*Estudantes:*

**Arthur Fernandes Minduca de Sousa –** [**fernandes.arthur@gmail.com**](mailto:fernandes.arthur@gmail.com)

**Carlos Eduardo Buarque Cruz Pimentel –** [**cebcp@cin.ufpe.br**](mailto:cebcp@cin.ufpe.br)

**Carlos Henrique Maciel Sobral Timóteo –** [**chmst@cin.ufpe.br**](mailto:chmst@cin.ufpe.br)

**Karina Rodrigues Pereira –** [**krp@cin.ufpe.br**](mailto:krp@cin.ufpe.br)

*Vínculo:*

**Mestrado Acadêmico**

*Disciplina:*

**Avaliação de Desempenho de Sistemas**

*Professor:*

**Paulo Maciel**

*Atividade:*

**Resolução da 5ª Lista de Exercícios**

# Estude os seguintes métodos de aderência e os descreva em detalhes:

**a. Metódo de Kolmogorov-Smirnov**

Em estatística, o teste Kolmogorov-Smirnov é usado para determinar se duas distribuições de probabilidade subjacentes diferem uma da outra ou se uma das distribuições de probabilidade subjacentes difere da distribuição em hipótese, em qualquer dos casos com base em amostras finitas. O nome é uma referência aos matemáticos russos Andrey Kolmogorov e Vladimir Ivanovich Smirnov.

A função distribuição acumulada Fn para n observações yi é definida por:

F_n(x)={1 \over n}\sum_{i=1}^n \left\{\begin{matrix}1 & \mathrm{if}\ y_i\leq x, \\ 0 & \mbox{caso contrario}.\end{matrix}\right.

As duas estatísticas de teste Kolmogorov-Smirnov de apenas um lado são dadas por:

D_n^{+}=\max(F_n(x)-F(x))\,

D_n^{-}=\max(F(x)-F_n(x))\,

onde F(x) é a distribuição em hipótese ou outra distribuição empírica. As distribuições de probabilidade destas duas estatísticas, dado que a hipótese nula de igualdade das distribuições é verdadeira, não depende daquilo que a distribuição em hipótese é, desde que ela seja contínua. Muitas pessoas usam max(Dn+, Dn−) alternativamente, mas a distribuição desta estatística é de uso mais difícil.

É importante observar que o teste Kolmogorov-Smirnov é mais sensível em pontos próximos da mediana da distribuição do que nas caudas. O [teste Anderson-Darling](http://pt.wikipedia.org/w/index.php?title=Teste_Anderson-Darling&action=edit&redlink=1) é um teste que providencia igual sensibilidade nas caudas.

Característica do teste de KS:

* Não estão associados a uma distribuição particular.
* Aplica-se para distribuições contínuas.
* Para distribuições discretas, os resultados são muito conservadores.

**b. Método Qui-Quadrado**

Teste de aderência utilizado para verificar se os dados coletados seguem alguma das distribuições (discretas ou contínuas), denominado Qui-Quadrado (χ2) com um nível de erro 5%, ou seja, um nível de confiança de 95%.

Considerando uma tabela de freqüências com g ≥ 2 categorias de resultados.

o*i* – freqüência de ocorrência dos eventos da categoria i (1 ≤ i ≤ g)

e*i* **–** freqüência teórica da categoria i (1 ≤ i ≤ g)

|  |  |  |
| --- | --- | --- |
| **Categorias** | **Freqüência Observada** | **Freqüência Esperada** |
| *1* | o1 | e1 |
| *2* | o2 | e2 |
| *...* | ... | ... |
| *G* | og | eg |
| Total | *n* | *n* |

Seja *pi* a probabilidade associada à categoria i (1 ≤ i ≤ g) e *P(xi)* a probabilidade associada à categoria *i* (1 ≤ i ≤ g),calculada através do modelo probabilístico de interesse.

O objetivo do teste de aderência é testar as hipóteses:

H0 : *pi = P(xi) , 1 ≤ i ≤ g* (segue o modelo)

H1 : ∃ pelo menos uma diferença. (não segue o modelo)

Se a hipótese H é verdadeira a freqüência esperada da categoria *i* é



Se H é verdadeiro temos:



Em outras palavras, se H é verdadeira, a variável aleatória χ2 tem distribuição aproximada qui-quadrado com *n* graus de liberdade.

O grau de liberdade *n* é dado pela expressão:

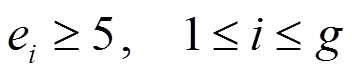


g – número de categorias ou grupos

A – número de parâmetros estimados no processo

|  |  |  |
| --- | --- | --- |
| **Distribuição** | ***A*** | **n** |
| Poisson | 1 (*l*) | *g-2* |

Obs: Este resultado é válido para *n* grande e



|  |  |  |
| --- | --- | --- |
| **Categorias** | **Freqüência Observada** | **Freqüência Esperada** |
| *1* | o1 | e1 |
| *2* | o2 | e2 |
| *...* | ... | ... |
| *G* | og | eg |
| Total | *n* | *n* |

Quantificação da distância entre as colunas de freqüências:



# Considere uma amostra aleatória simples (AAS) - apresentada na planilha em anexo - da medição do tempo de execução de um procedimento (Procedimento A) em departamento financeiro de uma organização. Analise a amostra, defina o procedimento adequado para estimar o tempo médio e o desvio padrão do tempo do procedimento com confiança. Posteriormente, estime o tempo médio e o desvio padrão do tempo de execução do Procedimento A com 95% de confiança.

Como nenhum dado acerca da população do tempo gasto em um procedimento no departamento financeiro de uma organização é dado, é necessário realizar análises estatísticas dos dados da amostra para tentar inferir alguma informação a respeito da distribuição dessa população.

Figura 1. Histograma dos dados da amostra comparada com a curva da distribuição normal.

A análise mais simples a ser realizada é a análise do histograma. Gerando o histograma juntamente com a curva da normal, como na Figura 1, é facilmente observável que a distribuição da amostra não segue a distribuição normal. Por exemplo, a cauda esquerda da curva da normal está truncada. Com os dados da estatística descritiva da amostra, Tabela 1, a inferência feita inicialmente a partir do histograma é corroborada, já que a característica mais básica de uma distribuição normal, média, mediana e moda iguais, não é obedecida. Entretanto, os valores desses parâmetros não estão muito diferentes, principalmente ao se considerar o valor do desvio padrão e o valor da média das juntas. Ainda sobre os dados da estatística, os valores da assimetria e da curtose são maiores que zero (quando deveriam ser igual a zero, caso a distribuição fosse uma normal).

|  |
| --- |
| Variável Média DesvPd Variância Mínimo Q1 Mediana Q3 Máximo  Procedimento A 95,3 60,2 3623,4 2,0 65,5 99,0 115,0 281,0  N para  Variável Moda Moda Assimetria Curtose MedJuntas  Procedimento A 112; 118 2 1,26 3,35 90,25 |

Tabela 1. Estatística descritiva para os dados da amostra.

Como forma de ter maiores evidências de que a distribuição da amostra não segue a distribuição da normal, o teste Anderson-Darling foi aplicado obtendo um valor de 1,23 para A² e que o valor p [[1]](#footnote-1), o *p-value*, era inferior a 0,005. Quando o teste Anderson-Darling é aplicado para a verificação da normalidade, um valor p menor que 0,05 rejeita a hipótese nula de que a distribuição da amostra é a normal [1]; além disso, para amostras de populações com média e variância desconhecidas, um valor de A² superior a 0,787 também rejeita, com significância de 0,05, a hipótese nula [2].

Diante desses dados que mostram que a distribuição da amostra não segue uma normal, é importante ressaltar que eles só dizem isso. O fato de uma distribuição não seguir uma normal não quer dizer que ela é completamente diferente e os valores da média, mediana e moda mostram isso. Sendo assim, e também ao levar-se em consideração a natureza contínua do problema, pode-se concluir que a amostra de tamanho 25 é relativamente pequena. Se o tamanho dessa amostra começasse a aumentar um pouco, o espaço em branco observado, provavelmente, iria diminuindo, bem como a cauda esquerda iria deixando de ser truncada. Para que isso aconteça, a noção do teorema do Limite Central será aplicada nos dados da amostra: serão retiradas, dos dados da planilha, 30 amostras cada uma com tamanho. Realizando essa amostragem, temos as seguintes médias:

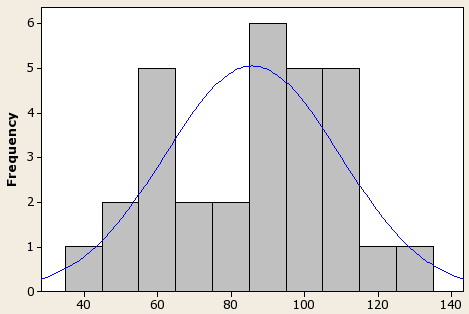
Tabela 2. Média das amostras dos dados da planilha.

|  |  |
| --- | --- |
| Amostra Média  1 112,0  2 35,0  3 114,8  4 95,00  5 97,00  6 87,6  7 62,4  8 85,4  9 109,2  10 114,4  11 48,8  12 117,8  13 103,8  14 45,6  15 97,4 | 16 64,0  17 83,8  18 129,0  19 95,2  20 69,6  21 86,0  22 90,8  23 111,4  24 94,80  25 75,4  26 64,6  27 67,8  28 59,0  29 94,2  30 64,4 |

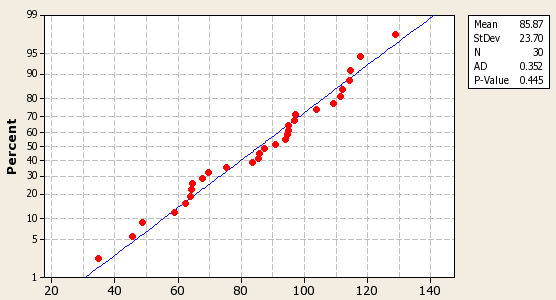
Analisando as estatísticas descritivas das médias das 30 amostras, tem-se o seguinte:

Tabela 3. Estatística descritiva para a média das médias.

|  |
| --- |
| Variável Média DesvPd Variância Mediana  Médias das 85,87 23,70 561,66 89,20  Amostras |

Figura 2. Histograma das médias das amostras comparada com a curva da distribuição normal.

Pelos dados das estatísticas e pela comparação da curva normal com o histograma das médias da amostra, pode parecer que de nada adiantou aplicar a abordagem do Limite Central. Entretanto, ao se aplicar um Q-Q plot aos dados, vê-se que a principal premissa do Limite Centra é obedecida (quanto maior o número de amostras, mais próxima de uma normal a distribuição fica).

Figura 3. Q-Q plot das médias das amostras.

O valor do *A2* e do valor p verificados na imagem corroboram com a afirmação da conformidade com o Teorema do Limite Central. Portanto, agora, uma estimativa de intervalo (o intervalo de confiança) bilateral para a média da população pode ser encontrado levando-se em consideração uma distribuição normal. Entretanto, como o valor da variância populacional também é desconhecido, ao invés de ser utilizada, para a construção do intervalo, a estatística *Z*, que corresponde à distribuição normal padrão, será usada a estatística *t* que correspondente à distribuição *t.* A distribuição *t* é muito parecida com a normal (é simétrica, unimodal e tem forma de sino) e a sua estatística difere da *Z* no uso que ela faz do desvio padrão da amostra em vez do desvio da população. Sendo assim, o intervalo pode ser construído de acordo com a Eq. 1.



, onde é a média amostral, *s* é o desvio padrão amostral, *tα/2,n-1* é o ponto superior com *n-1* graus de liberdade, sendo *n* o tamanho da amostra e *α* é o grau de significância[[2]](#footnote-2).

Substituindo os termos da equação pelos valores encontrados com análise da estatística descritiva e considerando uma confiança de 95%, ou seja, uma siginificância de 5% (ou 0,05), se tem o seguinte intervalo de confiança para a média da população:



Indo na tabela da distribuição t, que encontra-se no Anexo 1, tem-se que *t0,025;24 =* 2,462, então:

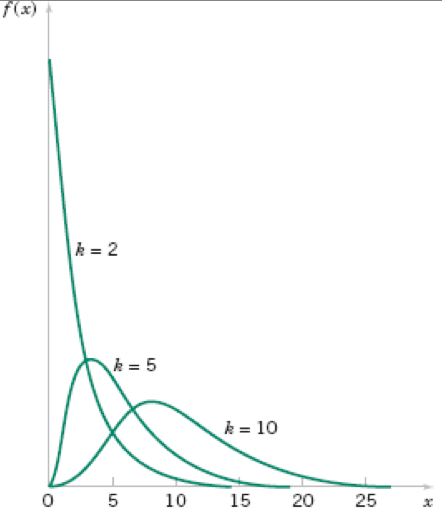




Dessa forma, a partir da amostra analisada para a questão, é possível afirmar, com 95% de confiança que o real valor da média da população do tempo gasto em um procedimento no departamento financeiro de uma organização está entre 70,45 e 120,15 minutos.

Deseja-se também que se estabeleça um intervalo de confiança para o desvio padrão populacional. Para tanto, é necessário uma estatística que relacione o desvio padrão populacional com o desvio padrão amostral (ou que relacione as respectivas variâncias). Tal relação é verificada na estatística da distribuição qui-quadrado:

A distribuição qui-quadrado aproxima-se de uma normal à medida que o tamanho de elementos da amostra aumenta. Pelo gráfico da Figura 3, pode ser verificado que para a amostra do problema que tem tamanho 25, usar a qui-quadrado como aproximação da normal já é plausível.

Figura 4. Gráfico da distribuição qui-quadrado para diferentes graus de liberdade (nesse caso, *k*).

Portanto, pela conformidade da distribuição da amostra com a normal pelo Teorema do Limite Central, será estabelecido um intervalo de confiança para a variância populacional do problema a partir da estatística *X2* , como mostrado pela Eq. 3:

, onde é a variância amostral, *s* é a variância amostral, *X2 1-α/2,n-1* é o ponto inferior com *n-1* graus de liberdade, sendo *n* o tamanho da amostra e *α* é o grau de significância.

Substituindo os termos da equação pelos valores encontrados com análise da estatística e levando em consideração uma confiança de 95%, ou seja, uma siginificância de 5% (ou 0,05), se tem o seguinte intervalo de confiança para a variância populacional:

Indo na tabela da distribuição *X2*, que encontra-se no Anexo 2, tem-se que *X2 0,025/29=* 16 e *X2 0,975/29=* 45,7, então:





Dessa forma, a partir das 30 amostras analisadas, é possível afirmar, com 95% de confiança que o real valor da variância da população do tempo gasto em um procedimento no departamento financeiro de uma organização está entre 356,4 e 1018 minutos2. Com isso, pode-se afirmar que o desvio padrão está entre 18,87 e 31,9.

# Elabore uma metodologia para avaliar e comparar o desempenho de dois computadores da família X86. Apresente o fluxo de atividades e o documento que descreve detalhadamente os pré-requisitos, os insumos, ações, produtos e condições que sinalizam a finalização de cada uma das atividades do fluxo de atividades. O documento deve descrever a carga adotada, as métricas analisadas, as ferramentas a serem utilizadas, as técnicas a serem adotadas, as fórmulas e os procedimentos de execução do processo de medição, análise e diagnóstico. Considere os aspectos discutidos em sala e ressaltados nas notas de aula.

A metodologia proposta pode ser representada pelo seguinte fluxograma:

Figura – Fluxograma de comparação da avaliação do desempenho de computadores

**Planejamento**

***Definição do objetivo***

Antes de realizar a avaliação do desempenho de sistemas, durante a fase de planejamento se devem determinar quais os parâmetros do sistema que são de interesse para a medição. A partir dos eventos observados do sistema pode-se derivar uma métrica para o estudo do desempenho. Uma métrica ideal satisfaz as seguintes características:

* Linearidade: Se o valor da métrica se modificar a uma determinada taxa, o desempenho da máquina deve se modificar à mesma taxa;
* Confiança: Uma métrica de desempenho é dita confiável se dados dois sistemas A e B, o sistema A tem desempenho superior ao sistema B, logo a métrica terá que indicar o mesmo;
* Repetibilidade: Uma métrica tem repetibilidade, se o mesmo valor é medido a cada vez que o experimento é executado. Isso significa que uma boa métrica deverá ser determinística;
* Fácil Medição: Se uma métrica é de difícil medição é improvável que desejem utilizá-la em algum estudo.

A maioria dos *benchmarks* baseia a medida do desempenho no tempo requerido para sua execução, mas há outras medidas como a medida da quantidade de computação em um determinado tempo e uma abordagem híbrida dessas.

O objetivo é realizar comparação da avaliação do desempenho de dois computadores da família x86 para identificar qual dos dois possui o melhor desempenho.

***Descrição do Cenário***

O cenário é composto por dois computadores com arquiteturas parecidas: processador Intel Core 2 Duo, memória DDR2, disco rígido SATA, chipset Intel. Os seguintes componentes das máquinas serão analisados, durante o funcionamento:

* Processadores
* Gráficos 2D
* Gráficos 3D
* Memória
* Disco Rígido

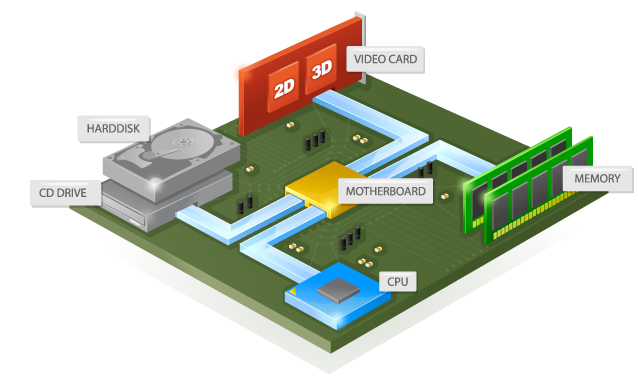


Figura – Sistema a ser avaliado

Devem ser considerados somente os processos vitais aos componentes dos sistemas avaliados, conforme Figura 2. O objetivo será dedicar a maior quantidade possível de processamento às cargas analisadas.

Os pré-requisitos, as pré-condições, os fornecedores, os insumos, as ações, os produtos, os consumidores e as pós-condições das atividades planejadas são apresentadas na Tabela 1.

As máquinas serão reiniciadas para limpar a memória *cache* do processador. As placas de rede serão desconectadas devido ao processamento que ocorre nos protocolos de comunicação. Os dois computadores estarão sobre temperatura ambiente, pois a temperatura causa alterações no desempenho dos circuitos integrados. Os dois sistemas deverão estar isolados, pois evita interferências do meio externo. A configuração do consumo de energia das máquinas foi alterada para não entrar em modo de espera, com o intuito de evitar que a execução dos processos seja interrompida.

O resumo da especificação do cenário quanto ao estado inicial das máquinas durante aplicação da metodologia é descrito na Tabela 2.

Tabela 1. Ações a serem realizadas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pré-requisito | Pré-condição | Insumo | Ação | Produto | Pós-condição |
| Ligação elétrica. | Máquina energizada. | Analista. | Ligar a máquina. | Máquina ligada. | Sistema Operacional inicializado. |
| Windows 7 instalado no sistema. | Windows 7 inicializado em conformidade. | Analista.  Teclado. | *Logon* no Windows 7. | Windows 7 pronto. | Máquina preparada para a execução dos *benchmarks*. |
| Os *benchmarks* deverão ser pré-instalados e configurados. | Nenhum outro programa do usuário tenha sido executado, previamente. | Analista.  Teclado.  Mouse.  Monitor. | Executar um *benchmark* e obter amostras de resultados definidos. | Amostras de resultados salvos num arquivo. | Resultados armazenados no computador. |
| Coleta de resultados anterior finalizada. | Resultados coletados em conformidade. | Analista. | Reiniciar o computador. | Computador no estado inicial. | Os componentes encontram-se no estado inicial. |
|  |  |  | Repetir a *Ação 2* até que todos os *benchmarks* tenham sido executados. |  |  |
|  | Procedimento finalizado. | Analista. | Desligar o computador. | Máquina desligada. |  |

Tabela 2. Descrição das condições gerais das máquinas para o processo.

|  |  |
| --- | --- |
| Pré-condição | Motivo |
| Reiniciar o SO | Limpar a cache do processador |
| Eliminar processos não vitais à máquina | Garantir que o processador estará alocado apenas para os processos importantes |
| Desconectar IOs e conexões não relevantes | Processamento é gasto com comunicação e dispositivos conectados |
| Manter os testes em temperatura ambiente | Temperatura pode resultar em alterações no resultado |
| Desativar modo de espera | Evita interrupção da medição |

***Definição da cargas de trabalho***

Nesta etapa, é preciso definir o que realmente deverá ser avaliado. O objetivo do estudo que está sendo realizado deve ficar claro. Para este trabalho, usaremos a técnica de *Benchmarking*, processo de comparação de desempenho em diferentes sistemas. No nosso caso, *benchmark* representa um software que realiza uma quantidade predefinida de operações e retorna um resultado em alguma métrica específica descrevendo o estado do sistema.

De acordo com a escolha das métricas, é possível definir a carga de trabalho a ser utilizada no experimento. As cargas gerarão, no sistema, os estímulos necessários para a obtenção das métricas que serão coletadas na próxima etapa da metodologia.

Nesse caso de estudo, as cargas de trabalhos utilizadas são padrões e são configuradas automaticamente pelos *benchmarks* adotados na avaliação do desempenho.

***Definição das Ferramentas***

A capacidade de leitura e escrita na memória principal influi diretamente no desempenho de microprocessadores. Dessa forma, a metodologia proposta envolve o uso da ferramenta PerformanceTest para medir a leitura da memória principal. Para aferir avaliação de operações aritméticas, decisões lógicas e uso de strings, utilizaremos o *benchmark* Dhrystone. Para medir o poder de computação de ponto flutuante do sistema, essa metodologia emprega o uso do Linpack. Após isso, será utilizado o *benchmark* Passmark PerformanceTest 7.0 para medir o desempenho do processador, do processamento gráfico 2D e 3D, da memória e do disco rígido.

**Dhrystone**

O Dhrystone foi publicado pela primeira vez em 1984 por Reinhold Weicker da Siemens em C, Pascal e ADA, na tentativa de medir e comparar o desempenho dos computadores. O objetivo seria que com o Dhrystone, fosse criado um benchmark pequeno e representativo para programação de sistemas (no caso de operações aritméticas). O código do Dhrystone é em grande parte representado por aritmética simples, operações com strings, decisões lógicas e acessos de leitura e escrita de memória com intenção de refletir as atividades da CPU nas aplicações de computação de propósito mais geral. O resultado é determinado através do cálculo do tempo médio que um processador leva para executar as todas as iterações de um laço que, por sua vez, contém uma seqüência fixa de instruções que o compõem.

O programa compara o desempenho do processador ao de uma “máquina de referência”, o que é considerado por muitos como uma vantagem em relação à utilização direta da métrica MIPS. A justificativa para tanto é que usar uma máquina de referência efetivamente compensa as diferenças na complexidade das instruções, onde comparar os números de MIPS de uma arquitetura RISC com os de uma CISC não é considerado válido por muitos pesquisadores. Como os processadores atuais são capazes de executar várias instruções por ciclo e dependem muito da velocidade e quantidade de memória cache, o resultado serve apenas como uma referência de desempenho bruto, que não indica necessariamente o desempenho do processador em aplicativos reais.

Inicialmente a indústria adotou o VAX 11/780 como uma máquina de referência de 1 MIPS. O VAX 11/780 alcança 1757 D/S (Dhrystone por segundo). O resultado do benchmark é calculado mensurando o número de D/S para o sistema, e dividindo-se este número por 1757 (da máquina de referência).

Podemos considerar que algumas das vantagens aparentes do Dhrystone também são fraquezas significativas. Os números do Dhrystone refletem na verdade o desempenho do compilador da linguagem C e suas bibliotecas, provavelmente mais do que o desempenho do próprio processador e seu projeto foi baseado na análise de vários outros programas escritos em diferentes linguagens e por diferentes autores, porém voltados à programação de sistemas operacionais e compiladores. Esta é uma característica bastante relevante, pois diferentes classes de aplicações enfatizam diferentes tipos de operações. De acordo com a freqüência das operações nos diferentes programas analisados, foram construídas algumas tabelas que deram origem ao Dhrystone. Em seu código original, em Ada, ele contém 100 sentenças entre o início e o fim do contador de tempo, balanceadas em relação a tipos de sentenças, tipos de dados (operandos) e sua localidade (global, local,parâmetro, constante). O programa contém uma distribuição de 51% de atribuições, 32% de sentenças de controle e 17% de chamadas de funções e procedimentos. O corpo do programa contém 12 procedimentos e, durante um laço (isto é, 1 Dhrystone), as 100 sentenças que compõem o laço são executadas.

O Dhrystone oferece como saída duas métricas: tempo em microsegundos e Dhrystone por segundo. As métricas são inversamente proporcionais, isto é, tempo em microsegundos – quanto maior o valor pior o desempenho; Dhrystone por segundo – quanto maior o valor melhor o desempenho. Seu objetivo é substituir o benchmark denominado Whetstone, mais antigo e menos confiável. O Dhrystone, como a maioria dos benchmarks, consiste em trechos de código padronizados, que são revistos periodicamente para minimizar vantagens que possam favorecer injustamente a determinadas combinações de hardware, compilador e ambiente. O Dhrystone se concentra no tratamento de strings e não utiliza operações com ponto flutuante. A exemplo da maioria dos testes de benchmark, ele é fortemente influenciado pelo projeto do hardware e do linker, a otimização do código, o uso de cache de memória, os estados de espera e os tipos de dados inteiros.

**Linpack**

Este benchmark foi produzido por Jack Dongarra, Jim Bunch, Cleve Moler e Gilbert Stewart, do pacote “LINPACK” das rotinas de álgebra linear. Tornou-se o benchmark básico para aplicações científicas desde a metade de 1980 com uma inclinação para o desempenho do supercomputador.

A versão original foi produzida em Fortran, porém uma versão em “C” apareceu posteriormente. A versão padrão “C” opera em matrizes 100x100, com precisão dupla ou simples e os tipos de loops podem ser rolled/unrolled. A versão pré-compilada é de precisão dupla, rolled, otimizada ou não otimizada. Também existem as versões para DOS e OS/2.

O Linpack faz uso das bibliotecas BLAS (Basic Linear Algebra Subprograms) para realização de vetores básicos e operações de matrizes. É uma medida da potência de computação de ponto flutuante de um sistema. Ele mede o quão rápido um computador resolve um denso N por N sistema de equações lineares Ax = b, o qual é uma tarefa comum em engenharia. A solução é obtida por eliminação Gaussiana com giro parcial, com operações de ponto flutuante 2/3·N3 + 2·N2. O resultado é reportado em milhões de operações de pontos flutuantes por segundo (MFLOP/s, as vezes simplesmente chamado FLOPS).

Para sistemas de memória distribuída em larga escala, Linpack de alto desempenho, uma implementação portátil do benchmarck Linpack de alto desempenho, é utilizado como uma medida de desempenho para qualificar supercomputadores no ranking da lista dos computadores mais rápidos do mundo, TOP500. Agora também existe o ranking Green500 que lista as máquinas baseadas em eficiência energética, em FLOPS por Watt. O benchmark HCP roda para matrizes de diferentes tamanhos N procurando pelo Nmax no qual o máximo desempenho Rmax é obtido. O benchmark também reporta o problema de tamanho N1/2 onde metade do desempenho (Rmax/2) é alcançado.

Uma das rotinas que mais consomem tempo no Linpack é a DAXPY. Arquitetos de computadores projetam sistemas para aperfeiçoar DAXPY com o intuito de obter uma alta avaliação no Linpack. Existem reclamações não verificadas de que o Linpack não seja um bom benchmark para supercondutores porque ele não estressa a interconexão entre nós, mas foca em unidades aritméticas de pontos flutuantes e memória cachê.

Uma característica interessante do Linpack é a referência de duas sub-rotinas : DGEFA e DGESL (estas são rotina que trabalha com ponto flutuante de 64 bits; já as SGEFA e SGESL trabalham normalmente com expressões de ponto flutuante de 32 bits). DGEFA realiza a decodificação parcial do vetor, e DGESL usa esse tipo de decodificação para resolver um determinado sistema de equações lineares. A maior parte das execuções de ponto flutuante gira em torno de O(n³), este é o tempo gasto em DGEFA. Uma vez que a matriz foi decomposta usando o DGESL, que trabalha com tempo de O(n²) operações de ponto flutuante.

Os resultados dos Benchmarks não devem ser utilizados como medidas de desempenho total do sistema (a não ser que a análise foi realizada suficiente para a carga de trabalho de um determinado interesse), mas sim como pontos de referências para novas avaliações.

O desempenho é freqüentemente medido em termos de Megaflops, Gigaflops, ou Teraflops. Incluem geralmente tanto adições e multiplicações na contagem das operações de ponto flutuante por segundo, e os valores dos operandos são assumidos de 64 bits de ponto flutuante. Atualmente, pode se medir o desempenho de um supercomputador chegando a Petaflops (1015flops/s).

Analisando o algoritmo Linpack e observando a forma de como os dados são referenciados, vemos que cada passo do processo de fatorização existe operações vetoriais que modificam completamente uma sub-matriz de dados. Esta atualização faz um bloco de dados ser lido, atualizando e escrevendo de volta para a memória principal.

O número de operações de ponto flutuante é 2 / 3n3, e o numero de dados referenciados, ambos loads e stores, é 2 / 3n3. Assim, para cada par soma/multiplicação temos de efetuar um load(carregar) e store(armazenar) dos elementos, infelizmente não obtendo a reutilização dos dados. Mesmo que as operações sejam completamente vetoriais, há uma movimentação significativa dos dados, resultando em um baixo desempenho. Em computadores vetoriais a um desempenho bem inferior ao pico estimado. Em computadores super-escalares isso resulta em uma grande quantidade de dados movimentados e atualizados. Para alcançar taxas de alto desempenho esta operação-memória-referência deve ser maior.

**Passmark PerformanceTest 7.0**

PerformanceTest é uma ferramenta de *benchmarking* rápida e fácil de usar que permite uma avaliação ágil do desempenho de computadores pessoais e compará-lo com um número de computadores de forma padrão.

Os resultados são apresentados em de uma forma fácil de ser compreendida através de gráficos de barra para cada métrica analisada, a avaliação de cada subsistema e a avaliação geral do desempenho do computador. A resolução dos contadores de tempo é na ordem de um milionésimo de segundos na maioria dos computadores. O resultado “PassMark rating” é uma média ponderada de todos os resultados dos subsistemas avaliados. Quanto maior o valor, mais rápido é o sistema. Esse valor só pode ser calculado se o resultado de todos os outros testes são disponíveis, com exceção do teste da unidade de disco óptico que pode ser omitido da avaliação. Os pesos das avaliações dos subsistemas são definidos de acordo com a Tabela 3.

Tabela 3. Pesos dos subsistemas no *PassMark rating.*

|  |  |
| --- | --- |
| Subsistema Testado | Peso |
| Disco Rígido | 21% |
| CD/ DVD | 5% |
| Memória | 19% |
| Gráficos 3D | 12% |
| Gráficos 2D | 14% |
| CPU | 29% |
| Total | 100% |

**Realização de medições**

Para esta metodologia, deve-se compilar os benchmarks sintéticos em cada máquina e gerar os executáveis compatíveis com a arquitetura do computador.

Então, o experimento será reproduzido um número finito de vezes para os *benchmarks* Dhrystone, Linpack e PerformanceTest 7, seguindo das configurações padrão.

O número de medições deve ser razoavelmente grande, trinta ou mais medições podem ser uma boa estimativa de população. Para esta metodologia, será adotado um total de 30(trinta) reproduções de cada *benchmark* em cada computador.

A resolução (menor alteração do dado que pode ser detectada pelo instrumento de medição) utilizada nesta metodologia será de duas casas decimais.

Além dos dados, o registro deverá conter as condições no momento da medição. O analista responsável deve ter domínio sobre a metodologia empregada e a ferramenta de medição.

A coleta de dados será registrada através de arquivo de texto. A carga de trabalho deverá ter registro das métricas de interesse, facilitando o trabalho de coleta.

Ao final da medição, o analista deve verificar se a resolução, precisão e exatidão estão aceitáveis. Caso não esteja, a ferramenta deve ser ajustada e o experimento revisto. O analista deve detectar os *outliers* e ser capaz de justificá-los. O produto desta etapa é o protocolo de medição preenchido.

**Análise dos dados**

A partir da interpretação dos dados, o analista pode inferir que sistema desempenhará melhor sobre determinado cenário com a carga de trabalho utilizada, justificando o padrão dos resultados obtidos.

Nesta metodologia, a análise dos dados obtidos deve ser realizada utilizando a ferramenta estatística Minitab. Utilizar as funções “*Graphical Summary*”, “*Decriptive Statistics*”, “*T Paired Test*” e “*2 Sample T*”.

**Validação dos Dados**

Os relatórios de *benchmarking* são concebidos de forma a mostrar os resultados do exercício de *benchmarking* realizado, ou seja, o desempenho relativo de computadores face a um grupo resultante dos critérios de *benchmarking* escolhidos, e são apresentados sob a forma de gráficos e de tabelas:

* Os gráficos mostram a posição relativa dos computadores para cada indicador; quanto mais elevada for a posição nos gráficos, melhor é o desempenho.
* As tabelas apresentam os dados reais dos computadores e da amostra, proporcionando uma desagregação dos dados em cinco intervalos de percentis:
  + Baixo:  valor correspondente ao percentil 5 da base de dados. Foi escolhido este critério para que os verdadeiros extremos da base de dados não sejam apresentados;
  + Quartil Inferior: valor correspondente ao percentil 25;
  + Média:  valor que representa o percentil 50 (também conhecido por Mediana) da base de dados;
  + Quartil Superior: valor que representa o percentil 75 (também conhecido por Quartil Superior) da base de dados;
  + Superior: valor correspondente ao percentil 95 da base de dados. Mais uma vez foi escolhido este critério para que os verdadeiros extremos da base de dados não sejam apresentados.

A validação dos dados é realizada para verificar a natureza dos dados e observar as suposições da comparação. Além disso, é possível para garantir que os resultados estão dentro de um intervalo de confiança válido.

# Seguindo a metodologia estabelecida, avaliem, apresentem os resultados e o diagnóstico do desempenho de dois computadores de fabricantes distintos (ou máquinas virtuais com diferentes configurações) X86. A avaliação deve considerar pelo menos os *benchmarks* *Dhrystone* e *Linpack*.

**Planejamento**

***Definição do objetivo***

O objetivo desse trabalho é avaliar e comparar o desempenho de dois computadores de fabricantes distintos com mesmo sistema operacional, cujos processadores devem ser da família X86, porém somente teve-se a disponibilidade de utilizar um computador com arquitetura de processador X86 e outro X64. Apesar disso, não se verifica problemas quanto ao uso das ferramentas e a aplicação do método, devido a disponibilidade de os *benckmarks* terem versões compatíveis com ambas arquiteturas e o método ser generalizado.

As métricas definidas para medição foram as seguintes:

* Processador: A métrica é uma composição das abaixo apresentadas.
  + CPU - Integer Math
  + CPU - Floating Point Math
  + CPU - Find Prime Numbers
  + CPU - Multimedia Instructions
  + CPU – Compression
  + CPU – Encryption
  + CPU – Physics
  + CPU - String Sorting
* Gráficos 2D: A métrica é uma composição das abaixo apresentadas.
  + Graphics 2D - Solid Vectors
  + Graphics 2D - Transparent Vectors
  + Graphics 2D - Complex Vectors
  + Graphics 2D - Fonts and Text
  + Graphics 2D - Windows Interface
  + Graphics 2D - Image Filters
  + Graphics 2D - Image Rendering
* Gráficos 3D: A métrica é uma composição das abaixo apresentadas.
  + Graphics 3D – Simple
  + Graphics 3D – Medium
  + Graphics 3D – Complex
* Memória: A métrica é uma composição das abaixo apresentadas.
  + Memory - Allocate Small Block
  + Memory - Read Cached
  + Memory - Read Uncached
  + Memory – Write
  + Memory - Large RAM
* Disco Rígido: A métrica é uma composição das abaixo apresentadas.
  + Disk - Sequential Read
  + Disk - Sequential Write
  + Disk - Random Seek + RW
* VAX MIPS rating: tempo de execução de um conjunto de instruções com tipos de dados inteiros.
* MFLOPS Speed: Velocidade da execução de um conjunto de instruções que operam tipos de dados numéricos com pontos flutuantes.

***Descrição do Cenário***

O cenário é composto por dois computadores com as seguintes descrições:

**Computador 1:**

*Sumário do Sistema:*

Windows 7 (32-bit)  
Intel Core2 Duo T6500 @ 2.10GHz  
2974 MB RAM  
Mobile Intel(R) 4 Series Express Chipset Family  
466GB HDD  
CD-RW/DVDRW

*Informações da Memória:*

Total Physical Memory: 2974 MB RAM  
Available Physical Memory: 2047 MB RAM  
Slot 1: DDR2, 2048MB, 800MHz  
Slot 2: DDR2, 2048MB, 800MHz  
Virtual Memory: C:\pagefile.sys (4461 MB)

*Informações da CPU:*

Manufacturer: GenuineIntel  
Type: Intel Core2 Duo T6500 @ 2.10GHz  
Number of CPU's: 1  
Cores per CPU: 2  
Hyperthreading: Not capable  
Measured Speed: 2095.1 MHz  
Multiplier: 11X  
Bus Speed: 200Mhz  
Front Side Bus Speed: 200Mhz  
L1 Instruction Cache: 2 x 32 KB  
L1 Data Cache: 2 x 32 KB  
L2 Cache Size: 1 x 2 MB  
L3 Cache: (N/A)

*Informações do Disco:*

Drive Letter (Number): C (Physical drive 0)  
Model Number: TOSHIBA MK5055GSX  
Disk Size (Free space): 465.7 GBytes (263.9 GBytes)  
Disk Cluster Size: 4 KBytes  
File System: NTFS

*Adaptadores de Vídeo:*

Description: Mobile Intel(R) 4 Series Express Chipset Family  
Chip Type: Mobile Intel(R) 4 Series Express Chipset Family  
DAC Type: Internal  
Memory: 1295MB  
Video BIOS: Intel Video BIOS  
Driver Provider: Intel Corporation  
Driver Version: 8.15.10.2202  
Driver Date: 8-25-2010  
Monitor 1: 1366x768x32 59Hz (Primary monitor)

**Computador 2:**

*Sumário do Sistema:*

PerformanceTest 7.0 1022 32-bit  
Windows 7 (64-bit)  
Intel Core i5 M 480 @ 2.67GHz  
5942 MB RAM  
Intel(R) HD Graphics  
466GB HDD  
CD-RW/DVDRW, CD/DVDRW/BD/HD DVD-ROM

*Informações da Memória:*

Total Physical Memory: 5942 MB RAM  
Available Physical Memory: 4643 MB RAM  
Slot 1: 2048MB, 1333MHz  
Slot 2: 2048MB, 1333MHz  
Virtual Memory: C:\pagefile.sys (5942 MB)

*Informações da CPU:*

Manufacturer: GenuineIntel  
Type: Intel Core i5 M 480 @ 2.67GHz  
Number of CPU's: 1  
Cores per CPU: 2  
Hyperthreading: Enabled  
Measured Speed:   
L1 Instruction Cache: 4 x 32 KB  
L1 Data Cache: 4 x 32 KB  
L2 Cache Size: 4 x 256 KB  
L3 Cache: 3 MB

*Informações do Disco:*

Drive Letter (Number): C (Physical drive 0)  
Model Number: ST9500325AS  
Disk Size (Free space): 453.4 GBytes (412.5 GBytes)  
Disk Cluster Size: 4 KBytes  
File System: NTFS

*Adaptadores de Vídeo:*

Description: Intel(R) HD Graphics  
Chip Type: Intel(R) HD Graphics (Core i5)  
DAC Type: Internal  
Memory: 2747MB  
Video BIOS: Intel Video BIOS  
Driver Provider: Intel Corporation  
Driver Version: 8.15.10.2202  
Driver Date: 8-25-2010  
Monitor 1: 1366x768x32 60Hz (Primary monitor)

Foram considerados somente os processos vitais aos componentes dos sistemas avaliados.

Os pré-requisitos, as pré-condições, os fornecedores, os insumos, as ações, os produtos, os consumidores e as pós-condições das atividades planejadas foram seguidas.

As máquinas foram reiniciadas para limpar a memória *cache* do processador. As placas de rede foram desconectadas devido ao processamento que ocorre nos protocolos de comunicação. Os dois computadores estarão sobre temperatura ambiente, pois a temperatura causa alterações no desempenho dos circuitos integrados. Os dois sistemas estão isolados, pois evita interferências do meio externo. A configuração do consumo de energia das máquinas foi alterada para não entrar em modo de espera, com o intuito de evitar que a execução dos processos seja interrompida.

Os computadores estão conectados ao carregador da bateria. Todos os processos indesejados que consumiam recursos foram finalizados. Não houve outro programa previamente executado. O computador não está conectado a nenhuma rede e a nenhum dispositivo de entrada e saída. Um computador tem 2 anos de uso (computador 1) e o outro tem 1 mês (computador 2). Há somente um sistema operacional instalado. Foi realizada a limpeza da cache de dados e da memória. Não há imcompatibilidade de drivers, pois todos os drivers instalados são do mesmo fabricante do dispositivo. Foi realizada a verificação de erros na memória e nenhuma irregularidade foi encontrada. Foi realizada a limpeza de arquivos desnecessários do disco. Não há fragmentação de arquivos do disco.

***Definição da cargas de trabalho***

Nesse caso de estudo, as cargas de trabalhos utilizadas são padrões e são configuradas automaticamente pelos *benchmarks* adotados na avaliação do desempenho.

Um ponto importante é que no PerformanceTest, as reproduções foram executadas estressando 2 núcleos de processadores em cada computador.

***Definição das Ferramentas***

As ferramentas adotadas foram compiladas nas duas arquiteturas de processadores.

**Dhrystone**

Dhrystone Benchmark Version 1.1 via C/C++ - 32 Bit Integers - Via MS 32-bit C/C++ Optimizing Compiler Version 13.10.3077 for 80x86.

Dhrystone Benchmark Version 2.1 via C/C++ - 32 Bit Integers - Via MS 32-bit C/C++ Optimizing Compiler Version 13.10.3077 for 80x86.

Dhrystone Benchmark Version 1.1 via C/C++ - 64 Bit Integers - Via MS C/C++ Compiler Version 14.00.40310.41 for AMD64.

Dhrystone Benchmark Version 2.1 via C/C++ - 64 Bit Integers - Via MS C/C++ Compiler Version 14.00.40310.41 for AMD64.

**Linpack**

Linpack SSE2 Double Precision Unrolled Benchmark n @ 100 Via Microsoft 32-bit C/C++ Optimizing Compiler Version 15.00.30729.207 for 80x86.

Linpack SSE2 Double Precision Unrolled Benchmark n @ 100 Via Microsoft C/C++ Optimizing Compiler Version 15.00.30729.207 for x64

**Passmark PerformanceTest 7.0**

PerformanceTest 7.0 1022 32-bit.

**Análise dos dados**

Nessa metodologia, a análise dos dados obtidos foi realizada utilizando a ferramenta estatística Minitab. Utilizando as funções “*Graphical Summary*”, “*Descriptive Statistics*”, “*T Paired Test*” e “*2 Sample T*”.

Os dados coletados no computador 1 são apresentados na Tabela 4, já os dados coletados no computador 2 são apresentados na Tabela 5.

Tabela 4. Resultados dos *benchmarks* para o computador 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Computador 1 | | | |
| PerformanceTest 7.0 | Dhrystone 1.1 | Dhrystone 2.1 | Linpack |
| 627,5 | 6936,03 | 4763,93 | 1289,17 |
| 625,9 | 7007,65 | 4790,07 | 1289,16 |
| 629,9 | 6940,65 | 4786,24 | 1288,9 |
| 631,7 | 6989,12 | 4769,09 | 1289,31 |
| 627,3 | 6948,86 | 4790,28 | 1288,59 |
| 627,3 | 6955,32 | 4793,01 | 1288,22 |
| 629,6 | 6955,63 | 4787,57 | 1288,07 |
| 623,8 | 6951,8 | 4747,31 | 1288,97 |
| 628,6 | 6983,46 | 4761,59 | 1287,74 |
| 623,6 | 6908,77 | 4778,24 | 1287,75 |
| 630 | 6953,38 | 4795,76 | 1287,44 |
| 624,3 | 6973,03 | 4781,09 | 1287,71 |
| 630,8 | 6964,49 | 4787,38 | 1286,79 |
| 633,6 | 6968,99 | 4764,46 | 1287,89 |
| 630,8 | 7014,68 | 4791,21 | 1288,34 |
| 626,6 | 6989,19 | 4776,89 | 1288,05 |
| 636,8 | 6959,88 | 4786,22 | 1288,17 |
| 622,8 | 6992,51 | 4774,1 | 1287,28 |
| 626,8 | 6970,28 | 4773,45 | 1288,4 |
| 628,4 | 6989,98 | 4748,41 | 1287,95 |
| 629,3 | 6969,18 | 4772,75 | 1288,58 |
| 623,6 | 6985,38 | 4782,7 | 1289,55 |
| 636,7 | 6967,98 | 4790,45 | 1289,09 |
| 633,1 | 6920,42 | 4791,19 | 1287,95 |
| 631 | 7000,13 | 4790,36 | 1288,6 |
| 622,6 | 6983,26 | 4787,2 | 1288,33 |
| 625,3 | 6997,91 | 4783,37 | 1287,68 |
| 628,7 | 6956,21 | 4767,16 | 1287,88 |
| 621,5 | 6931,34 | 4774,17 | 1288,08 |
| 623,9 | 6949,11 | 4790,83 | 1287,49 |

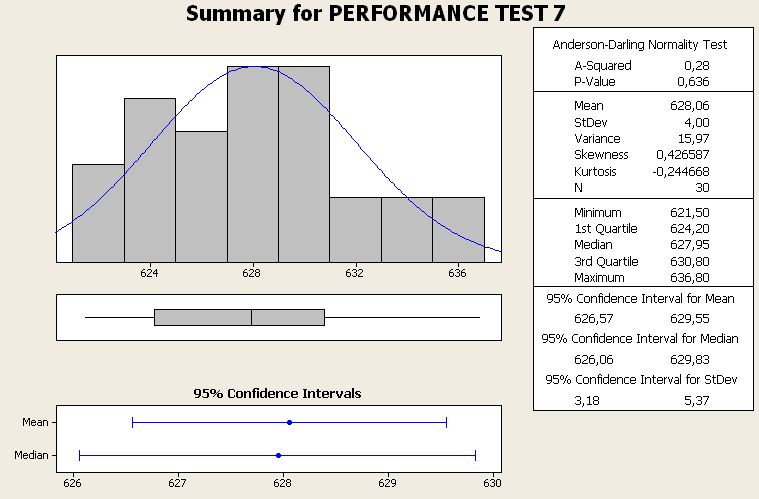
Tabela 5. Resultados dos *benchmarks* para o computador 2.

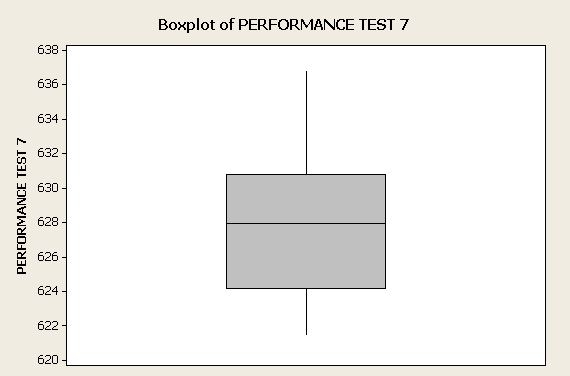
|  |  |  |  |
| --- | --- | --- | --- |
| Computador 2 | | | |
| PerformanceTest 7.0 | Dhrystone 1.1 | Dhrystone 2.1 | Linpack |
| 887,70 | 14924,16 | 9589,36 | 2089,34 |
| 880,30 | 15228,42 | 9600,95 | 2115,41 |
| 892,40 | 15228,65 | 9583,85 | 2124,95 |
| 842,40 | 15232,44 | 9586,92 | 2126,04 |
| 904,10 | 15237,90 | 9588,96 | 2120,05 |
| 882,60 | 15229,38 | 9592,57 | 2088,43 |
| 885,80 | 15176,28 | 9576,99 | 2103,43 |
| 885,80 | 15227,62 | 9588,65 | 2133,23 |
| 896,00 | 15225,97 | 9592,35 | 2124,11 |
| 907,40 | 15225,96 | 9591,91 | 2124,34 |
| 863,30 | 15225,84 | 9599,95 | 2118,44 |
| 889,00 | 15222,67 | 9592,08 | 2120,85 |
| 863,70 | 15226,87 | 9594,87 | 2124,81 |
| 892,40 | 15219,20 | 9595,07 | 2118,19 |
| 905,90 | 15228,85 | 9592,84 | 2110,80 |
| 890,60 | 15215,72 | 9592,70 | 2113,26 |
| 899,80 | 15225,85 | 9596,26 | 2111,57 |
| 906,60 | 15227,52 | 9460,61 | 2114,04 |
| 905,90 | 15233,25 | 9432,82 | 2108,23 |
| 901,30 | 15228,15 | 9593,68 | 2120,30 |
| 902,80 | 14967,63 | 9590,58 | 2117,93 |
| 901,90 | 15192,69 | 9594,01 | 2125,42 |
| 858,80 | 14945,29 | 9529,73 | 2103,10 |
| 913,30 | 15230,62 | 9585,84 | 2101,30 |
| 912,00 | 15226,27 | 9583,61 | 2118,60 |
| 867,30 | 15225,42 | 9588,90 | 2105,76 |
| 880,40 | 15221,71 | 9537,37 | 2115,15 |
| 921,00 | 15223,88 | 9581,98 | 2118,15 |
| 918,80 | 15228,28 | 9540,38 | 2116,02 |
| 872,20 | 15221,25 | 9588,60 | 2121,45 |

Olhando para os dados contidos nas Tabelas, é possível notar que o computador 2 tem o desempenho melhor que o computador 1. A abordagem que será realizada a seguir, serve para o caso da comparação do desempenho de um sistema que sofreu um *upgrade* ou dois computadores distintos.

**Computador 1:**

*Graphical Summary*:





O sumário gráfico revela que para o *benckmark* PerformanceTest 7.0, a amostra pode ser dita originar-se de uma distribuição normal, devido a média, a moda e a mediana serem relativamente próximas, a curtose e a assimetria serem próximas de zero e o teste de normalidade de Anderson-Darling ter um *P-value* acima de 0.05.

Variable Mean SE Mean StDev Variance Minimum Q1 Median

PERFORMANCE TEST 7 628,06 0,730 4,00 15,97 621,50 624,20 627,95

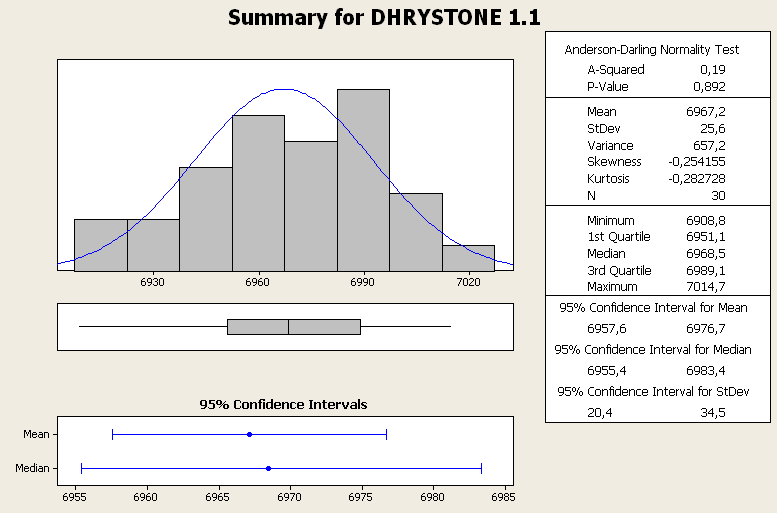
N for

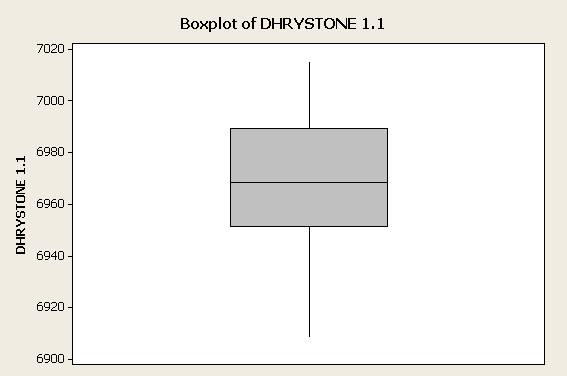
Variable Q3 Maximum Range IQR Mode Mode

PERFORMANCE TEST 7 630,80 636,80 15,30 6,60 623,6; 627,3; 630,8 2

Variable Skewness Kurtosis

PERFORMANCE TEST 7 0,43 -0,24





O sumário gráfico revela que para o *benckmark* Dhrystone v1.1, a amostra pode ser dita originar-se de uma distribuição normal, devido a média, a moda serem relativamente próximas, a curtose e a assimetria serem próximas de zero e o teste de normalidade de Anderson-Darling ter um *P-value* acima de 0.05.

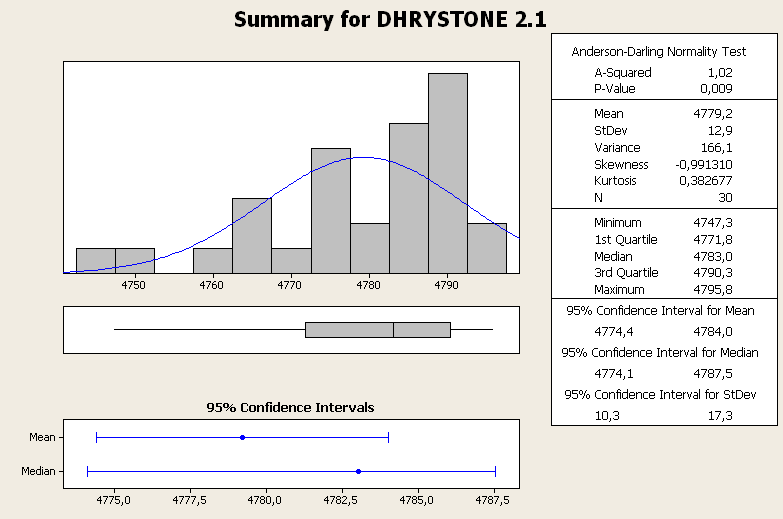
Variable Mean SE Mean StDev Variance Minimum Q1 Median

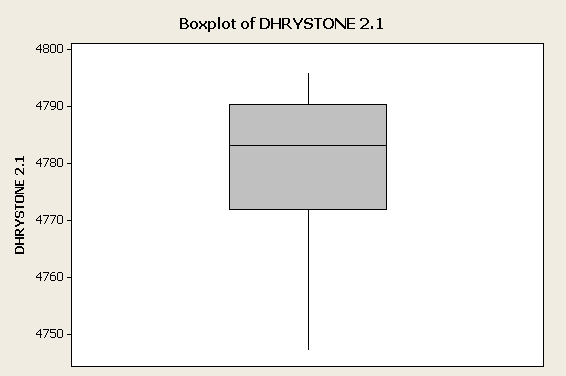
DHRYSTONE 1.1 6967,2 4,68 25,6 657,2 6908,8 6951,1 6968,5

N for

Variable Q3 Maximum Range IQR Mode Mode Skewness Kurtosis

DHRYSTONE 1.1 6989,1 7014,7 105,9 38,0 \* 0 -0,25 -0,28





O sumário gráfico revela que para o *benckmark* Dhrystone v2.1, a amostra pode ser dita originar-se de uma distribuição normal, devido a média e a mediana serem relativamente próximas, a curtose e a assimetria serem próximas de zero e o teste de normalidade de Anderson-Darling ter um *P-value* acima de 0.05, porém não muito.

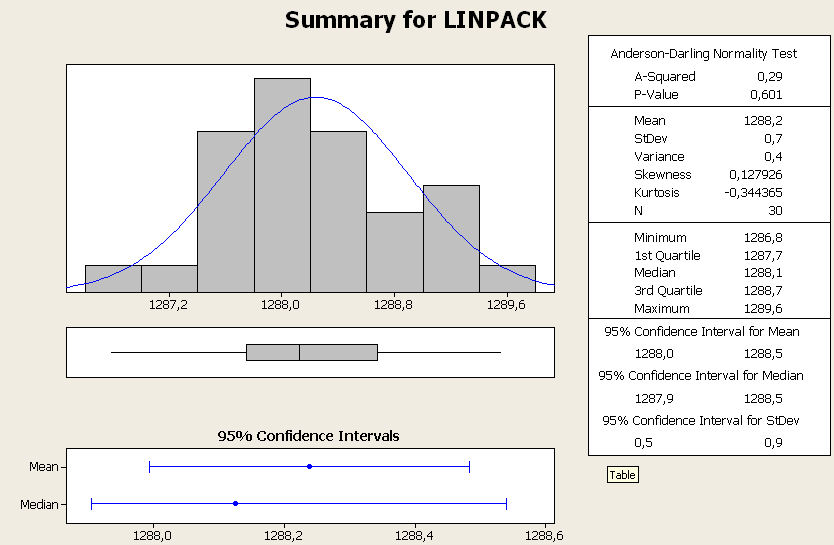
Variable Mean SE Mean StDev Variance Minimum Q1 Median

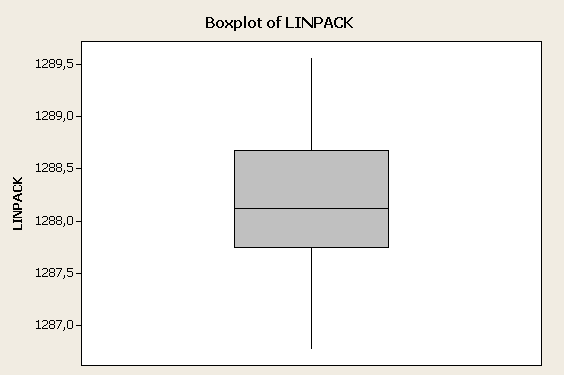
DHRYSTONE 2.1 4779,2 2,35 12,9 166,1 4747,3 4771,8 4783,0

N for

Variable Q3 Maximum Range IQR Mode Mode Skewness Kurtosis

DHRYSTONE 2.1 4790,3 4795,8 48,4 18,5 \* 0 -0,99 0,38





O sumário gráfico revela que para o *benckmark* Linpack, a amostra pode ser dita originar-se de uma distribuição normal, devido a média, a moda e a mediana serem relativamente próximas, a curtose e a assimetria serem próximas de zero e o teste de normalidade de Anderson-Darling ter um *P-value* acima de 0.05.

Variable Mean SE Mean StDev Variance Minimum Q1 Median Q3

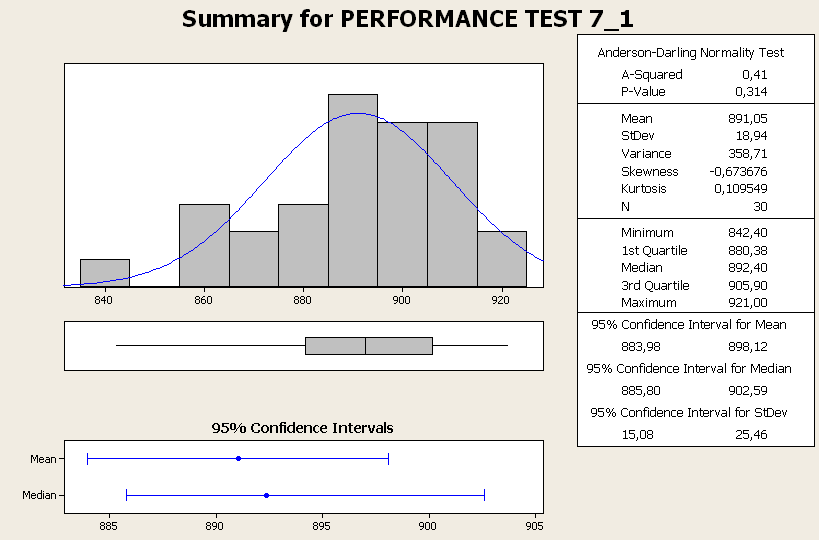
LINPACK 1288,2 0,120 0,657 0,431 1286,8 1287,7 1288,1 1288,7

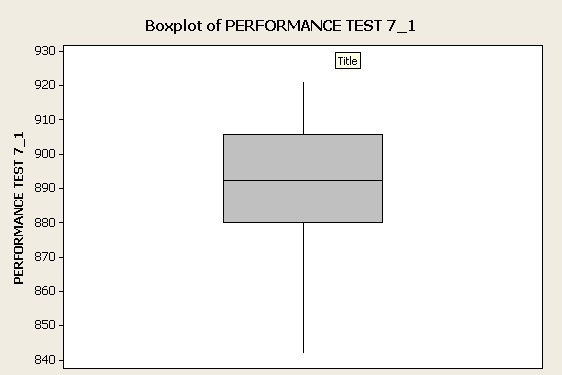
N for

Variable Maximum Range IQR Mode Mode Skewness Kurtosis

LINPACK 1289,6 2,76 0,928 1287,95 2 0,13 -0,34

**Computador 2:**





O sumário gráfico revela que para o *benckmark* PerformanceTest 7.0, a amostra pode ser dita originar-se de uma distribuição normal, devido a média, a moda e a mediana serem relativamente próximas, a curtose e a assimetria serem próximas de zero e o teste de normalidade de Anderson-Darling ter um *P-value* acima de 0.05.

Variable Mean SE Mean StDev Variance Minimum Q1 Median

PERFORMANCE TEST 7\_1 891,05 3,46 18,94 358,71 842,40 880,38 892,40

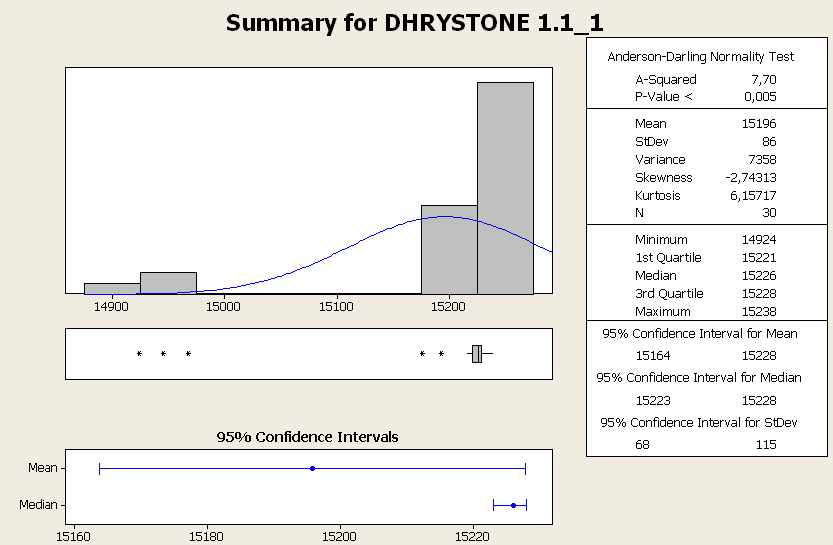
N for

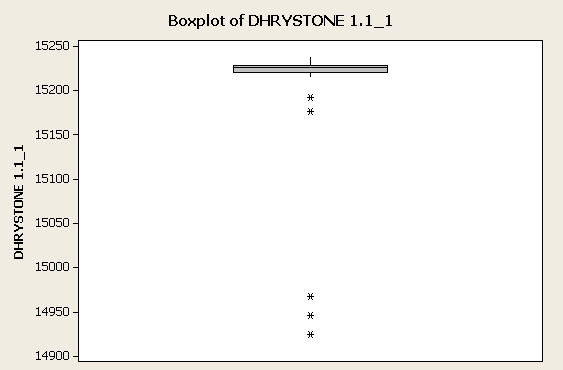
Variable Q3 Maximum Range IQR Mode Mode

PERFORMANCE TEST 7\_1 905,90 921,00 78,60 25,52 885,8; 892,4; 905,9 2

Variable Skewness Kurtosis

PERFORMANCE TEST 7\_1 -0,67 0,11





O sumário gráfico revela que para o *benckmark* Dhrystone 1.1, a amostra não pode ser dita originar-se de uma distribuição normal, apesar de a média e a mediana serem relativamente próximas, a curtose e a assimetria são diferentes de zero e o teste de normalidade de Anderson-Darling ter um *P-value* abaixo de 0.05.

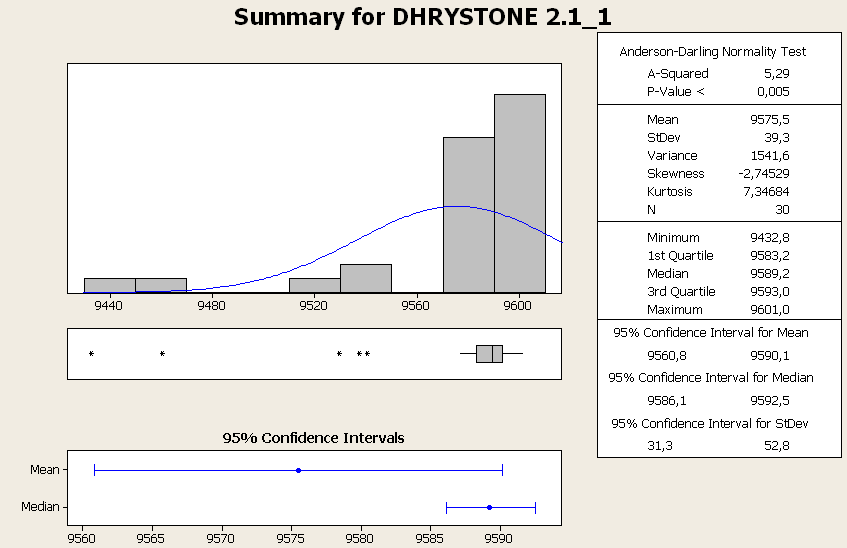
Variable Mean SE Mean StDev Variance Minimum Q1 Median Q3

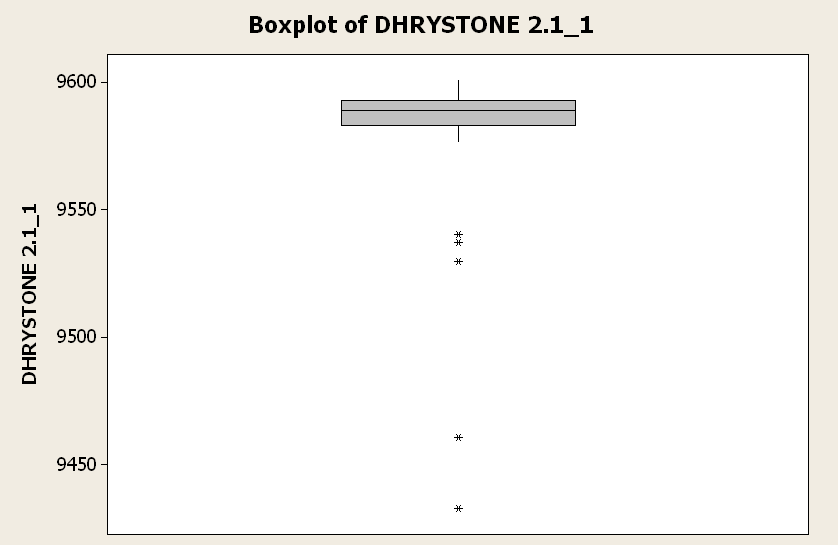
DHRYSTONE 1.1\_1 15196 15,7 85,8 7358 14924 15221 15226 15228

N for

Variable Maximum Range IQR Mode Mode Skewness Kurtosis

DHRYSTONE 1.1\_1 15238 314 7,7 \* 0 -2,74 6,16





O sumário gráfico revela que para o *benckmark* Dhrystone 2.1, a amostra não pode ser dita originar-se de uma distribuição normal, apesar de a média e a mediana serem relativamente próximas, a curtose e a assimetria são diferentes de zero e o teste de normalidade de Anderson-Darling ter um *P-value* abaixo de 0.05.

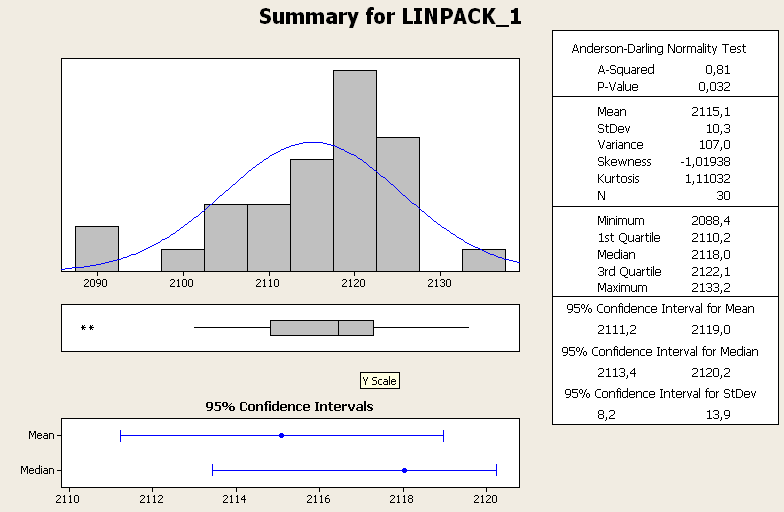
Variable Mean SE Mean StDev Variance Minimum Q1 Median

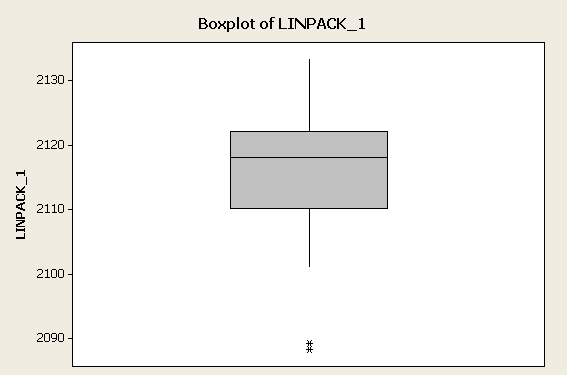
DHRYSTONE 2.1\_1 9575,5 7,17 39,3 1541,6 9432,8 9583,2 9589,2

N for

Variable Q3 Maximum Range IQR Mode Mode Skewness Kurtosis

DHRYSTONE 2.1\_1 9593,0 9601,0 168,1 9,85 \* 0 -2,75 7,35





O sumário gráfico revela que para o *benckmark* Linpack, a amostra não pode ser dita originar-se de uma distribuição normal, apesar de a média e a mediana serem relativamente próximas, a curtose e a assimetria são diferentes de zero e o teste de normalidade de Anderson-Darling ter um *P-value* abaixo de 0.05.

Variable Mean SE Mean StDev Variance Minimum Q1 Median Q3

LINPACK\_1 2115,1 1,89 10,3 107,0 2088,4 2110,2 2118,0 2122,1

N for

Variable Maximum Range IQR Mode Mode Skewness Kurtosis

LINPACK\_1 2133,2 44,8 12,0 \* 0 -1,02 1,11

*T Paired Test*:

Paired T for PERFORMANCE TEST 7 - PERFORMANCE TEST 7\_1

N Mean StDev SE Mean

PERFORMANCE TEST 7 30 628,06 4,00 0,73

PERFORMANCE TEST 7\_1 30 891,05 18,94 3,46

Difference 30 -262,99 20,00 3,65

95% CI for mean difference: (-270,46; -255,52)

T-Test of mean difference = 0 (vs not = 0): T-Value = -72,03 P-Value = 0,000

Paired T for DHRYSTONE 1.1 - DHRYSTONE 1.1\_1

N Mean StDev SE Mean

DHRYSTONE 1.1 30 6967,2 25,6 4,7

DHRYSTONE 1.1\_1 30 15195,8 85,8 15,7

Difference 30 -8228,6 86,3 15,8

95% CI for mean difference: (-8260,9; -8196,4)

T-Test of mean difference = 0 (vs not = 0): T-Value = -522,09 P-Value = 0,000

Paired T for DHRYSTONE 2.1 - DHRYSTONE 2.1\_1

N Mean StDev SE Mean

DHRYSTONE 2.1 30 4779,22 12,89 2,35

DHRYSTONE 2.1\_1 30 9575,48 39,26 7,17

Difference 30 -4796,26 40,48 7,39

95% CI for mean difference: (-4811,38; -4781,15)

T-Test of mean difference = 0 (vs not = 0): T-Value = -649,05 P-Value = 0,000

Paired T for LINPACK - LINPACK\_1

N Mean StDev SE Mean

LINPACK 30 1288,24 0,66 0,12

LINPACK\_1 30 2115,09 10,35 1,89

Difference 30 -826,85 10,42 1,90

95% CI for mean difference: (-830,74; -822,96)

T-Test of mean difference = 0 (vs not = 0): T-Value = -434,54 P-Value = 0,000

Realizando o Teste T pareado, verificamos que em todos os *benchmarks,* o desempenho do computador 2 é superior ao do computador 1. Além disso, como todos os intervalos de confiança da média das diferenças não contém zero, não há evidências para afirmar que o desempenho dos sistemas possam ser iguais.

*2 Sample T*:

Two-sample T for PERFORMANCE TEST 7 vs PERFORMANCE TEST 7\_1

N Mean StDev SE Mean

PERFORMANCE TEST 7 30 628,06 4,00 0,73

PERFORMANCE TEST 7\_1 30 891,0 18,9 3,5

Difference = mu (PERFORMANCE TEST 7) - mu (PERFORMANCE TEST 7\_1)

Estimate for difference: -262,99

95% CI for difference: (-270,20; -255,78)

T-Test of difference = 0 (vs not =): T-Value = -74,42 P-Value = 0,000 DF = 31

Two-sample T for DHRYSTONE 1.1 vs DHRYSTONE 1.1\_1

N Mean StDev SE Mean

DHRYSTONE 1.1 30 6967,2 25,6 4,7

DHRYSTONE 1.1\_1 30 15195,8 85,8 16

Difference = mu (DHRYSTONE 1.1) - mu (DHRYSTONE 1.1\_1)

Estimate for difference: -8228,6

95% CI for difference: (-8261,9; -8195,4)

T-Test of difference = 0 (vs not =): T-Value = -503,42 P-Value = 0,000 DF = 34

Two-sample T for DHRYSTONE 2.1 vs DHRYSTONE 2.1\_1

N Mean StDev SE Mean

DHRYSTONE 2.1 30 4779,2 12,9 2,4

DHRYSTONE 2.1\_1 30 9575,5 39,3 7,2

Difference = mu (DHRYSTONE 2.1) - mu (DHRYSTONE 2.1\_1)

Estimate for difference: -4796,26

95% CI for difference: (-4811,58; -4780,95)

T-Test of difference = 0 (vs not =): T-Value = -635,71 P-Value = 0,000 DF = 35

Two-sample T for LINPACK vs LINPACK\_1

N Mean StDev SE Mean

LINPACK 30 1288,238 0,657 0,12

LINPACK\_1 30 2115,1 10,3 1,9

Difference = mu (LINPACK) - mu (LINPACK\_1)

Estimate for difference: -826,85

95% CI for difference: (-830,72; -822,98)

T-Test of difference = 0 (vs not =): T-Value = -436,85 P-Value = 0,000 DF = 29

Realizando o Teste T para 2 Amostras, verificamos que em todos os *benchmarks,* o desempenho do computador 2 também é superior ao do computador 1. Além disso, como todos os intervalos de confiança da média das diferenças não contém zero, não há evidências para afirmar que o desempenho dos sistemas possam ser iguais. Porém, nesse teste a confiança é menor pois os intervalos de confiança são mais curtos que no Teste T Pareado.

Logo, podemos validar as amostras obtidas e verificamos que o computador 2 tem um desempenho acima do computador 1, como era de se esperar. Porém, em todos os casos, o computador 2 comportou-se de forma mais instável, o que diminui a precisão das amostras de todos *benchmarks*.

1. O valor p, ou *p-value*, indica o menor nível de significância com o qual a hipótese nula é rejeitada [↑](#footnote-ref-1)
2. O grau de siginificância, *α*, é o complemento do valor da confiança com a qual um intervalo de confiança é estabelecido. [↑](#footnote-ref-2)