

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**UM ESTUDO QUANTITATIVO SOBRE A VARIABILIDADE DOS  
TEMPOS DE EXECUÇÃO DE PROGRAMAS EM EXPERIMENTOS  
COMPUTACIONAIS**

Paulo Eduardo Nogueira

Uberlândia, Minas Gerais

2015

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**UM ESTUDO QUANTITATIVO SOBRE A VARIABILIDADE DOS TEMPOS DE  
EXECUÇÃO DE PROGRAMAS EM EXPERIMENTOS COMPUTACIONAIS**

Paulo Eduardo Nogueira

Dissertação de Mestrado apresentada à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como parte dos requisitos exigidos para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Engenharia de Software  
Orientador: Prof. Dr. Rivalino Matias Júnior

Uberlândia, Minas Gerais

2015

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da UFU, MG, Brasil.

---

N778e Nogueira, Paulo Eduardo, 1980-  
2015 Um estudo quantitativo sobre a variabilidade dos tempos de  
execução de programas em experimentos computacionais / Paulo  
Eduardo Nogueira. - 2015.  
186 f. : il.

Orientador: Rivalino Matias Júnior.  
Dissertação (mestrado) - Universidade Federal de Uberlândia,  
Programa de Pós-Graduação em Ciência da Computação.  
Inclui bibliografia.

1. Computação - Teses. 2. Sistemas operacionais (Computadores) -  
Teses. 3. Programas de computador - Teses. I. Matias Júnior, Rivalino.  
II. Universidade Federal de Uberlândia, Programa de Pós-Graduação em  
Ciência da Computação. III. Título.

---

CDU: 681.3

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “Um Estudo Quantitativo Sobre a Variabilidade dos Tempos de Execução de Programas em Experimentos Computacionais” por Paulo Eduardo Nogueira como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação.

Uberlândia, Março de 2015.

**Orientador:**

---

Prof. Dr. Rivalino Matias Júnior  
Universidade Federal de Uberlândia

**Banca Examinadora:**

---

Prof. Dr. Autran Macêdo  
Universidade Federal de Uberlândia

---

Prof. Dr. Lúcio Borges de Araújo  
Universidade Federal de Uberlândia

---

Prof. Dr. Paulo Romero Martins Maciel  
Universidade Federal de Pernambuco

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Data: Março de 2015.

Autor:

**Paulo Eduardo Nogueira**

Título:

**Um Estudo Quantitativo Sobre a Variabilidade dos Tempos de Execução de Programas em Experimentos Computacionais**

Faculdade:

**Faculdade de Computação**

Grau:

**Mestrado**

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão de cópias deste documento para propósitos exclusivamente acadêmicos, desde que o autor seja devidamente informado.

---

Autor

O AUTOR RESERVA PARA SI QUALQUER OUTRO DIREITO DE PUBLICAÇÃO DESTE DOCUMENTO, NÃO PODENDO O MESMO SER IMPRESSO E REPRODUZIDO, SEJA NA TOTALIDADE OU EM PARTES, SEM A PERMISSÃO ESCRITA DO AUTOR.

© Todos os direitos reservados a Paulo Eduardo Nogueira

## DEDICATÓRIA

*Aos meus pais Paulo César e Alcina e ao meu irmão Keppler César.*

*À minha amada noiva Rubiana.*

# AGRADECIMENTOS

A Deus,

por abençoar os meus dias, iluminar os meus caminhos e me conceder a oportunidade de contribuir para a evolução do planeta.

Ao meu orientador, Prof. Rivalino Matias Júnior,  
pelas oportunidades que me concedeu, pelo conhecimento que compartilhou e,  
principalmente, pela paciência que teve comigo ao longo dessa jornada. Prof.  
Rivalino, eu realmente serei eternamente grato.

Aos meus pais Paulo César e Alcina,  
pelo amor, pela educação e pelo apoio que sempre me deram.

Ao meu irmão Keppler César,  
pela a oportunidade de crescer ao seu lado. Que esta conquista seja inspiração para  
as suas, nunca desista, meu irmão.

A minha noiva Rubiana,  
pelo seu amor, pelo carinho, pela fidelidade, pela cumplicidade e por sua  
compreensão. Que possamos crescer juntos, construindo o nosso futuro.

Aos amigos da ETEc Pedro Badran e do Instituto Federal Goiano,  
pelo carinho, pela a amizade e por permitirem a conclusão deste sonho.

Ao Prof. Cacildo Devós,  
por ter sido um dos primeiros a acreditar e confiar no meu potencial.

Aos meus tantos amigos da UFU,  
pelas conversas, conselhos, piadas e por aí vai. Quero agradecer a todos os que  
estiveram comigo nesta jornada. Estarei sempre feliz com o sucesso de cada um de  
vocês.

*“Nem tudo que reluz é luz.”*

**Irmã Consuelo**

# RESUMO

Na pesquisa experimental em Ciência da Computação muitos trabalhos dependem da correta mensuração e da análise dos tempos de execução de programas de computador. Observa-se que nem todos consideram que repetidas execuções do mesmo programa possam produzir tempos de execução significativamente diferentes em termos estatísticos. Em virtude disso, realizou-se um estudo quantitativo sobre a variabilidade dos tempos de execução de programas, com o objetivo de estabelecer um protocolo para análise quantitativa comparativa de tempos de execução de programas em experimentos computacionais. Assim sendo, três experimentos foram configurados e executados de modo a permitirem a observação de alguns fatores presentes no ambiente computacional, em especial, relacionados com o sistema operacional. Os fatores escolhidos foram: o *Runlevel*, a *Otimização* do compilador, o Tamanho de uma Variável de Ambiente, o Número de *Threads* e as Estratégias de Alocação de *Threads*. Os resultados obtidos com o protocolo RTA demonstram que esses fatores são capazes de influenciar os tempos de execução dos programas, levando a conclusões equivocadas. Em alguns casos, a diferença estatística entre dois tratamentos de um experimento chegou a 100% das comparações. Além disso, observou-se que a distribuição dos tempos de execução nem sempre é aderente a uma distribuição Gaussiana. Observou-se, também, que, em análises de significância com múltiplos tratamentos, o problema conhecido como *familywise error rate* deve ser considerado e deve ser tratado, pois ele também pode levar a conclusões equivocadas. Assim, essas observações enfatizam o uso de um ambiente experimental controlado e o uso de uma metodologia estatística para a análise estatística de múltiplas comparações.

**Palavras-chave:** Tempos de Execução. Variabilidade. Sistemas Operacionais. Múltiplas Comparações.

## ABSTRACT

In experimental research in computer science many works depend on correct measurement and analysis of the execution times of computer programs. It is observed that not all hold that repeated executions of the same program can produce significantly different execution times in statistical terms. Because of this, there was a quantitative study on the variability of programs execution times, in order to establish a protocol for comparative quantitative analysis of execution times of programs in computational experiments. Subsequently, three experiments were set up and implemented so as to allow the observation of some factors present in the computing environment, in particular related to the operating system. The factors chosen were: Runlevel, compiler Optimization, the Size of an Environment Variable, the Number of Threads and Thread Allocation Strategies. The results obtained with the RTA protocol demonstrate that these factors can influence the program execution times, leading to erroneous conclusions. In some cases, the statistical difference between treatments in an experiment reached 100% of the comparisons. Furthermore, it was observed that the distribution of execution times is not always adhered to a Gaussian distribution. It was observed also that in significance analysis with multiple treatments, the problem known as familywise error rate should be considered and should be treated because it can also lead to wrong conclusions. Thus, these observations emphasize the use of a controlled computing environment and using a statistical method for statistical analysis of multiple comparisons.

**Keywords:** Execution Times. Variability. Operating Systems. Multiple Comparisons.

# SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 CONTEXTUALIZAÇÃO .....	1
1.2 RELEVÂNCIA DA PESQUISA .....	3
1.3 OBJETIVOS DA PESQUISA .....	4
<i>Geral:</i> .....	4
<i>Específicos:</i> .....	4
1.4 DESENVOLVIMENTO DA PESQUISA .....	4
1.5 ESTRUTURA DO DOCUMENTO .....	5
<b>2 REVISÃO DA LITERATURA.....</b>	<b>6</b>
2.1 INTRODUÇÃO.....	6
2.2 MÉTODOS ESTATÍSTICOS.....	6
2.2.1 <i>Teste t</i> .....	7
2.2.2 <i>Teste de Wilcoxon-Mann-Whitney</i> .....	9
2.2.3 <i>Testes de Aderência</i> .....	10
2.2.4 <i>Análise de Variância com o Teste F de Fisher</i> .....	12
2.2.5 <i>Múltiplas Comparações</i> .....	13
2.3 TRABALHOS RELACIONADOS.....	14
2.4 CONSIDERAÇÕES FINAIS .....	19
<b>3 ESTUDO EXPLORATÓRIO EXPERIMENTAL SOBRE VARIAÇÃO EM TEMPOS DE EXECUÇÃO DE PROGRAMAS .....</b>	<b>20</b>
3.1 INTRODUÇÃO.....	20
3.2 PLANEJAMENTO DE EXPERIMENTOS .....	21
3.2.1 <i>Método</i> .....	21
3.2.2 <i>Benchmarks</i> .....	21
3.2.3 <i>Ambiente de Teste e Instrumentação</i> .....	27
3.2.4 <i>Experimento #1</i> .....	29
3.2.5 <i>Experimento #2</i> .....	30
3.2.6 <i>Experimento #3</i> .....	31
3.2.7 <i>Experimento #4</i> .....	32
3.3 ANÁLISE DOS RESULTADOS .....	33
3.3.1 <i>Experimento #1</i> .....	33

3.3.2 Experimento #2 .....	38
3.3.3 Experimento #3 .....	44
3.3.4 Experimento #4 .....	49
3.4 CONSIDERAÇÕES FINAIS.....	51
<b>4 O PROTOCOLO SPEEDUP-TEST.....</b>	<b>52</b>
4.1 INTRODUÇÃO .....	52
4.2 PROTOCOLO SPEEDUP-TEST .....	52
4.2.1 Analisando a média dos tempos de execução .....	53
4.2.2 Analisando a mediana dos tempos de execução .....	55
4.3 ANÁLISE DE RESULTADOS.....	56
4.3.1 Experimento #1 .....	57
4.3.2 Experimento #2 .....	60
4.3.3 Experimento #3 .....	68
4.3.4 Experimento #4 .....	72
4.3.5 Críticas ao Protocolo Speedup-Test .....	74
<b>5 O PROTOCOLO RUNTIME ANALYSIS .....</b>	<b>76</b>
5.1 INTRODUÇÃO .....	76
5.2 PROPOSTAS DE MELHORAMENTO DO PROTOCOLO SPEEDUP-TEST .....	76
5.2.1 Integração da metodologia de análise .....	76
5.2.2 Múltiplas Comparações.....	77
5.3 ANÁLISE DE RESULTADOS.....	78
5.3.1 Experimento #1 .....	79
5.3.2 Experimento #2 .....	80
5.3.3 Experimento #3 .....	83
5.3.4 Experimento #4 .....	85
5.4 CONSIDERAÇÕES FINAIS.....	87
<b>6 CONCLUSÕES.....</b>	<b>88</b>
6.1 PRINCIPAIS RESULTADOS.....	88
6.2 LIMITAÇÕES DA PESQUISA .....	88
6.3 CONTRIBUIÇÕES PARA A LITERATURA.....	88
6.4 DIFICULDADES ENCONTRADAS.....	89
6.5 TRABALHOS FUTUROS .....	89
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>90</b>
<b>APÊNDICE A: CLASSE DE INPUT PARA CADA BENCHMARK NPB (ADAPTADO DE NASA ADVANCED SUPERCOMPUTING DIVISON (2015)) .....</b>	<b>95</b>

APÊNDICE A: CLASSE DE <i>INPUT</i> PARA CADA <i>BENCHMARK NPB</i> (ADAPTADO DE NASA ADVANCED SUPERCOMPUTING DIVISON (2015)) (CONT.).....	96
APÊNDICE B: CONJUNTO DE <i>INPUT</i> PARA CADA PROGRAMA <i>SPLASH-2X</i> (ADAPTADO DE GROUP (2011)) .....	97
APÊNDICE B: CONJUNTO DE <i>INPUT</i> PARA CADA PROGRAMA <i>SPLASH-2X</i> (ADAPTADO DE GROUP (2011)) (CONT.) .....	98
APÊNDICE C: CONJUNTO DE <i>INPUT</i> PARA CADA PROGRAMA <i>PARSEC</i> (ADAPTADO DE BIENIA (2011)) .....	99
APÊNDICE C: CONJUNTO DE <i>INPUT</i> PARA CADA PROGRAMA <i>PARSEC</i> (ADAPTADO DE BIENIA (2011)) (CONT.).....	100
APÊNDICE D: SERVIÇOS EM BACKGROUND DURANTE OS EXPERIMENTOS..	101
APÊNDICE D: SERVIÇOS EM BACKGROUND DURANTE OS EXPERIMENTOS (CONT.).....	102
APÊNDICE E: DISTRIBUIÇÃO DE THREADS PELOS NÚCLEOS SEGUNDO AS ESTRATÉGIAS DE THREAD AFFINITY (INTEL, 2015).....	103
APÊNDICE F: CONFIGURAÇÃO DOS TRATAMENTOS NO EXP. #4.....	104
APÊNDICE F: CONFIGURAÇÃO DOS TRATAMENTOS NO EXP. #4 (CONT.).....	105
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 .....	106
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	107
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	108
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	109
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	110
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	111
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	112
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	113
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	114
APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.).....	115
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 .....	116
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.).....	117

APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	118
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	119
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	120
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	121
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	122
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	123
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	124
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	125
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	126
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	127
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	128
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	129
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	130
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	131
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	132
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	133
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	134
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	135
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	136
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	137
APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.) .....	138

<b>APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.).....</b>	<b>139</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 .....</b>	<b>140</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>141</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>142</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>143</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>144</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>145</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>146</b>
<b>APÊNDICE I: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #3 (CONT.).....</b>	<b>147</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 .....</b>	<b>148</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 (CONT.).....</b>	<b>149</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 (CONT.).....</b>	<b>150</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 (CONT.).....</b>	<b>151</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 (CONT.).....</b>	<b>152</b>
<b>APÊNDICE J: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #4 (CONT.).....</b>	<b>153</b>
<b>APÊNDICE K: PROTOCOLO RUNTIME ANALYSIS.....</b>	<b>154</b>
<b>ANEXO A: DISTRIBUIÇÃO <i>T</i> DE STUDENT (SHESKIN, 2003).....</b>	<b>155</b>
<b>ANEXO B: DISTRIBUIÇÃO GAUSSIANA (GIBBONS; CHAKRABORTI, 2014). ...</b>	<b>156</b>
<b>ANEXO C: VALORES CRÍTICOS PARA <i>U</i> DO TESTE DE MANN-WHITNEY (SHESKIN, 2003). .....</b>	<b>157</b>
<b>ANEXO D: KOLMOGOROV-SMIRNOV UMA AMOSTRA (GIBBONS; CHAKRABORTI, 2014). .....</b>	<b>158</b>
<b>ANEXO E: KOLMOGOROV-SMIRNOV DUAS AMOSTRAS (SHESKIN, 2003).....</b>	<b>159</b>
<b>ANEXO F: DISTRIBUIÇÃO <i>F</i> PARA <math>\alpha = 0,05</math> (SHESKIN, 2003).....</b>	<b>161</b>
<b>ANEXO G: ALGORITMO DO PROGRAMA PROVA. (ADAPTADO DE (SOBRINHO, 2012)).....</b>	<b>162</b>

**ANEXO H: PROTOCOLO SPEEDUP-TEST PARA ANÁLISE DO MÉDIA. (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013))..... 163**

**ANEXO I: PROTOCOLO SPEEDUP-TEST PARA ANÁLISE DA MEDIANA. (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013))..... 164**

# LISTA DE TABELAS

TABELA 1.1. METODOLOGIA DE PESQUISA EM COMPUTAÇÃO (TRADUZIDO DE RAMESH; GLASS; VESSEY (2004)).....	2
TABELA 2.1. TIPOS DE ERROS (ADAPTADO DE BARBETTA; REIS; BORNIA (2010)).....	7
TABELA 2.2. OBSERVAÇÕES PAREADAS (ADAPTADO DE BARBETTA; REIS; BORNIA (2010)).....	7
TABELA 3.1. FATORES E NÍVEIS AVALIADOS NO EXP. #1.....	30
TABELA 3.2. CONFIGURAÇÃO DOS TRATAMENTOS DO EXP. #1.....	30
TABELA 3.3. FATORES E NÍVEIS AVALIADOS NO EXP. #2.....	31
TABELA 3.4. FATORES E NÍVEIS AVALIADOS NO EXP. #3.....	31
TABELA 3.5. CONFIGURAÇÃO DOS TRATAMENTOS DO EXP. #3.....	32
TABELA 3.6. FATORES E NÍVEIS AVALIADOS NO EXP. #4. (ADAPTADO DE (VICENTE; MATIAS JR., 2013)).....	32
TABELA 3.7. DISTRIBUIÇÃO DO MAIOR E MENOR TEMPO POR TRATAMENTO PARA O <i>NPB</i> .....	34
TABELA 3.8. DISTRIBUIÇÃO DO MAIOR E MENOR TEMPO POR TRATAMENTO PARA O <i>PARSEC</i> .....	34
TABELA 3.9. DISTRIBUIÇÃO DA VARIABILIDADE POR TRATAMENTO PARA O <i>PARSEC</i> .....	34
TABELA 3.10. DISTRIBUIÇÃO DO MAIOR E MENOR TEMPO POR TRATAMENTO PARA O <i>SPLASH-2X</i> .....	36
TABELA 3.11. DISTRIBUIÇÃO DO MAIOR E MENOR TEMPO POR ESTRATÉGIA DE <i>THREAD AFFINITY</i> PARA O <i>NPB</i> .....	44
TABELA 3.12. DISTRIBUIÇÃO DA VARIABILIDADE POR ESTRATÉGIA DE <i>THREAD AFFINITY</i> PARA O <i>NPB</i> .44	
TABELA 3.13. DISTRIBUIÇÃO DO MAIOR E MENOR TEMPO POR ESTRATÉGIA DE <i>THREAD AFFINITY</i> PARA O <i>PARSEC</i> .....	48
TABELA 3.14. DISTRIBUIÇÃO DA VARIABILIDADE POR ESTRATÉGIA DE <i>THREAD AFFINITY</i> DO <i>PARSEC</i> .48	
TABELA 4.1. RESULTADO DO TESTE DE ADERÊNCIA A UMA DISTRIBUIÇÃO GAUSSIANA PARA OS TEMPOS DE EXECUÇÃO DOS PROGRAMAS PARA O EXP. #1.....	58
TABELA 4.2. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE DOS PROGRAMAS PARA O EXP. #1. ....	59
TABELA 4.3. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS PARA O EXP. #1 (COM BASE NA MÉDIA).....	59
TABELA 4.4. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO DOS PROGRAMAS PARA O EXP. #1.....	60
TABELA 4.5. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS PARA O EXP. #1 (COM BASE NA MEDIANA).....	60
TABELA 4.6. TESTE DE ADERÊNCIA A UMA DISTRIBUIÇÃO GAUSSIANA PARA OS TEMPOS DE EXECUÇÃO DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO O2</i> PARA O EXP. #2.....	61
TABELA 4.7. TESTE DE ADERÊNCIA A UMA DISTRIBUIÇÃO GAUSSIANA PARA OS TEMPOS DE EXECUÇÃO DOS PROGRAMAS COM <i>OTIMIZAÇÃO O3</i> PARA O EXP. #2.....	62
TABELA 4.8. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO O2</i> PARA O EXP. #2.....	63
TABELA 4.9. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO O3</i> PARA O EXP. #2.....	64

TABELA 4.10. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 02 PARA O EXP. #2 (COM BASE NA MÉDIA).....	64
TABELA 4.11. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 03 PARA O EXP. #2 (COM BASE NA MÉDIA).....	65
TABELA 4.12. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 02 PARA O EXP. #2.....	65
TABELA 4.13. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 03 PARA O EXP. #2.....	66
TABELA 4.14. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 02 PARA O EXP. #2 (COM BASE NA MEDIANA).....	67
TABELA 4.15. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>OTIMIZAÇÃO</i> 03 PARA O EXP. #2 (COM BASE NA MEDIANA).....	67
TABELA 4.16. TESTE DE ADERÊNCIA A UMA DISTRIBUIÇÃO GAUSSIANA PARA OS TEMPOS DE EXECUÇÃO DOS PROGRAMAS PARA O EXP. #3.....	68
TABELA 4.17. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE PARA O EXP. #3.....	69
TABELA 4.18. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>NPB</i> PARA O EXP. #3 (COM BASE NA MÉDIA).....	70
TABELA 4.19. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>PARSEC</i> PARA O EXP. #3 (COM BASE NA MÉDIA).....	70
TABELA 4.20. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO DO EXP. #3.71	
TABELA 4.21. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>NPB</i> PARA O EXP. #3 (COM BASE NA MEDIANA).....	71
TABELA 4.22. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>PARSEC</i> PARA O EXP. #3 (COM BASE NA MEDIANA).....	72
TABELA 4.23. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM 32 TRATAMENTOS PARA O EXP. #4.....	72
TABELA 4.24. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 32 TRATAMENTOS PARA O EXP. #4 (COM BASE NA MÉDIA).....	72
TABELA 4.25. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO COM 32 TRATAMENTOS PARA O EXP. #4.....	73
TABELA 4.26. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 32 TRATAMENTOS PARA O EXP. #4 (COM BASE NA MEDIANA).....	73
TABELA 4.27. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM 64 TRATAMENTOS PARA O EXP. #4.....	73
TABELA 4.28. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 64 TRATAMENTOS PARA O EXP. #4 (COM BASE NA MÉDIA).....	74
TABELA 4.29. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS COM A MESMA DISTRIBUIÇÃO COM 64 TRATAMENTOS PARA O EXP. #4.....	74
TABELA 4.30. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 64 TRATAMENTOS PARA O EXP. #4 (COM BASE NA MEDIANA).....	74
TABELA 5.1. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE DOS PROGRAMAS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #1 .....	79
TABELA 5.2. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #1.....	80

TABELA 5.3. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM O FATOR DE <i>OTIMIZAÇÃO</i> 02 COM O PROTOCOLO <i>RTA</i> PARA O EXP. #2.....	81
TABELA 5.4. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM O FATOR DE <i>OTIMIZAÇÃO</i> 03 COM O PROTOCOLO <i>RTA</i> PARA O EXP. #2.....	82
TABELA 5.5. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>Otimização</i> 02 COM O PROTOCOLO <i>RTA</i> PARA O EXP. #2.....	82
TABELA 5.6. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES DOS PROGRAMAS COM O FATOR DE <i>Otimização</i> 03 COM O PROTOCOLO <i>RTA</i> PARA O EXP. #2.....	83
TABELA 5.7. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM O PROTOCOLO <i>RTA</i> PARA O EXP. #3.....	84
TABELA 5.8. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>NPB</i> COM O PROTOCOLO <i>RTA</i> PARA O EXP. #3.....	85
TABELA 5.9. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM O <i>PARSEC</i> COM O PROTOCOLO <i>RTA</i> PARA O EXP. #3.....	85
TABELA 5.10. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM 32 TRATAMENTOS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #4.....	85
TABELA 5.11. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 32 TRATAMENTOS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #4.....	86
TABELA 5.12. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS POR TIPO DE TESTE COM 64 TRATAMENTOS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #4.....	86
TABELA 5.13. PERCENTUAL DE COMPARAÇÕES DOIS A DOIS ESTATISTICAMENTE DIFERENTES COM 64 TRATAMENTOS COM O PROTOCOLO <i>RTA</i> PARA O EXP. #4.....	86
TABELA 6.1. PUBLICAÇÕES CIENTÍFICAS.....	89

# LISTA DE FIGURAS

FIGURA 3.1. TOPOLOGIA DE UM DOS QUATRO PROCESSADORES USADOS NOS EXP. #1, #2 E #3.....	28
FIGURA 3.2. <i>SCRIPT</i> PARA A COLETA DE TEMPO DE EXECUÇÃO DO PROGRAMA <i>BLACKSCHOLES</i> .....	29
FIGURA 3.3 TOPOLOGIA DOS PROCESSADORES UTILIZADOS NO EXPERIMENTO EXP. #3.....	29
FIGURA 3.4. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> NO EXP.#1.....	35
FIGURA 3.5. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>PARSEC</i> NO EXP.#1.....	37
FIGURA 3.6. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>SPLASH-2X</i> NO EXP.#1.....	38
FIGURA 3.7. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> COM OTIMIZAÇÃO O2 NO EXP.#2... ..	40
FIGURA 3.8. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> COM <i>OTIMIZAÇÃO O3</i> NO EXP.#2....	41
FIGURA 3.9. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>SPLASH-2X</i> COM <i>OTIMIZAÇÃO O2</i> NO EXP.#2.....	42
FIGURA 3.10. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>SPLASH-2X</i> COM OTIMIZAÇÃO O3 NO EXP.#2.....	43
FIGURA 3.11. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> COM 2 <i>THREADS</i> NO EXP. #3.....	45
FIGURA 3.12. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> COM 4 <i>THREADS</i> NO EXP. #3.....	46
FIGURA 3.13. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>NPB</i> COM 6 <i>THREADS</i> NO EXP. #3.....	47
FIGURA 3.14. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>PARSEC</i> COM 2 <i>THREADS</i> NO EXP. #3..	48
FIGURA 3.15. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>PARSEC</i> COM 4 <i>THREADS</i> NO EXP. #3..	49
FIGURA 3.16. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO <i>PARSEC</i> COM 6 <i>THREADS</i> NO EXP. #3.	49
FIGURA 3.17. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO PROG. PROVA PARA OS 32 TRATAMENTOS INICIAS DO EXP. #4.....	50
FIGURA 3.18. VARIABILIDADE DOS TEMPOS DE EXECUÇÃO DO PROG. PROVA PARA OS 64 TRATAMENTOS DO EXP. #4.....	50
FIGURA 4.1. PASSOS DO PROTOCOLO PARA VERIFICAR A ADERÊNCIA DOS DADOS A UMA DISTRIBUIÇÃO GAUSSIANA (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013)).....	54
FIGURA 4.2. PASSOS DO PROTOCOLO PARA ANALISAR A SIGNIFICÂNCIA (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013)). .....	54
FIGURA 4.3. PASSOS DO PROTOCOLO PARA APRESENTAR O RESULTADO DA ANÁLISE (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013)).....	55
FIGURA 4.4. PASSOS DO PROTOCOLO PARA ANALISAR A SIGNIFICÂNCIA (ADAPTADO DE TOUATI; WORMS; BRIAIS, 2013). .....	56
FIGURA 4.5. PASSOS DO PROTOCOLO PARA APRESENTAR O RESULTADO DA ANÁLISE (ADAPTADO DE TOUATI; WORMS; BRIAIS, 2013). .....	56
FIGURA 5.1. ETAPA PARA ANÁLISE DE ADERÊNCIA E COMPARAÇÃO DE TRATAMENTOS COM O PROTOCOLO <i>RTA</i> .....	77
FIGURA 5.2. ETAPA PARA A ANÁLISE DE MÚLTIPLAS COMPARAÇÕES COM O PROTOCOLO <i>RTA</i> .....	78

# LISTA DE ABREVIATURAS E SIGLAS

ANOVA	<i>Analysis of variance</i>
ARM	<i>Association Rule Mining</i>
AVC	<i>Advance Video Coding</i>
CFD	<i>Computational Fluid Dynamics</i>
CFS	<i>Completely Fair Scheduler</i>
ccNUMA	<i>Cache Coherent Non-Uniform Memory Access</i>
DOE	<i>Design of Experiments</i>
FIMI	<i>Frequent Itemset Mining</i>
FX	<i>Fixed Priority</i>
FP-growth	<i>Frequent Pattern-growth</i>
HJM	<i>Health Jarrow Morton</i>
HPC	<i>High Performance Computing</i>
MPI	<i>Message Passing Interface</i>
NPB	<i>NAS Parallel Benchmarks</i>
OOPSLA	<i>Object-Oriented Programming, Systems, Languages and Applications</i>
PDE	<i>Partial Differential Equation</i>
PLDI	<i>Programming Language Design and Implementation</i>
RTA	<i>Runtime Analysis</i>
SO	Sistema Operacional
SPH	<i>Smoothed Particle Hydrodynamics</i>
SPEC	<i>Standard Performance Evaluation Corporation</i>
VIPS	<i>Vasari Image Processing System</i>
WCET	<i>Worst Case Execution Time</i>



# 1. INTRODUÇÃO

## 1.1 Contextualização

A pesquisa experimental, no âmbito da Ciência da Computação, não tem sido uma prática tão explorada como ocorre nas ciências naturais e sociais, mas do que isso, há ainda um questionamento sobre se a Ciência da Computação é realmente uma ciência (DENNING, 2005; DENNING, 2013; MAXION; LONGSTAFF; MCHUGH, 2010; RAMESH; GLASS; VESSEY, 2004; TICHY, 1998; WAINER et al., 2009). Para Maxion Longstaff; Mchugh (2010), o problema pode ser o fato de que a origem da Ciência da Computação está relacionada com a Matemática e com a Engenharia, o que leva ao uso da lógica, da razão e da demonstração de resultados úteis.

Ramesh; Glass; Vessey (2004) dividem o campo da computação em três subdivisões mais comuns dentro da área acadêmica, Ciência da Computação, Engenharia de *Software* e Sistemas de Informação. Essa divisão foi feita para melhor compreender os trabalhos de pesquisa realizados na computação. A Tabela 1.1 ilustra as abordagens utilizadas pelos pesquisadores da computação, segundo os mesmos autores. Como pode ser visto, 79,1% das pesquisas em Ciência da Computação e 55,3% das pesquisas em Engenharia de *Software* se concentram na formulação de conceitos, *frameworks*, modelos etc. Enquanto isso, em relação ao que é realizado em Sistemas de Informação, 66,8% das pesquisas são aplicadas em avaliações. Já quanto aos métodos de pesquisa, em Ciência da Computação, 73,4% são análises conceituais e matemáticas e, apenas 1,8 e 1,9% são métodos relacionados a experimentos envolvendo humanos e *software*, respectivamente. Assim, ao observar esses números pode-se concluir que a experimentação, no início dos anos 2000, não era o ponto de maior interesse dos trabalhos de pesquisa em Ciência da Computação.

Segundo Tichy (1998), os cientistas clássicos estão acostumados a realizar os seus trabalhos utilizando matéria e energia, o que não ocorre na computação, pois o material de trabalho fornecido nesse caso é a informação. Para o autor, o cientista da computação deve observar fenômenos e testar observações e teorias a fim de compreender a natureza do processo de informação.

Assim, a experimentação traria benefícios tais como: (a) uma base confiável de conhecimento que reduziria as incertezas sobre quais teorias, métodos e ferramentas são adequadas; (b) conhecimentos novos, úteis, e inesperados que gerariam novas áreas de investigação; (c) a rápida aceleração do progresso a fim de eliminar abordagens infrutíferas, suposições errôneas, e modismos, além de orientar a engenharia e a teoria dentro de direções promissoras.

**Tabela 1.1. Metodologia de pesquisa em computação (traduzido de Ramesh; Glass; Vessey (2004)).**

Abordagem da Pesquisa	Ciência da Computação	Engenharia de Software	Sistemas de Informação
<b>Descritiva</b>	<b>9,8%</b>	<b>27,9%</b>	<b>9,0%</b>
Descrição de sistemas	4,1%	8,1%	2,7%
Revisão da literatura	0,6%	1,6%	-
Outros descritivos	5,1%	18,2%	6,3%
<b>Avaliativa</b>	<b>11,0%</b>	<b>13,8%</b>	<b>66,8%</b>
Avaliativa-dedutiva	1,1%	4,3%	46,7%
Avaliativa-interpretativa	-	<1%	4,7%
Avaliativa-crítica	-	1,4%	-
Avaliativa-outras	9,9%	7,3%	15,4%
<b>Formulada</b>	<b>79,1%</b>	<b>55,3%</b>	<b>24,2%</b>
Conceito-formulado	17,0%	3,0%	1,0%
Framework-formulado	2,4%	4,1%	2,5%
Guidelines/Standards-formulado	0,6%	4,3%	0,8%
Modelo-formulado	5,7%	9,8%	12,5%
Processo, Método, Algoritmo-formulado	52,6%	36,0%	4,7%
Classificação/Taxonomia-formulada	0,8%	1,1%	2,7%
<b>Método de Pesquisa</b>	<b>Ciência da Computação</b>	<b>Engenharia de Software</b>	<b>Sistemas de Informação</b>
Pesquisa de ação	-	0%	0,8%
Análise conceitual	15,1%	43,5%	14,7%
Análise/Matemática conceitual	73,4%	10,6%	12,1%
Implementação de conceito (Prova de conceito)	2,9%	17,1%	1,6%
Estudo de caso	0,2%	2,2%	12,5%
Análise de dados	0,2%	2,2%	5,3%
Etnografia	-	-	0,2%
Experimento de campo	-	<1%	1,6%
Estudo de campo	0,2%	<1%	24,5%
Teoria fundamentada	-	<1%	0,2%
Hermenêuticas	-	<1%	-
Desenvolvimento de instrumentos	-	-	3,5%
Experimento laboratorial (Humanos)	1,8%	3,0%	16,2%
Revisão/Análise da literatura	0,3%	1,1%	0,8%
Experimento laboratorial ( <i>Software</i> )	1,9%	<1%	0,6%
Prova Matemática	2,4%	<1%	0,2%
Análise de protocolo	-	-	1,2%
Simulação	1,8%	1,1%	1,4%
Pesquisa descritiva/exploratória	-	1,6%	2,7%

No entanto, o mesmo autor informa que a experimentação consiste em um trabalho minucioso de coleta de dados e no uso de métodos estatísticos para verificar os resultados obtidos. Portanto, reportar os resultados de um trabalho de experimentação exige uma descrição detalhada do ambiente e dos computadores utilizados, além dos métodos adotados, permitindo que a comunidade científica possa reproduzir os experimentos e comparar os seus resultados. Contudo, observa-se em Touati; Worms; Briais (2013) que, em grupos de pesquisa relacionados à computação de alto desempenho (HPC - *High Performance Computing*), na maioria das vezes, a comparação entre os tempos de execução obtidos na otimização de programas não pode ser reproduzida, devido à falta de informações essenciais de como os experimentos foram conduzidos originalmente. Após efetuar um levantamento bibliográfico em cerca de 50 artigos, realizado nos principais

congressos da área, tais como *Object-Oriented Programming, Systems, Languages and Applications* (OOPSLA), *Programming Language Design and Implementation* (PLDI), entre outros, Georges; Buytaert; Eeckhout (2007) reportam que dezesseis artigos (um terço) não descrevem a metodologia utilizada em seus experimentos, impossibilitando terceiros de replicá-los a fim de avaliar seus resultados.

## 1.2 Relevância da Pesquisa

Em algumas áreas da Ciência da Computação os tempos de execução de um programa fornecem importantes informações sobre comportamento e desempenho de um determinado sistema (GEORGES; BUYTAERT; EECKHOUT, 2007; TOUATI; WORMS; BRIAIS, 2013). Há, por exemplo, grupos de pesquisa que estudam a eficiência de algoritmos, estratégias e técnicas para acelerar o tempo de execução de programas de computador mediante a execução de experimentos computacionais (TOUATI; WORMS; BRIAIS, 2013). Esses grupos, geralmente, avaliam experimentalmente o desempenho de dois algoritmos, em um dado sistema, comparando os tempos de execução de suas implementações. Para esses casos, torna-se imprescindível reproduzir os resultados experimentais de forma consistente, a fim de se obter confiança estatística nos resultados.

Em sistemas computacionais, nem sempre execuções sucessivas de um mesmo programa produzem os mesmos tempos de execução. De fato, o que se observa, na prática, é que, muitas vezes, ocorre uma variação significativa nos tempos de execução do programa, durante sucessivas execuções (NOGUEIRA; MATIAS JR., 2014). Em pesquisa experimental, essa variação é conhecida como erro experimental e é causada por fatores não controláveis (MONTGOMERY, 2000). Em experimentos computacionais, esses fatores tais como a influência de mecanismos de interrupção, da arquitetura da memória *cache* e diferentes interferências provocadas por rotinas do sistema operacional (SO) conhecidas como *OS Jitter* (VICENTE; MATIAS JR., 2013), entre outros, podem estar relacionados com o *hardware* ou com o *software*.

A natureza estocástica das variações, presentes nos tempos de execução de um programa de computador, faz com que sua magnitude seja de difícil predição. Como consequência, trabalhos experimentais, especialmente aqueles que dependem de correta análise dos tempos de execução de programas, falham ao tratar esse problema. Devido ao problema de variação, fica claro que se basear em uma única execução não é uma abordagem confiável, uma vez que o único tempo de execução observado pode desviar-se,显著地, dos valores mais frequentes. Além disso, Touati; Worms; Briais (2013) argumentam que os tempos mínimo e máximo, ou mesmo a média, podem levar a conclusões incorretas sobre a comparação de tempos de execução de programas. Ao se considerarem os tempos mínimo e máximo não se pode informar, com certeza, que esses valores serão sempre encontrados; a respeito da média, é fato que essa é mais sensível a dados discrepantes (*outliers*). Touati; Worms; Briais (2013) apresentam, também, um esforço inicial ao proporem um protocolo para a avaliação de significância estatística dos valores de *speedup* de programas (HENNESSY; PATTERSON, 2011). *Speedup* é a relação entre o tempo de execução de um programa executado em um determinado computador e o tempo de execução desse mesmo programa executado em um computador com algum aprimoramento sobre o computador original (HENNESSY; PATTERSON, 2011).

Em experimentos preliminares descritos em Nogueira; Matias; Vicente (2014), foi possível verificar que o protocolo sugerido em Touati; Worms; Briais (2013) não contempla a análise de dados não paramétricos, ou seja, dados que não possuem uma distribuição de probabilidade gaussiana, ou mesmo que não possuem uma distribuição conhecida *a priori*. Além disso, o protocolo em questão contempla apenas a comparação entre duas amostras de tempos de execução. Quando se realiza um projeto de experimentos mais sofisticado, múltiplos tratamentos são executados, o que exige comparações sucessivas desses tratamentos. Neste caso, seguindo o enfoque de Touati; Worms; Briais (2013), há o risco de se obter um erro na análise devido ao problema conhecido como *familywise error rate* (HOWELL, 2009), o que implica o aumento da probabilidade de ocorrerem erros do tipo I (BARBETTA; REIS; BORNIA, 2010) nos resultados dos testes estatísticos. Desse modo, fica claro que o método em questão carece de melhoramentos.

Este trabalho visa a contribuir no campo da experimentação computacional da seguinte forma:

- Realizar um estudo exploratório para analisar dados de tempos de execução de programas de computador, obtidos experimentalmente, com vistas a melhor compreensão das suas propriedades estatísticas.
- Analisar a influência de diferentes fatores operacionais sobre os tempos de execução dos programas.
- Aprimorar o protocolo proposto em Touati; Worms; Briais (2013) em dois aspectos: *i)* considerar não apenas dados paramétricos e *ii)* incluir análise apropriada para a administração do *familywise error rate*, inerente às comparações sucessivas.

### 1.3 Objetivos da Pesquisa

Geral:

- Estabelecer um protocolo para análise quantitativa comparativa de tempos de execução de programas em experimentos computacionais.

Específicos:

- Investigar as propriedades estatísticas dos tempos de execução de programas voltados para computação científica (HPC).
- Analisar e avaliar quantitativamente diferentes fontes de variabilidade em tempos de execução.
- Avaliar a aderência de diferentes testes estatísticos aplicados para a comparação de tempos de execução.

### 1.4 Desenvolvimento da Pesquisa

Nesta seção, estão descritas, de modo geral, as principais atividades que foram realizadas nessa pesquisa de mestrado.

Como esta pesquisa é de natureza teórico-experimental (WAZLAWICK, 2009), inicialmente, foi realizada uma revisão da literatura para os principais tópicos necessários para o desenvolvimento deste trabalho. As principais fontes de consulta foram livros e artigos científicos, abrangendo trabalhos na área de experimentação computacional, variabilidade dos tempos de execução de programas de computador e testes estatísticos. Em seguida, foram definidos e realizados três experimentos com o objetivo de avaliar o impacto de determinados fatores na variabilidade dos tempos de execução dos programas. Esses fatores foram obtidos da literatura e por meio de experimentos preliminares realizados como parte deste trabalho.

Os experimentos foram planejados e executados de acordo com o método DOE (*Design of Experiment*) (MONTGOMERY, 2000), que requer modificações controladas dos fatores em estudo, a fim de observar os efeitos dessas alterações sobre a variável de resposta. Nesta pesquisa, definiu-se o tempo de execução dos programas como sendo a variável de resposta. Em relação ao fator, este, pode assumir diferentes níveis de operação e, ao combinar fator e nível tem-se um tratamento (MONTGOMERY, 2000; BARBETTA; REIS; BORNIA, 2010). Assim, testes estatísticos foram utilizados para caracterizar os tempos de execução de programas HPC. Para o estudo experimental, foram utilizados os *benchmarks* PARSEC versão 3.1 (BIENIA, 2011), SPLASH2x (BARROW-WILLIAMS; FENSCH; MOORE, 2009; WOO *et al.*, 1995). O SPLASH2x é uma variação do SPLASH2, e foi desenvolvido pelo grupo do PARSEC, para permitir o uso de *inputs* e o NPB (*NAS Parallel Benchmarks*) versão 3.3.1 (NASA ADVANCED SUPERCOMPUTING DIVISON, 2015). Após a caracterização dos tempos de execução, utilizou-se o protocolo *Speedup-Test*, proposto em Touati; Worms; Briais (2013) para comparar, aos pares, os tratamentos adotados. Desse modo, foi possível avaliar se houve diferença estatística entre os tratamentos. Ao final, foi realizado um estudo de caso para comparar o protocolo *RTA*, proposto neste trabalho, com o protocolo *Speedup-Test*.

## 1.5 Estrutura do Documento

Esse documento está estruturado em seis capítulos. O primeiro foi o de introdução; o segundo apresenta uma descrição dos testes estatísticos utilizados no estudo, além dos trabalhos relacionados com a presente pesquisa, com ênfase para estudos sobre a variabilidade de tempos de execução de programas. Já o Capítulo terceiro apresenta um estudo exploratório experimental do problema da variabilidade dos tempos de execução em programas de computador, a fim de caracterizar os tempos de execução em programas típicos da área de HPC. O Capítulo quarto descreve o protocolo *Speedup-Test*, como originalmente proposto por Touati; Worms; Briais (2013), discutindo os cenários não contemplados pelo protocolo. O Capítulo quinto, por sua vez, traz uma proposta de melhoria para o protocolo *Speedup-Test*, bem como um estudo de caso comparando o protocolo original com a proposta deste trabalho. Para finalizar, o Capítulo sexto apresenta as considerações finais do trabalho, destacando os principais resultados alcançados, as limitações da pesquisa, as dificuldades encontradas e os futuros trabalhos que podem ser desenvolvidos a partir desta pesquisa.

# 2. REVISÃO DA LITERATURA

## 2.1 Introdução

Em sistemas computacionais, fatores como o tamanho de uma variável de ambiente, a quantidade de processos administrativos executando em *background*, o número de *threads* de um programa, entre outros, podem ser responsáveis por variações nos tempos de execução dos programas. Por isso, avaliar o desempenho de um programa apenas observando uma única execução e determinando o melhor ou pior caso não é uma abordagem confiável.

Além disso, segundo Touati; Worms; Briais (2013), é comum utilizar a média na análise estatística, embora esta seja suscetível a *outliers*, ao passo que a mediana também é uma boa opção para a análise e não apresenta o mesmo problema. Essa seção descreve os métodos estatísticos que utilizam ambas em suas análises, o teste *t* de Student (ver Seção 2.2.1) (BARBETTA; REIS; BORNIA, 2010) e o teste de *Wilcoxon-Mann-Whitney* (ver Seção 2.2.2) (HOWELL, 2009), que são utilizados para a análise de significância dos tempos de execução no protocolo proposto por esta pesquisa. Contudo, há pressupostos que devem ser satisfeitos ao se aplicarem esses dois testes e, por isso, também são descritos outros testes estatísticos (ver Seções 2.2.3 e 2.2.4) que foram usados na verificação desses pressupostos.

O trabalho de revisão da literatura também permitiu identificar fatores que interferem nos tempos de execução dos programas e, portanto, podem influenciar os resultados de experimentos computacionais, levando o experimentador a possíveis conclusões incorretas. Assim sendo, este capítulo também aborda a variabilidade dos tempos de execução de programas.

## 2.2 Métodos Estatísticos

Ao aplicar um teste estatístico, o objetivo é testar se uma dada hipótese é verdadeira ou não. Segundo Barbetta; Reis; Bornia (2010), ao trabalhar com um problema de teste de hipóteses, deve-se formular a hipótese nula e a hipótese alternativa. A hipótese nula ( $H_0$ ) é aceita como verdadeira até que, estatisticamente, se prove o contrário. Essa hipótese é formulada em termos de igualdade entre parâmetros, ou entre um parâmetro e uma constante. No caso da hipótese alternativa ( $H_1$ ), é formulada em termos de desigualdade e é aceita quando os dados mostram evidência suficiente de que  $H_0$  é falsa (BARBETTA; REIS; BORNIA, 2010).

Assim, os testes estatísticos são utilizados para avaliar se os dados apresentam evidências suficientes para rejeitar  $H_0$  e aceitar a  $H_1$  com uma margem de segurança (BARBETTA; REIS; BORNIA, 2010). Contudo, é possível cometer um erro ao aceitar, ou rejeitar  $H_0$  (HOWELL, 2009) e, por isso, ao planejar o experimento, é comum definir qual o nível de significância do teste ( $\alpha$ ). Desse modo, se o teste rejeitar  $H_0$  o risco de erro está limitado pelo  $\alpha$  adotado. No entanto, se o teste aceitar  $H_0$  a probabilidade de erro é definida por  $\beta$  (BARBETTA; REIS; BORNIA, 2010). Os erros, possíveis, encontrados na análise estatística estão descritos na Tabela 2.1.

**Tabela 2.1. Tipos de erros (adaptado de Barbetta; Reis; Bornia (2010))**

	Aceita $H_0$	Rejeita $H_0$
$H_0$ Verdadeira	Correto, com prob. = $1 - \alpha$	Erro tipo I com prob. = $\alpha$
$H_0$ Falsa	Erro tipo II com prob. = $\beta$	Correto, com prob. = $1 - \beta$

## 2.2.1 Teste $t$

O teste  $t$  é um teste estatístico que compara duas amostras por meio de seus valores médios. Esse teste pode ser aplicado para avaliar comparações de tratamentos com amostras pareadas e com amostras independentes (BARBETTA; REIS; BORNIA, 2010; FIELD, 2009; HAIR *et al.*, 2009). Segundo Barbetta; Reis; Bornia (2010), tratamento é a combinação entre o fator, que está sendo estudado, e seus níveis de operação. Esses autores também ilustram que amostras pareadas possuem uma relação de dependência entre si, no que diz respeito à observação de seus dados, ao contrário das amostras independentes.

No caso das amostras pareadas, a análise das  $n$  observações pode ser feita pela diferença entre os tratamentos  $T_1$  e  $T_2$  por meio da variável  $D$  (ver Tabela 2.2).

**Tabela 2.2. Observações pareadas (adaptado de Barbetta; Reis; Bornia (2010)).**

Observação	T1	T2	$D = T_2 - T_1$
1	$x_1$	$y_1$	$D_1$
2	$x_2$	$y_2$	$D_2$
3	$x_3$	$y_3$	$D_3$
:	:	:	:
$n$	$x_n$	$y_n$	$D_n$

Barbetta; Reis; Bornia (2010) relatam que, o fato dos dados serem pareados permite a observação, em cada ensaio, de quão melhor um tratamento foi sobre o outro (ver Equação 2.1).

$$D = T_2 - T_1 \quad (2.1)$$

Tendo a diferença entre as amostras como base do teste de hipótese,  $H_0$  e  $H_1$  podem ser descritas de acordo com Barbetta; Reis; Bornia (2010):

$$H_0: \mu_D = 0 \text{ e } H_1: \mu_D > 0;$$

sendo  $\mu_D$  o valor esperado de  $D$ . Assim, a estatística do teste é dada pela Equação 2.2:

$$t = \frac{\bar{d} \times \sqrt{n}}{s_d}, \quad (2.2)$$

em que  $\bar{d}$  é a média das diferenças observadas (ver Equação 2.3),  $n$  é o tamanho da amostra, ou seja, o número de pares e  $s_d$  é o desvio padrão das diferenças observadas (Equação 2.4).

$$\bar{d} = \frac{\sum_{i=1}^n D_i}{n} \quad (2.3)$$

$$s_d = \sqrt{\frac{1}{n-1} \times \left( \sum_i D_i^2 - n \times \bar{d}^2 \right)} \quad (2.4)$$

Se for considerado que a amostra é pequena e seus valores provenientes de uma distribuição Gaussiana, a comparação da média dos tratamentos pode ser realizada utilizando uma distribuição  $t$  de *Student* com  $n - 1$  graus de liberdade ( $gl$ ) (BARBETTA; REIS; BORNIA, 2010). Os graus de liberdade, nesse caso, indicam a aproximação de uma distribuição  $t$  de *Student* de uma distribuição Gaussiana, ou seja, quanto maiores os graus de liberdade, mais próxima uma distribuição  $t$  de *Student* estará de uma distribuição Gaussiana (NAVIDI, 2012). Nesse caso, o resultado do teste deve ser comparado a um valor tabelado da distribuição  $t$  de *Student* (ver Anexo A), tendo como base os graus de liberdade obtidos. Para o caso de amostras suficientemente grandes, a distribuição  $t$  de *Student* se aproxima de uma distribuição Gaussiana (BARBETTA; REIS; BORNIA, 2010).

Todavia, há cenários em que as amostras são independentes, ou seja, em que uma amostra não depende da outra. Nesse caso, essas amostras precisam satisfazer três suposições básicas (BARBETTA; REIS; BORNIA, 2010): (1) As observações devem ser independentes; (2) As variâncias populacionais devem ser iguais nos dois grupos; (3) As duas amostras ser provenientes de uma distribuição Gaussiana.

A primeira suposição é atendida no momento em que se planeja o experimento e se define de que maneira as amostras serão obtidas. Para as suposições #2 e #3, recomenda-se o uso de testes estatísticos para confirmá-las. No caso da suposição #2, podem ocorrer três situações: (I) as variâncias são conhecidas; (II) as variâncias são desconhecidas, mas supostamente são iguais; e (III) as variâncias são desconhecidas, mas supostamente são diferentes. Assim, há diferentes cálculos estatísticos para o teste  $t$  (SEWARD; DOANE, 2014). Para o primeiro caso (I), em que a variância populacional  $\sigma$  é conhecida, a estatística do teste pode ser realizada pela Equação 2.5 utilizando a distribuição Gaussiana.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (2.5)$$

Assim,  $\bar{x}_1$  é a média da amostra 1,  $\bar{x}_2$  é a média da amostra 2,  $\sigma_1^2$  é a variância da população 1,  $\sigma_2^2$  é a variância da população 2,  $n_1$  é o tamanho da amostra 1 e  $n_2$  é o tamanho da amostra 2. O resultado do teste deve então ser comparado a um valor crítico de  $z$  (ver Anexo B).

Ainda sobre as variâncias, quando não se conhece as variâncias populacionais (o caso II citado anteriormente), pode-se utilizar uma média ponderada das variâncias de cada amostra,  $s_1^2$  e  $s_2^2$  respectivamente, e os pesos atribuídos às variâncias amostrais são baseados em seus respectivos graus de liberdade ( $n_1 - 1$ ) e ( $n_2 - 1$ ), como descrito na Equação 2.6 (SEWARD; DOANE, 2014).

$$s_p^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2} \quad (2.6)$$

Sendo  $s_p^2$  a variância combinada das duas amostras, pode-se, então, calcular a estatística do teste  $t$  para as médias das duas amostras,  $\bar{x}_1$  e  $\bar{x}_2$  respectivamente, com a Equação 2.7. Nesse caso, os graus de liberdade das duas amostras são combinados e tem-se a Equação 2.8.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad (2.7)$$

$$gl = n_1 + n_2 - 2. \quad (2.8)$$

Ainda, como descrito anteriormente, há casos em que não se conhece a variância populacional e também não se pode supor que as variâncias sejam iguais (caso III). Para essa situação, o correto é utilizar o teste  $t$  com a correção de *Welch* para os graus de liberdade (SEWARD; DOANE, 2014; HOWELL, 2009). A Equação 2.9 apresenta a correção de *Welch* para os graus de liberdade e a Equação 2.10 apresenta a estatística do teste  $t$  para esse caso.

$$gl' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \quad (2.9)$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2.10)$$

Diferente da Equação 2.7, a Equação 2.10 não utiliza uma variância combinada das amostras. Para os itens II e III, deve-se utilizar a distribuição  $t$  de *Student*. Assim, o resultado deve ser comparado seguindo essa distribuição com base nos graus de liberdade obtidos (ver Anexo A).

### 2.2.2 Teste de Wilcoxon-Mann-Whitney

O teste de *Wilcoxon-Mann-Withney*, segundo GIBBONS; CHAKRABORTI (2014) e KRISHNAMOORTHY (2006), é útil quando se deseja comparar a mediana de uma distribuição contínua simétrica. Esse teste pode ser considerado uma alternativa ao teste  $t$  para duas amostras quando as variâncias das amostras não são homogêneas e quando não se pode afirmar que os dados das amostras seguem uma distribuição Gaussiana. Entretanto, ele só é válido quando se pode supor que: (I) cada amostra foi aleatoriamente selecionada a partir da população que representa; (II) as amostras são independentes umas das outras; (III) a variável subsequentemente ordenada é uma variável aleatória contínua; e (IV) as amostras são provenientes de uma mesma distribuição. Sendo  $\eta_1$  a mediana da população 1 e  $\eta_2$  a mediana da população 2 as hipóteses para o teste podem ser dadas por:

$$H_0: \eta_1 = \eta_2 \text{ e } H_1: \eta_1 \neq \eta_2.$$

O teste de *Wilcoxon-Mann-Whitney* consiste em ordenar (*ranks*) as  $n$  observações e computar a soma de cada grupo. Uma vez que há uma relação entre os dois grupos (BARBETTA; REIS; BORNIA, 2010), a estatística do teste pode ser obtida apenas pela soma dos *ranks* da amostra 1 ( $M_1$ ) (ver Equação 2.11):

$$U = M_1 - \left[ \frac{n_1 \times (n_1 + 1)}{2} \right] \quad (2.11)$$

Uma tabela com valores críticos para a estatística do teste ( $U$ ) é fornecida quando a análise é realizada com amostras menores ou iguais a 20 (ver Anexo C). No caso de amostras grandes, maiores do que 20, a distribuição de  $U$  aproxima-se de uma Gaussiana e, portanto, sua estatística pode ser dada pela Equação 2.12. Assim, pode-se usar a distribuição Gaussiana padrão para obter o valor crítico de  $z$  (BARBETTA; REIS; BORNIA, 2010; GIBBONS; CHAKRABORTI, 2014) (ver Anexo B). A constante 12, presente na equação é obtida por meio de um esforço matemático, como descrito em (EMERSON, 2011).

$$z = \frac{U - \frac{n_1 \times n_2}{2}}{\sqrt{\frac{n_1 \times n_2 \times (n_1 + n_2 + 1)}{12}}} \quad (2.12)$$

## 2.2.3 Testes de Aderência

Os testes de aderência são aplicados para verificar se os dados de uma amostra seguem uma determinada distribuição teórica (GIBBONS; CHAKRABORTI, 2014; BARBETTA; REIS; BORNIA, 2010). Há vários testes de aderência, como o *Anderson Darling*, *Kolmogorov-Smirnov*, *Lilliefors*, *Qui-Quadrado*, *Shapiro-Wilk*, entre outros. Nesse tópico são apresentados dois testes de aderência que são utilizados durante a análise estatística desta pesquisa — o teste de *Shapiro Wilk*, que é específico para avaliar se os dados de uma amostra seguem uma distribuição Gaussiana e o teste de *Kolmogorov-Smirnov*, que pode ser utilizado para avaliar a aderência de uma amostra a uma determinada distribuição ou mesmo avaliar se duas amostras provêm de uma mesma distribuição. O teste de *Shapiro-Wilk* foi escolhido por ser um teste robusto para amostras pequenas e grandes ( RAZALI; WAH; SCIENCES, 2011; SHAPIRO; WILK, 1965) e o teste de *Kolmogorov-Smirnov* foi escolhido por seus valores críticos não dependerem de uma distribuição específica a ser testada tornando-o assim, livre de distribuição (BARBETTA; REIS; BORNIA, 2010; GIBBONS; CHAKRABORTI, 2014; NIST/SEMATECH, [s.d.]).

### 2.2.3.1 Teste de Shapiro Wilk

Como descrito na Seção 2.2.1, o teste  $t$  tem como uma de suas premissas que os dados sejam provenientes de uma distribuição Gaussiana. Por isso, antes de aplicar o teste  $t$ , é necessário que seja confirmada a aderência dos dados da amostra a essa distribuição. Um teste que pode ser usado para esse fim é o teste de *Shapiro Wilk* ( RAZALI; WAH; SCIENCES, 2011; ROYSTON, 1995; SHAPIRO; WILK, 1965). O teste de *Shapiro Wilk* baseia-se na estimativa dos mínimos quadrados ponderados

normalmente distribuídos e na variância amostral de outra população (RAZALI; WAH; SCIENCES, 2011; ROYSTON, 1995).

Sendo,  $x_1, x_2, \dots, x_n$  uma amostra aleatória independentemente identicamente distribuída e assumindo que  $x_{(1)} < x_{(2)} < \dots < x_{(n)}$  são seus valores ordenados (SEN; SRIVASTAVA, 1997), a estatística do teste é obtida pela Equação 2.13:

$$W = \frac{(\sum_{i=1}^n a_i \times x_i)^2}{(\sum_{i=1}^n x_i - \bar{x})^2} \quad (2.13)$$

em que,  $x_i$  é o  $i$ -ésimo valor ordenado,  $\bar{x}$  a média amostral,  $a$  é uma constante dada pela Equação 2.14:

$$(a_1, \dots, a_n) = \frac{m^T \times V^{-1}}{(m^T \times V^{-1} \times V^{-1} \times m)^{1/2}} \quad (2.14)$$

Onde se tem  $m = (m_1, \dots, m_n)^T$  e  $m_1, \dots, m_n$  sendo os valores esperados dos valores ordenados de uma distribuição Gaussiana padrão e  $V$  a matriz de covariância desses valores ordenados (RAZALI; WAH; SCIENCES, 2011). A hipótese nula será rejeitada dependendo do valor de  $W$ . Para valores menores que o nível de significância  $\alpha$ , rejeita-se  $H_0$  e aceita-se  $H_1$ . O teste de *Shapiro Wilk* é robusto para amostras com tamanho pequeno, no entanto, com o algoritmo *AS R94* (ROYSTON, 1995) as amostras podem variar de três até 5.000 valores.

#### 2.2.3.2 Teste de Kolmogorov-Smirnov

O teste de *Kolmogorov-Smirnov* pode ser utilizado para testar a aderência dos dados de uma amostra a uma determinada distribuição de dados e analisar se os dados de duas amostras são provenientes de uma mesma distribuição. Diferente do teste de *Shapiro Wilk* que é aplicado apenas à distribuição Gaussiana, o teste de *Kolmogorov-Smirnov* é livre de distribuição e pode ser aplicado a qualquer população de distribuição contínua (GIBBONS; CHAKRABORTI, 2014). Para uma única amostra as hipóteses são dadas por Barbetta; Reis; Bornia (2010):

$H_0$ : os dados provêm da distribuição especificada (há aderência); e

$H_1$ : os dados não provêm da distribuição especificada (não há aderência).

Quando utilizado para avaliar uma única amostra, o teste de *Kolmogorov-Smirnov* utiliza como base as distribuições de frequências acumuladas empírica  $S(y)$  e teórica  $F(y)$  ( BARBETTA; REIS; BORNIA, 2010; GIBBONS; CHAKRABORTI, 2014). Deve-se, então, ordenar os elementos da amostra e calcular o valor de  $S(y)$  e  $F(y)$  para cada um deles;  $S(y)$  pode ser obtido pela Equação 2.15:

$$S(y) = \frac{\text{número de valores } \leq y_i}{n} \quad (2.15)$$

em que,  $n$  é o tamanho da amostra e  $y_i$  é um valor qualquer da amostra. Para  $F(y)$ , o cálculo deve ser baseado na distribuição especificada em  $H_0$ . Após isso, a estatística do teste pode ser dada pela diferença absoluta máxima  $D$  ( BARBETTA; REIS; BORNIA, 2010; GIBBONS; CHAKRABORTI, 2014), Equação 2.16.

$$D = \max_i \{|F(y_i) - S(y_i)|, |F(y_i) - S(y_{i-1})|\} \quad (2.16)$$

Com a diferença absoluta máxima obtida,  $D$ , compara-se o valor tabelado  $D_\alpha$  (ver Anexo D) ao nível de significância  $\alpha$  com o tamanho  $n$  da amostra (BARBETTA; REIS; BORNIA, 2010). Assim, se:

$D < D_\alpha$  aceita  $H_0$ ; e, portanto, declara que há aderência a distribuição especificada com o nível de significância  $\alpha$ ; e

$D \geq D_\alpha$  rejeita  $H_0$ ; e, portanto, declara que não há aderência a distribuição especificada com o nível de significância  $\alpha$ .

No caso em que o teste de *Kolmogorov-Smirnov* é utilizado para avaliar se os dados de duas amostras são provenientes de uma mesma distribuição, a comparação é feita utilizando as funções de distribuição empírica das duas amostras (GIBBONS; CHAKRABORTI, 2014). Com  $x_1, x_2, \dots, x_m$  e  $w_1, w_2, \dots, w_m$  sendo os valores ordenados das amostras 1 e 2, respectivamente, de uma população contínua; então, suas respectivas funções de distribuição empírica,  $S_m(y)$  e  $S_n(y)$ , são:

$$S_m(y) = \begin{cases} 0, & \text{se } y < x_{(1)} \\ k/m, & \text{se } x_{(k)} \leq y < x_{(k+1)} \\ 1, & \text{se } y \geq x_{(m)} \end{cases} \quad \text{para } k = 1, 2, \dots, m-1.$$

e

$$S_n(y) = \begin{cases} 0, & \text{se } y < w_{(1)} \\ k/n, & \text{se } w_{(k)} \leq y < w_{(k+1)} \\ 1, & \text{se } y \geq w_{(n)} \end{cases} \quad \text{para } k = 1, 2, \dots, n-1.$$

Para o teste com duas amostras as hipóteses, nula e alternativa, são dadas por:

$$H_0: F_w(y) = F_x(y) \text{ e } H_1: F_w(y) \neq F_x(y).$$

O teste de *Kolmogorov-Smirnov* para duas amostras é baseado na máxima diferença absoluta entre as duas distribuições empíricas  $D_{m,n}$  e a estatística do teste é dada pela Equação 2.17 (GIBBONS; CHAKRABORTI, 2014). O resultado do teste deve, então, ser comparado com o valor tabelado  $D_\alpha$  (ver Anexo E).

$$D_{m,n} = \max_y |S_m(y) - S_n(y)| \quad (2.17)$$

## 2.2.4 Análise de Variância com o Teste *F* de *Fisher*

Para o teste  $t$ , a variância é uma das premissas que devem ser atendidas. O teste *F* de *Fisher* permite analisar se as variâncias de duas populações, supostamente com distribuição Gaussiana, são iguais, tendo a distribuição *F* como referência. Assim, as hipóteses podem ser formuladas (BARBETTA; REIS; BORNIA, 2010).

$$H_0: \sigma_1^2 = \sigma_2^2; \text{ e } H_1: \sigma_1^2 \neq \sigma_2^2.$$

Com as amostras das populações 1 e 2, a estatística do teste é dada pela Equação 2.18 e deve-se colocar a maior variância como numerador (BARBETTA; REIS; BORNIA, 2010).

$$f = \frac{s_1^2}{s_2^2} \quad (2.18)$$

O resultado do teste deve ser comparado ao valor tabelado da distribuição  $F$  (ver Anexo F), com base nos graus de liberdade do numerador  $gl = n_1 - 1$  e do denominador  $gl = n_2 - 1$ . A hipótese nula deve ser rejeitada caso o valor encontrado seja maior ou igual ao valor tabelado (BARBETTA; REIS; BORNIA, 2010).

## 2.2.5 Múltiplas Comparações

Em um planejamento experimental em que haja um grupo de tratamentos o intuito é compará-los par a par, de modo a identificar se há uma diferença significativa entre eles. Os resultados dessas comparações formam um conjunto conhecido como família (HOWELL, 2009). Quando se compararam apenas dois tratamentos, o resultado obtido é em função de um  $\alpha$  (geralmente com valor 0,05) que representa a probabilidade de se obter um erro do tipo I, para essa comparação. No entanto, quando se obtém uma família de tratamentos em que o número de tratamentos é maior do que dois, o correto é avaliar a probabilidade de se rejeitar, de forma incorreta, pelo menos uma das hipóteses nulas que compõe essa família (*familywise error rate*) (HOWELL, 2009) (ver Equação 2.19).

$$\alpha' = 1 - (1 - \alpha)^c \quad (2.19)$$

em que,  $c$  é o número de comparações realizadas dentro da família,  $\alpha$  é o nível de significância admitido e  $\alpha'$  é a probabilidade de se rejeitar, de forma incorreta, pelo menos uma das hipóteses nulas.

Há na literatura alguns testes que são específicos para trabalhar com múltiplas comparações. Entre eles, pode ser citado o teste de *Holm* (HOWELL, 2009; KEAREY; KLEPEIS; VINE, 2014). O teste de *Holm* pode ser considerado um refinamento da abordagem proposta por *Bonferroni* (KEAREY; KLEPEIS; VINE, 2014), ao apresentar uma posição menos conservadora; esse teste propõe-se a aceitar ou rejeitar um conjunto de hipóteses nulas ordenadas, iniciando pelo menor *p-value* (KEAREY; KLEPEIS; VINE, 2014). Assim, se há  $k$  comparações dois a dois, devem-se ordenar os  $k$  *p-values* em ordem crescente,  $(p_k, p_{k+1}, p_{k+2}, \dots)$ , sendo,  $p_k$  o menor valor *p* não ajustado,  $j$  é o  $j$ -ésimo teste de hipótese realizado ( $j \leq k$ ), e  $\alpha_\tau$  o nível de significância real. O teste de *Holm* pode ser dado pela Equação 2.20.

$$\alpha_j = \alpha_\tau / (k - j + 1) \quad (2.20)$$

Assim, se o *p-value* não ajustado for menor que  $\alpha_j$ , rejeita-se a hipótese nula e então compara-se o próximo *p-value* e, assim, sucessivamente. Quando for encontrado um *p-value* não ajustado maior que  $\alpha_j$  o teste deve ser interrompido, pois entende-se que as comparações subsequentes não são significativas (KEAREY; KLEPEIS; VINE, 2014).

## 2.3 Trabalhos Relacionados

Ao estudar a variabilidade dos tempos de execução de programas de computador, observou-se a sua presença em diversos cenários.

Em Georges; Buytaert; Eeckhout (2007), os autores concentram seus esforços na defesa do uso de uma metodologia rigorosa para avaliação de desempenho de programas *Java*. Eles alegam que as metodologias predominantes encontradas na literatura podem ser enganosas e levar a conclusões incorretas. Em relação ao desempenho dos programas *Java*, os autores realizaram experimentos levando em consideração aspectos como: o *start-up* do programa, com o objetivo de mensurar o quanto rapidamente uma máquina virtual pode executar um programa relativamente pequeno, e programas em *steady-state*, com o objetivo de mensurar o desempenho de programas que já se encontram executando de forma estável na máquina virtual. Para isso eles utilizaram a máquina virtual *Jikes Research Virtual Machine* (ALPERN *et al.*, 2005) com cinco estratégias de *garbage collection* (*CopyMS*, *GenCopy*, *GenMS*, *MarkSweep* e *SemiSpace*), todas fornecidas pelo *Jikes*. Como *benchmarks*, os autores utilizaram o *SPECjvm98* (SPEC, 2013a) e o *DaCapo* (BLACKBURN *et al.*, 2006). Em todos os experimentos considerou-se uma faixa de tamanho da *heap* por *benchmark*. O tamanho da *heap* foi alterado até seis vezes o mínimo, em intervalos de 0,25.

Os autores realizaram dois experimentos de variabilidade em tempos de execução. No primeiro, foram realizadas 30 invocações das máquinas virtuais e uma única interação de *benchmark* com o intuito de avaliar o desempenho de uma máquina virtual ao executar um programa relativamente pequeno. No segundo experimento, os autores, realizaram dez invocações das máquinas virtuais e 30 interações de *benchmark*, com o objetivo de avaliar o desempenho de programas, quando esses já se encontram executando de forma estável na máquina virtual. Para avaliar as metodologias predominantes, Georges *et al.* compararam (aos pares) o desempenho global dos *garbage collectors* e computaram os intervalos de confiança aplicando os métodos estatísticos *ANOVA* (LILJA, 2005) e *Tukey HSD* (SURHONE; TIMPLEDON; MARSEKEN, 2010).

Observou-se que a variabilidade dos tempos de execução foi significante tanto para o desempenho de *start-up*, quanto para o desempenho de programas em execução estável. O coeficiente de variação (definido como o desvio padrão dividido pela média) foi em torno de 2%, sendo maior para vários *benchmarks*, e que a diferença entre o máximo e o mínimo desempenho variou de acordo com os *benchmarks* e foi geralmente em torno de 8% para o desempenho de *start-up* e 20% para o desempenho de programas em execução estável. No entanto, ao analisar o desempenho de *start-up* com os métodos mais usados na literatura, os resultados observados podem ser enganosos em uma fração substancial das comparações (até 16%), e, no caso do desempenho de programas em execução estável, as observações feitas foram similares às do desempenho *start-up*. Georges *et al.* concluíram que a metodologia predominante na literatura está incorreta em mais de 20% dos casos. Outro fato apontado pelos autores é que algumas metodologias predominantes se mostram melhores que outras, como o caso da média e da mediana dos tempos, que são exemplos de boas metodologias.

Para compreender o WCET (*Worst Case Execution Time*), (HANSEN; HISSAM; MORENO, 2009) propõem um algoritmo que mensura tanto uma estimativa do

WCET, quanto uma predição da probabilidade que um futuro tempo de execução excederá essa estimativa. A abordagem estatística apresentada pelos autores utiliza a teoria dos valores extremos para construir um modelo do comportamento da “cauda” dos tempos de execução medidos.

Para validar sua estimativa de WCET, os autores utilizaram *traces* dos tempos de execução de um conjunto de 154 tarefas periódicas incorporando mais de 200 milhões de amostras combinadas de tempo de execução. Os autores coletaram aproximadamente 125 minutos de *traces* para cada tarefa em 25 execuções de cinco minutos. Os primeiros quinze minutos são utilizados para a estimação WCET e a segunda parte é utilizada para validação, ou seja, usada para avaliar a estimação WCET e consistem dos últimos 110 minutos de cada *trace*. A fim de assegurar um número suficiente de amostras para estimação e validação, foram excluídos os *traces* com menos de 75.000 amostras. Os tempos de execução foram coletados de um sistema embarcado com mais de um milhão de linhas de código. A tarefa a ser acompanhada foi marcada utilizando o *WindView* (WIND RIVER SYSTEMS, INC., 2002) e as amostras foram coletadas quando as tarefas já estavam em *steady-state*.

Segundo os autores, ao aplicar o algoritmo (referente à sua abordagem) nos quinze minutos iniciais (estimação de 122 *traces*), 75 *traces* passaram no teste de *Qui-Quadrado* (HOWELL, 2009) e alcançaram uma estimação WCET. Ao aumentar o tempo de quinze para 30 minutos, o número de *traces* passou para 95. O algoritmo foi aplicado em uma única tarefa e em um conjunto de tarefas. Os resultados apontaram que as probabilidades excedidas, medida e predita se combinam bem. Para um conjunto de tarefas, os autores compararam a estimação WCET realizada pelo seu algoritmo de estimação com um método de “Máximo Observado”. Os resultados obtidos apontaram que, no caso do “Máximo Observado”, a fração dos WCET excedentes é impraticável e tem maior variância, diferentemente da variabilidade da abordagem apresentada pelos autores que foi显著mente menor e cujos valores foram agrupados em torno do valor predito.

O trabalho realizado em Mytkowicz *et al.* (2009) demonstra, por meio de experimentação, que medidas tendenciosas são significantes e comuns e, por isso, é necessário ser atencioso ao planejar um experimento computacional, uma vez que o próprio planejamento pode levar a medições tendenciosas. Na otimização de programas, os autores destacam duas fontes de medidas tendenciosas: (I) o tamanho de variáveis do ambiente operacional e (II) a ordem em que os arquivos objeto são ligados pelo *linker*. Alguns cuidados foram tomados para garantir a integridade dos resultados e os experimentos foram conduzidos em um ambiente mínimo, exceto naqueles em que variáveis de ambiente foram adicionadas.

Além disso, nota-se que alterações no tamanho da variável de ambiente afetam o alinhamento da pilha do programa e, por isso, o alinhamento das estruturas alocadas na *heap* influencia o tempo de execução. A ordem de entrega dos arquivos objeto ao *linker* pode afetar diferentes programas de diferentes maneiras e, em determinados casos, pode afetar o alinhamento do código na fila de instruções do processador (*instruction queue*) e, em outros, a *cache* de instruções (*instruction cache*), o que modifica a forma pela qual ocorrem as faltas de *cache* (*cache misses*), que, em última análise, têm influência significativa nos tempos de execução. No entanto, segundo os autores, é possível evitar e detectar a ocorrência de medidas tendenciosas, ao se utilizarem boas práticas no planejamento dos experimentos.

Mazouz; Touati; Barthou (2010) demonstram que a variabilidade dos tempos de execução de programas *OpenMP* é mais significativa em relação a programas não paralelos. Os autores também relatam que os dados obtidos durante o experimento com programas *OpenMP* não seguem uma distribuição Gaussiana e que, por isso, ela não é indicada como uma função geral de distribuição para tempos de execução. Para os experimentos, os serviços do sistema operacional, com exceção dos básicos (ex.: *sshd*) e a alocação dinâmica aleatória de pilha foram desativados. Os programas disponibilizados pelos *benchmarks* *SPEC CPU2006* (SPEC, 2013b) e *SPEC OMP2001* (SPEC, 2013c) foram compilados com opções de otimização e os seus tempos de execução foram medidos utilizando a função *gettimeofday*. Para o *benchmark* *SPEC CPU2006*, os autores variaram o tamanho do ambiente *shell UNIX*, experimentando dois níveis de otimização de código (-O3 e -O2). Para o *SPEC OMP2001* o ambiente *shell UNIX* foi fixo, mas variaram os números de *threads*, como: sem *thread*, uma, duas, quatro, seis e oito *threads*.

O *SPEC CPU2006* foi configurado para executar 30 vezes, sendo cada execução sequencial em um modo *back-to-back*, em que não mais que um programa foi executado por vez e o *input* empregado para os experimentos foi o *train set*. O teste de aderência para a distribuição Gaussiana foi realizado em uma amostra com 30 execuções com o teste de *Shapiro-Wilk* sob três níveis de confiança (95%, 90% e 50%), tanto nos programas *SPEC CPU2006*, quanto nos programas *SPEC OMP2001*. Para os testes de variabilidade dos tempos de execução dos programas *SPEC CPU2006* e *SPEC OMP2001* (paralelo com uma, duas, quatro e oito *threads*), os autores utilizaram o teste *t* de *Student* para reportar o intervalo de confiança da média. Além disso, os programas foram compilados utilizando dois compiladores, *gcc* e *icc*, com dois níveis de otimização. As variáveis de ambiente também foram alteradas e desabilitaram o endereço inicial aleatório da pilha.

A análise de aderência dos dados demonstrou, com um nível de confiança de 95%, que a maioria dos programas *SPEC CPU2006* apresentou amostras com distribuição Gaussiana, ao contrário dos programas do *SPEC OMP2001* que não apresentaram essa tendência. Os resultados dos testes apontaram que foi possível encontrar distribuição Gaussiana nos tempos de execução de três programas do *SPEC CPU2006* e do *SPEC OMP2001* quando o nível de confiança é reduzido para 50%. Assim, a distribuição Gaussiana foi considerada não indicada como uma função geral de distribuição para tempos de execução de programas. Os dados gerados pela análise da variabilidade dos tempos de execução dos programas *SPEC CPU2006* permitiram confirmar que o tamanho do ambiente *shell UNIX* pode influenciar e que, quando o ambiente de execução é fixo, as variações são menores.

Já Mazouz; Touati; Barthou (2011a) investigam as causas da variabilidade dos tempos de execução de programas *OpenMP* em arquiteturas *multicore*. Os experimentos foram realizados em um computador Intel dedicado. Foram desativados: o *dynamic voltage scaling*, variáveis de ambiente desnecessárias (o restante teve seu tamanho total constante), todos os serviços desnecessários (exceto o *sshd*) do sistema operacional e o endereço inicial aleatório da pilha. Os autores otimizaram e compilaram os programas do *benchmark* *SPEC OMP2001* 31 vezes. As execuções foram sequenciais e consecutivas, com os programas executando individualmente, a não ser ao estudar o efeito de processos executando concomitantemente. Para analisar o impacto do *thread affinity* na variabilidade do desempenho dos programas do *SPEC OMP2001*, cada programa foi executado

respectivamente com duas, quatro e seis *threads* sob três tipos de configurações: sem *scheduling affinity* (LOOSEMORE *et al.*, 2013), sob o *icc compiler compact affinity strategy* e sob *icc compiler scatter strategy*.

No caso da variabilidade de desempenho dos programas *SPEC OMP2001* com processos executando concomitantemente, o *benchmark* foi iniciado com oito *threads* para ocupar todos os núcleos do sistema, em que o *benchmark* foi executado ou como uma única aplicação no computador, ou em paralelo com oito, 16, 24 ou 32 processos executando concomitantemente, respectivamente. Os resultados foram reportados quando o *SPEC OMP2001* e os processos executando concomitantemente foram iniciados sem *scheduling affinity* para o sistema de núcleos. O número de *OpenMP threads* e os processos executando concomitantemente em cada núcleo foram, respectivamente, um, dois, três, quatro e cinco. Os resultados dos experimentos apontam que, quando o recurso de *scheduling affinity* está desabilitado, ou seja, a alocação das *threads* nos diferentes núcleos/processadores é definida pelo sistema operacional, há uma alta variação nos tempos de execução das *threads*, o que afeta diretamente a *cache* de processador e causa uma alta variação no número de *cache misses*.

Além disso, há o fato de ocorrer migração de *threads*, o que também impacta de forma negativa no desempenho. Enfim, o estudo apresentado aponta que mesmo um computador com baixo *overhead*, ou seja, com pouco processamento, com o *dynamic voltage scaling* (controle dinâmico de voltagem do computador) inativo e com a busca antecipada de dados e instruções na memória (*automatic hardware prefetcher*) desabilitada, possui tempos de execução muito instáveis em programas *OpenMP* em plataformas *multicore*. Os autores também informam que fixar *threads* nos núcleos remove a variabilidade do desempenho em muitos dos casos e ainda mostram que executar em separado outros processos em paralelo com as *threads* de um programa *OpenMP* pode ser benéfico para o tempo de execução em *user level*, mas não para todo o tempo de execução decorrido.

Outro trabalho similar ao anterior é apresentado em Mazouz; Touati; Barthou (2011), no qual foi realizado um estudo experimental completo do desempenho de várias estratégias de alocação de *thread* para programas *OpenMP*. O trabalho não se limita a encontrar a melhor alocação de *thread*, mas investiga os desempenhos globais de um programa *multithread* sob um conjunto predominante de escalonadores de *threads*. Os experimentos foram conduzidos utilizando todos os programas do *SPEC OMP2001* em três computadores distintos, sendo um *desktop*, um servidor e um HPC *Cache Coherent Non-Uniform Memory Access (ccNUMA)*. Os programas foram compilados com quatro, seis, oito, dezesseis e 96 *threads*. Por suportar um maior grau de paralelismo, o computador HPC *ccNUMA* foi o único a executar os programas compilados com 16 e 96 *threads*.

Métodos estatísticos foram utilizados para validar as conclusões obtidas. As estratégias de alocação de *thread* foram classificadas dentro de duas famílias: “Independente da Aplicação” e “Dependente da Aplicação”. A família “Independente da Aplicação” não diferencia os programas; as *threads* são alocadas da mesma forma (independente de *affinity*). Nessa família estão contidas quatro estratégias de *thread affinity* (sem *affinity*, *random*, *compact icc* e *scatter icc*).

A família “Dependente da Aplicação” aloca as *threads* de acordo com a característica do programa (depende da *affinity*). A melhor alocação para as *threads* foi baseada no compartilhamento de dados entre elas. Nessa família estão contidas 5 estratégias de *thread affinity* (GP, LP *compact*, LP *scatter*, LPGP e GPLP). Os resultados demonstram que em ambientes computacionais voltados para *desktop* e servidor, alocar uma *thread* em um processador específico leva a valores marginais de *speedups* e *slowdowns*. De forma contrária, em computadores *ccNUMA* os autores observaram *speedups* para todas as estratégias de *thread affinity*. Assim sendo, em termos de média ou de mediana dos tempos de execução, deixar o sistema operacional decidir sobre a alocação da *thread* não é um estratégia ruim. No entanto, há uma alta variação do desempenho.

Oliveira *et al.* (2013) apontam que a comunidade *Linux* tem realizado experimentos a fim de medir a latência de escalonamento e processamento computacional de dois conhecidos escalonadores de processos (*CFS* e *BFS*). No entanto, alguns fatores não percebidos podem interferir nos resultados dos tempos de execução dos programas, ou mesmo impedir a reproduzibilidade desses valores. Segundo os mesmos autores, ao reproduzir os experimentos propostos pela comunidade, não foi possível obter as mesmas conclusões, o que, muito provavelmente, se deva ao fato de a comunidade não ter utilizado uma metodologia estatística para obter e publicar os seus resultados. A fim de produzir e reproduzir os experimentos, Oliveira *et al.* (2013) realizaram um projeto experimental fatorial e uma análise de variância (*ANOVA*) (LILJA, 2005).

Contudo, o método *ANOVA* requer que os dados satisfaçam algumas propriedades estatísticas como, por exemplo, seguirem uma distribuição Gaussiana e serem independentes. Como não foi possível atender essas suposições, os autores utilizaram o método *Quantile Regression* (ROGER KOENKER, 2005), que se mostra adequado e eficiente ao permitir uma análise mais detalhada sem requerer as suposições exigidas pela *ANOVA*. Segundo Oliveira *et al.* (2013), o método *Quantile Regression* permite avaliar a quantificação dos efeitos do fator em qualquer dado *quantil*, sendo ele a mediana ou o 99º *percentil*.

O estudo de variação de desempenho realizado por Pusukuri, Gupta; Bhuyan (2012) apontou como diferentes políticas de alocação de memória junto com a migração de *threads* afetam o desempenho de programas de uso intensivo da memória e como o efeito de políticas de escalonamento de CPU afetam o desempenho de programas de uso intensivo da CPU. O estudo foi conduzido usando migração de *threads* e as políticas de alocação de memória *Next-Touch* (padrão), *Random* e *Round Robin*, além de apresentar prioridades dinâmicas e trocas de contexto involuntárias e a combinação de alocação de memória e políticas de escalonamento. Os autores utilizaram quinze programas *multithread*, incluindo *TATP* (programa de banco de dados), *SPECjbb2005*, *PARSEC* e *SPEC OMP*.

Os resultados apontaram uma redução significativa da variação de até 98% (média de 68%) e um aumento de até 43% (média de até 15%) no desempenho das políticas padrão do sistema operacional OpenSolaris. No sistema operacional Linux não foi possível utilizar a combinação *Random* e *Fixed Priority (FX)*, visto que o Linux não possui suporte para essas políticas. Nesse caso, utilizou-se a política de *Round Robin* para a alocação de memória junto com a política padrão *Next-Touch* para programas de uso intensivo da memória. Para essa categoria de programa, os

resultados apontaram uma redução de variação no desempenho de até 91% e um aumento do desempenho de até 53%.

O trabalho realizado por Wright *et al.* (2009) tem por objetivo alcançar um melhor entendimento sobre a variação dos tempos de execução de programas aparentemente idênticos. Para isso, os autores mediram os tempos de execução de dois programas, *PARATEC* e *WRF*, em três computadores HPC. Examinaram exatamente com que frequência as variações ocorrem nos tempos de execução e olharam para a distribuição de tempos obtida.

As medições sobre as variações de desempenho foram obtidas usando um conjunto de *benchmarks* que representam programas reais de HPC. Os programas foram executados com 256 *MPI* (*Message Passing Interface*) (GROPP; LUSK; SKJELLUM, 1999) *tasks* usando todos os núcleos disponíveis dentro do nó. Os computadores foram configurados para executarem duas vezes ao dia, uma a partir das seis horas da manhã e outra à noite, a partir das seis horas da tarde.

Em uma inspeção visual, os tempos de execução obtidos dos programas do *PARATEC* e do *WRF* apresentaram uma distribuição LogNormal. Além disso, a faixa de amplitude dos tempos de execução observada para o *PARATEC* esteve em torno de 17%, ao passo que, para o *WRF*, a execução mais lenta foi 66% maior do que a mais rápida. Isto ocorreu porque o *WRF* tem maior necessidade de *i/o* do que o *PARATEC* e alguns *jobs* que sofreram *delay* foram devidos a uma quantidade significante de *i/o*. Os resultados apontam maior variação em uma das máquinas com aproximadamente 25% das execuções demorando mais do que 10% acima do resultado mais rápido medido. Para os autores, a variação no tempo de execução foi causada pelo desempenho do *MPI*, muito provavelmente, pela contenção da rede.

## 2.4 Considerações Finais

Neste capítulo obteve-se os subsídios necessários para a compreensão da variabilidade que há nos tempos de execução dos programas de computador. Assim, sabe-se que, ao executar um programa várias vezes com o mesmo dado de entrada em um mesmo computador, dificilmente se observará tempos de execuções iguais. O que mais se observa são pequenas variações decorrentes da presença de fatores de *Hardware* ou *Software*, o que, na pesquisa experimental, é conhecido como **erro experimental**. Não tratar adequadamente esse erro experimental leva a resultados que, posteriormente, podem não ser reproduzidos. Além disso, identificou-se a importância da escolha dos parâmetros observados, uma vez que parâmetros que contemplam uma única execução não devem ser utilizados devido à natureza estocástica das variações dos tempos de execução dos programas.

Como pode ser visto, há vários trabalhos na literatura que estudam a variabilidade nos tempos de execução de programas e que sugerem o uso de uma metodologia estatística. No entanto, não há um padrão no uso dos testes estatísticos (TOUATI; WORMS; BRIAIS, 2013; GEORGES; BUYTAERT; EECKHOUT, 2007). Assim sendo, este trabalho de pesquisa difere dos trabalhos apresentados, por contribuir no sentido de aprimorar o protocolo proposto em Touati; Worms; Briais (2013) de modo a contemplar a análise de vários tratamentos utilizando um planejamento experimental, a análise de dados não paramétricos, além de evitar a ocorrência do problema *familywise error rate*, inerente às múltiplas comparações.

# 3. ESTUDO EXPLORATÓRIO EXPERIMENTAL SOBRE VARIAÇÃO EM TEMPOS DE EXECUÇÃO DE PROGRAMAS

## 3.1 Introdução

O tempo de execução de um programa é uma variável de estudo muito importante para algumas áreas da Ciência da Computação. Um exemplo é o estudo de desempenho de dois algoritmos, em que o pesquisador executa esses dois algoritmos e coleta o tempo de execução obtido para comparar qual deles é o melhor, segundo os seus critérios. No entanto, executar cada um deles uma única vez não é aconselhado, devido ao fato de que o tempo obtido pode ser fruto do acaso e não necessariamente será observado em uma execução futura (TOUATI; WORMS; BRIAIS, 2013). A maneira correta de avaliar esses dois algoritmos é produzir sucessivas execuções a fim de obter resultados mais consistentes e permitir uma análise estatística confiável. Contudo, ao executar um programa por um determinado número de vezes, o que se observa é que há uma variação dos tempos de execução obtidos (ver Seção 2.3).

Essa variação, conhecida como erro experimental, pode ser inerente à própria execução do programa ou pode ser causada por fatores não controláveis (MONTGOMERY, 2000). No caso de experimentos computacionais, esses fatores podem estar relacionados com o *hardware* e o *software*. Como as variações provocadas por esses fatores possuem uma natureza estocástica, medir a sua magnitude torna-se um trabalho de difícil predição. Por isso, trabalhos que dependem de uma correta análise dos tempos de execução dos programas, muitas vezes, falham em seu objetivo por não tratarem de forma correta esse problema.

Isso posto, com o intuito de demonstrar a variação nos tempos de execução de programas de computador decorrente de fatores não controlados, realizaram-se três experimentos computacionais. No Experimento #1, observou-se a variação dos tempos de execução dos programas decorrentes do *runlevel* do SO e da otimização do compilador. No Experimento #2, avaliou-se o impacto que o tamanho de uma variável de ambiente do SO causa nos tempos de execução dos programas. E no Experimento #3, avaliou-se a variação nos tempos de execução dos programas decorrente do número e das estratégias de alocação de *threads*. Além desses três, há o Experimento #4 que está descrito em Vicente; Matias Jr. (2013), que avalia a influência do *OSjitter* nos tempos de execução dos programas de computador. Nesta pesquisa, analisaram-se apenas os resultados obtidos por esses autores.

Neste capítulo é apresentada uma caracterização dos tempos de execução de programas de computador, a fim de se demonstrar a importância de controlar os fatores presentes no ambiente experimental, que interferem na análise dos resultados obtidos em experimentos computacionais.

## 3.2 Planejamento de Experimentos

### 3.2.1 Método

Nesta pesquisa, foi adotado o método estatístico de planejamento experimental DOE (BARBETTA; REIS; BORNIA, 2010; MONTGOMERY, 2000) para conduzir os experimentos, bem como analisar os seus resultados. Esse método requer modificações controladas dos fatores estudados, a fim de observar os efeitos dessas alterações sobre a variável resposta. Para esta pesquisa, definiu-se o tempo de execução dos programas como a variável resposta de interesse. Cada fator (fonte de variabilidade dos tempos de execução) assume diferentes níveis de operação e a combinação entre um fator e esses níveis configura um tipo de tratamento (BARBETTA; REIS; BORNIA, 2010). Para a definição dos tratamentos do Experimentos #1, foi utilizado o método da matriz de sinais (JAIN, 1991) configurada de acordo com a ordem de Yates (MONTGOMERY, 2000). No caso dos Experimentos #2 e #3, o método da matriz de sinais não pôde ser adotado, pois os seus tratamentos são compostos por mais de dois níveis por fator, o que inviabiliza o uso desse método. Assim, para o Experimento #2, foi descrito em que tamanho da variável de ambiente e em que tipo de otimização o tratamento foi replicado e, para o Experimento #3, foi adotada a seguinte definição: 0 para o nível mais baixo, 1 para o nível médio e 2 para nível mais alto.

Para todos os experimentos, o SO foi reinicializado a cada novo tratamento e, desse modo, evitou-se que a execução de um tratamento influenciasse nos resultados do seguinte.

### 3.2.2 Benchmarks

Segundo SAAVEDRA; SMITH (1996), *benchmarking* é a técnica utilizada para avaliar o desempenho de uma máquina ou sistema, através da execução de um programa. Esse programa pode ser uma aplicação completa ou, apenas, as partes mais executadas (*kernel*) (SAAVEDRA; SMITH, 1996).

Neste estudo, os experimentos foram realizados com *benchmarks* compostos por programas de código aberto que representam aplicações HPC. Os *benchmarks* selecionados foram o *NPB* (*NAS Parallel Benchmarks*) versão 3.3.1 (NASA ADVANCED SUPERCOMPUTING DIVISON, 2015) e o *SPLASH-2x* ( BARROW-WILLIAMS; FENSCH; MOORE, 2009; WOO *et al.*, 1995). O *SPLASH-2x* é uma variação do *SPLASH-2*, e foi desenvolvido por um grupo de pesquisa da Universidade de Princeton, para permitir o uso de *inputs*. O *benchmark PARSEC* versão 3.1 (BIENIA, 2011), não é exatamente um *benchmark HPC*, ele foi concebido com o intuito de disponibilizar programas para a avaliação de desempenho de computadores com multiprocessadores *multi-core*. Contudo, esse *benchmark* também foi selecionado pelo fato de oferecer um conjunto diversificado de programas.

### 3.2.2.1 NPB - NAS Parallel Benchmark

O *benchmark NPB* (NASA ADVANCED SUPERCOMPUTING DIVISON, 2015) formado por *benchmarks* derivados de aplicações *Computational Fluid Dynamics* (CFD) e que foram projetados para avaliar o desempenho de supercomputadores paralelos. A versão 3.3.1 é composta de cinco *kernels* e cinco *pseudo-programas*. O conjunto de *kernels* é composto pelos *benchmarks CG, EP, FT, IS* e *MG*. Com exceção do *EP* (*cpu-bound*), todos os outros *kernels* são *memory-bound*. Os cinco *pseudo-programas* que compõe a versão 3.3.1 do NPB são: *BT, DC, LU, SP* e *UA*. Os *benchmarks BT* e *LU* são *cpu-bound*, ao passo que os *benchmarks SP* e *UA* são *memory-bound*, e o *DC* é *i/o-bound*.

O fato do programa ser *cpu-bound* indica que ele faz um uso intensivo da *cpu* para realizar o seu processamento, no caso de programas *memory bound*, o tempo de processamento do programa está ligado a quantidade de memória necessária para manter os seus dados e, no caso de *i/o-bound*, indica que o programa faz um uso intensivo de operações de entrada/saída.

A seguir é apresentada uma descrição de cada *kernel* e *pseudoprograma* que compõem este *benchmark*:

- O *CG* é utilizado para computar uma aproximação para um menor autovalor de uma matriz, grande, dispersa e simétrica positiva;
- O *EP* fornece uma estimativa dos limites superiores alcançáveis para o desempenho de ponto flutuante, por exemplo, o desempenho sem comunicação significante entre processadores;
- O *FT* resolve parte de uma equação diferencial tridimensional, utilizando transformada de *Fourier*, sendo um rigoroso teste de desempenho de uma carregada comunicação de longa distância. O problema a ser resolvido, por este *kernel*, é uma discreta versão do original *Poisson Partial Differential Equation* (PDE);
- O *IS* por meio de acesso aleatório a memória, desempenha uma grande operação de ordenação de inteiros, testando a velocidade de computação e o desempenho de comunicação;
- O *MG* realiza um cálculo *multigrid* simplificado, que requer uma comunicação de dados de longa distância altamente estruturada e testes tanto para comunicações de dados de curta e longa distância;
- O *BT* desempenha um sintético problema de CFD, ao resolver, de forma múltipla, sistemas independentes não diagonalmente dominantes, com blocos de equações tri-diagonais (5x5);
- O *DC* foi proposto para testar as capacidades da *grid* em manipular grandes conjuntos de dados distribuídos. Esse programa estressa todos os níveis da memória da *grid*, uma vez que produz 2<sup>d</sup> visões de um

conjunto de dados aritméticos de *d-tuplas* descritos por um pequeno número de parâmetros inteiros;

- O *LU* desempenha um sintético cálculo de CFD, ao resolver uma regular matriz esparsa em sistemas triangulares superiores e inferiores com bloco (5x5);
- O *SP* desempenha um sintético problema de CFD, ao resolver de forma múltipla, sistemas independentes não diagonalmente dominantes, escalares e equações penta-diagonais, e;
- *UA* mede o desempenho de sistemas computacionais modernos quando necessitam resolver problemas científicos caracterizando acesso irregular a memória dinâmica. Esse programa envolve a solução de um estilizado problema de transferência em um cubo discretizado em uma refinada rede não estruturada.

Os *benchmarks* disponibilizados pelo *NPB* podem ser compilados para diversas plataformas e essa compilação pode ser feita utilizando um dos vários arquivos *make*, oferecidos pelo *NPB* e que contém os parâmetros necessários para cada plataforma. Além disso, também é possível escolher três tipos de execução: serial, com *MPI* ou com *OpenMP* e não é necessário que os *benchmarks* sejam executados em conjunto, eles podem ser utilizados individualmente.

Os *inputs*, que representam as entradas para o processamento, estão divididos em classes. Cada classe contém uma configuração específica de cada parâmetro presente no problema a ser resolvido pelo *benchmark*. O Apêndice A apresenta os parâmetros e as configurações de cada classe para cada *benchmark NPB*. Para alterar o tipo de *input*, basta alterar o arquivo de configuração *suite*.

### 3.2.2.2 SPLASH-2x

O *benchmark SPLASH-2x* ( BARROW-WILLIAMS; FENSCH; MOORE, 2009; WOO *et al.*, 1995) , foi elaborado por um grupo de pesquisa da Universidade de Princeton (grupo PARSEC). Esse grupo alterou o *benchmark SPLASH-2* de modo a permitir que ele aceitasse o uso de *inputs*. O Apêndice B apresenta os parâmetros e as configurações de cada conjunto de *input* para cada programa do *SPLASH-2x*. Esses dois *benchmarks*, *SPLASH-2* e *SPLASH-2x* são disponibilizados em conjunto com o *benchmark PARSEC*. O *SPLASH-2x* possui um conjunto de programas paralelos e foi disponibilizado no intuito de facilitar o estudo de multiprocessadores com espaço de endereçamento compartilhado centralizado e compartilhado distribuído; seus programas possuem características tanto de *cpu-bound*, quanto de *memory bound*.

O *SPLASH-2x* é composto por cinco *kernels* e nove programas e representa uma ampla faixa de programas científicos, de Engenharia e de gráficos. Os cinco *kernels* que o *SPLASH-2x* disponibiliza são: *cholesky*, *fft*, *lu* (versão *cb* e *ncb*) e o *radix*, e os nove programas são: *barnes*, *fmm*, *ocean* (*cp* e *ncp*), *radiosity*, *raytrace*, *volrend*, *water\_nsquared* e *water\_spatial*. A seguir segue uma descrição de cada programa disponibilizado por esse *benchmark*:

- O *cholesky* fatora uma dada matriz esparsa dentro do produto de matriz triangular inferior e sua transposta. Este *kernel* opera em matrizes

esparsas, que têm uma grande razão de comunicação no processamento para problemas de tamanhos comparáveis;

- O *fft* é uma complexa versão unidimensional de uma raiz; ele foi otimizado para minimizar a comunicação interprocessadores. O conjunto de dados desse *kernel* consiste de  $n$  ponteiros de dados complexos para serem transformados e  $n$  ponteiros para referenciar como unidades da raiz. Ambos os conjuntos de dados são organizados como matrizes:  $\sqrt{n} \times \sqrt{n}$ , particionadas tal que todo processador é associado a um conjunto contíguo de linhas as quais são alocadas dentro da memória local;
- O *lu* fatora uma matriz densa dentro do produto de uma triangular inferior e em uma matriz triangular superior. A matriz  $n \times n A$ , é dividida dentro de um *array*  $N \times N$  de  $B \times B$  blocos ( $n = N \times B$ ) para explorar a localidade temporal em elementos de uma submatriz. O bloco de tamanho  $B$  deve ser grande o suficiente para manter o *cache miss rate* baixo, e pequeno o suficiente para manter um bom balanço de carga. O SPLASH-2x conta com duas versões desse *kernel*; a primeira é a *lu\_cb* (*lu blocos contíguos*); essa versão indica que o *lu* utiliza um *array* quadridimensional para representar a matriz bidimensional, assim, o bloco é contíguo no espaço de endereço virtual e cada página de dados é alocada em um único processador. A segunda versão *lu\_ncb* (*lu blocos não contíguos*) utiliza um *array* bidimensional para representar a matriz, assim, as sucessivas sublinhas de um bloco não são contíguas com um outro bloco no espaço de endereço. Nesta versão, os dados são escritos por múltiplos processadores e abrangem uma página;
- O *radix* possui um algoritmo interativo que desempenha uma interação para cada raiz de tamanho  $r$ . Para cada interação, um processador passa sobre uma chave associada e gera um histograma local. Os histogramas locais são então acumulados dentro de um histograma global. Finalmente, cada processador usa o histograma global para permutar as chaves dentro de um novo *array* para a próxima interação. Assim, tem-se uma comunicação dos interprocessadores, todos com todos;
- O *barnes* utiliza o método *Barnes-Hut hierarchical N-body* para simular a interação de sistemas de corpos (galáxia ou partículas, por exemplo) em três dimensões sobre um número de *time-steps*;
- O *fmm*, como o *barnes*, também simula um sistema de corpos sobre um número de *time-steps*, no entanto, as interações são simuladas em duas dimensões usando o *Fast Multipole Method*. O padrão de comunicação é bem desestruturado e não tenta fazer uma distribuição inteligente da partícula do dado na memória principal;
- O *ocean* estuda em grande escala os movimentos do oceano baseado em correntes de contorno e redemoinhos. Essa aplicação partitiona as *grids* dentro de *subgrids* quadradas em vez de grupos de colunas para aumentar a comunicação. Na versão contígua, as *grids* bidimensionais são representadas como *arrays* quadridimensionais, com todas as

*subgrids* alocadas contiguamente e localmente nos nós pertencentes a ela. Na versão não contígua os *arrays* são bidimensionais;

- O *radiosity* calcula a distribuição de equilíbrio da luz em uma cena, usando o método de hierarquia interativa difusa *radiosity*. Uma cena é inicialmente modelada com um número grande de polígonos. As interações de transporte da luz são computadas entre estes polígonos, e os polígonos são hierarquicamente subdivididos dentro de *patches* necessários para aumentar a precisão. A estrutura de processamento e os padrões de acesso as estruturas de dados são altamente irregulares e o paralelismo é gerenciado pelas filas de tarefas distribuídas, sendo uma por processador;
- O *raytrace* renderiza uma cena tridimensional usando o traço de raios. Uma *grid* uniforme hierárquica é usada para representar a cena e, em breve, a terminação do raio e a suavização são implementadas. Um raio é traçado por meio de cada *pixel* na imagem e se reflete em imprevisíveis maneiras fora dos objetos com os quais colide. A imagem é então particionada entre os processadores em blocos de grupos de *pixels* contíguos;
- O *volrend* renderiza um volume tridimensional usando uma técnica de *ray casting*. O volume é representado como um cubo de *voxels* e uma estrutura de dados *octree* é usada para permitir que o volume seja transversal rapidamente. Este programa renderiza vários *frames* de mudança de ponto de vista. As principais estruturas de dados são os *voxels*, *octree* e *pixels*. Os acessos aos dados são independentes e irregulares e não há uma distribuição de dados inteligente;
- O *water\_nsquared* avalia forças e potências que ocorrem em todo tempo em um sistema de moléculas de água. As forças e os potenciais são computados usando um algoritmo  $O(n^2)$  e um método *predictor-corrector* é usado para integrar o movimento das moléculas da água, e;
- O *water\_spatial* resolve o mesmo problema do *water\_nsquared*, mas utiliza um algoritmo mais eficiente. Ele impõe uma *grid* tridimensional de células em um domínio, e usa um algoritmo  $O(n)$ , que é mais eficiente que o anterior para grandes números de moléculas.

O grupo *PARSEC* oferece um conjunto de *scripts* que permite “gerenciar” o uso dos *benchmarks* que ele disponibiliza. Por exemplo, para compilar todos os programas do *SPLASH-2x*, é preciso utilizar o programa *parsecmgmt* com o comando *build* e o parâmetro *splash2x*. Os programas do *SPLASH-2x* podem ser utilizados individualmente, não sendo necessário que eles executem em conjunto.

Para compilar especificamente um programa, é preciso utilizar o programa *parsecmgmt* passando como parâmetro o comando *build* e o programa que se deseja compilar, por exemplo, *parsecmgmt build splash2x.barnes*. Para executar o programa, basta alterar o comando *build* para *run*. O grupo *PARSEC* disponibilizou uma coletânea de arquivos de configuração *make* que permite a compilação desses *benchmarks* para diferentes plataformas.

### 3.2.2.3 PARSEC

O PARSEC (BIENIA, 2011) é composto por programas do *RMS* suíte (*benchmark* da Intel), da Universidade de Princeton, da Universidade de Stanford e programas de código aberto. O intuito foi disponibilizar uma suíte que pudesse auxiliar as pesquisas em computadores com multiprocessadores *multi-core*. Uma das metas foi unir uma seleção de programas suficientemente grande e diversa para ser representativa para estudos científicos.

Os programas deste *benchmark* são considerados mais *cpu-bound* do que *memory-bound* (BHADAURIA; WEAVER; MCKEE, 2009). Ao todo, o PARSEC disponibiliza três *kernels* (*canneal*, *dedup* e *streamcluster*) e nove programas (*blackscholes*, *bodytrack*, *facesim*, *ferret*, *fluidanimate*, *freqmine*, *swaptions*, *vips* e *x264*). A seguir segue uma descrição de cada programa disponibilizado pelo PARSEC:

- O *canneal* usa *cache-aware simulated annealing* (SA) para minimizar o custo de roteamento do projeto de um chip. Esse *kernel* pseudorandomicamente seleciona pares de elementos e tenta trocá-los;
- O *dedup* comprime uma *stream* de dados com uma combinação de compressão global e uma local, buscando alcançar altas taxas de compressão. Esse *kernel* usa um modelo de programação *pipelined* para paralelizar a compressão imitando implementações reais;
- O *streamcluster* resolve o problema de *clustering online*: Para um *stream* de ponteiros de entrada, ele localiza um número predeterminado de medianas, tal que cada ponto é associado a um centro mais próximo. A qualidade do *clustering* é medida pela soma da distância dos quadrados. *Stream clustering* é uma operação comum pela qual grande quantidade de dados ou dados produzidos continuamente têm de ser organizados sob condições de tempo real;
- O *blackscholes* calcula o preço para um portfólio de ações europeias analiticamente com a equação *Back-Scholes Partial Differential*. Este programa divide o portfólio dentro de um número de unidades de trabalho igual ao número de *threads* e processa-as concorrentemente;
- O *bodytrack* traça uma pose tridimensional de um corpo humano sem marcadores usando apenas múltiplas câmeras por meio de uma sequência de imagens. Este programa emprega um filtro de partícula temperada para traçar a pose usando bordas e a silhueta em primeiro plano como características da imagem;
- O *facesim* pega o modelo de um rosto humano em uma sequência de ativações musculares e computa uma animação visualmente realística da face modelada por simular a física básica;
- O *ferret* é uma aplicação baseada no *Ferret toolkit* e é utilizado em busca de similaridade baseada em conteúdo. No PARSEC, este programa foi configurado para busca de similaridade de imagem;

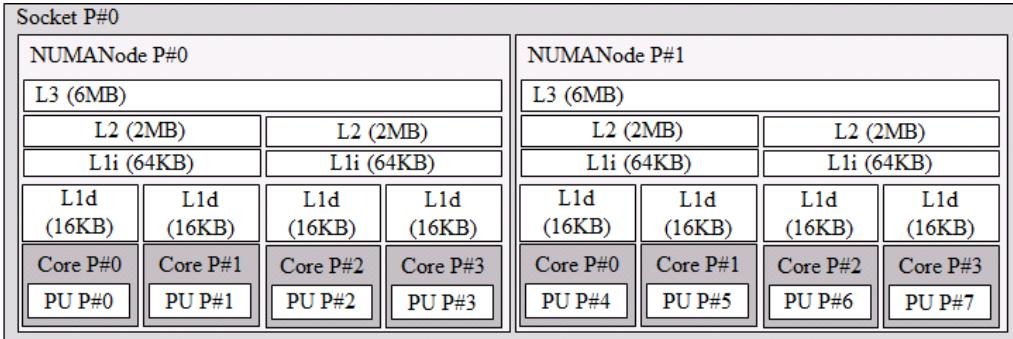
- O *fluidanimate* usa uma extensão do método *Smoothed Particle Hydrodynamics* (SPH) para simular um fluído incompreensível para propósitos de animação interativa. Sua saída pode ser visualizada pela detecção e renderização da superfície do fluído. A cena geométrica empregada pela aplicação é uma caixa na qual o fluído reside. Todas as colisões são manipuladas pela adição de forças a fim de alterar a direção do movimento da partícula envolvida ao invés de modificar a velocidade diretamente;
- O *freqmine* emprega uma versão baseada em *array* do método *FP-growth* (*Frequent Pattern-growth*) para o *Frequent Itemset Mining* (FIMI). FIMI é a base do *Association Rule Mining* (ARM), um problema de *data mining* muito comum o qual é relevante para áreas tais como sequência de proteínas, dados de mercado ou análises de *log*;
- O *swaptions* utiliza o *framework Heath-Jarrow-Morton* (HJM) colocar preço em um portfólio de opções de *swap*. Para computar os preços, este programa emprega uma simulação *Monte Carlo*;
- O *vips* é baseado no *VASARI Image Processing System* (VIPS), este programa inclui operações fundamentais com imagem, tais como, *affine transformation* e convolução, e;
- O *x264* é um H.264/AVC (*Advance Video Coding*), baseado no padrão ITU-T H.264 o qual foi completado em maio de 2003 e que agora também é parte ISO/IEC MPEG-4. Os *encoders* e *decoders* H.264 operam em macro blocos de *pixels* com tamanho fixo de 16x16 pixels. Várias técnicas são usadas para detectar e eliminar redundância de dados.

Os programas do *PARSEC*, como os programas do *SPLASH-2x*, podem ser compilados e executados individualmente, não é necessário que sejam utilizados em conjunto. A compilação destes programas é realizada por meio de um *script* de gerenciamento (*parsecmgmt*), o qual é disponibilizado pelo grupo mantenedor do *benchmark*. Os conjuntos de *inputs* para cada programa deste *benchmark* são apresentados no Apêndice C.

### 3.2.3 Ambiente de Teste e Instrumentação

Os experimentos #1, #2 e #3 foram realizados em um computador com quatro processadores *AMD Opteron(TM)* 6212 de 1,40GHz e 32 núcleos, com três níveis de *cache* e 64 GB de memória *RAM*. A Figura 3.1 ilustra, com detalhes, a topologia dos processadores utilizados nos experimentos. É possível observar que cada processador (*socket*) é composto por dois nós NUMA e cada um desses nós é composto por quatro núcleos com três níveis de *cache*. O último nível de *cache* (L3) possui um tamanho de 6 MB e é compartilhada pelos quatro núcleos do mesmo *socket*. O nível de *cache* L2 possui dois módulos com tamanho de 2 MB e cada um deles é compartilhado por dois núcleos. O primeiro nível de *cache* (L1) é dividido em módulos de instrução e módulo de dados. No caso da *cache* L1 de instrução (L1i), são dois módulos com tamanho de 64KB, em que cada módulo é compartilhado entre dois núcleos. No caso da *cache* L1 de dados (L1d), é um único módulo por núcleo com tamanho de 16KB. O SO utilizado foi o *OpenSuse 13.1* com uma instalação

de interface gráfica padrão e apenas os serviços padrão do SO, compatíveis com cada nível de execução, estavam em execução (ver Apêndice D) durante os experimentos. Além disso, o início aleatório da pilha de memória foi desabilitado, utilizando o comando `echo 0 > /proc/sys/kernel/randomize_va_space`.



**Figura 3.1. Topologia de um dos quatro processadores usados nos Exp. #1, #2 e #3.**

Para o Experimento #1 utilizou-se a versão não paralela dos programas que compõe os *benchmarks* PARSEC versão 3.1, SPLASH-2x e o NPB (*NAS Parallel Benchmarks*) versão 3.3.1. Esses *benchmarks* foram escolhidos devido a diversidade de seus programas, assim, os efeitos dos fatores poderiam ser observados em uma gama maior de programas. O Experimento #2 foi realizado com a versão não paralela dos programas disponibilizados pelos *benchmarks* do NPB e pelos programas do SPLASH-2x. Neste caso, o intuito foi observar apenas o comportamento de programas HPC, uma vez que, os programas do PARSEC não são voltados para esse tipo de programa, não fez sentido utilizá-los. No caso do Experimento #3, por ser necessário o uso de paralelismo, foram utilizados os *benchmarks* disponibilizados pelo NPB e os programas *blackscholes*, *bodytrack* e *freqmine* do PARSEC. Esses programas foram escolhidos por serem os únicos, entre os outros programas do PARSEC e do SPLASH-2x, que oferecem uma versão OpenMP. Nesta pesquisa, os *benchmarks* foram compilados com o *gcc* 4.7.2 (GCC, 2014).

Nos experimentos #1, #2 e #3 cada tratamento foi replicado 31 vezes; a primeira replicação foi descartada, afim de, se analisar os dados do programa na *cache* e sem maior influência do *buffer* de disco. Cada programa de cada *benchmark* foi executado como se fosse um serviço do SO, de modo que, a cada término de replicação de um tratamento o SO fosse reiniciado, sem a intervenção humana, evitando assim, que a execução de um tratamento influenciasse a execução do próximo. Tanto para o SPLASH-2x, quanto para o PARSEC, foi adotado o *input* do tipo *native* (ver Apêndices B e C); no caso do NPB foi adotada a classe *B* (ver Apêndice A). Além disso, a coleta dos tempos de execução foi realizada de forma padronizada utilizando o comando *time* do UNIX (KERRISK, 2010), o qual retorna o tempo de execução de um comando, e a unidade métrica definida foi o segundo.

O NPB utiliza a função *gettimeofday* (KERRISK, 2010) para coletar seus tempos de execução, porém, o PARSEC e o SPLASH-2x utilizam o programa *time* e alterar os códigos fonte para o uso de outro método poderia comprometer o comportamento dos programas. A Figura 3.2 descreve o *script* utilizado para obter o tempo de execução do programa *blackscholes* do *benchmark* PARSEC no Experimento #1.

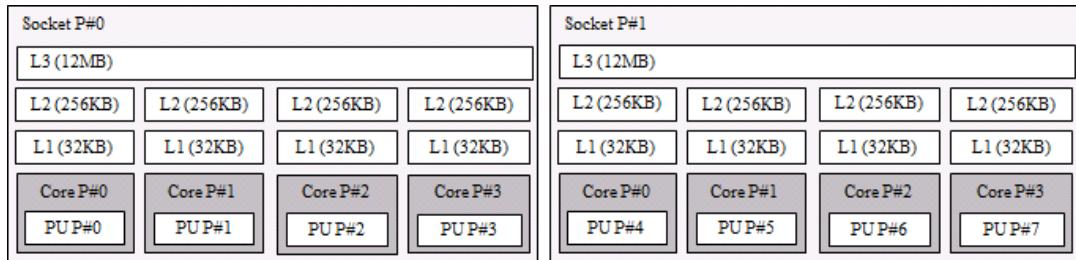
```

1 #!/bin/bash
2 :
3 count=0
4 while [ $count -le 30 ]
5 Do
6     /usr/bin/time -p -a -o
7     /home/paulo/HPDCS/workspace/PARSECRL/resultados/
8     blackscholesr11_03.time.out /home/paulo/HPDCS(parsec-
9     3.0/pkgs/apps/blackscholes/inst/amd64-linux.gcc-
     serial/bin/blackscholes 1 /home/paulo/HPDCS(parsec-
     3.0/pkgs/apps/blackscholes/run/in_10M.txt prices.txt
7     ((count++))
8 done
9 :

```

**Figura 3.2. Script para a coleta de tempo de execução do programa *blackscholes*.**

O Experimento # 4 é descrito em Vicente; Matias Jr. (2013) e seus resultados foram analisados neste estudo. Os autores realizaram esse experimento em um computador com dois processadores *Quad-core Intel Xeon E5620* de *2,40GHz*, com *24 GB* de memória *RAM* e com um disco rígido de *1 TB*. Cada processador (*socket*) é composto por três níveis de *cache*. O primeiro e o segundo níveis (L1 e L2 respectivamente) são individuais para cada núcleo e o terceiro nível (L3) é compartilhado entre todos os núcleos do mesmo processador (ver Figura 3.3).



**Figura 3.3 Topologia dos processadores utilizados no Exp. #4.**

O SO utilizado foi o *Linux kernel 2.6.32.28*, distribuição *OpenSUSE 11.4* com uma instalação de interface gráfica padrão, no entanto, para melhor controlar as fontes de *OS Jitter*, os autores aplicaram algumas alterações no *kernel*. Para esse experimento os autores utilizaram um típico programa HPC que consiste em multiplicar matrizes. Esse programa recebeu o nome de *programa prova* (ver Anexo G). Os tratamentos foram replicados 53 vezes e as três primeiras replicações foram descartadas por causa da possibilidade de influência do *file system* e das *caches* do processador. A coleta dos tempos de execução foi realizada utilizando a *syscall gettimeofday* (KERRISK, 2010) e o segundo também foi a unidade métrica escolhida.

### 3.2.4 Experimento #1

Com o Experimento #1 teve-se o objetivo de observar o impacto do *runlevel* e da otimização do compilador sobre o tempo de execução de um programa, quando

estes estão em um ambiente real de execução. Os níveis foram escolhidos por, geralmente, serem os utilizados nos ambientes computacionais (ver Tabela 3.1).

**Tabela 3.1. Fatores e níveis avaliados no Exp. #1.**

Fatores		Nível (-)	Nível (+)
	Runlevel (RL)	5	3
Otimização (O)	02	03	

No nível (-) o fator *Runlevel* assume o valor 5, o que indica um número maior de processos administrativos (serviços) do SO executando em segundo plano (*background*); o nível (+) configura este fator para 3, o que resulta em um número reduzido de processos administrativos. Variando esse fator, buscou-se verificar o efeito dos processos administrativos do SO no tempo de execução dos programas. Já o fator *Otimização* refere-se ao nível de otimização aplicada a cada programa durante sua compilação. No nível (-), os programas foram compilados com a otimização 02 e no nível (+) com a otimização 03. A 03 não prioriza o tamanho do arquivo executável, mas sim, a velocidade. O uso do parâmetro *-finline-functions*, por parte da 03, permite o aumento de desempenho do executável, no entanto, dependendo da função *inline* o tamanho do arquivo executável pode aumentar consideravelmente. Isso porque, esse parâmetro integra o código, da função chamada, ao código de “quem” a chamou, aumentando assim o código de um programa (STALLMAN; COMMUNITY, 2012).

Assim, embora os executáveis se tornem rápidos, se o tamanho do arquivo exceder o tamanho da *instruction cache* poderá ocorrer um aumento de *cache misses*. Por isso, compilar com otimização 02 pode melhorar as chances do executável se ajustar ao tamanho da *instruction cache*. A Tabela 3.2 apresenta como ficou definida a configuração dos tratamentos.

**Tabela 3.2. Configuração dos tratamentos do Exp. #1.**

Tratamento	Descrição
<b>RL302</b>	<i>Runlevel 3, Otimização 02</i>
<b>RL303</b>	<i>Runlevel 3, Otimização 03</i>
<b>RL502</b>	<i>Runlevel 5, Otimização 02</i>
<b>RL503</b>	<i>Runlevel 5, Otimização 03</i>

### 3.2.5 Experimento #2

O Experimento #2 reproduz o experimento descrito em (MYTKOWICZ *et al.*, 2009) e consiste em avaliar a variabilidade dos tempos de execução de programas causada por alterações no tamanho de uma variável de ambiente do SO. Para esse experimento definiu-se o *runlevel* em 1, reduzindo o número de serviços do SO em *background* e permitindo uma melhor observação dos efeitos causados pelos fatores aqui estudados (ver Tabela 3.3).

**Tabela 3.3. Fatores e níveis avaliados no Exp. #2.**

Fatores	Níveis	
	<i>Tamanho da Variável de Ambiente (T)</i>	0, 64, 128, 256, 512, 1024, 2048 e 4096
	<i>Otimização (O)</i>	02 e 03

O fator *Tamanho da Variável de Ambiente* está relacionado com as variáveis de ambiente presentes no SO. Para esse experimento foi criada uma variável que recebe um valor do tipo *string* e não está relacionada com os programas utilizados nesse estudo. O tamanho dessa variável foi alterado em número de *bytes*, de acordo com os níveis adotados. Assim como no Exp. #1, o fator *Otimização* refere-se ao nível de otimização aplicada a cada programa durante sua compilação.

### 3.2.6 Experimento #3

O experimento “*Analyzing the Impact of Thread Affinity on SPEC OMP Performance Variability*” descrito no trabalho de (MAZOUZ; BARTHOU, 2011), tem por objetivo avaliar a variabilidade dos tempos de execução dos programas devido as estratégias de alocação de *threads* em programas paralelos *OpenMP*. Nesse experimento, os autores compilaram os programas do *benchmark* SPEC OMP2001 (SPEC, 2013c) utilizando o *gcc* com o parâmetro *-O3 -fopenmp* e o *icc* com o parâmetro *-O3 -openmp*, além disso, utilizaram três configurações: sem *affinity*, *compact affinity* e *scatter affinity*. Na primeira configuração o SO é livre para alocar as *threads* nos processadores de acordo com sua disponibilidade. Na segunda, o SO é instruído a alocar as *threads* em cada núcleo de modo sequencial, tão próximas quanto possível, para aumentar as chances de compartilhamento da cache de processador. Na terceira, o SO é instruído a distribuir as *threads* pelos núcleos o mais uniformemente possível (NOGUEIRA; MATIAS JR., 2014). O Apêndice E ilustra a distribuição das *threads* com as estratégias *compact* e *affinity*.

Nesta pesquisa, a reprodução destas três estratégias de afinidade de *threads* foi possível utilizando as funcionalidades implementadas pela *libgomp* (GNU, 2014) e pelo *gcc*. Nesse caso, a escolha da estratégia é feita pela variável de ambiente *GOMP\_CPU\_AFFINITY*. A configuração do *gcc+libgomp* compatível com a *icc-compact* é *GOMP\_CPU\_AFFINITY=0-7*. Para implementar a *icc-scatter*, utilizou-se *GOMP\_CPU\_AFFINITY=0 4 2 6 1 5 3 7*. Para esse experimento definiu-se o *runlevel* em 1 reduzindo o número de serviços do SO em *background* e permitindo uma melhor observação dos efeitos causados pelos fatores aqui estudados. A Tabela 3.4 apresenta os níveis e fatores para esse experimento e a Tabela 3.5 apresenta como ficou definida a configuração dos tratamentos. Uma vez que cada fator foi avaliado em três níveis, foram realizados  $3^2$  tratamentos.

**Tabela 3.4. Fatores e níveis avaliados no Exp. #3.**

Fatores		0	1	2
	<i>Número de Threads (NT)</i>	2	4	6
	<i>Estratégias de Affinity (EA)</i>	<i>sem affinity</i>	<i>compact</i>	<i>scatter</i>

**Tabela 3.5. Configuração dos tratamentos do Exp. #3.**

Tratamento	Descrição
<b>SA2TH</b>	Sem <i>Affinity</i> com duas <i>threads</i>
<b>SA4TH</b>	Sem <i>Affinity</i> com quatro <i>threads</i>
<b>SA6TH</b>	Sem <i>Affinity</i> com seis <i>threads</i>
<b>CP2TH</b>	<i>Compact</i> com duas <i>threads</i>
<b>CP4TH</b>	<i>Compact</i> com quatro <i>threads</i>
<b>CP6TH</b>	<i>Compact</i> com seis <i>threads</i>
<b>SC2TH</b>	<i>Scatter</i> com duas <i>threads</i>
<b>SC4TH</b>	<i>Scatter</i> com quatro <i>threads</i>
<b>SC6TH</b>	<i>Scatter</i> com seis <i>threads</i>

### 3.2.7 Experimento #4

O Experimento #4 é descrito em Vicente; Matias Jr. (2013) e seu objetivo foi avaliar o impacto causado pelas rotinas internas do SO (*OS Jitter*) no tempo de execução dos programas de usuários em HPC. Este experimento concentrou-se apenas em analisar os resultados obtidos em Vicente; Matias Jr. (2013). Esse experimento foi dividido em dois. Em um primeiro momento, os autores avaliam a execução do *programa prova* sob a influência de fatores presentes no SO, sem a execução de um processo concorrente. A Tabela 3.6 apresenta os fatores selecionados pelos autores para realizar a primeira parte desse experimento.

**Tabela 3.6. Fatores e níveis avaliados no Exp. #4. (adaptado de (VICENTE; MATIAS JR., 2013)).**

		Nível (-)	Nível (+)
Fatores	<i>Runlevel</i>	5	1
	<i>Kernel Timers</i>	<i>Off</i>	<i>On</i>
	<i>IRQ Handling</i>	<i>Off</i>	<i>On</i>
	<i>Processor Affinity</i>	<i>Off</i>	<i>On</i>
	<i>Timer Interrupt Handling</i>	<i>Off</i>	<i>On</i>

Como no Experimento #1, o nível (-) do fator *Runlevel* assume o valor 5, o que indica um número maior de processos administrativos (serviços) do SO executando em segundo plano (*background*); todavia, diferentemente do Exp. #1, o nível (+) configura este fator para 1, o que resulta em um número mínimo de processos administrativos. No entanto, o objetivo aqui é o mesmo do Exp. #1, observar o impacto causado pela execução de vários processos de forma concorrente com o programa em estudo. O nível (-) do fator *Kernel Timers* indica que os *timers* estão desligados, tanto no nível do *kernel*, quanto no do usuário no mesmo processador que executa o *programa prova* (*PU #1*). O nível (+) indica que é possível executar os *timers* no mesmo processador que está executando o *programa prova*. Assim, é possível observar o efeito causado por possíveis interferências desse fator no tempo de execução do *programa prova*. Com o fator *IRQ Handling*, os autores observaram o efeito causado pelas manipulações das interrupções de *hardware*, por isso, o nível (-) indica que todas as requisições de interrupção de *hardware IRQs* são direcionadas para o *PU #0* (ver Figura 3.3) e o nível (+) indica que todas as *IRQs* são manipuladas pelo *PU #1* concomitantemente com a execução do *programa prova*. O nível (-) do fator *Process Affinity* indica que ele está desabilitado e, portanto, os processos podem ser executados em qualquer processador. O nível (+) indica que esse fator está habilitado e, conforme configuração dos autores, todos os processos são executados no *PU #0*. Segundo os autores isso permite observar a interferência

direta de outros processos em um dado *runlevel* sobre o *programa prova*. Para finalizar, o nível (-) do fator *Timer Interrupt* indica que esse fator está desabilitado no *PU #1*. Para desabilitar e habilitar esse fator, os autores criaram uma chamada de sistema que é invocada de dentro do *programa prova*.

Em um segundo momento, outro fator foi adicionado para representar um processo executando em *background* concorrendo, contra o *programa prova*, por recursos do computador. O objetivo, nesse caso, é observar como o subsistema de gerenciamento de memória do SO lida com o aumento de *cache misses* em L3. O fator que representa essa concorrência é aqui chamado de *Shared Cache L3*. Quando o fator *Shared Cache L3* está em nível (-) significa que o processo concorrente está executando em *PU #5* (ver Figura 3.3) e, portanto, não compartilha a *cache L3* com o *programa prova*, uma vez que, esse executa em *PU #1*. No nível (+), o programa concorrente passa a executar em *PU #2*, o qual compartilha L3 com o *programa prova* (VICENTE; MATIAS JR., 2013). Com a inclusão desse sexto fator, o Exp. #4 passa a contar com 64 tratamentos. O Apêndice F apresenta uma legenda contendo a combinação dos fatores presente em cada tratamento.

### 3.3 Análise dos Resultados

As amostras dos tempos de execução dos programas obtidas em cada tratamento de cada experimento foram tabuladas utilizando o *Software* de estatística *R* (R Core Team, 2014). Os resultados são apresentados na forma de gráficos do tipo *violin plot*. Esses gráficos são parecidos com os do tipo *boxplot*, a diferença está no fato de o *violin plot* incorporar a densidade de probabilidade dos dados (HINTZE; NELSON, 1998). O ponto branco no *violin plot* indica a mediana observada, o retângulo preto indica o intervalo interquartílico, as extremidades da linha indicam os valores mais ao extremo e as pontas do gráfico indicam a presença de valores discrepantes (HINTZE; NELSON 1998). O eixo X do gráfico representa os tratamentos definidos para o experimento e o eixo Y os 30 tempos de execução observados para cada tratamento.

#### 3.3.1 Experimento #1

As tabelas contendo a análise descritiva de cada *benchmark* para esse experimento estão disponibilizadas no Apêndice G.

Ao analisar os tempos de execução dos *benchmarks* disponibilizados pelo *NPB*, observou-se que, em 80,00% dos casos os *benchmarks*, alcançaram o menor tempo de execução nos tratamentos com o fator *Otimização* em 02 (40,00% com o tratamento RL302 e 40,00% com o tratamento RL502). A Tabela 3.7 sumariza o resultado para cada tratamento avaliado.

Quando se observa a diferença entre o maior e o menor tempo entre os tratamentos, é possível notar que no caso dos *benchmarks* *BT*, *CG*, *IS* e *MG* essa diferença é pequena, mas no caso dos *benchmarks* *DC*, *EP*, *LU* e *SP* essa diferença é maior, em torno de 70 segundos no caso do *DC* com tratamento RL503, em torno de doze segundos no caso do *EP* com os tratamentos RL302 e RL502 e próximo de sete segundos no caso do *SP* para o tratamento RL302 (ver Apêndice G).

**Tabela 3.7. Distribuição do maior e menor tempo por tratamento para o NPB.**

	<b>RL302</b>	<b>RL303</b>	<b>RL502</b>	<b>RL503</b>
<b>Maior Tempo</b>	40,00%	10,00%	30,00%	30,00%
<b>Menor Tempo</b>	40,00%	20,00%	40,00%	20,00%

Ao realizar uma análise descritiva observou-se que 60% dos *benchmarks* disponibilizados pelo *NPB* não apresentaram variâncias que fossem superiores a um segundo, dos outros 40% é preciso destacar que o *benchmark DC* apresentou uma variância de 170 segundos no tratamento RL503. No caso dos outros *benchmarks*, o *EP* apresentou uma variância de 18 segundos no tratamento RL302, o *SP* uma variância de aproximadamente 5 segundos também no tratamento RL302 e o *UA* uma variância de aproximadamente 6 segundos no tratamento RL502. Essa análise indica que alguns programas não são afetados por esses fatores e que programas que realizam operações de entrada e saída, como é o caso do *DC*, podem apresentar uma alta variabilidade em seus tempos de execução quando estão em *Runlevel* 5 e são otimizadas com configuração 03. A Figura 3.4 apresenta os gráficos, de cada *benchmark NPB*, gerados a partir das amostras coletadas para cada tratamento.

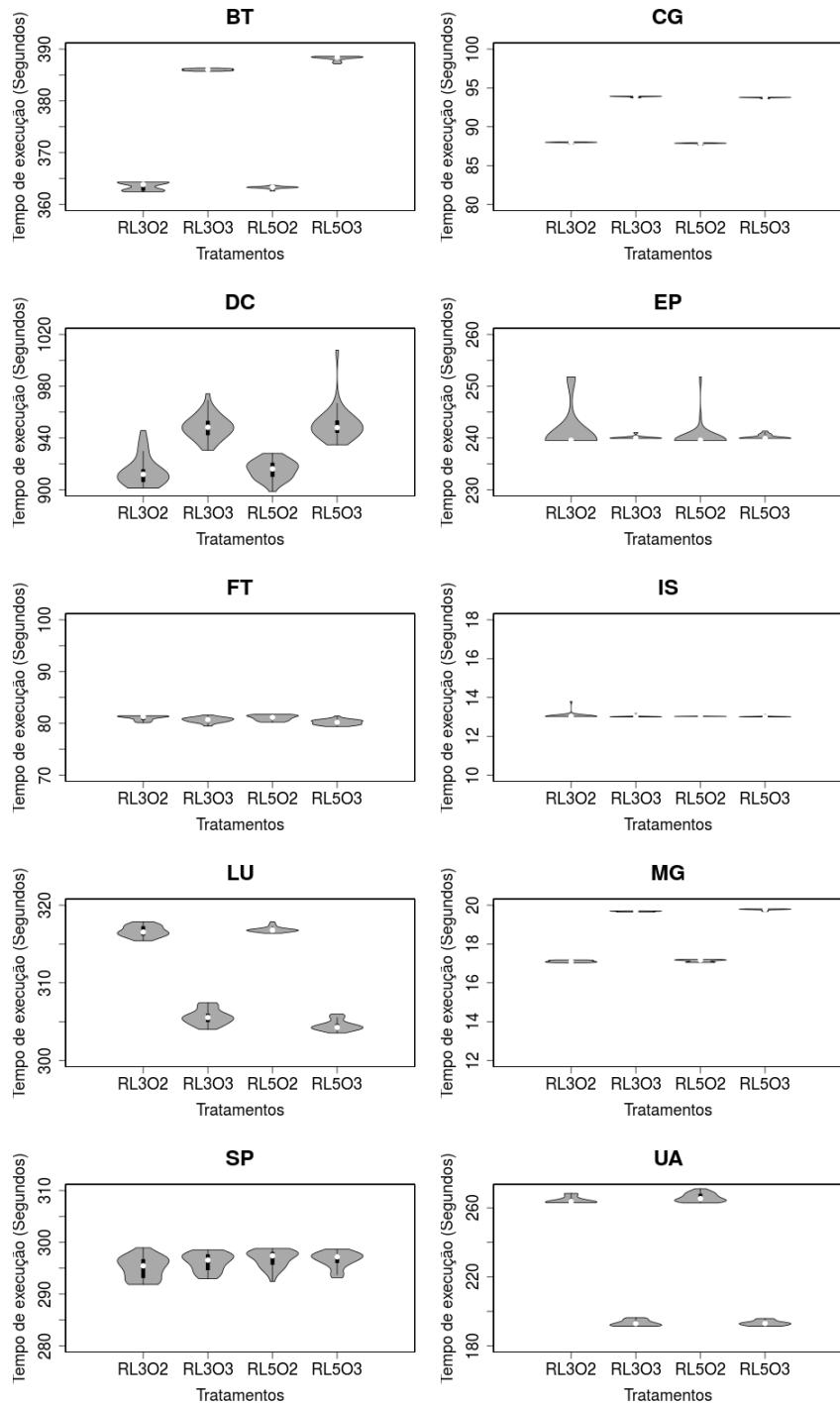
Ao analisar os tempos de execução dos programas disponibilizados pelo *PARSEC*, observou-se que em 41,67% dos casos os programas alcançaram o menor tempo de execução no tratamento RL303. Além disso, notou-se que em 75% dos casos os maiores tempos de execução foram obtidos quando o fator *Runlevel* estava em 5. Em relação à variância, apenas o programa *bodytrack* não apresentou variância em seus resultados. Ao avaliar os outros programas, observou-se que em 90,90% dos casos a menor variância foi obtida quando o fator *Runlevel* estava em 3. Assim, nesse caso, entende-se que os serviços que executam em *background* podem influenciar na variabilidade dos tempos de execução dos programas. A Tabela 3.8 sumariza o resultado referente ao maior e menor tempo para cada tratamento avaliado e a Tabela 3.9 sumariza o resultado referente a variância.

**Tabela 3.8. Distribuição do maior e menor tempo por tratamento para o PARSEC.**

	<b>RL302</b>	<b>RL303</b>	<b>RL502</b>	<b>RL503</b>
<b>Maior Tempo</b>	8,33%	16,67%	41,67%	33,33%
<b>Menor Tempo</b>	16,67%	41,67%	33,33%	8,33%

**Tabela 3.9. Distribuição da variabilidade por tratamento para o PARSEC.**

	<b>RL302</b>	<b>RL303</b>	<b>RL502</b>	<b>RL503</b>
<b>Maior Variabilidade</b>	18,18%	18,18%	27,27%	36,36%
<b>Menor Variabilidade</b>	54,54%	36,36%	18,18%	0,00%



**Figura 3.4. Variabilidade dos tempos de execução do NPB no Exp.#1.**

Ao realizar uma análise descritiva, foi observada uma alta variância nos tempos de execução dos programas *blackscholes* e *swaptions*, alcançando aproximadamente 231 e 395 segundos, respectivamente, no tratamento RL5O2. Além disso, foi observada também uma alta diferença entre o menor e o maior tempo desses programas para esse mesmo tratamento, aproximadamente 74 e 63 segundos. No caso dos programas *canneal* e *facesim* a maior variância foi registrada com, aproximadamente, 76 e 80 segundos, respectivamente, no tratamento RL5O3. O restante dos programas apresentou variâncias com valores menores. No caso do *bodytrack*, sua variância permaneceu em 0 segundos para todos os tratamentos, com uma diferença entre o menor e maior tempo em torno de 0,20 segundos. A Figura

3.5 apresenta os gráficos, de cada programa do *PARSEC*, gerados a partir das amostras coletadas para cada tratamento.

Ao analisar os tempos de execução dos programas do *benchmark SPLASH-2x*, identificou-se que os programas *ocean\_cp* e *radix* obtiveram tempos abaixo do que pode ser mensurado pelo programa *time*. Assim, a avaliação desse *benchmark* se concentra nos outros programas. Ao avaliar os tempos de execução dos programas restantes, observou-se que o maior tempo de execução dos programas *ocean\_ncp* e *volrend*, foi, respectivamente, o mesmo para todos os tratamentos e que o menor tempo de execução dos programas *cholesky* e *water\_nsquared*, foi, respectivamente, o mesmo para todos os tratamentos. Assim, nesse caso, a análise do maior e do menor tempo fica restrita aos programas restantes. Em relação ao maior e menor tempo, notou-se que, em 50% dos casos, o maior tempo foi obtido com o tratamento RL503 e, em 70% dos casos, o menor tempo foi alcançado com o tratamento RL502. A Tabela 3.10 sumariza o resultado para cada tratamento avaliado.

**Tabela 3.10. Distribuição do maior e menor tempo por tratamento para o *SPLASH-2x*.**

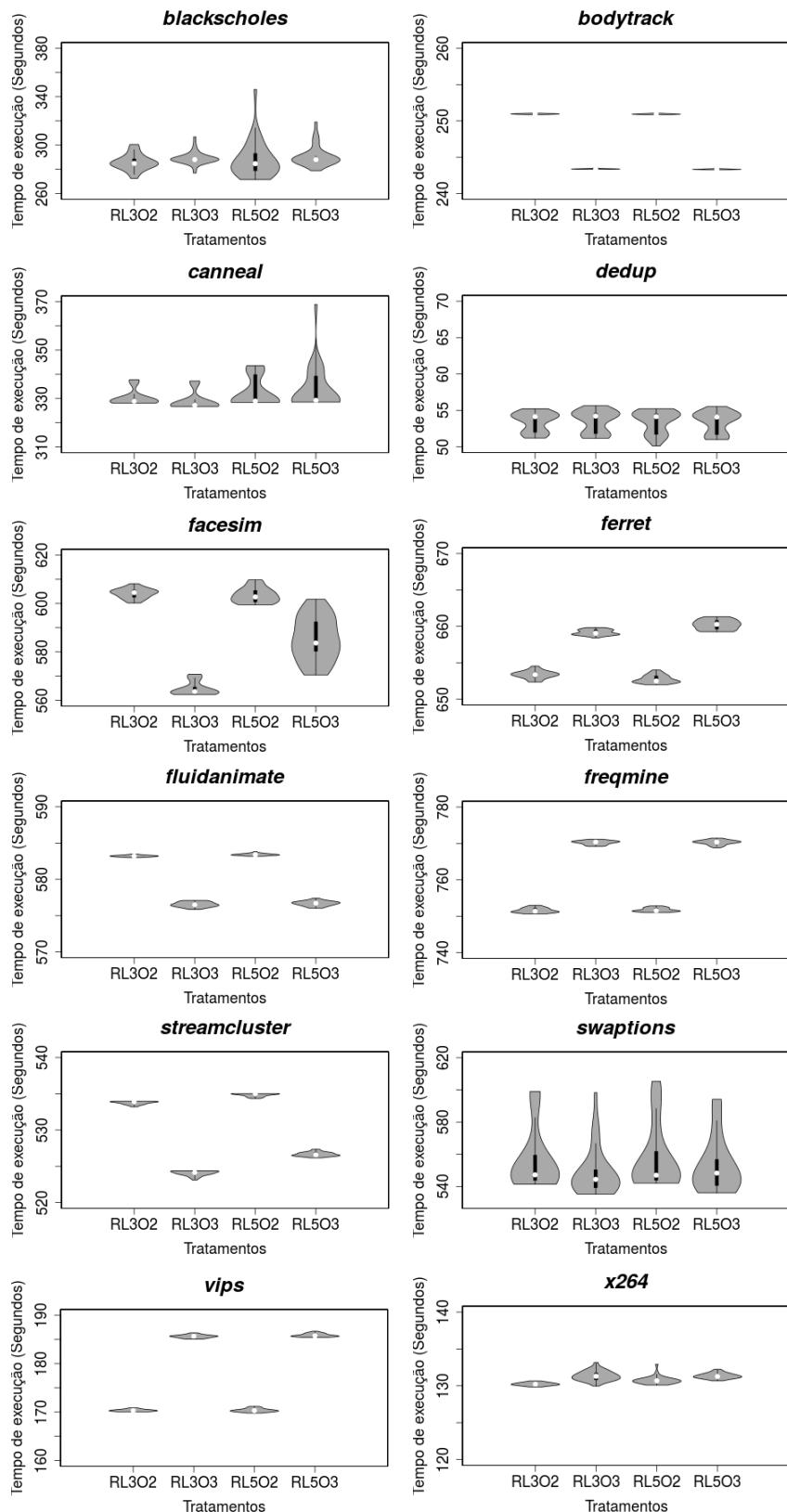
	RL302	RL303	RL502	RL503
<b>Maior Tempo</b>	40,00%	20,00%	40,00%	50,00%
<b>Menor Tempo</b>	10,00%	30,00%	70,00%	30,00%

Ao realizar uma análise descritiva, observou-se que a variância 58,33% dos programas não apresentou variância em seus tempos de execução. Dos programas restantes a variância não chegou a um segundo, com exceção do programa *fft* que obteve uma variância em torno de três segundos nos tratamentos RL302, RL303 e RL502. A Figura 3.6 apresenta os gráficos, de cada programa do *SPLASH-2x*, gerados a partir das amostras coletadas para cada tratamento.

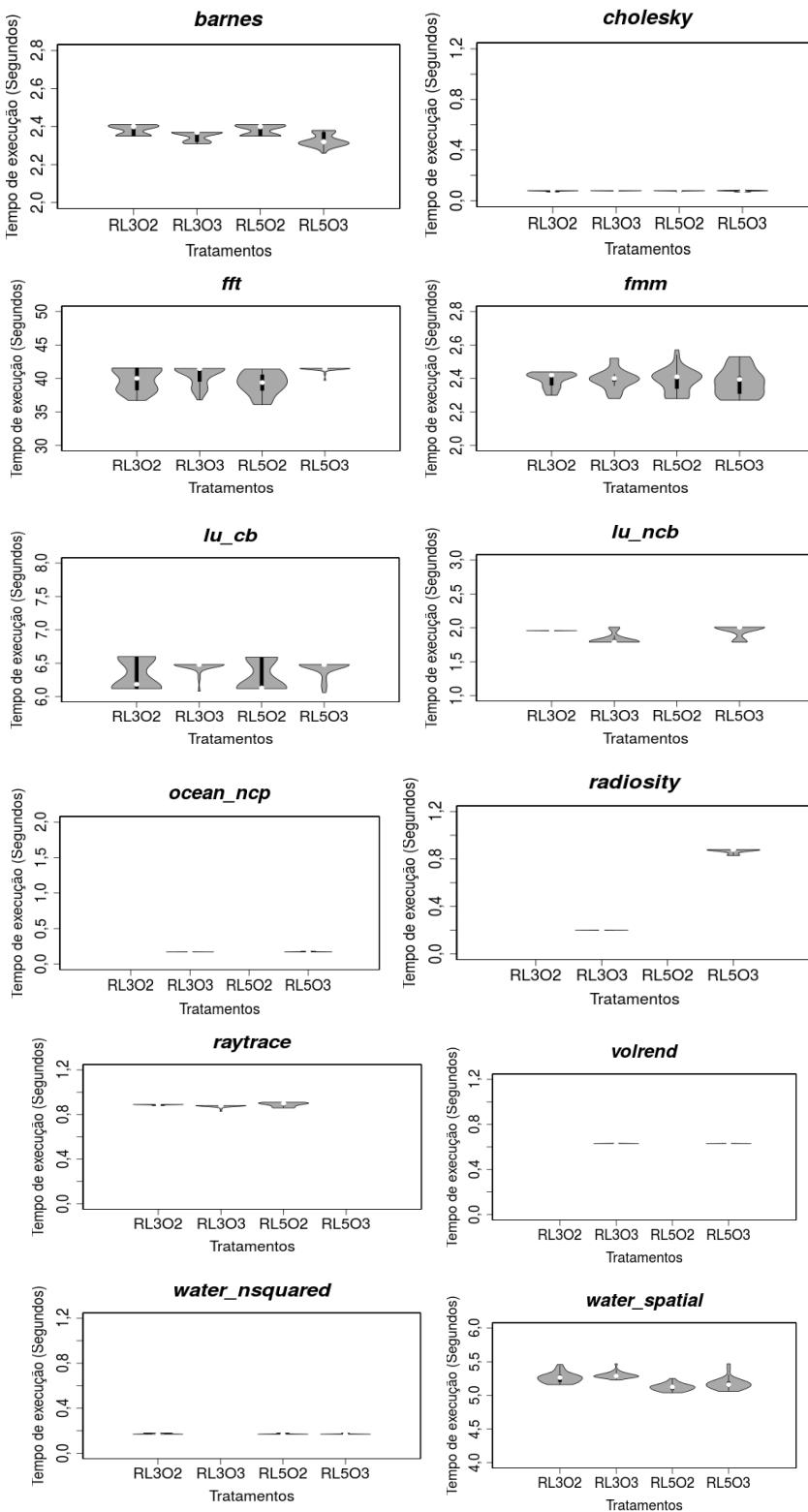
As observações realizadas neste experimento fornecem subsídios para indicar que esses fatores influenciam os tempos de execução dos programas de computador. Observou-se que programas do tipo *cpu-bound* apresentaram menor variância quando são executados em *runlevel* 3. Esse resultado já era esperado pois, uma vez que, em *runlevel* 3 o número de serviços do SO é menor, a concorrência pelo uso do processador também é menor, permitindo assim um melhor uso deste recurso. No caso de programas do tipo *memory bound*, observou-se uma baixa variação em seus tempos de execução, isso também já era esperado, pois, o computador no qual o experimento foi realizado oferece uma boa quantidade de memória e permite que as requisições por esse recurso sejam atendidas sem maiores penalidades. No caso de programas *i/o-bound*, observou-se uma alta variação nos tempos de execução quando o *runlevel* está em nível 5 e a otimização em nível 03. O fato do *runlevel* estar em nível 5 significa que há mais serviços do SO executando e que esses podem estar concorrendo por recursos de *i/o*.

Observou-se também, que programas com características comerciais, como o *blackscholes* e *bodytrack*, podem apresentar uma variação nos seus tempos de execução. Além disso, programas que possuem características HPC (*NPB* e *SPLASH-2x*) apresentaram melhor desempenho quando são otimizados em nível 02 e programas desenvolvidos para computadores com múltiplos processadores (*PARSEC*) alcançaram melhor desempenho quando o *runlevel* está em nível 3. Assim,

entende-se que controlar esses fatores ao realizar experimentos computacionais reduz a possibilidade de erros ao se analisar os tempos de execução dos programas.



**Figura 3.5. Variabilidade dos tempos de execução do PARSEC no Exp.#1.**



**Figura 3.6. Variabilidade dos tempos de execução do SPLASH-2x no Exp.#1.**

### 3.3.2 Experimento #2

As tabelas contendo a análise descritiva de cada *benchmark* para esse experimento estão disponibilizadas no Apêndice H.

Ao analisar os tempos de execução dos *benchmarks* disponibilizados pelo *NPB*, observou-se que os *benchmarks* *IS* e *MG* não apresentaram variância em seus tempos de execução quando o fator *Otimização* está em 03. Quando esse mesmo fator estava em 02, além desses dois *benchmarks* já citados, observou-se que o *benchmark* *CG* também não apresentou nenhuma variância em seus tempos de execução. No caso do *EP*, a variância dos seus tempos de execução estava em torno de 21 segundos, quando o fator *Otimização* estava em 02 e o tamanho da variável estava em 4096 *bytes*, porém, essa variância chegou a 0 quando o tamanho da variável estava em 128 *bytes*. Contudo, quando o fator *Otimização* estava em 03, a variância estava abaixo de 1 segundo para todos os tamanhos da variável.

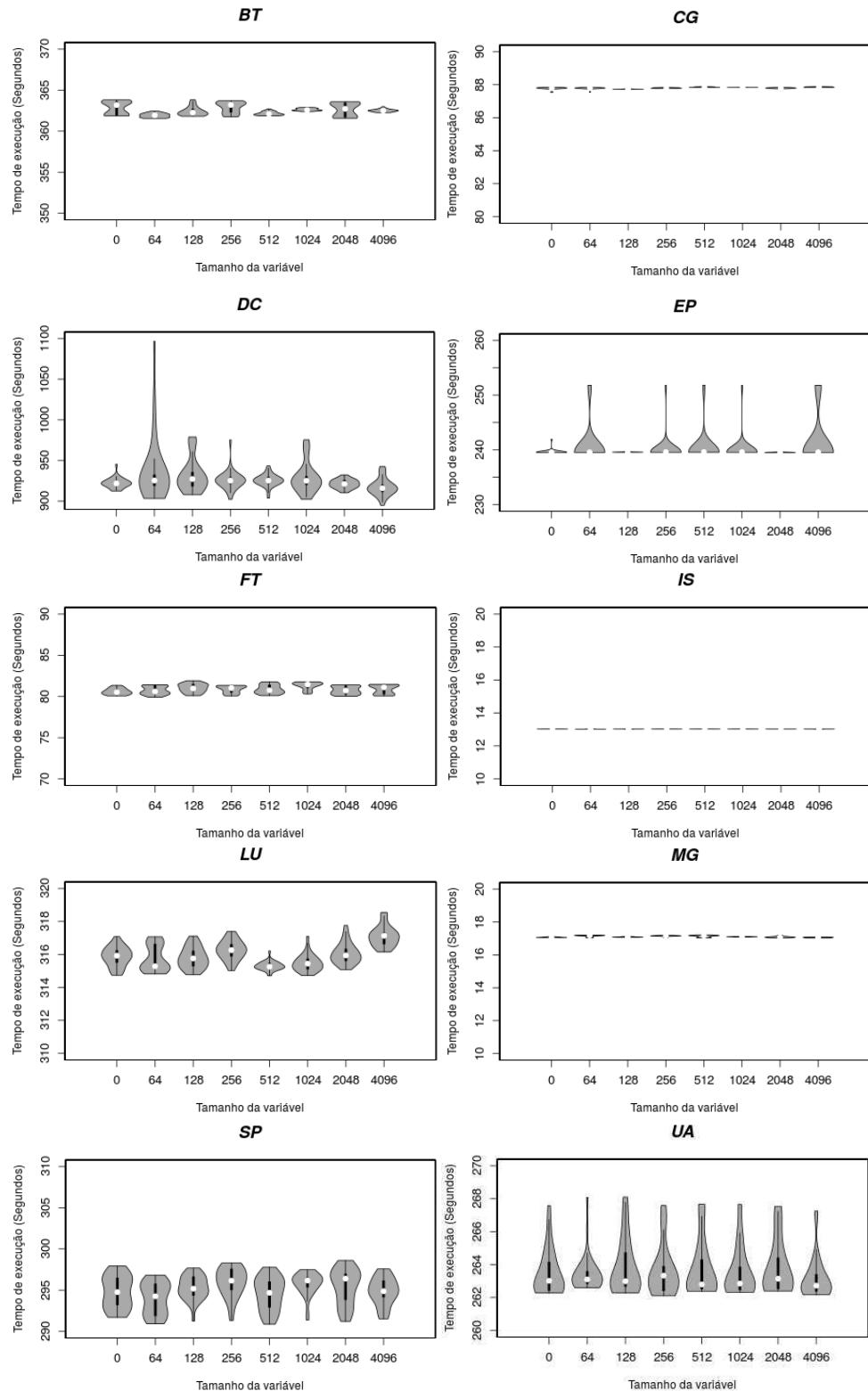
O *benchmark* *DC* foi o que apresentou a maior variância. Com o fator *Otimização* em 02 e o tamanho da variável em 1024 *bytes*, a variância de seus tempos de execução estava próxima a 990 segundos. Já quando o fator *Otimização* estava em 03 e o tamanho da variável estava em 64 *bytes*, a amostra dos tempos de execução desse programa apresentou uma variância em torno de 1622 segundos, todavia, essa variância ficou próxima de 20 segundos quando o tamanho da variável estava em 2048 *bytes*. Além disso, é importante ressaltar que, com exceção do tamanho 64 *bytes*, a variação dos tempos de execução desse *benchmark* reduziu-se consideravelmente, para todos os demais tamanhos de variável, quando este foi compilado com a optimização 02.

No caso dos *benchmarks* *SP* e *UA*, as suas variâncias não foram maiores do que 4 segundos para todos os tamanhos de variável e para os dois níveis do fator *Otimização*. Os outros *benchmarks* não apresentaram variância maior do que 1 segundo para todos os tamanhos de variável e para os dois níveis do fator *Otimização*. A Figura 3.7 apresenta os gráficos para cada *benchmark* do *NPB* compilados com o fator *Otimização* em 02 e a Figura 3.8 apresenta os gráficos, para esse mesmo *benchmark*, com o fator *Otimização* em 03.

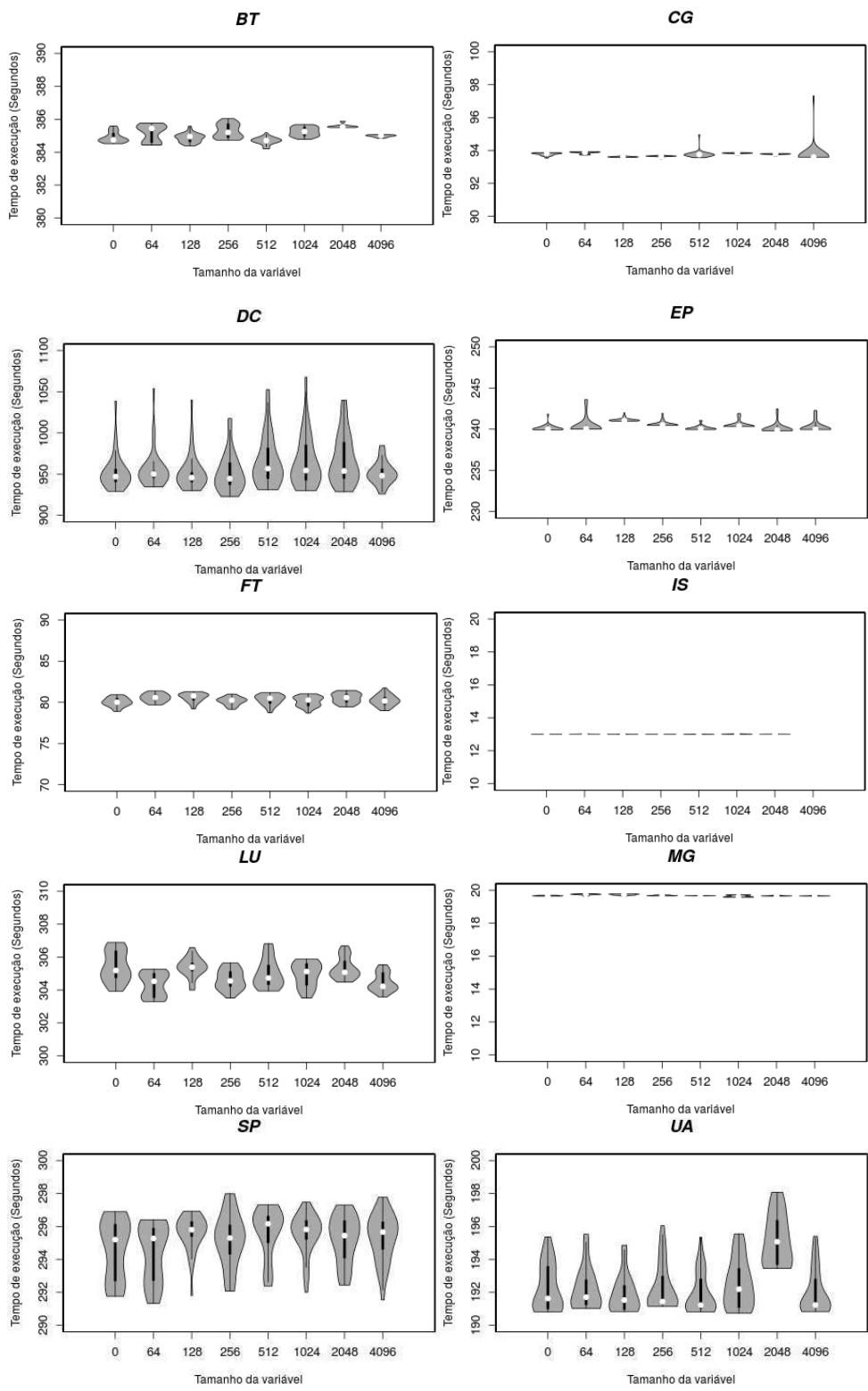
Ao considerar a mediana dos tempos de execução de cada *benchmark* (o ponto branco presente nos gráficos), observa-se que não há um tamanho da variável de ambiente que se destaca no que diz respeito a apresentar maior e menor tempo de execução. O que se observa em alguns casos (*BT*, *LU*, *SP* e *UA*), é que a mediana apresenta uma variação entre os tamanhos da variável de ambiente.

Os resultados obtidos para o *benchmark* *SPLASH-2x* são apresentados nos gráficos da Figura 3.9 (fator *Otimização* em 02) e Figura 3.10 (fator *Otimização* em 03). Ao analisar os tempos de execução dos programas disponibilizados pelo *SPLASH-2x*, observou-se que, para todos os tamanhos de variável e para ambos os níveis do fator *Otimização*, a variância dos tempos de execução dos programas não é superior a quatro segundos. Do mesmo modo que no *benchmark* anterior, é possível ver que, em alguns programas (*fft*, *lu\_cb* e *water\_spatial*), a mediana também apresentou variação entre os tamanhos da variável de ambiente.

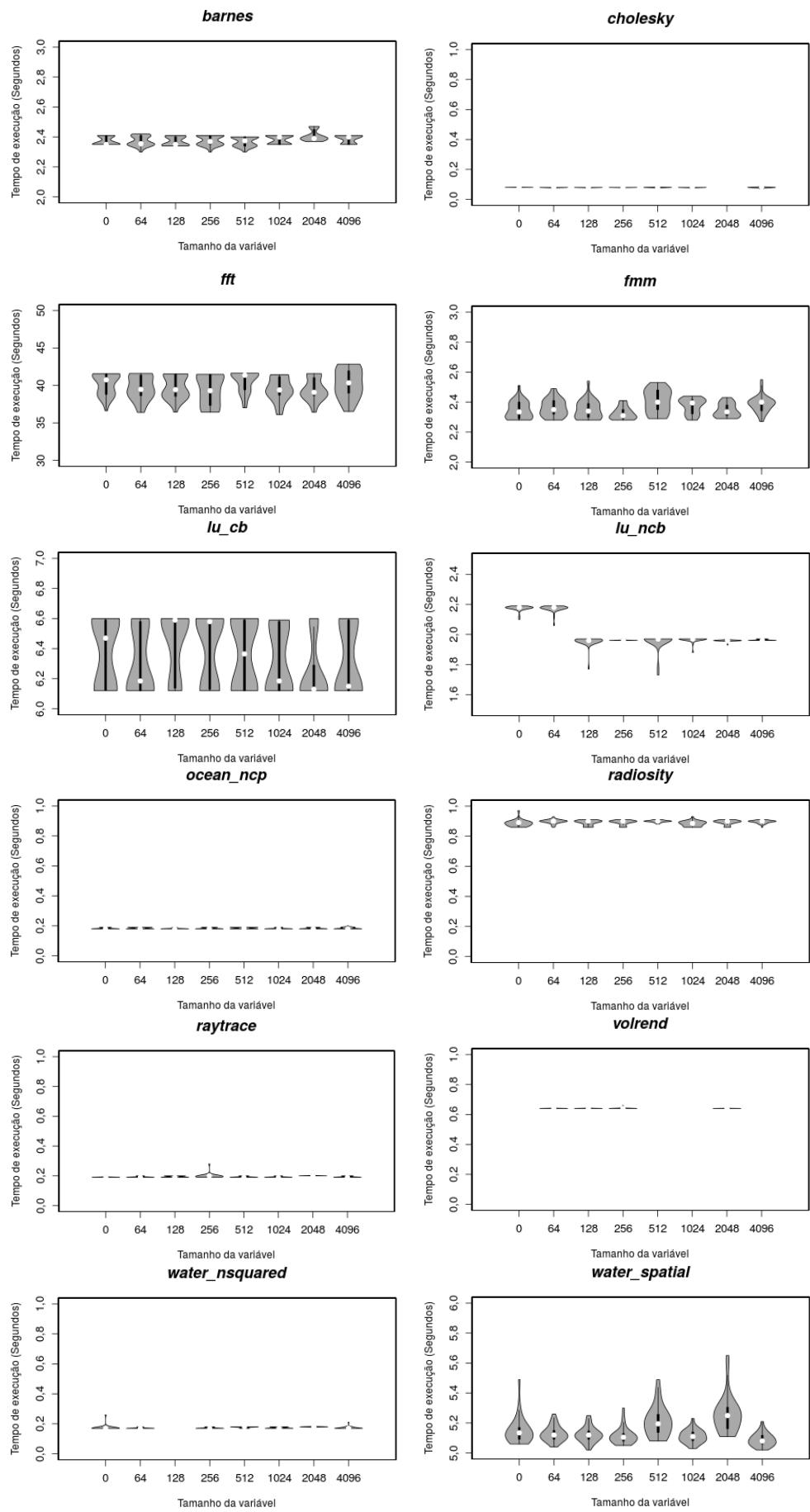
Assim, independentemente do tipo do programa (*cpu-bound*, *memory-bound* ou *i/o-bound*), as evidências observadas nesse experimento apontam que executar um programa em um mesmo computador, mas com diferentes tamanhos de uma ou mais variáveis de ambiente, pode acarretar em tempos de execução diferentes, comprometendo a análise desses tempos caso essa influência não seja tratada.



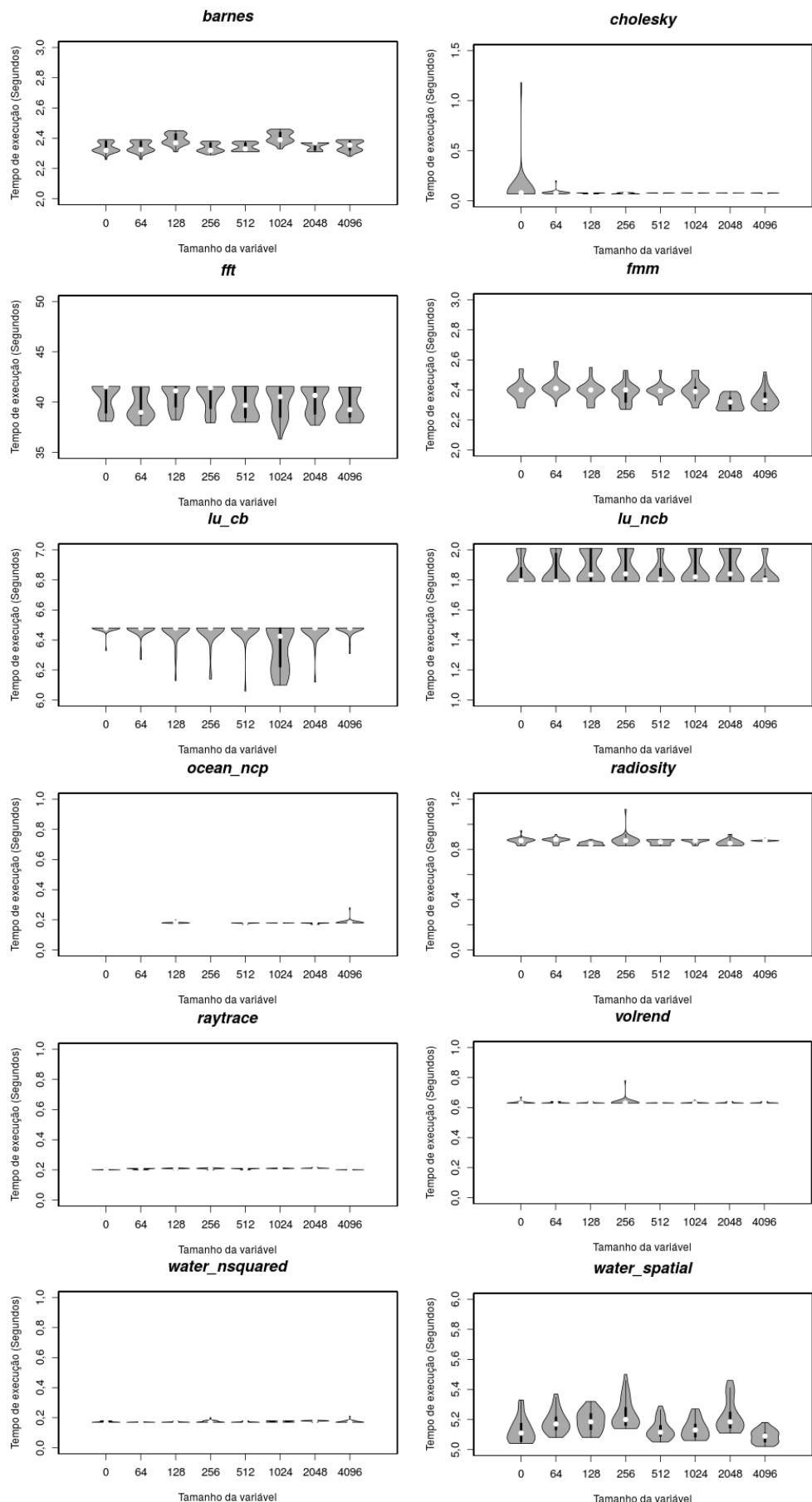
**Figura 3.7. Variabilidade dos tempos de execução do NPB com Otimização O2 no Exp.#2.**



**Figura 3.8. Variabilidade dos tempos de execução do NPB com *Otimização O3* no Exp.#2.**



**Figura 3.9. Variabilidade dos tempos de execução do SPLASH-2x com Optimização O2 no Exp.#2.**



**Figura 3.10. Variabilidade dos tempos de execução do SPLASH-2x com Otimização O3 no Exp.#2.**

### 3.3.3 Experimento #3

As tabelas contendo a análise descritiva de cada *benchmark* para esse experimento estão disponibilizadas no Apêndice I. No caso do *benchmark NPB*, ao analisar os tempos de execução dos programas, observou-se que permitir que o sistema operacional distribua as *threads* pelos núcleos (estratégia Sem *Affinity*) mostra ser a melhor opção, pois em 70% dos casos o menor tempo de execução foi encontrado com esta estratégia. Para o maior tempo encontrado, observa-se que a estratégia *Compact* se destacou em 90% dos casos. A Tabela 3.11 sumariza o resultado para cada estratégia de *thread affinity* avaliada.

**Tabela 3.11. Distribuição do maior e menor tempo por estratégia de *thread affinity* para o *NPB*.**

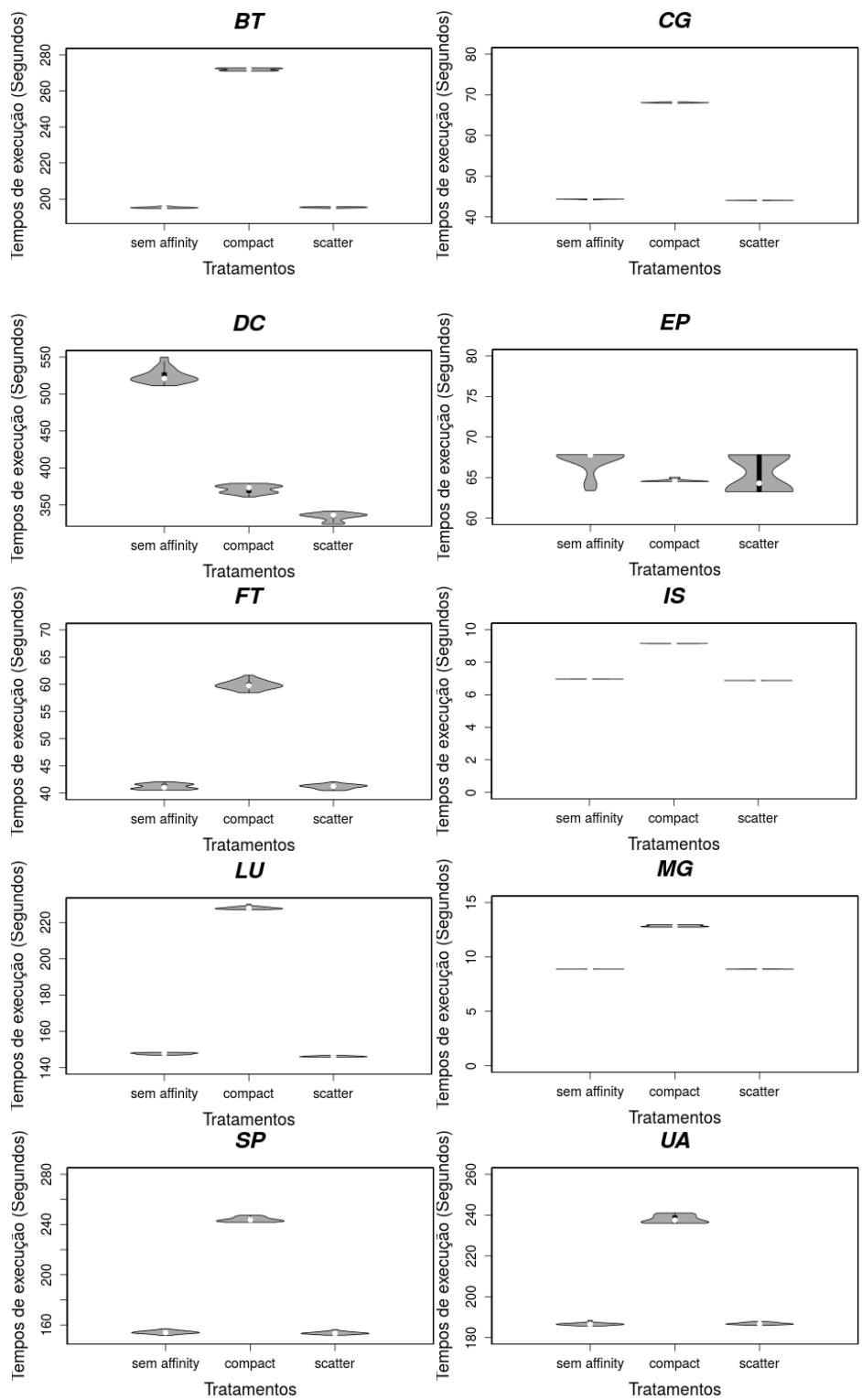
	sem affinity	compact	scatter
Maior Tempo	10,00%	90,00%	0,00%
Menor Tempo	70,00%	0,00%	33,33%

Além disso, observou-se, também, que, em relação à variabilidade, tanto a estratégia *scatter*, quanto a *compact* alcançaram baixa variabilidade em 40% dos casos. Contudo, a estratégia *compact* também apresentou alta variabilidade em 53,33% dos casos, em relação as outras estratégias. Assim, a estratégia *scatter* se mostra mais interessante quando se desejam resultados com menor variabilidade nos tempos de execução dos programas. A Tabela 3.12 sumariza o resultado para cada estratégia de *thread affinity* avaliada.

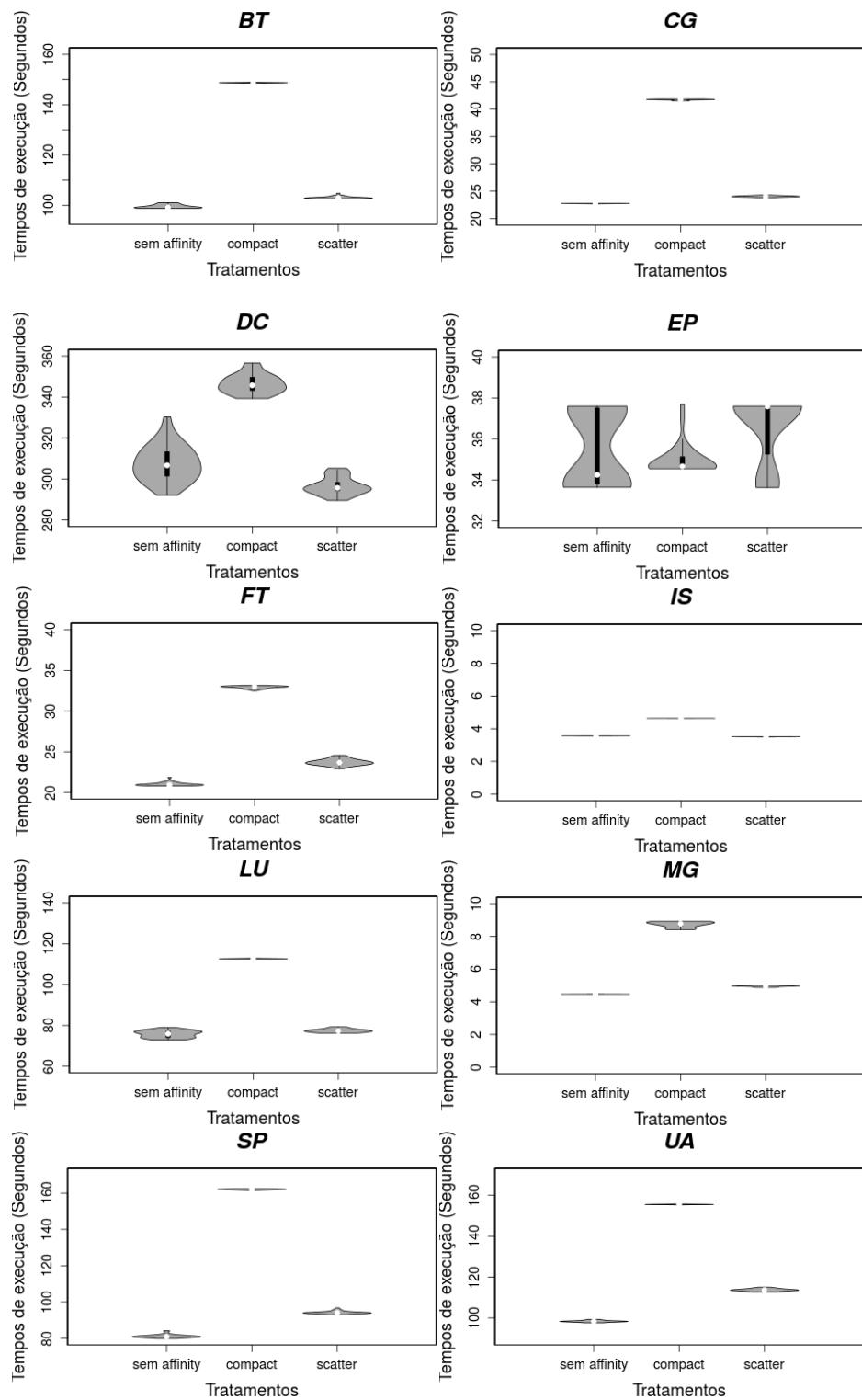
**Tabela 3.12. Distribuição da variabilidade por estratégia de *thread affinity* para o *NPB*.**

	sem affinity	compact	scatter
Maior Variabilidade	40,00%	53,33%	20,00%
Menor Variabilidade	36,67%	40,00%	40,00%

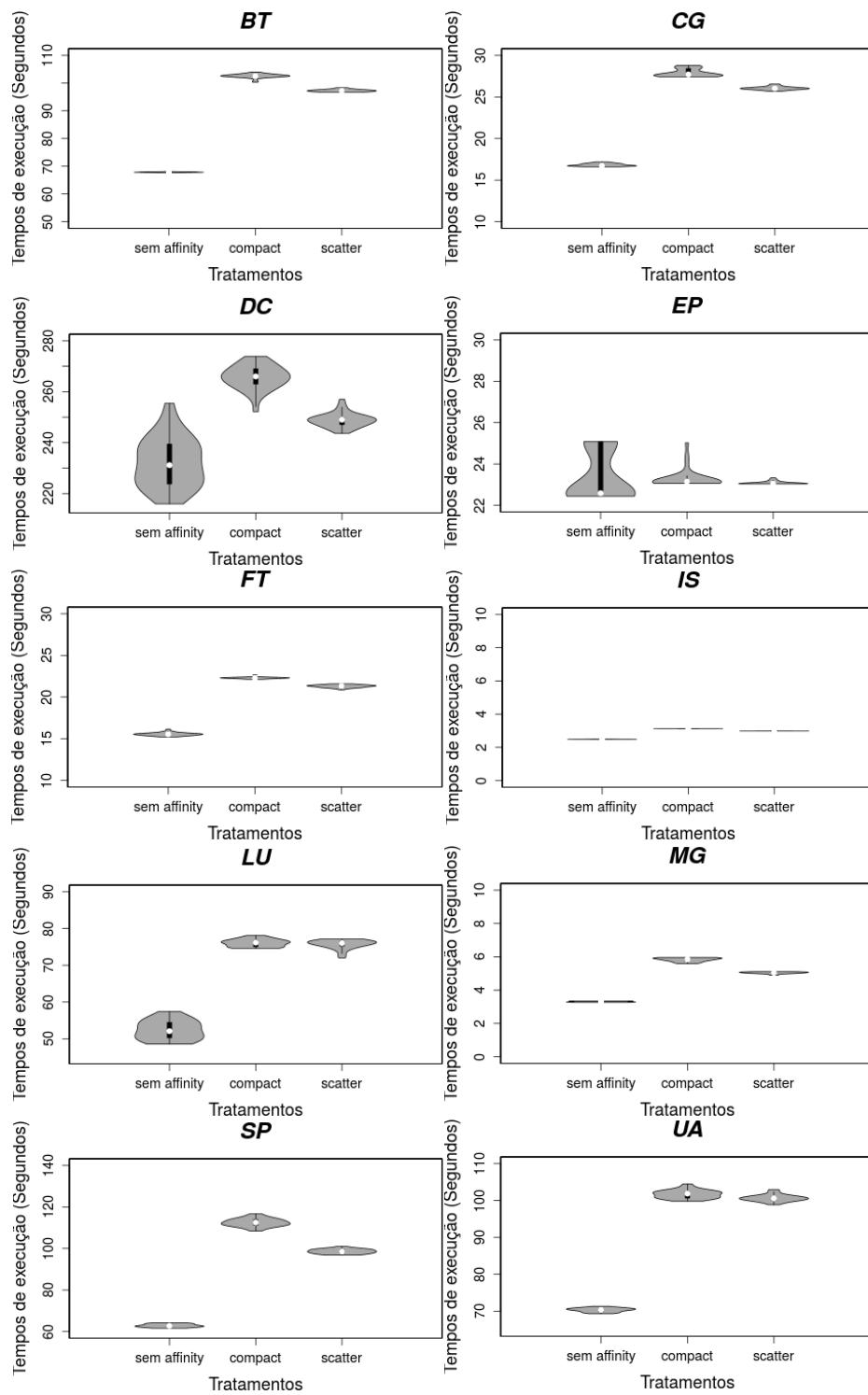
Ao analisar especificamente a estratégia *scatter*, observou-se que as execuções com duas threads apresentaram menor variabilidade, 60% dos casos em relação as execuções com quatro e seis *threads*. No entanto, o que chama a atenção, é que, em 40% dos casos, a menor variabilidade foi encontrada com a execução de seis *threads*. Esse resultado é maior que o encontrado para quatro *threads* (10%). As Figuras 3.11, 3.12 e 3.13 apresentam os gráficos obtidos para os *benchmarks NPB* com duas, quatro e seis *threads*, respectivamente.



**Figura 3.11. Variabilidade dos tempos de execução do NPB com 2 threads no Exp. #3.**



**Figura 3.12. Variabilidade dos tempos de execução do NPB com 4 threads no Exp. #3.**



**Figura 3.13. Variabilidade dos tempos de execução do *NPB* com 6 threads no Exp. #3.**

Ao analisar os tempos de execução dos programas do *benchmark PARSEC*, observou-se que permitir que o sistema operacional distribua as *threads* pelos núcleos (estratégia *sem affinity*) mostra ser a melhor opção, pois em 77,78% dos casos, o menor tempo de execução foi encontrado com essa estratégia. Para o maior tempo encontrado, observa-se que a um empate entre as estratégias *compact* e *scatter* que se destacam com 66,67% dos casos. A Tabela 3.13 sumariza o resultado para cada estratégia de *thread affinity* avaliada.

**Tabela 3.13. Distribuição do maior e menor tempo por estratégia de *thread affinity* para o PARSEC.**

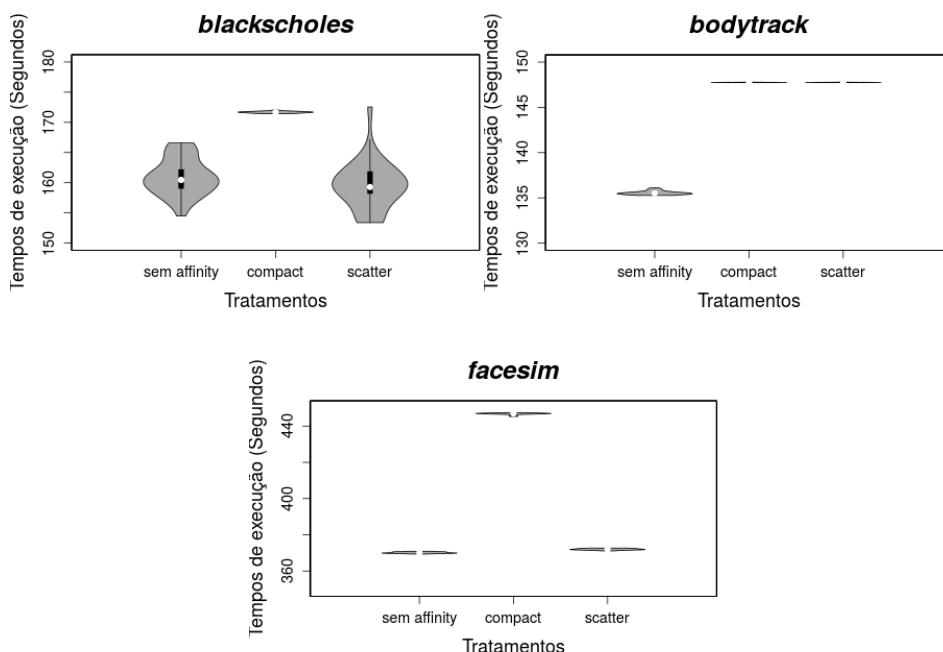
	<i>sem affinity</i>	<i>compact</i>	<i>scatter</i>
<b>Maior Tempo</b>	11,11%	66,67%	66,67%
<b>Menor Tempo</b>	77,78%	0,00%	28,57%

No caso dos programas do *PARSEC*, em relação à variabilidade estratégia *compact* alcançou uma baixa variabilidade em 77,78% dos casos. A maior variabilidade foi encontrada quando os programas executaram com estratégia sem *affinity* e *scatter*, 44,44% dos casos. Desse modo, apesar de alcançar tempos maiores de execução, a estratégia *compact* se mostra mais interessante quando se desejam resultados com menor variabilidade nos tempos de execução dos programas. A Tabela 3.14 sumariza o resultado para cada estratégia de *thread affinity* avaliada.

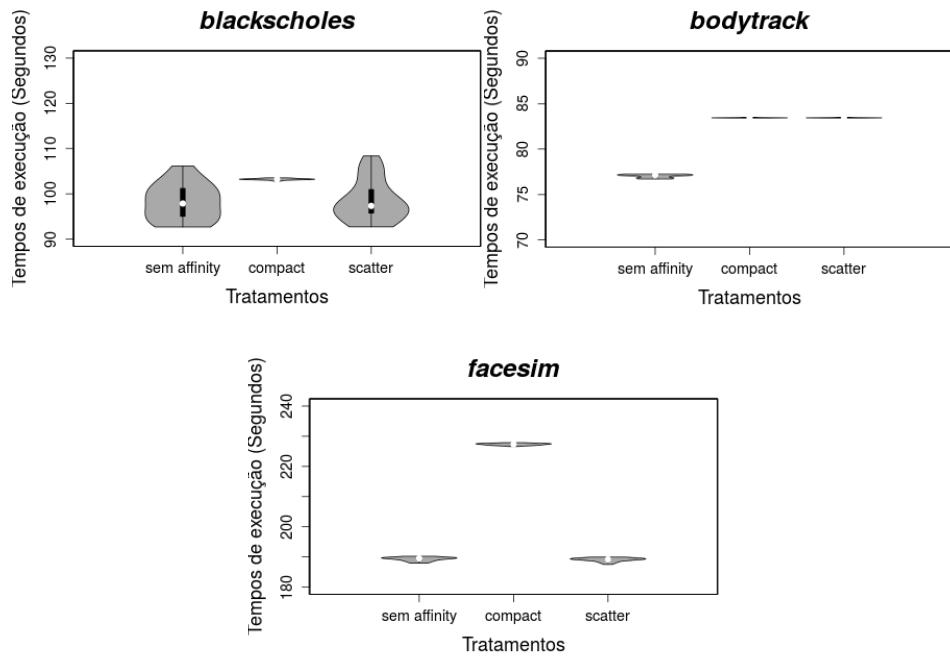
**Tabela 3.14. Distribuição da variabilidade por estratégia de *thread affinity* do PARSEC.**

	<i>sem affinity</i>	<i>compact</i>	<i>scatter</i>
<b>Maior Variabilidade</b>	44,44%	11,11%	44,44%
<b>Menor Variabilidade</b>	22,22%	77,78%	33,33%

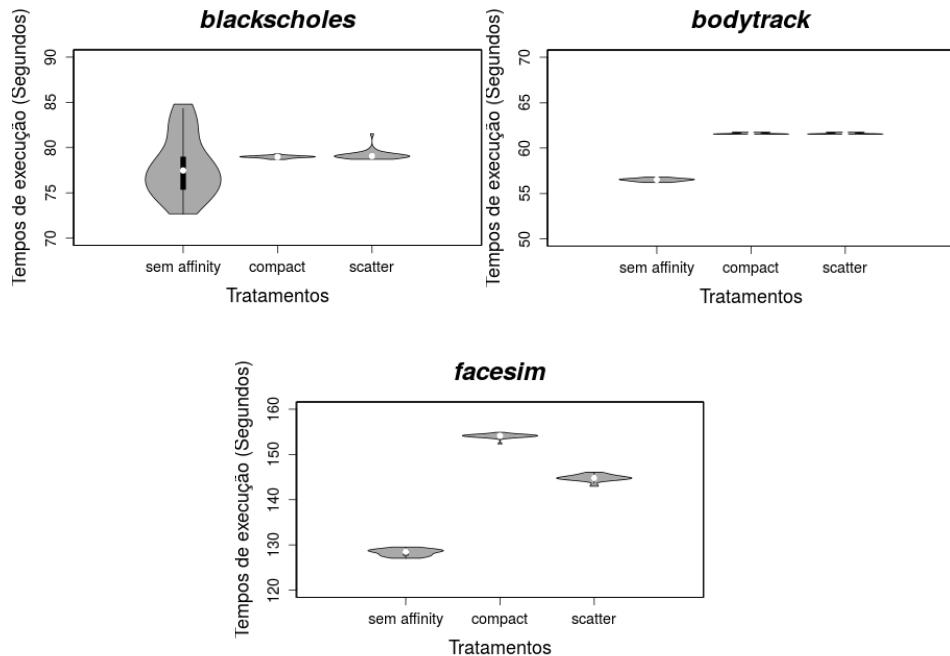
As Figuras 3.14, 3.15 e 3.16 apresentam os gráficos obtidos para os programas *blackscholes*, *bodytrack* e *facesim* do benchmark *PARSEC* com duas, quatro e seis *threads*, respectivamente.



**Figura 3.14. Variabilidade dos tempos de execução do PARSEC com 2 threads no Exp. #3.**



**Figura 3.15. Variabilidade dos tempos de execução do PARSEC com 4 threads no Exp. #3.**



**Figura 3.16. Variabilidade dos tempos de execução do PARSEC com 6 threads no Exp. #3.**

### 3.3.4 Experimento #4

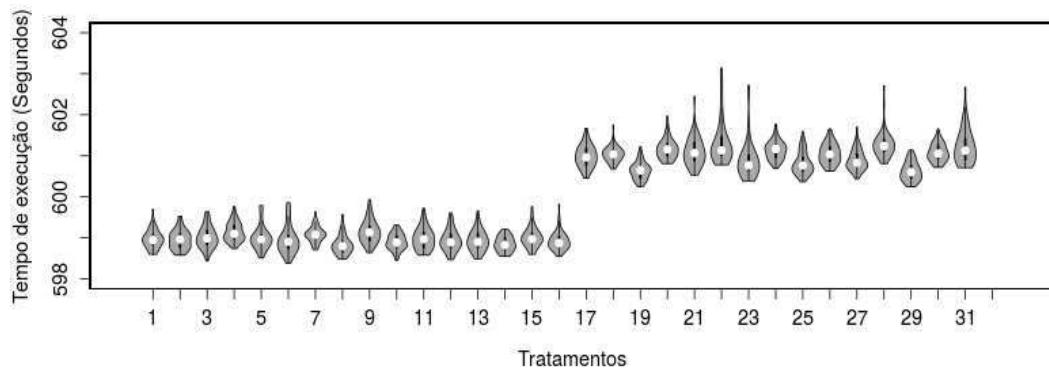
As tabelas contendo a análise descritiva do programa aqui utilizado estão disponibilizadas no Apêndice J.

Na Seção 3.2.5 está descrito que esse experimento está dividido em duas partes, sendo assim, a análise de seus resultados também segue essa divisão. Ao analisar os tempos de execução, observou-se que, a partir do 16º tratamento, o

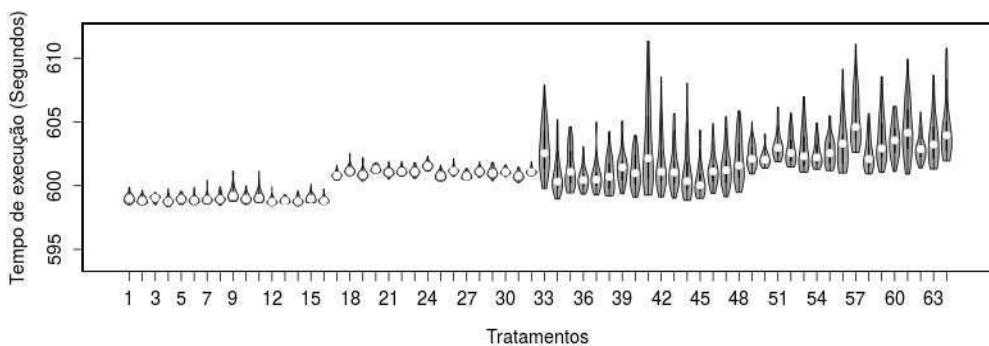
tempo de execução do *programa prova* teve uma alta em relação aos tempos de execução dos dezesseis primeiros. Esse aumento pode ser atribuído ao fator *Timer Interrupt Handling* que está presente nos últimos dezesseis tratamentos. Observou-se também que a variância dos tempos de execução do *programa prova* está abaixo de um segundo para todos os tratamentos. Ao avaliar o gráfico nota-se que, a diferença entre o maior e o menor tempo para os tratamentos 21, 22, 23, 28, 31 e 32 está em torno de dois segundos, ao passo que, para os outros tratamentos, essa diferença está próxima de um segundo. A Figura 3.17 apresenta o gráfico obtido com os dados dos 32 tratamentos iniciais.

Se for considerada a mediana dos tratamentos, é possível notar, no gráfico, que, nos primeiros dezesseis tratamentos, a variação entre os tratamentos praticamente se manteve inalterada, ao contrário do que se observa quando se analisam os dezesseis últimos.

A Figura 3.18 ilustra o gráfico da segunda parte desse experimento, contendo os 64 tratamentos obtidos da configuração dos seis fatores com os seus dois níveis de operação. Os primeiros 32 tratamentos são referentes à primeira parte do experimento, em que não há a presença do fator *Shared Cache L3*. Nos últimos 32 tratamentos, há a presença desse fator e, do tratamento 33 ao 48, ele está em nível (-) e do 49 ao 64 ele está em nível (+).



**Figura 3.17. Variabilidade dos tempos de execução do prog. prova para os 32 tratamentos iniciais do Exp. #4.**



**Figura 3.18. Variabilidade dos tempos de execução do prog. prova para os 64 tratamentos do Exp. #4.**

Ao observar o gráfico, é possível identificar que, a partir do tratamento 33, há um aumento na variabilidade dos tempos de execução dos programas. Essa variabilidade está acima dos cinco segundos na maioria dos tratamentos. Além

disso, se for considerado a mediana dos tempos de execução dos programas, em cada tratamento, pode-se notar que há uma variação entre os tratamentos. Ao realizar uma análise descritiva, foi encontrado que a maioria dos tratamentos apresentaram uma variância acima dos 2 segundos e no caso do tratamento 41, essa variância chegou em torno de 11 segundos.

As observações realizadas nesse experimento dão subsídios para concluir que a presença de processos executando em *background* pode contribuir para o aumento da variabilidade nos tempos de execução dos programas. Assim sendo, essa informação confirma a necessidade de se utilizar equipamentos dedicados durante a realização de experimentos computacionais.

### 3.4 Considerações Finais

Neste capítulo, foi apresentada uma caracterização dos tempos de execução de programas de computador para demonstrar como alguns fatores podem influenciar nos tempos de execução dos programas.

Os resultados obtidos confirmam a variabilidade dos tempos de execução dos programas devido à presença de fatores inerentes ao ambiente computacional. Caso esses fatores não sejam devidamente controlados, eles podem influenciar nas conclusões dificultando a reproduzibilidade dos experimentos.

Com esses resultados, não é possível afirmar se há diferença entre os tratamentos replicados no experimento, pois a observação de apenas um valor não reflete a realidade. No próximo capítulo é abordado o protocolo *Speedup-Test* (TOUATI; WORMS; BRIAIS, 2013), que foi proposto justamente para auxiliar na comparação dos tratamentos.

# 4. O PROTOCOLO SPEEDUP-TEST

## 4.1 Introdução

A caracterização dos tempos de execução dos programas de computador é importante para indicar o comportamento dos tempos obtidos. No entanto, muitas vezes, o que se deseja é justamente comparar os tratamentos definidos no experimento. Em seu trabalho, Touati; Worms; Briaïs (2013) apresentam um protocolo para comparar tratamentos voltados para a otimização de programas. Segundo os autores, a simples observação dos tempos de execução de dois programas não são suficientes para indicar se a otimização aplicada foi eficaz ou não. A variabilidade dos tempos de execução dos programas pode influenciar as observações, levando a conclusões incorretas. Assim, os autores destacam a importância dos métodos estatísticos durante o processo de análise dos resultados.

Neste capítulo, é apresentado o protocolo *Speedup-Test* (TOUATI; WORMS; BRIAIS, 2013) utilizado para a comparação de duas amostras de tempos de execução de programas de computador.

## 4.2 Protocolo *Speedup-Test*

Dentro da área de otimização de algoritmos,  $P$  é o algoritmo original de um programa e  $P'$  sua versão otimizada. Com uma entrada  $I$ , para ambos os algoritmos, é possível obter um conjunto de amostras distintas dos tempos de suas execuções.

Desse modo, sendo  $X$  a variável aleatória representando os tempos de execução de  $P(I)$  e  $Y$  a variável aleatória representando os tempos de execução de  $P'(I)$ , é possível obter respectivamente  $n$  tempos de execução ( $X = \{x_1, \dots, x_n\}$ ) e  $m$  tempos de execução ( $Y = \{y_1, \dots, y_m\}$ ) (TOUATI; WORMS; BRIAIS, 2013). A média teórica das variáveis aleatórias  $X$  e  $Y$  pode ser definida, como sendo  $\mu_X$  e  $\mu_Y$ , respectivamente; a mediana teórica de  $X$  e  $Y$  como  $\theta_X$  e  $\theta_Y$ ; as variâncias teóricas por  $\sigma_X^2$  e  $\sigma_Y^2$ ; e as médias amostrais de  $X$  e  $Y$ , pelas Equações 4.1 e 4.2:

$$\bar{X} = 1/n \sum_{i=1}^n x_i \quad (4.1)$$

$$\bar{Y} = 1/m \sum_{i=1}^m y_i \quad (4.2)$$

em que,  $x_i$  é o  $i$ -ésimo elemento pertencente a  $X$  e  $y_i$  o  $i$ -ésimo elemento de  $Y$ . Além disso, é possível definir as variâncias amostrais como  $s_X^2$  e  $s_Y^2$ ; e as medianas amostrais de  $X$  e  $Y$  por  $\overline{\text{med}}(X)$  e  $\overline{\text{med}}(Y)$ . Com essas amostras, o *speedup* de um programa pode ser calculado (ver Equação 4.3) (HENNESSY; PATTERSON, 2011).

$$\text{speedup} = \frac{\text{tempo de execução do programa}}{\text{tempo de execução do programa otimizado}} \quad (4.3)$$

Além disso, é possível obter também o *speedup* para o menor tempo de execução (*spmin*), assim como para a média (*spmean*) e para a mediana dos tempos de execução (*spmedian*) de uma determinada amostra, conforme apresentado pelas Equações 4.4, 4.5 e 4.6 (TOUATI; WORMS; BRIAIS, 2013).

$$spmin(P, I) = \frac{\min_i x_i}{\min_j y_j} \quad (4.4)$$

$$spmean(P, I) = \frac{\bar{X}}{\bar{Y}} = \frac{\sum_{1 \leq i \leq n} x_i}{\sum_{1 \leq j \leq n} y_j} \times \frac{m}{n} \quad (4.5)$$

$$spmedian(P, I) = \frac{\overline{\text{med}}(X)}{\overline{\text{med}}(Y)} \quad (4.6)$$

Sendo que,  $\min_i x_i$  é o elemento com o menor tempo de execução dentro da amostra  $X$  e  $\min_j y_j$  o elemento com o menor tempo de execução dentro da amostra  $Y$ . O protocolo *Speedup-Test* proposto por (TOUATI; WORMS; BRIAIS, 2013) descreve uma metodologia estatística para avaliar a significância do *speedup* obtido por meio de duas métricas, a saber: média e mediana. Segundo os mesmos autores, a média é bem conhecida na análise estatística, no entanto, ela é fortemente influenciada por dados discrepantes (*outliers*), ao passo que a mediana é menos sensível aos *outliers* e uma das mais utilizadas pelos padrões *SPEC* (SPEC, 2013d).

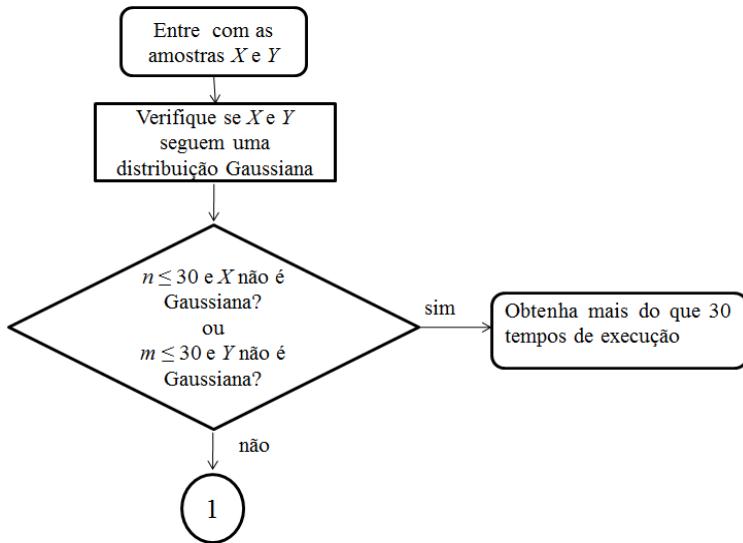
#### 4.2.1 Analisando a média dos tempos de execução

Sendo  $X$  e  $Y$  duas amostras dos tempos de execução de duas versões de um mesmo programa, com a mesma entrada  $I$  e um nível de confiança ( $1-\alpha$ ) (no caso, o protocolo adota o  $\alpha$  de 5%), é possível determinar por meio de um teste de hipótese, se a média teórica de  $X$  ( $\mu_X$ ) é maior que a média teórica de  $Y$  ( $\mu_Y$ ).

$$H_0: \mu_X \leq \mu_Y; \text{ e } H_1: \mu_X > \mu_Y.$$

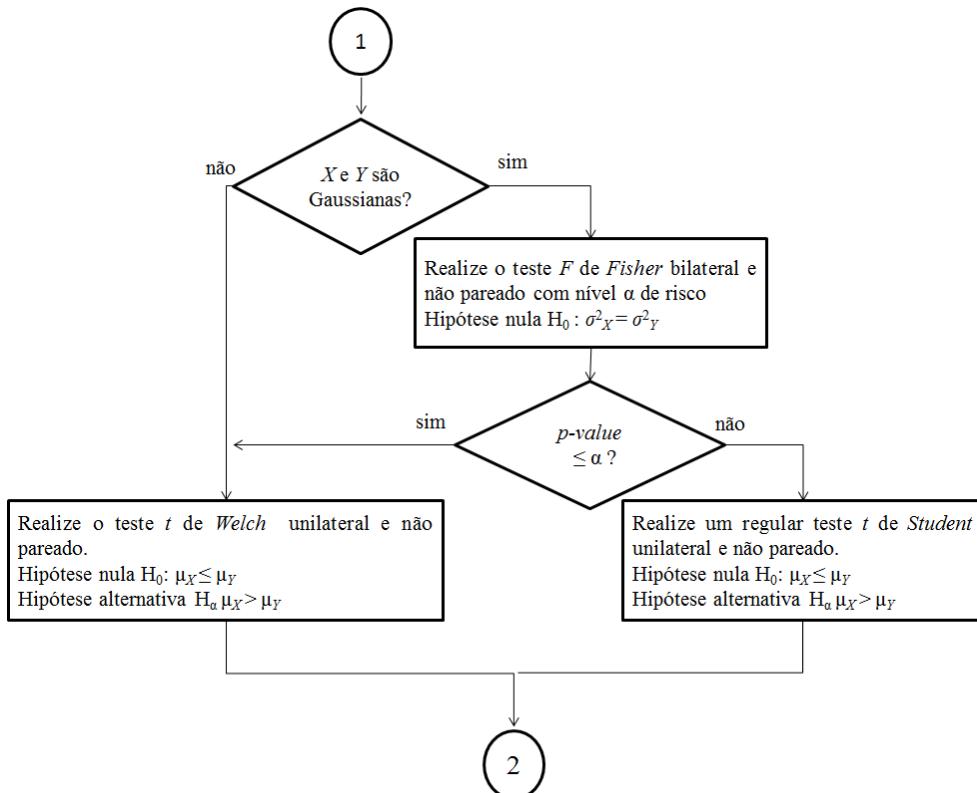
O protocolo então, propõe o uso do teste  $t$  de *Student* (BARBETTA; REIS; BORNIA, 2010). No entanto, esse teste exige que duas premissas sejam atendidas: (I) os dados devem seguir uma distribuição Gaussiana e (II) as amostras devem possuir a mesma variância. Por isso, o protocolo adota o teste de *Shapiro-Wilk* (SHAPIRO; WILK, 1965) como teste de aderência e o teste  $F$  de *Fisher* para verificar se as amostras possuem a mesma variância (ver Figura 4.1). Ao obter as amostras é necessário aplicar o teste de aderência a uma distribuição Gaussiana (teste de *Shapiro-Wilk*), e avaliar se o *p-value* está dentro do nível de confiança ( $1-\alpha$ ) preestabelecido, o que, no caso, indica uma distribuição Gaussiana. Além disso, é preciso observar se há mais do que 30 tempos de execução em cada amostra. Isso, porque, caso o teste não indique uma aderência dos dados a uma distribuição Gaussiana, ainda é possível analisá-las, uma vez que o Teorema do Limite Central (ROSS, 2014) demonstra que, para amostras grandes, as distribuições das médias amostrais se aproxima de uma distribuição Gaussiana.

Contudo, se os dados não seguem uma distribuição Gaussiana, ou mesmo se as amostras não possuem a mesma variância, ainda assim é possível avaliar a significância do *speedup* por meio do teste  $t$  de *Welch* (HOWELL, 2009). Esse teste é uma variação do teste  $t$  de *Student* (ver Figura 4.2).

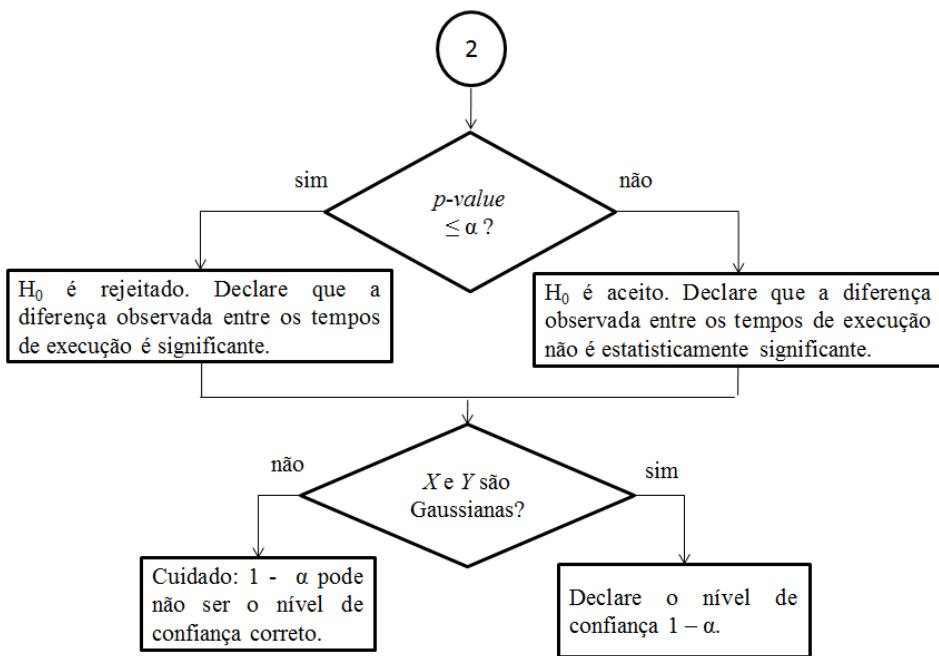


**Figura 4.1. Passos do protocolo para verificar a aderência dos dados a uma distribuição Gaussiana (adaptado de (TOUATI; WORMS; BRIAIS, 2013)).**

De acordo com os autores, o resultado encontrado pode não ser correto para o nível de confiança especificado, porque, para grandes amostras, o Teorema do Limite Central indica uma aproximação à distribuição Gaussiana e o cálculo da estatística pode não ser preciso. Por fim, ao se obter o resultado do *p-value* ou do teste *t* de *Student*, ou do teste *t* de *Welch*, pode-se determinar, ou não, a significância do *speedup* (ver Figura 4.3). O Anexo H ilustra as etapas do protocolo para a média.



**Figura 4.2. Passos do protocolo para analisar a significância (adaptado de (TOUATI; WORMS; BRIAIS, 2013)).**



**Figura 4.3. Passos do protocolo para apresentar o resultado da análise (adaptado de (TOUATI; WORMS; BRIAIS, 2013)).**

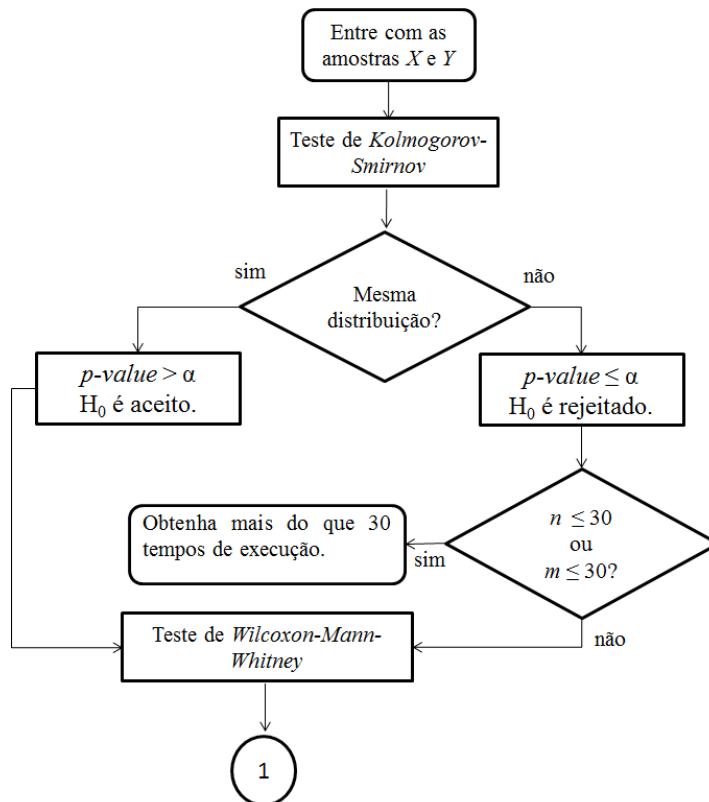
#### 4.2.2 Analisando a mediana dos tempos de execução

Como o teste  $t$  de *Student* analisa a média das amostras de tempos, é necessário, então, o uso de outro teste para avaliar a significância da diferença entre amostras. Assim sendo, o protocolo utiliza o teste de *Wilcoxon-Mann-Whitney* (SHESKIN, 2003). Esse teste não exige que os dados sejam provenientes de uma distribuição Gaussiana, ou que as amostras possuam a mesma variância.

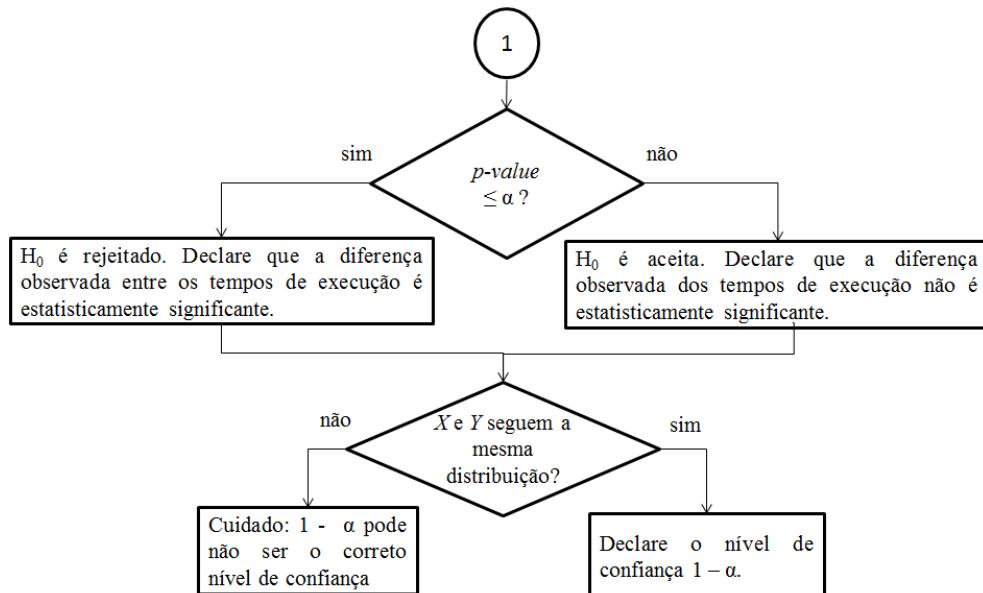
O teste de *Wilcoxon-Mann-Whitney* possui quatro premissas: (I) cada amostra foi selecionada, aleatoriamente, a partir da população que representa; (II) as amostras são independentes umas das outras; (III) a variável subsequentemente ordenada é uma variável aleatória contínua; e (IV) as amostras são provenientes de uma mesma distribuição. As três primeiras premissas (I, II e III) são satisfeitas ao planejar o experimento. Para a última (IV), o correto é utilizar um teste estatístico que avalie a sua veracidade.

Portanto, segundo a metodologia descrita em Touati; Worms; Briais (2013), o teste de *Kolmogorov-Smirnov* (K-S) (GIBBONS; CHAKRABORTI, 2014) deve ser aplicado a fim de verificar essa suposição de mesma distribuição (ver Figura 4.4).

Se o teste K-S apresentar que as amostras não seguem a mesma distribuição, ainda assim é possível realizar o teste de *Wilcoxon-Mann-Whitney* para avaliar a significância dos *speedups*, no entanto, é importante ressaltar que o resultado pode não ser correto para o nível de confiança de 95% (valor adotado pelo protocolo). A Figura 4.5 ilustra o fluxo para a apresentação do resultado obtido. O Anexo I apresenta uma visão completa do protocolo para a mediana.



**Figura 4.4. Passos do protocolo para analisar a significância (adaptado de TOUATI; WORMS; BRIAIS, 2013).**



**Figura 4.5. Passos do protocolo para apresentar o resultado da análise (adaptado de TOUATI; WORMS; BRIAIS, 2013).**

### 4.3 Análise de Resultados

Nesta seção, são apresentadas as análises de significância dos tempos de execução de programas, obtidos nos experimentos #1, #2, #3 e #4 (ver Seção 3.3). Nesta análise, aplicou-se o protocolo *Speedup-Test* (TOUATI; WORMS; BRIAIS, 2013) com o intuito de comparar os tratamentos e observar se há diferença

estatisticamente significativa entre eles. Isso significa que as amostras de tempos de execução de todos os programas, dos três *benchmarks*, foram comparadas em cada tratamento estudado. Assim, o teste de hipótese pode ser dado por:

$$H_0: \mu_X = \mu_Y; \text{ e } H_1: \mu_X \neq \mu_Y.$$

Os programas *ocean\_cp* e *radix* disponibilizados pelo *SPLASH-2x* apresentaram tempos abaixo de 0,01 segundos, portanto, não foi possível mensurá-los com o programa *time* e, por isso, eles não foram analisados. Para todos os testes estatísticos adotou-se o nível de confiança de 95%.

#### 4.3.1 Experimento #1

A análise apresentada nesta seção é referente ao Experimento #1, descrito na Seção 3.2.4. Ao aplicar a abordagem do protocolo que faz uso da média, observou-se que apenas o programa *fluidanimate* do *benchmark PARSEC* apresentou aderência a uma distribuição Gaussiana em todos os tratamentos do Exp. #1. No caso do *benchmark SPLASH-2x*, observou-se que, para os tratamentos RL303 e RL503, nenhum de seus programas apresentou aderência a uma distribuição Gaussiana.

O fato de algumas amostras não seguirem uma distribuição Gaussiana sugere duas coisas. A primeira, que o *Runlevel* e o tipo de *Otimização* do compilador influenciam, de forma significativa, a variabilidade dos tempos de execução, alterando as distribuições dos tempos em diferentes tratamentos. A segunda, que nem sempre os tempos de execução seguem uma distribuição Gaussiana. A Tabela 4.1 sumariza os resultados obtidos para todos os programas dos três *benchmarks* presentes neste experimento. O símbolo “\*” indica os tratamentos em que as amostras dos programas obtiveram aderência a distribuição Gaussiana.

Ao analisar as comparações dois a dois entre os tratamentos para cada programa, observou-se que o teste *t* de *Welch* prevaleceu na maioria dos casos (ver Tabela 4.2). Esse resultado já era esperado, pois, segundo o protocolo *Speedup-Test*, o teste *t* de *Welch* deve ser utilizado sempre que pelo menos uma das amostras não apresentar aderência a uma distribuição Gaussiana ou quando as amostras não apresentarem a mesma variância.

Os resultados de todas as comparações dois a dois, entre os tratamentos, para os três *benchmarks* estão dispostos na Tabela 4.3. Observou-se que, com exceção do par RL302-RL502 (*benchmark SPLASH-2x*), todos os outros pares apresentaram diferença significativa em mais de 30% das comparações. Se forem considerados os pares RL302-RL303, RL302-RL503, RL303-RL502 e RL502-RL503 para os três *benchmarks*, observa-se que essa diferença está acima de 65%.

Entretanto, ao se considerarem apenas os pares RL302-RL502, RL303-RL503, observa-se que a diferença está abaixo de 60%, no caso do par RL302-RL502 do *benchmark PARSEC*, especificamente, essa diferença está abaixo dos 10%. Isso demonstra uma influência maior do fator 0, em relação ao fator RL, sobre os tempos de execução dos programas. Contudo, não se deve desconsiderar o fator RL, pois também apresentou valores altos. Assim, fica evidente a necessidade e a importância de se relatar o ambiente utilizado durante os experimentos computacionais, uma vez que, configurações diferentes produzem resultados diferentes. Contudo, conforme indicado pelo protocolo *Speedup-Test*, o fato de os

dados não serem aderentes a uma distribuição Gaussiana sugere que os resultados obtidos durante a análise de significância podem não ser corretos para o nível de confiança adotado. Assim, uma vez que a maioria das amostras obtidas neste estudo não apresentaram evidência estatística de que seus dados são provenientes de uma distribuição Gaussiana, não se pode afirmar que os resultados dessa análise estão corretos para o nível de confiança adotado.

**Tabela 4.1. Resultado do teste de aderência a uma distribuição Gaussiana para os tempos de execução dos programas para o Exp. #1.**

		RL302	RL303	RL502	RL503
<b>NPB</b>	<i>BT</i>	0,00	0,17*	0,00	0,00
	<i>CG</i>	0,00	0,00	0,00	0,00
	<i>DC</i>	0,00	0,80*	0,83*	0,00
	<i>EP</i>	0,00	0,00	0,00	0,00
	<i>FT</i>	0,00	0,35*	0,02	0,43*
	<i>IS</i>	0,00	0,00	0,00	0,00
	<i>LU</i>	0,30*	0,09	0,00	0,00
	<i>MG</i>	0,03	0,00	0,00	0,00
	<i>SP</i>	0,06*	0,04	0,01	0,00
	<i>UA</i>	0,00	0,00	0,03	0,03
<b>PARSEC</b>	<i>Blackscholes</i>	0,07*	0,00	0,00	0,00
	<i>Bodytrack</i>	0,98*	0,00	0,31*	0,00
	<i>Canneal</i>	0,00	0,00	0,00	0,00
	<i>Dedup</i>	0,00	0,00	0,00	0,00
	<i>Facesim</i>	0,58*	0,00	0,03	0,19*
	<i>Ferret</i>	0,40*	0,01	0,00	0,07*
	<i>Fluidanimate</i>	0,92*	0,43*	0,16*	0,78*
	<i>Freqmine</i>	0,03	0,14*	0,01	0,28*
	<i>Streamcluster</i>	0,00	0,00	0,00	0,03
	<i>Swaptions</i>	0,00	0,00	0,00	0,00
	<i>Vips</i>	0,15*	0,72*	0,14*	0,03
	<i>x264</i>	0,85*	0,60*	0,00	0,61*
<b>SPLASH-2X</b>	<i>Barnes</i>	0,00	0,00	0,00	0,00
	<i>Cholesky</i>	0,00	0,00	0,00	0,00
	<i>Fft</i>	0,00	0,00	0,04	0,00
	<i>Fmm</i>	0,00	0,00	0,00	0,03
	<i>lu_cb</i>	0,00	0,00	0,00	0,00
	<i>lu_ncb</i>	0,00	0,00	0,00	0,00
	<i>ocean_ncp</i>	0,00	0,00	0,00	0,00
	<i>Radiosity</i>	0,00	0,00	0,00	0,00
	<i>Raytrace</i>	0,00	0,00	0,00	0,00
	<i>Volrend</i>	0,00	0,00	0,00	0,00
	<i>water_nsquared</i>	0,00	0,00	0,00	0,00
	<i>water_spatial</i>	0,08*	0,00	0,49*	0,00

\*  $p\text{-value} \geq \alpha (\alpha = 0,05)$ .

Subsequentemente, aplicou-se a segunda abordagem do protocolo *Speedup-Test*, que faz referência ao parâmetro mediana. Ao analisar os resultados do teste de aderência (*Kolmogorov-Smirnov*), observou-se que a maior parte das amostras são aderentes a mesma distribuição de dados (ver Tabela 4.4). No entanto, em alguns casos, o percentual de amostras que não apresentam aderência a uma mesma distribuição está acima dos 25%, chegando a 50% no caso da comparação RL302-RL502 para o *NPB*. Isso significa que, diferente da primeira abordagem, a análise de significância com o nível de confiança adotado estava correta na maioria dos casos.

**Tabela 4.2. Percentual de comparações dois a dois por tipo de teste dos programas para o Exp. #1.**

	Comparações	Teste <i>t</i> de Student	Teste <i>t</i> de Welch
<i>NPB</i>	<b>RL302/RL303</b>	10,00	90,00
	<b>RL302/RL502</b>	0,00	100,00
	<b>RL302/RL503</b>	0,00	100,00
	<b>RL303/RL502</b>	10,00	90,00
	<b>RL303/RL503</b>	10,00	90,00
	<b>RL502/RL503</b>	0,00	100,00
<i>PARSEC</i>	<b>RL302/RL303</b>	0,00	100,00
	<b>RL302/RL502</b>	16,67	83,33
	<b>RL302/RL503</b>	8,33	91,67
	<b>RL303/RL502</b>	8,33	91,67
	<b>RL303/RL503</b>	16,67	83,33
	<b>RL502/RL503</b>	0,00	100,00
<i>SPLASH-2x</i>	<b>RL302/RL303</b>	0,00	100,00
	<b>RL302/RL502</b>	8,33	91,67
	<b>RL302/RL503</b>	0,00	100,00
	<b>RL303/RL502</b>	0,00	100,00
	<b>RL303/RL503</b>	0,00	100,00
	<b>RL502/RL503</b>	0,00	100,00

Os valores estão em %.

**Tabela 4.3. Percentual de comparações dois a dois estatisticamente diferentes dos programas para o Exp. #1 (com base na média).**

<i>NPB</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	80,00	40,00	90,00
<b>RL303</b>		70,00	50,00
<b>RL502</b>			80,00

<i>PARSEC</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	75,00	33,33	75,00
<b>RL303</b>		75,00	50,00
<b>RL502</b>			66,67

<i>SPLASH-2x</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	75,00	8,33	75,00
<b>RL303</b>		75,50	58,33
<b>RL502</b>			66,67

Os valores estão em %.

Ao realizar uma análise de significância com o teste de *Wilcoxon-Mann-Whitney*, como sugerido pelo protocolo, observou-se que, no caso do *NPB*, houve diferença entre as comparações, em relação ao que foi observado com a média. O par RL303-RL503 foi o que apresentou a maior diferença (40%). No caso do *PARSEC* e *SPLASH-2x*, notou-se que as comparações RL302-RL502 e RL303-RL503 alcançaram o mesmo percentual encontrado para a média (ver Tabela 4.5). Essa informação reforça a afirmação de que o fator O apresenta maior influência sobre os tempos de execução dos programas do que o fator RL.

**Tabela 4.4. Percentual de comparações dois a dois com a mesma distribuição dos programas para o Exp. #1.**

	Comparações	Mesma Distribuição	Distribuições Diferentes
<b>NPB</b>	<b>RL302/RL303</b>	60,00	40,00
	<b>RL302/RL502</b>	50,00	50,00
	<b>RL302/RL503</b>	70,00	30,00
	<b>RL303/RL502</b>	80,00	20,00
	<b>RL303/RL503</b>	90,00	10,00
	<b>RL502/RL503</b>	90,00	10,00
<b>PARSEC</b>	<b>RL302/RL303</b>	91,67	8,33
	<b>RL302/RL502</b>	91,67	8,33
	<b>RL302/RL503</b>	75,00	25,00
	<b>RL303/RL502</b>	83,33	16,67
	<b>RL303/RL503</b>	75,00	25,00
	<b>RL502/RL503</b>	66,67	33,33
<b>SPLASH-2x</b>	<b>RL302/RL303</b>	66,67	33,33
	<b>RL302/RL502</b>	91,67	8,33
	<b>RL302/RL503</b>	66,67	33,33
	<b>RL303/RL502</b>	75,00	25,00
	<b>RL303/RL503</b>	75,00	25,00
	<b>RL502/RL503</b>	58,33	41,67

Os valores estão em %.

**Tabela 4.5. Percentual de comparações dois a dois estatisticamente diferentes dos programas para o Exp. #1 (com base na mediana).**

<b>NPB</b>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	60,00	50,00	70,00
<b>RL303</b>		80,00	90,00
<b>RL502</b>			90,00

<b>PARSEC</b>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	91,67	33,33	83,33
<b>RL303</b>		91,67	50,00
<b>RL502</b>			75,00

<b>SPLASH-2X</b>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	66,67	8,33	66,67
<b>RL303</b>		75,00	58,33
<b>RL502</b>			66,67

Os valores estão em %.

Portanto, há uma diferença entre a análise realizada com a média e com a mediana. Essa diferença pode ser explicada pelo fato de a maior parte das amostras não indicar aderência a distribuição Gaussiana, portanto, o uso de métodos estatísticos que contemplem essa suposição se mostra inadequado. Desse modo, utilizar testes estatísticos que utilizam a mediana se mostra uma melhor escolha.

#### 4.3.2 Experimento #2

A análise apresentada nesta seção é referente ao Experimento #2, descrito na Seção 3.2.5. Ao analisar a aderência das amostras, dos dois *benchmarks* (*NPB* e *SPLASH-2x*), a uma distribuição Gaussiana, quando o fator *Otimização* estava no

nível O2, observou-se que em apenas 14% da amostra essa aderência foi confirmada. No caso do *benchmark NPB*, observou-se que com o tamanho de variável em 1024, nenhuma das amostras apresentou aderência a referida distribuição. No caso do *SPLASH-2x*, observou-se que este fato ocorreu com os tamanhos 0, 256 e 2048. A Tabela 4.6 sumariza os resultados obtidos para todos os programas dos dois *benchmarks* presentes neste experimento. O símbolo “\*” indica os tratamentos em que as amostras dos programas obtiveram aderência a distribuição Gaussiana.

**Tabela 4.6. Teste de aderência a uma distribuição Gaussiana para os tempos de execução dos programas com o fator de *Otimização O2* para o Exp. #2.**

	0	64	128	256	512	1024	2048	4096
<i>NPB</i>	<i>BT</i>	0,00	0,11*	0,00	0,00	0,01	0,04	0,00
	<i>CG</i>	0,00	0,00	0,00	0,02	0,00	0,00	0,00
	<i>DC</i>	0,00	0,00	0,00	0,00	0,09*	0,00	0,92* 0,08*
	<i>EP</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>FT</i>	0,14*	0,00	0,15*	0,00	0,04	0,00	0,00
	<i>IS</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>LU</i>	0,72*	0,00	0,42*	0,99*	0,02	0,02	0,07* 0,05*
	<i>MG</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,01
	<i>SP</i>	0,06*	0,01	0,27*	0,08*	0,13*	0,00	0,02 0,37*
	<i>UA</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<i>SPLASH-2x</i>	<i>barnes</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>cholesky</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>fft</i>	0,00	0,03	0,00	0,00	0,00	0,02	0,01 0,08*
	<i>fmm</i>	0,01	0,09*	0,01	0,00	0,03	0,00	0,01 0,00
	<i>lu_cb</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>lu_ncb</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>ocean_ncp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>radiosity</i>	0,00	0,00	0,00	0,00	0,00	0,01	0,00 0,00
	<i>raytrace</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>volrend</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>water_nsquared</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<i>water_spatial</i>	0,00	0,22*	0,31*	0,00	0,06*	0,48*	0,00 0,20*

Ao realizar a mesma análise, mas com as amostras obtidas com o fator de *Otimização O3*, observou-se que em apenas 11% dos casos foi confirmado que as amostras apresentam aderência a uma distribuição Gaussiana. Observou-se também que no caso do *benchmark SPLASH-2x*, com exceção do programa *water\_spatial*, todas as amostras dos outros programas não apresentaram aderência a referida distribuição. No caso do *benchmark NPB*, também pode ser observada uma alteração nas amostras que apresentam aderência a uma distribuição Gaussiana (ver Tabela 4.7). Essa evidência sugere que o fator *Otimização* tem influência sobre o tempo de execução dos programas, o que corrobora com os resultados encontrados no experimento #1.

Ao analisar as comparações dois a dois entre os tratamentos para cada programa de cada *benchmark*, observou-se que o teste *t* de Welch prevaleceu na maioria dos casos (ver Tabela 4.8 e Tabela 4.9). No caso do *benchmark SPLASH-2x*, em que os programas foram compilados com o fator de *Otimização* em O3, com exceção da comparação 64-128, todas as outras comparações foram realizadas com teste *t* de Welch em 100% dos casos. Esse resultado se deve ao fato de que a maioria das amostras não apresentou aderência a uma distribuição Gaussiana. Para o restante das comparações, a prevalência do teste *t* de Welch esteve acima de 70%.

**Tabela 4.7. Teste de aderência a uma distribuição Gaussiana para os tempos de execução dos programas com *Otimização O3* para o Exp. #2.**

	<b>0</b>	<b>64</b>	<b>128</b>	<b>256</b>	<b>512</b>	<b>1024</b>	<b>2048</b>	<b>4096</b>	
<b>NPB</b>	<b>BT</b>	0,00	0,00	0,15*	0,00	0,09*	0,05*	0,00	0,07*
	<b>CG</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	<b>DC</b>	0,00	0,00	0,00	0,00	0,00	0,01	0,13*	
	<b>EP</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>FT</b>	0,46*	0,40*	0,02	0,39*	0,09*	0,07*	0,14*	0,92*
	<b>IS</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>LU</b>	0,04	0,00	0,14*	0,18*	0,01	0,04	0,00	0,01
	<b>MG</b>	0,01	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>SP</b>	0,00	0,00	0,00	0,28*	0,00	0,00	0,06*	0,25*
	<b>UA</b>	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00
<b>SPLASH-2x</b>	<b>barnes</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>cholesky</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>fft</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>fmm</b>	0,00	0,00	0,00	0,00	0,00	0,01	0,02	
	<b>lu_cb</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>lu_ncb</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>ocean_ncp</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>radiosity</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>raytrace</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>volrend</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>water_nsquared</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
	<b>water_spatial</b>	0,00	0,06*	0,32*	0,00	0,01	0,02	0,00	0,17*

A Tabela 4.10 mostra os resultados de todas as comparações dois a dois entre os tratamentos, para todos os programas dos dois *benchmarks* presentes neste experimento. Observou-se no caso do *benchmark NPB*, compilado com o fator *Otimização* em 02, com exceção das comparações 0-2048 e 256-2048, o percentual das comparações estatisticamente significantes estiveram acima de 30%, no caso do *SPLASH-2x*, também compilado com o fator *Otimização* em 02, com exceção da comparação 512-1024, as comparações estatisticamente significantes estiveram abaixo de 34%, no caso da comparação 128-512, o número de comparações estatisticamente significante esteve em 0%.

Quando os *benchmarks* foram compilados com o fator *Otimização* 03, no caso do *NPB*, observou-se que, com exceção das comparações 0-4096 e 256-1024, o percentual de comparações estatisticamente significante esteve acima de 30%, chegando a 90% na comparação 256-2048. No caso do *SPLASH-2x*, as comparações estatisticamente significantes estiveram abaixo de 50%, chegando a 0% no caso da comparação 0-256 (ver Tabela 4.11).

No entanto, como a maior parte das amostras não apresentou aderência a uma distribuição Gaussiana, é possível que os resultados encontrados com o *Speedup-Test* não estejam corretos para o nível de confiança adotado. Além disso, identificou-se que a alteração do fator *Otimização* também contribuiu para o aumento ou para a diminuição do percentual de análises estatisticamente significantes; um exemplo pode ser observado na comparação 256-2048 do *benchmark NPB*: quando essa comparação é realizada com o fator de *Otimização* em 02 o percentual de comparações estatisticamente significante está em 10%, contudo, quando o fator é alterado para 03, esse percentual aumenta para 90%.

**Tabela 4.8. Percentual de comparações dois a dois por tipo de teste dos programas com o fator de *Otimização* O2 para o Exp. #2.**

	<i>NPB</i>		<i>SPLASH-2x</i>	
	Teste <i>t</i> de <i>Student</i>	Teste <i>t</i> de <i>Welch</i>	Teste <i>t</i> de <i>Student</i>	Teste <i>t</i> de <i>Welch</i>
<b>0/64</b>	0,00	100,00	0,00	100,00
<b>0/128</b>	30,00	70,00	0,00	100,00
<b>0/256</b>	20,00	80,00	0,00	100,00
<b>0/512</b>	10,00	90,00	0,00	100,00
<b>0/1024</b>	0,00	100,00	0,00	100,00
<b>0/2048</b>	10,00	90,00	0,00	100,00
<b>0/4096</b>	10,00	90,00	0,00	100,00
<b>64/128</b>	0,00	100,00	8,33	91,67
<b>64/256</b>	0,00	100,00	0,00	100,00
<b>64/512</b>	0,00	100,00	0,00	100,00
<b>64/1024</b>	0,00	100,00	8,33	91,67
<b>64/2048</b>	0,00	100,00	0,00	100,00
<b>64/4096</b>	0,00	100,00	8,33	91,67
<b>128/256</b>	20,00	80,00	0,00	100,00
<b>128/512</b>	10,00	90,00	0,00	100,00%
<b>128/1024</b>	0,00	100,00	8,33	91,67
<b>128/2048</b>	10,00	90,00	0,00	100,00
<b>128/4096</b>	10,00	90,00	8,33	91,67
<b>256/512</b>	10,00	90,00	0,00	100,00
<b>256/1024</b>	0,00	100,00	0,00	100,00
<b>256/2048</b>	10,00	90,00	0,00	100,00
<b>256/4096</b>	10,00	90,00	0,00	100,00
<b>512/1024</b>	0,00	100,00	0,00	100,00
<b>512/2048</b>	10,00	90,00	0,00	100,00
<b>512/4096</b>	20,00	80,00	0,00	100,00
<b>1024/2048</b>	0,00	100,00	0,00	100,00
<b>1024/4096</b>	0,00	100,00	8,33	91,67
<b>2048/4096</b>	0,00	100,00	0,00	100,00

Os valores estão em %.

Ao aplicar a segunda abordagem do protocolo *Speedup-Test*, que faz referência ao parâmetro mediana, observou-se que os resultados do teste de aderência (*Kolmogorov-Smirnov*) indicaram que a maior parte das amostras é aderente à mesma distribuição de dados (ver Tabela 4.12 e Tabela 4.13). Isso significa que, diferente da primeira abordagem, a análise de significância com o nível de confiança adotado, estava correta na maioria das comparações. No entanto, foi possível observar que houve diferença nos percentuais, quando se compararam os resultados com o fator de *Otimização* em O2 em relação ao mesmo fator em O3. No caso do *benchmark NPB*, diferente do que foi observado com fator de *Otimização* em O2, as comparações 128-512, 256-512, 256-512 e 2048-4096, com o fator de *Otimização* em O3, não apresentaram indícios de que as amostras são provenientes de uma mesma distribuição em mais de 50% das comparações.

**Tabela 4.9. Percentual de comparações dois a dois por tipo de teste dos programas com o fator de Otimização O3 para o Exp. #2.**

	NPB		SPLASH-2x	
	Teste <i>t</i> de Student	Teste <i>t</i> de Welch	Teste <i>t</i> de Student	Teste <i>t</i> de Welch
<b>0/64</b>	10,00	90,00	0,00	100,00
<b>0/128</b>	0,00	100,00	0,00	100,00
<b>0/256</b>	10,00	90,00	0,00	100,00
<b>0/512</b>	10,00	90,00	0,00	100,00
<b>0/1024</b>	10,00	90,00	0,00	100,00
<b>0/2048</b>	10,00	90,00	0,00	100,00
<b>0/4096</b>	10,00	90,00	0,00	100,00
<b>64/128</b>	0,00	100,00	8,33	91,67
<b>64/256</b>	10,00	90,00	0,00	100,00
<b>64/512</b>	10,00	90,00	0,00	100,00
<b>64/1024</b>	10,00	90,00	0,00	100,00
<b>64/2048</b>	10,00	90,00	0,00	100,00
<b>64/4096</b>	10,00	90,00	0,00	100,00
<b>128/256</b>	10,00	90,00	0,00	100,00
<b>128/512</b>	10,00	90,00	0,00	100,00
<b>128/1024</b>	10,00	90,00	0,00	100,00
<b>128/2048</b>	0,00	100,00	0,00	100,00
<b>128/4096</b>	0,00	100,00	0,00	100,00
<b>256/512</b>	10,00	90,00	0,00	100,00
<b>256/1024</b>	10,00	90,00	0,00	100,00
<b>256/2048</b>	20,00	80,00	0,00	100,00
<b>256/4096</b>	20,00	80,00	0,00	100,00
<b>512/1024</b>	20,00	80,00	0,00	100,00
<b>512/2048</b>	10,00	90,00	0,00	100,00
<b>512/4096</b>	10,00	90,00	0,00	100,00
<b>1024/2048</b>	10,00	90,00	0,00	100,00
<b>1024/4096</b>	10,00	90,00	0,00	100,00
<b>2048/4096</b>	20,00	80,00	0,00	100,00

Os valores estão em %.

**Tabela 4.10. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de Otimização O2 para o Exp. #2 (com base na média).**

	Comparações	64	128	256	512	1024	2048	4096
NPB	<b>0</b>	40,00	50,00	40,00	50,00	50,00	10,00	70,00
	<b>64</b>		60,00	40,00	40,00	60,00	50,00	60,00
	<b>128</b>			40,00	40,00	40,00	30,00	50,00
	<b>256</b>				40,00	40,00	10,00	60,00
	<b>512</b>					50,00	50,00	40,00
	<b>1024</b>						40,00	70,00
	<b>2048</b>							40,00
SPLASH-2x	<b>0</b>	8,33	16,67	25,00	8,33	33,33	25,00	16,67
	<b>64</b>		33,33	16,67	8,33	25,00	8,33	16,67
	<b>128</b>			8,33	0,00	8,33	8,33	16,67
	<b>256</b>				8,33	16,67	8,33	16,67
	<b>512</b>					50,00	33,33	16,67
	<b>1024</b>						16,67	8,33
	<b>2048</b>							33,33

Os valores estão em %.

**Tabela 4.11. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de *Otimização O3* para o Exp. #2 (com base na média).**

Comparações	64	128	256	512	1024	2048	4096
<b>NPB</b>	<b>0</b>	60,00	60,00	50,00	30,00	50,00	50,00
	<b>64</b>		60,00	40,00	60,00	50,00	60,00
	<b>128</b>			60,00	50,00	60,00	60,00
	<b>256</b>				50,00	20,00	90,00
	<b>512</b>					40,00	30,00
	<b>1024</b>						60,00
	<b>2048</b>						50,00
<b>SPLASH-2x</b>	<b>0</b>	16,67	8,33	0,00	16,67	8,33	25,00
	<b>64</b>		25,00	16,67	41,67	25,00	25,00
	<b>128</b>			8,33	41,67	16,67	25,00
	<b>256</b>				16,67	16,67	25,00
	<b>512</b>					8,33	25,00
	<b>1024</b>						33,33
	<b>2048</b>						41,67

Os valores estão em %.

**Tabela 4.12. Percentual de comparações dois a dois com a mesma distribuição dos programas com o fator de *Otimização O2* para o Exp. #2.**

Comparações	NPB		SPLASH-2X	
	Mesma Distribuição	Distribuições Diferentes	Mesma Distribuição	Distribuições Diferentes
<b>0/64</b>	80,00	20,00	75,00	25,00
<b>0/128</b>	90,00	10,00	83,33	16,67
<b>0/256</b>	90,00	10,00	58,33	41,67
<b>0/512</b>	70,00	30,00	66,67	33,33
<b>0/1024</b>	70,00	30,00	83,33	16,67
<b>0/2048</b>	80,00	20,00	91,67	8,33
<b>0/4096</b>	60,00	40,00	66,67	33,33
<b>64/128</b>	90,00	10,00	91,67	8,33
<b>64/256</b>	60,00	40,00	75,00	25,00
<b>64/512</b>	80,00	20,00	66,67	33,33
<b>64/1024</b>	70,00	30,00	83,33	16,67
<b>64/2048</b>	90,00	10,00	83,33	16,67
<b>64/4096</b>	90,00	10,00	75,00	25,00
<b>128/256</b>	60,00	40,00	75,00	25,00
<b>128/512</b>	80,00	20,00	66,67	33,33
<b>128/1024</b>	80,00	20,00	50,00	50,00
<b>128/2048</b>	70,00	30,00	83,33	16,67
<b>128/4096</b>	70,00	30,00	75,00	25,00
<b>256/512</b>	70,00	30,00	58,33	41,67
<b>256/1024</b>	80,00	20,00	58,33	41,67
<b>256/2048</b>	80,00	20,00	75,00	25,00
<b>256/4096</b>	90,00	10,00	75,00	25,00
<b>512/1024</b>	60,00	40,00	50,00	50,00
<b>512/2048</b>	90,00	10,00	66,67	33,33
<b>512/4096</b>	60,00	40,00	58,33	41,67
<b>1024/2048</b>	60,00	40,00	66,67	33,33
<b>1024/4096</b>	90,00	10,00	83,33	16,67
<b>2048/4096</b>	80,00	20,00	83,33	16,67

Os valores estão em %.

Ao realizar uma análise de significância com o teste de *Wilcoxon-Mann-Whitney*, como sugerido pelo protocolo em análise, observou-se que, no caso do *benchmark NPB* compilado com o fator de *Otimização 02*, o percentual de comparações estatisticamente significativas ficou acima de 20%, chegando a 70% em alguns casos. Se esses resultados forem comparados com os encontrados no parâmetro média, para o mesmo *benchmark* e o mesmo nível do fator *Otimização*, a diferença entre os resultados chega a 20% em algumas comparações. No caso do *benchmark SPLASH-2x*, o percentual de comparações estatisticamente significantes ficou abaixo dos 50%, quando o fator de *Otimização* estava em 02, as comparações 128-256 e 128-512, não apresentaram comparações estatisticamente significantes. Se esses resultados forem comparados aos resultados obtidos com o parâmetro média, com o fator de *Otimização* no mesmo nível, pode ser observado que os resultados são os mesmos para as comparações 0-512, 0-2048, 64-128, 64-512, 128-256, 128-1024, 256-512, 512-4096, 1024-4096 e 2048-4096. Para os outros resultados encontrou uma diferença de até 8,33% (ver Tabela 4.14).

**Tabela 4.13. Percentual de comparações dois a dois com a mesma distribuição dos programas com o fator de *Otimização 03* para o Exp. #2.**

	<i>NPB</i>		<i>SPLASH-2X</i>		
	Mesma Distribuição	Distribuições Diferentes	Mesma Distribuição	Distribuições Diferentes	
Comparações	<b>0/64</b>	90,00	10,00	58,33	41,67
	<b>0/128</b>	80,00	20,00	83,33	16,67
	<b>0/256</b>	80,00	20,00	83,33	16,67
	<b>0/512</b>	70,00	30,00	66,67	33,33
	<b>0/1024</b>	100,00	0,00	66,67	33,33
	<b>0/2048</b>	80,00	20,00	75,00	25,00
	<b>0/4096</b>	70,00	30,00	75,00	25,00
	<b>64/128</b>	70,00	30,00	66,67	33,33
	<b>64/256</b>	60,00	40,00	58,33	41,67
	<b>64/512</b>	60,00	40,00	66,67	33,33
	<b>64/1024</b>	60,00	40,00	58,33	41,67
	<b>64/2048</b>	70,00	30,00	58,33	41,67
	<b>64/4096</b>	50,00	50,00	66,67	33,33
	<b>128/256</b>	60,00	40,00	83,33	16,67
	<b>128/512</b>	40,00	60,00	75,00	25,00
	<b>128/1024</b>	60,00	40,00	66,67	33,33
	<b>128/2048</b>	80,00	20,00	83,33	16,67
	<b>128/4096</b>	60,00	40,00	66,67	33,33
	<b>256/512</b>	50,00	50,00	58,33	41,67
	<b>256/1024</b>	90,00	10,00	58,33	41,67
	<b>256/2048</b>	80,00	20,00	75,00	25,00
	<b>256/4096</b>	50,00	50,00	58,33	41,67
	<b>512/1024</b>	100,00	0,00	58,33	41,67
	<b>512/2048</b>	80,00	20,00	75,00	25,00
	<b>512/4096</b>	60,00	40,00	83,33	16,67
	<b>1024/2048</b>	80,00	20,00	58,33	41,67
	<b>1024/4096</b>	60,00	40,00	58,33	41,67
	<b>2048/4096</b>	30,00	70,00	75,00	25,00

Os valores estão em %.

Ao realizar uma análise de significância com o teste de *Wilcoxon-Mann-Whitney*, como sugerido pelo protocolo, observou-se que, no caso do *NPB* compilado com o fator de *Otimização 03* o percentual de comparações estatisticamente

significantes ficou acima de 30%, chegando a 90% (256-2048). No caso do *benchmark* *SPLASH-2x*, para o mesmo nível do fator *Otimização*, o percentual de comparações estatisticamente significantes ficou abaixo dos 50% (ver Tabela 4.15).

**Tabela 4.14. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de *Otimização* O2 para o Exp. #2 (com base na mediana).**

Comparações	64	128	256	512	1024	2048	4096
<i>NPB</i>	<b>0</b>	50,00	40,00	60,00	70,00	50,00	20,00
	<b>64</b>		70,00	60,00	40,00	70,00	60,00
	<b>128</b>			40,00	30,00	50,00	40,00
	<b>256</b>				40,00	40,00	30,00
	<b>512</b>					50,00	50,00
	<b>1024</b>						50,00
	<b>2048</b>						60,00
<i>SPLASH-2x</i>	<b>0</b>	8,33	16,67	25,00	16,67	33,33	8,33
	<b>64</b>		25,00	16,67	16,67	25,00	8,33
	<b>128</b>			0,00	0,00	16,67	8,33
	<b>256</b>				16,67	16,67	8,33
	<b>512</b>					50,00	33,33
	<b>1024</b>						16,67
	<b>2048</b>						25,00

Os valores estão em %.

**Tabela 4.15. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de *Otimização* O3 para o Exp. #2 (com base na mediana)**

Comparações	64	128	256	512	1024	2048	4096
<i>NPB</i>	<b>0</b>	60,00	50,00	50,00	50,00	30,00	70,00
	<b>64</b>		60,00	50,00	60,00	70,00	60,00
	<b>128</b>			60,00	60,00	70,00	70,00
	<b>256</b>				60,00	20,00	90,00
	<b>512</b>					40,00	60,00
	<b>1024</b>						60,00
	<b>2048</b>						60,00
<i>SPLASH-2x</i>	<b>0</b>	25,00	16,67	16,67	16,67	16,67	41,67
	<b>64</b>		25,00	25,00	33,33	33,33	25,00
	<b>128</b>			16,67	25,00	25,00	33,33
	<b>256</b>				16,67	16,67	25,00
	<b>512</b>					8,33	8,33
	<b>1024</b>						33,33
	<b>2048</b>						41,67

Os valores estão em %.

Quando foram observados os resultados obtidos com o fator de *Otimização* em O3, no caso do *NPB*, notou-se que em algumas comparações o resultado é igual ao encontrado com o parâmetro média (por exemplo: 0-64, 0-128 e 64-128), para o restante das comparações identificou-se uma diferença de até 20%. No caso do *SPLASH-2x*, também se observaram resultados iguais, quando se compararam os resultados obtidos com o parâmetro mediana em relação ao parâmetro média (por exemplo: 0-512, 64-128, 64-1024 e 256-4096); para o restante das comparações a diferença esteve em até 8,33%. Quando se compararam os resultados obtidos com o fator *Otimização* em O2 em relação a O3, no caso do *NPB*, observou-se uma diferença entre os resultados, chegando a 60% no caso da comparação 256-2048.

No caso do *SPLASH-2x*, também foi identificada uma diferença, chegando a 33,34% no caso da comparação 0-2048.

Essas observações evidenciam que há uma diferença entre a análise realizada com a média e a mediana. Essa diferença pode ser explicada pelo fato de a maior parte das amostras não indicar aderência a uma distribuição Gaussiana, e em alguns casos, não ser proveniente de uma mesma distribuição. Contudo, neste experimento, observou-se que o fator *Otimização* também contribuiu para as diferenças encontradas nos resultados. Assim, pode-se concluir que esse fator possui um grande impacto nos tempos de execução dos programas e deve ser controlado para se evitarem erros durante a realização de um experimento computacional.

#### 4.3.3 Experimento #3

Ao analisar a aderência das amostras do Experimento #3 a uma distribuição Gaussiana, considerando os programas dos dois *benchmarks*, observou-se que 41% das amostras apresentaram aderência a referida distribuição. Desse montante, 33% são amostras obtidas com a estratégia *scatter*, 39% sem estratégia e 28% com a estratégia *compact*. Assim, as evidências experimentais sugerem que o número de *threads* e as estratégias de alocação de *thread* influenciam na distribuição dos tempos de execução e que não escolher uma estratégia de alocação de *thread* resulta em uma possibilidade maior da distribuição dos tempos de execução ser aderente a uma distribuição Gaussiana (ver Tabela 4.16).

**Tabela 4.16. Teste de aderência a uma distribuição Gaussiana para os tempos de execução dos programas para o Exp. #3.**

<b>NPB</b>	<b>SA 2TH</b>	<b>SA 4TH</b>	<b>SA 6TH</b>	<b>CP 2TH</b>	<b>CP 4TH</b>	<b>CP 6TH</b>	<b>SC 2TH</b>	<b>SC 4TH</b>	<b>SC 6TH</b>
<b>BT</b>	0,23*	0,00	0,45*	0,00	0,64*	0,00	0,69*	0,00	0,08*
<b>CG</b>	0,00	0,01	0,01	0,00	0,00	0,00	0,00	0,90*	0,07*
<b>DC</b>	0,00	0,77*	0,35*	0,01	0,18*	0,52*	0,00	0,03	0,15*
<b>EP</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>FT</b>	0,01	0,00	0,06*	0,87*	0,00	0,00	0,43*	0,80*	0,58*
<b>IS</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>LU</b>	0,28*	0,12*	0,20*	0,02	0,13*	0,16*	0,05*	0,17*	0,00
<b>MG</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,01
<b>SP</b>	0,99*	0,00	0,12*	0,01	0,25*	0,87*	0,36*	0,00	0,44*
<b>UA</b>	0,06*	0,20*	0,22*	0,00	0,88*	0,32*	0,52*	0,48*	0,21*
<b>PARSEC</b>	<b>SA 2TH</b>	<b>SA 4TH</b>	<b>SA 6TH</b>	<b>CP 2TH</b>	<b>CP 4TH</b>	<b>CP 6TH</b>	<b>SC 2TH</b>	<b>SC 4TH</b>	<b>SC 6TH</b>
<b>blackscholes</b>	0,29*	0,14*	0,07*	0,62*	0,36*	0,43*	0,01	0,01	0,00
<b>bodytrack</b>	0,01	0,00	0,59*	0,02	0,02	0,00	0,02	0,02	0,00
<b>facesim</b>	0,17*	0,01	0,18*	0,00	0,52*	0,00	0,36*	0,01	0,03

Ao analisar as comparações dois a dois entre os tratamentos para cada programa de cada *benchmark*, observou-se que o teste *t* de *Welch* prevaleceu na maioria dos casos. No caso do *NPB*, o uso do teste *t* de *Welch* esteve acima de 80%, no caso do *PARSEC*, o uso deste teste esteve acima dos 66% (ver Tabela 4.17). Esse resultado é reflexo dos resultados encontrados na análise de aderência dos testes.

Ao analisar se houve diferença estatisticamente significante entre os tratamentos observou-se que no caso do *benchmark NPB*, com exceção da comparação SA2TH-SC2TH, o percentual de diferença estatisticamente significante

esteve acima de 90% das comparações. No caso do *benchmark PARSEC*, com exceção da comparação SA4TH-SC4TH, o percentual de diferença estatisticamente significante esteve acima de 66% das comparações (ver Tabela 4.18 e Tabela 4.19). Esses resultados indicam que há diferença entre as estratégias de alocação de *thread*. Obviamente, o número de *threads* melhora o desempenho dos programas e contribui para essa diferença, mas se for observado apenas os pares em que apenas a estratégia de alocação é alterada, percebe-se que também há um percentual elevado, chegando a 100% das comparações em alguns casos.

**Tabela 4.17. Percentual de comparações dois a dois por tipo de teste para o Exp. #3.**

Comparações	NPB		PARSEC	
	Teste <i>t</i> de Student	Teste <i>t</i> de Welch	Teste <i>t</i> de Student	Teste <i>t</i> de Welch
SA2TH/SA4TH	10,00	90,00	33,33	66,67
SA2TH/SA6TH	10,00	90,00	33,33	66,67
SA2TH/CP2TH	0,00	100,00	0,00	100,00
SA2TH/CP4TH	0,00	100,00	33,33	66,67
SA2TH/CP6TH	0,00	100,00	0,00	100,00
SA2TH/SC2TH	30,00	70,00	33,33	66,67
SA2TH/SC4TH	10,00	90,00	0,00	100,00
SA2TH/SC6TH	20,00	80,00	0,00	100,00
SA4TH/SA6TH	20,00	80,00	33,33	66,67
SA4TH/CP2TH	0,00	100,00	0,00	100,00
SA4TH/CP4TH	0,00	100,00	0,00	100,00
SA4TH/CP6TH	0,00	100,00	0,00	100,00
SA4TH/SC2TH	10,00	90,00	0,00	100,00
SA4TH/SC4TH	10,00	90,00	0,00	100,00
SA4TH/SC6TH	0,00	100,00	0,00	100,00
SA6TH/CP2TH	0,00	100,00	0,00	100,00
SA6TH/CP4TH	10,00	90,00	0,00	100,00
SA6TH/CP6TH	0,00	100,00	0,00	100,00
SA6TH/SC2TH	20,00	80,00	0,00	100,00
SA6TH/SC4TH	10,00	90,00	0,00	100,00
SA6TH/SC6TH	20,00	80,00	0,00	100,00
CP2TH/CP4TH	0,00	100,00	33,33	66,67
CP2TH/CP6TH	0,00	100,00	33,33	66,67
CP2TH/SC2TH	0,00	100,00	0,00	100,00
CP2TH/SC4TH	0,00	100,00	0,00	100,00
CP2TH/SC6TH	0,00	100,00	0,00	100,00
CP4TH/CP6TH	10,00	90,00	33,33	66,67
CP4TH/SC2TH	0,00	100,00	33,33	66,67
CP4TH/SC4TH	0,00	100,00	0,00	100,00
CP4TH/SC6TH	0,00	100,00	0,00	100,00
CP6TH/SC2TH	0,00	100,00	0,00	100,00
CP6TH/SC4TH	10,00	90,00	0,00	100,00
CP6TH/SC6TH	10,00	90,00	0,00	100,00
SC2TH/SC4TH	20,00	80,00	0,00	100,00
SC2TH/SC6TH	10,00	90,00	0,00	100,00
SC4TH/SC6TH	0,00	100,00	0,00	100,00

Os valores estão em %.

Posteriormente, as amostras foram avaliadas com o protocolo referente ao parâmetro mediana e ao analisar a aderência das amostras a uma mesma distribuição de dados, observou-se que boa parte das amostras não seguem a mesma distribuição, o que evidencia a influência na distribuição dos tempos de

execução por parte do número de *threads* e pela estratégia de alocação de *threads* (ver Tabela 4.20). Ao analisar o percentual de tratamentos estatisticamente diferentes, no caso do *NPB*, observou-se que com exceção da comparação SA2TH-CP4TH, todos os outros resultados foram iguais aos encontrados com o parâmetro média (ver Tabela 4.21). No caso do *benchmark SPLASH-2x* todos os resultados foram iguais aos encontrados com o parâmetro média (ver Tabela 4.22).

**Tabela 4.18. Percentual de comparações dois a dois estatisticamente diferentes com o *NPB* para o Exp. #3 (com base na média).**

	SA 4TH	SA 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
<b>SA2TH</b>	100,00	100,00	100,00	100,00	100,00	60,00	100,00	100,00
<b>SA4TH</b>		100,00	100,00	90,00	90,00	100,00	100,00	90,00
<b>SA6TH</b>			100,00	100,00	90,00	100,00	100,00	90,00
<b>CP2TH</b>				100,00	100,00	100,00	100,00	100,00
<b>CP4TH</b>					100,00	100,00	100,00	100,00
<b>CP6TH</b>						100,00	100,00	90,00
<b>SC2TH</b>							100,00	100,00
<b>SC4TH</b>								100,00

Os valores estão em %.

**Tabela 4.19. Percentual de comparações dois a dois estatisticamente diferentes com o *PARSEC* para o Exp. #3 (com base na média).**

	SA 4TH	SA 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
<b>SA2TH</b>	100,00	100,00	100,00	100,00	100,00	66,67	100,00	100,00
<b>SA4TH</b>		100,00	100,00	100,00	100,00	100,00	33,33	100,00
<b>SA6TH</b>			100,00	100,00	100,00	100,00	100,00	100,00
<b>CP2TH</b>				100,00	100,00	66,67	100,00	100,00
<b>CP4TH</b>					100,00	100,00	66,67	100,00
<b>CP6TH</b>						100,00	100,00	66,67
<b>SC2TH</b>							100,00	100,00
<b>SC4TH</b>								100,00

Os valores estão em %.

Essas observações, no entanto, podem não ser corretas para o nível de confiança adotado, isso porque, em alguns casos as amostras não são provenientes de uma mesma distribuição e, segundo os autores do protocolo investigado, não se pode garantir que os resultados estão corretos para o nível de confiança adotado. Contudo, é certo em afirmar que os fatores estudados nestes experimentos influenciaram nos tempos de execução dos programas. Portanto, recomenda-se que em um experimento computacional esses fatores sejam controlados.

**Tabela 4.20. Percentual de comparações dois a dois com a mesma distribuição do Exp. #3.**

Comparações	NPB		PARSEC	
	Mesma Distribuição	Distribuições Diferentes	Mesma Distribuição	Distribuições Diferentes
<b>SA2TH/SA4TH</b>	70,00	30,00	100,00	0,00
<b>SA2TH/SA6TH</b>	50,00	50,00	100,00	0,00
<b>SA2TH/CP2TH</b>	30,00	70,00	33,33	66,67
<b>SA2TH/CP4TH</b>	30,00	70,00	33,33	66,67
<b>SA2TH/CP6TH</b>	40,00	60,00	33,33	66,67
<b>SA2TH/SC2TH</b>	70,00	30,00	66,67	33,33
<b>SA2TH/SC4TH</b>	70,00	30,00	66,67	33,33
<b>SA2TH/SC6TH</b>	70,00	30,00	33,33	66,67
<b>SA4TH/SA6TH</b>	60,00	40,00	100,00	0,00
<b>SA4TH/CP2TH</b>	10,00	90,00	33,33	66,67
<b>SA4TH/CP4TH</b>	50,00	50,00	33,33	66,67
<b>SA4TH/CP6TH</b>	50,00	50,00	33,33	66,67
<b>SA4TH/SC2TH</b>	60,00	40,00	66,67	33,33
<b>SA4TH/SC4TH</b>	50,00	50,00	66,67	33,33
<b>SA4TH/SC6TH</b>	60,00	40,00	33,33	66,67
<b>SA6TH/CP2TH</b>	20,00	80,00	33,33	66,67
<b>SA6TH/CP4TH</b>	40,00	60,00	33,33	66,67
<b>SA6TH/CP6TH</b>	70,00	30,00	33,33	66,67
<b>SA6TH/SC2TH</b>	50,00	50,00	66,67	33,33
<b>SA6TH/SC4TH</b>	80,00	20,00	66,67	33,33
<b>SA6TH/SC6TH</b>	70,00	30,00	33,33	66,67
<b>CP2TH/CP4TH</b>	20,00	80,00	100,00	0,00
<b>CP2TH/CP6TH</b>	60,00	40,00	100,00	0,00
<b>CP2TH/SC2TH</b>	50,00	50,00	66,67	33,33
<b>CP2TH/SC4TH</b>	40,00	60,00	66,67	33,33
<b>CP2TH/SC6TH</b>	50,00	50,00	100,00	0,00
<b>CP4TH/CP6TH</b>	40,00	60,00	100,00	0,00
<b>CP4TH/SC2TH</b>	50,00	50,00	66,67	33,33
<b>CP4TH/SC4TH</b>	50,00	50,00	66,67	33,33
<b>CP4TH/SC6TH</b>	40,00	60,00	100,00	0,00
<b>CP6TH/SC2TH</b>	50,00	50,00	66,67	33,33
<b>CP6TH/SC4TH</b>	70,00	30,00	66,67	33,33
<b>CP6TH/SC6TH</b>	80,00	20,00	100,00	0,00
<b>SC2TH/SC4TH</b>	80,00	20,00	100,00	0,00
<b>SC2TH/SC6TH</b>	60,00	40,00	66,67	33,33
<b>SC4TH/SC6TH</b>	90,00	10,00	66,67	33,33

Os valores estão em %.

**Tabela 4.21. Percentual de comparações dois a dois estatisticamente diferentes com o NPB para o Exp. #3 (com base na mediana).**

	SA 4TH	SA 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
SA2TH	100,00	100,00	100,00	90,00	100,00	60,00	100,00	100,00
SA4TH		100,00	100,00	90,00	90,00	100,00	100,00	90,00
SA6TH			100,00	100,00	90,00	100,00	100,00	90,00
CP2TH				100,00	100,00	90,00	100,00	100,00
CP4TH					100,00	100,00	100,00	100,00
CP6TH						100,00	100,00	90,00
SC2TH							100,00	100,00
SC4TH								100,00

Os valores estão em %.

**Tabela 4.22. Percentual de comparações dois a dois estatisticamente diferentes com o PARSEC para o Exp. #3 (com base na mediana).**

	SA 4TH	SA 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
<b>SA2TH</b>	100,00	100,00	100,00	100,00	100,00	66,67	100,00	100,00
<b>SA4TH</b>		100,00	100,00	100,00	100,00	100,00	33,33	100,00
<b>SA6TH</b>			100,00	100,00	100,00	100,00	100,00	100,00
<b>CP2TH</b>				100,00	100,00	66,67	100,00	100,00
<b>CP4TH</b>					100,00	100,00	66,67	100,00
<b>CP6TH</b>						100,00	100,00	66,67
<b>SC2TH</b>							100,00	100,00
<b>SC4TH</b>								100,00

Os valores estão em %.

#### 4.3.4 Experimento #4

Os tratamentos do Experimento #4 foram divididos em quatro grupos experimentais. Essa divisão permitiu uma análise melhor dos seus resultados. Cada grupo é composto por dezesseis tratamentos. Os dois primeiros grupos são referentes à primeira parte do experimento, em que o sexto fator (*Shared Cache L3*) não está presente e, portanto, o experimento é composto por 32 tratamentos.

Inicialmente, ao avaliar a primeira parte desse experimento utilizando a primeira abordagem do protocolo, observou-se que o teste *t* de Welch prevaleceu em todas as comparações (ver Tabela 4.23). Isso significa que, ou as amostras não são aderentes a distribuição Gaussiana, ou elas não possuem a mesma variância.

Ao analisar as comparações dois a dois entre os 32 tratamentos, observou-se que houve diferença estatisticamente significante nos tempos de execução do mesmo programa dentro e fora do grupo experimental (ver Tabela 4.24).

**Tabela 4.23. Percentual de comparações dois a dois por tipo de teste com 32 tratamentos para o Exp. #4.**

	Teste <i>t</i> de Student	Teste <i>t</i> de Welch
<b>G1-G1</b>	36,30	63,70
<b>G1-G2</b>	27,70	72,30
<b>G2-G2</b>	19,10	80,90

Os valores estão em %.

**Tabela 4.24. Percentual de comparações dois a dois estatisticamente diferentes com 32 tratamentos para o Exp. #4 (com base na média).**

	G1	G2
<b>G1</b>	38,30	100,00
<b>G2</b>		77,50

Os valores estão em %.

Ao comparar as amostras, considerando a mediana, observou-se que a maior parte segue a mesma distribuição. Assim, subentende-se que a maior parte dos resultados obtidos com o teste de *Wilcoxon-Mann-Whitney* estão corretos para o nível de confiança adotado (ver Tabela 4.25).

**Tabela 4.25. Percentual de comparações dois a dois com a mesma distribuição com 32 tratamentos para o Exp. #4.**

	Mesma Distribuição	Distribuições Diferentes
G1-G1	100,00	0,00
G1-G2	99,60	0,40
G2-G2	97,70	2,30

Os valores estão em %.

Ao comparar os tratamentos dois a dois com o teste de *Wilcoxon-Mann-Whitney*, observou-se uma diferença entre os valores encontrados com a média e a mediana. Para o par G1-G1 essa diferença foi de 12,5% e para o par G2-G2 essa diferença foi de 3,3% (ver Tabela 4.26).

**Tabela 4.26. Percentual de comparações dois a dois estatisticamente diferentes com 32 tratamentos para o Exp. #4 (com base na mediana).**

	G1	G2
G1	50,80	100,00
G2		74,20

Os valores estão em %.

Os resultados baseados na média foram obtidos a partir de amostras que não seguem uma distribuição Gaussiana e, portanto, podem não ser precisos, utilizar os valores da mediana parece ser mais confiável.

Na sequência, avaliou-se os 64 tratamentos contidos na segunda parte desse experimento. Ao analisá-los levando em consideração a média, notou-se que novamente o teste *t* de *Welch* se destacou em relação ao teste *t* de *Student*. Em algumas comparações, o teste *t* de *Welch* foi utilizado 100% das vezes (G1-G3, G1-G4, G2-G3 e G2-G4). Com exceção da comparação G2-G2 (81,6%) o teste *t* de *Welch* é aplicado em mais de 90% das análises (ver Tabela 4.27).

**Tabela 4.27. Percentual de comparações dois a dois por tipo de teste com 64 tratamentos para o Exp. #4.**

	Teste <i>t</i> de <i>Student</i>	Teste <i>t</i> de <i>Welch</i>
G1-G1	1,60	98,40
G1-G2	5,10	94,90
G1-G3	0,00	100,00
G1-G4	0,00	100,00
G2-G2	18,40	81,60
G2-G3	0,00	100,00
G2-G4	0,00	100,00
G3-G3	0,80	99,20
G3-G4	0,80	99,20
G4-G4	1,20	98,80

Os valores estão em %.

Além disso, observou-se que o percentual das comparações dois a dois intergrupos G1-G1 e G2-G2 apresentou uma diferença de 28,3% e 8,4%, respectivamente. Essa diferença indica uma possível influência do fator *Shared Cache L3*, mesmo quando o seu nível está em *off* (ver Tabela 4.28).

**Tabela 4.28. Percentual de comparações dois a dois estatisticamente diferentes com 64 tratamentos para o Exp. #4 (com base na média).**

	G1	G2	G3	G4
G1	66,60	100,00	100,00	100,00
G2		69,10	59,70	100,00
G3			61,60	90,60
G4				75,80

Os valores estão em %.

Ao se avaliarem os resultados dos 64 tratamentos observando a mediana, notou-se que, em alguns grupos (G1-G3, G1-G4, G2-G3, G2-G4), o percentual de amostras que não seguem a mesma distribuição esteve acima de 90%. Essa observação indica que o fator *Shared Cache L3* foi influente o suficiente para alterar a distribuição dos tempos de execução dos programas (ver Tabela 4.29). Sendo assim, a maior parte das comparações realizadas entre esses grupos pode não ser precisa para o nível de confiança adotado.

**Tabela 4.29. Percentual de comparações dois a dois com a mesma distribuição com 64 tratamentos para o Exp. #4.**

	Mesma Distribuição	Distribuições Diferentes
<b>G1-G1</b>	93,70	6,30
<b>G1-G2</b>	97,70	2,30
<b>G1-G3</b>	2,70	97,30
<b>G1-G4</b>	5,10	94,90
<b>G2-G2</b>	100,00	0,00
<b>G2-G3</b>	1,60	98,40
<b>G2-G4</b>	5,10	94,90
<b>G3-G3</b>	88,30	11,70
<b>G3-G4</b>	80,90	19,10
<b>G4-G4</b>	71,90	28,10

Os valores estão em %.

Em sequência, aplicou-se o teste de *Wilcoxon-Mann-Whitney* para avaliar se houve diferença estatisticamente significante entre os tratamentos. Como esperado, observou-se uma diferença entre os valores encontrados com a média e a mediana. Contudo, como em alguns grupos (G1-G3, G1-G4, G2-G3, G2-G4) as comparações foram realizadas entre amostras provenientes de distribuições diferentes, o resultado pode não ser preciso para o nível de confiança adotado (ver Tabela 4.30).

**Tabela 4.30. Percentual de comparações dois a dois estatisticamente diferentes com 64 tratamentos para o Exp. #4 (com base na mediana).**

	G1	G2	G3	G4
G1	68,30	100,00	100,00	100,00
G2		70,80	58,50	100,00
G3			65,80	92,20
G4				69,20

Os valores estão em %.

#### 4.3.5 Críticas ao Protocolo *Speedup-Test*

O protocolo *Speedup-Test* cumpre o seu propósito de apresentar uma metodologia para a comparação de duas amostras. No entanto, esse protocolo fica limitado à análise de apenas duas amostras, porque os testes definidos são

conhecidos na literatura como testes de comparação de duas amostras e o seu uso na comparação de três ou mais amostras leva ao problema conhecido como *familywise error rate*, que é o aumento da probabilidade de se encontrar um erro do tipo I (HOWELL, 2009). Além disso, o protocolo é dividido em duas partes, uma que analisa a média das amostras e outra que avalia a mediana, e deixa a cargo do experimentador escolher qual delas utilizar. Ao se analisarem os tratamentos de cada experimento descrito neste estudo, observou-se que o uso da mediana é aconselhado na maioria das vezes, porque dificilmente se observa que as amostras seguem uma distribuição paramétrica. Isso pode ser explicado pelo fato de que, variáveis que envolvem tempo, normalmente, são distribuições assimétricas. Portanto, uma correção no protocolo, que indique o uso de testes não paramétricos, quando a aderência uma distribuição Gaussiana não for confirmada, pode evitar conclusões equivocadas. Outra questão é o fato de utilizar o teste *t* de Welch mesmo quando não foi identificado que as amostras são aderentes a uma distribuição Gaussiana. Na literatura (SEWARD; DOANE, 2014; HOWELL, 2009), esse teste é indicado para amostras provenientes de uma distribuição Gaussiana, mas que não apresentam a mesma variância.

Deste modo, justifica-se a elaboração de um outro protocolo que atenda aos requisitos de experimentos computacionais mais elaborados.

# 5. O PROTOCOLO RUNTIME ANALYSIS

## 5.1 Introdução

Ao estudar o protocolo *Speedup-Test* (TOUATI; WORMS; BRIAIS, 2013) (ver Seção 4.2) observou-se que ele é inadequado para tratar casos em que as amostras são obtidas de múltiplos tratamentos. Isso ocorre, porque, em uma pesquisa experimental, é comum a elaboração de dois ou mais tratamentos para observar os efeitos que cada fator tem sobre uma determinada variável resposta. Os testes  $t$  e *Wilcoxon-Mann-Whitney* (BARBETTA; REIS; BORNIA, 2010), utilizados pelo protocolo, são indicados para comparar dois tratamentos e avaliar a significância estatística das diferenças nos tempos de execução, pela média e mediana, respectivamente. No entanto, utilizar esses testes para avaliar três ou mais tratamentos pode aumentar a probabilidade de se rejeitar a hipótese nula ( $H_0$ ) quando esta é verdadeira (*familywise error rate*). Uma alternativa é utilizar testes para múltiplas comparações. Esses testes consideram o número de comparações realizadas e permitem avaliar se o *p-value* obtido está realmente correto.

Neste capítulo, é apresentado o protocolo *Runtime Analysis* (RTA) (ver Apêndice K), que apresenta os melhoramentos sugeridos ao protocolo *Speedup-Test* (TOUATI; WORMS; BRIAIS, 2013) e é utilizado para comparar duas ou mais amostras de tempos de execução de programas de computador.

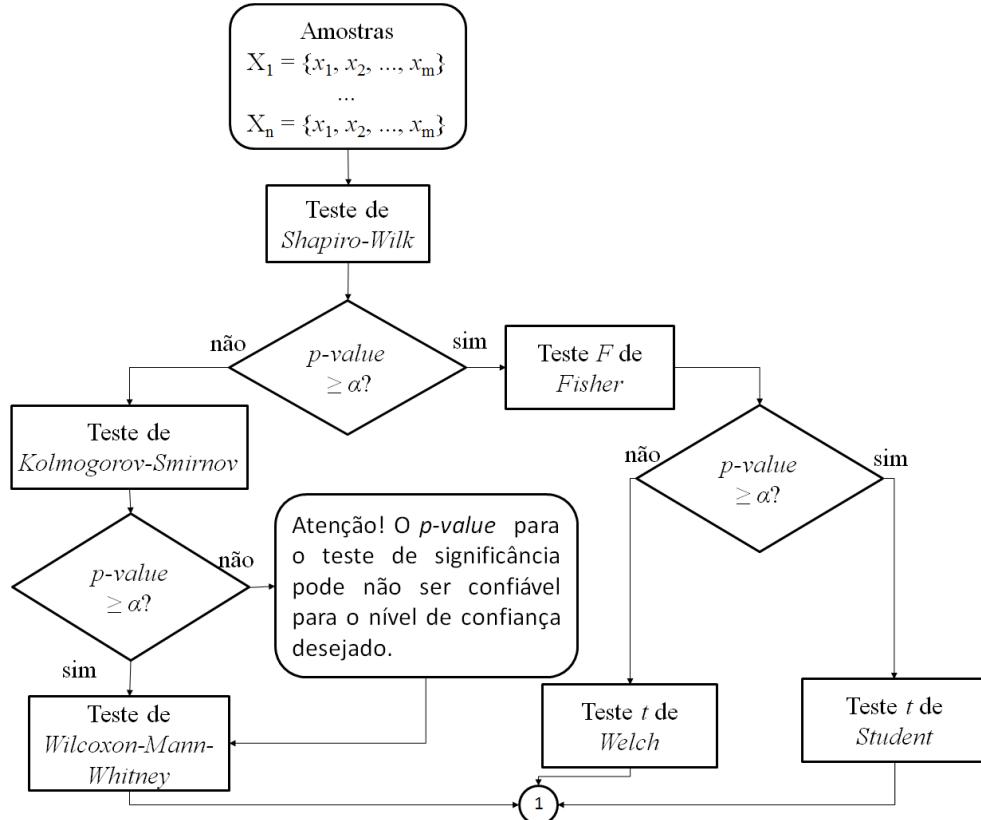
## 5.2 Propostas de Melhoramento do Protocolo *Speedup-Test*

### 5.2.1 Integração da metodologia de análise

O protocolo *Speedup-Test* oferece duas opções de metodologia para analisar as amostras de dois algoritmos, uma para analisar a média e outra para analisar a mediana. No caso da análise pela média, o protocolo sugere o uso do teste  $t$  de *Welch*, sempre que a amostra não apresentar aderência a uma distribuição Gaussiana. No entanto, a literatura (HOWELL, 2009) indica que esse teste deve ser usado quando as amostras apresentem aderência a uma distribuição Gaussiana, mas não apresentem a mesma variância. Na Seção 4.3, mostrou-se que a maior parte das amostras de tempos de execução de programas de computador não apresenta aderência a uma distribuição Gaussiana. No caso da análise da mediana das amostras, o protocolo sugere o uso do teste de *Kolmogorov-Smirnov* para avaliar se as amostras apresentam aderência a uma mesma distribuição. Nesse caso, a aderência a uma distribuição Gaussiana não é uma exigência, pois o teste de *Wilcoxon-Mann-Whitney* para duas amostras não tem como premissa que as amostras apresentem aderência a uma distribuição específica.

Por isso, neste capítulo apresenta-se uma proposta de correção para o protocolo *Speedup-Test*, de modo a integrar a metodologia de análise pela média com a metodologia de análise pela mediana. Assim sendo, nos experimentos em que as amostras apresentam aderência a uma distribuição Gaussiana, o protocolo segue

como o sugerido pelo *Speedup-Test*, no entanto, caso as amostras não apresentem aderência a uma distribuição Gaussiana, o protocolo *Runtime Analysis* sugere o uso do teste de aderência *Kolmogorov-Smirnov* para avaliar se as amostras são provenientes da mesma distribuição. Caso as amostras não apresentem aderência a mesma distribuição, o protocolo informa que o resultado obtido pode não estar correto para o nível de confiança adotado. Em seguida, o protocolo sugere o teste de *Wilcoxon-Mann-Whitney* para comparar, dois a dois, as amostras de cada tratamento. A Figura 5.1 ilustra esta etapa da proposta para o novo protocolo.



**Figura 5.1. Etapa para análise de aderência e comparação de tratamentos com o protocolo RTA.**

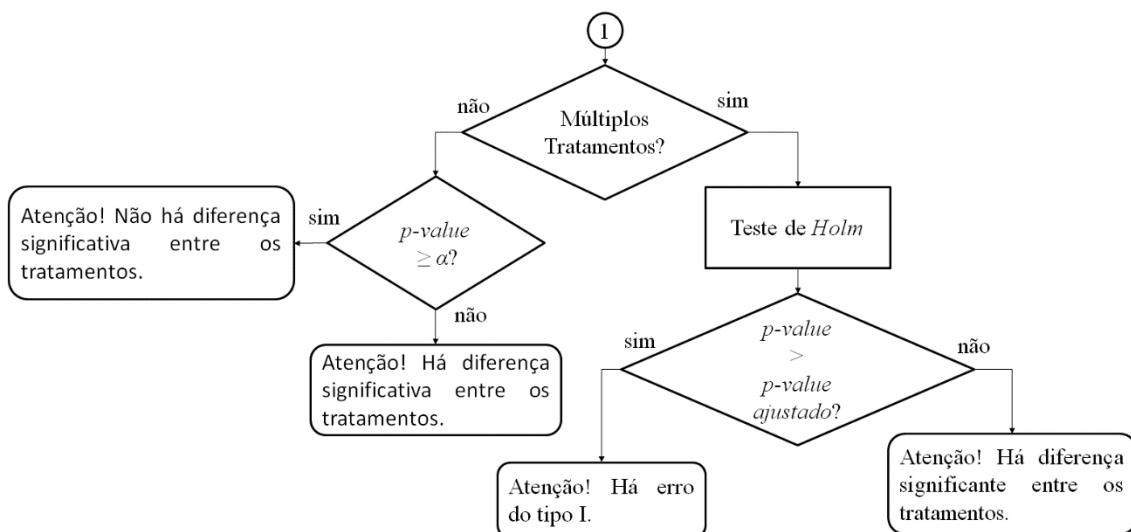
### 5.2.2 Múltiplas Comparações

O fato de o protocolo *Speedup-Test* sugerir testes de comparação de duas amostras torna-se um problema, quando se tem um experimento computacional com mais de dois tratamentos e se quer avaliar se há diferença estatisticamente significante entre eles. O uso de testes como a *ANOVA* (BARBETTA; REIS; BORNIA, 2010) ou mesmo o de *Kruskal-Wallis* (GIBBONS; CHAKRABORTI, 2014) é indicado para substituir o teste *t* e o teste de *Wilcoxon-Mann-Whitney* ao se comparar mais do que três tratamentos. No entanto, esses testes só indicam se há diferença entre os tratamentos, mas não aponta em qual comparação essa diferença foi encontrada.

Assim, para múltiplas comparações cujo objetivo é comparar os tratamentos dois a dois, o correto é utilizar um teste específico para múltiplas comparações. Neste estudo, optou-se pelo teste de *Holm* por ser um teste *post hoc*, ou seja, ele é aplicado após se obter o *p-value* de cada comparação (ver Seção 2.2.5).

A correção proposta neste estudo consiste em avaliar se os *p-values* encontrados nas comparações, dois a dois, dos tratamentos são realmente significantes. Por tanto, se forem realizadas mais do que duas comparações, o protocolo sugere que seja aplicado o teste de *Holm*, de modo que os *p-values* obtidos, menores do que  $\alpha$ , possam ser comparados a um *p-value ajustado*. Seguindo o teste de *Holm*, os *p-values* menores do que  $\alpha$ , devem ser ordenados de forma crescente e comparados, sequencialmente, a um *p-value* ajustado (ver Equação 2.20). Se o *p-value* for maior que o *p-value* ajustado o protocolo indica que há a presença do erro do tipo I para aquela comparação e que, a partir dela, todos os outros *p-values* estarão errados. Caso o contrário, o teste indica que os *p-values* estão corretos e que a diferença entre os tratamentos é estatisticamente significante.

No caso de haver apenas a comparação entre dois tratamentos, o protocolo apenas compara o *p-value* ao nível de significância ( $\alpha$ ) adotado e se o *p-value* for maior ou igual a  $\alpha$ , o protocolo indica que não há diferença estatisticamente significante entre os tratamentos. Caso contrário, o protocolo indica que a diferença é estatisticamente significante. A Figura 5.2 ilustra esta etapa para o novo protocolo.



**Figura 5.2. Etapa para a análise de múltiplas comparações com o protocolo RTA.**

### 5.3 Análise de Resultados

Nesta seção são apresentadas as análises de significância dos tempos de execução de programas, obtidos nos experimentos #1, #2, #3 e #4 (ver Seção 3.3). Nessa análise, aplicou-se o protocolo *RTA* com o intuito de comparar os tratamentos, observar se há diferença estatisticamente significante entre eles e analisar a presença de erro do tipo I nas comparações.

De todos os programas utilizados neste estudo, os programas *ocean\_cp* e *radix* disponibilizados pelo *benchmark SPLASH-2x* apresentaram tempos abaixo de 0,01 segundos, portanto, não foi possível mensurá-los com o programa *time* e, por isso, eles não foram analisados aqui. Para todos os testes estatísticos realizados adotou-se o nível de confiança de 95%.

### 5.3.1 Experimento #1

O experimento #1 teve por objetivo avaliar a influência dos fatores *Runlevel* e *Otimização* sobre os tempos de execução de um programa de computador (ver Seção 3.2.4). Ao comparar os tratamentos desse experimento, para cada *benchmark* com o protocolo *RTA*, observou-se que a maioria das comparações foram realizadas com o teste de *Wilcoxon-Mann-Whitney*, o que significa que a maioria das amostras não apresentou aderência a uma distribuição Gaussiana ou não apresentou a mesma variância. Dos *benchmarks* utilizados neste experimento, observou-se que apenas algumas amostras do *PARSEC* apresentaram diferença em sua variância. No caso das comparações RL302-RL303 e RL302-RL503 essa diferença foi observada em 25,00% das amostras. A Tabela 5.1 sumariza os resultados obtidos, com o protocolo *RTA*, para todos os programas dos três *benchmarks* presentes neste experimento.

**Tabela 5.1. Percentual de comparações dois a dois por tipo de teste dos programas com o protocolo *RTA* para o Exp. #1.**

	Comparações	Teste <i>t</i> de <i>Student</i>	Teste <i>t</i> de <i>Welch</i>	Teste de <i>Wilcoxon-Mann-Whitney</i>
<i>NPB</i>	<b>RL302/RL303</b>	10,00	0,00	90,00
	<b>RL302/RL502</b>	0,00	0,00	100,00
	<b>RL302/RL503</b>	0,00	0,00	100,00
	<b>RL303/RL502</b>	10,00	0,00	90,00
	<b>RL303/RL503</b>	10,00	0,00	90,00
	<b>RL502/RL503</b>	0,00	0,00	100,00
<i>PARSEC</i>	<b>RL302/RL303</b>	0,00	25,00	75,00
	<b>RL302/RL502</b>	16,67	8,33	75,00
	<b>RL302/RL503</b>	8,33	25,00	66,67
	<b>RL303/RL502</b>	8,33	8,33	83,33
	<b>RL303/RL503</b>	16,67	8,33	75,00
	<b>RL502/RL503</b>	0,00	8,33	91,67
<i>SPLASH-2x</i>	<b>RL302/RL303</b>	0,00	0,00	100,00
	<b>RL302/RL502</b>	8,33	0,00	91,67
	<b>RL302/RL503</b>	0,00	0,00	100,00
	<b>RL303/RL502</b>	0,00	0,00	100,00
	<b>RL303/RL503</b>	0,00	0,00	100,00
	<b>RL502/RL503</b>	0,00	0,00	100,00

Os valores estão em %.

Ao analisar se houve diferença estatisticamente significante entre os tratamentos, observou-se que, com exceção das comparações RL302-RL502, tanto para o *benchmark* *PARSEC*, quanto para o *benchmark* *SPLASH-2x*, a diferença entre os demais tratamentos esteve acima dos 50,00% (ver Tabela 5.2). Os resultados encontrados indicam que o fator *Otimização* apresenta uma influência maior sobre os tempos de execução dos programas do que o fator *Runlevel*.

No entanto, ao aplicar o teste de *Holm* ao conjunto de comparações com diferença estatisticamente significante, ou seja, que apresentaram *p-values* menores do que 0,05, para cada programa, observou-se que no caso do *benchmark* *NPB* apenas o programa *SP* apresentou um resultado (RL302-RL303) com erro do tipo I, o que representa 33,33% das comparações realizadas. No caso do *benchmark* *PARSEC*, observou-se que apenas os programas *blackscholes* e *swaptions* apresentaram resultados com erro do tipo I. No *blackscholes* verificou-se que 50,00% das comparações com diferença estatisticamente significativas estavam

erradas (RL303-RL502 e RL502-RL503), ao passo que no *swaptions* verificou-se que esse erro estava presente em 100,00% das comparações com diferença estatisticamente significante (RL302-RL303 e RL303-RL502).

**Tabela 5.2. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o protocolo RTA para o Exp. #1.**

<i>NPB</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	100,00	50,00	100,00
<b>RL303</b>		90,00	50,00
<b>RL502</b>			90,00
<i>PARSEC</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	91,67	33,33	83,33
<b>RL303</b>		91,67	50,00
<b>RL502</b>			75,00
<i>SPLASH-2x</i>			
Comparações	RL303	RL502	RL503
<b>RL302</b>	66,67	8,33	66,67
<b>RL303</b>		75,00	58,33
<b>RL502</b>			66,67

Os valores estão em %.

### 5.3.2 Experimento #2

Ao comparar os tratamentos desse experimento (ver Seção 3.2.5) com o protocolo *RTA*, para cada *benchmark* com o fator *Otimização* em 02, observou-se que a maioria das comparações foi realizada com o teste de *Wilcoxon-Mann-Whitney*, o que significa que a maioria das amostras não apresentou aderência a uma distribuição Gaussiana ou não apresentou a mesma variância.

No caso das comparações 64-4096 e 2048-4096, do *benchmark NPB* e das comparações 64-512, 128-512, 512-1024 e 512-4096, do *benchmark SPLASH-2x*, observou-se que em algumas comparações, as amostras, apresentaram aderência a uma distribuição Gaussiana, contudo, não apresentaram a mesma variância, o que levou essas comparações a serem analisadas com o teste *t* de *Welch* (ver Tabela 5.3).

Ao realizar a mesma análise com o fator *Otimização* em 03, observou-se que as comparações, em sua maioria, foram realizadas com o teste de *Wilcoxon-Mann-Whitney*. No caso das comparações 128-4096, 512-4096 e 1024-4096, do *benchmark NPB* e das comparações 64-4096 e 128-4096, do *benchmark SPLASH-2x*, observou-se que as amostras comparadas apresentaram aderência a uma distribuição Gaussiana, contudo, não apresentaram a mesma variância (ver Tabela 5.4). Essa evidência reforça que o fator de *Otimização* influencia a distribuição dos tempos de execução dos programas.

Ao analisar as comparações dois a dois entre os tratamentos, observou-se que há diferença estatisticamente significante acima dos 16,67%, tanto quando o fator *Otimização* estava em 02 quanto quando esse mesmo fator estava em 03 (ver Tabelas 5.5 e 5.6).

**Tabela 5.3. Percentual de comparações dois a dois por tipo de teste com o fator de *Otimização* O2 com o protocolo *RTA* para o Exp. #2.**

	Teste <i>t</i> de Student		Teste <i>t</i> de Welch		Teste de Wilcoxon-Mann-Whitney	
	<i>NPB</i>	<i>SPLASH-2x</i>	<i>NPB</i>	<i>SPLASH-2x</i>	<i>NPB</i>	<i>SPLASH-2x</i>
<b>0/64</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>0/128</b>	30,00	0,00	0,00	0,00	70,00	100,00
<b>0/256</b>	20,00	0,00	0,00	0,00	80,00	100,00
<b>0/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/1024</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>0/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/4096</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/128</b>	0,00	8,33	0,00	0,00	100,00	91,67
<b>64/256</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>64/512</b>	0,00	0,00	0,00	8,33	100,00	91,67
<b>64/1024</b>	0,00	8,33	0,00	0,00	100,00	91,67
<b>64/2048</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>64/4096</b>	0,00	8,33	10,00	0,00	90,00	91,67
<b>128/256</b>	20,00	0,00	0,00	0,00	80,00	100,00
<b>128/512</b>	10,00	0,00	0,00	8,33	90,00	91,67
<b>128/1024</b>	0,00	8,33	0,00	0,00	100,00	91,67
<b>128/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>128/4096</b>	10,00	8,33	0,00	0,00	90,00	91,67
<b>256/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>256/1024</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>256/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>256/4096</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>512/1024</b>	0,00	0,00	0,00	8,33	100,00	91,67
<b>512/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>512/4096</b>	20,00	0,00	0,00	8,33	80,00	91,67
<b>1024/2048</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>1024/4096</b>	0,00	8,33	0,00	0,00	100,00	91,67
<b>2048/4096</b>	0,00	0,00	10,00	0,00	90,00	100,00

Os valores estão em %.

Ao aplicar o teste de *Holm* ao conjunto de comparações com diferença estatisticamente significante, ou seja, que apresentaram *p-values* menores do que 0,05, para cada programa em que o fator *Otimização* estava em O2, observou-se que, com exceção dos *benchmarks* *CG* e *IS*, todos os outros *benchmarks* disponibilizados pelo *NPB* apresentaram erro do tipo I em suas comparações. No caso do *benchmark* *BT*, 47,05% de suas comparações com diferença estatisticamente significante apresentaram erro do tipo I, no caso do *DC* foram 69,23% das comparações, para o *EP* foram 33,33%, para o *FT* 45,45%, para o *LU* 27,78%, para o *MG* 17,39%, para o *SP* 77,78% e para o *UA* 100,00% das comparações estatisticamente significantes estavam com erro do tipo I.

Para o *benchmark* *SPLASH-2x*, observou-se que, com exceção do programa *volrend*, todos os outros programas apresentaram o erro do tipo I em pelos menos um dos seus tratamentos. No caso do programa *barnes*, 90,90% de suas comparações com diferença estatisticamente significante apresentaram erro do tipo I, no caso do *cholesky* foram 100,00% das comparações, do *fft* 66,67%, do *fmm* 78,57%, do *lu\_cb* 100,00%, do *lu\_ncb* 4,54%, do *ocean\_ncp* 100,00%, do *radiosity* 54,54%, do *raytrace* 38,46%, do *water\_nsquared* 33,33% e do *water\_spatial* 27,78%.

**Tabela 5.4. Percentual de comparações dois a dois por tipo de teste com o fator de *Otimização O3* com o protocolo *RTA* para o Exp. #2.**

	Teste <i>t</i> de Student		Teste <i>t</i> de Welch		Teste de Wilcoxon-Mann-Whitney	
	<i>NPB</i>	<i>SPLASH-2x</i>	<i>NPB</i>	<i>SPLASH-2x</i>	<i>NPB</i>	<i>SPLASH-2x</i>
<b>0/64</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/128</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>0/256</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/1024</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>0/4096</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/128</b>	0,00	8,33	0,00	0,00	100,00	91,67
<b>64/256</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/1024</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>64/4096</b>	10,00	0,00	0,00	8,33	90,00	91,67
<b>128/256</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>128/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>128/1024</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>128/2048</b>	0,00	0,00	0,00	0,00	100,00	100,00
<b>128/4096</b>	0,00	0,00	10,00	8,33	90,00	91,67
<b>256/512</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>256/1024</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>256/2048</b>	20,00	0,00	0,00	0,00	80,00	100,00
<b>256/4096</b>	20,00	0,00	0,00	0,00	80,00	100,00
<b>512/1024</b>	20,00	0,00	0,00	0,00	80,00	100,00
<b>512/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>512/4096</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>1024/2048</b>	10,00	0,00	0,00	0,00	90,00	100,00
<b>1024/4096</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>2048/4096</b>	20,00	0,00	0,00	0,00	80,00	100,00

Os valores estão em %.

**Tabela 5.5. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de *Otimização O2* com o protocolo *RTA* para o Exp. #2.**

<i>NPB</i>	Comparações	<b>64</b>	<b>128</b>	<b>256</b>	<b>512</b>	<b>1024</b>	<b>2048</b>	<b>4096</b>
<i>NPB</i>	<b>0</b>	50,00	40,00	60,00	70,00	50,00	20,00	50,00
<i>NPB</i>	<b>64</b>		70,00	60,00	40,00	70,00	60,00	70,00
<i>NPB</i>	<b>128</b>			40,00	30,00	50,00	40,00	50,00
<i>NPB</i>	<b>256</b>				40,00	40,00	30,00	60,00
<i>NPB</i>	<b>512</b>					50,00	50,00	50,00
<i>NPB</i>	<b>1024</b>						50,00	70,00
<i>NPB</i>	<b>2048</b>							50,00
<i>SPLASH-2x</i>	<b>0</b>	25,00	25,00	50,00	58,33	41,67	58,33	41,67
<i>SPLASH-2x</i>	<b>64</b>		41,67	33,33	50,00	25,00	58,33	25,00
<i>SPLASH-2x</i>	<b>128</b>			16,67	58,33	50,00	50,00	50,00
<i>SPLASH-2x</i>	<b>256</b>				75,00	41,67	41,67	33,33
<i>SPLASH-2x</i>	<b>512</b>					58,33	58,33	41,67
<i>SPLASH-2x</i>	<b>1024</b>						66,67	33,33
<i>SPLASH-2x</i>	<b>2048</b>							50,00

Os valores estão em %.

**Tabela 5.6. Percentual de comparações dois a dois estatisticamente diferentes dos programas com o fator de Otimização O3 com o protocolo RTA para o Exp. #2.**

	Comparações	64	128	256	512	1024	2048	4096
NPB	<b>0</b>	60,00	50,00	50,00	40,00	30,00	70,00	40,00
	<b>64</b>		60,00	50,00	60,00	70,00	60,00	60,00
	<b>128</b>			60,00	60,00	70,00	70,00	60,00
	<b>256</b>				60,00	20,00	90,00	40,00
	<b>512</b>					40,00	60,00	50,00
	<b>1024</b>						60,00	40,00
	<b>2048</b>							60,00
SPLASH-2x	<b>0</b>	33,33	41,67	41,67	16,67	33,33	66,67	16,67
	<b>64</b>		50,00	50,00	33,33	50,00	50,00	33,33
	<b>128</b>			16,67	33,33	58,33	58,33	58,33
	<b>256</b>				16,67	33,33	58,33	58,33
	<b>512</b>					33,33	33,33	50,00
	<b>1024</b>						66,67	50,00
	<b>2048</b>							50,00

Os valores estão em %.

No caso em que o fator *Otimização* estava em O3, com exceção do *benchmark EP* todos os outros *benchmarks* disponibilizados pelo *NPB* apresentaram erro do tipo I em suas comparações. No caso do *benchmark BT*, 30,00% de suas comparações com diferença estatisticamente significante apresentaram erro do tipo I, no caso do *CG* foram 7,41% das comparações, para o *DC* foram 100,00%, para o *FT* 54,54%, para o *IS* 35,00%, para o *LU* 33,33%, para o *MG* 21,05%, para o *SP* 80,00% e para o *UA* 77,78% das comparações estatisticamente significantes estavam com erro do tipo I.

Para o *benchmark SPLASH-2x*, observou-se que, com exceção do programa *volrend*, todos os outros programas apresentaram o erro do tipo I em, pelos menos, um dos seus tratamentos. No caso do programa *barnes*, 26,67% de suas comparações com diferença estatisticamente significante apresentaram erro do tipo I, no caso do *cholesky* foram 72,72% das comparações, do *fft* 80,00%, do *fmm* 5,88%, do *lu\_cb* 14,28%, do *lu\_ncb* 100,00%, do *ocean\_ncp* 100,00%, do *radiosity* 50,00%, do *raytrace* 30,00%, do *water\_nsquared* 27,27% e do *water\_spatial* 21,05%.

### 5.3.3 Experimento #3

O experimento #3 teve por objetivo analisar a influência do número e da estratégia de alocação de *threads* sobre os tempos de execução de um programa de computador (ver Seção 3.2.6). Ao analisar as comparações dois a dois entre os tratamentos, com o protocolo *RTA*, observou-se que a maior parte das comparações foi realizada com o teste de *Wilcoxon-Mann-Whitney*. No entanto, no caso do *benchmark NPB*, observou-se que em alguns casos o teste *t* de *Student* foi utilizado em até 30,00% das comparações e o teste *t* de *Welch* em até 40,00%. No caso do *benchmark PARSEC* observou-se que em alguns casos o teste *t* de *Student* foi utilizado em até 33,33% das comparações e o teste *t* de *Welch* em até 66,67% (ver Tabelas 5.7). Esses resultados indicam que o número e a estratégia de alocação de *threads* adotados alteram a distribuição dos tempos de execução dos programas.

Ao avaliar se houve diferença estatisticamente significante entre os tratamentos, observou-se que, com exceção da comparação SA4TH-SC4TH para o

*benchmark PARSEC*, todas as comparações apresentaram diferença estatisticamente significantes acima de 60,00% (ver Tabelas 5.8 e 5.9). Ao aplicar o teste de *Holm* ao conjunto de comparações com diferença estatisticamente significante, ou seja, que apresentaram *p-values* menores do que 0,05, para cada programa, observou-se que apenas o *benchmark MG* (3,03% das comparações) do *NPB* e o programa *blackscholes* (3,12% das comparações) do *PARSEC* apresentaram erro do tipo I. Para todos os outros programas os *p-values* encontrados estavam corretos.

**Tabela 5.7. Percentual de comparações dois a dois por tipo de teste com o protocolo *RTA* para o Exp. #3.**

	Teste <i>t</i> de Student		Teste <i>t</i> de Welch		Teste de Wilcoxon-Mann-Whitney	
	<i>NPB</i>	<i>PARSEC</i>	<i>NPB</i>	<i>PARSEC</i>	<i>NPB</i>	<i>PARSEC</i>
<b>SA2TH/SA4TH</b>	10,00	33,33	10,00	0,00	80,00	66,67
<b>SA2TH/SA6TH</b>	10,00	33,33	30,00	33,33	60,00	33,33
<b>SA2TH/CP2TH</b>	0,00	0,00	0,00	33,33	100,00	66,67
<b>SA2TH/CP4TH</b>	0,00	33,33	40,00	33,33	60,00	33,33
<b>SA2TH/CP6TH</b>	0,00	0,00	30,00	33,33	70,00	66,67
<b>SA2TH/SC2TH</b>	30,00	33,33	10,00	0,00	60,00	66,67
<b>SA2TH/SC4TH</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>SA2TH/SC6TH</b>	20,00	0,00	10,00	0,00	70,00	100,00
<b>SA4TH/SA6TH</b>	20,00	33,33	10,00	0,00	70,00	66,67
<b>SA4TH/CP2TH</b>	0,00	0,00	0,00	33,33	100,00	66,67
<b>SA4TH/CP4TH</b>	0,00	0,00	30,00	33,33	70,00	66,67
<b>SA4TH/CP6TH</b>	0,00	0,00	30,00	33,33	70,00	66,67
<b>SA4TH/SC2TH</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>SA4TH/SC4TH</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>SA4TH/SC6TH</b>	0,00	0,00	20,00	0,00	80,00	100,00
<b>SA6TH/CP2TH</b>	0,00	0,00	10,00	33,33	90,00	66,67
<b>SA6TH/CP4TH</b>	10,00	0,00	40,00	66,67	50,00	33,33
<b>SA6TH/CP6TH</b>	0,00	0,00	40,00	33,33	60,00	66,67
<b>SA6TH/SC2TH</b>	20,00	0,00	30,00	33,33	50,00	66,67
<b>SA6TH/SC4TH</b>	10,00	0,00	20,00	0,00	70,00	100,00
<b>SA6TH/SC6TH</b>	20,00	0,00	30,00	0,00	50,00	100,00
<b>CP2TH/CP4TH</b>	0,00	33,33	0,00	0,00	100,00	66,67
<b>CP2TH/CP6TH</b>	0,00	33,33	0,00	0,00	100,00	66,67
<b>CP2TH/SC2TH</b>	0,00	0,00	10,00	0,00	90,00	100,00
<b>CP2TH/SC4TH</b>	0,00	0,00	10,00	0,00	90,00	100,00
<b>CP2TH/SC6TH</b>	0,00	0,00	10,00	0,00	90,00	100,00
<b>CP4TH/CP6TH</b>	10,00	33,33	30,00	0,00	60,00	66,67
<b>CP4TH/SC2TH</b>	0,00	33,33	40,00	0,00	60,00	66,67
<b>CP4TH/SC4TH</b>	0,00	0,00	20,00	0,00	80,00	100,00
<b>CP4TH/SC6TH</b>	0,00	0,00	40,00	0,00	60,00	100,00
<b>CP6TH/SC2TH</b>	0,00	0,00	30,00	0,00	70,00	100,00
<b>CP6TH/SC4TH</b>	10,00	0,00	10,00	0,00	80,00	100,00
<b>CP6TH/SC6TH</b>	10,00	0,00	20,00	0,00	70,00	100,00
<b>SC2TH/SC4TH</b>	20,00	0,00	10,00	0,00	70,00	100,00
<b>SC2TH/SC6TH</b>	10,00	0,00	30,00	0,00	60,00	100,00
<b>SC4TH/SC6TH</b>	0,00	0,00	30,00	0,00	70,00	100,00

Os valores estão em %.

**Tabela 5.8. Percentual de comparações dois a dois estatisticamente diferentes com o NPB com o protocolo RTA para o Exp. #3.**

	SA 4TH	SA 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
<b>SA2TH</b>	100,00	100,00	100,00	90,00	100,00	60,00	100,00	100,00
<b>SA4TH</b>		100,00	100,00	90,00	90,00	100,00	100,00	90,00
<b>SA6TH</b>			100,00	100,00	90,00	100,00	100,00	90,00
<b>CP2TH</b>				100,00	100,00	90,00	100,00	100,00
<b>CP4TH</b>					100,00	100,00	100,00	100,00
<b>CP6TH</b>						100,00	100,00	90,00
<b>SC2TH</b>							100,00	100,00
<b>SC4TH</b>								100,00

Os valores estão em %.

**Tabela 5.9. Percentual de comparações dois a dois estatisticamente diferentes com o PARSEC com o protocolo RTA para o Exp. #3.**

	SA 4TH	AS 6TH	CP 2TH	CP 4TH	CP 6TH	SC 2TH	SC 4TH	SC 6TH
<b>SA2TH</b>	100,00	100,00	100,00	100,00	100,00	66,67	100,00	100,00
<b>SA4TH</b>		100,00	100,00	100,00	100,00	100,00	33,33	100,00
<b>SA6TH</b>			100,00	100,00	66,67	100,00	100,00	100,00
<b>CP2TH</b>				100,00	100,00	66,67	100,00	100,00
<b>CP4TH</b>					100,00	100,00	66,67	100,00
<b>CP6TH</b>						100,00	100,00	66,67
<b>SC2TH</b>							100,00	100,00
<b>SC4TH</b>								100,00

Os valores estão em %.

### 5.3.4 Experimento #4

Os tratamentos do Experimento #4 (ver Seção 3.2.7) foram divididos em quatro grupos. Cada grupo é composto por dezesseis tratamentos. Os dois primeiros grupos são referentes a primeira parte do experimento, em que o sexto fator (*Shared Cache L3*) não está presente e, portanto, o experimento é composto por 32 tratamentos. Em um primeiro momento, ao avaliar a primeira parte desse experimento utilizando o protocolo RTA, observou-se que o teste *Wilcoxon-Mann-Whitney* prevaleceu em todas as comparações (ver Tabela 5.10). Contudo, também se observou que as amostras do G1 apresentaram um alto índice de aderência a uma distribuição Gaussiana, 45,84%. Entretanto, 11,67% destas amostras não apresentaram a mesma variância. Diferentemente, os tratamentos do G2 apresentaram um baixo índice de aderência a uma distribuição Gaussiana, 17,50%. Assim, uma vez que o fator *Timer Interrupt Handling* está em nível *On* nos tratamentos deste grupo, entende-se que esse fator foi o responsável pela alteração na distribuição dos tempos de execução dos programas.

**Tabela 5.10. Percentual de comparações dois a dois por tipo de teste com 32 tratamentos com o protocolo RTA para o Exp. #4.**

	Teste <i>t</i> de Student	Teste <i>t</i> de Welch	Teste de Wilcoxon-Mann-Whitney
<b>G1-G1</b>	34,17	11,67	54,17
<b>G1-G2</b>	28,13	1,95	69,92
<b>G2-G2</b>	17,50	0,00	82,50

Os valores estão em %.

Ao analisar as comparações dois a dois entre os 32 tratamentos, observou-se que houve diferença estatisticamente significante nos tempos de execução do mesmo programa dentro e fora do grupo experimental (ver Tabela 5.11). Essa diferença chegou a 100% quando se compararam os tratamentos do G1 com os do G2. Ao aplicar o teste de *Holm*, ao conjunto de comparações com diferença estatisticamente significante, ou seja, que apresentaram *p-values* menores do que 0,05, observou-se que 15,15% dessas comparações apresentaram o erro do tipo I.

**Tabela 5.11. Percentual de comparações dois a dois estatisticamente diferentes com 32 tratamentos com o protocolo RTA para o Exp. #4.**

	G1	G2
G1	46,67%	100,0%
G2		70,00%

Os valores estão em %.

Em sequência, avaliou-se os 64 tratamentos contidos na segunda parte desse experimento. Ao analisá-los, notou-se que novamente o teste *Wilcoxon-Mann-Whitney* se destacou em relação aos outros testes. Com exceção da comparação G2-G2 (82,50%) o teste *Wilcoxon-Mann-Whitney* foi aplicado em mais de 90,00% das análises (ver Tabela 5.12). Ao analisar as comparações dois a dois entre os 64 tratamentos, observou-se que houve diferença estatisticamente significantes nos tempos de execução do mesmo programa dentro e fora do grupo experimental (ver Tabela 5.13). Essa diferença chegou a 100,00% nos pares G1-G2, G1-G3, G1-G4, G2-G4. No entanto, ao aplicar o teste de *Holm*, notou-se que 10,40% das comparações com diferença estatisticamente significante apresentaram erro do tipo I.

**Tabela 5.12. Percentual de comparações dois a dois por tipo de teste com 64 tratamentos com o protocolo RTA para o Exp. #4.**

	Teste <i>t</i> de Student	Teste <i>t</i> de Welch	Teste de Wilcoxon-Mann-Whitney
<b>G1-G1</b>	0,83	0,00	99,17
<b>G1-G2</b>	5,08	0,39	94,53
<b>G1-G3</b>	0,00	1,56	98,44
<b>G1-G4</b>	0,00	2,34	97,66
<b>G2-G2</b>	16,67	0,83	82,50
<b>G2-G3</b>	0,00	5,47	94,53
<b>G2-G4</b>	0,00	8,20	91,80
<b>G3-G3</b>	0,00	0,83	99,17
<b>G3-G4</b>	0,78	1,56	97,66
<b>G4-G4</b>	0,00	2,50	97,50

Os valores estão em %.

**Tabela 5.13. Percentual de comparações dois a dois estatisticamente diferentes com 64 tratamentos com o protocolo RTA para o Exp. #4.**

	G1	G2	G3	G4
G1	65,00	100,00	100,00	100,00
G2		67,50	52,73	100,00
G3			56,67	90,63
G4				60,00

Os valores estão em %.

## 5.4 Considerações Finais

Neste capítulo analisaram-se as amostras de cada experimento descrito no Capítulo terceiro, com base na metodologia empregada pelo protocolo *RTA*.

O protocolo *RTA* cumpre o seu objetivo, ao oferecer uma metodologia para análise de significância de múltiplos tratamentos de um elaborado experimento computacional. Os testes indicados nesse protocolo permitem avaliar amostras tanto de uma distribuição Gaussiana, quanto de outras não conhecidas *a priori*, além de identificar a presença do erro do tipo I.

As análises realizadas com o teste *Holm* na Seção 5.3, indicam que experimentos computacionais com múltiplos tratamentos que não consideram uma avaliação global sobre todos os tratamentos correm o risco de apresentarem conclusões erradas sobre os resultados obtidos e assim, replicar o seu erro para outros experimentos computacionais. Assim, para a análise de múltiplos tratamentos, é importante considerar o uso de testes que atendam essa condição. Caso contrário, as conclusões podem ser equivocadas e uma hipótese nula ( $H_0$ ) pode ser rejeitada, em detrimento de uma hipótese alternativa ( $H_1$ ), de forma incorreta.

# 6 CONCLUSÕES

## 6.1 Principais Resultados

Por não considerar um plano experimental adequado, muitos trabalhos erram ao mensurar e analisar os tempos de execução dos programas e, também erram, ao não descreverem o ambiente experimental em que os seus trabalhos foram realizados, dificultando assim a reproduzibilidade de seus resultados.

Em virtude disso, este estudo propõe um protocolo contendo diretrizes para a correta análise de significância da diferença entre os tempos de execução dos programas de computador. A contribuição desse protocolo está no fato de ele apresentar um controle para o problema conhecido como *familywise error rate*, presente na análise de múltiplas comparações de tratamentos em experimentos computacionais. Os resultados obtidos neste estudo enfatizam a variabilidade presente nos tempos de execução dos programas de computador decorrentes de fatores presentes no ambiente computacional, em especial, relacionados ao SO, além disso, enfatizam a presença do problema do *familywise error rate* decorrente da análise de significância de múltiplos tratamentos.

## 6.2 Limitações da Pesquisa

Neste estudo, devido às restrições de tempo e de equipamento, os experimentos #2 e #3 foram realizados com algumas limitações.

Com o Experimento #2, o intuito inicial era analisar os tempos de execução dos programas com uma variável de ambiente, em que, o seu tamanho variasse de 0 a 4096 bytes. No entanto, a reprodução desse experimento, com 30 execuções de cada programa com essa variação, consumiria um tempo exacerbado, inviabilizando assim a sua realização neste estudo. Por isso, o Experimento #2 foi limitado a sete tamanhos de variável. Esse número, apesar de pequeno, permitiu observar a variação dos tempos de execução dos programas conforme a alteração do tamanho de uma variável de ambiente.

No caso do Experimento #3, devido ao fato de o computador utilizado não possuir tecnologia Intel, o uso do compilador *icc* ficou inviável. Por isso, os programas deste experimento foram compilados apenas com o *gcc*. Além disso, a execução do experimento ficou restrita aos *benchmarks NPB* e aos programas *blackscholes*, *bodytrack* e *facesim* do *benchmark PARSEC*, pois eram os únicos programas que dispunham de uma versão *OpenMP*.

## 6.3 Contribuições para a Literatura

O estudo realizado nesta pesquisa fomentou a base para a produção e publicação de dois artigos científicos. A Tabela 6.1 apresenta os veículos nos quais os artigos foram publicados.

**Tabela 6.1. Publicações Científicas.**

Conferência	Qualis
ACM SAC - Symposium On Applied Computing	A1
SBESC – Simpósio Brasileiro de Engenharia de Sistemas Computacionais	B4

Além disso, o estudo permitiu disponibilizar um protocolo de comparação de tratamentos mais adequado para um estudo experimental com múltiplos tratamentos.

## 6.4 Dificuldades Encontradas

Por causa da quantidade de experimentos e de *benchmarks*, a grande dificuldade foi organizar e caracterizar os tempos de execução dos programas. Os *scripts* utilizados para executar os experimentos armazenavam os tempos de execução em arquivos textos. No entanto, esses arquivos continham outras informações além dos tempos, assim, foi necessário criar outro *script*, específico para retirar apenas os tempos de execução desses arquivos. Além disso, as análises de significância foram realizadas utilizando o R *Software* (R Core Team, 2014), que não possui um *frontend* amigável, toda a operação é realizada por comandos em seu *prompt* de comando. Para melhorar a interação com esse *Software* e facilitar o manuseio do grande número de dados, que foram gerados durante os experimentos, optou-se por utilizar o *frontend RKWard* (TEAM, 2013).

## 6.5 Trabalhos Futuros

Como trabalho futuro, pretende-se avaliar a variabilidade inerente ao método de coleta de tempos de execução dos programas. Observou-se, durante este estudo que há vários métodos de coleta de tempo, mas não se sabe ao certo qual deles tem o menor impacto ao coletar os tempos de execução. Pretende-se também avaliar o uso do método de regressão quantílica no estudo de variância de dados não paramétricos. Além disso, pretende-se incorporar ao protocolo métodos de análise de variância como ANOVA, Kruskal Wallis, ou mesmo regressão quantílica. A inclusão desses métodos permitirá avaliar, antecipadamente, se há diferença nos tempos de execução em um conjunto de tratamentos. Assim, se não houver diferença não será necessário comparar os tratamentos dois a dois.

# REFERÊNCIAS BIBLIOGRÁFICAS

- ALPERN, B. et al. The Jikes Research Virtual Machine project: Building an open-source research community. **IBM Systems Journal**, v. 44, p. 399–417, 2005.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. **Estatística Para Cursos de Engenharia e Informática**. 3. ed. [s.l.] Atlas, 2010. p. 410
- BARROW-WILLIAMS, N.; FENSCH, C.; MOORE, S. **A communication characterisation of Splash-2 and Parsec**2009 IEEE International Symposium on Workload Characterization (IISWC). **Anais...IEEE**, out. 2009Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5306792>>. Acesso em: 18 fev. 2015
- BHADAURIA, M.; WEAVER, V. M.; MCKEE, S. A. Understanding parsec performance on contemporary CMPS. **Proceedings of the 2009 IEEE International Symposium on Workload Characterization, IISWC 2009**, p. 98–107, 2009.
- BIENIA, C. **Benchmarking Modern Multiprocessors**. [s.l.] Princeton, 2011.
- BLACKBURN, S. M. et al. **The DaCapo benchmarks**ACM SIGPLAN Notices. **Anais...2006**
- BOLDT, M.; TRAULSEN, C.; HANXLEDEN, R. VON. Worst case reaction time analysis of concurrent reactive programs. **Electronic Notes in Theoretical ...**, v. 203, n. 4, p. 65–79, jun. 2008.
- DENNING, P. J. Is computer science science? **Communications of the ACM**, v. 48, n. 4, p. 27, 1 abr. 2005.
- DENNING, P. J. The science in computer science. **Communications of the ACM**, v. 56, n. 5, p. 35, 1 maio 2013.
- EMERSON, S. S. **Some Observations on the Wilcoxon Rank Sum Test**: UW Biostatistics Working Paper Series. [s.l.: s.n.]. Disponível em: <<http://biostats.bepress.com/uwbiostat/paper380>>.
- FIELD, A. **Descobrindo a estatística usando o SPSS**. 2. ed. [s.l.]: Bookman, 2009. 688 p. ISBN 9788536320182.
- GCC. **the GNU Compiler Collection**. Disponível em: <<https://gcc.gnu.org/>>. Acesso em: 20 jul. 2014.
- GEORGES, A.; BUYTAERT, D.; EECKHOUT, L. Statistically rigorous java performance evaluation. **ACM SIGPLAN Notices**, v. 42, n. 10, p. 57, 21 out. 2007.
- GIBBONS, J. D.; CHAKRABORTI, S. **Nonparametric Statistical Inference, Fourth Edition: Revised and Expanded**. 4 revisada ed. [s.l.] Taylor & Francis, 2014. p. 680

- GNU. **GNU libgomp**. Disponível em: <<https://gcc.gnu.org/onlinedocs/libgomp/>>. Acesso em: 20 jul. 2014.
- GROPP, W.; LUSK, E.; SKJELLUM, A. **Using MPI: Portable Parallel Programming with the Message-Passing Interface**. MIT Press, 2014. p. 336
- HAIR, J. F. et al. **Análise Multivariada de Dados**. 6. ed. [s.l.] Bookman, 2009. p. 688
- HANSEN, J.; HISSAM, S.; MORENO, G. A. StatisticalBased WCET Estimation and Validation. **9th International Workshop on Worst-Case Execution Time Analysis**, n. January, p. 123–133, 2009.
- HENNESSY, J. L.; PATTERSON, D. A. **Computer Architecture, Fifth Edition: A Quantitative Approach**. 5. ed. [s.l.] Morgan Kaufmann, 2011. p. 856
- HINTZE, J. L.; NELSON, R. D. Violin plots: A box plot-density trace synergism. **American Statistician**, v. 52, p. 181–184, 1998.
- HOWELL, D. C. **Statistical Methods For Psychology**. Seventh ed. [s.l.] Cengage Wadsworth, 2009. p. 792
- INTEL. **Thread Affinity Interface (Linux\* and Windows\*)**. Disponível em: <<https://software.intel.com/en-us/node/522691>>. Acesso em: 20 abr. 2015.
- JAIN, R. **The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling**. [s.l.] John Wiley, 1991. p. 720
- KEAREY, P.; KLEPEIS, K. A.; VINE, F. J. **Princípios de Bioestatística**. [s.l.] AMGH, 2014. p. 314
- KERRISK, M. **The Linux Programming Interface**. 1. ed. [s.l.] No Starch Press, 2010. p. 1552
- KOENKER, R. **Quantile Regression**. [s.l.] Cambridge University Press, 2005. p. 349
- KRISHNAMOORTHY, K. **Handbook of Statistical Distributions with Applications**. Chapman and Hall/CRC, 2006. p. 376 p
- LILJA, D. J. **Measuring Computer Performance: A Practitioner's Guide**. [s.l.] Cambridge University Press, 2005. p. 261
- LOOSEMORE, S. et al. **The GNU C Library Reference Manual**. Disponível em: <<http://www.gnu.org/software/libc/manual/>>. Acesso em: 19 out. 2013.
- MAXION, R. A.; LONGSTAFF, T. A.; MCHUGH, J. Why is there no science in cyber science? **Proceedings of the 2010 workshop on New security paradigms - NSPW '10**, p. 1, 2010.
- MAZOUZ, A.; BARTHOU, D. Analysing the variability of openMP programs performances on multicore architectures. **Fourth workshop on ...**, p. 1–14, 2011.

MAZOUZ, A.; TOUATI, S.; BARTHOU, D. Study of variations of native program execution times on multi-core architectures. **Complex, Intelligent and ...**, 2010.

MAZOUZ, A.; TOUATI, S.-A.-A.; BARTHOU, D. Performance evaluation and analysis of thread pinning strategies on multi-core platforms: Case study of SPEC OMP applications on intel architectures. **2011 International Conference on High Performance Computing & Simulation**, p. 273–279, jul. 2011.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 5. ed. [s.l.] John Wiley, 2000. p. 699

MYTKOWICZ, T. et al. Producing wrong data without doing anything obviously wrong! **ACM SIGPLAN Notices**, v. 44, n. 3, p. 265, 2009.

NASA ADVANCED SUPERCOMPUTING DIVISON. **NAS Parallel Benchmarks**. Disponível em: <<http://www.nas.nasa.gov/publications/npb.html>>. Acesso em: 18 fev. 2015.

NAVIDI, W. **Probabilidade e Estatística para Ciências Exatas**. [s.l.] McGraw Hill Brasil, 2012. p. 616

NIST/SEMATECH. **e-Handbook of Statistical Methods**.

NOGUEIRA, P. E.; MATIAS JR., R. **Estudo Quantitativo da Variabilidade nos Tempos de Execução de Experimentos Computacionais**Sociedade Brasileira de Engenharia de Sistemas Computacionais SBESC. **Anais...**2014

NOGUEIRA, P. E.; MATIAS, R.; VICENTE, E. **An experimental study on execution time variation in computer experiments**Proceedings of the 29th Annual ACM Symposium on Applied Computing - SAC '14. **Anais...**New York, New York, USA: ACM Press, 2014Disponível em: <<http://dl.acm.org/citation.cfm?doid=2554850.2555022>>. Acesso em: 18 fev. 2015

OLIVEIRA, A. B. et al. Why you should care about quantile regression. **ACM SIGPLAN Notices**, v. 48, n. 4, p. 207, 23 abr. 2013.

PUSUKURI, K. K.; GUPTA, R.; BHUYAN, L. N. Thread Tranquilizer. **ACM Transactions on Architecture and Code Optimization**, v. 8, n. 4, p. 1–21, 1 jan. 2012.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org>>.

RAMESH, V.; GLASS, R. L.; VESSEY, I. Research in computer science: an empirical study. **Journal of Systems and Software**, v. 70, n. 1-2, p. 165–176, fev. 2004.

RAZALI, N. M.; WAH, Y. B.; SCIENCES, M. Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests. v. 2, n. 1, p. 21–33, 2011.

- ROSS, S. M. **Introduction to Probability and Statistics for Engineers and Scientists**. 5. ed. [s.l.] Academic Press, 2014. p. 686
- ROYSTON, P. Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 44, p. 547–551, 1995.
- SAAVEDRA, R. H.; SMITH, A. J. Analysis of benchmark characteristics and benchmark performance prediction. **ACM Transactions on Computer Systems**, v. 14, n. 4, p. 344–384, 1996.
- SEN, A.; SRIVASTAVA, M. **Regression Analysis: Theory, Methods, and Applications**. [s.l.] Springer, 1997. p. 348
- SEWARD, L. E.; DOANE, D. P. **Estatística Aplicada à Administração e à Economia**. 4<sup>a</sup>. ed. [s.l.]: McGraw Hill Brasil, 2014. p. 860
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591–611, 1 dez. 1965.
- SHESKIN, D. J. **Handbook of Parametric and Nonparametric Statistical Procedures**. [s.l.] CRC Press, 2003. p. 1193
- SOBRINHO, E. V. DE P. **Estudo exploratório sobre fontes de**. [s.l.] Universidade Federal de Uberlândia, 2012.
- SPEC. **Standard Performance Evaluation Corporation - SPEC JVM, 98**. Disponível em: <<http://www.spec.org/jvm98/>>. Acesso em : 19 out. 2013a.
- SPEC. **Standard Performance Evaluation Corporation - SPEC CPU, 2006**. Disponível em: <<http://www.spec.org/cpu2006/>>. Acesso em: 19 out. 2013b.
- SPEC. **Standard Performance Evaluation Corporation - SPEC OMP, 2001**. Disponível em: <<http://www.spec.org/omp2001/>>. Acesso em: 19 out. 2013c.
- SPEC. **Standard Performance Evaluation Corporation**. Disponível em: <<http://www.spec.org/>>. Acesso em: 19 out. 2013d.
- STALLMAN, R. M.; COMMUNITY, T. G. D. **Using the GNU Compiler CollectionDevelopment** BostonGNU Press, 2012.
- STAPPERT, F.; ALTENBERND, P. Complete worst-case execution time analysis of straight-line hard real-time programs. **Journal of Systems Architecture**, p. 1–20, 2000.
- SURHONE, L. M.; TIMPLEDON, M. T.; MARSEKEN, S. F. **Tukey's Range Test**. [s.l.] VDM Publishing, 2010. p. 128
- TEAM, RKW. **RKWARD**. Disponível em: <[https://rkward.kde.org/Main\\_Page](https://rkward.kde.org/Main_Page)>. Acesso em: 19 out. 2013.

TICHY, W. F. Should computer scientists experiment more? **Computer**, v. 31, n. 5, p. 32–40, maio 1998.

TOUATI, S.; WORMS, J.; BRIAIS, S. The Speedup-Test: a statistical methodology for programme speedup analysis and computation. **Concurrency and Computation: Practice and Experience**, v. 25, n. 10, p. 1410–1426, 15 jul. 2013.

VICENTE, E.; MATIAS JR., R. **Modeling and Simulating the Effects of OS Jitter** Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. **Anais...** jan. 2013

WAINER, J. et al. Empirical evaluation in Computer Science research published by ACM. **Information and Software Technology**, v. 51, n. 6, p. 1081–1085, jun. 2009.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. [s.l.] Campus, 2009. p. 159

WIND RIVER SYSTEMS, INC., **Tornado Getting Started Guide, 2.2, Windows Version**, Wind River Systems, Inc., Alameda, CA, 2002

WOO, S. C. et al. **The SPLASH-2 programs** Proceedings of the 22nd annual international symposium on Computer architecture - ISCA '95. **Anais...** New York, New York, USA: ACM Press, 1995 Disponível em: <<http://portal.acm.org/citation.cfm?doid=223982.223990>>. Acesso em: 18 fev. 2015

WRIGHT, N. J. et al. **Measuring and understanding variation in benchmark performance** Department of Defense Proceedings of the High Performance Computing Modernization Program - Users Group Conference, HPCMP-UGC 2009. **Anais...** 2009

# APÊNDICE A: CLASSE DE *INPUT* PARA CADA *BENCHMARK NPB* (ADAPTADO DE NASA ADVANCED SUPERCOMPUTING DIVISON (2015))

		Classes						
		S	W	A	B	C	D	E
BT	Tamanho da grid	12 x 12 x 12	24 x 24 x 24	64 x 64 x 64	102 x 102 x 102	162 x 162 x 162	408 x 408 x 408	1020 x 1020 x 1020
	Número de Interações	60	200	200	200	200	250	250
	Time step	10 <sup>-2</sup>	8x10 <sup>-4</sup>	8x10 <sup>-4</sup>	3x10 <sup>-4</sup>	10 <sup>-4</sup>	2x10 <sup>-5</sup>	4x10 <sup>-6</sup>
CG	Número de Linhas	1,4x10 <sup>3</sup>	7x10 <sup>3</sup>	1,4x10 <sup>4</sup>	7,5x10 <sup>4</sup>	1,5x10 <sup>5</sup>	1,5x10 <sup>6</sup>	9x10 <sup>6</sup>
	Número de não zeros	7	8	11	13	15	21	26
	Número de interações	15	15	15	75	75	100	100
	Troca de auto valores	10	12	20	60	110	500	1500
DC	Tuplas	103	105	106	107	-	-	-
	Número de dimensões	5	10	15	20	-	-	-
EP	Pares de números randômicos	2 <sup>24</sup>	2 <sup>25</sup>	2 <sup>28</sup>	2 <sup>30</sup>	2 <sup>32</sup>	2 <sup>36</sup>	2 <sup>40</sup>
	Tamanho da grid	64 x 64 x 64	128 x 128 x 32	256 x 256 x 128	512 x 256 x 256	512 x 512 x 512	2048 x 1024 x 1024	4096 x 2048 x 2048
FT	Número de Interações	6	6	6	20	20	25	25
	Número de Chaves	216	220	223	225	227	231	-
	Máximo valor chave	211	216	219	221	223	227	-

(Continua)

# APÊNDICE A: CLASSE DE *INPUT* PARA CADA *BENCHMARK NPB* (ADAPTADO DE NASA ADVANCED SUPERCOMPUTING DIVISON (2015)) (CONT.)

		Classes						
		S	W	A	B	C	D	E
<i>LU</i>	Tamanho da grid	12 x 12 x 12	33 x 33 x 33	64 x 64 x 64	102 x 102 x 102	162 x 162 x 162	408 x 408 x 408	1020 x 1020 x 1020
	Número de interações	50	300	250	250	250	300	300
	Time step	0,5	0,0015	2,0	2,0	2,0	1,0	0,5
<i>MG</i>	Tamanho da grid	32 x 32 x 32	128 x 128 x 128	256 x 256 x 256	256 x 256 x 256	512 x 512 x 512	1024 x 1024 x 1024	2048 x 2048 x 2048
	Número de Interações	4	4	4	20	20	50	50
<i>SP</i>	Tamanho da grid	12 x 12 x 12	36 x 36 x 36	64 x 64 x 64	102 x 102 x 102	162 x 162 x 162	408 x 408 x 408	1020 x 1020 x 1020
	Número de Interações	$10^2$	$4 \times 10^2$	$4 \times 10^2$	$4 \times 10^2$	$4 \times 10^2$	$5 \times 10^2$	$5 \times 10^2$
	Time step	$1,5 \times 10^{-2}$	$1,5 \times 10^{-3}$	$1,5 \times 10^{-3}$	$10^{-3}$	$6,7 \times 10^{-4}$	$3 \times 10^{-4}$	$10^{-4}$
<i>UA</i>	Número de elementos	250	700	$2,4 \times 10^3$	$8,8 \times 10^3$	$3,35 \times 10^4$	$5,15 \times 10^5$	-
	Número de mortar points	$1,16 \times 10^4$	$2,67 \times 10^4$	$9,27 \times 10^4$	$3,346 \times 10^5$	$1,2621 \times 10^6$	$1,95 \times 10^7$	-
	Níveis de refinamento	4	5	6	7	8	10	-
	Número de interações	50	100	200	200	200	250	-
	Heat source radius	$4 \times 10^{-2}$	$6 \times 10^{-2}$	$7,6 \times 10^{-2}$	$7,6 \times 10^{-2}$	$6,7 \times 10^{-2}$	$4,6 \times 10^{-2}$	-

# APÊNDICE B: CONJUNTO DE *INPUT* PARA CADA PROGRAMA *SPLASH-2X* (ADAPTADO DE GROUP (2011))

Programas	Input	Tamanho	
		Complexo	Linear
barnes	Simsmall	16K Particles, Timestep = 0,25	Tolerance = 1,0
	Simmediu	32K Particles, Timestep = 0,25	Tolerance = 1,0
	Simlarge	256K Particles, Timestep = 0,25	Tolerance = 1,0
	Native	2M Particles, Timestep = 0,15	Tolerance = 0,5
	Simsmall	13992x13992, NZ=316740	
	Simmediu	13992x13992, NZ=316740	
	Simlarge	13992x13992, NZ=316740	
	Native	13992x13992, NZ=316740	
	Simsmall	20 <sup>2</sup> total complex data points	
	Simmediu	22 <sup>2</sup> total complex data points	
	Simlarge	24 <sup>2</sup> total complex data points	
	Native	28 <sup>2</sup> total complex data points	
fmm	Simsmall	16K Particles	Timestep = 5
	Simmediu	64K Particles	Timestep = 5
	Simlarge	256K Particles	Timestep = 5
	Native	4M Particles	Timestep = 5
lu_cb	Simsmall	512x512 Matrix, Block = 16	
	Simmediu	1Kx1K Matrix, Block = 16	
	Simlarge	2Kx2K Matrix, Block = 16	
	Native	8Kx8K Matrix, Block = 32	
lu_ncb	Simsmall	512x512 Matrix, Block = 16	
	Simmediu	1Kx1K Matrix, Block = 16	
	Simlarge	2Kx2K Matrix, Block = 16	
	Native	8Kx8K Matrix, Block = 32	
ocean_cp	Simsmall	514x514 Grid	Distance=20000,Timestep=28800
	Simmediu	1026x1026 Grid	Distance=20000,Timestep=28800
	Simlarge	2050x2050 Grid	Distance=20000,Timestep=28800
	Native	4098X4098 Grid	Distance=10000,Timestep=14400
ocean_ncp	Simsmall	514x514 Grid	Distance=20000,Timestep=28800
	Simmediu	1026x1026 Grid	Distance=20000,Timestep=28800
	Simlarge	2050x2050 Grid	Distance=20000,Timestep=28800
	Native	4098X4098 Grid	Distance=10000,Timestep=14400

(Contínua)

# APÊNDICE B: CONJUNTO DE *INPUT* PARA CADA PROGRAMA *SPLASH-2X* (ADAPTADO DE GROUP (2011)) (CONT.)

Programas	<i>Input</i>	Tamanho	
		Complexo	Linear
radix	Simsmall	4M Keys, Radix = 4K	
	Simmedium	16M Keys, Radix = 4K	
	Simlarge	64M Keys, Radix = 4K	
	Native	256M Keys, Radix = 4K	
radiosity	Simsmall	BF refinement =1,5e-1	Room
	Simmedium	BF refinement =1,5e-2	Room
	Simlarge	BF refinement =1,5e-3	Room
	Native	BF refinement =1,5e-4	Largeroom
raytrace	Simsmall	Teapot	Antialiasing w/ 8 subpixels
	Simmedium	Balls4	Antialiasing w/ 2 subpixels
	Simlarge	Balls4	Antialiasing w/ 8 subpixels
	Native	Car	Antialiasing w/ 128 subpixels
volrend	Simsmall	Head-Scaledown4	Rotate Step = 20
	Simmedium	Head-Scaledown2	Rotate Step = 50
	Simlarge	Head-Scaledown2	Rotate Step = 100
	Native	Head	Rotate Step = 1000
water_nsquared	Simsmall	$8^3$ Molecules	Timestep = 3
	Simmedium	$15^3$ Molecules	Timestep = 3
	Simlarge	$20^3$ Molecules	Timestep = 3
	Native	$32^3$ Molecules	Timestep = 7
water_spatial	Simsmall	$15^3$ Molecules	Timestep = 3
	Simmedium	$20^3$ Molecules	Timestep = 3
	Simlarge	$32^3$ Molecules	Timestep = 3
	Native	$100^3$ Molecules	Timestep = 3

# APÊNDICE C: CONJUNTO DE *INPUT* PARA CADA PROGRAMA *PARSEC* (ADAPTADO DE BIENIA (2011))

Programas	Input	Tamanho
<i>blockscholes</i>	Simsmall	4.096 opções
	Simmedium	16.384 opções
	Simlarge	65.536 opções
	Native	10.000.000 opções
	Simsmall	4 câmeras, 1 quadros, 1.000 partículas, 5 emparelhamentos
	Simmedium	4 câmeras, 2 quadros, 2.000 partículas, 5 emparelhamentos
	Simlarge	4 câmeras, 4 quadros, 4.000 partículas, 5 emparelhamentos
	Native	4 câmeras, 261 quadros, 4.000 partículas, 5 emparelhamentos
	Simsmall	10.000 trocas por passo de temperatura, 2.000° temperatura inicial, 100.000 elementos da lista, 32 passos da temperatura
	Simmedium	15.000 trocas por passo de temperatura, 2.000° temperatura inicial, 200.000 elementos da lista, 64 passos da temperatura
<i>canneal</i>	Simlarge	15.000 trocas por passo de temperatura, 2.000° temperatura inicial, 400.000 elementos da lista, 128 passos da temperatura
	Native	15.000 trocas por passo de temperatura, 2.000° temperatura inicial, 2.500.000 elementos da lista, 6.000 passos da temperatura
	Simsmall	10 MB
	Simmedium	31 MB
	Simlarge	184 MB
<i>dedup</i>	Native	72 MB
	Simsmall	80.598 partículas, 372.126 <i>tetrahedra</i> , 1 quadros
	Simmedium	80.598 partículas, 372.126 <i>tetrahedra</i> , 1 quadros
	Simlarge	80.598 partículas, 372.126 <i>tetrahedra</i> , 1 quadros
	Native	80.598 partículas, 372.126 <i>tetrahedra</i> , 100 quadros
<i>facesim</i>	Simsmall	consulta com 16 imagens, base com 3.544 imagens, encontrar topo10 imagens
	Simmedium	consulta com 64 imagens, base com 13.787 imagens, encontrar topo 10 imagens
	Simlarge	consulta com 256 imagens, base com 34.973 imagens, encontrar topo 10 imagens
	Native	consulta com 3.500 imagens, base com 59.695 imagens, encontrar topo 50 imagens
<i>ferret</i>	Simsmall	35.000 partículas, 5 quadros
	Simmedium	100.000 partículas, 5 quadros
	Simlarge	300.000 partículas, 5 quadros
	Native	500.000 partículas, 500 quadros
<i>fluidanimate</i>	Simsmall	35.000 partículas, 5 quadros
	Simmedium	100.000 partículas, 5 quadros
	Simlarge	300.000 partículas, 5 quadros
	Native	500.000 partículas, 500 quadros

(Continua)

# APÊNDICE C: CONJUNTO DE *INPUT* PARA CADA PROGRAMA *PARSEC* (ADAPTADO DE BIENIA (2011)) (CONT.)

	<b>Input</b>	<b>Tamanho</b>
Programas	<i>freqmine</i>	Simsmall      Base com 250.000 anônimos <i>click streams</i> de um portal de notícias Húngaro, suporte mínimo 220.
		Simmedium      Base com 500.000 anônimos <i>click streams</i> de um portal de notícias Húngaro, suporte mínimo 410.
		Simlarge      Base com 990.000 anônimos <i>click streams</i> de um portal de notícias Húngaro, suporte mínimo 790.
		Native      Base composta por uma coletânea de 250.000 documentos HTML, suporte mínimo 11.000.
	<i>streamcluster</i>	Simsmall      4.096 <i>input points</i> , tamanho do bloco 4.096 <i>points</i> , 32 <i>point dimensions</i> , 10–20 centros, até 1.000 centros intermediários permitidos
		Simmedium      8.192 <i>input points</i> , tamanho do bloco 8.192 <i>points</i> , 64 <i>point dimensions</i> , 10–20 centros, até 1.000 centros intermediários permitidos
		Simlarge      16.384 <i>input points</i> , tamanho do bloco 16.384 <i>points</i> , 128 <i>point dimensions</i> , 10–20 centers, até 1.000 centros intermediários permitidos
		Native      1.000.000 <i>input points</i> , tamanho do bloco 200.000 <i>points</i> , 128 <i>point dimensions</i> , 10–20 centros, até 5.000 centros intermediários permitidos
	<i>swaptions</i>	Simsmall      16 opções de <i>swap</i> , 5.000 simulações
		Simmedium      32 opções de <i>swap</i> , 10.000 simulações
		Simlarge      64 opções de <i>swap</i> , 20.000 simulações
		Native      128 opções de <i>swap</i> , 1.000.000 simulações
	<i>vips</i>	Simsmall      1.600 × 1.200 <i>pixels</i>
		Simmedium      1.600 × 1.200 <i>pixels</i>
		Simlarge      2.662 × 5.500 <i>pixels</i>
		Native      18.000 × 18.000 <i>pixels</i>
	<i>x264</i>	Simsmall      640 × 360 <i>pixels</i> (1/3 HDTV resolução), 8 quadros
		Simmedium      640 × 360 <i>pixels</i> (1/3 HDTV resolução), 32 quadros
		Simlarge      640 × 360 <i>pixels</i> (1/3 HDTV resolução), 128 quadros
		Native      1.920 × 1.080 <i>pixels</i> (HDTV resolução), 512 quadros

## APÊNDICE D: SERVIÇOS EM BACKGROUND DURANTE OS EXPERIMENTOS

Runlevel 1	
apparmor.service	loaded active exited LSB: AppArmor initialization
cycle.service	loaded active exited LSB: Set default boot entry if called
kmmod-static-nodes.service	loaded active exited Create list of required static device nodes for the current kernel
rescue.service	loaded active running Rescue Shell
systemd-fsck@dev-disk-by\x2did-ata\x2dCorsair_Force_3_SSD_121779010000148000CF\x2dpart3.service	loaded active exited File System Check on /dev/disk/by-id/ata-WDC_WD10EARS-00Y5B1_WD-WMAV52584480-part1
systemd-fsck@dev-disk-by\x2did-ata\x2dWDC_WD10EARS\x2d00Y5B1_WD\x2dWMAV52584480\x2dpart1.service	loaded active exited File System Check on /dev/disk/by-id/ata-Corsair_Force_3_SSD_121779010000148000CF-part3
systemd-journald.service	loaded active running Journal Service
systemd-modules-load.service	loaded active exited Load Kernel Modules
systemd-random-seed.service	loaded active exited Load/Save Random Seed
systemd-readahead-collect.service	loaded active exited Collect Read-Ahead Data
systemd-readahead-replay.service	loaded active exited Replay Read-Ahead Data
systemd-remount-fs.service	loaded active exited Remount Root and Kernel File Systems
systemd-sysctl.service	loaded active exited Apply Kernel Variables
systemd-tmpfiles-setup-dev.service	loaded active exited Create static device nodes in /dev
systemd-tmpfiles-setup.service	loaded active exited Recreate Volatile Files and Directories
systemd-udev-root-symlink.service	loaded active exited Create dynamic rule for /dev/root link
systemd-udev-trigger.service	loaded active exited udev Coldplug all Devices
systemd-udevd.service	loaded active running udev Kernel Device Manager
systemd-update-utmp.service	loaded active exited Update UTMP about System Reboot/Shutdown
systemd-vconsole-setup.service	loaded active exited Setup Virtual Console

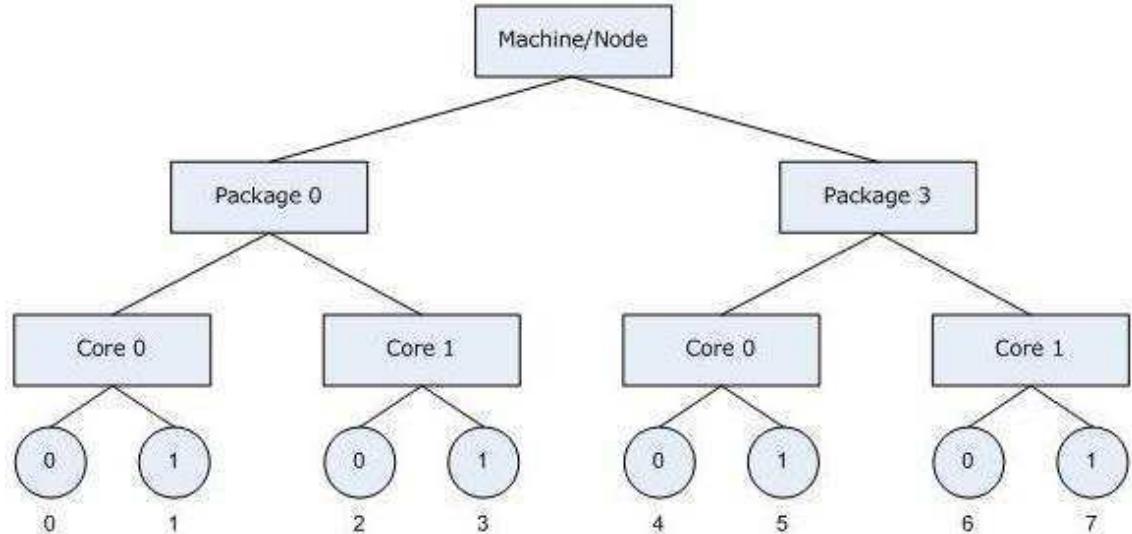
(Continua)

## APÊNDICE D: SERVIÇOS EM BACKGROUND DURANTE OS EXPERIMENTOS (CONT.)

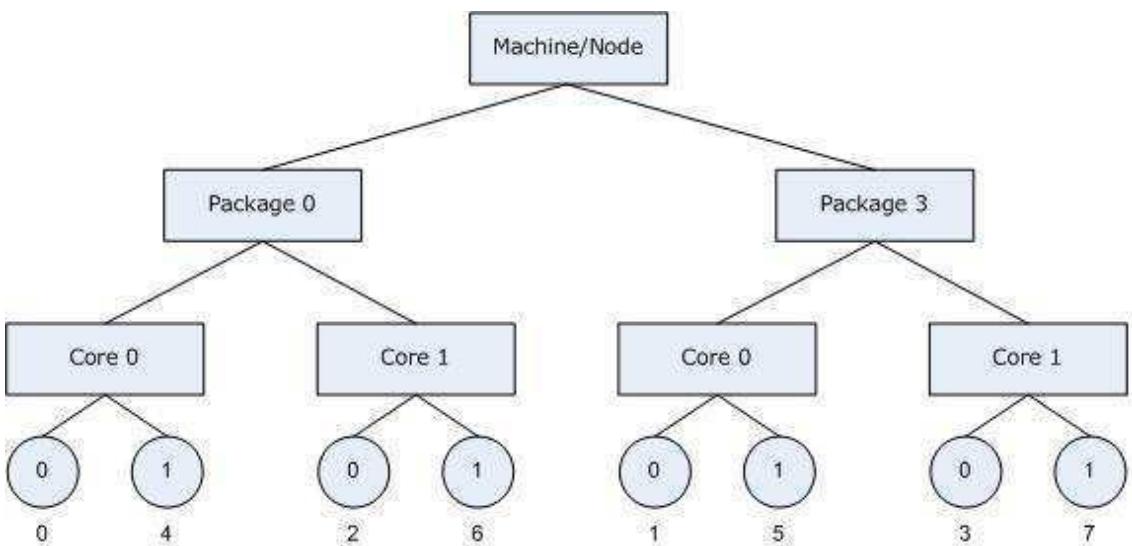
Runlevel 3	
Todos os serviços do <b>Runlevel 1</b> (menos o <i>rescue.service</i> ), além dos seguintes serviços:	
after-local.service	loaded active exited /etc/init.d/after.local Compatibility
avahi-daemon.service	loaded active running Avahi mDNS/DNS-SD Stack
cron.service	loaded active running Command Scheduler
dbus.service	loaded active running D-Bus System Message Bus
getty@tty1.service	loaded active running Getty on tty1
ModemManager.service	loaded active running Modem Manager
network.service	loaded active exited LSB: Configure network interfaces and set up routing
network@enp2s0f0.service	loaded active exited ifup managed network interface enp2s0f0
nscd.service	loaded active running Name Service Cache Daemon
polkit.service	loaded active running Authorization Manager
postfix.service	loaded active running Postfix Mail Transport Agent
rc-local.service	loaded active exited /etc/init.d/boot.local Compatibility
rsyslog.service	loaded active running System Logging Service
sshd.service	loaded active running OpenSSH Daemon
systemd-logind.service	loaded active running Login Service
systemd-user-sessions.service	loaded active exited Permit User Sessions
user@0.service	loaded active running User Manager for 0
user@1003.service	loaded active running User Manager for 1003
wpa_supplicant.service	loaded active running WPA Suplicant daemon

Runlevel 5	
Todos os serviços do <b>Runlevel 1</b> (menos o <i>rescue.service</i> ) e do <b>Runlevel 3</b> , além dos seguintes serviços:	
xdm.service	loaded active running LSB: X Display Manager

## APÊNDICE E: DISTRIBUIÇÃO DE THREADS PELOS NÚCLEOS SEGUNDO AS ESTRATÉGIAS DE THREAD AFFINITY (INTEL, 2015)



Distribuição das threads segundo a estratégia *compact*.



Distribuição das threads segundo a estratégia *scatter*.

## APÊNDICE F: CONFIGURAÇÃO DOS TRATAMENTOS NO EXP. #4

Tratamento	Descrição
<b>1</b>	Runlevel 5, Timers off, IRQ off, ProcessAffinity off, Clock_Tick off
<b>2</b>	Runlevel 1, Timers off, IRQ off, ProcessAffinity off, Clock_Tick off
<b>3</b>	Runlevel 5, Timers on, IRQ off, ProcessAffinity off, Clock_Tick off
<b>4</b>	Runlevel 1, Timers on, IRQ off, ProcessAffinity off, Clock_Tick off
<b>5</b>	Runlevel 5, Timers off, IRQ on, ProcessAffinity off, Clock_Tick off
<b>6</b>	Runlevel 1, Timers off, IRQ on, ProcessAffinity off, Clock_Tick off
<b>7</b>	Runlevel 5, Timers on, IRQ on, ProcessAffinity off, Clock_Tick off
<b>8</b>	Runlevel 1, Timers on, IRQ on, ProcessAffinity off, Clock_Tick off
<b>9</b>	Runlevel 5, Timers off, IRQ off, ProcessAffinity on, Clock_Tick off
<b>10</b>	Runlevel 1, Timers off, IRQ off, ProcessAffinity on, Clock_Tick off
<b>11</b>	Runlevel 5, Timers on, IRQ off, ProcessAffinity on, Clock_Tick off
<b>12</b>	Runlevel 1, Timers on, IRQ off, ProcessAffinity on, Clock_Tick off
<b>13</b>	Runlevel 5, Timers off, IRQ on, ProcessAffinity on, Clock_Tick off
<b>14</b>	Runlevel 1, Timers off, IRQ on, ProcessAffinity on, Clock_Tick off
<b>15</b>	Runlevel 5, Timers on, IRQ on, ProcessAffinity on, Clock_Tick off
<b>16</b>	Runlevel 1, Timers on, IRQ on, ProcessAffinity on, Clock_Tick off
<b>17</b>	Runlevel 5, Timers off, IRQ off, ProcessAffinity off, Clock_Tick on
<b>18</b>	Runlevel 1, Timers off, IRQ off, ProcessAffinity off, Clock_Tick on
<b>19</b>	Runlevel 5, Timers on, IRQ off, ProcessAffinity off, Clock_Tick on
<b>20</b>	Runlevel 1, Timers on, IRQ off, ProcessAffinity off, Clock_Tick on
<b>21</b>	Runlevel 5, Timers off, IRQ on, ProcessAffinity off, Clock_Tick on
<b>22</b>	Runlevel 1, Timers off, IRQ on, ProcessAffinity off, Clock_Tick on
<b>23</b>	Runlevel 5, Timers on, IRQ on, ProcessAffinity off, Clock_Tick on
<b>24</b>	Runlevel 1, Timers on, IRQ on, ProcessAffinity off, Clock_Tick on
<b>25</b>	Runlevel 5, Timers off, IRQ off, ProcessAffinity on, Clock_Tick on
<b>26</b>	Runlevel 1, Timers off, IRQ off, ProcessAffinity on, Clock_Tick on
<b>27</b>	Runlevel 5, Timers on, IRQ off, ProcessAffinity on, Clock_Tick on
<b>28</b>	Runlevel 1, Timers on, IRQ off, ProcessAffinity on, Clock_Tick on
<b>29</b>	Runlevel 5, Timers off, IRQ on, ProcessAffinity on, Clock_Tick on
<b>30</b>	Runlevel 1, Timers off, IRQ on, ProcessAffinity on, Clock_Tick on
<b>31</b>	Runlevel 5, Timers on, IRQ on, ProcessAffinity on, Clock_Tick on
<b>32</b>	Runlevel 1, Timers on, IRQ on, ProcessAffinity on, Clock_Tick on

(Continua)

## APÊNDICE F: CONFIGURAÇÃO DOS TRATAMENTOS NO EXP. #4 (CONT.)

Tratamento	Descrição
33	Runlevel 5, Timers off, IRQ off, ProcessAffinity off, Clock_Tick off, Shared Cache on
34	Runlevel 1, Timers off, IRQ off, ProcessAffinity off, Clock_Tick off, Shared Cache on
35	Runlevel 5, Timers on, IRQ off, ProcessAffinity off, Clock_Tick off, Shared Cache on
36	Runlevel 1, Timers on, IRQ off, ProcessAffinity off, Clock_Tick off, Shared Cache on
37	Runlevel 5, Timers off, IRQ on, ProcessAffinity off, Clock_Tick off, Shared Cache on
38	Runlevel 1, Timers off, IRQ on, ProcessAffinity off, Clock_Tick off, Shared Cache on
39	Runlevel 5, Timers on, IRQ on, ProcessAffinity off, Clock_Tick off, Shared Cache on
40	Runlevel 1, Timers on, IRQ on, ProcessAffinity off, Clock_Tick off, Shared Cache on
41	Runlevel 5, Timers off, IRQ off, ProcessAffinity on, Clock_Tick off, Shared Cache on
42	Runlevel 1, Timers off, IRQ off, ProcessAffinity on, Clock_Tick off, Shared Cache on
43	Runlevel 5, Timers on, IRQ off, ProcessAffinity on, Clock_Tick off, Shared Cache on
44	Runlevel 1, Timers on, IRQ off, ProcessAffinity on, Clock_Tick off, Shared Cache on
45	Runlevel 5, Timers off, IRQ on, ProcessAffinity on, Clock_Tick off, Shared Cache on
46	Runlevel 1, Timers off, IRQ on, ProcessAffinity on, Clock_Tick off, Shared Cache on
47	Runlevel 5, Timers on, IRQ on, ProcessAffinity on, Clock_Tick off, Shared Cache on
48	Runlevel 1, Timers on, IRQ on, ProcessAffinity on, Clock_Tick off, Shared Cache on
49	Runlevel 5, Timers off, IRQ off, ProcessAffinity off, Clock_Tick on, Shared Cache on
50	Runlevel 1, Timers off, IRQ off, ProcessAffinity off, Clock_Tick on, Shared Cache on
51	Runlevel 5, Timers on, IRQ off, ProcessAffinity off, Clock_Tick on, Shared Cache on
52	Runlevel 1, Timers on, IRQ off, ProcessAffinity off, Clock_Tick on, Shared Cache on
53	Runlevel 5, Timers off, IRQ on, ProcessAffinity off, Clock_Tick on, Shared Cache on
54	Runlevel 1, Timers off, IRQ on, ProcessAffinity off, Clock_Tick on, Shared Cache on
55	Runlevel 5, Timers on, IRQ on, ProcessAffinity off, Clock_Tick on, Shared Cache on
55	Runlevel 1, Timers on, IRQ on, ProcessAffinity off, Clock_Tick on, Shared Cache on
57	Runlevel 5, Timers off, IRQ off, ProcessAffinity on, Clock_Tick on, Shared Cache on
58	Runlevel 1, Timers off, IRQ off, ProcessAffinity on, Clock_Tick on, Shared Cache on
59	Runlevel 5, Timers on, IRQ off, ProcessAffinity on, Clock_Tick on, Shared Cache on
60	Runlevel 1, Timers on, IRQ off, ProcessAffinity on, Clock_Tick on, Shared Cache on
61	Runlevel 5, Timers off, IRQ on, ProcessAffinity on, Clock_Tick on, Shared Cache on
62	Runlevel 1, Timers off, IRQ on, ProcessAffinity on, Clock_Tick on, Shared Cache on
63	Runlevel 5, Timers on, IRQ on, ProcessAffinity on, Clock_Tick on, Shared Cache on
64	Runlevel 1, Timers on, IRQ on, ProcessAffinity on, Clock_Tick on, Shared Cache on

## APÊNDICE G: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #1

	<i>Runlevel 3 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	363,52	363,87	0,80	0,64	364,33	362,46	1,87	[363,22 - 363,82]	[363,11 - 364,63]
<b>CG</b>	87,99	88,01	0,06	0,00	88,05	87,76	0,29	[87,96 - 88,01]	[87,99 - 88,02]
<b>DC</b>	914,40	912,03	11,24	126,23	945,83	901,41	44,42	[910,21 - 918,60]	[909,14 - 914,91]
<b>EP</b>	241,37	239,58	4,24	18,02	251,80	239,53	12,27	[239,78 - 242,95]	[239,57 - 239,59]
<b>FT</b>	81,06	81,32	0,46	0,21	81,47	80,13	1,34	[80,89 - 81,23]	[81,09 - 81,55]
<b>IS</b>	13,06	13,03	0,14	0,02	13,80	13,02	0,78	[13,00 - 13,11]	[13,03 - 13,03]
<b>LU</b>	316,64	316,55	0,69	0,48	317,87	315,41	2,46	[316,38 - 316,90]	[316,10 - 316,99]
<b>MG</b>	17,11	17,10	0,05	0,00	17,18	17,03	0,15	[17,09 - 17,12]	[17,06 - 17,14]
<b>SP</b>	295,11	295,47	2,08	4,32	298,95	291,86	7,09	[294,33 - 295,88]	[294,21 - 296,72]
<b>UA</b>	264,44	263,76	1,87	3,51	268,37	262,90	5,47	[263,74 - 265,14]	[263,02 - 264,50]
	<i>Runlevel 3 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	386,01	386,04	0,19	0,04	386,34	385,69	0,65	[385,94 - 386,08]	[385,90 - 386,17]
<b>CG</b>	93,89	93,92	0,06	0,00	93,94	93,73	0,21	[93,86 - 93,91]	[93,92 - 93,92]
<b>DC</b>	948,68	948,35	9,40	88,43	974,23	930,62	43,61	[945,16 - 952,19]	[943,94 - 952,75]
<b>EP</b>	240,10	239,98	0,27	0,07	241,05	239,95	1,10	[240,00 - 240,20]	[239,95 - 240,01]
<b>FT</b>	80,66	80,72	0,50	0,25	81,62	79,47	2,15	[80,47 - 80,84]	[80,45 - 80,99]
<b>IS</b>	13,01	13,01	0,03	0,00	13,17	13,00	0,17	[13,00 - 13,03]	[13,01 - 13,01]
<b>LU</b>	305,63	305,55	0,96	0,93	307,44	304,00	3,44	[305,27 - 305,99]	[305,20 - 305,90]
<b>MG</b>	19,68	19,68	0,02	0,00	19,71	19,65	0,06	[19,67 - 19,68]	[19,67 - 19,69]
<b>SP</b>	296,16	296,58	1,68	2,83	298,55	292,98	5,57	[295,53 - 296,79]	[295,63 - 297,53]
<b>UA</b>	193,27	192,98	1,66	2,76	196,38	191,46	4,92	[192,65 - 193,89]	[191,94 - 194,01]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

(Continua)

# APÊNDICE G: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	Runlevel 5 e Otimização 02								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	363,28	363,34	0,23	0,05	363,73	362,57	1,16	[363,20 - 363,37]	[363,29 - 363,38]
<b>CG</b>	87,86	87,88	0,07	0,00	87,94	87,65	0,29	[87,84 - 87,89]	[87,88 - 87,88]
<b>DC</b>	915,66	916,22	7,31	53,44	928,17	898,79	29,38	[912,93 - 918,39]	[912,76 - 919,68]
<b>EP</b>	240,13	239,59	2,32	5,38	251,80	239,54	12,26	[239,26 - 241,00]	[239,58 - 239,60]
<b>FT</b>	81,10	81,17	0,50	0,25	81,79	80,24	1,55	[80,91 - 81,28]	[80,87 - 81,47]
<b>IS</b>	13,03	13,02	0,01	0,00	13,10	13,02	0,08	[13,02 - 13,03]	[13,02 - 13,02]
<b>LU</b>	316,86	316,79	0,34	0,11	317,87	316,38	1,49	[316,73 - 316,98]	[316,69 - 316,88]
<b>MG</b>	17,14	17,18	0,06	0,00	17,21	17,05	0,16	[17,12 - 17,16]	[17,13 - 17,23]
<b>SP</b>	296,79	297,41	1,63	2,65	298,80	292,42	6,38	[296,18 - 297,39]	[296,50 - 298,32]
<b>UA</b>	265,94	265,31	2,39	5,71	271,03	262,82	8,21	[265,04 - 266,83]	[263,97 - 266,64]
	Runlevel 5 e Otimização 03								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	388,28	388,47	0,43	0,18	388,61	387,17	1,44	[388,12 - 388,44]	[388,43 - 388,51]
<b>CG</b>	93,76	93,78	0,05	0,00	93,80	93,62	0,18	[93,74 - 93,78]	[93,76 - 93,79]
<b>DC</b>	950,70	948,09	13,06	170,56	1007,92	934,78	73,14	[945,83 - 955,58]	[944,12 - 952,06]
<b>EP</b>	240,21	239,98	0,42	0,18	241,35	239,94	1,41	[240,05 - 240,37]	[239,94 - 240,02]
<b>FT</b>	80,21	80,20	0,49	0,24	81,44	79,39	2,05	[80,03 - 80,39]	[79,90 - 80,50]
<b>IS</b>	13,02	13,01	0,02	0,00	13,14	13,00	0,14	[13,01 - 13,02]	[13,01 - 13,01]
<b>LU</b>	304,54	304,29	0,66	0,44	305,98	303,54	2,44	[304,29 - 304,79]	[304,04 - 304,53]
<b>MG</b>	19,78	19,79	0,04	0,00	19,81	19,68	0,13	[19,77 - 19,79]	[19,79 - 19,79]
<b>SP</b>	296,65	297,22	1,51	2,30	298,70	293,17	5,53	[296,09 - 297,22]	[296,55 - 297,89]
<b>UA</b>	193,21	193,07	1,35	1,82	195,87	191,46	4,41	[192,71 - 193,71]	[192,53 - 193,61]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 3 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	286,02	284,91	6,55	42,91	300,56	272,40	28,16	[283,57 - 288,46]	[283,74 - 286,07]
<i>bodytrack</i>	250,96	250,96	0,05	0,00	251,07	250,86	0,21	[250,94 - 250,97]	[250,93 - 250,98]
<i>canneal</i>	330,31	328,92	3,19	10,15	337,78	328,09	9,69	[329,12 - 331,50]	[328,57 - 329,26]
<i>dedup</i>	53,44	54,14	1,30	1,70	55,22	51,20	4,02	[52,95 - 53,93]	[53,15 - 55,13]
<i>facesim</i>	604,17	604,50	2,02	4,09	608,18	600,22	7,96	[603,42 - 604,93]	[603,41 - 605,59]
<i>ferret</i>	653,36	653,37	0,54	0,29	654,55	652,36	2,19	[653,16 - 653,56]	[653,20 - 653,53]
<i>fluidanimate</i>	583,21	583,20	0,11	0,01	583,47	582,99	0,48	[583,17 - 583,25]	[583,14 - 583,26]
<i>freqmine</i>	751,64	751,41	0,66	0,43	752,99	750,63	2,36	[751,40 - 751,89]	[751,03 - 751,78]
<i>streamcluster</i>	533,75	533,88	0,23	0,05	533,98	533,20	0,78	[533,67 - 533,84]	[533,73 - 534,02]
<i>swaptions</i>	556,61	547,28	18,93	358,31	599,05	541,41	57,64	[549,54 - 563,68]	[543,43 - 551,12]
<i>vips</i>	170,36	170,32	0,22	0,05	170,90	170,05	0,85	[170,28 - 170,44]	[170,21 - 170,42]
<i>x264</i>	130,23	130,23	0,21	0,04	130,65	129,83	0,82	[130,15 - 130,31]	[130,12 - 130,33]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 3 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	289,24	288,10	5,12	26,20	306,75	276,91	29,84	[287,33 - 291,15]	[287,83 - 288,37]
<i>bodytrack</i>	243,39	243,38	0,05	0,00	243,58	243,34	0,24	[243,37 - 243,41]	[243,36 - 243,40]
<i>canneal</i>	328,92	327,22	3,64	13,26	337,22	326,72	10,50	[327,56 - 330,28]	[327,01 - 327,43]
<i>dedup</i>	53,51	54,23	1,49	2,23	55,64	51,17	4,47	[52,95 - 54,07]	[52,97 - 55,48]
<i>facesim</i>	564,92	563,63	2,85	8,11	570,77	562,35	8,42	[563,85 - 565,98]	[562,98 - 564,28]
<i>ferret</i>	659,19	659,07	0,38	0,14	659,84	658,42	1,42	[659,05 - 659,33]	[658,79 - 659,34]
<i>fluidanimate</i>	576,54	576,51	0,35	0,12	577,09	575,85	1,24	[576,41 - 576,67]	[576,32 - 576,70]
<i>freqmine</i>	770,24	770,38	0,55	0,30	771,17	769,18	1,99	[770,04 - 770,44]	[770,10 - 770,66]
<i>streamcluster</i>	524,07	524,15	0,35	0,12	524,40	523,10	1,30	[523,94 - 524,20]	[523,95 - 524,35]
<i>swaptions</i>	548,75	544,46	14,87	221,00	598,27	535,30	62,97	[543,20 - 554,30]	[540,90 - 548,01]
<i>vips</i>	185,63	185,68	0,33	0,11	186,37	185,07	1,30	[185,51 - 185,76]	[185,56 - 185,79]
<i>x264</i>	131,35	131,30	0,67	0,45	133,17	129,96	3,21	[131,10 - 131,61]	[130,95 - 131,64]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 5 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	288,14	284,70	15,19	230,59	345,98	271,55	74,43	[282,47 - 293,81]	[279,10 - 290,29]
<i>bodytrack</i>	250,94	250,93	0,05	0,00	251,05	250,86	0,19	[250,92 - 250,96]	[250,91 - 250,95]
<i>canneal</i>	332,72	328,94	5,97	35,64	343,58	328,37	15,21	[330,49 - 334,95]	[326,66 - 331,22]
<i>dedup</i>	53,45	54,15	1,51	2,28	55,23	50,11	5,12	[52,89 - 54,02]	[52,96 - 55,34]
<i>facesim</i>	603,49	602,73	3,07	9,41	609,74	599,45	10,29	[602,34 - 604,63]	[600,90 - 604,55]
<i>ferret</i>	652,77	652,49	0,54	0,29	654,01	651,99	2,02	[652,57 - 652,97]	[652,17 - 652,81]
<i>fluidanimate</i>	583,42	583,39	0,14	0,02	583,83	583,18	0,65	[583,36 - 583,47]	[583,33 - 583,45]
<i>freqmine</i>	751,78	751,55	0,54	0,29	752,84	751,00	1,84	[751,58 - 751,98]	[751,18 - 751,92]
<i>streamcluster</i>	534,83	535,00	0,22	0,05	535,02	534,35	0,67	[534,75 - 534,92]	[534,89 - 535,10]
<i>swaptions</i>	557,31	547,02	19,86	394,59	605,38	542,10	63,28	[549,89 - 564,72]	[541,93 - 552,11]
<i>vips</i>	170,38	170,32	0,35	0,13	171,18	169,78	1,40	[170,25 - 170,51]	[170,17 - 170,47]
<i>x264</i>	130,75	130,69	0,54	0,29	132,92	130,09	2,83	[130,55 - 130,95]	[130,52 - 130,85]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 5 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	290,78	287,98	8,26	68,30	319,07	278,93	40,14	[287,69 - 293,86]	[287,70 - 288,26]
<i>bodytrack</i>	243,34	243,33	0,05	0,00	243,51	243,28	0,23	[243,33 - 243,36]	[243,32 - 243,34]
<i>canneal</i>	333,61	329,27	8,70	75,77	368,89	328,50	40,39	[330,36 - 336,86]	[328,67 - 329,87]
<i>dedup</i>	53,36	54,10	1,47	2,17	55,54	50,99	4,55	[52,81 - 53,91]	[52,95 - 55,24]
<i>facesim</i>	585,34	583,68	8,94	79,90	601,75	570,49	31,26	[582,00 - 588,68]	[578,48 - 588,87]
<i>ferret</i>	660,22	660,29	0,66	0,44	661,33	659,26	2,07	[659,98 - 660,47]	[659,91 - 660,66]
<i>fluidanimate</i>	576,67	576,71	0,33	0,11	577,40	576,02	1,38	[576,55 - 576,80]	[576,56 - 576,86]
<i>freqmine</i>	770,32	770,38	0,64	0,41	771,46	768,82	2,64	[770,08 - 770,56]	[770,19 - 770,57]
<i>streamcluster</i>	526,67	526,58	0,29	0,08	527,35	526,16	1,19	[526,56 - 526,78]	[526,38 - 526,78]
<i>swaptions</i>	553,30	548,36	17,42	303,32	594,19	536,02	58,17	[546,80 - 559,80]	[541,53 - 555,19]
<i>vips</i>	185,87	185,78	0,34	0,12	186,70	185,39	1,31	[185,75 - 186,00]	[185,62 - 185,93]
<i>x264</i>	131,35	131,28	0,37	0,14	132,25	130,68	1,57	[131,21 - 131,49]	[131,11 - 131,44]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 3 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,38	2,40	0,03	0,00	2,41	2,35	0,06	[2,37 - 2,39]	[2,38 - 2,42]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,79	40,03	1,69	2,85	41,57	36,72	4,85	[39,16 - 40,42]	[38,49 - 41,56]
<i>fmm</i>	2,39	2,42	0,04	0,00	2,44	2,30	0,14	[2,38 - 2,41]	[2,39 - 2,45]
<i>lu_cb</i>	6,33	6,19	0,23	0,05	6,60	6,12	0,48	[6,24 - 6,41]	[5,96 - 6,41]
<i>lu_ncb</i>	1,96	1,96	0,00	0,00	1,96	1,94	0,02	[1,96 - 1,96]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,18	0,00	[0,00 - 0,00]	[0,18 - 0,18]
<i>radiosity</i>	0,19	0,19	0,00	0,00	0,19	0,19	0,00	[0,19 - 0,19]	[0,19 - 0,19]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,89	0,89	0,00	0,00	0,89	0,88	0,01	[0,89 - 0,89]	[0,89 - 0,89]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,26	5,27	0,07	0,01	5,46	5,16	0,30	[5,24 - 5,29]	[5,24 - 5,29]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 3 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,35	2,37	0,03	0,00	2,37	2,31	0,06	[2,34 - 2,36]	[2,35 - 2,39]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,49	41,48	1,46	2,12	41,50	36,80	4,70	[39,95 - 41,03]	[40,76 - 42,20]
<i>fmm</i>	2,39	2,40	0,06	0,00	2,52	2,28	0,24	[2,37 - 2,42]	[2,40 - 2,40]
<i>lu_cb</i>	6,46	6,48	0,08	0,01	6,48	6,08	0,40	[6,43 - 6,49]	[6,48 - 6,48]
<i>lu_ncb</i>	1,83	1,79	0,08	0,01	2,01	1,79	0,22	[1,80 - 1,86]	[1,78 - 1,80]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>radiosity</i>	0,20	0,20	0,00	0,00	0,21	0,20	0,01	[0,20 - 0,20]	[0,20 - 0,20]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,88	0,88	0,01	0,00	0,88	0,83	0,05	[0,87 - 0,88]	[0,88 - 0,88]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,17	0,17	0,00	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,30	5,29	0,04	0,00	5,47	5,23	0,24	[5,28 - 5,32]	[5,28 - 5,30]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 5 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,38	2,40	0,03	0,00	2,41	2,35	0,06	[2,37 - 2,39]	[2,38 - 2,42]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,26	39,40	1,68	2,81	41,40	36,09	5,31	[38,63 - 39,88]	[38,50 - 40,30]
<i>fmm</i>	2,38	2,41	0,07	0,01	2,57	2,28	0,29	[2,36 - 2,41]	[2,38 - 2,44]
<i>lu_cb</i>	6,31	6,13	0,22	0,05	6,59	6,12	0,47	[6,22 - 6,39]	[5,91 - 6,35]
<i>lu_ncb</i>	1,96	1,96	0,00	0,00	1,96	1,96	0,00	[1,96 - 1,96]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,18	0,00	[0,00 - 0,00]	[0,18 - 0,18]
<i>radiosity</i>	0,19	0,19	0,00	0,00	0,19	0,19	0,00	[0,19 - 0,19]	[0,19 - 0,19]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,89	0,90	0,02	0,00	0,91	0,86	0,05	[0,88 - 0,90]	[0,89 - 0,91]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,12	5,13	0,05	0,00	5,25	5,04	0,21	[5,10 - 5,14]	[5,10 - 5,15]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE G: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #1 (CONT.)

	<i>Runlevel 5 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,33	2,32	0,03	0,00	2,38	2,26	0,12	[2,32 - 2,34]	[2,32 - 2,32]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,09	0,07	0,02	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	41,41	41,50	0,35	0,12	41,51	39,73	1,78	[41,28 - 41,54]	[41,50 - 41,50]
<i>fmm</i>	2,38	2,40	0,08	0,01	2,53	2,27	0,26	[2,35 - 2,41]	[2,35 - 2,44]
<i>lu_cb</i>	6,43	6,48	0,12	0,01	6,48	6,06	0,42	[6,39 - 6,47]	[6,48 - 6,48]
<i>lu_ncb</i>	1,96	2,01	0,09	0,01	2,01	1,79	0,22	[1,92 - 1,99]	[2,01 - 2,01]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>radiosity</i>	0,87	0,88	0,02	0,00	0,88	0,83	0,05	[0,86 - 0,87]	[0,88 - 0,88]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,20	0,20	0,00	0,00	0,20	0,20	0,00	[0,20 - 0,20]	[0,20 - 0,20]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,17	5,16	0,09	0,01	5,47	5,06	0,41	[5,14 - 5,21]	[5,13 - 5,19]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2

	<i>Tamanho de Variável 0 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,85	363,19	0,79	0,63	363,81	361,85	1,96	[362,55 - 363,14]	[362,45 - 363,93]
<b>CG</b>	87,78	87,79	0,07	0,00	87,84	87,53	0,31	[87,76 - 87,81]	[87,78 - 87,80]
<b>DC</b>	922,26	921,86	6,47	41,90	945,34	912,46	32,88	[919,84 - 924,67]	[920,25 - 923,46]
<b>EP</b>	239,67	239,57	0,44	0,19	241,96	239,53	2,43	[239,51 - 239,84]	[239,57 - 239,57]
<b>FT</b>	80,60	80,53	0,38	0,15	81,35	80,07	1,28	[80,46 - 80,75]	[80,33 - 80,73]
<b>IS</b>	13,03	13,03	0,00	0,00	13,04	13,02	0,02	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	315,86	315,93	0,59	0,35	317,08	314,74	2,34	[315,64 - 316,08]	[315,67 - 316,18]
<b>MG</b>	17,07	17,06	0,02	0,00	17,12	17,05	0,07	[17,06 - 17,08]	[17,05 - 17,07]
<b>SP</b>	294,77	294,77	1,96	3,84	297,93	291,69	6,24	[294,04 - 295,50]	[293,38 - 296,16]
<b>UA</b>	263,50	263,02	1,39	1,93	267,58	262,28	5,30	[262,98 - 264,02]	[262,49 - 263,55]
	<i>Tamanho de Variável 64 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	361,92	361,94	0,25	0,06	362,45	361,55	0,90	[361,83 - 362,02]	[361,79 - 362,09]
<b>CG</b>	87,79	87,80	0,05	0,00	87,83	87,55	0,28	[87,77 - 87,81]	[87,79 - 87,81]
<b>DC</b>	938,46	925,42	40,27	1621,94	1096,54	903,27	193,27	[923,43 - 953,50]	[920,88 - 929,96]
<b>EP</b>	240,81	239,58	3,73	13,90	251,80	239,54	12,26	[239,41 - 242,20]	[239,57 - 239,59]
<b>FT</b>	80,75	80,61	0,50	0,25	81,41	79,91	1,50	[80,56 - 80,93]	[80,19 - 81,02]
<b>IS</b>	13,02	13,02	0,00	0,00	13,03	13,02	0,01	[13,02 - 13,02]	[13,02 - 13,02]
<b>LU</b>	315,78	315,29	0,78	0,61	317,07	314,82	2,25	[315,49 - 316,07]	[314,68 - 315,89]
<b>MG</b>	17,16	17,18	0,05	0,00	17,21	17,04	0,17	[17,14 - 17,18]	[17,18 - 17,18]
<b>SP</b>	293,98	294,26	1,92	3,70	296,83	290,93	5,90	[293,26 - 294,70]	[292,97 - 295,55]
<b>UA</b>	263,46	263,11	1,13	1,28	268,08	262,59	5,49	[263,04 - 263,88]	[262,81 - 263,40]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

(Continua)

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 128 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,42	362,27	0,56	0,32	363,83	361,81	2,02	[362,21 - 362,63]	[362,09 - 362,44]
<b>CG</b>	87,73	87,72	0,01	0,00	87,74	87,72	0,02	[87,72 - 87,73]	[87,72 - 87,72]
<b>DC</b>	932,48	927,26	20,97	439,61	978,92	907,97	70,95	[924,65 - 940,31]	[920,64 - 933,87]
<b>EP</b>	239,62	239,60	0,05	0,00	239,78	239,57	0,21	[239,60 - 239,64]	[239,59 - 239,61]
<b>FT</b>	81,02	80,97	0,52	0,27	81,91	80,09	1,82	[80,82 - 81,21]	[80,64 - 81,30]
<b>IS</b>	13,03	13,03	0,00	0,00	13,03	13,02	0,01	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	315,79	315,76	0,64	0,41	317,11	314,77	2,34	[315,55 - 316,03]	[315,39 - 316,13]
<b>MG</b>	17,09	17,09	0,01	0,00	17,12	17,07	0,05	[17,08 - 17,09]	[17,08 - 17,09]
<b>SP</b>	295,31	295,18	1,43	2,03	297,69	291,25	6,44	[294,78 - 295,84]	[294,59 - 295,77]
<b>UA</b>	263,85	262,99	1,69	2,86	268,10	262,28	5,82	[263,22 - 264,48]	[262,36 - 263,62]
	<i>Tamanho de Variável 256 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,87	363,21	0,70	0,48	363,70	361,75	1,95	[362,61 - 363,13]	[362,69 - 363,72]
<b>CG</b>	87,79	87,79	0,02	0,00	87,83	87,77	0,06	[87,79 - 87,80]	[87,78 - 87,80]
<b>DC</b>	925,84	925,31	11,95	142,69	975,33	902,25	73,08	[921,38 - 930,30]	[922,81 - 927,81]
<b>EP</b>	240,02	239,61	2,23	4,96	251,81	239,56	12,25	[239,19 - 240,85]	[239,59 - 239,62]
<b>FT</b>	80,86	81,03	0,46	0,21	81,35	80,07	1,28	[80,69 - 81,03]	[80,70 - 81,36]
<b>IS</b>	13,03	13,03	0,00	0,00	13,04	13,02	0,02	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	316,27	316,27	0,59	0,35	317,40	315,02	2,38	[316,05 - 316,49]	[316,00 - 316,53]
<b>MG</b>	17,15	17,16	0,04	0,00	17,19	17,04	0,15	[17,13 - 17,16]	[17,16 - 17,16]
<b>SP</b>	295,96	296,16	1,77	3,12	298,29	291,32	6,97	[295,30 - 296,62]	[295,36 - 296,95]
<b>UA</b>	263,59	263,33	1,54	2,39	267,59	262,11	5,48	[263,01 - 264,17]	[262,95 - 263,71]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

<i>Tamanho de Variável 512 e Otimização 02</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,15	362,11	0,22	0,05	362,71	361,87	0,84	[362,07 - 362,24]	[361,98 - 362,23]
<b>CG</b>	87,85	87,84	0,02	0,00	87,88	87,83	0,05	[87,84 - 87,85]	[87,83 - 87,85]
<b>DC</b>	925,53	925,40	7,50	56,22	943,36	903,80	39,56	[922,73 - 928,33]	[923,25 - 927,55]
<b>EP</b>	240,44	239,61	3,09	9,56	251,82	239,58	12,24	[239,29 - 241,60]	[239,58 - 239,64]
<b>FT</b>	80,89	80,75	0,54	0,29	81,78	80,10	1,68	[80,69 - 81,09]	[80,44 - 81,05]
<b>IS</b>	13,03	13,03	0,00	0,00	13,04	13,02	0,02	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	315,30	315,26	0,27	0,07	316,21	314,71	1,50	[315,20 - 315,41]	[315,16 - 315,36]
<b>MG</b>	17,14	17,17	0,06	0,00	17,22	17,04	0,18	[17,11 - 17,16]	[17,13 - 17,21]
<b>SP</b>	294,53	294,69	1,97	3,88	297,79	290,87	6,92	[293,79 - 295,27]	[293,62 - 295,75]
<b>UA</b>	263,75	262,80	1,74	3,04	267,66	262,38	5,28	[263,10 - 264,40]	[262,03 - 263,56]
<i>Tamanho de Variável 1024 e Otimização 02</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,64	362,57	0,15	0,02	362,89	362,38	0,51	[362,58 - 362,69]	[362,46 - 362,68]
<b>CG</b>	87,83	87,83	0,01	0,00	87,85	87,83	0,02	[87,83 - 87,84]	[87,83 - 87,83]
<b>DC</b>	929,13	925,13	18,91	357,71	975,47	902,32	73,15	[922,07 - 936,19]	[922,11 - 928,14]
<b>EP</b>	240,02	239,61	2,22	4,95	251,80	239,56	12,24	[239,19 - 240,85]	[239,60 - 239,62]
<b>FT</b>	81,30	81,47	0,46	0,21	81,78	80,31	1,47	[81,13 - 81,47]	[81,29 - 81,65]
<b>IS</b>	13,03	13,03	0,00	0,00	13,04	13,02	0,02	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	315,47	315,45	0,51	0,26	317,10	314,72	2,38	[315,29 - 315,66]	[315,26 - 315,63]
<b>MG</b>	17,10	17,10	0,01	0,00	17,12	17,08	0,04	[17,10 - 17,11]	[17,10 - 17,10]
<b>SP</b>	295,77	296,17	1,37	1,87	297,48	291,37	6,11	[295,26 - 296,28]	[295,75 - 296,59]
<b>UA</b>	263,56	262,85	1,52	2,30	267,65	262,32	5,33	[263,00 - 264,13]	[262,28 - 263,42]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 2048 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,64	362,73	0,82	0,67	363,58	361,57	2,01	[362,34 - 362,95]	[361,96 - 363,49]
<b>CG</b>	87,79	87,79	0,03	0,00	87,83	87,65	0,18	[87,78 - 87,80]	[87,78 - 87,80]
<b>DC</b>	921,26	921,08	5,50	30,26	932,28	910,36	21,92	[919,21 - 923,32]	[917,67 - 924,49]
<b>EP</b>	239,56	239,54	0,07	0,01	239,79	239,51	0,28	[239,54 - 239,59]	[239,52 - 239,56]
<b>FT</b>	80,76	80,72	0,49	0,24	81,40	80,03	1,37	[80,57 - 80,94]	[80,33 - 81,10]
<b>IS</b>	13,03	13,03	0,00	0,00	13,03	13,02	0,01	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	316,06	315,95	0,63	0,40	317,76	315,08	2,68	[315,82 - 316,29]	[315,65 - 316,24]
<b>MG</b>	17,09	17,08	0,04	0,00	17,20	17,03	0,17	[17,07 - 17,11]	[17,06 - 17,10]
<b>SP</b>	295,42	296,40	2,16	4,64	298,60	291,21	7,39	[294,62 - 296,23]	[295,23 - 297,57]
<b>UA</b>	263,87	263,15	1,67	2,80	267,53	262,39	5,14	[263,24 - 264,49]	[262,49 - 263,80]
	<i>Tamanho de Variável 4096 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	362,51	362,50	0,17	0,03	362,99	362,25	0,74	[362,45 - 362,58]	[362,42 - 362,57]
<b>CG</b>	87,85	87,85	0,02	0,00	87,88	87,83	0,05	[87,84 - 87,86]	[87,84 - 87,86]
<b>DC</b>	917,35	915,98	10,79	116,49	942,68	894,82	47,86	[913,32 - 921,38]	[913,10 - 918,86]
<b>EP</b>	241,61	239,59	4,60	21,16	251,80	239,54	12,26	[239,89 - 243,33]	[239,57 - 239,61]
<b>FT</b>	80,87	81,11	0,56	0,31	81,47	80,07	1,40	[80,66 - 81,07]	[80,64 - 81,58]
<b>IS</b>	13,03	13,03	0,00	0,00	13,03	13,02	0,01	[13,03 - 13,03]	[13,03 - 13,03]
<b>LU</b>	317,11	317,14	0,63	0,40	318,55	316,15	2,40	[316,87 - 317,34]	[316,86 - 317,41]
<b>MG</b>	17,06	17,06	0,03	0,00	17,10	17,02	0,08	[17,05 - 17,07]	[17,05 - 17,07]
<b>SP</b>	294,84	294,88	1,67	2,78	297,58	291,52	6,06	[294,22 - 295,46]	[294,01 - 295,74]
<b>UA</b>	263,16	262,73	1,26	1,59	267,25	262,17	5,08	[262,69 - 263,63]	[262,39 - 263,06]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 0 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	384,90	384,75	0,32	0,10	385,60	384,52	1,08	[384,78 - 385,02]	[384,67 - 384,83]
<b>CG</b>	93,78	93,82	0,09	0,01	93,87	93,54	0,33	[93,74 - 93,81]	[93,77 - 93,87]
<b>DC</b>	952,46	946,97	21,71	471,46	1038,50	929,04	109,46	[944,35 - 960,57]	[943,20 - 950,74]
<b>EP</b>	240,15	240,01	0,36	0,13	241,83	239,93	1,90	[240,01 - 240,28]	[239,98 - 240,04]
<b>FT</b>	80,03	80,01	0,54	0,29	80,92	78,90	2,02	[79,83 - 80,23]	[79,80 - 80,22]
<b>IS</b>	13,01	13,01	0,00	0,00	13,01	13,00	0,01	[13,01 - 13,01]	[13,01 - 13,01]
<b>LU</b>	305,40	305,19	0,92	0,84	306,89	303,93	2,96	[305,06 - 305,75]	[304,44 - 305,94]
<b>MG</b>	19,67	19,67	0,02	0,00	19,70	19,65	0,05	[19,67 - 19,68]	[19,66 - 19,68]
<b>SP</b>	294,59	295,21	1,80	3,25	296,90	291,76	5,14	[293,92 - 295,26]	[293,62 - 296,79]
<b>UA</b>	192,28	191,63	1,41	2,00	195,37	190,82	4,55	[191,75 - 192,80]	[190,69 - 192,57]
	<i>Tamanho de Variável 64 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	385,15	385,45	0,51	0,26	385,77	384,43	1,34	[384,96 - 385,34]	[385,01 - 385,89]
<b>CG</b>	93,87	93,90	0,07	0,01	93,93	93,72	0,21	[93,84 - 93,89]	[93,88 - 93,91]
<b>DC</b>	955,70	950,12	23,58	555,82	1053,71	934,35	119,36	[946,89 - 964,50]	[947,06 - 953,17]
<b>EP</b>	240,45	240,06	0,96	0,92	243,60	240,01	3,59	[240,09 - 240,81]	[240,00 - 240,11]
<b>FT</b>	80,56	80,59	0,50	0,25	81,38	79,70	1,68	[80,37 - 80,74]	[80,31 - 80,87]
<b>IS</b>	13,01	13,01	0,01	0,00	13,03	13,01	0,02	[13,01 - 13,01]	[13,01 - 13,01]
<b>LU</b>	304,32	304,54	0,71	0,51	305,27	303,30	1,97	[304,06 - 304,59]	[303,96 - 305,11]
<b>MG</b>	19,74	19,76	0,04	0,00	19,79	19,65	0,14	[19,73 - 19,76]	[19,75 - 19,77]
<b>SP</b>	294,41	295,28	1,74	3,03	296,41	291,33	5,08	[293,76 - 295,06]	[294,01 - 296,54]
<b>UA</b>	192,21	191,71	1,28	1,65	195,53	191,02	4,51	[191,73 - 192,69]	[191,06 - 192,36]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 128 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	384,86	384,96	0,27	0,07	385,58	384,39	1,19	[384,76 - 384,97]	[384,80 - 385,11]
<b>CG</b>	93,62	93,62	0,03	0,00	93,76	93,58	0,18	[93,60 - 93,63]	[93,60 - 93,63]
<b>DC</b>	952,94	945,78	22,72	516,05	1039,86	929,77	110,09	[944,46 - 961,42]	[941,91 - 949,64]
<b>EP</b>	241,15	241,07	0,22	0,05	241,98	240,98	1,00	[241,07 - 241,24]	[241,06 - 241,08]
<b>FT</b>	80,60	80,80	0,53	0,29	81,30	79,20	2,10	[80,40 - 80,80]	[80,56 - 81,03]
<b>IS</b>	13,01	13,01	0,00	0,00	13,01	13,00	0,01	[13,01 - 13,01]	[13,01 - 13,01]
<b>LU</b>	305,39	305,39	0,60	0,36	306,58	304,00	2,58	[305,16 - 305,61]	[305,18 - 305,60]
<b>MG</b>	19,74	19,74	0,04	0,00	19,78	19,65	0,13	[19,72 - 19,75]	[19,71 - 19,77]
<b>SP</b>	295,62	295,81	1,08	1,17	296,93	291,81	5,12	[295,22 - 296,03]	[295,50 - 296,11]
<b>UA</b>	191,90	191,53	1,15	1,31	194,86	190,83	4,03	[191,47 - 192,33]	[190,86 - 192,20]
	<i>Tamanho de Variável 256 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	385,29	385,21	0,42	0,17	386,04	384,73	1,31	[385,14 - 385,45]	[384,90 - 385,51]
<b>CG</b>	93,65	93,65	0,04	0,00	93,71	93,49	0,22	[93,64 - 93,67]	[93,65 - 93,65]
<b>DC</b>	951,60	944,51	23,40	547,67	1017,44	922,63	94,81	[942,86 - 960,34]	[935,94 - 953,07]
<b>EP</b>	240,64	240,51	0,32	0,11	241,92	240,46	1,46	[240,52 - 240,76]	[240,48 - 240,53]
<b>FT</b>	80,17	80,24	0,48	0,23	81,00	79,15	1,85	[79,99 - 80,35]	[80,06 - 80,42]
<b>IS</b>	13,01	13,01	0,00	0,00	13,01	13,00	0,01	[13,01 - 13,01]	[13,01 - 13,01]
<b>LU</b>	304,65	304,56	0,61	0,37	305,65	303,51	2,14	[304,43 - 304,88]	[304,16 - 304,95]
<b>MG</b>	19,69	19,68	0,02	0,00	19,72	19,66	0,06	[19,68 - 19,69]	[19,67 - 19,69]
<b>SP</b>	295,16	295,30	1,62	2,63	297,99	292,07	5,92	[294,56 - 295,77]	[294,77 - 295,83]
<b>UA</b>	192,27	191,44	1,43	2,05	196,05	191,14	4,91	[191,74 - 192,80]	[190,87 - 192,01]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 512 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	384,69	384,71	0,22	0,05	385,18	384,21	0,97	[384,60 - 384,77]	[384,63 - 384,79]
<b>CG</b>	93,77	93,75	0,23	0,05	94,96	93,56	1,40	[93,69 - 93,86]	[93,74 - 93,76]
<b>DC</b>	965,26	956,68	30,56	933,62	1052,61	930,91	121,70	[953,85 - 976,67]	[948,68 - 964,67]
<b>EP</b>	240,16	240,05	0,27	0,08	241,04	239,96	1,08	[240,06 - 240,26]	[240,02 - 240,07]
<b>FT</b>	80,32	80,50	0,63	0,40	81,18	78,73	2,45	[80,09 - 80,56]	[80,15 - 80,84]
<b>IS</b>	13,00	13,00	0,00	0,00	13,01	13,00	0,01	[13,00 - 13,01]	[13,00 - 13,00]
<b>LU</b>	305,02	304,73	0,87	0,75	306,82	303,94	2,88	[304,70 - 305,35]	[304,22 - 305,24]
<b>MG</b>	19,67	19,67	0,01	0,00	19,70	19,66	0,04	[19,67 - 19,68]	[19,67 - 19,67]
<b>SP</b>	295,62	296,17	1,53	2,35	297,32	292,38	4,94	[295,04 - 296,19]	[295,83 - 296,50]
<b>UA</b>	191,90	191,23	1,23	1,51	195,34	190,82	4,52	[191,44 - 192,36]	[190,79 - 191,66]
	<i>Tamanho de Variável 1024 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	385,26	385,26	0,27	0,07	385,68	384,78	0,90	[385,16 - 385,36]	[385,07 - 385,44]
<b>CG</b>	93,83	93,85	0,04	0,00	93,88	93,70	0,18	[93,82 - 93,85]	[93,84 - 93,85]
<b>DC</b>	964,81	954,37	31,44	988,78	1067,60	929,72	137,88	[953,06 - 976,55]	[942,77 - 965,96]
<b>EP</b>	240,54	240,37	0,44	0,19	241,92	240,33	1,59	[240,37 - 240,70]	[240,35 - 240,38]
<b>FT</b>	80,11	80,30	0,61	0,37	81,03	78,70	2,33	[79,88 - 80,34]	[79,89 - 80,70]
<b>IS</b>	13,01	13,01	0,01	0,00	13,09	13,00	0,09	[13,01 - 13,02]	[13,01 - 13,01]
<b>LU</b>	304,93	305,13	0,73	0,54	305,88	303,51	2,37	[304,65 - 305,20]	[304,68 - 305,57]
<b>MG</b>	19,65	19,62	0,06	0,00	19,73	19,58	0,15	[19,63 - 19,67]	[19,57 - 19,67]
<b>SP</b>	295,55	295,82	1,23	1,52	297,48	292,00	5,48	[295,09 - 296,01]	[295,47 - 296,17]
<b>UA</b>	192,52	192,20	1,49	2,22	195,55	190,73	4,82	[191,96 - 193,08]	[191,30 - 193,09]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 2048 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	385,59	385,55	0,12	0,01	385,88	385,50	0,38	[385,54 - 385,63]	[385,53 - 385,56]
<b>CG</b>	93,78	93,79	0,04	0,00	93,82	93,63	0,19	[93,76 - 93,80]	[93,79 - 93,79]
<b>DC</b>	967,41	953,84	29,92	895,25	1039,66	928,48	111,18	[956,24 - 978,58]	[936,09 - 971,59]
<b>EP</b>	240,13	239,89	0,61	0,37	242,47	239,82	2,65	[239,90 - 240,36]	[239,78 - 239,99]
<b>FT</b>	80,51	80,58	0,61	0,37	81,45	79,44	2,01	[80,29 - 80,74]	[80,24 - 80,92]
<b>IS</b>	13,01	13,01	0,01	0,00	13,03	13,01	0,02	[13,01 - 13,02]	[13,01 - 13,01]
<b>LU</b>	305,33	305,08	0,64	0,41	306,68	304,48	2,20	[305,09 - 305,57]	[304,89 - 305,26]
<b>MG</b>	19,67	19,66	0,01	0,00	19,70	19,65	0,05	[19,66 - 19,67]	[19,66 - 19,66]
<b>SP</b>	295,24	295,45	1,45	2,09	297,30	292,44	4,86	[294,70 - 295,78]	[294,82 - 296,07]
<b>UA</b>	195,20	195,08	1,54	2,36	198,08	193,45	4,63	[194,62 - 195,77]	[194,03 - 196,12]
	<i>Tamanho de Variável 4096 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	384,99	385,01	0,07	0,00	385,08	384,84	0,24	[384,96 - 385,01]	[384,98 - 385,03]
<b>CG</b>	93,76	93,60	0,67	0,46	97,31	93,59	3,72	[93,50 - 94,01]	[93,59 - 93,61]
<b>DC</b>	951,32	948,05	13,88	192,73	984,71	925,88	58,83	[946,13 - 956,50]	[943,41 - 952,68]
<b>EP</b>	240,31	240,01	0,63	0,40	242,28	239,96	2,32	[240,07 - 240,54]	[239,97 - 240,04]
<b>FT</b>	80,22	80,17	0,65	0,43	81,74	78,98	2,76	[79,97 - 80,46]	[79,96 - 80,38]
<b>IS</b>	13,01	13,01	0,00	0,00	13,01	13,01	0,00	[13,01 - 13,01]	[13,01 - 13,01]
<b>LU</b>	304,47	304,23	0,54	0,29	305,54	303,58	1,96	[304,27 - 304,67]	[303,95 - 304,50]
<b>MG</b>	19,66	19,66	0,01	0,00	19,70	19,65	0,05	[19,66 - 19,67]	[19,66 - 19,66]
<b>SP</b>	295,32	295,67	1,45	2,11	297,78	291,54	6,24	[294,78 - 295,87]	[294,98 - 296,35]
<b>UA</b>	192,00	191,24	1,33	1,78	195,40	190,83	4,57	[191,50 - 192,50]	[190,71 - 191,76]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 0 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,37	2,35	0,03	0,00	2,41	2,35	0,06	[2,36 - 2,38]	[2,32 - 2,38]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,09	0,08	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,17	40,76	1,45	2,10	41,57	36,63	4,94	[39,63 - 40,71]	[39,68 - 41,83]
<i>fmm</i>	2,35	2,34	0,06	0,00	2,51	2,28	0,23	[2,32 - 2,37]	[2,30 - 2,37]
<i>lu_cb</i>	6,38	6,47	0,22	0,05	6,60	6,12	0,48	[6,29 - 6,46]	[6,25 - 6,69]
<i>lu_ncb</i>	2,17	2,18	0,02	0,00	2,19	2,10	0,09	[2,17 - 2,18]	[2,18 - 2,18]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,19	0,18	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,89	0,89	0,02	0,00	0,97	0,86	0,11	[0,88 - 0,89]	[0,88 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,00	0,00	0,20	0,19	0,01	[0,19 - 0,19]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,02	0,00	0,26	0,17	0,09	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,15	5,14	0,09	0,01	5,49	5,06	0,43	[5,12 - 5,18]	[5,11 - 5,16]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	Tamanho de Variável 64 e Otimização 02								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,37	2,36	0,03	0,00	2,42	2,30	0,12	[2,36 - 2,38]	[2,33 - 2,38]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,65	39,48	1,59	2,52	41,60	36,40	5,20	[39,06 - 40,25]	[38,33 - 40,62]
<i>fmm</i>	2,36	2,35	0,06	0,00	2,49	2,28	0,21	[2,34 - 2,38]	[2,31 - 2,39]
<i>lu_cb</i>	6,31	6,19	0,21	0,05	6,60	6,12	0,48	[6,23 - 6,39]	[5,96 - 6,41]
<i>lu_ncb</i>	2,17	2,18	0,02	0,00	2,19	2,06	0,13	[2,16 - 2,18]	[2,18 - 2,18]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,01	0,00	0,19	0,18	0,01	[0,18 - 0,19]	[0,18 - 0,18]
<i>radiosity</i>	0,90	0,90	0,01	0,00	0,93	0,86	0,07	[0,89 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,00	0,00	0,20	0,19	0,01	[0,19 - 0,19]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,65	0,64	0,01	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,13	5,12	0,05	0,00	5,26	5,04	0,22	[5,11 - 5,15]	[5,10 - 5,14]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 128 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,37	2,35	0,03	0,00	2,41	2,34	0,07	[2,36 - 2,38]	[2,33 - 2,37]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,79	39,44	1,52	2,32	41,55	36,44	5,11	[39,22 - 40,36]	[38,04 - 40,83]
<i>fmm</i>	2,35	2,34	0,06	0,00	2,54	2,28	0,26	[2,33 - 2,37]	[2,30 - 2,38]
<i>lu_cb</i>	6,43	6,59	0,21	0,05	6,60	6,12	0,48	[6,35 - 6,51]	[6,47 - 6,71]
<i>lu_ncb</i>	1,95	1,96	0,03	0,00	1,97	1,77	0,20	[1,94 - 1,97]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,19	0,18	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,89	0,90	0,02	0,00	0,91	0,86	0,05	[0,88 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,01	0,00	0,20	0,19	0,01	[0,19 - 0,20]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,65	0,64	0,01	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,17	0,17	0,00	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,12	5,12	0,05	0,00	5,25	5,02	0,23	[5,10 - 5,14]	[5,11 - 5,13]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 256 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,38	2,37	0,03	0,00	2,41	2,30	0,11	[2,37 - 2,39]	[2,34 - 2,40]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,34	39,30	2,00	4,02	41,48	36,43	5,05	[38,59 - 40,09]	[37,72 - 40,88]
<i>fmm</i>	2,33	2,31	0,04	0,00	2,41	2,28	0,13	[2,31 - 2,34]	[2,29 - 2,33]
<i>lu_cb</i>	6,40	6,58	0,23	0,05	6,60	6,12	0,48	[6,31 - 6,48]	[6,36 - 6,80]
<i>lu_ncb</i>	1,96	1,96	0,00	0,00	1,97	1,96	0,01	[1,96 - 1,96]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,19	0,18	0,01	[0,18 - 0,19]	[0,18 - 0,18]
<i>radiosity</i>	0,89	0,90	0,02	0,00	0,91	0,86	0,05	[0,89 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,20	0,20	0,02	0,00	0,28	0,19	0,09	[0,19 - 0,20]	[0,20 - 0,20]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,66	0,64	0,02	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,11	5,11	0,06	0,00	5,30	5,05	0,25	[5,09 - 5,14]	[5,09 - 5,12]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 512 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,37	2,38	0,03	0,00	2,40	2,30	0,10	[2,36 - 2,38]	[2,35 - 2,40]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,09	0,07	0,02	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,50	41,37	1,45	2,11	41,63	37,02	4,61	[39,96 - 41,05]	[40,33 - 42,40]
<i>fmm</i>	2,41	2,40	0,08	0,01	2,53	2,29	0,24	[2,38 - 2,44]	[2,35 - 2,45]
<i>lu_cb</i>	6,36	6,37	0,23	0,05	6,60	6,12	0,48	[6,28 - 6,45]	[6,14 - 6,59]
<i>lu_ncb</i>	1,96	1,97	0,05	0,00	1,97	1,73	0,24	[1,94 - 1,98]	[1,97 - 1,97]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,19	0,19	0,01	0,00	0,19	0,18	0,01	[0,18 - 0,19]	[0,18 - 0,19]
<i>radiosity</i>	0,90	0,90	0,01	0,00	0,91	0,88	0,03	[0,90 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,00	0,00	0,20	0,19	0,01	[0,19 - 0,19]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,18	0,18	0,00	0,00	0,18	0,17	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>water_spatial</i>	5,21	5,20	0,10	0,01	5,49	5,08	0,41	[5,17 - 5,25]	[5,16 - 5,23]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 1024 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,38	2,40	0,03	0,00	2,41	2,35	0,06	[2,37 - 2,39]	[2,37 - 2,43]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,47	39,40	1,56	2,43	41,44	36,11	5,33	[38,89 - 40,05]	[38,83 - 39,97]
<i>fmm</i>	2,37	2,40	0,05	0,00	2,44	2,28	0,16	[2,35 - 2,39]	[2,36 - 2,43]
<i>lu_cb</i>	6,32	6,19	0,21	0,05	6,59	6,12	0,47	[6,24 - 6,40]	[5,98 - 6,39]
<i>lu_ncb</i>	1,97	1,97	0,02	0,00	1,97	1,88	0,09	[1,96 - 1,97]	[1,97 - 1,97]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,19	0,18	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,88	0,89	0,02	0,00	0,93	0,86	0,07	[0,88 - 0,89]	[0,88 - 0,89]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,00	0,00	0,20	0,19	0,01	[0,19 - 0,19]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,11	5,11	0,04	0,00	5,23	5,03	0,20	[5,10 - 5,13]	[5,09 - 5,13]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 2048 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,41	2,39	0,04	0,00	2,47	2,37	0,10	[2,40 - 2,42]	[2,39 - 2,39]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,08	0,00	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,48	39,08	1,47	2,16	41,58	36,42	5,16	[38,93 - 40,02]	[38,49 - 39,67]
<i>fmm</i>	2,35	2,34	0,04	0,00	2,43	2,29	0,14	[2,33 - 2,36]	[2,30 - 2,37]
<i>lu_cb</i>	6,24	6,13	0,20	0,04	6,60	6,12	0,48	[6,17 - 6,32]	[6,13 - 6,13]
<i>lu_ncb</i>	1,96	1,96	0,01	0,00	1,97	1,93	0,04	[1,96 - 1,96]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,19	0,18	0,01	[0,18 - 0,19]	[0,18 - 0,18]
<i>radiosity</i>	0,89	0,90	0,02	0,00	0,91	0,86	0,05	[0,89 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,20	0,20	0,00	0,00	0,21	0,20	0,01	[0,20 - 0,20]	[0,20 - 0,20]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,65	0,64	0,01	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,18	0,18	0,00	0,00	0,19	0,17	0,02	[0,18 - 0,18]	[0,18 - 0,18]
<i>water_spatial</i>	5,26	5,25	0,12	0,02	5,65	5,11	0,54	[5,21 - 5,31]	[5,20 - 5,30]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 4096 e Otimização 02</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,39	2,40	0,03	0,00	2,41	2,35	0,06	[2,38 - 2,40]	[2,38 - 2,42]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,09	0,07	0,02	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,25	40,35	1,95	3,80	42,84	36,52	6,32	[39,52 - 40,98]	[39,33 - 41,37]
<i>fmm</i>	2,39	2,40	0,05	0,00	2,55	2,27	0,28	[2,37 - 2,41]	[2,38 - 2,42]
<i>lu_cb</i>	6,32	6,15	0,22	0,05	6,60	6,12	0,48	[6,23 - 6,40]	[5,92 - 6,38]
<i>lu_ncb</i>	1,96	1,96	0,00	0,00	1,97	1,96	0,01	[1,96 - 1,96]	[1,96 - 1,96]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,01	0,00	0,20	0,18	0,02	[0,18 - 0,19]	[0,18 - 0,18]
<i>radiosity</i>	0,90	0,90	0,01	0,00	0,91	0,86	0,05	[0,89 - 0,90]	[0,90 - 0,90]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,19	0,19	0,00	0,00	0,20	0,19	0,01	[0,19 - 0,19]	[0,19 - 0,19]
<i>volrend</i>	0,64	0,64	0,00	0,00	0,64	0,64	0,00	[0,64 - 0,64]	[0,64 - 0,64]
<i>water_nsquared</i>	0,17	0,17	0,01	0,00	0,21	0,17	0,04	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,09	5,08	0,04	0,00	5,21	5,02	0,19	[5,07 - 5,10]	[5,07 - 5,09]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 0 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,34	2,32	0,03	0,00	2,39	2,26	0,13	[2,33 - 2,35]	[2,29 - 2,35]
<i>cholesky</i>	0,12	0,08	0,20	0,04	1,18	0,07	1,11	[0,04 - 0,19]	[0,08 - 0,08]
<i>fft</i>	40,42	41,55	1,40	1,97	41,56	38,09	3,47	[39,90 - 40,95]	[40,44 - 42,65]
<i>fmm</i>	2,39	2,40	0,06	0,00	2,54	2,28	0,26	[2,37 - 2,42]	[2,40 - 2,40]
<i>lu_cb</i>	6,48	6,48	0,03	0,00	6,48	6,33	0,15	[6,46 - 6,49]	[6,48 - 6,48]
<i>lu_ncb</i>	1,85	1,80	0,09	0,01	2,01	1,79	0,22	[1,82 - 1,89]	[1,77 - 1,82]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,18	0,00	[0,00 - 0,00]	[0,18 - 0,18]
<i>radiosity</i>	0,87	0,87	0,02	0,00	0,95	0,83	0,12	[0,86 - 0,88]	[0,87 - 0,87]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,20	0,20	0,00	0,00	0,21	0,20	0,01	[0,20 - 0,20]	[0,20 - 0,20]
<i>volrend</i>	0,63	0,63	0,01	0,00	0,67	0,63	0,04	[0,63 - 0,64]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,13	5,11	0,08	0,01	5,33	5,04	0,29	[5,10 - 5,16]	[5,07 - 5,15]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 64 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,34	2,33	0,03	0,00	2,39	2,26	0,13	[2,33 - 2,36]	[2,30 - 2,35]
<i>cholesky</i>	0,08	0,08	0,02	0,00	0,20	0,07	0,13	[0,08 - 0,09]	[0,08 - 0,08]
<i>fft</i>	39,67	38,98	1,33	1,76	41,53	37,66	3,87	[39,17 - 40,16]	[38,04 - 39,92]
<i>fmm</i>	2,43	2,41	0,06	0,00	2,59	2,29	0,30	[2,41 - 2,45]	[2,41 - 2,41]
<i>lu_cb</i>	6,47	6,48	0,05	0,00	6,48	6,27	0,21	[6,45 - 6,48]	[6,48 - 6,48]
<i>lu_ncb</i>	1,86	1,79	0,10	0,01	2,01	1,79	0,22	[1,82 - 1,89]	[1,76 - 1,82]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,18	0,00	[0,00 - 0,00]	[0,18 - 0,18]
<i>radiosity</i>	0,87	0,88	0,02	0,00	0,92	0,83	0,09	[0,87 - 0,88]	[0,88 - 0,88]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,00	0,00	0,21	0,20	0,01	[0,20 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,18	5,17	0,07	0,01	5,37	5,08	0,29	[5,16 - 5,21]	[5,14 - 5,20]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 128 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,39	2,37	0,04	0,00	2,45	2,31	0,14	[2,38 - 2,41]	[2,34 - 2,40]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,07 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,51	41,14	1,19	1,41	41,56	38,23	3,33	[40,07 - 40,96]	[40,30 - 41,97]
<i>fmm</i>	2,39	2,40	0,06	0,00	2,55	2,28	0,27	[2,37 - 2,42]	[2,40 - 2,40]
<i>lu_cb</i>	6,46	6,48	0,08	0,01	6,48	6,13	0,35	[6,43 - 6,48]	[6,48 - 6,48]
<i>lu_ncb</i>	1,89	1,84	0,10	0,01	2,01	1,79	0,22	[1,85 - 1,93]	[1,73 - 1,94]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,20	0,17	0,03	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,85	0,85	0,01	0,00	0,88	0,83	0,05	[0,84 - 0,85]	[0,84 - 0,85]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,00	0,00	0,22	0,20	0,02	[0,21 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,19	5,19	0,07	0,00	5,32	5,08	0,24	[5,16 - 5,21]	[5,15 - 5,22]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCRIPTIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 256 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,34	2,32	0,03	0,00	2,38	2,29	0,09	[2,33 - 2,35]	[2,30 - 2,34]
<i>cholesky</i>	0,07	0,07	0,01	0,00	0,09	0,07	0,02	[0,07 - 0,08]	[0,07 - 0,07]
<i>fft</i>	40,46	41,44	1,31	1,72	41,54	37,94	3,60	[39,97 - 40,95]	[40,83 - 42,05]
<i>fmm</i>	2,38	2,40	0,07	0,00	2,53	2,27	0,26	[2,36 - 2,41]	[2,40 - 2,40]
<i>lu_cb</i>	6,45	6,48	0,09	0,01	6,48	6,14	0,34	[6,42 - 6,49]	[6,48 - 6,48]
<i>lu_ncb</i>	1,88	1,84	0,10	0,01	2,01	1,79	0,22	[1,84 - 1,92]	[1,74 - 1,94]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,18	0,00	[0,00 - 0,00]	[0,18 - 0,18]
<i>radiosity</i>	0,87	0,87	0,05	0,00	1,12	0,83	0,29	[0,85 - 0,89]	[0,86 - 0,88]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,01	0,00	0,22	0,20	0,02	[0,21 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,64	0,63	0,03	0,00	0,78	0,63	0,15	[0,63 - 0,65]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,01	0,00	0,20	0,17	0,03	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,24	5,20	0,10	0,01	5,50	5,14	0,36	[5,20 - 5,27]	[5,16 - 5,24]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 512 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,34	2,33	0,03	0,00	2,38	2,31	0,07	[2,33 - 2,35]	[2,31 - 2,35]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,88	39,69	1,44	2,08	41,58	38,00	3,58	[39,34 - 40,42]	[38,19 - 41,19]
<i>fmm</i>	2,40	2,40	0,05	0,00	2,53	2,30	0,23	[2,38 - 2,42]	[2,39 - 2,40]
<i>lu_cb</i>	6,46	6,48	0,08	0,01	6,48	6,06	0,42	[6,43 - 6,49]	[6,48 - 6,48]
<i>lu_ncb</i>	1,85	1,81	0,08	0,01	2,01	1,79	0,22	[1,82 - 1,88]	[1,77 - 1,84]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,17	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,86	0,86	0,02	0,00	0,88	0,83	0,05	[0,85 - 0,87]	[0,85 - 0,87]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,00	0,00	0,21	0,20	0,01	[0,21 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,00	0,00	0,18	0,17	0,01	[0,17 - 0,17]	[0,17 - 0,17]
<i>water_spatial</i>	5,13	5,12	0,06	0,00	5,29	5,05	0,24	[5,11 - 5,16]	[5,10 - 5,13]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 1024 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,41	2,39	0,04	0,00	2,46	2,33	0,13	[2,39 - 2,42]	[2,36 - 2,42]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,05	40,55	1,57	2,47	41,56	36,33	5,23	[39,46 - 40,64]	[39,54 - 41,55]
<i>fmm</i>	2,40	2,39	0,07	0,00	2,53	2,28	0,25	[2,38 - 2,43]	[2,39 - 2,39]
<i>lu_cb</i>	6,34	6,43	0,15	0,02	6,48	6,10	0,38	[6,29 - 6,40]	[6,31 - 6,54]
<i>lu_ncb</i>	1,87	1,82	0,09	0,01	2,01	1,79	0,22	[1,84 - 1,91]	[1,77 - 1,87]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,17	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,87	0,87	0,02	0,00	0,88	0,83	0,05	[0,86 - 0,87]	[0,87 - 0,87]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,00	0,00	0,22	0,20	0,02	[0,21 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,65	0,63	0,02	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,18	0,18	0,01	0,00	0,18	0,17	0,01	[0,17 - 0,18]	[0,17 - 0,18]
<i>water_spatial</i>	5,14	5,13	0,06	0,00	5,27	5,06	0,21	[5,11 - 5,16]	[5,10 - 5,16]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 2048 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,35	2,37	0,03	0,00	2,37	2,31	0,06	[2,34 - 2,36]	[2,35 - 2,39]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	40,20	40,67	1,34	1,80	41,54	37,71	3,83	[39,70 - 40,70]	[39,55 - 41,79]
<i>fmm</i>	2,31	2,32	0,04	0,00	2,39	2,26	0,13	[2,30 - 2,33]	[2,29 - 2,35]
<i>lu_cb</i>	6,46	6,48	0,07	0,00	6,48	6,12	0,36	[6,43 - 6,48]	[6,48 - 6,48]
<i>lu_ncb</i>	1,89	1,84	0,10	0,01	2,01	1,79	0,22	[1,85 - 1,93]	[1,74 - 1,94]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,00	0,00	0,18	0,17	0,01	[0,18 - 0,18]	[0,18 - 0,18]
<i>radiosity</i>	0,86	0,85	0,02	0,00	0,92	0,83	0,09	[0,85 - 0,86]	[0,84 - 0,86]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,21	0,21	0,00	0,00	0,22	0,21	0,01	[0,21 - 0,21]	[0,21 - 0,21]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,18	0,18	0,00	0,00	0,19	0,17	0,02	[0,18 - 0,18]	[0,18 - 0,18]
<i>water_spatial</i>	5,22	5,19	0,11	0,01	5,46	5,11	0,35	[5,18 - 5,26]	[5,15 - 5,22]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE H: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #2 (CONT.)

	<i>Tamanho de Variável 4096 e Otimização 03</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>barnes</i>	2,35	2,36	0,03	0,00	2,39	2,28	0,11	[2,34 - 2,36]	[2,33 - 2,38]
<i>cholesky</i>	0,08	0,08	0,00	0,00	0,08	0,07	0,01	[0,08 - 0,08]	[0,08 - 0,08]
<i>fft</i>	39,79	39,25	1,44	2,07	41,50	37,92	3,58	[39,25 - 40,33]	[37,87 - 40,63]
<i>fmm</i>	2,34	2,33	0,06	0,00	2,52	2,26	0,26	[2,32 - 2,37]	[2,31 - 2,35]
<i>lu_cb</i>	6,47	6,48	0,04	0,00	6,48	6,31	0,17	[6,45 - 6,48]	[6,48 - 6,48]
<i>lu_ncb</i>	1,84	1,80	0,08	0,01	2,01	1,79	0,22	[1,81 - 1,87]	[1,79 - 1,81]
<i>ocean_cp</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>ocean_ncp</i>	0,18	0,18	0,02	0,00	0,28	0,18	0,10	[0,18 - 0,19]	[0,18 - 0,18]
<i>radiosity</i>	0,87	0,87	0,01	0,00	0,89	0,86	0,03	[0,87 - 0,87]	[0,87 - 0,87]
<i>radix</i>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	[0,00 - 0,00]	[0,00 - 0,00]
<i>raytrace</i>	0,20	0,20	0,00	0,00	0,21	0,20	0,01	[0,20 - 0,20]	[0,20 - 0,20]
<i>volrend</i>	0,63	0,63	0,00	0,00	0,64	0,63	0,01	[0,63 - 0,63]	[0,63 - 0,63]
<i>water_nsquared</i>	0,17	0,17	0,01	0,00	0,21	0,17	0,04	[0,17 - 0,18]	[0,17 - 0,17]
<i>water_spatial</i>	5,09	5,09	0,04	0,00	5,18	5,02	0,16	[5,07 - 5,10]	[5,07 - 5,11]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3

	Sem Affinity com 2 Threads								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	195,27	195,20	0,33	0,11	196,07	194,77	1,30	[195,15 - 195,39]	[195,05 - 195,35]
<b>CG</b>	44,33	44,34	0,05	0,00	44,39	44,20	0,19	[44,31 - 44,34]	[44,33 - 44,35]
<b>DC</b>	524,65	521,06	9,45	89,36	549,77	511,04	38,73	[521,12 - 528,18]	[517,64 - 524,48]
<b>EP</b>	67,03	67,75	1,42	2,01	67,82	63,39	4,43	[66,50 - 67,56]	[67,66 - 67,84]
<b>FT</b>	41,19	41,05	0,47	0,22	42,07	40,56	1,51	[41,02 - 41,36]	[40,67 - 41,43]
<b>IS</b>	6,96	6,96	0,00	0,00	6,97	6,96	0,01	[6,96 - 6,97]	[6,96 - 6,96]
<b>LU</b>	147,71	147,79	0,46	0,21	148,48	146,82	1,66	[147,54 - 147,88]	[147,50 - 148,08]
<b>MG</b>	8,88	8,88	0,00	0,00	8,89	8,87	0,02	[8,88 - 8,88]	[8,88 - 8,88]
<b>SP</b>	154,20	154,00	1,21	1,46	157,06	151,65	5,41	[153,75 - 154,65]	[153,35 - 154,65]
<b>UA</b>	186,54	186,49	0,56	0,32	188,42	185,64	2,78	[186,33 - 186,75]	[186,30 - 186,67]
	Sem Affinity com 4 Threads								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	99,55	99,23	0,73	0,53	101,04	98,72	2,32	[99,28 - 99,82]	[98,83 - 99,63]
<b>CG</b>	22,77	22,77	0,02	0,00	22,79	22,72	0,07	[22,76 - 22,77]	[22,76 - 22,78]
<b>DC</b>	307,62	306,69	8,75	76,53	330,30	292,14	38,16	[304,36 - 310,89]	[302,25 - 311,13]
<b>EP</b>	35,51	34,25	1,85	3,43	37,60	33,64	3,96	[34,82 - 36,21]	[32,47 - 36,03]
<b>FT</b>	21,03	20,95	0,21	0,05	21,87	20,80	1,07	[20,95 - 21,11]	[20,89 - 21,01]
<b>IS</b>	3,56	3,56	0,00	0,00	3,57	3,55	0,02	[3,56 - 3,56]	[3,56 - 3,56]
<b>LU</b>	75,72	75,91	1,69	2,87	79,01	72,91	6,10	[75,09 - 76,35]	[74,70 - 77,11]
<b>MG</b>	4,47	4,47	0,00	0,00	4,48	4,47	0,01	[4,47 - 4,47]	[4,47 - 4,47]
<b>SP</b>	81,21	80,96	0,92	0,84	84,24	79,87	4,37	[80,86 - 81,55]	[80,74 - 81,17]
<b>UA</b>	98,36	98,35	0,43	0,18	99,34	97,62	1,72	[98,20 - 98,52]	[98,24 - 98,45]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

(Continua)

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

	Sem Affinity com 6 Threads								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	67,88	67,87	0,10	0,01	68,10	67,70	0,40	[67,85 - 67,92]	[67,80 - 67,93]
<b>CG</b>	16,84	16,75	0,16	0,03	17,19	16,60	0,59	[16,78 - 16,90]	[16,65 - 16,85]
<b>DC</b>	232,21	231,18	9,58	91,71	255,46	215,95	39,51	[228,63 - 235,78]	[224,85 - 237,50]
<b>EP</b>	23,40	22,59	1,16	1,34	25,08	22,44	2,64	[22,97 - 23,83]	[22,00 - 23,18]
<b>FT</b>	15,54	15,54	0,19	0,04	16,15	15,20	0,95	[15,47 - 15,62]	[15,47 - 15,61]
<b>IS</b>	2,49	2,49	0,01	0,00	2,52	2,48	0,04	[2,49 - 2,49]	[2,49 - 2,49]
<b>LU</b>	52,41	52,10	2,59	6,72	57,40	48,59	8,81	[51,44 - 53,37]	[50,35 - 53,84]
<b>MG</b>	3,31	3,31	0,03	0,00	3,35	3,27	0,08	[3,30 - 3,32]	[3,29 - 3,33]
<b>SP</b>	62,86	62,74	0,78	0,61	64,14	61,57	2,57	[62,56 - 63,15]	[62,37 - 63,10]
<b>UA</b>	70,33	70,38	0,54	0,30	71,32	69,31	2,01	[70,13 - 70,53]	[70,15 - 70,60]
	Compact com 2 Threads								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	271,93	272,10	0,63	0,40	272,75	271,05	1,70	[271,70 - 272,17]	[271,58 - 272,61]
<b>CG</b>	68,08	68,07	0,09	0,01	68,25	67,73	0,52	[68,05 - 68,12]	[68,04 - 68,10]
<b>DC</b>	371,35	373,50	5,06	25,64	379,07	360,90	18,17	[369,46 - 373,25]	[369,85 - 377,15]
<b>EP</b>	64,61	64,52	0,16	0,03	65,02	64,51	0,51	[64,55 - 64,67]	[64,49 - 64,55]
<b>FT</b>	59,85	59,74	0,78	0,61	61,69	58,44	3,25	[59,56 - 60,14]	[59,44 - 60,04]
<b>IS</b>	9,15	9,15	0,01	0,00	9,15	9,14	0,01	[9,14 - 9,15]	[9,15 - 9,15]
<b>LU</b>	228,01	227,84	0,66	0,43	230,10	227,02	3,08	[227,76 - 228,25]	[227,50 - 228,18]
<b>MG</b>	12,85	12,84	0,08	0,01	12,96	12,74	0,22	[12,82 - 12,88]	[12,78 - 12,89]
<b>SP</b>	243,91	243,73	1,71	2,93	247,29	241,72	5,57	[243,27 - 244,55]	[242,83 - 244,63]
<b>UA</b>	238,03	237,51	1,83	3,36	241,02	235,97	5,05	[237,35 - 238,72]	[236,22 - 238,80]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

	<i>Compact com 4 Threads</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	148,71	148,72	0,09	0,01	148,90	148,48	0,42	[148,68 - 148,74]	[148,66 - 148,77]
<b>CG</b>	41,75	41,79	0,09	0,01	41,85	41,55	0,30	[41,72 - 41,79]	[41,77 - 41,81]
<b>DC</b>	346,47	345,73	4,73	22,35	356,62	339,26	17,36	[344,70 - 348,23]	[342,93 - 348,52]
<b>EP</b>	34,97	34,67	0,73	0,53	37,69	34,55	3,14	[34,70 - 35,24]	[34,50 - 34,83]
<b>FT</b>	32,94	33,00	0,16	0,02	33,16	32,50	0,66	[32,88 - 32,99]	[32,93 - 33,07]
<b>IS</b>	4,64	4,64	0,00	0,00	4,65	4,63	0,02	[4,64 - 4,64]	[4,64 - 4,64]
<b>LU</b>	112,61	112,61	0,08	0,01	112,80	112,51	0,29	[112,59 - 112,64]	[112,57 - 112,64]
<b>MG</b>	8,74	8,79	0,18	0,03	8,92	8,42	0,50	[8,67 - 8,80]	[8,69 - 8,88]
<b>SP</b>	162,05	162,12	0,22	0,05	162,38	161,53	0,85	[161,97 - 162,14]	[162,02 - 162,21]
<b>UA</b>	155,55	155,55	0,09	0,01	155,75	155,31	0,44	[155,51 - 155,58]	[155,50 - 155,59]
	<i>Compact com 6 Threads</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	102,59	102,55	0,71	0,51	103,97	100,29	3,68	[102,32 - 102,86]	[102,35 - 102,75]
<b>CG</b>	27,90	27,71	0,48	0,23	28,80	27,42	1,38	[27,72 - 28,08]	[27,56 - 27,86]
<b>DC</b>	265,71	266,02	4,72	22,27	273,90	252,17	21,73	[263,95 - 267,47]	[264,32 - 267,72]
<b>EP</b>	23,29	23,15	0,44	0,19	25,02	23,06	1,96	[23,12 - 23,45]	[23,10 - 23,20]
<b>FT</b>	22,30	22,29	0,10	0,01	22,66	22,13	0,53	[22,26 - 22,33]	[22,27 - 22,31]
<b>IS</b>	3,13	3,13	0,00	0,00	3,13	3,12	0,01	[3,13 - 3,13]	[3,13 - 3,13]
<b>LU</b>	76,06	76,18	1,03	1,06	78,11	74,58	3,53	[75,67 - 76,44]	[75,58 - 76,77]
<b>MG</b>	5,83	5,86	0,12	0,01	5,96	5,59	0,37	[5,78 - 5,87]	[5,79 - 5,92]
<b>SP</b>	112,48	112,53	1,95	3,82	116,70	108,42	8,28	[111,75 - 113,21]	[111,42 - 113,64]
<b>UA</b>	101,68	101,87	1,14	1,30	104,49	99,81	4,68	[101,25 - 102,10]	[101,28 - 102,46]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

	<i>Scatter com 2 Threads</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	195,36	195,38	0,26	0,07	195,78	194,80	0,98	[195,26 - 195,45]	[195,24 - 195,52]
<b>CG</b>	44,04	44,05	0,03	0,00	44,08	43,93	0,15	[44,03 - 44,05]	[44,04 - 44,05]
<b>DC</b>	333,88	336,16	5,11	26,09	341,66	324,02	17,64	[331,97 - 335,78]	[334,81 - 337,50]
<b>EP</b>	65,46	64,31	2,18	4,74	67,80	63,26	4,54	[64,65 - 66,27]	[62,20 - 66,41]
<b>FT</b>	41,21	41,26	0,39	0,15	42,07	40,46	1,61	[41,06 - 41,35]	[41,09 - 41,43]
<b>IS</b>	6,88	6,88	0,00	0,00	6,88	6,87	0,01	[6,88 - 6,88]	[6,88 - 6,88]
<b>LU</b>	146,22	146,19	0,26	0,07	146,70	145,85	0,85	[146,12 - 146,31]	[146,04 - 146,34]
<b>MG</b>	8,88	8,88	0,01	0,00	8,90	8,87	0,03	[8,88 - 8,89]	[8,87 - 8,89]
<b>SP</b>	153,54	153,31	1,02	1,05	156,40	151,76	4,64	[153,16 - 153,92]	[152,90 - 153,71]
<b>UA</b>	186,79	186,70	0,49	0,24	187,82	185,94	1,88	[186,61 - 186,98]	[186,45 - 186,94]
	<i>Scatter com 4 Threads</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	103,07	102,92	0,47	0,22	104,77	102,65	2,12	[102,90 - 103,25]	[102,77 - 103,06]
<b>CG</b>	24,05	24,05	0,11	0,01	24,29	23,83	0,46	[24,01 - 24,09]	[23,99 - 24,10]
<b>DC</b>	296,62	295,71	4,26	18,18	305,24	289,52	15,72	[295,03 - 298,22]	[294,73 - 296,68]
<b>EP</b>	36,60	37,59	1,66	2,76	37,60	33,63	3,97	[35,98 - 37,22]	[37,55 - 37,63]
<b>FT</b>	23,72	23,69	0,38	0,14	24,56	22,93	1,63	[23,58 - 23,86]	[23,51 - 23,86]
<b>IS</b>	3,51	3,51	0,00	0,00	3,52	3,50	0,02	[3,51 - 3,51]	[3,51 - 3,51]
<b>LU</b>	77,44	77,35	0,86	0,73	79,26	76,23	3,03	[77,12 - 77,76]	[77,00 - 77,70]
<b>MG</b>	4,96	4,97	0,04	0,00	5,02	4,88	0,14	[4,95 - 4,98]	[4,96 - 4,98]
<b>SP</b>	94,29	94,11	0,78	0,61	96,83	93,11	3,72	[93,99 - 94,58]	[93,91 - 94,30]
<b>UA</b>	113,72	113,67	0,60	0,37	115,08	112,72	2,36	[113,50 - 113,95]	[113,38 - 113,96]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

	<i>Scatter com 6 Threads</i>								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>BT</b>	97,30	97,21	0,43	0,18	98,31	96,70	1,61	[97,14 - 97,46]	[97,03 - 97,39]
<b>CG</b>	26,05	26,03	0,20	0,04	26,55	25,66	0,89	[25,97 - 26,12]	[25,97 - 26,09]
<b>DC</b>	248,87	249,04	2,80	7,87	257,05	243,63	13,42	[247,82 - 249,92]	[248,01 - 250,07]
<b>EP</b>	23,09	23,05	0,08	0,01	23,33	23,04	0,29	[23,06 - 23,12]	[23,05 - 23,05]
<b>FT</b>	21,31	21,34	0,17	0,03	21,59	20,87	0,72	[21,24 - 21,37]	[21,27 - 21,41]
<b>IS</b>	2,98	2,98	0,00	0,00	2,99	2,98	0,01	[2,98 - 2,98]	[2,98 - 2,98]
<b>LU</b>	75,71	76,04	1,26	1,59	77,16	72,01	5,15	[75,24 - 76,18]	[75,68 - 76,39]
<b>MG</b>	5,04	5,05	0,05	0,00	5,11	4,90	0,21	[5,02 - 5,06]	[5,02 - 5,08]
<b>SP</b>	98,59	98,54	1,09	1,19	101,02	96,81	4,21	[98,18 - 99,00]	[97,79 - 99,28]
<b>UA</b>	100,66	100,58	0,93	0,86	102,97	98,84	4,13	[100,32 - 101,01]	[100,24 - 100,92]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

<i>Sem Affinity com 2 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	161,00	160,47	3,01	9,04	166,59	154,51	12,08	[159,87 - 162,12]	[159,42 - 161,51]
<i>bodytrack</i>	135,55	135,53	0,21	0,04	136,11	135,27	0,84	[135,48 - 135,63]	[135,45 - 135,60]
<i>facesim</i>	370,14	370,07	0,39	0,15	370,90	369,38	1,52	[369,99 - 370,29]	[369,77 - 370,36]
<i>Sem Affinity com 4 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	98,21	97,88	3,99	15,92	106,18	92,67	13,51	[96,72 - 99,70]	[95,70 - 100,05]
<i>bodytrack</i>	77,03	77,11	0,18	0,03	77,23	76,72	0,51	[76,97 - 77,10]	[76,98 - 77,24]
<i>facesim</i>	189,31	189,54	0,64	0,41	190,23	187,94	2,29	[189,07 - 189,55]	[189,14 - 189,93]
<i>Sem Affinity com 6 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	77,85	77,48	3,26	10,62	84,80	72,68	12,12	[76,64 - 79,07]	[76,20 - 78,75]
<i>bodytrack</i>	56,51	56,52	0,17	0,03	56,82	56,21	0,61	[56,44 - 56,57]	[56,44 - 56,60]
<i>facesim</i>	128,34	128,50	0,71	0,50	129,52	127,06	2,46	[128,07 - 128,60]	[128,05 - 128,94]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

<i>Compact com 2 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	171,67	171,68	0,14	0,02	172,05	171,43	0,62	[171,62 - 171,73]	[171,61 - 171,74]
<i>bodytrack</i>	147,76	147,76	0,01	0,00	147,79	147,74	0,05	[147,75 - 147,76]	[147,76 - 147,76]
<i>facesim</i>	446,85	446,97	0,49	0,24	447,46	445,32	2,14	[446,67 - 447,03]	[446,89 - 447,05]
<i>Compact com 4 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	103,24	103,26	0,16	0,03	103,53	102,83	0,70	[103,18 - 103,30]	[103,20 - 103,32]
<i>bodytrack</i>	83,44	83,44	0,02	0,00	83,49	83,41	0,08	[83,43 - 83,45]	[83,43 - 83,45]
<i>facesim</i>	227,32	227,36	0,31	0,10	227,88	226,60	1,28	[227,20 - 227,43]	[227,21 - 227,51]
<i>Compact com 6 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>blackscholes</i>	78,98	78,97	0,14	0,02	79,31	78,69	0,62	[78,92 - 79,03]	[78,91 - 79,03]
<i>bodytrack</i>	61,61	61,56	0,09	0,01	61,75	61,52	0,23	[61,58 - 61,64]	[61,48 - 61,64]
<i>facesim</i>	154,10	154,14	0,43	0,18	154,90	152,34	2,56	[153,94 - 154,26]	[153,95 - 154,32]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE I: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #3 (CONT.)

<i>Scatter com 2 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>black scholes</i>	159,54	159,29	3,95	15,62	172,54	153,39	19,15	[158,06 - 161,01]	[158,25 - 160,33]
<i>bodytrack</i>	147,76	147,76	0,01	0,00	147,79	147,74	0,05	[147,75 - 147,76]	[147,76 - 147,76]
<i>facesim</i>	372,01	371,98	0,42	0,18	372,66	371,11	1,55	[371,85 - 372,17]	[371,74 - 372,21]
<i>Scatter com 4 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>black scholes</i>	98,39	97,35	4,47	19,97	108,38	92,71	15,67	[96,72 - 100,06]	[95,43 - 99,26]
<i>bodytrack</i>	83,44	83,44	0,02	0,00	83,49	83,41	0,08	[83,43 - 83,45]	[83,43 - 83,45]
<i>facesim</i>	189,13	189,24	0,60	0,36	189,95	187,51	2,44	[188,91 - 189,35]	[189,04 - 189,44]
<i>Scatter com 6 Threads</i>									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<i>black scholes</i>	79,17	79,07	0,50	0,25	81,53	78,71	2,82	[78,98 - 79,36]	[79,02 - 79,12]
<i>bodytrack</i>	61,61	61,56	0,09	0,01	61,75	61,52	0,23	[61,58 - 61,64]	[61,48 - 61,64]
<i>facesim</i>	144,84	144,82	0,68	0,46	146,06	142,96	3,10	[144,59 - 145,09]	[144,57 - 145,06]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4

	Experimento com 32 Tratamentos								
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>1</b>	598,98	598,95	0,22	0,05	599,70	598,60	1,10	[598,89 - 599,06]	[598,89 - 599,01]
<b>2</b>	598,96	598,97	0,23	0,05	599,53	598,58	0,95	[598,87 - 599,04]	[598,85 - 599,08]
<b>3</b>	599,01	598,99	0,26	0,07	599,64	598,44	1,20	[598,91 - 599,11]	[598,91 - 599,07]
<b>4</b>	599,15	599,10	0,23	0,05	599,76	598,74	1,02	[599,06 - 599,23]	[598,99 - 599,21]
<b>5</b>	599,01	598,96	0,29	0,09	599,79	598,51	1,28	[598,90 - 599,12]	[598,90 - 599,02]
<b>6</b>	598,96	598,91	0,35	0,12	599,86	598,38	1,48	[598,83 - 599,09]	[598,82 - 599,00]
<b>7</b>	599,08	599,08	0,19	0,04	599,64	598,71	0,93	[599,01 - 599,15]	[599,05 - 599,11]
<b>8</b>	598,83	598,80	0,23	0,05	599,57	598,48	1,09	[598,75 - 598,92]	[598,75 - 598,85]
<b>9</b>	599,15	599,14	0,28	0,08	599,93	598,64	1,29	[599,05 - 599,26]	[599,05 - 599,22]
<b>10</b>	598,91	598,89	0,20	0,04	599,31	598,46	0,85	[598,84 - 598,98]	[598,81 - 598,97]
<b>11</b>	598,99	598,97	0,27	0,07	599,72	598,58	1,14	[598,89 - 599,09]	[598,87 - 599,07]
<b>12</b>	598,94	598,90	0,26	0,07	599,61	598,47	1,14	[598,84 - 599,04]	[598,82 - 598,97]
<b>13</b>	598,94	598,91	0,25	0,06	599,66	598,49	1,17	[598,84 - 599,03]	[598,82 - 598,99]
<b>14</b>	598,86	598,83	0,18	0,03	599,21	598,55	0,66	[598,79 - 598,92]	[598,75 - 598,91]
<b>15</b>	599,01	598,97	0,24	0,06	599,76	598,60	1,16	[598,92 - 599,10]	[598,90 - 599,04]
<b>16</b>	598,93	598,88	0,23	0,05	599,82	598,55	1,27	[598,84 - 599,01]	[598,79 - 598,96]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

(Continua)

## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4 (CONT.)

Experimento com 32 Tratamentos									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
17	600,99	600,97	0,27	0,07	601,67	600,45	1,22	[600,89 - 601,09]	[600,86 - 601,07]
18	601,04	601,04	0,19	0,04	601,75	600,68	1,07	[600,97 - 601,11]	[600,98 - 601,10]
19	600,64	600,65	0,21	0,04	601,22	600,25	0,97	[600,56 - 600,72]	[600,58 - 600,72]
20	601,17	601,16	0,25	0,06	601,97	600,81	1,16	[601,08 - 601,26]	[601,07 - 601,25]
21	601,10	601,07	0,35	0,12	602,45	600,53	1,92	[600,97 - 601,23]	[600,95 - 601,19]
22	601,33	601,14	0,53	0,29	603,14	600,78	2,36	[601,13 - 601,53]	[601,06 - 601,21]
23	600,90	600,77	0,47	0,22	602,72	600,38	2,34	[600,72 - 601,07]	[600,65 - 600,89]
24	601,13	601,17	0,22	0,05	601,77	600,69	1,08	[601,05 - 601,22]	[601,09 - 601,25]
25	600,82	600,77	0,28	0,08	601,59	600,37	1,22	[600,72 - 600,92]	[600,69 - 600,84]
26	601,05	601,04	0,25	0,06	601,65	600,63	1,02	[600,96 - 601,14]	[600,93 - 601,14]
27	600,90	600,83	0,25	0,06	601,70	600,44	1,26	[600,80 - 600,99]	[600,74 - 600,92]
28	601,28	601,24	0,32	0,10	602,71	600,81	1,90	[601,16 - 601,40]	[601,18 - 601,30]
29	600,62	600,61	0,23	0,05	601,14	600,25	0,89	[600,53 - 600,70]	[600,51 - 600,70]
30	601,08	601,06	0,22	0,05	601,65	600,72	0,93	[601,00 - 601,16]	[600,98 - 601,13]
31	601,23	601,13	0,42	0,17	602,67	600,71	1,96	[601,08 - 601,39]	[600,97 - 601,29]
32	601,14	601,15	0,21	0,04	601,58	600,72	0,86	[601,06 - 601,22]	[601,07 - 601,22]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4 (CONT.)

Experimento com 64 Tratamentos									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
<b>1</b>	599,00	599,00	0,30	0,09	599,86	598,50	1,36	[598,89 - 599,11]	[598,87 - 599,12]
<b>2</b>	598,88	598,83	0,24	0,06	599,65	598,50	1,15	[598,79 - 598,97]	[598,71 - 598,94]
<b>3</b>	599,01	599,06	0,21	0,04	599,47	598,51	0,96	[598,94 - 599,09]	[598,99 - 599,12]
<b>4</b>	598,79	598,77	0,28	0,08	599,76	598,37	1,39	[598,69 - 598,89]	[598,71 - 598,83]
<b>5</b>	598,98	598,96	0,26	0,07	599,55	598,53	1,02	[598,88 - 599,07]	[598,87 - 599,05]
<b>6</b>	598,89	598,86	0,26	0,07	599,86	598,52	1,34	[598,79 - 598,99]	[598,77 - 598,94]
<b>7</b>	598,96	598,91	0,35	0,12	600,43	598,54	1,89	[598,83 - 599,09]	[598,84 - 598,97]
<b>8</b>	599,03	598,95	0,32	0,10	599,91	598,48	1,43	[598,91 - 599,15]	[598,89 - 599,00]
<b>9</b>	599,37	599,21	0,56	0,31	601,15	598,76	2,39	[599,16 - 599,58]	[599,07 - 599,34]
<b>10</b>	599,03	598,98	0,32	0,10	599,98	598,53	1,45	[598,91 - 599,15]	[598,90 - 599,06]
<b>11</b>	599,21	599,04	0,49	0,24	601,15	598,68	2,47	[599,03 - 599,40]	[598,92 - 599,15]
<b>12</b>	598,80	598,74	0,25	0,06	599,89	598,42	1,47	[598,71 - 598,89]	[598,66 - 598,82]
<b>13</b>	598,86	598,82	0,18	0,03	599,29	598,59	0,70	[598,79 - 598,93]	[598,72 - 598,91]
<b>14</b>	598,82	598,78	0,26	0,07	599,59	598,41	1,18	[598,72 - 598,92]	[598,71 - 598,85]
<b>15</b>	599,02	598,96	0,30	0,09	600,12	598,67	1,45	[598,91 - 599,14]	[598,83 - 599,09]
<b>16</b>	598,88	598,83	0,24	0,06	599,74	598,53	1,21	[598,79 - 598,97]	[598,75 - 598,90]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4 (CONT.)

Experimento com 64 Tratamentos									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
17	600,81	600,80	0,25	0,06	601,58	600,42	1,16	[600,72 - 600,90]	[600,75 - 600,85]
18	601,22	601,14	0,34	0,12	602,53	600,79	1,74	[601,09 - 601,35]	[601,06 - 601,22]
19	600,95	600,85	0,39	0,15	602,18	600,39	1,79	[600,80 - 601,10]	[600,73 - 600,96]
20	601,34	601,29	0,22	0,05	601,76	601,00	0,76	[601,25 - 601,42]	[601,19 - 601,39]
21	601,06	601,05	0,25	0,06	601,87	600,51	1,36	[600,97 - 601,16]	[600,94 - 601,16]
22	601,13	601,11	0,24	0,06	601,88	600,73	1,15	[601,04 - 601,22]	[601,03 - 601,19]
23	601,12	601,10	0,29	0,08	601,79	600,58	1,21	[601,01 - 601,23]	[601,00 - 601,20]
24	601,56	601,56	0,25	0,06	602,32	601,13	1,19	[601,47 - 601,65]	[601,46 - 601,65]
25	600,79	600,78	0,27	0,07	601,60	600,37	1,23	[600,69 - 600,89]	[600,67 - 600,88]
26	601,18	601,17	0,26	0,07	602,11	600,72	1,39	[601,09 - 601,28]	[601,10 - 601,23]
27	600,80	600,76	0,22	0,05	601,35	600,47	0,88	[600,72 - 600,89]	[600,67 - 600,84]
28	601,12	601,11	0,24	0,06	601,88	600,66	1,22	[601,03 - 601,21]	[601,03 - 601,18]
29	601,06	600,99	0,33	0,11	601,82	600,39	1,43	[600,94 - 601,18]	[600,84 - 601,13]
30	601,11	601,08	0,21	0,04	601,62	600,70	0,92	[601,04 - 601,19]	[601,01 - 601,14]
31	600,77	600,77	0,28	0,08	601,44	600,25	1,19	[600,67 - 600,87]	[600,66 - 600,88]
32	601,11	601,08	0,23	0,05	601,88	600,76	1,12	[601,02 - 601,20]	[601,01 - 601,14]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4 (CONT.)

Experimento com 64 Tratamentos									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
33	602,88	602,54	1,90	3,61	607,87	599,77	8,10	[602,17 - 603,59]	[601,85 - 603,23]
34	600,78	600,32	1,40	1,96	605,18	598,96	6,22	[600,25 - 601,30]	[599,98 - 600,65]
35	601,38	601,09	1,46	2,13	604,61	599,44	5,17	[600,83 - 601,92]	[600,52 - 601,66]
36	600,55	600,45	0,81	0,66	603,06	599,34	3,72	[600,25 - 600,85]	[600,22 - 600,67]
37	600,79	600,49	1,25	1,55	604,98	599,27	5,71	[600,32 - 601,25]	[600,30 - 600,67]
38	600,86	600,70	1,28	1,63	604,23	599,21	5,02	[600,38 - 601,34]	[600,05 - 601,34]
39	601,45	601,45	1,32	1,75	605,06	599,37	5,69	[600,96 - 601,94]	[601,11 - 601,78]
40	601,18	600,99	1,29	1,67	603,93	599,10	4,83	[600,70 - 601,66]	[600,50 - 601,47]
41	603,27	602,16	3,31	10,96	611,34	599,26	12,08	[602,03 - 604,50]	[600,63 - 603,69]
42	601,80	601,09	2,07	4,29	608,53	599,12	9,41	[601,03 - 602,57]	[600,53 - 601,65]
43	601,37	601,05	1,60	2,57	605,65	599,03	6,62	[600,77 - 601,97]	[600,62 - 601,48]
44	601,14	600,36	2,17	4,73	608,04	598,85	9,19	[600,33 - 601,96]	[600,03 - 600,68]
45	600,43	600,03	1,35	1,81	604,35	599,01	5,34	[599,93 - 600,94]	[599,65 - 600,41]
46	601,21	601,13	1,22	1,49	604,88	599,39	5,49	[600,75 - 601,67]	[600,82 - 601,43]
47	601,44	601,22	1,49	2,21	605,40	599,14	6,26	[600,88 - 601,99]	[600,67 - 601,77]
48	602,01	601,58	1,71	2,92	605,86	599,48	6,38	[601,37 - 602,65]	[600,83 - 602,33]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

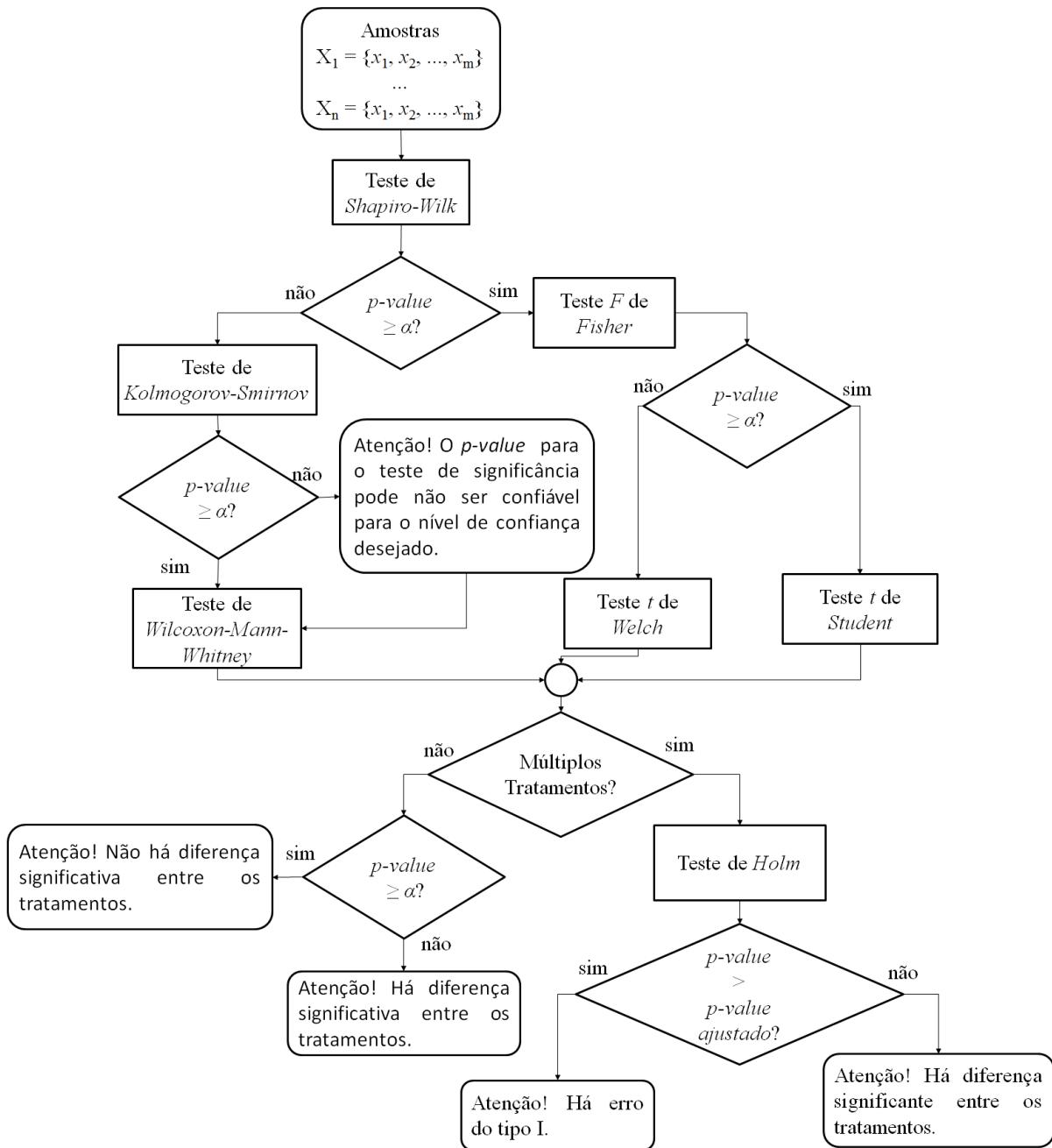
## APÊNDICE J: RESULTADO DA ANÁLISE DESCritIVA DOS TRATAMENTOS DO EXP. #4 (CONT.)

Experimento com 64 Tratamentos									
	Média (S)	Mediana (S)	DP (S)	Variância	Maior Tempo (S)	Menor Tempo (S)	Dif. (S)	Intervalo de Confiança (Média) (S)	Intervalo de Confiança (Mediana) (S)
49	602,32	602,10	0,90	0,81	604,97	600,92	4,05	[601,99 - 602,66]	[601,73 - 602,47]
50	602,17	602,05	0,62	0,38	604,05	601,36	2,69	[601,93 - 602,40]	[601,85 - 602,24]
51	603,13	602,97	1,01	1,01	606,16	601,87	4,29	[602,76 - 603,51]	[602,70 - 603,24]
52	602,83	602,57	1,08	1,16	605,69	601,46	4,23	[602,43 - 603,23]	[602,24 - 602,90]
53	602,93	602,31	1,57	2,46	606,98	601,04	5,94	[602,35 - 603,52]	[601,55 - 603,06]
54	602,47	602,22	0,92	0,85	604,91	601,29	3,62	[602,13 - 602,82]	[601,93 - 602,50]
55	602,69	602,56	1,08	1,18	605,46	601,18	4,28	[602,28 - 603,09]	[602,06 - 603,05]
56	603,36	603,31	1,96	3,83	609,13	601,00	8,13	[602,63 - 604,09]	[602,57 - 604,04]
57	605,31	604,62	2,06	4,23	611,09	602,60	8,49	[604,54 - 606,08]	[603,61 - 605,62]
58	602,50	602,10	1,22	1,50	605,64	600,92	4,72	[602,04 - 602,96]	[601,68 - 602,51]
59	603,70	602,92	1,89	3,56	608,54	601,04	7,50	[603,00 - 604,41]	[602,02 - 603,82]
60	603,60	603,56	1,34	1,79	606,21	601,12	5,09	[603,10 - 604,10]	[603,04 - 604,08]
61	604,35	604,16	2,09	4,37	609,92	600,91	9,01	[603,57 - 605,13]	[603,52 - 604,79]
62	602,90	602,87	0,97	0,93	605,79	601,39	4,40	[602,54 - 603,27]	[602,53 - 603,20]
63	603,58	603,24	1,81	3,29	608,67	601,31	7,36	[602,91 - 604,26]	[602,56 - 603,91]
64	604,41	603,95	2,20	4,83	610,77	601,94	8,83	[603,59 - 605,23]	[603,30 - 604,60]

DP – Desvio Padrão; Dif. – Diferença entre o maior e o menor tempo.

(S) – Segundos.

# APÊNDICE K: PROTOCOLO RUNTIME ANALYSIS



# ANEXO A: DISTRIBUIÇÃO *T* DE STUDENT (SHESKIN, 2003).

Two-tailed	.80	.50	.20	.10	.05	.02	.01	.001
One-tailed	.40	.25	.10	.05	.025	.01	.005	.0005
<i>p</i>	.60	.75	.90	.95	.975	.99	.995	.9995
<i>df</i>								
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.289	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	3.373
$\infty$	.253	.674	1.282	1.645	1.960	2.326	2.576	3.291

## ANEXO B: DISTRIBUIÇÃO GAUSSIANA (GIBBONS; CHAKRABORTI, 2014).

Each table entry is the tail probability  $P$ , right tail from the value of  $z$  to plus infinity, and also left tail from minus infinity to  $-z$ , for all  $P \leq .50$ . Read down the first column to the first decimal value of  $z$ , and over to the correct column for the second decimal value; the number at the intersection is  $P$ .

$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002

Source: Adapted from Table 1 of Pearson, E. S., and H. O. Hartley, eds. (1954), *Biometrika Tables for Statisticians*, vol. 1, Cambridge University Press, Cambridge, England, with permission of the Biometrika Trustees.

# ANEXO C: VALORES CRÍTICOS PARA $U$ DO TESTE DE *MANN-WHITNEY* (SHESKIN, 2003).

Bicaudal

$n_1 \backslash n_2$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2								0	0	0	0	1	1	1	1	2	2	2	2	
3					0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	
4				0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	
5			0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	
6		1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	
7	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34		
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41	
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48	
10	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55	
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62	
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69	
13	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76	
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83	
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90	
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98	
17	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105	
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112	
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119	
20	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127	

# ANEXO D: *KOLMOGOROV-SMIRNOV* UMA AMOSTRA (GIBBONS; CHAKRABORTI, 2014).

<i>n</i>	.200	.100	.050	.020	.010	<i>n</i>	.200	.100	.050	.020	.010
1	.900	.950	.975	.990	.995	21	.226	.259	.287	.321	.344
2	.684	.776	.842	.900	.929	22	.221	.253	.281	.314	.337
3	.565	.636	.780	.785	.829	23	.216	.247	.275	.307	.330
4	.493	.565	.624	.689	.734	24	.212	.242	.269	.301	.323
5	.447	.509	.563	.627	.669	25	.208	.238	.264	.295	.317
6	.410	.468	.519	.577	.617	26	.204	.233	.259	.290	.311
7	.381	.436	.483	.538	.576	27	.200	.229	.254	.284	.305
8	.358	.410	.454	.507	.542	28	.197	.225	.250	.279	.300
9	.339	.387	.430	.480	.513	29	.193	.221	.246	.275	.295
10	.323	.369	.409	.457	.489	30	.190	.218	.242	.270	.290
11	.308	.352	.391	.437	.468	31	.187	.214	.238	.266	.285
12	.296	.338	.375	.419	.449	32	.184	.211	.234	.262	.281
13	.285	.325	.361	.404	.432	33	.182	.208	.231	.258	.277
14	.275	.314	.349	.390	.418	34	.179	.205	.227	.254	.273
15	.266	.304	.338	.377	.404	35	.177	.202	.224	.251	.269
16	.258	.295	.327	.366	.392	36	.174	.199	.221	.247	.265
17	.250	.286	.318	.355	.381	37	.172	.196	.218	.244	.262
18	.244	.279	.309	.346	.371	38	.170	.194	.215	.241	.258
19	.237	.271	.301	.337	.361	39	.168	.191	.213	.238	.255
20	.232	.265	.294	.329	.352	40	.165	.189	.210	.235	.252

For  $n > 40$ , right-tail critical values based on the asymptotic distribution can be calculated as follows:

$$\begin{array}{ccccc} .200 & .100 & .050 & .020 & .010 \\ 1.07/\sqrt{n} & 1.22/\sqrt{n} & 1.36/\sqrt{n} & 1.52/\sqrt{n} & 1.63/\sqrt{n} \end{array}$$

Source: Adapted from L. H. Miller (1956), Table of percentage points of Kolmogorov statistics, *Journal of the American Statistical Association*, 51, 111–121, with permission.

## ANEXO E: *KOLMOGOROV-SMIRNOV* DUAS AMOSTRAS (SHESKIN, 2003).

<b><i>n</i><sub>1</sub></b>	<b><i>n</i><sub>2</sub></b>	<b>One-tailed</b>	.10	.05	.025	.01	.005
		<b>Two-tailed</b>	.20	.10	.05	.02	.01
3	3	.667	.667				
3	4	.750	.750				
3	5	.667	.800	.800			
3	6	.667	.667	.833			
3	7	.667	.714	.857	.857		
3	8	.625	.750	.750	.875		
3	9	.667	.667	.778	.889	.889	
3	10	.600	.700	.800	.900	.900	
3	12	.583	.667	.750	.833	.917	
4	4	.750	.750	.750			
4	5	.600	.750	.800	.800		
4	6	.583	.667	.750	.833	.833	
4	7	.607	.714	.750	.857	.857	
4	8	.625	.625	.750	.875	.875	
4	9	.556	.667	.750	.778	.889	
4	10	.550	.650	.700	.800	.800	
4	12	.583	.667	.667	.750	.833	
4	16	.563	.625	.688	.750	.812	
5	5	.600	.600	.800	.800	.800	
5	6	.600	.667	.667	.833	.833	
5	7	.571	.657	.714	.829	.857	
5	8	.550	.625	.675	.800	.800	
5	9	.556	.600	.689	.778	.800	
5	10	.500	.600	.700	.700	.800	
5	15	.533	.600	.667	.733	.733	
5	20	.500	.550	.600	.700	.750	
6	6	.500	.667	.667	.833	.833	
6	7	.548	.571	.690	.714	.833	
6	8	.500	.583	.667	.750	.750	
6	9	.500	.556	.667	.722	.778	
6	10	.500	.567	.633	.700	.733	
6	12	.500	.583	.583	.667	.750	
6	18	.444	.556	.611	.667	.722	
6	24	.458	.500	.583	.625	.667	
7	7	.571	.571	.714	.714	.714	
7	8	.482	.589	.625	.732	.750	
7	9	.492	.556	.635	.714	.746	
7	10	.471	.557	.614	.700	.714	
7	14	.429	.500	.571	.643	.714	
7	28	.429	.464	.536	.607	.643	
8	8	.500	.500	.625	.625	.750	
8	9	.444	.542	.625	.667	.750	
8	10	.475	.525	.575	.675	.700	
8	12	.458	.500	.583	.625	.667	
8	16	.438	.500	.563	.625	.625	
8	32	.406	.438	.500	.563	.594	

(Continua)

# ANEXO E: *KOLMOGOROV-SMIRNOV* DUAS AMOSTRAS (SHESKIN, 2003) (CONT.).

<b>One-tailed</b>		.10	.05	.025	.01	.005
<b>Two-tailed</b>		.20	.10	.05	.02	.01
<b><i>n</i><sub>1</sub></b>	<b><i>n</i><sub>2</sub></b>					
9	9	.444	.556	.556	.667	.667
9	10	.467	.500	.578	.667	.689
9	12	.444	.500	.556	.611	.667
9	15	.422	.489	.533	.600	.644
9	18	.389	.444	.500	.556	.611
9	36	.361	.417	.472	.528	.556
10	10	.400	.500	.600	.600	.700
10	15	.400	.467	.500	.567	.633
10	20	.400	.450	.500	.550	.600
10	40	.350	.400	.450	.500	.576
11	11	.454	.454	.545	.636	.636
12	12	.417	.417	.500	.583	.583
12	15	.383	.450	.500	.550	.583
12	16	.375	.438	.479	.542	.583
12	18	.361	.417	.472	.528	.556
12	20	.367	.417	.467	.517	.567
13	13	.385	.462	.462	.538	.615
14	14	.357	.429	.500	.500	.571
15	15	.333	.400	.467	.467	.533
16	16	.375	.375	.438	.500	.563
17	17	.353	.412	.412	.471	.529
18	18	.333	.389	.444	.500	.500
19	19	.316	.368	.421	.473	.473
20	20	.300	.350	.400	.450	.500
21	21	.286	.333	.381	.429	.476
22	22	.318	.364	.364	.454	.454
23	23	.304	.348	.391	.435	.435
24	24	.292	.333	.375	.417	.458
25	25	.280	.320	.360	.400	.440
For all other sample sizes		1.07 <i>K</i>	1.22 <i>K</i>	1.36 <i>K</i>	1.52 <i>K</i>	1.63 <i>K</i>
Where: $i = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$						

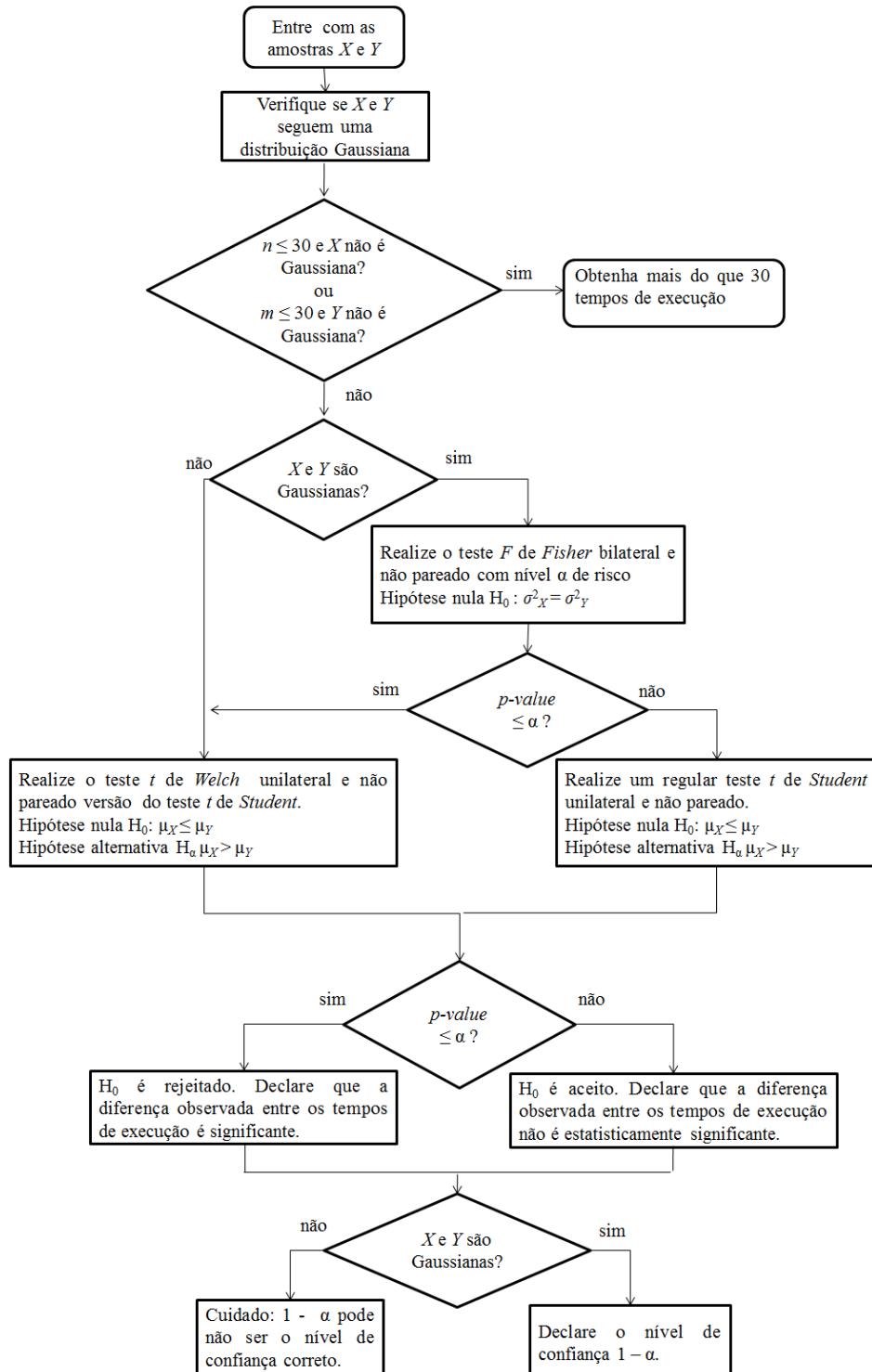
# ANEXO F: DISTRIBUIÇÃO $F$ PARA $\alpha = 0,05$ (SHESKIN, 2003).

$\frac{df_{\text{num}}}{df_{\text{den}}}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	60	120	$\infty$
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	252,2	253,3	254,3
2	18,51	19,0	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,63	2,60	2,53	2,46	2,39	2,35	2,31	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,78	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,74	1,68	1,62
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,43	1,35	1,25
$\infty$	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,32	1,22	1,00

# ANEXO G: ALGORITMO DO PROGRAMA PROVA. (ADAPTADO DE (SOBRINHO, 2012)).

```
1 Max: tamanho das matrizes
2 MultMat: numero de vezes para multiplicar as matrizes
3 M, R, T: matrizes de processamento
4 i, j, k, z: contadores
5
6
7 sched_setaffinity(PU#1);
8 MountIPC; monta comunicação interprocessos com o ipcserver
9 Max = 1000; MultMat=48;
10 AlocMem(M,R,T); aloca[Max*4, Max*4] bytes de memória
11 SetsFactors(D,B,E); configura o nível dos fatores D, B e E
12 GetsTimeI; ipcserver pega o tempo inicial
13
14 for (z = 0; z < MultMat; z++)
15     for (j = 0; j < Max; j++)
16         for (i = 0; i < Max; i++)
17             for (k = 0; k < Max; k++)
18                 if ( z == 0 ) then
19                     M[k][i])
20                 else
21                     M[k][i])
22                 if (z < (MultMat - 1)) then
23                     for (j = 0; j < Max; j++)
24                         for (i = 0; i < Max; i++)
25                             T[j][i] = R[j][i]
26                             R[j][i] = 0;
27                         end for
28                     end for
29 GetsTimeF; ipcserver pega o tempo final
30             if ( o fator E esta no nível -) then configura o fator E no
31             nível +;
32 FreeMem(M,R,T); desaloca a memoria
33 UnmountIPC; Desmonta comunicação interprocessos com o ipcserver
```

# ANEXO H: PROTOCOLO SPEEDUP-TEST PARA ANÁLISE DO MÉDIA. (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013).



# ANEXO I: PROTOCOLO *SPEEDUP-TEST* PARA ANÁLISE DA MEDIANA. (ADAPTADO DE (TOUATI; WORMS; BRIAIS, 2013)).

