf{x}_n) \cdot [-\Delta_i, \Delta_i],$$

onde $\tilde{x}_i = (\underline{x}_i + \bar{x}_i)/2$ é o ponto médio e $\Delta_i = (\underline{x}_i - \bar{x}_i)/2$ é o raio do intervalo \mathbf{x} . A forma centrada sempre produz um intervalo centrado no ponto $f(\tilde{x}_1, \dots, \tilde{x}_n)$. Dessa forma, como resultado de aplicação da forma centrada, obtém-se um intervalo centrado em 0 (zero), isto é, o extremo inferior do intervalo é negativo. O extremo superior produzido pela forma centrada é diferente do extremo superior do intervalo desejado.

Com o objetivo de tornar possível a computação dos indicadores estatísticos variância, covariância e coeficiente de correlação com entradas intervalares, escolhe-se a extensão intervalar por se adequar às expressões dos indicadores estatísticos com abordagem intervalar (Capítulo 3), e por tornar mais acessível a construção de algoritmos imediatos (TOSCANI; VELOSO, 2001). Segundo Ferson *et al* (2002), em alguns casos o intervalo solução é exato, porém em outros casos mais complexos (como o problema de computar o intervalo da variância, por exemplo) obtém-se intervalos solução com amplitude grande, ou seja, intervalos superestimados. Preocupados com a possível ocorrência de intervalos superestimados, representam-se os valores reais $\{x_1, \dots, x_n\}$ de amostras aleatórias de uma população em intervalos $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ considerando uma margem de precisão δ , isto é, $\mathbf{x}_1 = [x_1 - \delta, x_1 + \delta], \dots, \mathbf{x}_n = [x_n - \delta, x_n + \delta]$. Em virtude desta representação, podem ocorrer amostras intervalares $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ onde todos os intervalos podem ser disjuntos, ou todos encaixados ou, ainda, alguns intervalos encaixados e alguns disjuntos. Salienta-se que o principal foco desta representação, de reais em intervalos, está relacionado ao erro numérico, isto é, erros oriundos de medições.

2.2 Complexidade Computacional

A seguir são apresentados os conceitos básicos sobre os quais está fundamentada a Teoria da Complexidade Computacional.

2.2.1 Algoritmos

Um algoritmo é, no caso geral, uma descrição passo a passo de como um problema é solucionado. A descrição deve ser finita e os passos devem ser bem definidos, sem ambigüidades, além de executáveis computacionalmente. Diz-se que um algoritmo resolve um problema p se este algoritmo recebe qualquer instância I de p e sempre fornece uma solução para esta instância I . Observa-se que, para qualquer instância I do problema p definida como uma entrada, o algoritmo deverá ser executável em tempo finito e, além disso, produzir como saída uma solução correta para o problema p (TOSCANI; VELOSO, 2001).

Formalmente algoritmo é definido como uma máquina de Turing (MT) que sempre pára. A definição de MT apresentada aqui segue a definição de Lewis (1981). A máquina de Turing consiste de uma fita e uma unidade de controle que reflete o estado corrente da máquina. A comunicação entre a fita e a unidade de controle é feita através de um cabeçote que se move ao longo desta fita, de uma célula para outra. Este cabeçote lê símbolos da fita e também é usado para mudar os símbolos na fita. A unidade de controle opera em passos discretos; cada passo realiza duas funções de maneira dependente do estado e do

símbolo corrente da fita verificado pelo cabeçote leitura/gravação: 1) coloca a unidade de controle em novo estado e 2) realiza uma das seguintes operações a) escreve um símbolo na célula corrente, substituindo o que já tinha; b) move o cabeçote leitura/gravação uma célula para esquerda ou direita da fita. Como a máquina de Turing pode escrever sobre a fita, isto pode permitir uma resposta sobre a fita e o fim da computação. Existe um estado especial *de parada* que é usado para avisar o fim da computação. Desde que este estado de parada seja o mesmo para todas as máquinas de Turing, denota-se este estado por h ; este símbolo não será usado daqui pra diante para qualquer outro propósito. Similarmente, o símbolo em branco será fixado como $\#$. Usa-se, também, os símbolos L e R para denotar o movimento do cabeçote para esquerda ou direita; assume-se que estes dois símbolos não são elementos de qualquer alfabeto que pode ser considerado.

A seguir apresentamos a definição formal da máquina de Turing (LEWIS; PAPADIMITRIOU, 1981):

Definição 2.8. MÁQUINA DE TURING DETERMINÍSTICA: *A máquina de Turing determinística (MTD) é uma quádrupla (K, Σ, δ, s) onde*

K é um conjunto finito de estados, não contendo o estado de parada h ;

Σ é um alfabeto, contendo o símbolo branco $\#$, sem os símbolos L e R ;

$s \in K$ é o estado inicial;

δ é uma função de $K \times \Sigma$ para $(K \cup \{h\}) \times (\Sigma \cup \{L, R\})$.

Se $q \in K$, $a \in \Sigma$, e $\delta(q, a) = (p, b)$, então a máquina M , quando no estado q for identificado símbolo a , passará para o estado p , e (1) se b é um símbolo em Σ , reescreve a como b , ou (2) se b é L ou R , move o cabeçote na direção esquerda ou direita, respectivamente. Sendo δ uma função, a operação de M é determinística e somente pára quando M entra no estado de parada ou tenta mover para esquerda no final da fita.

Para especificar o estatus da computação da máquina de Turing, necessita-se especificar o estado, o conteúdo da fita, e a posição do cabeçote. Mesmo que uma porção inicial finita da fita seja branca, os conteúdos da fita podem ser especificados por uma *string* finita. Lewis (1981) escolheu dividir a *string* em dois pedaços: a parte á esquerda da célula lida, incluindo o símbolo simples ali contido; e a parte, possivelmente vazia, á direita da célula lida. Além disso, considerando que dois pares de (*string* irão corresponder á mesma combinação de posição e conteúdo do cabeçote e da fita, ressalta-se que a segunda *string* não acaba com um espaço em branco (todos as células da fita á direita do último explicitamente representado são assumidos como contendo espaços em branco de qualquer modo). Com estas considerações, Lewis (1981) formulou a seguinte definição:

Definição 2.9. CONFIGURAÇÃO: *Uma configuração da máquina de Turing $M = (K, \Sigma, \delta, s)$ é um elemento de*

$$(K \cup \{h\}) \times \Sigma^* \times \Sigma \times (\Sigma^*(\Sigma - \{\#\}) \cup \{e\}).$$

Uma configuração cujo estado componente é h será chamada uma configuração de parada.

Para a Máquina de Turing tem-se claramente os passos definidos (uma aplicação de δ , assim pode-se definir complexidade computacional (tempo de computação) $t_a(x)$ de um algoritmo a sobre uma entrada x como o número de passos da máquina de Turing com entrada x , até parar. Esta é a definição padrão da complexidade computacional na teoria da computação.

Segundo Toscani (2001) a complexidade, quantidade de trabalho requerido por um algoritmo, não pode ser descrita simplesmente por um número, porque o número de operações básicas efetuadas em geral não é o mesmo para qualquer entrada (depende do tamanho da entrada). Mesmo para entradas do mesmo tamanho, o número de operações efetuadas pelo algoritmo pode variar.

Questões relativas à complexidade de um algoritmo em termos do tempo de computação e espaço de memória são determinantes para o julgamento da eficiência do mesmo. Um algoritmo, para ser razoável ou não, vai depender de quantos passos computacionais ele necessita para chegar a solução de um problema. Um algoritmo é considerado razoável quando obtém a solução de um problema em tempo polinomial (KREINOVICH et al, 1998).

A caracterização de algoritmo de tempo polinomial está baseada em duas noções fundamentais: tamanho da entrada, $tam(x)$, e tempo de computação, $t_a(x)$. Para formalizar a noção de algoritmo razoável, deve-se formalizar as noções de tamanho da entrada e tempo de computação.

A noção do tamanho da entrada em diferentes computadores está relacionada com a representação dos dados. Embora os computadores possuem diferentes *hardware*, a maioria deles utilizam a mesma forma de representação de todos seus dados, cada elemento do dado é representado por uma seqüência de 0's e 1's. Desta forma, sempre podemos medir o tamanho de cada entrada pelo número de *binary units* (também conhecido como bits), isto é, pelo número de 0's ou 1's que são necessários para representá-lo (CORMEN; LEISERSON; RIVEST, 1990).

Para definir as classes de complexidade é preciso definir algoritmos não-determinísticos.

Garey e Johnson (1979), definem que os algoritmos não-determinísticos são compostos de dois estágios separados: um primeiro estágio de conjectura (ou estimativa) e um segundo estágio de checagem (ou reconhecimento). Dada uma instância I de um problema, o primeiro estágio apenas conjectura alguma proposta de solução S. Para o estágio de checagem, ambos I e S são dados como entrada, e a computação prossegue de maneira determinística, parando com a resposta sim ou com a resposta não.

2.2.2 Problema Computacional

Sob um ponto de vista computacional, Garey e Johnson (1979) descrevem informalmente um problema como uma questão genérica a ser respondida, geralmente possuindo vários parâmetros, ou variáveis livres.

Para Kreinovich *et al* (1998), quando dizemos que existe um problema, queremos dizer que: (i) temos alguma informação x e (ii) conhecemos a relação $R(x, y)$ entre a informação conhecida x e o objeto desejado y .

Uma instância (I) de um problema é obtida através da especificação de valores particulares para todos os parâmetros do problema. A descrição da instância de um problema, fornecida como entrada (*input*) para o computador, pode ser vista como uma cadeia (*string*) finita de símbolos escolhidos de um alfabeto finito.

Um problema é chamado de problema computável se existir um procedimento efetivo que o resolva em um número finito de passos, ou seja, se existe um algoritmo que leve à sua solução. Observa-se, contudo, que um problema considerado “em princípio” computável pode não ser tratável na prática, devido às limitações dos recursos computacionais para executar o algoritmo implementado (TOSCANI; VELOSO, 2001).

Se existe um algoritmo de tempo polinomial que resolve todas as instâncias de um

problema, este problema é tratável, caso contrário diz-se que é intratável (TOSCANI; VELOSO, 2001).

Conforme Szwarcfiter (1984), existem classes de problemas: os Problemas de Decisão, os de Localização e os de Otimização. Um problema de decisão tem a resposta sim ou não. Um problema de localização requer a localização de uma resposta que satisfaça um conjunto de propriedades dadas. Os problemas de otimização têm o objetivo de encontrar um ponto ótimo (mínimo ou máximo) de uma função definida sobre um certo domínio. Apesar da classificação, tais classes apresentam um relacionamento entre si, ou seja, é possível associar problemas de decisão, otimização e localização, de tal forma que um problema possa ser derivado do outro. Exemplificando: se um problema de otimização pede por uma estrutura de um certo tipo que tem custo mínimo entre todas as outras, pode-se associá-lo com aquele problema de decisão que inclui um limite numérico L como um parâmetro adicional e que pergunta se existe uma estrutura do tipo requerido que tenha um custo não maior que L .

2.2.3 Complexidade

O termo *Complexidade* refere-se, em geral, aos requerimentos de recursos necessários para que um algoritmo possa resolver um problema sob o ponto de vista computacional, ou seja, é a quantidade de trabalho despendido pelo algoritmo.

Se a complexidade é tomada como a máxima para qualquer entrada de um dado “tamanho”, a complexidade é chamada Complexidade no pior caso ou simplesmente Complexidade, como será chamada neste trabalho. O critério de avaliação mais utilizado, entre as medidas de complexidade, é a complexidade no pior caso (TOSCANI; VELOSO, 2001).

A complexidade também pode ser vista como uma propriedade do problema, o que significa dar uma medida independente do tratamento dado ao problema, independente do caminho percorrido na busca da solução, portanto independente de algoritmos. Alguns problemas são bem comportados, isto é, permitem chegar a limites de complexidades bem definidos, outros estão em classes com contornos não bem claros (TOSCANI; VELOSO, 2001).

Os problemas, como vimos podem ser classificados como tratáveis e intratáveis sob o ponto de vista computacional. Há ainda a considerar problemas cujo status é desconhecido, ou seja, nenhum algoritmo de tempo polinomial é conhecido, nem tampouco foi provada a não existência de um algoritmo de tempo polinomial que os resolva. A Teoria da Complexidade está estruturada matematicamente de modo a classificar problemas computacionais relativamente a sua dificuldade intrínseca (GAREY; JOHNSON, 1979).

Segundo Garey e Johnson (1979), a teoria da complexidade se aplica principalmente a problemas de decisão. Problemas de decisão são formalmente representados como linguagens, de tal forma que a solução de um problema de decisão equivale ao reconhecimento da linguagem correspondente por uma máquina de Turing. Em geral, diz-se que um programa MTD (Máquina de Turing Determinística) M com um alfabeto de entrada Σ aceita $x \in \Sigma^*$ se, e somente se, M pára quando aplicado à entrada x . A linguagem L_M reconhecida pelo programa M é dada por $L_M = \{x \in \Sigma^* \mid M \text{ aceita } x\}$ (TOSCANI; VELOSO, 2001).

Quanto à complexidade do problema de decisão, são definidas classes: P, NP, NP-Completo e NP-Difícil, entre outras desses.

2.2.4 Classes de Complexidade

Nesta seção são definidas as principais classes de complexidade relativamente à performance de seus algoritmos com respeito a complexidade de tempo.

- **CLASSE P:** A classe P é definida como o conjunto de todos os problemas de decisão resolvíveis por um algoritmo determinístico em tempo polinomial (GAREY; JOHNSON, 1979).

A importância de tal classe deriva do fato de ela conter todos os problemas “simples”, isto é, aqueles que são computacionalmente tratáveis, isto é, tem algoritmo razoável (polinomial). Observa-se, no entanto, que um algoritmo razoável não é necessariamente “eficiente”. De fato, devido a classificação em relação a complexidade de tempo não ser mais do que uma definição de um limite superior para a ordem de complexidade, é possível que hajam exceções, no sentido de que existam alguns algoritmos classificados como exponenciais que apenas para alguma instância apresentam esse tipo de complexidade, enquanto que outros, classificados como polinomiais, tenham um grau tão alto que se tornam ineficientes, na prática.

- **CLASSE NP:** A classe NP é definida como o conjunto de todos os problemas de decisão resolvíveis por algoritmo não-determinístico em tempo polinomial. Uma definição alternativa para a classe NP é como a classe de problemas de decisão para os quais a verificação de que uma solução estimada para uma dada entrada satisfaz todos os requerimentos do problema, pode ser certificada rapidamente, isto é, em tempo polinomial (GAREY; JOHNSON, 1979).

A relação entre as classes de problemas P e NP é um problema em aberto na Ciência da Computação. Certamente $P \subseteq NP$, mas não se sabe se $P = NP$. Existe uma forte conjectura de que $P \neq NP$ (TOSCANI; VELOSO, 2001).

- **RELAÇÃO ENTRE P E NP:** A partir da conjectura $P \neq NP$, a distinção entre os conjuntos P e NP-P é muito importante e significativa: todos os problemas em P podem ser resolvidos com algoritmo de tempo polinomial, enquanto que todos os problemas em NP-P são intratáveis. Com certeza este é um dos assuntos que têm preocupado muitos pesquisadores desta área. Em muitos casos, é possível responder uma pergunta através da idéia de redução de um problema em outro. A redutibilidade dá uma noção precisa do que significa um problema ser pelo menos tão difícil quanto o outro. Por essa razão a noção de redutibilidade entre problemas tornou-se tão importante (TOSCANI; VELOSO, 2001).

Um tipo de redução, introduzida por (GAREY; JOHNSON, 1979), chamada *transformação polinomial* é definida como segue.

Definição 2.10. TRANSFORMAÇÃO POLINOMIAL: *Uma transformação polinomial de uma linguagem $L_1 \subseteq \Sigma^*$ para uma linguagem $L_2 \subseteq \Sigma^*$, é uma função $f : \Sigma^* \rightarrow \Sigma^*$ que satisfaz as duas condições seguintes:*

- existe um programa para MTD (Máquina de Turing Determinística) de complexidade polinomial que computa f .
- para todo $x \in \Sigma_1^*$, $x \in L_1$ se, e somente se $f(x) \in L_2$.

Se existe uma redução polinomial de L_1 para L_2 , escreve-se $L_1 \propto L_2$, e diz-se que L_1 é redutível a L_2 .

Definição 2.11. (Função Computável) Uma função $f : x \rightarrow y$ é computável quando possui um algoritmo que computa um número racional de $f(x_1, \dots, x_n)$ com números racionais arbitrários x_1, \dots, x_n .

Da mesma forma, se existe uma redução polinomial de um problema Π_1 para um problema Π_2 , escreve-se $\Pi_1 \propto \Pi_2$ e diz-se que Π_1 é redutível a Π_2 .

Convém ressaltar que a redução polinomial preserva a complexidade polinomial.

Existe um grande número de problemas do mundo real para os quais não se conhecem algoritmos computacionais eficientes para resolvê-los e cuja complexidade intrínseca ainda não foi provada, entre eles os NP-Completo.

- **CLASSE NP-COMPLETO:** Um problema Π pertence à classe NP-Completo se: primeiro, pertence a classe NP; e em segundo, todos os problemas da classe NP puderem ser redutíveis a ele.

Os problemas NP-Completo tem a interessante propriedade de que, se um problema NP-Completo tiver solução determinística polinomial, qualquer outro problema NP também tem. Esta propriedade é posta a partir do conceito de redução polinomial. O fato de não estar provado que $P=NP$ e acreditar-se que $P \neq NP$ torna a identificação de um problema na classe NP-Completa uma questão na prática muito importante, pois determina o não conhecimento de solução eficiente para o problema (TOSCANI; VELOSO, 2001).

Há problemas que estão numa classe “meio nebulosa” como o de decidir se duas expressões regulares são equivalentes e o de decidir se uma palavra é gerada por uma certa gramática sensível ao contexto. Esses problemas são polinomialmente redutíveis de problemas NP-Completo, entretanto não se conhece algoritmo não determinístico de tempo polinomial que os resolva, e portanto não se sabe se pertencem ou não a P. Esses problemas são chamados NP-Difíceis (TOSCANI; VELOSO, 2001).

- **CLASSE NP-DIFÍCIL:** Um problema Π é NP-Difícil se para qualquer problema $\Pi' \in NP$, Π' é redutível ao problema Π .

Os problemas NP-Difíceis possuem a propriedade de não serem resolvíveis deterministicamente em tempo polinomial exceto se $P = NP$. Esses problemas são potencialmente mais difíceis que os problemas NP-Completo.

Na presente seção, definições formais e aspectos da teoria da complexidade foram deixados de lado não por serem menos importantes, mas por não apresentarem relevância no direcionamento do trabalho. Optou-se pela apresentação daqueles resultados que efetivamente pudessem contribuir para o entendimento do trabalho a ser desenvolvido. Para uma lista mais completa de resultados sugere-se a consulta à literatura considerada clássica (GAREY; JOHNSON, 1979) (HOROWITZ; SAHNI, 1978) (PAPADIMITRIOU, 1994) (TOSCANI; VELOSO, 2001), entre outros.

2.3 Complexidade de Problemas Intervalares

Conforme descrito na Seção 1.1, um dos primeiros problemas a ser analisado quanto a complexidade computacional foi o PBCI, ou seja, o problema de computar o intervalo imagem de uma função contínua e computável f .

A seguir, apresentamos as definições de topologia, função contínua (RUDIN, 1970) e função computável (KREINOVICH et al, 1998):

Definição 2.12. (Topologia) Uma coleção τ de subconjuntos de um conjunto X é dito ser uma topologia em X se τ possui as seguintes três propriedades: (i) $\emptyset \in \tau$ e $X \in \tau$; (ii) se $V_i \in \tau$, $i = 1, \dots, n$, então $V_1 \cap V_2 \cap \dots \cap V_n \in \tau$; (iii) se $\{V_\alpha\}$ é uma coleção arbitrária de membros de τ (finito, enumerável ou não enumerável), então $\cup_\alpha V_\alpha \in \tau$.

Definição 2.13. (Função Contínua) Se X e Y são espaços topológicos e se f é um mapeamento de X em Y , então f é dito contínuo se $f^{-1}(V)$ é um conjunto aberto em X para todo conjunto aberto V em Y .

Definição 2.14. (Função Intervalar) Seja $f : \mathbf{x} \rightarrow \mathbf{y}$ uma função. Se $\mathbf{x} = \text{Dom}(f) \subseteq \mathbb{IR}$ e $\mathbf{y} = \text{Cod}(f) \subseteq \mathbb{IR}$, então, dizemos que f é uma função intervalar de uma variável intervalar.

Observação: As funções intervalares dependem da sua expressão, pois em \mathbb{IR} , não vale a distributividade da soma em relação a multiplicação (OLIVEIRA; DIVERIO; CLAUDIO, 1997).

Definição 2.15. (Imagem intervalar de uma função real) Sejam f uma função real de variável real e \mathbf{x} um intervalo tal que $\mathbf{x} \subseteq \text{Dom}(f)$ e f é contínua em \mathbf{x} . Define-se a imagem intervalar da função f em \mathbf{x} (ou simplesmente imagem de f em \mathbf{x}) como sendo o intervalo definido por (OLIVEIRA; DIVERIO; CLAUDIO, 1997):

$$I = I(f, \mathbf{x}) = [\min\{f(x) \mid x \in \mathbf{x}\}, \max\{f(x) \mid x \in \mathbf{x}\}].$$

Definição 2.16. (Avaliação intervalar de uma função real) Sejam f uma função real de variável real x e \mathbf{x} um intervalo. Define-se a avaliação intervalar de f em \mathbf{x} (ou extensão intervalar de f) como sendo a função intervalar $F(\mathbf{x})$, definida da seguinte maneira: - cada ocorrência da variável real x é substituída pela variável intervalar \mathbf{x} e cada operação real (+, -, *, /) é substituída pela respectiva operação intervalar de tal modo que, quando $\mathbf{x} = [x, x]$ (intervalo pontual) então $F(\mathbf{x}) = f(x)$.

OBSERVAÇÃO: A definição avaliação intervalar depende da expressão funcional da respectiva função real $f(x)$ correspondente. Por exemplo, em \mathbb{R} tem-se $f(x) = ax^2 + bx + c = c + x(b + xa) = g(x)$, mas em \mathbb{IR} , $F(\mathbf{x}) = A\mathbf{x}^2 + B\mathbf{x} + C$ pode ser diferente de $G(\mathbf{x}) = C + \mathbf{x}(B + \mathbf{x}A)$, dependendo do valor de \mathbf{x} . Porém sempre vale que $I(f, \mathbf{x}) = I(g, \mathbf{x})$ (OLIVEIRA; DIVERIO; CLAUDIO, 1997).

Definição 2.17. (Problema básico da computação intervalar: PBCI) O problema básico da computação intervalar é definido por:

DADOS:

- n intervalos racionais \mathbf{x}_i (intervalos com extremos racionais), e
- uma função contínua e computável f que transforma n números reais x_1, \dots, x_n em um número real $y = f(x_1, \dots, x_n)$.

CALCULAR: o intervalo dos possíveis valores de y :

$$\mathbf{y} = [\underline{y}, \bar{y}] = f(\mathbf{x}_1, \dots, \mathbf{x}_n) = \{y \mid y = f(x_1, \dots, x_n), x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n\}.$$

Moore (1976) foi o primeiro a considerar a computação do intervalo imagem de uma função polinomial com entradas intervalares. Mais tarde, A.A. Gaganov (1985) provou que o problema de calcular a imagem de um polinômio com dados intervalares é de fato computacionalmente intratável. Gaganov baseou-se no problema formalizado por Matiyasevich (1970), de que nenhum algoritmo podia encontrar a imagem de um polinômio com entradas intervalares.

Com o propósito de responder a questão: “É possível ter um algoritmo que sempre calcula a imagem exata (isto é, os extremos inferior e superior do intervalo) em tempo razoável?”, Kreinovich *et al* (1998) realizaram a análise da complexidade do PBCI, verificando que pertence a classe de problemas NP-Difícil.

É importante observar que a NP-dificuldade do PBCI está relacionada com o processamento dos dados de entrada do problema. No PBCI, os dados de entrada de uma função contínua f são valores intervalares, e a forma de calcular o intervalo imagem do PBCI é através da imagem intervalar (OLIVEIRA; DIVERIO; CLAUDIO, 1997).

Na imagem intervalar de f em \mathbf{x} , calcula-se o intervalo definido por $\mathbf{Im} = I(f, \mathbf{x}) = [\min\{f(x)|x \in \mathbf{x}\}, \max\{f(x)|x \in \mathbf{x}\}]$. Neste caso calcula-se os valores dos extremos inferior e superior do intervalo \mathbf{Im} , isto é, no pior caso quando tem-se m número arbitrário de valores e deve-se considerar todos os valores compreendidos entre \underline{x} e \bar{x} do intervalo \mathbf{x} para encontrar estes extremos. Como no problema de computar a imagem de uma função de várias variáveis $f(x_1, \dots, x_n)$, tem-se \mathbf{x}_i intervalos, $i = 1, \dots, n$, calcula-se o extremo inferior \underline{y} avaliando todos os valores de todos os n intervalos (da mesma forma para \bar{y}), totalizando n^m avaliações numéricas (operações fundamentais), o que caracteriza uma complexidade exponencial. Conforme o tamanho da entrada do problema, o tempo de processamento da computação cresce exponencialmente, e pela dificuldade do processamento de todos os dados, não se tem algoritmo de tempo de processamento polinomial conhecido.

Outro fato importante, e que completa a caracterização do problema como NP-Difícil, é que o problema em questão é um problema de decisão, pois deseja-se saber se existe o intervalo $\mathbf{y} = [\underline{y}, \bar{y}]$ que contenha a solução aproximada do problema. Observa-se, ainda, que o problema PBCI é computável; que não foi provado que o problema PBCI é NP; e que a questão em relação a complexidade deste problema está em aberto.

2.3.1 O PBCI com restrições

A seguir apresentamos, de forma sucinta, os principais resultados obtidos por meio da análise da complexidade do PBCI (KREINOVICH *et al*, 1998) (KREINOVICH, 2003).

Como primeira restrição, o PBCI foi analisado considerando o intervalo imagem ε -aproximável.

Definição 2.18. ε -APROXIMÁVEL: *Define-se ε -aproximável como sendo o valor (ε) da diferença entre um valor aproximado \tilde{y} e o valor real y . Para o PBCI ε -aproximado, escolhe-se um número racional $\varepsilon > 0$ e calcula-se os números racionais $\tilde{\underline{y}}$ e $\tilde{\bar{y}}$ ε -aproximável dos extremos da imagem, ou seja, $|\tilde{\underline{y}} - \underline{y}| \leq \varepsilon$ e $|\tilde{\bar{y}} - \bar{y}| \leq \varepsilon$, onde: $\mathbf{y} = [\underline{y}, \bar{y}] = f(\mathbf{x}_1, \dots, \mathbf{x}_n) = \{y|y = f(x_1, \dots, x_n), x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n\}$.*

Tal problema mostrou-se NP-Difícil para o cálculo da imagem exata e ε -aproximada ($\varepsilon > 0$) da função polinomial $f(x_1, \dots, x_n)$.

Como segunda restrição, foi considerado PBCI com entradas intervalares estreitas (ou intervalos com amplitude pequena). As entradas intervalares, usualmente provem de medidas de tal modo que $\mathbf{x}_i = [\underline{x}_i, \bar{x}_i] = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$. Em termos destas

medidas, estes intervalos extensos correspondem a uma baixa exatidão. Para se obter uma exatidão perfeita $\Delta_i = 0$, se considera todas as entradas intervalares degeneradas $\mathbf{x}_i = [\tilde{x}_i, \tilde{x}_i]$, e a imagem desejada y consiste de um simples ponto computável $f(\tilde{x}_1, \dots, \tilde{x}_n)$. É natural esperar que para uma boa exatidão, quando os valores de Δ_i estão próximos de 0, as correspondentes entradas intervalares sejam estreitas e o PBCI assim é mais fácil que o caso da primeira restrição. Entretanto, mesmo considerando entradas intervalares *estreitas*, foi provado que o PBCI permanece NP-Difícil (KREINOVICH et al, 1998).

A terceira restrição, considerada sobre a função polinomial, foi em relação ao número de variáveis, juntamente com a ordem do polinômio: foi verificado que o PBCI pertence a classe P quando a função polinomial tem uma variável ($n = 1$), duas variáveis ($n = 2$) e número fixo de variáveis (n); e NP-Difícil quando a função polinomial possui número arbitrário de variáveis (KREINOVICH et al, 1998).

Além da função polinomial, considerada como caso geral, foram analisados outros tipos de funções no PBCI, sendo elas funções algébricas, função polinomial, splines, função g -polinomial, função g -racional, função racional-complexa

As Tabelas 2.1, 2.2, 2.3, 2.4 e 2.5 apresentam os resultados de complexidade computacional do PBCI considerando os tipos de funções citados anteriormente. Devemos ressaltar que os ífens “-” presentes na Tabela 2.1 significam que estes casos não foram analisados. Na Tabela 2.3 evidenciamos que ficou em aberto a complexidade da função quadrática com coeficientes 0 e 1, e na Tabela 2.4 mostramos que a complexidade ficou em aberto para a função quadrática em geral.

Tabela 2.1: Tipos de Funções, Complexidade do PBCI com n Variáveis e com Polinômio de Ordem Fixa

Tipos de Funções	Complexidade do PBCI com n Variáveis	Complexidade do PBCI com Polinômio de Ordem Fixa
Função Polinomial com coeficientes racionais	Polinomial ($n = 1$) Polinomial ($n = 2$) Polinomial (n fixo) NP-Difícil (n arbitrário)	Tempo Linear (Linear) Polinomial (Quadrático) Polinomial (Cúbico e ordem superior)
Função Algébrica	Polinomial ($n = 1$) Polinomial ($n = 2$) Polinomial (n fixo)	-
Função Polinomial com descrição Compacta	Exponencial ($n = 1$) Exponencial ($n = 2$) Exponencial (n fixo) Exponencial (n arbitrário)	-
Splines $f(x)$	-	NP-Difícil (Linear) NP-Difícil (Quadrática) NP-Difícil (Cúbico e ordem superior)
Exp-polinomial	Exponencial ($n \geq 1$)	-
Seno-polinomial	NP-Difícil ($n = 1$) NP-Difícil ($n \geq 2$)	-
Funções Racionais	Polinomial ($n = 1$) Exponencial ($n \geq 2$)	-
Função Racional-complexa	Exponencial ($n = 1$) Exponencial ($n \geq 2$)	-

Tabela 2.2: Função Polinomial de Ordem Fixa e Complexidade do PBCI com Polinômio de Ordem Fixa

Função Polinomial de Ordem Fixa	Complexidade do PBCI com Polinômio de Ordem Fixa
Função linear	Tempo Linear
Função quadrática	NP-Difícil
Função cúbica e ordem superior	NP-Difícil

Tabela 2.3: Tipos de Funções Polinomiais e Complexidade do PBCI com Coeficientes Restritos

Tipos de Funções Polinomiais	Complexidade do PBCI com Coeficientes Restritos
Função quadrática com coeficientes 0, 1, 2 e 3	NP-Difícil
Função quadrática com coeficientes 0 e 1	-
Função cúbica (e ordem superior) com coeficientes 0 e 1	NP-Difícil

Tabela 2.4: Tipos de Funções Polinomiais com x_i arbitrários e Complexidade do PBCI com Polinômios Fixos e Intervalos Estreitos

Tipos de Funções Polinomiais com x_i arbitrários	Complexidade do PBCI com Polinômios Fixos e Intervalos Estreitos
Linear $f_n(x_1, \dots, x_n)$	Tempo Linear
Quadrática $f_n(x_1, \dots, x_n)$	-
Cúbica (e ordem superior) $f_n(x_1, \dots, x_n)$	NP-Difícil

Tabela 2.5: Tipos de Funções Polinomiais com x_i fixos e Complexidade do PBCI com Polinômios Arbitrários e Intervalos Fixos

Tipos de Funções Polinomiais com x_i fixos	Complexidade do PBCI com Polinômios Arbitrários e Intervalos Fixos
f_n linear	Tempo Linear
f_n quadrática	NP-Difícil
f_n cúbica (e ordem superior)	NP-Difícil

2.4 Comentários Finais

O presente capítulo teve como objetivo revisar conceitos conhecidos da área de Complexidade Computacional, os quais formam um conjunto mínimo de informações necessárias para o desenvolvimento do trabalho. Além disso, descrevem-se os principais resultados de complexidade de problemas da computação intervalar.

Na maioria dos casos considerados sobre o PBCI, verificou-se que o mesmo pertence à classe de problemas NP-Difícil. Por conseguinte, conclui-se que não se pode esperar um algoritmo que resolva todos os problemas da computação intervalar em tempo razoável.

Entretanto, resultados de NP-Dificuldade também tem seu lado investigativo, ou seja, sempre existirá o desafio de encontrar algoritmos para os problemas que são NP-Difíceis, mesmo porque, a questão em relação a complexidade destes problemas está em aberto.

Nesterov (1996), salienta que, do ponto de vista técnico, muitos problemas práticos e importantes da computação intervalar são conhecidos como computacionalmente intratáveis, ou conhecidos como computacionalmente razoáveis. Existem poucos problemas dos quais não se sabe se são razoáveis ou não. Assim, pode-se ter a impressão de que esta área de pesquisa está quase finalizada, o que é uma falsa impressão. Acredita-se que existem muitos problemas da computação intervalar em aberto e que as pesquisas nesta área mostram-se recentes e atuais.

O próximo capítulo apresentará uma abordagem intervalar construída para indicadores estatísticos. A definição de indicadores estatísticos em nível intervalar faz-se necessária devido a ausência de tais definições na literatura.

3 ESTATÍSTICA DESCRITIVA INTERVALAR

O presente capítulo tem como objetivo apresentar: uma forma de representação dos valores reais de uma amostra aleatória em valores intervalares, a relação de ordem quando os intervalos são todos disjuntos ou todos encaixados e uma proposta intervalar para os indicadores estatísticos usuais ou os estatísticos descritivos (FERSON et al., 2002)(FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002)(FERSON et al., 2002)(KREINOVICH et al, 1998).

Os indicadores estatísticos com dados reais, como definidos na literatura (CHRISTMANN, 1978), encontram-se no Anexo A.

3.1 Representação dos reais em intervalos

Na estatística, a computação do valor das medidas de tendência central média, mediana e moda, das medidas de dispersão amplitude total, variância, desvio padrão, coeficiente de variação, covariância e coeficiente de correlação e das separatrizes quartil, decil e percentil considera, nas suas fórmulas (descritas no anexo A), n variáveis reais x_i com n entradas reais. Ou seja, realiza-se o processamento de dados aplicando operações aritméticas reais (+, -, *, /) sobre os dados de entrada para obter os valores destes indicadores estatísticos.

Na abordagem intervalar, desenvolvida na presente tese, considera-se o seguinte: sejam X e Y variáveis aleatórias populacionais seguindo a lei da probabilidade $f_X(\theta_X)$ e $f_Y(\theta_Y)$, respectivamente. Sejam μ_x , μ_y , σ_x^2 e σ_y^2 as esperanças matemáticas e variâncias das variáveis aleatórias X e Y . Adicionalmente sejam $\{x_1, \dots, x_n\}$ e $\{y_1, \dots, y_n\}$ amostras aleatórias de X e Y .

Uma forma de se considerar os erros numéricos inerentes aos valores x_1, \dots, x_n e y_1, \dots, y_n é considerar que, para $i = 1, \dots, n$ e $j = 1, \dots, n$, $x_i \in [\underline{x}_i, \bar{x}_i]$ e $y_j \in [\underline{y}_j, \bar{y}_j]$.

Neste caso, consideram-se os dados de entrada

$$\{x_1, \dots, x_n\}, x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n,$$

ou seja, domínios intervalares onde $\mathbf{x}_1 = [\underline{x}_1, \bar{x}_1], \dots, \mathbf{x}_n = [\underline{x}_n, \bar{x}_n]$.

Representa-se cada um dos n valores reais $\{x_1, \dots, x_n\}$ de uma amostra aleatória de uma população por intervalos $\mathbf{x}_i = [x_i - \delta, x_i + \delta]$, onde δ é a margem de precisão escolhida para estes valores.

Nestas situações, para diferentes valores possíveis $x_i \in \mathbf{x}_i$, obtém-se diferentes valores da média me , mediana md , moda mo , amplitude total at , variância va , desvio padrão dp , coeficiente de variação cv , covariância co , coeficiente de correlação cc , quartil q , decil d e percentil p .

Considerando que tem-se dados de entrada intervalares, deseja-se conhecer quais são os intervalos da média intervalar ME_v , mediana intervalar MD_v , moda intervalar MO_v , amplitude total intervalar AT_v , variância intervalar VA_v , desvio padrão intervalar DP_v , coeficiente de variação intervalar CV_v , covariância intervalar CO_v , coeficiente de correlação intervalar CC_v , quartil intervalar Q_v , decil intervalar D_v e percentil intervalar P_v dos possíveis valores reais da média me , mediana md , moda mo , amplitude total at , variância va , desvio padrão dp , coeficiente de variação cv , covariância co , coeficiente de correlação cc , quartil q , decil d e percentil p .

3.2 Relação de ordem em Intervalos

A abordagem intervalar elaborada para os indicadores estatísticos permite a ocorrência de intervalos disjuntos ou encaixados (não disjuntos). Este fato é devido aos valores reais, de amostras aleatórias de uma população, serem representados através de intervalos para considerar erros numéricos de medidas.

Para calcular os valores dos indicadores estatísticos mediana, amplitude total e as separatrizes quartil, decil e percentil, os dados devem estar ordenados em ordem ascendente.

Levando em consideração que os dados intervalares x_1, \dots, x_n podem ser todos disjuntos, ou todos encaixados, ou alguns intervalos disjuntos e alguns encaixados, ordena-se os dados intervalares de acordo com os seguintes casos:

- **TODOS OS INTERVALOS DISJUNTOS:** considera-se a ordem intervalar de Kulisch (1981):

$$x_1 \leq x_2 \leq \dots \leq x_n,$$

onde $x_1 \leq x_2 \iff \underline{x}_1 \leq \underline{x}_2 \wedge \bar{x}_1 \leq \bar{x}_2$;

- **TODOS OS INTERVALOS ENCAIXADOS:** neste caso, para a ordenação dos elementos, considera-se a ordem de inclusão (KULISCH; MIRANKER, 1981):

$$x_{(1)} \subseteq x_{(2)} \subseteq \dots \subseteq x_{(n)},$$

onde $x_{(1)} \subseteq x_{(2)} \iff \underline{x}_1 \geq \underline{x}_2 \wedge \bar{x}_1 \leq \bar{x}_2$.

- **INTERVALOS DISJUNTOS E ENCAIXADOS AO MESMO TEMPO:** não existe na literatura a relação de ordem para este caso, sendo tema para trabalhos futuros (Seção 6.2).

Observa-se que as definições intervalares elaboradas para os indicadores estatísticos, média, moda, variância, desvio padrão, coeficiente de variação, covariância e coeficiente de correlação, suportam dados intervalares todos disjuntos ou todos encaixados.

O cálculo da mediana e demais separatrizes está em estudo quando nem todos os intervalos são ordenáveis segundo a ordem “ \leq ” ou encaixada “ \subseteq ”.

Devido a representação dos reais em intervalos, abordada neste trabalho (Seção 3.1), todos os intervalos terão o mesmo diâmetro, podendo não ocorrer, portanto, intervalos totalmente contidos uns nos outros.

3.3 Nível Intervalar dos Indicadores Estatísticos

As definições dos indicadores estatísticos média intervalar, moda intervalar, variância intervalar e desvio padrão intervalar foram retiradas de Campos e Faria (1988), os indicadores covariância intervalar e coeficiente de correlação intervalar foram pesquisadas em (FERSON et al., 2002) e (FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002). As definições dos indicadores coeficiente de variação intervalar, mediana intervalar, quartil, decil e percentil intervalares foram formuladas pela autora da tese.

As fórmulas com abordagem intervalar, dos indicadores estatísticos, foram elaboradas a partir das fórmulas definidas na literatura (CHRISTMANN, 1978), trocando as variáveis reais por variáveis intervalares. Em consequência, as operações aritméticas reais também foram substituídas por operações intervalares (MOORE, 1966).

3.3.1 Média aritmética Intervalar

Considere

$$\sum_{i=1}^n \mathbf{x}_i = \sum_{i=1}^n [\underline{x}_i, \bar{x}_i] = \left[\sum_{i=1}^n \underline{x}_i, \sum_{i=1}^n \bar{x}_i \right].$$

Logo

$$\mathbf{ME}_v = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \left[\sum_{i=1}^n \underline{x}_i, \sum_{i=1}^n \bar{x}_i \right].$$

Esta última expressão pode ser resolvida de duas formas distintas: i) usando a multiplicação de um real por um intervalo, como $1/(n) > 0$ não haverá alteração nos limites do intervalo resultante e ii) definindo o intervalo degenerado $[1/(n), 1/(n)]$ e usando a multiplicação de intervalos. Em ambos os casos tem-se:

Definição 3.1. MÉDIA ARITMÉTICA INTERVALAR

$$\mathbf{ME}_v = [\underline{me}, \bar{me}] = \frac{1}{n} \left[\sum_{i=1}^n \underline{x}_i, \sum_{i=1}^n \bar{x}_i \right].$$

Se for necessário um único valor para a média, pode-se calcular o ponto médio (MOORE, 1979) do intervalo \mathbf{ME}_v , usando

$$m(\mathbf{ME}_v) = (\underline{me} + \bar{me})/2.$$

3.3.2 Mediana Intervalar

Primeiramente deve-se ordenar os dados intervalares $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ de acordo com os casos descritos na Seção 3.2.

A mediana intervalar \mathbf{MD}_v corresponderá ao intervalo mediano. Calcula-se a posição deste valor intervalar da mesma forma que é calculada com valores reais.

Definição 3.2. MEDIANA INTERVALAR

$$\mathbf{MD}_v = [\underline{md}, \bar{md}] = \begin{cases} \left(\mathbf{x}_{(\frac{n}{2})} + \mathbf{x}_{(\frac{n}{2}+1)} \right) / 2 & \text{se } n \text{ for par,} \\ \mathbf{x}_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar.} \end{cases}$$

3.3.3 Moda Intervalar

Define-se o intervalo modal por

Definição 3.3. MODA INTERVALAR. *Se existe um valor modal para os dados reais, então*

$$\mathbf{MO}_v = [\underline{mo}, \overline{mo}] = [mo_{01 \leq i \leq n} \{\underline{x}_i\}, mo_{01 \leq i \leq n} \{\overline{x}_i\}].$$

Neste caso, deve-se primeiro representar os valores reais da amostra em valores intervalares e, procurar pelo valor modal intervalar nos valores intervalares da amostra.

3.3.4 Amplitude total Intervalar

Com os dados intervalares ordenados segundo a ordem para intervalos disjuntos “ \leq ” ou intervalos encaixados “ \subseteq ” (Seção 3.2), considera-se :

- Mínimo Intervalar: $MI = [\underline{mi}, \overline{mi}] = [\min_{1 \leq i \leq n} \{\underline{x}_i\}, \min_{1 \leq i \leq n} \{\overline{x}_i\}]$,
- Máximo Intervalar: $MA = [\underline{ma}, \overline{ma}] = [\max_{1 \leq i \leq n} \{\underline{x}_i\}, \max_{1 \leq i \leq n} \{\overline{x}_i\}]$.

Define-se a amplitude total intervalar como sendo:

Definição 3.4. AMPLITUDE TOTAL INTERVALAR

$$\mathbf{AT}_v = [\underline{at}, \overline{at}] = \begin{cases} [\underline{ma} - \overline{mi}, \overline{ma} - \underline{mi}] & \text{se } \underline{ma} \geq \overline{mi}, \\ [0, \overline{ma} - \underline{mi}] & \text{se } \underline{ma} < \overline{mi}. \end{cases}$$

3.3.5 Variância Intervalar

Tem-se que

$$\mathbf{x}_i - \mathbf{ME} = [\underline{x}_i, \overline{x}_i] - [\underline{me}, \overline{me}] = [\underline{x}_i - \overline{me}, \overline{x}_i - \underline{me}],$$

e

$$(\mathbf{x}_i - \mathbf{ME})^2 = \begin{cases} [(\underline{x}_i - \overline{me})^2, (\overline{x}_i - \underline{me})^2] & \text{se } \underline{x}_i - \overline{me} \geq 0, \\ [(\overline{x}_i - \underline{me})^2, (\underline{x}_i - \overline{me})^2] & \text{se } \overline{x}_i - \underline{me} < 0, \\ [0, \max((\underline{x}_i - \overline{me})^2, (\overline{x}_i - \underline{me})^2)] & \text{se } \underline{x}_i - \overline{me} < 0 \leq \overline{x}_i - \underline{me}. \end{cases}$$

Então define-se a variância intervalar como:

Definição 3.5. VARIÂNCIA INTERVALAR

$$\mathbf{VA}_v = [\underline{va}, \overline{va}] = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{ME})^2.$$

3.3.6 Desvio padrão Intervalar

Como o desvio padrão é a raiz quadrada positiva da variância, a obtenção da raiz quadrada de cada limite do intervalo \mathbf{VA}_v , fornecerá o intervalo \mathbf{DP}_v do desvio padrão intervalar:

Definição 3.6. DESVIO PADRÃO INTERVALAR

$$\mathbf{DP}_v = [\underline{dp}, \overline{dp}] = \sqrt{\mathbf{VA}_v} = +\sqrt{[\underline{va}, \overline{va}]} = [+ \sqrt{\underline{va}}, + \sqrt{\overline{va}}].$$

3.3.7 Coeficiente de variação Intervalar

Definição 3.7. (COEFICIENTE DE VARIAÇÃO INTERVALAR) Se $0 \notin [\underline{me}, \overline{me}]$, então

$$CV_v = [\underline{cv}, \overline{cv}] = \frac{DP_v}{ME_v} = \frac{[dp, \overline{dp}]}{[\underline{me}, \overline{me}]} = [dp, \overline{dp}] * \left(\frac{1}{[\underline{me}, \overline{me}]} \right).$$

3.3.8 Covariância Intervalar

Tem-se que:

$$x_i - ME_X = [\underline{x}_i, \overline{x}_i] - [\underline{me}_X, \overline{me}_X] = [\underline{x}_i - \overline{me}_X, \overline{x}_i - \underline{me}_X]$$

e

$$y_i - ME_Y = [\underline{y}_i, \overline{y}_i] - [\underline{me}_Y, \overline{me}_Y] = [\underline{y}_i - \overline{me}_Y, \overline{y}_i - \underline{me}_Y],$$

onde ME_X é a média intervalar dos valores de X e ME_Y a média intervalar dos valores de Y .

Então, define-se a covariância intervalar como:

Definição 3.8. COVARIÂNCIA INTERVALAR

$$CO_v = [\underline{co}, \overline{co}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - ME_X)(y_i - ME_Y).$$

3.3.9 Coeficiente de correlação Intervalar

O coeficiente de correlação intervalar é definido da seguinte maneira:

Definição 3.9. (COEFICIENTE DE CORRELAÇÃO INTERVALAR) Se $0 \notin DP_X DP_Y$, então

$$CC_v = [\underline{cc}, \overline{cc}] = \frac{CO_v}{DP_X DP_Y} = [\underline{co}, \overline{co}] * \left(\frac{1}{[dp_X, \overline{dp}_X][dp_Y, \overline{dp}_Y]} \right),$$

onde CO é covariância intervalar, DP_X o desvio padrão intervalar dos valores de X e DP_Y o desvio padrão intervalar dos valores de Y .

3.3.10 Separatrizes Intervalares

As separatrizes quartis, decis e percentis se comportam de forma similar à separatriz mediana, pois para estes indicadores deve-se ordenar os valores da amostra e depois dividir em partes iguais a seqüência de valores: Quartis divide em 4 partes (calculando 3 quartis), Decis divide em 10 partes (calculando 9 decis) e Percentis em 100 partes (precisando de 99 valores que separam uma série).

Para calcular os valores da mediana intervalar MD_v (também considerada uma separatriz), do quartil intervalar $Q_v = [q, \overline{q}]$, decil intervalar $D_v = [d, \overline{d}]$ e percentil intervalar $P_v = [p, \overline{p}]$ não é necessário realizar operações aritméticas entre os valores da amostra. Apenas deve-se classificar os valores da amostra em ordem ascendente. Como tem-se amostra com valores intervalares, deve-se ordenar os valores conforme descrito na Seção 3.2. Os valores intervalares das separatrizes são encontrados conforme a posição calculada na seqüência de valores intervalares $\{x_1, \dots, x_n\}$.

A fórmula para calcular as posições dos valores das separatrizes é devido a Jain (1991):

$$pos = (n - 1)\alpha + 1, \quad (3.1)$$

onde pos é a posição do valor da separatriz, n é o número de valores da amostra e $\alpha = \frac{ns}{nd}$ (ns representa o número ordinal da separatriz desejada e nd o número de divisões de cada separatriz (na mediana o $nd = 2$, no quartil o $nd = 4$, no decil o $nd = 10$ e no percentil o $nd = 100$)).

Exemplo: Se deseja-se calcular D_8 , isto é, o oitavo decil intervalar de uma amostra com 30 valores intervalares, tem-se $n = 30$ e $\alpha = \frac{8}{10}$. Assim, a posição do oitavo decil é $pos = 24.2 = 24^0$.

A seguir apresentamos o Quadro 3.1 contendo todas as expressões dos indicadores intervalares definidos no presente capítulo.

Média Intervalar:	$ME_v = [\underline{me}, \overline{me}] = \frac{1}{n} \left[\sum_{i=1}^n x_i, \sum_{i=1}^n \bar{x}_i \right]$
Mediana Intervalar:	$MD_v = [\underline{md}, \overline{md}] = \begin{cases} \frac{(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})}{2}, & \text{se } n \text{ for par} \\ (x_{(\frac{n+1}{2})}), & \text{se } n \text{ for ímpar.} \end{cases}$
Moda Intervalar:	$MO_v = [\underline{mo}, \overline{mo}] = [mo_{01 \leq i \leq n} \{x_i\}, mo_{01 \leq i \leq n} \{\bar{x}_i\}]$
Amplitude total Intervalar:	$AT_v = [\underline{at}, \overline{at}] = \begin{cases} [\underline{ma} - \overline{mi}, \overline{ma} - \underline{mi}], & \text{se } \underline{ma} \geq \overline{mi} \\ [0, \overline{ma} - \underline{mi}], & \text{se } \underline{ma} < \overline{mi}. \end{cases}$
Variância Intervalar:	$VA_v = [\underline{va}, \overline{va}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - ME)^2$
Desvio padrão Intervalar:	$DP_v = [\underline{dp}, \overline{dp}] = \sqrt{VA} = +\sqrt{[\underline{va}, \overline{va}]} = [+ \sqrt{\underline{va}}, + \sqrt{\overline{va}}]$
Coefficiente de variação Intervalar:	$CV_v = [\underline{cv}, \overline{cv}] = \frac{DP}{ME} = \frac{[\underline{dp}, \overline{dp}]}{[\underline{me}, \overline{me}]}$
Covariância Intervalar:	$CO_v = [\underline{co}, \overline{co}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - ME_X)(y_i - ME_Y)$
Coefficiente de correlação Intervalar:	$CC_v = [\underline{cc}, \overline{cc}] = \frac{CO}{DP_X DP_Y} = \left(\frac{[\underline{co}, \overline{co}]}{[\underline{dp}_X, \overline{dp}_X][\underline{dp}_Y, \overline{dp}_Y]} \right)$
Separatrizes Intervalares:	$Q_v = [q, \bar{q}], D_v = [d, \bar{d}], P_v = [p, \bar{p}], pos = (n - 1)\alpha + 1$

Figura 3.1: Formulário dos Indicadores Estatísticos Intervalares

3.3.11 Estado da Arte dos problemas da Estatística Descritiva Intervalar

Neste caso foi considerado como entrada x_1, \dots, x_n com $x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n$, ou seja, domínios intervalares onde $\mathbf{x}_1 = [x_1, \bar{x}_1], \dots, \mathbf{x}_n = [x_n, \bar{x}_n]$.

Ferson *et al* (2002) provaram que o problema de computar o extremo superior \overline{va} da variância intervalar VA_v é NP-Difícil; para computar o extremo inferior \underline{va} da variância intervalar, desenvolveram um algoritmo razoável, mais precisamente um algoritmo de tempo quadrático, ou seja, um algoritmo que requer $O(n^2)$ passos computacionais para n intervalos de entrada $\mathbf{x}_i = [x_i, \bar{x}_i]$. Considerando o caso quando os pontos médios

$\tilde{x}_i = \frac{(x_i + \bar{x}_i)}{2}$ de todos os intervalos são diferentes uns dos outros, no sentido em que os intervalos estreitos $[\frac{\tilde{x}_i + \Delta_i}{n}, \frac{\tilde{x}_i - \Delta_i}{n}]$ (onde $\Delta_i = \frac{(x_i - \bar{x}_i)}{2}$ é o raio do intervalo) não se interceptam uns com os outros, propõem um algoritmo que computa \overline{va} em tempo quadrático.

Sobre o problema de computar o desvio padrão intervalar \mathbf{DP}_v com n variáveis reais x_i e n entradas intervalares $\mathbf{x}_i = [x_i, \bar{x}_i]$, não é realizada nenhuma análise de complexidade devido aos resultados de complexidade serem equivalentes aos obtidos da análise da complexidade do problema da variância intervalar.

Em (FERSON et al., 2002)(FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002), é provado que o problema de computar os extremos inferior \underline{co} e superior \overline{co} do intervalo da covariância intervalar \mathbf{CO}_v , com n entradas intervalares $\mathbf{x}_i = [x_i, \bar{x}_i]$, $\mathbf{y}_i = [y_i, \bar{y}_i]$ é pertencente à classe de problemas NP-Difícil. O problema de computar o intervalo imagem da correlação intervalar, com n entradas intervalares $[x_i, \bar{x}_i]$, $[y_i, \bar{y}_i]$ também é NP-Difícil. Ferson *et al* (2002) comentam que os problemas de computar a covariância e a correlação por serem NP-Difíceis, não existe algoritmo razoável que pode sempre computar os limites desejados para \mathbf{CO}_v e \mathbf{CC}_v . Resultado similar de NP-dificuldade foi provado para a variância intervalar, porém neste caso, desenvolveram um algoritmo razoável que funciona em alguns casos. Neste sentido, algoritmos similares para covariância e correlação intervalares podem ser desenvolvidos, considerando casos particulares em relação aos dados de entrada intervalares.

Kreinovich (2003) mostra que o problema de computar o extremo superior da variância intervalar \overline{va} é NP-Difícil e para computar o extremo inferior \underline{va} fornece um algoritmo razoável de tempo quadrático. Ainda, considerando condições sobre os dados de entrada, em particular intervalos incertos para manter privacidade no banco de dados estatísticos, fornece algoritmos razoáveis que computam \underline{va} .

Na procura de melhores resultados sobre a computação do valor intervalar da variância, Wu *et al* (2003) descrevem novos algoritmos para o processamento estatístico de dados intervalares, algoritmos em que, adicionando-se um novo valor nos dados de entrada, requerem somente $O(n)$ passos computacionais. Ou seja, se temos um novo resultado de medição, podemos modificar o valor da variância em tempo constante.

Em recente pesquisa, Xiang *et al* (2004) consideraram os casos quando todas as medições são feitas usando o mesmo instrumento de medida e quando é limitado o número de diferentes instrumentos de medição. Para o primeiro caso, quando todas as medições são feitas com o mesmo instrumento, nenhuma de duas entradas intervalares \mathbf{x}_i podem ser o próprio subconjunto de uma outra entrada, e como resultado, foi encontrado o intervalo “exato” da variância em tempo $O(n \log n)$. Já no segundo caso, quando se limita o número ($m > 1$) de diferentes tipos de instrumentos de medição, foi verificado que é possível computar o intervalo da variância \mathbf{VA}_v em tempo polinomial $O(n^{m+1})$.

Ferson *et al* (2004) recentemente pesquisaram sobre a computação dos extremos do intervalo da variância considerando população finita de dados intervalares. Os resultados obtidos neste caso são os mesmos encontrados quando considerados amostras intervalares como dados de entrada (FERSON et al., 2002)(FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002).

3.4 Comentários Finais

Apresentou-se neste capítulo uma abordagem intervalar para os indicadores estatísticos média, mediana, moda, amplitude total, variância, desvio padrão, coeficiente de variação, covariância, coeficiente de correlação e separatrizes quartil, decil e percentil.

Espera-se que com definições intervalares para os indicadores estatísticos, tenha-se uma definição para a estatística descritiva a nível intervalar.

Com o objetivo de considerar erros numéricos inerentes aos valores de uma amostra aleatória, realizou-se a representação dos valores reais $\{x_1, \dots, x_n\}$ em valores intervalares considerando uma margem de precisão δ .

Considerando que os valores intervalares possam ser todos disjuntos ou todos encaixados, descreveu-se a relação de ordem considerada neste trabalho para cada caso.

Finalizando o capítulo, apresentou-se o estado da arte dos problemas da estatística descritiva intervalar.

O próximo capítulo apresenta a análise da complexidade dos problemas de computar os valores dos indicadores estatísticos com entradas intervalares.

4 ANÁLISE DA COMPLEXIDADE DOS PROBLEMAS DE ESTATÍSTICA DESCRITIVA INTERVALAR

O presente capítulo investiga a complexidade computacional dos problemas de computar os valores das medidas de tendência central média, mediana e moda, das medidas de dispersão amplitude, variância, desvio padrão, coeficiente de variação, covariância, coeficiente de correlação e das separatrizes quartil, decil e percentil com entradas intervalares.

4.1 Redefinindo os Problemas

Na literatura (FERSON et al., 2002)(FERSON; GINZBURG; KREINOVICH; LONGPRE; AVILES, 2002)(KREINOVICH, 2003) verifica-se que para o problema de computar o intervalo da variância intervalar, considera-se n variáveis reais x e, como entrada, n valores reais x_1, \dots, x_n onde $x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n$, sendo $\mathbf{x}_i = [\underline{x}_i, \bar{x}_i] \in \mathbb{IR}, i = 1, \dots, n$. Para os problemas dos indicadores estatísticos covariância intervalar e coeficiente de correlação intervalar, considera-se n variáveis reais x e n variáveis reais y e, como entrada, n valores reais x_1, \dots, x_n , onde $x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n$ e n valores reais y_1, \dots, y_n , onde $y_1 \in \mathbf{y}_1, \dots, y_n \in \mathbf{y}_n$. A busca da solução destes problemas ocorre através da computação da imagem intervalar.

Dessa forma, pode-se afirmar que os problemas definidos na literatura consideram intervalos \mathbf{x}_i contidos no domínio da função f , isto é, $Dom(f) = \mathbb{R}^n$, onde n é o número de argumentos da função.

Neste trabalho, considera-se como domínio dos problemas de computar os intervalos dos indicadores estatísticos intervalares, um conjunto de valores intervalares, ou seja, $Dom(f) = \mathbb{IR}^n$ (n é o número de argumentos da função). Ressalta-se que, quando se altera a definição do domínio (dados de entrada) de um problema, define-se um novo problema.

Sendo assim, a presente tese investiga a complexidade computacional dos problemas de computar os intervalos dos indicadores estatísticos considerando variáveis intervalares x e y , n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$ e n valores intervalares $\mathbf{y}_1, \dots, \mathbf{y}_n$. Para a computação dos intervalos dos indicadores estatísticos intervalares, considera-se a extensão intervalar (MOORE, 1979).

Na bibliografia pesquisada, ressaltamos três aspectos que devem ser considerados: (i) apenas os indicadores estatísticos variância, covariância e coeficiente de correlação foram abordados em nível intervalar; (ii) por meio da análise da complexidade computacional realizada nos problemas de computar os intervalos das medidas de dispersão variância, covariância e coeficiente de correlação com variáveis reais x e y e valores intervalares x_1, \dots, x_n , onde $x_1 \in \mathbf{x}_1, \dots, x_n \in \mathbf{x}_n$ e y_1, \dots, y_n , onde $y_1 \in \mathbf{y}_1, \dots, y_n \in \mathbf{y}_n$, prova-

se que estes são problemas pertencentes à classe de problemas NP-Difícil, considerando a computação da imagem intervalar para obter soluções com intervalos de amplitude pequena e (iii) ao utilizar a computação intervalar na solução destes problemas, segundo Ferson *et al* (2002), sempre obtém-se soluções com intervalos superestimados (ou intervalos com amplitude grande).

4.2 Complexidade dos Problemas de computar Indicadores Estatísticos Intervalares

Nesta seção investiga-se a complexidade dos problemas de computar os valores de medidas de tendência central, de medidas de dispersão e separatrizes com entradas intervalares. Realiza-se uma comparação entre nossos resultados com os encontrados na bibliografia.

É importante ressaltar que os problemas de computar os valores de medidas de tendência central intervalar, dispersão intervalar e separatrizes intervalares são problemas de localização associado ao problema de decisão, pois deseja-se saber se os intervalos solução contém a solução real para estes problemas.

A partir das expressões definidas para os indicadores estatísticos descritivos intervalares (Seção 3.3), desenvolve-se algoritmos para a solução dos problemas de computar os intervalos de medidas de tendência central intervalar, dispersão intervalar e separatrizes intervalares. Estes algoritmos utilizam a aritmética intervalar definida por Moore (1966) e a extensão intervalar (MOORE, 1979) como método da computação intervalar. Para todos os algoritmos propostos, realiza-se a análise da complexidade para determinar a ordem de complexidade dos mesmos (todos os algoritmos e a análise da complexidade encontram-se no Anexo B).

Neste contexto, salienta-se que existe a possibilidade de desenvolver algoritmos para ambientes não intervalares, e neste caso, aconselha-se a utilização da multiplicação intervalar definida por Vaccaro (2001). A multiplicação intervalar definida em (VACCARO, 2001) baseia-se em um mapeamento particular de $\mathbb{I}\mathbb{R}$ que permite isolar características importantes dos intervalos de cada região. Este mapeamento considera não somente os sinais dos extremos dos intervalos, conforme sugerido por Moore (1966), como também o sinal do ponto médio do intervalo. Através deste mapeamento define-se expressões que permitem obter a multiplicação de intervalos mais exatos que na formulação proposta por Moore.

Os algoritmos propostos foram construídos para serem executados em ambientes intervalares como Maple Intervalar (MAPLE, 2005) e/ou IntLab (MATLAB, 2005).

Em alguns algoritmos propostos para os problemas de medidas de tendência central intervalar e dispersão intervalar, temos que efetuar comparações entre os extremos do intervalo, isto é, sejam x e y intervalos, consideramos $x < y$ se $sup(x)$ é menor que $inf(y)$, isto reduz a comparações de reais.

4.2.1 Problema: Média Intervalar

PROBLEMA: Computar o intervalo $ME_v = [me, \overline{me}]$ da média aritmética intervalar com variáveis intervalares x e n valores intervalares x_1, \dots, x_n .

Para computar a média aritmética intervalar com a expressão definida na Seção 3.3.1, projeta-se o algoritmo MI, descrito no anexo B.1, Figura B.1.

COMPLEXIDADE: Como o problema em questão é um problema de localização asso-

ciado ao problema de decisão e o algoritmo MI é um algoritmo determinístico de tempo linear, pode-se afirmar que o problema de computar o intervalo da média aritmética intervalar pertence à classe de problemas P.

Nas pesquisas realizadas sobre este problema, verifica-se que o indicador estatístico média intervalar não foi abordado como um problema da computação intervalar estatística. Conseqüentemente, não se encontra resultados de complexidade do mesmo. Segundo Ferson *et al* (2002) a computação intervalar fornece o intervalo “exato” para o caso quando a função $f(x_1, \dots, x_n)$ é uma simples operação aritmética, ou seja, obtém-se o mesmo intervalo da média ME_v ao utilizar a computação da imagem intervalar ou a extensão intervalar.

4.2.2 Problema: Mediana Intervalar

PROBLEMA: Computar o intervalo $MD_v = [\underline{md}, \overline{md}]$ da mediana intervalar com variáveis intervalares \mathbf{x} e n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Calcula-se a mediana da mesma forma que as separatrizes, isto é, através do cálculo da posição do elemento, equação 3.1. Desenvolve-se o algoritmo SI (anexo B.4, Figura B.4) baseado na expressão definida na Seção 3.3.

COMPLEXIDADE: O problema de computar o intervalo da mediana intervalar é resolvido por um algoritmo determinístico de ordem $O(n^2)$. Por ser um problema de localização associado ao problema de decisão, o problema em questão pertence a classe de problemas P.

A mediana não requer operações aritméticas entre intervalos, pois é o valor que fica na posição central de uma seqüência de valores intervalares da amostra (classificados na ordem ascendente).

O resultado de complexidade do algoritmo SI é devido a operação fundamental considerada ser a comparação entre os valores, isto é, tem-se um problema de classificação onde o limite de complexidade no pior caso é $O(n^2)$.

Na literatura não foram encontrados resultados de análise de complexidade do problema de computar o valor da mediana intervalar.

4.2.3 Problema: Moda Intervalar

PROBLEMA: Computar o intervalo $MO_v = [\underline{mo}, \overline{mo}]$ da moda intervalar com variáveis intervalares \mathbf{x} e n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Para computar o valor modal intervalar com a expressão definida na Seção 3.3.3, propõe-se o algoritmo MOI (descrito no anexo B.3, Figura B.3). Os dados intervalares são ordenados na ordem ascendente pelo algoritmo B.2.

COMPLEXIDADE: O problema de computar o valor modal intervalar com dados intervalares pertence a classe de problemas P, pois o problema em estudo é um problema de localização associado ao problema de decisão e o algoritmo (MOI), proposto como solução, é um algoritmo determinístico de tempo quadrático ($O(n^2)$).

A moda, bem como a mediana, não requer operações aritméticas entre intervalos. As pesquisas realizadas sobre problemas de computação intervalar estatística não abordaram o problema de computar o valor modal intervalar. Conseqüentemente, não é possível realizar uma comparação com resultados já existentes.

4.2.4 Problema: Separatrizes - Quartil, Decil e Percentil Intervalares

PROBLEMA: Computar os intervalos $Q_v = [q, \bar{q}]$ do quartil intervalar, $D_v = [d, \bar{d}]$ do decil intervalar e $P_v = [p, \bar{p}]$ do percentil intervalar com variáveis intervalares x e n valores intervalares x_1, \dots, x_n .

Para computar os valores intervalares das separatrizes quartil, decil e percentil, desenvolve-se o algoritmo SI (descrito no anexo B.4, Figura B.4) baseado na expressão definida na Seção 3.3.10.

COMPLEXIDADE: O problema de computar os valores intervalares das separatrizes pertence a classe de problemas P, pois trata-se de problema de decisão e o algoritmo (SI), proposto como solução, é um algoritmo determinístico de $O(n^2)$.

As separatrizes quartil, decil e percentil se comportam de forma similar à separatriz mediana, pois para estes indicadores deve-se ordenar os valores da amostra e depois calcular a posição da seqüência dos valores. O valor intervalar do quartil, decil, percentil ou mediana será o correspondente valor da posição calculada. Conforme o problema da mediana intervalar, a solução dos problemas de computar o valor das separatrizes intervalares dá-se através de algoritmo de classificação baseado em comparação, cuja complexidade é de $O(n^2)$.

4.2.5 Problema: Amplitude total Intervalar

PROBLEMA: Computar o intervalo $AT_v = [at, \bar{at}]$ da amplitude total intervalar com variáveis intervalares x e n valores intervalares x_1, \dots, x_n .

Para computar a amplitude total intervalar com a expressão definida na Seção 3.3.4, desenvolve-se o algoritmo ATI, descrito no anexo B.5, Figura B.5.

COMPLEXIDADE: Como o problema em questão é um problema de localização associado ao problema de decisão e o algoritmo ATI é um algoritmo determinístico de ordem $O(n^2)$, pode-se afirmar que o problema de computar o intervalo da amplitude total intervalar pertence à classe de problemas P.

Para esta medida de dispersão deve-se calcular, primeiramente, o mínimo intervalar MI e o máximo intervalar MA conforme definido na Seção 3.3.4.

A amplitude intervalar é a diferença entre o máximo intervalar e o mínimo intervalar. Esta diferença intervalar é calculada conforme definida por Moore (1966). O problema de computar o intervalo da amplitude total intervalar coincide com o problema de classificação baseado em comparações. Para este problema, tem-se uma solução igual ao problema da mediana intervalar, ou seja, algoritmo de ordem $O(n^2)$.

Embora este indicador envolva a operação de subtração entre intervalos, esta não é considerada a operação fundamental na análise da complexidade do algoritmo que resolve o problema.

4.2.6 Problema: Variância Intervalar

PROBLEMA: Computar o intervalo $VA_v = [va, \bar{va}]$ da variância intervalar com variáveis intervalares x e n valores intervalares x_1, \dots, x_n .

Com a expressão definida na Seção 3.3.5, desenvolve-se o algoritmo VI descrito no anexo B.6, Figura B.6.

COMPLEXIDADE: Considerando que o problema analisado é um problema de locali-

zação associado ao problema de decisão e o algoritmo VI é um algoritmo determinístico de tempo linear, pode-se afirmar que o problema de computar o intervalo da variância intervalar pertence à classe de problemas P.

O problema de computar o intervalo da variância intervalar foi o primeiro problema estudado e analisado por pesquisadores da área da computação intervalar. Para este problema foram encontrados dois resultados de complexidade devido a utilização da computação da imagem intervalar. Foi analisada a complexidade do problema ao computar o extremo inferior \underline{va} e o extremo superior \overline{va} do intervalo VA_v .

Através da Tabela 4.1 pode-se comparar os resultados de complexidade quando considerado a utilização da imagem intervalar (estado da arte) e nosso resultado de complexidade obtido ao considerarmos a extensão intervalar.

Tabela 4.1: Computação Intervalar e Complexidade do Problema Variância Intervalar

Computação Intervalar	Complexidade do Problema Variância Intervalar
Imagem Intervalar	computar $\underline{va} \rightarrow P$ (ordem $O(n^2)$) computar $\overline{va} \rightarrow NP$ -Difícil
Extensão Intervalar	P (ordem $O(n)$)

4.2.7 Problema: Desvio Padrão Intervalar

PROBLEMA: Computar o intervalo $DP_v = [\underline{dp}, \overline{dp}]$ do desvio padrão intervalar com variáveis intervalares \mathbf{x} e n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Desenvolve-se o algoritmo DPI para a expressão do desvio padrão intervalar definida na Seção 3.3.6. O algoritmo DPI encontra-se no anexo B.7, Figura B.7.

COMPLEXIDADE: O problema de computar o intervalo do desvio padrão intervalar é resolvido por um algoritmo determinístico de complexidade de ordem linear, $O(n)$. Sendo um problema de localização associado ao problema de decisão, o problema em questão pertence à classe de problemas P.

Não foram encontrados na bibliografia resultados de análise de complexidade do problema do desvio padrão intervalar.

4.2.8 Problema: Coeficiente de Variação Intervalar

PROBLEMA: Computar o intervalo $CV_v = [\underline{cv}, \overline{cv}]$ do coeficiente de variação intervalar com variáveis intervalares \mathbf{x} e n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Desenvolve-se o algoritmo CVI (descrito no anexo B.8, Figura B.8) para a expressão do coeficiente de variação intervalar definida na Seção 3.3.7.

COMPLEXIDADE: Como o problema de computar o intervalo do coeficiente de variação intervalar é resolvido por um algoritmo determinístico de complexidade de ordem linear, $O(n)$, pode-se concluir o problema em questão pertence à classe de problemas P.

Da mesma forma que os problemas da mediana intervalar, moda intervalar, amplitude total intervalar e desvio padrão intervalar, o problema do coeficiente de variação intervalar

não foi estudado e analisado por pesquisadores da área da computação intervalar. Em consequência, não é possível realizar uma comparação com outros resultados.

4.2.9 Problema: Covariância Intervalar

PROBLEMA: Computar o intervalo $\text{CO}_v = [\underline{co}, \overline{co}]$ da covariância intervalar com variáveis intervalares \mathbf{x} e \mathbf{y} , n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$ e n valores intervalares $\mathbf{y}_1, \dots, \mathbf{y}_n$.

Para a expressão definida na Seção 3.3.8, desenvolve-se o algoritmo COI (descrito no anexo B.9, Figura B.9).

COMPLEXIDADE: Sendo o problema analisado um problema de localização associado ao problema de decisão, e o algoritmo COI um algoritmo determinístico de ordem linear, $O(n)$, pode-se afirmar que o problema de computar o intervalo da covariância intervalar pertence à classe de problemas P.

Na literatura, o problema da covariância intervalar foi analisado considerando o problema de computar o extremo inferior \underline{co} e o problema de computar o extremo superior \overline{co} do intervalo CO (isto devido a utilização da computação da imagem intervalar). A Tabela 4.2 apresenta os resultados de complexidade destes problemas e o nosso resultado de complexidade obtido ao considerar a extensão intervalar para a computação do intervalo CO_v .

Tabela 4.2: Computação Intervalar e Complexidade do Problema Covariância Intervalar

Computação Intervalar	Complexidade do Problema Covariância Intervalar
Imagem Intervalar	computar \underline{co} → NP-Difícil computar \overline{co} → NP-Difícil
Extensão Intervalar	P

4.2.10 Problema: Coeficiente de Correlação Intervalar

PROBLEMA: Computar o intervalo $\text{CC}_v = [\underline{cc}, \overline{cc}]$ do coeficiente da correlação intervalar com variáveis intervalares \mathbf{x} e \mathbf{y} , n valores intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$ e n valores intervalares $\mathbf{y}_1, \dots, \mathbf{y}_n$.

Para a expressão definida na Seção 3.3.9, desenvolve-se o algoritmo CCI, descrito no anexo B.10, Figura B.10.

COMPLEXIDADE: O problema de computar o intervalo CC_v do coeficiente de correlação intervalar pertence a classe de problemas P, pois o problema em questão é um problema de localização associado ao problema de decisão e o algoritmo que o resolve é um algoritmo determinístico de ordem linear, $O(n)$.

O problema do coeficiente de correlação intervalar foi analisado da mesma forma que o problema da covariância intervalar. Através da Tabela 4.3 pode-se comparar os resultados de complexidade já existentes e o nosso resultado de complexidade.

Tabela 4.3: Computação Intervalar e Complexidade do Problema Coeficiente de Correlação Intervalar

Computação Intervalar	Complexidade do Problema Coeficiente de Correlação Intervalar
Imagem Intervalar	computar $cc \rightarrow$ NP-Difícil computar $\overline{cc} \rightarrow$ NP-Difícil
Extensão Intervalar	P

4.3 Resultados de Complexidade

A partir da investigação da complexidade dos problemas baseados em computar os indicadores estatísticos descritivos intervalares, realiza-se a classificação destes problemas quanto a classe de complexidade. Conhecendo as classes de complexidade a que os problemas pertencem, os projetistas de algoritmos, e até mesmo os profissionais da área de estatística, podem ter uma medida real quanto às soluções disponíveis e a expectativa de melhorar esses resultados.

Como existem problemas de indicadores estatísticos que requerem ou não operações aritméticas, separamos nossos resultados de complexidade em duas tabelas. A primeira tabela (Tabela 4.4) refere-se aos problemas que não possuem operações aritméticas intervalares, a segunda tabela (Tabela 4.5) apresenta os problemas que possuem operações aritméticas intervalares.

Na Tabela 4.4 apresenta-se os nossos resultados de complexidade dos problemas de computar os valores da mediana, moda e separatrizes com entradas intervalares.

Tabela 4.4: Indicadores Estatísticos Intervalares e Complexidade dos Problemas dos Indicadores Estatísticos Intervalares

Indicadores Estatísticos Intervalares	Complexidade dos Problemas dos Indicadores Estatísticos Intervalares
Mediana Intervalar	P
Moda Intervalar	P
Separatrizes Intervalares	P

Na Tabela 4.5 apresenta-se a classificação quanto a classe de complexidade dos problemas descritos na seção anterior, e, também, descreve-se todos os resultados de complexidade destes problemas encontrados na bibliografia, realizando uma comparação com nossos resultados.

OBSERVAÇÃO: Nas tabelas a seguir utiliza-se o símbolo “-” para indicar que não foram encontrados resultados de complexidade para o referido problema.

Tabela 4.5: Problemas da Estatística Descritiva Intervalar, complexidade dos Problemas com extensão intervalar e com imagem intervalar

Problemas da Estatística Descritiva Intervalar	Complexidade dos Problemas com extensão intervalar (nossos resultados)	Complexidade dos Problemas com Imagem Intervalar (estado da arte)
Média Intervalar	P	-
Amplitude Intervalar	P	-
Variância Intervalar	P	$\underline{va} \rightarrow P$ $\overline{va} \rightarrow \text{NP-Difícil}$
Desvio Padrão	P	-
Coefficiente de Variação Intervalar	P	-
Covariância Intervalar	P	$\underline{co} \rightarrow \text{NP-Difícil}$ $\overline{co} \rightarrow \text{NP-Difícil}$
Coefficiente de Correlação Intervalar	P	$\underline{cc} \rightarrow \text{NP-Difícil}$ $\overline{cc} \rightarrow \text{NP-Difícil}$

4.4 Comentários Finais

O presente capítulo teve como objetivo mostrar os resultados de complexidade dos problemas de computar os valores intervalares de medidas de tendência central, dispersão e separatrizes.

Por meio da análise da complexidade computacional classifica-se quanto a classe de complexidade os problemas de computar os valores dos indicadores estatísticos com entradas intervalares.

Enfim, o capítulo apresentou algoritmos de tempo polinomial que calculam os intervalos dos indicadores estatísticos média, mediana, moda, amplitude total, variância, desvio padrão, coeficiente de variação, covariância, coeficiente de correlação, e separatrizes quartil, decil e percentil com entradas intervalares.

Para computar os valores da mediana, quartil, decil e percentil aplicou-se a mesma fórmula para calcular a posição dos valores desejados (fórmula descrita na Seção 3.3.10).

Observamos que, com os resultados de complexidade obtidos para os problemas de computar os valores intervalares dos indicadores estatísticos, tornamos a estatística descritiva viável computacionalmente. Além disso, os problemas de computar a variância, covariância e coeficiente de correlação com dados intervalares, que eram NP-Difíceis, foram reclassificados quanto a classe de complexidade, passando a pertencer a classe de problemas P.

O próximo capítulo apresenta a certificação da não ocorrência de intervalos solução superestimados.

5 VERIFICAÇÃO DO INTERVALO SOLUÇÃO

O presente capítulo tem como objetivo mostrar a qualidade de aproximação nos intervalos solução para os indicadores estatísticos intervalares. Realiza-se a verificação da qualidade de aproximação em relação a resposta real exata, através de exemplos numéricos, isto é, através da aplicação dos valores de índice de massa corporal de alunos do ensino fundamental. A efetiva certificação é realizada através do cálculo da medida de erro.

5.1 Qualidade da solução intervalar

A qualidade de um intervalo está relacionada com o diâmetro do mesmo (RATSCHEK; ROKNE, 1988). Quanto menor for o tamanho do diâmetro, melhor será a qualidade do intervalo.

No presente trabalho, representa-se os valores reais de uma amostra em valores intervalares considerando uma margem de precisão δ . Esta margem de precisão deve ser de valor pequeno, de tal forma que considere os erros inerentes aos valores da amostra e proporcione intervalos de boa qualidade.

Baseados nas fórmulas descritas no Capítulo 3, observa-se que são necessárias operações aritméticas simples para calcular o valor do intervalo solução, não altera a qualidade do intervalo. Nestas condições, a solução encontrada para os problemas de computar os valores intervalares dos indicadores estatísticos não apresentam intervalos superestimados.

5.2 Medida de correção de uma resposta intervalar

Segundo Ratschek (1988), os computadores utilizam uma aritmética chamada aritmética de ponto flutuante. Nesta aritmética números reais são aproximados por um subconjunto de números reais chamados representação numérica da máquina. Devido esta representação são gerados dois tipos de erros. O primeiro tipo de erro ocorre quando uma entrada de valor real é aproximada por um número de máquina. O segundo tipo de erro é causado por resultados intermediários aproximados pelos números de máquina. A aritmética intervalar fornece uma ferramenta para estimar e controlar esses erros *automaticamente*. No lugar de aproximar um valor real x por um número de máquina, o valor real x , usualmente desconhecido, é aproximado por um intervalo x tendo número de máquina nos extremos inferior e superior. O intervalo x contém o valor x . O comprimento (ou diâmetro) deste intervalo pode ser usado como medida para qualidade da aproximação da resposta real exata. Os cálculos são executados usando intervalos em vez de números re-

ais e, conseqüentemente, a aritmética real é substituída pela aritmética intervalar. Quando se realiza computação com número de máquina \tilde{x} não existe estimativa do erro $|\tilde{x} - x|$. A computação com utilização de intervalos fornece as seguintes estimativas para o erro:

- *Erro Absoluto*: $|x_k - m(\mathbf{x}_k)| < w(\mathbf{x}_k)/2$,
onde $m(\mathbf{x}_k)$ é o ponto médio do intervalo \mathbf{x}_k e $w(\mathbf{x}_k) = \bar{x}_k - \underline{x}_k$ é o diâmetro do intervalo \mathbf{x}_k ;
- *Erro Relativo*: $\left| \frac{x_k - m(\mathbf{x}_k)}{x_k} \right| \leq \frac{w(\mathbf{x}_k)}{2 \min|\mathbf{x}_k|}$ se $0 \notin \mathbf{x}_k$, onde $|\mathbf{x}_k| = \{ |x| : x \in \mathbf{x}_k \}$.

Neste trabalho os valores reais $\{x_1, \dots, x_n\}$, das amostras aleatórias de uma população, são representados por intervalos $\mathbf{x}_i = [x_i - \delta, x_i + \delta]$, onde δ é a margem de precisão escolhida para os valores reais.

Para a verificação da qualidade de aproximação no intervalo solução, deve-se considerar o valor $x_k \in \mathbb{R}$, o intervalo $\mathbf{x}_k = [\underline{x}_k, \bar{x}_k]$ e uma dada exatidão ε . Realiza-se a análise do erro através das seguintes medidas de erros:

Observa-se que nas medidas de erros absoluto e relativo, citadas acima, utiliza-se o ponto médio $m(\mathbf{x})$ do intervalo \mathbf{x} para medir a distância do valor real em relação ao valor pontual (ponto médio) do intervalo. Tais medidas de erros podem gerar questões como: “O valor real está no intervalo solução?”; “A utilização do ponto médio representa um indicador confiável na análise do erro?”.

Segundo Sunaga (1958) a interpretação usualmente aceita para um intervalo no contexto da aritmética intervalar é a de envoltória intervalar de um número real. Esta semântica, de envoltória intervalar, sugere a representação dos intervalos na forma

$$m(\mathbf{x}) \pm w(\mathbf{x})/2,$$

aludindo à idéia que o ponto médio $m(\mathbf{x})$ seria o número real “medido” e o raio $w(\mathbf{x})/2$ indicaria a incerteza gerada pelas restrições de precisão e ambientais existentes. Dessa forma o valor exato estaria limitado pelo intervalo apresentado.

Em (VACCARO, 2001) verifica-se que à interpretação de envoltória intervalar são associados os seguintes pressupostos:

- Qualquer real pertencente à envoltória intervalar de reais é um possível representante do valor real exato associado ao intervalo: como o valor exato do número real que se deseja representar não é explicitamente conhecido, então todos os números reais contidos na envoltória intervalar são possíveis representantes deste valor real. Isto é, o valor real continua indeterminado, porém restrito ao domínio dos valores reais que compõem a envoltória intervalar;
- Com a utilização de envoltórias intervalares espera-se modelar e determinar o efeito da propagação de erros em procedimentos de cálculo numérico em ponto flutuante.

Aplica-se estas medidas de erros com o objetivo de verificar a qualidade do intervalo solução, obtido após o processamento de operações aritméticas intervalares. Através das medidas de erros verifica-se se o método da extensão intervalar fornece como resposta intervalos que englobam a resposta real exata.

5.3 Estudo de Caso

Como estudo de caso, considera-se o cálculo do IMC (Índice de Massa Corporal) intervalar dos alunos do ensino fundamental de 5a. a 8a. série da escola Nossa Senhora Imaculada Conceição, localizada na cidade de Cachoeira do Sul/RS.

O índice de Massa Corporal (IMC) é uma fórmula que indica se uma pessoa está acima do peso, se está obeso ou com peso ideal considerado saudável. A fórmula para calcular o Índice de Massa Corporal é: $IMC = peso/(altura)^2$ (IMC, 2005).

A fim de desenvolver uma definição aceitável para sobrepeso e obesidade em crianças e adolescentes, foi desenvolvido um estudo internacional envolvendo seis países: Brasil, Grã-Bretanha, China (Hong Kong), Holanda, Cingapura e Estados Unidos. Foram 97.876 meninos e 94.851 meninas, acompanhados do nascimento até os 25 anos de idade.

Para cada faixa etária e sexo foram desenhadas curvas de tal sorte que, na idade de 18 anos, passassem pelos pontos de corte largamente aceitos para sobrepeso e obesidade em adultos (IMC de 25 e 30 Kg/m²). Essas curvas foram arranjadas para fornecer esses pontos de corte para cada sexo e idade, de 2 a 18 anos. Maiores informações podem ser encontradas em (VIUNISKI, 2005).

A Tabela 5.1, por ser menos arbitrária, mais universal e mais representativa, é o instrumento utilizado para fornecer um levantamento internacional da prevalência da obesidade em crianças e adolescentes (VIUNISKI, 2005). A Tabela 5.1 não se encontra completa, de 2 a 18 anos, devido o nosso estudo de caso possuir alunos com idades entre 9 e 16 anos.

Tabela 5.1: Idade, IMC Meninas e IMC Meninos.

Idade	IMC Meninas			IMC Meninos		
	Saudável	Sobrepeso	Obeso	Saudável	Sobrepeso	Obeso
9	< 19.1	19.1-22.7	≥ 22.8	< 19.1	19.1-22.8	≥ 22.8
10	< 19.9	19.9-24.0	≥ 24.1	< 19.8	19.8-23.9	≥ 24.0
11	< 20.7	20.7-25.3	≥ 25.4	< 20.6	20.6-25.0	≥ 25.1
12	< 21.7	21.7-26.6	≥ 26.7	< 21.2	21.2-25.9	≥ 26.0
13	< 22.6	22.6-27.7	≥ 27.8	< 21.9	21.9-26.7	≥ 26.8
14	< 23.3	23.3-28.5	≥ 28.6	< 22.6	22.6-27.5	≥ 27.6
15	< 23.9	23.9-29.0	≥ 29.1	< 23.3	23.3-28.2	≥ 28.3
16	< 24.4	24.4-29.3	≥ 29.4	< 23.9	23.9-28.8	≥ 28.9

É importante salientar que o objetivo do capítulo não é realizar uma análise estatística sobre o IMC dos alunos do ensino fundamental de 5a. a 8a. série. O foco é mostrar, através da aplicação do IMC intervalar, cálculos numéricos para os indicadores estatísticos intervalares, os quais retornam intervalos solução com qualidade de aproximação em relação a solução real exata), utilizando a extensão intervalar.

Na próxima seção apresentam-se os cálculos realizados para os indicadores estatísticos utilizando o IMC com valores reais e com valores intervalares.

5.4 Verificação da Qualidade de Aproximação

Para verificação da qualidade de aproximação nos intervalos solução, apresentam-se exemplos de cálculos numéricos dos indicadores estatísticos com IMC (Índice de Massa Corporal) intervalares (estudo de caso seção 5.3) usando sistema de ponto flutuante F(10, 5, -10, 10) e arredondamento direcionado (KULISCH; MIRANKER, 1981).

As operações intervalares envolvidas podem ser encontradas em Moore (1966).

Para a validação das soluções intervalares encontradas na tese, realiza-se um comparativo entre as soluções dos cálculos reais dos indicadores estatísticos, realizados no software NetBook (CAMPOS et al. , 2004), e as soluções dos cálculos intervalares dos mesmos indicadores estatísticos intervalares, realizados no software IntLab (MATLAB , 2005).

O NetBook é uma ferramenta gratuita desenvolvida no CIn/UFPE, com o objetivo de suportar análise de desempenho de sistemas de comunicação, em particular, redes de computadores. É composto por quatro módulos: Estatística, Geração, Transformação e Gráfico. Exceto pelo módulo de geração de tráfego auto-similar, que foi implementado em C++, todos os demais módulos do NetBook foram implementados em Java.

O IntLab contém tipos de dados básicos e operadores para aritmética intervalar, bem como uma variedade de métodos numéricos usando intervalos.

Com os dados da altura, peso e idade (encontram-se no Anexo C) calcula-se o Índice de Massa Corporal (IMC) para todos os alunos da escola.

Como o IMC está relacionado com o sexo e a idade, separamos os alunos de cada turma (5a., 6a., 7a. e 8a.) em dois conjuntos de dados: conjunto do IMC das meninas e conjunto do IMC dos meninos. A partir destes conjuntos realizamos a representação intervalar com precisão $\delta = 0.005$ e calculam-se os indicadores estatísticos intervalares abordados neste trabalho. A margem de precisão δ é de livre escolha e está relacionada com a exatidão que se deseja obter nos extremos inferior e superior do intervalo solução.

5.4.1 IMC dos alunos da 5a. série

Considera-se o IMC da turma de 21 alunos da 5a. série da escola Imaculada Conceição.

IMC das meninas: {16.8786, 14.2687, 20.7834, 18.2961, 19.8179, 27.0258, 22.6562, 22.0711, 26.9127, 21.0517, 24.6094, 25.1095, 16.0697, 21.8146}.

Representação intervalar: {[16.874, 16.884], [14.264, 14.274], [20.778, 20.788], [18.291, 18.301], [19.813, 19.823], [27.021, 27.031], [22.651, 22.661], [22.066, 22.076], [26.908, 26.918], [21.047, 21.057], [24.604, 24.614], [25.104, 25.114], [16.065, 16.075], [21.810, 21.820]}.

Para calcular os indicadores estatísticos covariância e coeficiente de correlação precisam-se de dois conjuntos de dados. Dessa forma, consideram-se o IMC dos meninos da 5a. série da mesma escola e alguns valores do IMC das meninas para completar o total de valores necessários: {16.8786, 15.1506, 20.7834, 19.0342, 19.8179, 17.3469, 22.6562, 15.7342, 26.9127, 20.8889, 21.0517, 25.1095, 16.0697, 21.8146}. Representam-se estes valores em intervalos com a mesma margem de precisão $\delta = 0.005$: {[16.874, 16.884], [15.146, 15.156], [20.778, 20.788], [19.029, 19.039], [19.813, 19.823], [17.342, 17.352], [22.651, 22.661], [15.729, 15.739], [26.908, 26.918], [20.884, 20.894], [21.047, 21.057], [25.104, 25.114], [16.065, 16.075], [21.810, 21.820]}.

Na Tabela 5.2 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC das meninas da 5a. série.

Tabela 5.2: Indicadores Estatísticos, dados com IMC reais (meninas da 5a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninas da 5a. série)	Dados com IMC Intervalares
Média	22.874	[22.868, 22.880]
Mediana	21.052	[21.047, 21.057]
Moda	-	-
Amplitude total	12.757	[12.747, 12.767]
Variância	18.282	[18.206, 18.357]
Desvio Padrão	4.2756	[4.2668, 4.2845]
Coef. de Variação	0.18692	[0.18649, 0.18735]
Covariância	11.835	[11.762, 11.906]
Coef. de Correlação	0.71544	[0.70802, 0.72290]

Na Tabela 5.3 apresentam-se os erros absoluto e relativo obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.3: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 5a. série) .

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	0.0000077 < 0.0053846	0.00000033662 ≤ 0.00023545
Mediana	0 < 0.005	0 ≤ 0.0002375
Moda	-	-
Amplitude total	0 < 0.01	0 ≤ 0.0007844
Variância	0.00023068 < 0.075398	0.000012618 ≤ 0.0041414
Desvio Padrão	0.000056624 < 0.0088172	0.000013243 ≤ 0.0020664
Coef. de Variação	0.000000291 < 0.00042946	0.0000015568 ≤ 0.0023029
Covariância	0.00062319 < 0.071746	0.000052657 ≤ 0.0060996
Coef. de Correlação	0.000020583 < 0.0074402	0.000028769 ≤ 0.010508

Pelos resultados apresentados na Tabela 5.2 e os erros calculados na Tabela 5.3 observa-se que os intervalos solução, obtidos para os indicadores estatísticos, englobam a resposta real exata. Os valores dos indicadores estatísticos com dados reais estão bem representados nos valores intervalares, isto é, os intervalos solução contém os valores reais.

IMC dos meninos: {15.1506, 14.9515, 19.0342, 15.3739, 17.3469, 15.7342, 20.8889}.

Representação intervalar: {[15.146, 15.156],[14.946, 14.956],[19.029, 19.039],[15.369, 15.379], [17.342, 17.352],[15.729, 15.739],[20.884, 20.894]}.

Para os indicadores estatísticos covariância e coeficiente de correlação considera-se o IMC das meninas da 5a. série da mesma escola: {16.8786, 14.2687, 20.7834, 18.2961, 19.8179, 27.0258, 22.6562}. Representam-se estes valores em intervalos com a mesma precisão $\delta = 0.005$: {[16.874, 16.884],[14.264, 14.274], [20.778, 20.788],[18.291, 18.301],[19.813, 19.823],[27.021, 27.031],[22.651, 22.661]}.

Na Tabela 5.4 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC dos meninos da 5a. série.

Tabela 5.4: Indicadores Estatísticos, dados com IMC reais (meninos da 5a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninos da 5a. série)	Dados com IMC Intervalares
Média	19.747	[19.740, 19.753]
Mediana	15.734	[15.729, 15.739]
Moda	-	-
Amplitude total	5.9374	[5.9274, 5.9474]
Variância	14.485	[14.405, 14.565]
Desvio Padrão	3.8059	[3.7955, 3.8164]
Coef. de Variação	0.1924	[0.19215, 0.19333]
Covariância	14.82073	[14.725, 14.916]
Coef. de Correlação	0.71092	[0.70309, 0.71878]

Na Tabela 5.5 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.5: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 5a.série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	0.00003334 < 0.0058333	0.0000016884 ≤ 0.00029549
Mediana	0 < 0.005	0 ≤ 0.0003178
Moda	-	-
Amplitude total	0 < 0.01	0 ≤ 0.001687
Variância	0.00018654 < 0.079558	0.000012878 ≤ 0.0055226
Desvio Padrão	0.000052877 < 0.010452	0.000013893 ≤ 0.0027537
Coef. de Variação	0.00033918 < 0.00058623	0.0017629 ≤ 0.0030508
Covariância	0.00040846 < 0.095326	0.000027560 ≤ 0.0064737
Coef. de Correlação	0.000017663 < 0.0078411	0.000024846 ≤ 0.011152

Observa-se, novamente pelas Tabelas 5.4 e 5.5, a qualidade de aproximação nos intervalos solução.

5.4.2 IMC dos alunos da 6a. série

Considera-se o IMC da turma de 25 alunos da 6a. série da escola Imaculada Conceição.

IMC das meninas: {26.936, 16.397, 17.778, 16.797, 15.089, 22.921, 19.396, 21.644, 22.000, 16.667, 15.753, 19.107, 19.979, 22.214, 22.893}.

Representação intervalar: {[26.931, 26.941], [16.392, 16.402], [17.773, 17.783], [16.792, 16.802], [15.084, 15.094], [22.916, 22.926], [19.391, 19.401], [21.639, 21.649], [21.995, 22.005], [16.661, 16.671], [15.748, 15.758], [19.102, 19.112], [19.974, 19.984], [22.209, 22.219], [22.888, 22.898]}.

Para calcular os indicadores estatísticos covariância e coeficiente de correlação consideram-se o IMC dos meninos da 6a. série da mesma escola e alguns valores do IMC das meninas para completar o total de valores necessários: {21.228, 16.436, 27.630, 26.927, 18.667, 16.435, 17.854, 19.724, 16.000, 14.382, 22.893, 22.214, 19.979, 19.107, 15.753}. Representam-se estes valores em intervalos com a mesma precisão $\delta = 0.005$: {[21.223, 21.233], [16.431, 16.441], [27.625, 27.635], [26.922, 26.932], [18.662, 18.672], [16.430, 16.440], [17.849, 17.859], [19.719, 19.729], [15.995, 16.005], [14.377, 14.387], [22.888, 22.898], [22.209, 22.219], [19.974, 19.984], [19.102, 19.112], [15.748, 15.758]}.

A Tabela 5.6 apresenta os resultados dos indicadores estatísticos, reais e intervalares, com IMC das meninas da 6a. série.

Tabela 5.6: Indicadores Estatísticos, dados com IMC reais (meninas da 6a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninas da 6a. série)	Dados com IMC Intervalares
Média	21.112	[21.106, 21.118]
Mediana	19.396	[19.391, 19.401]
Moda	-	-
Amplitude total	11.846	[11.836, 11.856]
Variância	13.402	[13.335, 13.469]
Desvio Padrão	3.661	[3.6518, 3.6700]
Coef. de Variação	0.17340	[0.17292, 0.17388]
Covariância	-0.73801	[-0.80977, -0.66628]
Coef. de Correlação	-0.047957	[-.0052866, -0.043096]

Na Tabela 5.7 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.7: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 6a. série).

Indicadores Estatísticos	Erro Absoluto	Erro Relativo
	$ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	$ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	$0.00014286 < 0.0053571$	$0.0000067668 \leq 0.00025381$
Mediana	$0 < 0.005$	$0 \leq 0.0002578$
Moda	-	-
Amplitude total	$0 < 0.01$	$0 \leq 0.0008448$
Variância	$0.00040946 < 0.066556$	$0.000030552 \leq 0.0049907$
Desvio Padrão	$0.000081148 < 0.0090900$	$0.000022165 \leq 0.0024892$
Coef. de Variação	$0.0000035915 < 0.00047456$	$0.000020712 \leq 0.0027442$
Covariância	$0.000018348 < 0.071743$	$0.000024861 \leq 0.10767$
Coef. de Correlação	$0.000024367 < 0.0048856$	$0.00050809 \leq 0.10182$

IMC dos meninos: {21.228, 16.436, 27.630, 26.927, 18.667, 16.435, 17.854, 19.724, 16.000, 14.382}.

Representação intervalar: {[21.223, 21.233], [16.431, 16.441], [27.625, 27.635], [26.922, 26.932], [18.662, 18.672], [16.430, 16.440], [17.849, 17.859], [19.719, 19.729], [15.995, 16.005], [14.377, 14.387]}.

Para os indicadores estatísticos covariância e coeficiente de correlação considera-se o IMC das meninas da 6a. série da mesma escola: {26.936, 16.397, 17.778, 16.797, 15.089, 22.921, 19.396, 21.644, 22.000, 16.666}.

Representação intervalar: {[26.931, 26.941], [16.392, 16.402], [17.773, 17.783], [16.792, 16.802], [15.084, 15.094], [22.916, 22.926], [19.391, 19.401], [21.639, 21.649], [21.995, 22.005], [16.661, 16.671]}.

Na Tabela 5.8 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC dos meninos da 6a. série.

Tabela 5.8: Indicadores Estatísticos, dados com IMC reais (meninos da 6a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninos da 6a. série)	Dados com IMC Intervalares
Média	21.698	[21.693, 21.704]
Mediana	17.854	[17.849, 17.859]
Moda	-	-
Amplitude total	13.248	[13.238, 13.258]
Variância	25.777	[25.6822, 25.878]
Desvio Padrão	5.0771	[5.0677, 5.0870]
Coef. de Variação	0.23399	[0.23349, 0.23449]
Covariância	3.8908	[3.7986, 3.9841]
Coef. de Correlação	0.17507	[0.17022, 0.17998]

Na Tabela 5.9 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.9: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 6a. série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	0.00066668 < 0.005	0.000030725 ≤ 0.00023048
Mediana	0 < 0.005	0 ≤ 0.0002801
Moda	-	-
Amplitude total	0 < 0.01	0 ≤ 0.0007554
Variância	0.0030828 < 0.097829	0.00011959 ≤ 0.0038092
Desvio Padrão	0.00029992 < 0.0096338	0.000059073 ≤ 0.0019009
Coef. de Variação	0.0000060555 < 0.00049789	0.000025879 ≤ 0.0021323
Covariância	0.00055156 < 0.092718	0.00014176 ≤ 0.024408
Coef. de Correlação	0.000033821 < 0.0048794	0.00019318 ≤ 0.028664

Os indicadores estatísticos, calculados para os IMC intervalares dos alunos da 6a. série, também apresentaram intervalos solução com qualidade de aproximação. Observa-se que os intervalos solução de todos os indicadores estatísticos contém os valores reais dos respectivos indicadores.

5.4.3 IMC dos alunos da 7a. série

Para a realização dos cálculos dos valores dos indicadores estatísticos, com valores reais e valores intervalares, considera-se o IMC da turma de 29 alunos da 7a. série da escola Imaculada Conceição.

IMC das meninas: {20.173, 20.343, 21.077, 21.218, 24.244, 20.957, 16.447, 17.111, 21.228, 15.601, 15.601, 18.222, 15.556, 31.163, 13.963, 16.647, 20.312, 17.592, 18.256, 15.616, 17.625}.

Representa-se os dados reais em intervalos com a margem de precisão $\delta = 0.005$ intervalar: {[20.168, 20.178], [20.328, 20.348], [21.072, 21.082], [21.213, 21.223], [24.239, 24.249], [20.952, 20.962], [16.442, 16.452], [17.106, 17.116], [21.223, 21.233], [15.596, 15.606], [15.596, 15.606], [18.217, 18.227], [15.551, 15.561], [31.158, 31.168], [13.958, 13.968], [16.642, 16.652], [20.307, 20.317], [17.587, 17.597], [18.251, 18.261], [15.611, 15.621], [17.620, 17.630]}.

Calcula-se os indicadores estatísticos covariância e coeficiente de correlação considerando o IMC dos meninos da 7a. série da mesma escola (e alguns valores do IMC das meninas para completar o total de valores necessários): {17.982, 20.333, 15.556, 21.096, 21.218, 17.348, 17.087, 29.202, 17.625, 15.616, 18.256, 17.592, 20.312, 16.647, 13.963, 31.163, 15.556, 18.222, 15.601, 15.601, 21.228}. Representam-se estes valores em intervalos com a mesma margem de precisão $\delta = 0.005$: {[17.977, 17.987], [20.328, 20.338], [15.551, 15.561], [21.091, 21.101], [21.213, 21.223], [17.343, 17.353], [17.082, 17.092], [29.197, 29.207], [17.620, 17.630], [15.611, 15.621], [18.251, 18.261], [17.587, 17.597], [20.307, 20.317], [16.642, 16.652], [13.958, 13.968], [31.158, 31.168], [15.551, 15.561], [18.217, 18.227], [15.596, 15.606], [15.596, 15.606], [21.223, 21.233]}.

Os resultados dos indicadores estatísticos com dados reais e dados intervalares encontram-se na Tabela 5.10.

Tabela 5.10: Indicadores Estatísticos, dados com IMC reais (meninas da 7a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninas da 7a. série)	Dados com IMC Intervalares
Média	19.948	[19.941, 19.953]
Mediana	18.222	[18.217, 18.21]
Moda	15.601	[15.596, 15.606]
Amplitude total	17.200	[17.19, 17.21]
Variância	15.488	[15.421, 15.553]
Desvio Padrão	3.9355	[3.9270, 3.9437]
Coef. de Variação	0.19729	[0.19681, 0.19776]
Covariância	-0.29078	[-0.36042, -0.22185]
Coef. de Correlação	-0.016731	[-0.020821, -0.012714]

Os erros obtidos ao calcular os indicadores estatísticos com valores intervalares encontram-se na Tabela 5.11.

Tabela 5.11: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 7a. série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	$0.0006500 < 0.0055$	$0.000032585 \leq 0.00027580$
Mediana	$0 < 0.005$	$0 \leq 0.0002744$
Moda	$0 < 0.005$	$0 \leq 0.0003205$
Amplitude total	$0 < 0.01$	$0 \leq 0.0002908$
Variância	$0.00061522 < 0.065626$	$0.000039722 \leq 0.0042554$
Desvio Padrão	$0.00010736 < 0.0083379$	$0.000027279 \leq 0.0021232$
Coef. de Variação	$0.0000008743 < 0.00047239$	$0.0000044315 \leq 0.0024002$
Covariância	$0.00035720 < 0.069289$	$0.0012284 \leq 0.31233$
Coef. de Correlação	$0.000036588 < 0.0040535$	$0.0021868 \leq 0.31882$

IMC dos meninos: {17.982, 20.333, 15.556, 21.096, 21.218, 17.348, 17.087, 29.202}.

Representação intervalar: {[17.977, 17.987], [20.328, 20.338], [15.551, 15.561], [21.091, 21.101], [21.213, 21.223], [17.343, 17.353], [17.082, 17.092], [29.197, 29.207]}.

Para os indicadores estatísticos covariância e coeficiente de correlação considera-se o IMC das meninas da 7a. série da mesma escola: {20.173, 20.343, 21.077, 21.218, 24.244, 20.957, 16.447, 17.111}.

Representação intervalar: {[20.168, 20.178], [20.328, 20.348], [21.072, 21.082], [21.213, 21.223], [24.239, 24.249], [20.952, 20.962], [16.442, 16.452], [17.106, 17.116]}.

Na Tabela 5.12 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC dos meninos da 7a. série.

Tabela 5.12: Indicadores Estatísticos, dados com IMC reais (meninos da 7a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninos da 7a. série)	Dados com IMC Intervalares
Média	22.832	[22.826, 22.837]
Mediana	20.333	[20.328, 20.338]
Moda	-	-
Amplitude total	13.646	[13.636, 13.656]
Variância	27.424	[27.315, 27.533]
Desvio Padrão	5.2368	[5.2264, 5.24721]
Coef. de Variação	0.22936	[0.22885, 0.22988]
Covariância	6.94556	[6.8463, 7.0438]
Coef. de Correlação	0.33614	[0.32977, 0.34254]

Na Tabela 5.13 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.13: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 7a. série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	0.00028572 < 0.0057142	0.000012514 ≤ 0.00025034
Mediana	0 < 0.005	0 ≤ 0.0002459
Moda	-	-
Amplitude total	0 < 0.01	0 ≤ 0.0007333
Variância	0.00034586 < 0.10889	0.000012612 ≤ 0.0039866
Desvio Padrão	0.000015610 < 0.010397	0.0000029808 ≤ 0.0019893
Coef. de Variação	0.0000059809 < 0.00051278	0.000026076 ≤ 0.0022406
Covariância	0.00051167 < 0.0987510	0.000073668 ≤ 0.014424
Coef. de Correlação	0.000018835 < 0.0063848	0.000056034 ≤ 0.019361

Para os dados de IMC dos alunos da 7a. série verifica-se a existência do valor modal real e intervalar. Conforme ocorreram nos exemplos anteriores, Seções 5.4.1 e 5.4.2, os valores intervalares de todos os indicadores estatísticos possuem qualidade de aproximação em relação a resposta real exata. Para finalizar a seqüência de cálculos dos valores intervalares dos indicadores estatísticos, apresenta-se na próxima seção os dados de IMC dos alunos da 8a. série.

5.4.4 IMC dos alunos da 8a. série

Nesta seção considera-se o IMC de 41 alunos da 8a. série da escola Imaculada Conceição, para a realização dos cálculos dos valores dos indicadores estatísticos, com valores reais e valores intervalares.

IMC das meninas: {16.023, 19.959, 26.446, 20.866, 16.797, 20.976, 21.294, 19.143, 24.022, 18.468, 16.836, 22.408, 17.799, 18.325, 17.709, 20.776, 19.355, 23.327, 21.613, 18.226, 24.219, 18.359, 18.444}.

Representam-se os dados reais em intervalos com a margem de precisão $\delta = 0.005$ intervalar: {[16.018, 16.028], [19.954, 19.964], [26.441, 26.451], [20.861, 20.871], [16.792, 16.802], [20.971, 20.981], [21.288, 21.298], [19.138, 19.148], [24.017, 24.027], [18.463, 18.473], [16.831, 16.841], [22.403, 22.413], [17.994, 18.004], [18.320, 18.330], [17.704, 17.714], [20.770, 20.780], [19.350, 19.360], [23.322, 23.332], [21.608, 21.618], [18.221, 18.231], [24.214, 24.224], [18.354, 18.364], [18.439, 18.449]}.

Calculam-se os indicadores estatísticos covariância e coeficiente de correlação considerando o IMC dos meninos da 8a. série da mesma escola (e alguns valores do IMC das meninas para completar o total de valores necessários): {16.396, 28.026, 16.107, 26.367, 20.742, 17.999, 32.562, 28.556, 17.667, 22.128, 24.913, 20.347, 18.549, 18.069, 20.724, 18.467, 16.512, 19.608, 16.023, 19.959, 26.446, 20.866, 16.797}.

Representação intervalar: {[16.391, 16.401], [28.021, 28.031], [16.102, 16.112], [26.362, 26.372], [20.737, 20.747], [17.994, 18.004], [32.557, 32.567], [28.551, 28.561], [17.662, 17.672], [22.123, 22.133], [24.908, 24.918], [20.342, 20.352], [18.544, 18.554], [18.064, 18.074], [20.719, 20.729], [18.462, 18.472], [16.507, 16.517], [19.603, 19.613], [16.018, 16.028], [19.954, 19.964], [26.441, 26.451], [20.861, 20.871], [16.792, 16.802]}.

Na Tabela 5.14 encontram-se os resultados dos indicadores estatísticos com dados reais e dados intervalares.

Tabela 5.14: Indicadores Estatísticos, dados com IMC reais (meninas da 8a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninas da 8a. série)	Dados com IMC Intervalares
Média	20.972	[20.976, 20.986]
Mediana	19.355	[19.35, 19.36]
Moda	-	-
Amplitude total	10.423	[10.413, 10.433]
Variância	8.1255	[8.0366, 8.1369]
Desvio Padrão	2.8505	[2.8349, 2.8525]
Coef. de Variação	0.13592	[0.13508, 0.13599]
Covariância	0.87359	[0.7831, 0.91885]
Coef. de Correlação	0.06503	[0.058139, 0.068907]

Através da Tabela 5.15 verifica-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.15: Indicadores Estatísticos, erros absolutos e erro relativo (meninas 8a. série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	$0.0092727 < 0.0052273$	$0.00044215 \leq 0.0002492$
Mediana	$0 < 0.005$	$0 \leq 0.0002583$
Moda	-	-
Amplitude total	$0 < 0.01$	$0 \leq 0.0009603$
Variância	$0.038717 < 0.050114$	$0.0047648 \leq 0.0062354$
Desvio Padrão	$0.0067867 < 0.00888113$	$0.0023801 \leq 0.0031082$
Coef. de Variação	$0.0003841 < 0.00045373$	$0.0028259 \leq 0.0033589$
Covariância	$0.022613 < 0.067873$	$0.025885 \leq 0.086672$
Coef. de Correlação	$0.0015068 < 0.0053836$	$0.0231716 \leq 0.092598$

IMC dos meninos: {16.396, 28.026, 16.107, 26.367, 20.742, 17.999, 32.562, 28.556, 17.667, 22.128, 24.913, > 20.347, 18.549, 18.069, 20.724, 18.467, 16.512, 19.608}.

Representação intervalar: {[16.391,16.401], [28.021, 28.031], [16.102, 16.112], [26.362, 26.372], [20.737, 20.747], [17.994, 18.004], [32.557, 32.567], [28.551, 28.561], [17.662, 17.672], [22.123, 22.133], [24.908, 24.918], [20.342, 20.352], [18.544, 18.554], [18.064, 18.074], [20.719, 20.729], [18.462, 18.472], [16.507, 16.517], [19.603, 19.613]}.

Para os indicadores estatísticos covariância e coeficiente de correlação considera-se o IMC das meninas da 8a. série da mesma escola: {16.023, 19.959, 26.446, 20.866, 16.797, 20.976, 21.294, 19.143, 24.022, 18.468, 16.836, 22.408, 17.799, 18.325, 17.709, 20.776, 19.355, 23.327}.

Representação intervalar: {[16.018, 16.028], [19.954, 19.964], [26.441, 26.451], [20.861, 20.871], [16.792, 16.802], [20.971, 20.981], [21.288, 21.298], [19.138, 19.148], [24.017, 24.027], [18.463, 18.473], [16.831, 16.841], [22.403, 22.413], [17.994, 18.004], [18.320, 18.330], [17.704, 17.714], [20.770, 20.780], [19.350, 19.360], [23.322, 23.332]}.

Na Tabela 5.16 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC dos meninos da 8a. série.

Na Tabela 5.17 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 5.16: Indicadores Estatísticos, dados com IMC reais (meninos da 8a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninos da 8a. série)	Dados com IMC Intervalares
Média	22.573	[22.567, 22.578]
Mediana	19.608	[19.603, 19.613]
Moda	-	-
Amplitude total	16.455	[16.445, 16.465]
Variância	24.760	[24.666, 24.854]
Desvio Padrão	4.9759	[4.9664, 4.9854]
Coef. de Variação	0.22044	[0.21996, 0.22091]
Covariância	0.12278	[0.016866, 0.16427]
Coef. de Correlação	0.0081444	[0.0011164, 0.01979]

Tabela 5.17: Indicadores Estatísticos, erros absolutos e erro relativo (meninos 8a. série).

Indicadores Estatísticos	Erro Absoluto $ x_k - m(\mathbf{x}_k) < w(\mathbf{x}_k)/2$	Erro Relativo $ \frac{x_k - m(\mathbf{x}_k)}{x_k} \leq \frac{w(\mathbf{x}_k)}{2\min \mathbf{x}_k }$
Média	0.00011764 < 0.0052941	0.0000052115 ≤ 0.00023459
Mediana	0 < 0.005	0 ≤ 0.000255
Moda	-	-
Amplitude total	0 < 0.01	0 ≤ 0.000608
Variância	0.00000013440 < 0.094091	0.00000542811 ≤ 0.003814589
Desvio Padrão	0.000305336 < 0.0094546	0.000061363 ≤ 0.0019037
Coef. de Variação	0.0000007553 < 0.00047055	0.0000034263 ≤ 0.0021391
Covariância	0.032212 < 0.073701	0.26235 ≤ 4.3696
Coef. de Correlação	0.0020967 < 0.0049312	0.25745 ≤ 4.41676

Nas Tabelas 5.16 e 5.17 verificam-se que os valores intervalares dos indicadores estatísticos possuem qualidade de aproximação em relação a resposta real exata.

5.5 Comentários Finais

O presente capítulo teve como objetivo mostrar a qualidade de aproximação em relação a resposta real exata dos intervalos solução. Verificamos que a utilização da extensão intervalar retorna como resposta intervalos que englobam a resposta real exata.

A verificação da qualidade de aproximação nos intervalos solução ocorreu através de cálculos das medidas de erros. Em todos os exemplos considerados neste trabalho, verificou-se que manteve-se a qualidade no intervalo solução. Apesar de várias operações aritméticas entre os intervalos, os diâmetros dos intervalos solução não aumentaram em

comparação ao diâmetro dos intervalos iniciais, o qual é duas vezes a margem de precisão escolhida para os mesmos.

6 CONCLUSÃO

Este capítulo apresenta um fechamento das principais idéias desenvolvidas na tese, destacando os resultados obtidos e os pontos positivos, identificando possíveis pontos negativos e descrevendo o prosseguimento do trabalho.

6.1 Resultados Obtidos

O presente trabalho teve como objetivo analisar a complexidade computacional dos problemas de computar os valores dos indicadores estatísticos com entradas intervalares, obtendo uma classificação quanto a classe complexidade destes problemas.

As pesquisas na área da estatística descritiva intervalar são muito recentes. Atualmente, pesquisadores como Ferson (?), Wu (2003), Kreinovich (2003) e Xiang (2004) encontram-se trabalhando nesta área.

As pesquisas desenvolvidas estão concentradas em alguns indicadores estatísticos como variância, covariância e coeficiente de correlação, talvez por serem os indicadores mais utilizados ou mais comuns na área de estatística.

Salienta-se que em estudos desenvolvidos sobre a complexidade dos problemas de calcular os valores da variância, covariância e coeficiente de correlação com entradas intervalares, considera-se, nas expressões das medidas de dispersão, n variáveis reais x_i e n variáveis y_i com n entradas intervalares $[x_i, \bar{x}_i], [y_i, \bar{y}_i]$. A maioria dos artigos afirmam que, ao utilizar a extensão intervalar para computar os valores das medidas de dispersão variância, covariância e coeficiente de correlação, com dados intervalares, sempre obtém-se resultados com intervalos superestimados. E que, ao considerar a computação de tais valores através da imagem intervalar, pode-se obter intervalos exatos como resultados. Porém, devido ao processamento dos dados de entrada, prova-se que o problema de computar o intervalo imagem das medidas de dispersão (variância, covariância e correlação) é pertencente a classe de problemas NP-Difícil.

Preocupados com os resultados de NP-Dificuldade obtidos para os problemas de computar os intervalos da variância, covariância e coeficiente de correlação, concentramos nossas pesquisas em métodos da computação intervalar que tornasse possível a computação dos intervalos destes indicadores, e dos demais que compõem a estatística descritiva (média, moda, mediana, amplitude, desvio padrão, coeficiente de variação e separatrizes) e que não retornassem intervalos superestimados. Dos métodos de computação intervalar existentes na bibliografia, escolhemos a extensão intervalar por permitir a construção de algoritmos eficientes para os problemas abordados.

No trabalho realizado e apresentado na presente tese, observamos que:

- O valor da média intervalar é o mesmo se considerar a computação da imagem

intervalar ou a extensão intervalar, ou seja, não existe intervalo com problema de superestimação. Isto devido a média apresentar uma única operação aritmética (adição);

- Para as separatrizes mediana, decil, quartil, percentil e o indicador moda intervalar, não necessitamos utilizar métodos de computação intervalar (imagem intervalar ou a extensão intervalar). Estes indicadores estatísticos não apresentam operações aritméticas entre intervalos. Os valores intervalares para estes indicadores são os intervalos dos dados de entrada, não ocorrendo problema de intervalo superestimado;
- A amplitude intervalar se comporta de maneira similar a média intervalar por apresentar uma única operação aritmética (subtração). A operação principal operação envolvida neste indicador é a classificação entre intervalos, o que não fornece intervalos superestimados;
- Os indicadores variância, covariância e coeficiente de correlação (analisados na literatura) tornam-se possíveis de serem calculados utilizando a extensão intervalar. Para estes indicadores verificamos a não ocorrência de superestimação nos intervalos calculados;
- O desvio padrão e o coeficiente de variação, que antes não eram possíveis de serem calculados devido a variância ser um problema NP-Difícil, com os resultados obtidos no presente trabalho passam a ser calculados através de operações aritméticas intervalares e os valores intervalares obtidos não são superestimados.

Se estes indicadores fossem analisados como os indicadores estatísticos variância, covariância e coeficiente de correlação, ou seja, considerando a computação da imagem intervalar para computar os valores intervalares, a expectativa de solução seria a de problemas NP-Difíceis. Entretanto, a solução encontrada neste trabalho coloca os problemas destes indicadores estatísticos pertencentes à classe de problemas P.

As principais contribuições do presente trabalho referem-se a: (i) definição da estatística descritiva a nível intervalar; (ii) possibilidade de calcular os valores intervalares da variância, covariância e coeficiente de correlação eficientemente; (iii) análise da complexidade dos problemas dos indicadores estatísticos descritivos; (iv) classificação quanto a classe de complexidade dos problemas da estatística descritiva intervalar e (v) possibilidade de obter intervalos não superestimados com a extensão intervalar devido a representação dos valores reais em intervalos com amplitude pequena (margem de precisão δ).

Além das contribuições apresentadas pelo desenvolvimento do trabalho, observamos que os resultados obtidos são relevantes por tornar a Estatística Descritiva Intervalar viável computacionalmente, ou seja, os problemas de calcular os valores dos indicadores estatísticos, com entradas intervalares, passam a ser tratáveis.

Adicionalmente, aos resultados obtidos, reclassificamos quanto a classe de complexidade, os problemas de calcular os valores da variância, covariância e coeficiente de correlação com entrada intervalares, que eram NP-Difíceis, passam a pertencer a classe de problemas P.

Destacamos, como pontos positivos inerentes aos resultados obtidos, a possibilidade de aplicação da estatística descritiva intervalar em diversas áreas, como por exemplo, a

da Bioinformática. Na Bioinformática existem muitos dados, a maioria com erros de medição devido os processos serem imprecisos. Existe uma necessidade de análises estatísticas para identificar conceitos importantes como onde estão os genes em uma sequência de DNA, quais genes estão sendo expressos em um determinado organismo em um determinado tempo, quais são as vias biológicas dos organismos, etc. As técnicas que existem hoje para essas informações em Bioinformática são baseadas em estatística.

6.2 Prosseguimento do Trabalho

6.2.1 Ordenação de dados intervalares

Devido aos valores reais, de amostras aleatórias de uma população, serem representados através de intervalos para considerar erros numéricos de medidas, estes intervalos podem ser todos disjuntos (isto é, seus extremos inferior e superior não coincidem), podem ser todos encaixados (ou seja, o extremo superior de alguns intervalos são menores que os extremos inferiores dos demais intervalos), ou ainda, os intervalos podem ser alguns disjuntos e alguns encaixados. Para os dois primeiros casos, conforme descrito na Seção 3.2, existe relação de ordem para realizar a ordenação dos dados intervalares. Porém, não existe uma relação de ordem quando tem-se o caso onde alguns intervalos são disjuntos e alguns são encaixados. Em virtude deste fato, o cálculo da mediana e demais separatrizes está em estudo quando os dados intervalares possuem esta característica.

6.2.2 Intervalos de confiança intervalar

Com o objetivo de dar segmento a abordagem intervalar na estatística, propõe-se trabalhar com intervalos de confiança, para os indicadores estatísticos abordados neste trabalho, com dados reais e dados intervalares.

Com a intenção de investigar a qualidade e confiabilidade nos intervalos solução, obtidos neste trabalho, já realizamos alguns cálculos de intervalos de confiança com dados reais.

Inicialmente, calculamos os valores dos intervalos de confiança para a média com dados reais. O objetivo foi verificar se os valores intervalares da média intervalar, possuem "tanta confiança" quanto os valores calculados para a média com dados reais.

O intervalo de confiança consiste de duas partes: uma estimativa pontual e um valor \pm que descreve a precisão da estimativa (?). Chama-se o valor \pm de margem de erro.

Calcula-se o intervalo de confiança para média por:

$$\text{IC} = \left[me - \frac{t * dp}{\sqrt{n}}, me + \frac{t * dp}{\sqrt{n}} \right], \quad (6.1)$$

onde IC é o intervalo de confiança, me é a média da amostra, t é a margem de erro, dp o desvio padrão e n o tamanho da amostra.

Para calcular o valor de t , utiliza-se a distribuição *t-Student*. A distribuição *t-Student*, com $n - 1$ graus de liberdade, se aproxima da distribuição normal padrão quando n for considerado grande (n maior ou igual a 30). A distribuição t é simétrica em torno da média.

Segundo Lipschutz (1972), em geral, quando se trabalha com amostras só é conhecido o desvio padrão da amostra, e em consequência deve-se utilizar a distribuição *t-Student*.

Na Tabela D.1 (Anexo D) apresenta-se o valor intervalar da média, calculado com entradas intervalares (Seção 5.4), o intervalo de confiança para a média com variância

desconhecida e os graus de liberdade considerado para 5a., 6a., 7a. e 8a. séries dos alunos da escola Imaculada Conceição.

Considera-se 95% de confiança (na tabela de distribuição *t* Student (LIPSCHUTZ, 1972), utiliza-se a coluna de 0,975) para todos os intervalos de confiança calculados para a média.

A partir da Tabela D.1 observa-se que os valores da média intervalar estão contidos nos intervalos de confiança calculados para a média com valores reais. Conseqüentemente, pode-se afirmar que os valores da média intervalar possuem “confiabilidade” em relação a representação dos valores da média real, e que os extremos dos intervalos da média satisfazem (ou não ultrapassam) os limites dos intervalos de confiança.

Em relação ao diâmetro do intervalo (o qual é uma medida de qualidade para o intervalo (RATSCHEK; ROKNE, 1988)), verifica-se que os intervalos de confiança possuem diâmetros maiores que os valores intervalares, o que significa afirmar que os valores intervalares da média possuem boa qualidade de aproximação e representação do valor real da média.

Em seguimento ao trabalho iniciado, pode-se calcular o intervalo de confiança para todos os indicadores estatísticos abordados neste trabalho, e verificar a qualidade dos resultados intervalares obtidos. Além disso, seria interessante construir uma abordagem intervalar para o intervalo de confiança, ou seja, calcular o intervalo de confiança com dados intervalares. No caso do intervalo de confiança para a média, considera como entrada o desvio padrão intervalar e a média intervalar.

REFERÊNCIAS

ACIOLY, B. M. **Fundamentação Computacional da Matemática Intervalar**. 1991. 263p. Tese (Doutorado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. **Estatística Aplicada à Administração e Economia**. São Paulo: Thomson Learning, 2003.

CAMPOS, M. A. **Uma Extensão Intervalar para a Probabilidade Real**. 1997. 127p. Tese (Doutorado em Informática) - Departamento de Informática, UFPE, Recife.

CAMPOS, M. A.; FARIA, R. A. Definição de alguns indicadores estatísticos usando intervalos. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, CNMAC, 11., 1988, Ouro Preto. **Resumo dos trabalhos**. Ouro Preto: SBMAC, 1988.

CAMPOS, M. A. et al. Netbook: uma ferramenta para avaliação de desempenho em redes de comunicação. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES, SBRC, 22., 2004, Gramado. **Anais...** Gramado: II/UFRGS, 2004. p. 967-974.

CHRISTMANN, R. U. **Estatística Aplicada**. São Paulo: E. Blucher, 1978.

CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L. **Introduction to Algorithms**. 2nd ed. Massachusetts: MIT, 1990.

DENNIS, D.; KREINOVICH, V.; RUMP, S. M. An overview of computation complexity intervals and the origins of calculus. **Reliable Computing**, Lafayette, v.4, p. 191-197, 1998.

FERSON, S. et al. Exact bounds on sample variance of interval data. In: SIAM WORKSHOP ON VALIDATED COMPUTING, 2002, Toronto. **Extended abstracts**. El Paso:[SIAM], 2002.

FERSON, S. et al. Absolute bounds on the mean of sum, product, etc.: A probabilistic extension of interval arithmetic. In: SIAM WORKSHOP ON VALIDATED COMPUTING, 2002, Toronto. **Extended abstracts**. El Paso:[SIAM], 2002.

FERSON, S.; GINZBURG, L.; KREINOVICH, V.; LONGPRE, L.; AVILES, M. Computing variance for interval data is NP-Hard. **ACM SIGACT News**, [S.l.], v.2, p. 108-118, 2002.

FERSON, S.; GINZBURG, L.; KREINOVICH, V.; LONGPRE, L.; AVILES, M. Exact bounds on finite populations of interval data. **Reliable Computing**, Lafayette, v.3, p. 207-233, 2004.

GAGANOV, A. A. Computational complexity of the range of the polynomial in several variables. **Cybernetics**, [S.l.], v.11, p. 418-421, 1985.

GAREY, M. E.; JOHNSON, D. S. **Computers and intractability**: a guide to the theory of NP-completeness. San Francisco: Freeman, 1979.

HOROWITZ, E.; SAHNI, S. **Fundamentals of Computer Algorithms**. [S.l.]: Computer Science Press, 1978. 626p.

IMC Índice de Massa Corporal. Disponível em:<www.copacabanarunners.net/imc.html>. Acesso em: 12 jul. 2005.

JAIN, R. **The art ocomputer systems performance analysis**: Techniques for experimental design, measurement, simulation and modeling. São Paulo: E. Blucher, 1978.

KEARFOTT, R. B. Interval computations: Introduction, uses, and resources. **Euromath Bulletin**, [S.l.], v.2, p. 95-112, 1996.

KEARFOTT, R. B.; KREINOVICH, V. **Applications of Interval Computations**: an introduction. Dordrecht: Kluwer, 1996.

KREINOVICH, V.; LAKEYEV, A.; ROHN, J.; KAHL, P. **Computational Complexity and Feasibility of Data Processing and Interval Computations**. Dordrecht: Kluwer, 1998.

KREINOVICH, V. Probabilities, intervals, what next? Optimization problems related to extension of interval computations to situations with partial information about probabilities. **Global Optimization**, [S.l.], v.3, p. 265-280, 2003.

KULISCH, U. W.; MIRANKER, W. L. **Computer Arithmetic in Theory and Practice**. New York: Academic Press, 1981.

LEWIS, H. R.; PAPADIMITRIOU, C. **Elements of the Theory of Computations**. Englewood Cliff: Prentice Hall, 1981.

LIPSCHUTZ, S. **Probabilidade**. São Paulo: McGraw-Hill, 1972.

LORETO, A. B. et al. Is to estimate the expected value of an interval function NP-Hard?. In: INTERNATIONAL SYMPOSIUM ON SCIENTIFIC COMPUTING, COMPUTER ARITHMETIC AND VALIDATED NUMERICS, SCAN, 2004, Fukuoka. **Proceedings...** Fukuoka: Kyshu University, 2004.

LORETO, A. B. et al. Analisando a complexidade computacional de medidas de tendência central e dispersão. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, SBMAC, 28., 2005, São Paulo. **Anais...** São Paulo: SENAC, 2005. p. 01-06.

LORETO, A. B.; SILVA, R.; CLAUDIO, D. M.; TOSCANI, L. V.; LEAL, L. A.; RIBEIRO, L. Amostragem por importância para estimar valores esperados: uma abordagem com heurísticas para problemas intervalares NP-Difíceis. **Tendências em Matemática Aplicada e Computacional**, TEMA, [S.l.], v.1, n.6, 2005.

MAPLE Power Interval Arithmetic. Disponível em :<www.math.uni-wuppertal.de/wrswt/software/intpakX>. Acesso em: 31 out. 2005.

MATIYASEVICH, Y. V. Enumerable sets are diophantine. **Soviet Math. Doklady**, [S.l.], v.11, p. 354-357, 1970.

MATLAB Intilab. Disponível em :<www.mathworks.com>. Acesso em: 20 nov. 2005.

MEYER, P. L. **Probabilidade, aplicações à Estatística**. Rio de Janeiro: LTC, 1968.

MOORE, R. E. **Interval Analysis**. Englewood Cliffs: Prentice Hall, 1966.

MOORE, R. E. On computing the range of a rational function of variables over a bounded region. **Computing**, [S.l.], v.16, p. 1-15, 1976.

MOORE, R. E. **Methods and Applications of Interval Analysis**. Philadelphia: Society for Industrial and Applied Mathematics, 1979.

NESTEROV, S.; KREINOVICH, V. The worse, the better: a survey of paradoxical computational complexity of interval computations. In: WORKSHOP ON COMPUTER ARITHMETIC, INTERVAL AND SYMBOLIC COMPUTATIONS-WAI, 1996, Recife. **Proceedings . . .**. Recife: SBC, 1988.

OLIVEIRA, P. W.; DIVERIO, T. A.; CLAUDIO, D. M. **Fundamentos de Matemática Intervalar**. Porto Alegre: II/UFRGS, Sagra-Luzzato, 1997.

PAPADIMITRIOU, C. **Computational Complexity**. Reading: Addison-Wesley, 1994.

RATSCHEK, H.; ROKNE, J. **New Computer Methods for Global Optimization**. Chichester: Ellis Horwood, 1988.

RUDIN, W. **Real and Complex Analysis**. Great Britain: McGraw-Hill, 1970.

SUNAGA, T. Theory of an interval algebra and its applications to numerical analysis. **RAAG Memoirs**, [S.l.], v.2, p. 547-564, 1958.

SZWARCFITER, J. **Grafos e Algoritmos Computacionais**. Rio de Janeiro: Campus, 1984.

TOSCANI, L. V.; VELOSO, P. A. **Complexidade de Algoritmos: análise, projetos e métodos**. Porto Alegre: II/UFRGS, Sagra-Luzzato, 2001.

TRAYLOR, B.; KREINOVICH, V. A bright side of NP-harness of interval computations: interval heuristics applied to NP-problems. **Reliable Computing**, Lafayette, v.3, p. 343-359, 1995.

VACCARO, G. L. R. **Soluções de Equações Intervalares**. 2001. 450p. Tese (Doutorado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.

VIUNISKI, N. Pontos de corte de IMC para sobrepeso e obesidade em crianças e adolescentes. **Revista Abeso**. Disponível em: <www.abeso.org.br/revista/revista3/imc.html>. Acesso em: 15 jul. 2005.

VUOLO, J. H. **Fundamentos da Teoria dos Erros**. São Paulo: Blücher, 1996.

WU, B. et al. Real-time algorithms for statistical analysis of interval data. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY, INTECH, 2003, Thailand. **Proceedings ...** Thailand: [s.n.], 2003. p. 483-490.

XIANG, G. et al. New algorithms for statistical analysis of interval data. In: WORKSHOP ON STATE-OF-THE-ART IN SCIENTIFIC COMPUTING, PARA, 2004, Lyngby. **Proceeding ...** [Lyngby, Denmark: Technical University of Denmark], 2004. p. 123-129.

ANEXO A INDICADORES ESTATÍSTICOS

A.1 Média Aritmética

$$me = \frac{1}{n} \sum_{i=1}^n x_i. \quad (\text{A.1})$$

A.2 Mediana

$$md = \begin{cases} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) / 2, & \text{se } n \text{ for par} \\ x_{(\frac{n+1}{2})}, & \text{se } n \text{ for ímpar.} \end{cases} \quad (\text{A.2})$$

A.3 Moda

$$mo = x_i \quad (\text{A.3})$$

onde $f_i = \max_{j=1, \dots, n} f_j$ e f_j representa a frequência de x_j .

A.4 Amplitude Total

$$at = ma - mi, \quad (\text{A.4})$$

onde $mi = \min\{x_i\}_{1 \leq i \leq n}$ e $ma = \max\{x_i\}_{1 \leq i \leq n}$.

A.5 Variância

$$va = \frac{1}{n-1} \sum_{i=1}^n (x_i - me)^2. \quad (\text{A.5})$$

A.6 Desvio padrão

$$dp = +\sqrt{va}. \quad (\text{A.6})$$

A.7 Coeficiente de Variação

$$cv = \frac{dp}{me}. \quad (\text{A.7})$$

A.8 Covariância

$$co = \frac{1}{n-1} \sum_{i=1}^n (x_i - me_X)(y_i - me_Y). \quad (\text{A.8})$$

Na Equação A.8, utiliza-se a notação me_X para a média aritmética da variável X e me_Y para a média da variável Y . A covariância e o coeficiente de correlação são medidas de associação entre duas variáveis.

A.9 Coeficiente de Correlação

$$cc = \frac{co}{dp_X dp_Y}, \quad (\text{A.9})$$

onde co é covariância, dp_X o desvio padrão de X e dp_Y o desvio padrão de Y .

A.10 Separatrizes

As separatrizes não são medidas de tendência central, mas estão ligadas à mediana relativamente à sua característica de separar a série em duas partes que apresentam o mesmo número de valores.

Essas medidas - os quartis, os decis e os percentis - são, juntamente com a mediana, conhecidas pelo nome genérico de separatrizes.

A.10.1 Quartis

São os valores de uma série que a dividem em quatro partes iguais. Precisamos portanto de 3 quartis (Q1, Q2 e Q3) para dividir a série em quatro partes iguais.

A.10.2 Decis

A definição dos decis obedece ao mesmo princípio dos quartis, com a modificação da porcentagem de valores que ficam aquém e além do decil que se pretende calcular. Indicamos os decis : D1, D2, ... , D9. Deste modo precisamos de 9 decis para dividirmos uma série em 10 partes iguais.

A.10.3 Percentil ou Centil

Denominamos percentis ou centis como sendo os noventa e nove valores que separam uma série em 100 partes iguais. Indicamos: P_1, P_2, \dots, P_{99} .

ANEXO B ALGORITMOS E ANÁLISE DA COMPLEXIDADE

B.1 Algoritmo da Média Intervalar

ALGORITMO MI	
1.	$s=[0,0];$
2.	para i de 1 até n faça
3.	$s = [\inf(s)+\inf(x[i]); \sup(s)+ \sup(x[i])];$
4.	fim-para;
5.	$ME = [\inf(x)/(n-1); \sup(s)/(n-1)];$
6.	pare-com-saída(ME).

Figura B.1: Algoritmo do cálculo da Média Intervalar

Análise da complexidade do algoritmo MI: o algoritmo começa atribuindo valores nulos aos intervalos S e ME . Toma-se como operação fundamental a adição (+) e como tamanho da entrada o número n de valores intervalares. Verificamos que MI executa $n - 1$ vezes a linha 3, fazendo $n - 1$ adições.

Com isso, podemos afirmar a complexidade do algoritmo MI é de ordem linear, $O(n)$, ou seja, requer $O(n)$ passos computacionais (adição) para n valores intervalares $x_i = [\underline{x}_i, \overline{x}_i]$.

B.2 Algoritmo Ordenação de Intervalos

```

                                ALGORITMO ORD

1. temp;
2. ValOrd = vetor[1, . . . , n];
3. x=vetor[1, . . . , n] {de intervalos};

4. Procedimento Classif-Intervalo Disjunto(ValOrd:x);
5. para i de 1 até n-1 faça
6.     para j de 1 até n-i faça
7.         se  $\text{inf}(x[j]) > \text{inf}(x[j+1])$  e  $\text{sup}(x[j]) > \text{sup}(x[j+1])$ 
8.             então  $\text{temp} = x[j]$ ;
9.                  $x[j] = x[j+1]$ ;
10.                 $x[j+1] = \text{temp}$ ;
11.     fim-para;
12. fim-para;
13. retorne-saída(ValOrd);
14. fim Procedimento;

15. Procedimento Classif-Intervalo Encaixado(ValOrd:x);
16. para i de 1 até n-1 faça
17.     para j de 1 até n-i faça
18.         se  $\text{inf}(x[j]) < \text{inf}(x[j+1])$  e  $\text{sup}(x[j]) \geq \text{sup}(x[j+1])$ 
19.             então  $\text{temp} = x[j]$ ;
20.                  $x[j] = x[j+1]$ ;
21.                  $x[j+1] = \text{temp}$ ;
22.     fim-para;
23. fim-para;
24. retorne-saída(ValOrd);
25. fim Procedimento;
26. pare-com-saída(ValOrd).
```

Figura B.2: Algoritmo de Ordenação Intervalar

Análise da complexidade do algoritmo ORD: O algoritmo ORD classifica em ordem ascendente um vetor x de intervalos de dimensão n , varrendo-o n vezes. Consiste de dois procedimentos, conforme forem os intervalos, isto é, se todos os intervalos do vetor forem disjuntos executa o procedimento Classif-Intevalo Disjuntos e se todos os intervalos forem encaixados, executa o procedimento Classif-Intervalo Encaixado. Tomamos para o tamanho da entrada a dimensão n do vetor x . As linhas 6-11, ou 17-22 (conforme o procedimento) são executadas $n - i$ vezes, portanto depende do valor de i , que varia em cada iteração das linhas 5-12, ou 16-23. A complexidade das linhas 7 e 18

é a complexidade da comparação mais a complexidade das linhas 8,9 e 10, e 19, 20 e 21 respectivamente. O algoritmo ORD possui, portanto, complexidade de ordem $O(n^2)$.

B.3 Algoritmo da Moda Intervalar

Análise da complexidade do algoritmo MOI: o algoritmo começa atribuindo valores nulos às variáveis de controle tempinf, tempsup, cont1 e cont2 e valor unitário à variável t. Define o valor modal intervalar MO como sendo um vetor de tamanho n e o vetor auxiliar MOAUX. Toma-se como operações fundamentais a adição (+) e a comparação, e como tamanho da entrada o número n de valores intervalares. Verificamos na linha 14 uma condição com duas comparações e uma adição para esta condição. No pior caso, o algoritmo MOI executa n-1 vezes as linhas 14-15. Nas linhas 17-22 temos complexidade constante, e nas linhas 10-23 e 27-32 temos n-1 vezes a execução das linhas 11-22 e 28-31 respectivamente (no pior caso). Considerando que nas linhas 13-16 temos complexidade $O(n)$, nas linhas 17-22 temos complexidade constante e nas linhas 10-23 e 27-32 temos complexidade $O(n)$, podemos concluir que o algoritmo MOI tem complexidade $O(n^2)$.

B.4 Algoritmo Separatrizes Intervalares

Análise da complexidade do algoritmo SI: o algoritmo SI recebe o vetor de intervalos ordenados através do vetor ValOrd e apresenta apenas operações aritméticas na linha 4. A complexidade do algoritmo é de ordem $O(n^2)$ devido a operação fundamental ser a comparação entre os intervalos, realizado no algoritmo B.2.

B.5 Algoritmo Amplitude Intervalar

Análise da complexidade do algoritmo ATI: o algoritmo ATI recebe o vetor de intervalos ordenados através do vetor ValOrd e apresenta apenas a operação de subtração (complexidade constante) na linha 2. A complexidade do algoritmo é de ordem $O(n^2)$ devido a operação fundamental ser a comparação entre os intervalos, realizado no algoritmo B.2.

B.6 Algoritmo da Variância Intervalar

Análise da complexidade do algoritmo VI: o algoritmo começa atribuindo valores nulos aos intervalos S, VA e V. Nas linhas 4-6 toma-se como operação fundamental a adição (+) e como tamanho da entrada o número n de valores intervalares. Verificamos que VI executa n-1 vezes a linha 5 fazendo n-1 adições. Nas linhas 8-15, se considerarmos a comparação e operação ao quadrado como operações fundamentais, no pior caso, temos duas condições com uma comparação cada e duas operações ao quadrado para cada condição, realizadas em n-1 passos computacionais. Dessa forma, o algoritmo VI faz (n-1) adições e (n-1)(2 comparações + 2 operações ao quadrado), ou seja, o algoritmo realiza um número linear de operações fundamentais em tempo linear.

Com isso, podemos afirmar que o algoritmo VI tem complexidade de ordem linear, $O(n)$, ou seja, requer $O(n)$ passos computacionais (adição + comparação + operação ao

quadrado) para n valores intervalares $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.

B.7 Algoritmo do Desvio Padrão Intervalar

Análise da complexidade do algoritmo DPI: para computar o intervalo do desvio padrão intervalar, primeiramente devemos computar o intervalo da variância intervalar utilizando o algoritmo VI descrito no anexo B.6. O algoritmo DPI consiste de apenas dois passos, começa atribuindo os valores do intervalo da variância intervalar e calcula o intervalo do desvio padrão intervalar fazendo duas operações de raiz quadrada positiva (do limite inferior da variância intervalar e do limite superior da variância intervalar). Para a complexidade do algoritmo DPI deve-se considerar a complexidade do algoritmo VI. Dessa forma, o algoritmo DPI tem complexidade de ordem $O(n)$ (complexidade do algoritmo VI) + $O(1)$ (ordem constante devido as duas operações de raiz quadrada).

B.8 Algoritmo do Coeficiente de Variação Intervalar

Observamos que o intervalo do desvio padrão intervalar será sempre positivo (pois o desvio padrão é a raiz quadrada positiva da variância) $DP \geq [0, 0]$.

Análise da complexidade do algoritmo CVI: como o intervalo do coeficiente de variação é calculado a partir dos intervalos do desvio padrão intervalar e do intervalo da média intervalar, devemos primeiramente calcular estes intervalos através dos algoritmos descritos nos anexos B.1 e B.7.

A complexidade do algoritmo CVI é de ordem $O(n)$ (complexidade do algoritmo MI) + $O(n)$ (complexidade do algoritmo DPI) + $O(1)$ (ordem constante devido a operação de divisão), ou seja, $O(n)$.

B.9 Algoritmo da Covariância Intervalar

Análise da complexidade do algoritmo COI: Nas linhas 4-7 toma-se como operação fundamental a adição (+) e como tamanho da entrada o número n de valores intervalares. Verificamos que COI executa $n - 1$ vezes as linhas 5 e 6, fazendo $n - 1$ vezes 4 adições. Nas linhas 8 e 9 temos 4 operações de divisão.

Como temos de realizar $n - 1$ vezes as linhas 10-15, podemos afirmar que o algoritmo COI tem complexidade de ordem $O(n)$.

B.10 Algoritmo do Coeficiente de Correlação Intervalar

Devido o intervalo do desvio padrão intervalar ser sempre positivo (pois o desvio padrão é a raiz quadrada positiva da variância), $DPX \geq [0, 0]$ e $DPY \geq [0, 0]$, o produto entre estes dois intervalos também será positivo.

Análise da complexidade do algoritmo CCI: como o intervalo do coeficiente de correlação é calculado a partir dos intervalos do desvio padrão intervalar dos dados intervalares $\mathbf{x}_1, \dots, \mathbf{x}_n$, do desvio padrão intervalar dos dados $\mathbf{y}_1, \dots, \mathbf{y}_n$ e do intervalo da covariância intervalar, devemos primeiramente calcular estes intervalos através dos algo-

ritmos descritos nos anexos B.7 e B.9. Analisando o algoritmo CCI verificamos que na linha 5 temos uma multiplicação e uma divisão. Neste caso, estamos considerando que a divisão é a operação fundamental, então a complexidade da linha 5 é constante.

Como temos de considerar as execuções dos algoritmos da covariância intervalar COI, que tem complexidade de $O(n)$, e do desvio padrão intervalar DPI, que tem complexidade de $O(n)$, podemos afirmar que o algoritmo CCI tem complexidade de ordem $O(n)$.

ALGORITMO MOI

1. ValOrd=ORD[x1, . . . , xn];
2. tempinf=0;
3. tempsup=0;
4. temp=0;
5. cont1=0;
6. cont2=0;
7. t=1;
8. MO = vetor[1 . . . n];
9. MOAUX=vetor[1 . . . n];
10. para j de 1 até n faça
11. tempinf=inf(x[j]);
12. tempsup=sup(x[j]);
13. para i de 1 até n faça
14. se tempinf==inf(x[i]) e tempsup==sup(x[i])
15. então cont1= cont1 + 1;
16. fim-para;
17. se cont1 <> 1 e cont1 >=cont2
18. então MOAUX[t]=x[j];
19. cont2=cont1;
20. t = t+1;
21. cont2=cont1;
22. cont1=0;
23. fim-para;
24. MO[1]=MOAUX[1];
25. temp=MOAUX[1];
26. j=1;
27. para i de 1 até n faça
28. se MOAUX[i] <> temp
29. então j = j+1;
30. MO[j]=MOAUX[i];
31. temp=MOAUX[i];
32. fim-para;
33. pare-com-saída(MO).

Figura B.3: Algoritmo do cálculo da Moda Intervalar

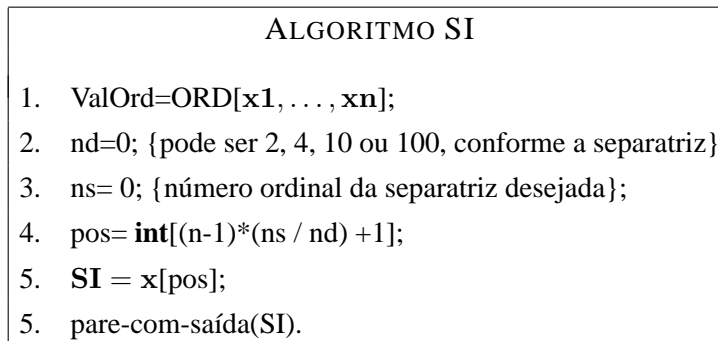


Figura B.4: Algoritmo Separatriz Intervalar

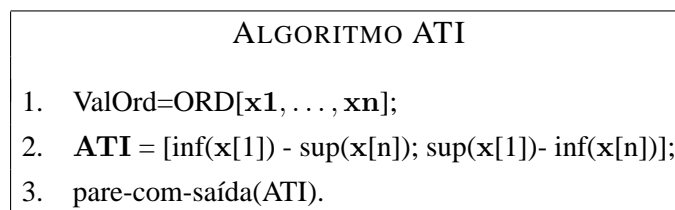


Figura B.5: Algoritmo Amplitude Intervalar

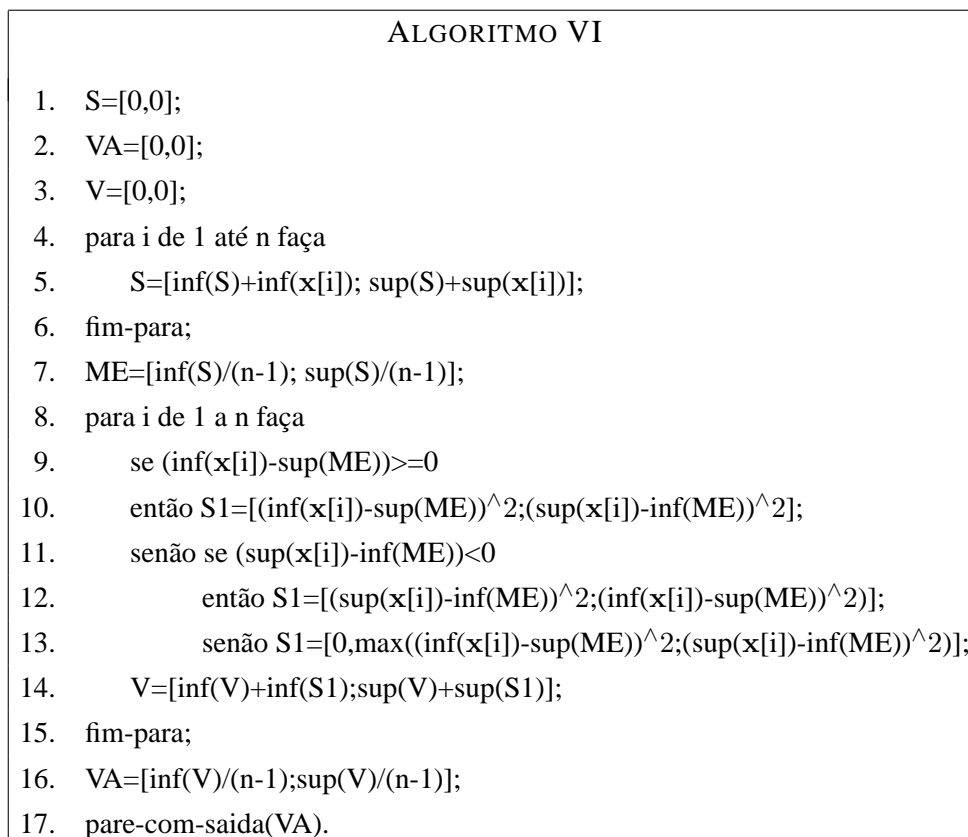


Figura B.6: Algoritmo do cálculo da Variância Intervalar

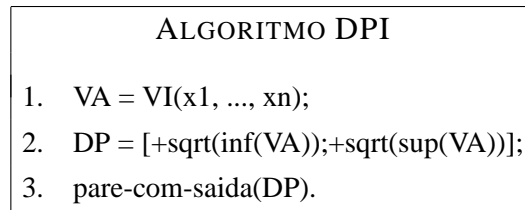


Figura B.7: Algoritmo do cálculo do Desvio padrão Intervalar

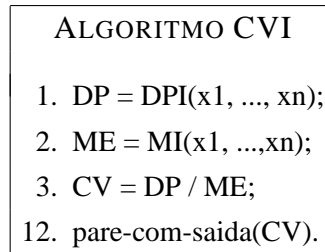


Figura B.8: Algoritmo do cálculo do Coeficiente de Variação Intervalar

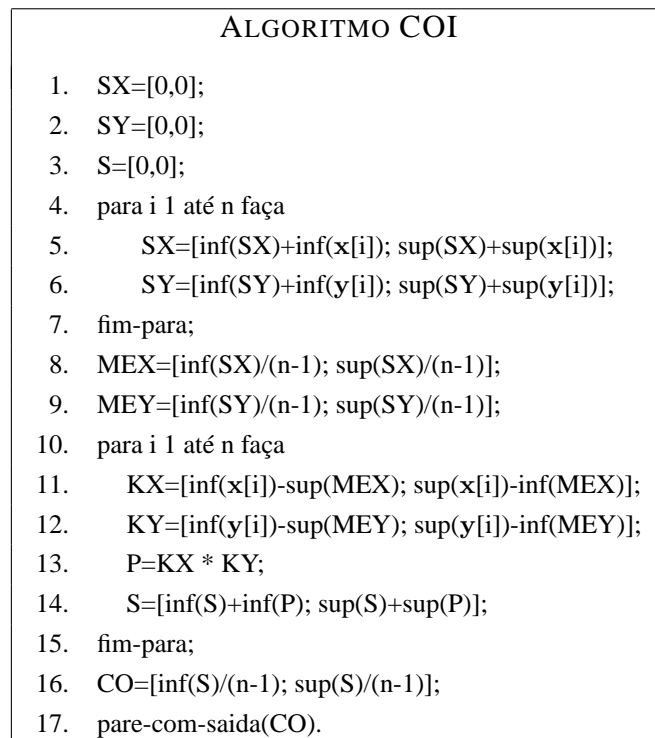


Figura B.9: Algoritmo do cálculo da Covariância Intervalar

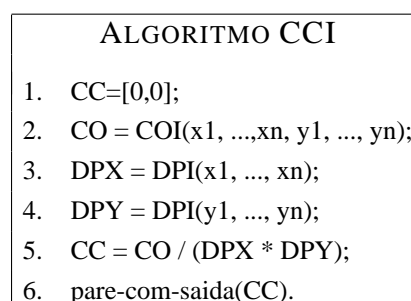


Figura B.10: Algoritmo do cálculo do Coeficiente de Correlação Intervalar

ANEXO C ALTURA, PESO E IDADE DOS ALUNOS DA ESCOLA IMACULADA CONCEIÇÃO

C.1 Altura, peso e idade da 5a. série

Tabela C.1: Altura, peso e idade 5a. série.

Meninas			Meninos		
Altura	Peso	Idade	Altura	Peso	Idade
1.44	35	10	1.31	26	9
1.45	30	10	1.53	35	11
1.43	42.5	11	1.51	43.4	10
1.46	39	11	1.42	31	10
1.41	39.4	10	1.40	34	11
1.49	60	10	1.47	34	10
1.60	58	10	1.50	47	10
1.49	49	10			
1.53	63	11			
1.51	48	10			
1.60	63	11			
1.48	55	10			
1.32	28	10			
1.46	46.5	10			

C.2 Altura, peso e idade da 6a. série

C.3 Altura, peso e idade da 7a. série

C.4 Altura, peso e idade da 8a. série

Tabela C.2: Altura, peso e idade 6a. série.

Meninas			Meninos		
Altura	Peso	Idade	Altura	Peso	Idade
1.48	59	12	1.55	51	12
1.44	34	12	1.56	40	12
1.50	40	12	1.51	63	12
1.60	43	12	1.68	76	13
1.41	30	12	1.50	42	12
1.76	71	13	1.48	36	12
1.54	46	13	1.42	36	12
1.55	52	12	1.56	48	12
1.53	51.5	13	1.50	36	12
1.49	37	12	1.56	35	12
1.38	30	13			
1.56	46.5	12			
1.55	48	12			
1.53	52	12			
1.55	55	12			

Tabela C.3: Altura, peso e idade 7a. série.

Meninas			Meninos		
Altura	Peso	Idade	Altura	Peso	Idade
1.59	51	12	1.51	41	12
1.52	47	13	1.42	41	12
1.63	56	12	1.50	35	11
1.61	55	15	1.57	52	12
1.56	59	12	1.61	55	12
1.56	51	13	1.48	38	12
1.53	38	11	1.53	40	12
1.50	38	12	1.58	72.9	13
1.55	51.	12			
1.54	37	13			
1.54	37	13			
1.50	41	12			
1.50	35	12			
1.52	72	12			
1.49	31	12			
1.45	35	12			
1.60	52	14			
1.46	37.5	12			
1.57	45	12			
1.64	42	12			
1.58	44	13			

Tabela C.4: Altura, peso e idade 8a. série.

Meninas			Meninos		
Altura	Peso	Idade	Altura	Peso	Idade
1.58	40	15	1.64	44.1	13
1.66	55	14	1.65	76.5	14
1.65	72	14	1.54	38.2	13
1.66	57.5	14	1.71	77.1	14
1.60	43	14	1.60	53.1	14
1.67	58.5	14	1.67	50.2	14
1.54	50.5	13	1.69	93	14
1.49	42.5	14	1.71	83.5	16
1.52	55.5	15	1.65	48.1	16
1.49	41	15	1.79	70.9	15
1.57	41.5	15	1.70	72	15
1.69	64	13	1.62	53.4	14
1.59	45	14	1.65	50.5	15
1.61	47.5	13	1.68	51	14
1.54	42	14	1.71	60.6	14
1.52	48	14	1.71	54	14
1.55	46.5	14	1.61	42.8	14
1.57	57.5	15	1.53	45.9	14
1.68	61	14			
1.58	45.5	14			
1.60	62	15			
1.60	47	14			
1.50	41.5	14			

ANEXO D INTERVALO DE CONFIANÇA

Tabela D.1: Graus de liberdade ($n - 1$), Intervalo de confiança e Valor Intervalar da média.

Turmas	Graus de liberdade ($n - 1$)	Intervalo de confiança	Valor Intervalar da média
5a. série - meninas	13	[20.4055, 25.3435]	[22.868, 22.880]
5a. série - meninos	6	[16.2268, 23.2666]	[19.740, 19.753]
6a. série - meninas	14	[19.085, 23.139]	[21.106, 21.118]
6a. série - meninos	9	[18.0660, 25. 330]	[21.693, 21.704]
7a. série - meninas	20	[18.156, 21.739]	[19.941, 19.953]
7a. série -meninos	7	[18.454, 27.209]	[22.826, 22.837]
8a. série - meninas	22	[19.739, 22.205]	[20.976, 20.986]
8a. série - meninos	17	[20.0982, 25.0476]	[22.567, 22.578]