

COMPARAÇÃO DO USO DO SOFTWARE R E MATLAB NA ESTATÍSTICA

BRINKHUS, Raquel Naile Bacharelando em Matemática no Centro Universitário
Internacional Uninter

PIANEZZER, Guilherme Augusto doutor em Métodos Numéricos em Engenharia pela
UFPR

RESUMO

O Presente trabalho comparou as linguagens R e Matlab no campo da estatística, ou seja, quais são as considerações importantes sobre o software R e o Matlab para o uso na estatística. Este trabalho se justifica pela necessidade de ter-se informações suficientes sobre as linguagens para que as pessoas possamos escolher qual é a melhor a ser utilizada em cada situação. Inicialmente foram realizadas pesquisas bibliográficas, e concluiu-se que ambos os softwares são parecidos (fácil utilização e boa interface gráfica), e como desvantagem o Matlab tem um alto custo de aquisição, enquanto o software R tem um tempo de processamento maior se comparado a outras linguagens. Após foi realizado um levantamento de algumas das principais funções na análise estatística: a média, mediana, valor máximo, valor mínimo, desvio padrão e variância. Na sequência realizou-se uma média de 10 testes de tempo de processamento de cada uma destas rotinas no software R e Matlab para duas populações de dados. Conclui-se que para o cálculo da média, mediana, valor máximo e valor mínimo o tempo de exceção foi menor no software Matlab, e para o cálculo do desvio padrão e variância, o tempo de execução foi menor no software R. Para os casos estudados, a diferença entre os resultados aumenta com o crescimento da população, ou seja, caso seja necessário analisar mais dados, a tendência é que a diferença no tempo de execução seja maior ainda.

Palavras-chave: Linguagem R. Linguagem Matlab. Estatística. Velocidade de processamento.

1. Introdução

Existem inúmeros softwares na atualidade e que podem resolver os mais diversos problemas, porém todos têm aspectos diferentes, tais como, facilidade de utilização, performance, quantidade de funções disponíveis, entre outros. Escolher um software ao acaso pode ocasionar na demora para obter as soluções desejadas ou até mesmo inviabilizar a busca pela solução de um problema. Com o crescimento da tecnologia a quantidade de dados disponíveis é cada vez maior, por isso, há uma crescente necessidade de ferramentas que não só analisem estas grandes quantidades de dados, mas que o faça no menor tempo possível. Dentre as linguagens mais utilizados para análise de dados e estatística está o software R e o Matlab, que possuem funções prontas e permitem a criação de rotinas que podem ir desde análises simples até muito complexas. Este trabalho tem como tema a comparação entre o software R e o Matlab no que se refere a facilidade de uso e velocidade de processamento quando são utilizados em problemas de estatística. Desta forma, a problematização deste trabalho é: Quais são as considerações importantes sobre o software R e o Matlab para o uso na estatística?

O objetivo geral deste trabalho é realizar uma análise comparativa entre o software R e Matlab no campo da estatística, com o intuito de comparar a facilidade de uso dos mesmos, suas vantagens e desvantagens, e tempo de processamento de funções estatísticas.

Os objetivos específicos serão: realizar uma pesquisa bibliográfica sobre o software R; realizar uma pesquisa bibliográfica sobre o software Matlab; documentar as vantagens e desvantagens destes dois softwares; e comparar a performance de cada um dos dois softwares para problemas comuns da estatística.

Para obter o embasamento teórico necessário sobre os dois softwares, inicialmente buscou-se sobre estes na literatura existente, artigos e materiais. Desta forma foi possível definir os dois softwares e descrever suas vantagens e desvantagens. A partir destas informações, propõe-se uma rotina com funções estatística para analisar o tempo de processamento de duas populações de dados em cada um dos dois softwares, e por fim comparar os resultados encontrados.

O autor Gilat (2012) apresenta o software Matlab, enquanto Chapman (2016) pontua as vantagens no seu uso. Com relação ao software R, Oliveira (2018) faz a apresentação, Barros (2018) aponta as vantagens de uso e Faria e Parga (2020) as desvantagens.

Este trabalho está organizado da seguinte forma: inicialmente serão apresentados os softwares Matlab e R e as suas vantagens de uso; após serão apresentadas as funções estatísticas disponíveis nos softwares; e para finalizar, será mostrado e comparado o tempo de processamento de cada função em cada software para cada uma das populações.

Na seção a seguir faz-se a metodologia e revisão básica da literatura que servirá para embasamento do presente estudo. Foram utilizados livros, artigos e trabalhos acadêmicos que abordam sobre os softwares R e Matlab, e sobre funções estatísticas.

2. Metodologia

A pesquisa usa de uma abordagem qualitativa de investigação e quantitativa de estudo de caso. Foram analisados os principais livros e artigos que tratem dos temas, investigação sobre as definições e vantagens das linguagens R e Matlab, e sobre as funções estatísticas.

Para o estudo quantitativo, foi realizado a comparação entre o tempo de execução de rotinas que usem funções comumente utilizadas na estatística para uma população com uma mesma quantidade de valores e precisão, em ambas as linguagens.

O primeiro cuidado que se teve para realizar está foi na medição do tempo de processamento, pois existem várias funções que retornam o tempo de processamento de uma função. Na linguagem R por exemplo, a função “system.time” retorna um vetor com 3 diferentes valores:

- Primeira coluna: tempo de processamento do usuário, ou seja, o tempo gasto no CPU apenas pelo processamento do software R;
- Segunda coluna: tempo do sistema, ou seja, tempo utilizado por outros processos que não são do software R (por exemplo, o sistema operacional do Windows ou outros programas abertos pelo usuário no computador);
- Terceira coluna: tempo decorrido, ou seja, tempo total sentido pelo usuário (soma da primeira e segunda coluna).

O tempo de processamento correto para comparação entre os softwares é aquele utilizado apenas pelo software R, desconsiderando outras aplicações que o usuário possa estar utilizando, por isto, foi utilizado os valores retornados da primeira coluna da função “system.time”.

No software Matlab as funções de tempo são um pouco diferentes, não existe uma função que retorna diretamente o tempo utilizado apenas pelo software para a execução de uma função, mas existe uma função chamada “cputime” que retorna o tempo utilizado apenas pelo software Matlab desde o início da sessão corrente. Desta forma, foi criada uma variável chamada to (representando o tempo inicial), nela foi utilizada a função “cputime” imediatamente antes de chamar a função que se desejava analisar, desta forma, a variável guardava o tempo de processamento gasto pelo Matlab até o momento anterior a chamada da função a ser analisada. Também foi criada outra variável chamada tf (representando o tempo final), nesta variável também era chamada à função “cputime”, porém imediatamente após realizar a função que se desejava analisar. Fazendo a diferença entre o tempo final e o tempo inicial, foi possível então obter o tempo gasto de processamento apenas pelo software Matlab para calcular a função desejada.

A tabela 1 apresenta algumas das rotinas de estatística disponíveis pelos softwares e o nome do comando utilizado para chamar cada uma destas rotinas em cada um dos dois softwares.

Tabela 1 – Funções estatísticas no software R e Matlab

Função	Software	
	Matlab	R
Média	mean	mean
Mediana	median	median
Moda	mode	*
Valor Máximo	max	max
Valor Mínimo	min	min
Desvio Padrão	std	sd
Variância	var	var
Coeficiente de Variação	vc	*

* Sem função pronta (apenas com uso de bibliotecas)

Fonte: da autora (2021).

Como o objetivo é de comparar as funções prontas e de fácil acesso dos softwares, foi descartado o cálculo da moda e do coeficiente de variação. Desta forma, foram analisadas as seguintes funções: Média (aritmética), Mediana, Valor máximo, Valor mínimo, Desvio padrão e Variância.

O tamanho das populações (quantidade de dados a serem analisados) foram escolhidos de forma que o tempo de processamento para execução de nenhuma das rotinas fosse muito pequeno, ou seja, nenhum dos tempos retornasse zero (as funções retornam o tempo de processamento em segundos). Desta forma, foram escolhidos dois casos de populações:

- Caso A: uma população de $1e8$ (cem milhões) de dados;
- Caso B: uma população de $1e9$ (um bilhão) de dados.

Outra preocupação foi com relação precisão dos dados. Para garantir que a comparação dos dados fosse igual entre dois softwares, utilizou-se de dados que variavam de 0 a 1 e com uma precisão de 4 casas decimais gerados aleatoriamente em ambos os softwares. Inicialmente pensou-se em utilizar exatamente os mesmos dados, ou seja, gerar os dados em um dos softwares, salvar estes dados em um arquivo de texto e após ler os dados no outro software. Porém a quantidade de dados era muito grande e acabou inviabilizando o processo, desta forma, optou-se em utilizar valores diferentes, mas, como citado, valores com a mesma precisão de casas decimais.

Além destas preocupações, também foram consideradas as quantidades de testes para uma mesma situação. É natural que mesmo se executando testes iguais, de uma mesma população de dados e em um mesmo software, os resultados de tempo de processamento não sejam exatamente os mesmos de um teste para outro. Desta forma, optou-se em repetir cada um dos testes 10 vezes e gerar uma média destes testes.

Após todas estas definições, foram realizados os roteiros para análise de tempo de execução das funções em cada um dos softwares. A se próxima seção apresenta a revisão bibliográfica em que foi feito um estudo das linguagens, uma apresentação das principais funções estatísticas, os resultados encontrados para o tempo de processamento em cada linguagem e as comparações destes resultados.

3. Revisão bibliográfica/ Estado da arte

3.1 Software R e Matlab

Os softwares R e Matlab são linguagens de programação criadas para análise matemática, ou seja, neles é possível desenvolver rotinas e utilizar funções matemáticas para resolver os mais diversos problemas.

Segundo Gilat (2012, p. 12) o Matlab “é bastante versátil em cálculos matemáticos, modelagens e simulações, análises numéricas e processamentos, visualização e gráficos, desenvolvimentos de algoritmos, etc.”.

As vantagens de uso do Matlab são muitas, segundo Chapman (2016, p. 2) podem ser destacadas a facilidade de uso, independência da plataforma, funções predefinidas, representações gráficas independentes de dispositivos, interface gráfica e o compilador Matlab. O mesmo autor ainda apresenta algumas das desvantagens, que são a linguagem interpretada e o alto custo de aquisição.

O software R amplamente utilizado para análise estatística, o mesmo foi criado justamente com este propósito, segundo Oliveira (2018, p. 9) “Podemos entender o R também como um conjunto de pacotes e ferramentas estatísticas, munido de funções que facilitam sua utilização, desde a criação de simples rotinas até análises de dados complexas, com visualizações bem acabadas”.

Barros (2018, p. 1) cita algumas vantagens da linguagem R, tais como uma linguagem simples, eficaz e com instalações gráficas para análise, a autora também complementa “grande coleção coerente e integrada de ferramentas intermediárias para análise de dados”. (BARROS, 2018, p. 1)

Oliveira (2018, p. 9) também apresenta várias vantagens da linguagem R, a citar: “é completamente gratuito e de livre distribuição; muito fácil de se aprender; enorme quantidade de tutoriais e ajuda disponíveis gratuitamente na internet; amplamente utilizado pela comunidade acadêmica e pelo mercado”. (OLIVEIRA, 2018, p. 9)

Como desvantagens da linguagem R os autores Faria e Parga (2020, p. 5) citam “essas linguagens (Python e R) não são particularmente rápidas se comparadas com outras linguagens que lhe obrigam a especificar cada componente de sua análise”.

As duas linguagens são utilizadas na criação de rotinas matemáticas, e tem como vantagens a fácil utilização e boa interface gráfica. O Matlab tem como desvantagem o seu elevado custo de aquisição, enquanto o software R é gratuito, porém este tem como desvantagem não ser tão rápido se comparado a outras linguagens.

Estas ferramentas possuem diversas funções matemáticas e estatísticas prontas, e para os problemas mais complexos em que uma única função não resolva, é possível realizar uma rotina (sequência) de comandos.

3.2 Funções estatísticas

Com relação a estatística, Silva, Grams e Silveira (2018, p.13) apontam “Toda evolução humana dá-se em virtude de descobertas e invenções, que podem ser criadas ou adaptadas para contribuir e descomplicar a vida do homem, seja na área da saúde, engenharia, economia, comunicação, entre outras.”. Desta forma percebemos que todas as áreas necessitam para o avanço da descoberta e invenções. As autoras complementam “Essa evolução se deve em grande parte à análise de dados coletados nas mais diversas áreas” (SILVA, GRAMS e SILVEIRA, 2018, p.13), assim concluímos que a análise de dados tem a contribuir com todas as áreas.

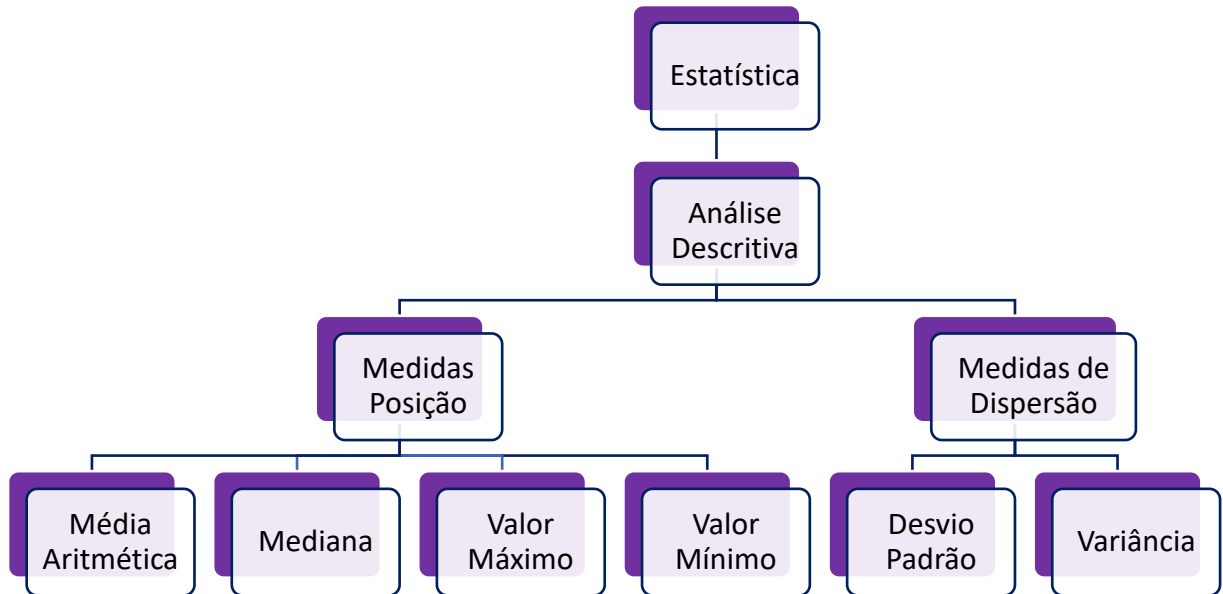
Silva, Grams e Silveira (2018, p.13) também comentam “coletar e analisar tais dados são funções da estatística, embasando decisões, planejamentos, sabendo como obter dados úteis e, principalmente, o que fazer com eles”. Podemos perceber que a estatística é fundamental para nortear as tomadas de decisões e para a melhor garantia de resultados.

Dentro da estatística há diversas técnicas que buscam auxiliar desde o processo de captação de dados até a sintetize e análise destes. A estatística descritiva é aquela auxilia no que fazer com os dados coletados, ou seja, como sintetizar e analisar o comportamento de um grupo de elementos.

As técnicas da estatística descritiva são divididas em dois grupos, as medidas de posição e as medidas de dispersão. Conforme Virgillito (2017, p. 77) “medidas de posição que ajudam a interpretar inicialmente a maioria dos eventos estudados”, e como exemplos tem-se os cálculos de médias, mediana, valor máximo e valor mínimo. O mesmo autor também comenta sobre as medidas de dispersão, “quando se fala em variabilidade ou oscilação” (VIRGILLITO, 2017, p. 77), e como exemplos tem-se o cálculo do desvio padrão e da variância. Cada uma destas técnicas será mais bem apresentada na sequência. A Figura 1 mostra esta área da estatística e apresenta algumas das técnicas.

Para Virgillito (2017, p. 77) “a média é o cálculo estatístico mais elementar”, o autor também complementa “talvez o cálculo mais instintivo do pesquisador e também dos profissionais” (VIRGILLITO, 2017, p. 77). Desta forma, talvez a média seja a função mais buscada, conhecida e esperada quando o objetivo é analisar inicialmente um conjunto de dados.

Figura 1: Estatística descritiva



Fonte: da autora (2021).

A média aritmética, ou apenas média como é frequentemente chamada, é definida por Costa Neto (2002, p. 21) “A média caracteriza o centro da distribuição de frequências, sendo, por isso, uma medida de posição. Em uma analogia de massas, a média corresponderia ao centro de gravidade da distribuição de frequências.”, e a mesma pode ser calculada conforme a equação 1.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Em que:

\bar{x} é a média;

n é o número total de valores;

x_i é cada valor.

A média não é melhor solução para todos os casos, quando deseja-se encontrar o valor central de (ideia do valor típico), a Mediana é mais recomendada. Costa Neto (2002, p. 21) defini “A mediana pode ser usada como alternativa, em relação à média, para caracterizar o centro do conjunto e dados. Em certos casos, efetivamente, seu uso é mais

conveniente.”. Caso o conjunto de valores analisados n seja ímpar a Mediana pode ser calculada conforme a equação 2.

$$md = x_{(n+1)/2} \quad (2)$$

Em que:

md é a mediana.

Caso o conjunto de valores analisados n seja par a Mediana pode ser calculada conforme a equação 3.

$$md = \frac{x_{(n/2)} + x_{(n/2+1)}}{2} \quad (3)$$

Além destas duas funções, dentro da estatística, ainda temos outras duas medidas que são importantes a serem definidas: os valores máximos e mínimos de um conjunto de dados.

O Valor mínimo pode ser definido conforme Becker (2015, p. 70) “O mínimo de um conjunto de dados é definido simplesmente como o seu menor valor.” Formalmente pode ser definido conforme a equação abaixo:

$$\min(X) = \min_i x_i, \text{ para } i = 1, \dots, n \quad (4)$$

Em que:

\min é o valor mínimo.

O Valor máximo também é definido conforme Becker (2015, p. 70) “O máximo de um conjunto de dados é definido simplesmente como o seu maior valor.” Formalmente pode ser definido conforme a equação abaixo:

$$\max(X) = \max_i x_i, \text{ para } i = 1, \dots, n \quad (5)$$

Em que:

\max é o valor máximo.

Apesar da média ser o cálculo mais elementar e intuitivo, nem sempre apenas essas informações de medidas de posição são suficientes. Virgillito (2017, p. 81) afirma “Em muitos casos, o simples cálculo da média não proporciona uma visão do que realmente acontece com o comportamento dos dados observados e, portanto, não oferece

informações para a tomada de decisão”. Desta forma, pode ser necessário utilizar também dos cálculos de medidas de dispersão (com por exemplo desvio padrão e variância).

O autor Virgillito (2017, p. 81) apresenta o conceito do desvio padrão: “medir o afastamento dos dados observados em relação à média da distribuição”. O mesmo autor também apresenta a sua definição “O desvio-padrão é definido como a “raiz quadrada da média, do quadrado dos desvios dos dados observados, em relação à média da distribuição destes”, conforme a equação 6.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (6)$$

Em que:

σ é o desvio padrão.

O valor do desvio padrão também está diretamente ligado a variância. Costa Neto (2002, p. 25) comenta “A variância é uma medida de dispersão extremamente importante na teoria estatística. Do ponto de vista prático, ela tem o inconveniente de se expressar numa unidade quadrática em relação à variável em questão”. A equação 7 apresenta a relação entre variância e desvio padrão.

$$var = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (7)$$

Em que:

var é a variância.

3.3 Tempo de Processamento

Após todas as considerações apresentada na metodologia e revisão bibliográfica, foram implementadas as rotinas computacionais que medem o tempo de processamento de cada função em ambos os softwares, fazendo uma média de 10 execuções para cada teste. A Tabela 2 apresenta os resultados de tempo encontrados para o Caso A e a Tabela 3 apresenta os resultados para o Caso B.

Tabela 2 – Tempo de execução Caso A

Função	Software	
	R (seg)	Matlab (seg)
Média	0,211	0,156
Mediana	1,988	1,436
Máximo	0,166	0,097
Mínimo	0,16	0,108
Desvio Padrão	0,459	0,648
Variância	0,464	0,586

Fonte: da autora (2021).

Tabela 3 - Tempo de execução Caso B

Função	Software	
	R (seg)	Matlab (seg)
Média	2,047	1,373
Mediana	31,019	25,922
Máximo	1,696	0,891
Mínimo	1,613	0,900
Desvio Padrão	4,475	19,153
Variância	4,474	18,597

Fonte: da autora (2021).

Como o objetivo do trabalho é a comparação entre os softwares, cabe ressaltar que durante a execução dos testes de desvio padrão e variância do Caso B de ambos os softwares levaram o uso da memória RAM ao máximo. A memória RAM utilizada para o processamento destes testes foi de 15,9 GB dos 16 GB (99,4% da memória disponível).

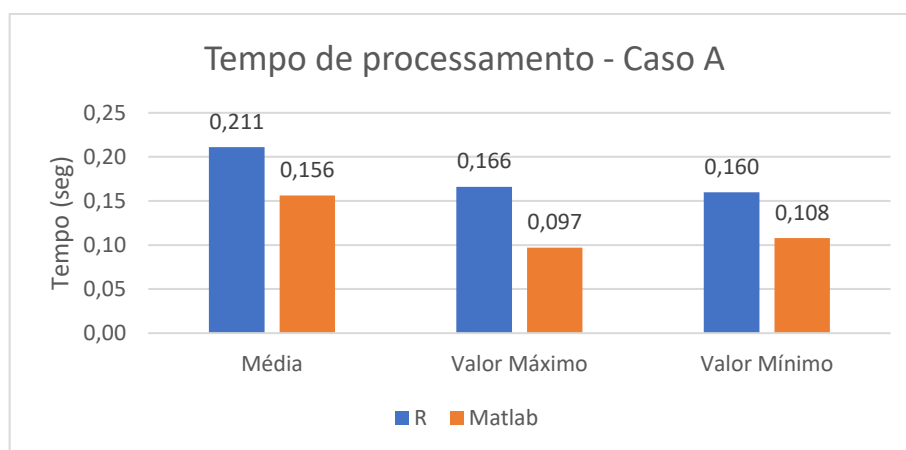
No software Matlab, o gerenciamento da utilização da memória é administrado pelo próprio software, ou seja, o usuário desenvolvedor da rotina não necessita fazer nenhuma adaptação no código ou intervenção. Já o software R não faz este gerenciamento sozinho, por padrão, o software permite a utilização apenas de uma parte da memória, quando esta memória limite é alcançada, o software para a execução e apresenta uma mensagem. Para contornar este problema e liberar o uso da memória, é necessário utilizar o comando “memory.limit” que expande o uso da memória.

Como os testes exigiram o uso de praticamente toda a memória RAM, no caso do software R houve a necessidade de realizar a adaptação na rotina para liberar o uso da memória, já no software Matlab não houve necessidade de realizar nenhuma alteração.

Ter a quantidade de uso da memória bloqueada tem aspectos positivos e negativos. O aspecto negativo, conforme já citado, é a interrupção da execução da rotina e adaptação do código. Já o lado positivo, é que evita que um software comprometa sozinho todo o uso da memória RAM. Nos casos em que a memória é utilizada apenas parcialmente, o usuário pode utilizar do restante da memória em outros softwares (realizando outras tarefas no computador); porém se a memória está totalmente liberada para o uso do software, o computador fica sem memória extra para utilizar em outras tarefas, fazendo com que o usuário tenha problemas caso tente realizar outras tarefas (demora excessiva).

A Figura 2 compara os resultados encontrados para o Caso A (cem milhões de dados). Para a função de média, o software R teve um tempo de processamento 35% mais lento se comparado ao Matlab, para a função de valor máximo, o software R teve um tempo de processamento 71,31% mais lento e para a função de valor mínimo, o software R teve um tempo de processamento 48,42% mais lento. Para este caso A, todas as funções analisadas tiveram um tempo de processamento menor no software Matlab.

Figura 2 – Tempo de processamento das funções média, valor máximo, valor mínimo para o Caso A

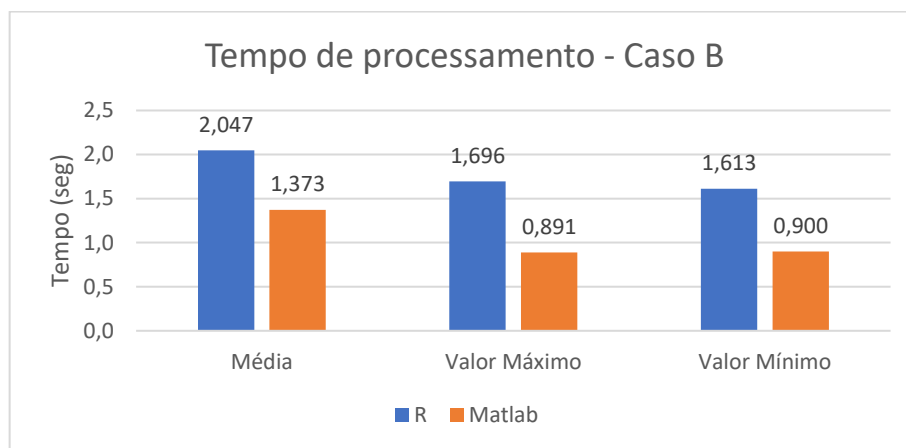


Fonte: da autora (2021).

A Figura 3 compara os resultados encontrados para o Caso B (um bilhão de dados). Para a função de média, o software R teve um tempo de processamento 49,05% mais lento se comparado ao Matlab, para a função de valor máximo, o software R teve um tempo de processamento 90,43% mais lento e para a função de valor mínimo, o software R teve um tempo de processamento 79,22% mais lento. Novamente todas as funções analisadas

tiveram um tempo de processamento menor no software Matlab, e além disto, também se percebeu que a comparação entre o tempo de processamento aumentou, para uma população maior, maior foi a diferença entre os tempos de execução.

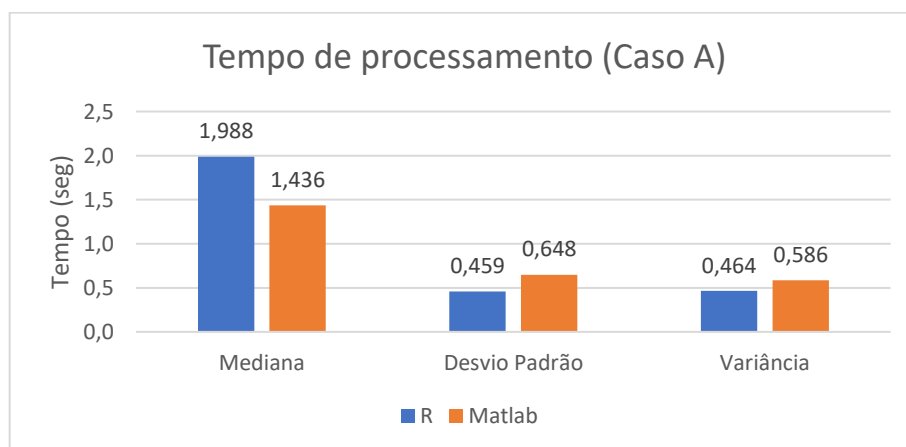
Figura 3 – Tempo de processamento das funções média, valor máximo, valor mínimo para o Caso B



Fonte: da autora (2021).

A Figura 4 compara os resultados encontrados para o Caso A (cem milhões de dados). Para a função de mediana, o software R teve um tempo de processamento 38,45% mais lento se comparado ao Matlab, para a função de desvio padrão, o software R teve um tempo de processamento 29,21% mais rápido e para a função de variância, o software R teve um tempo de processamento 20,81% mais rápido.

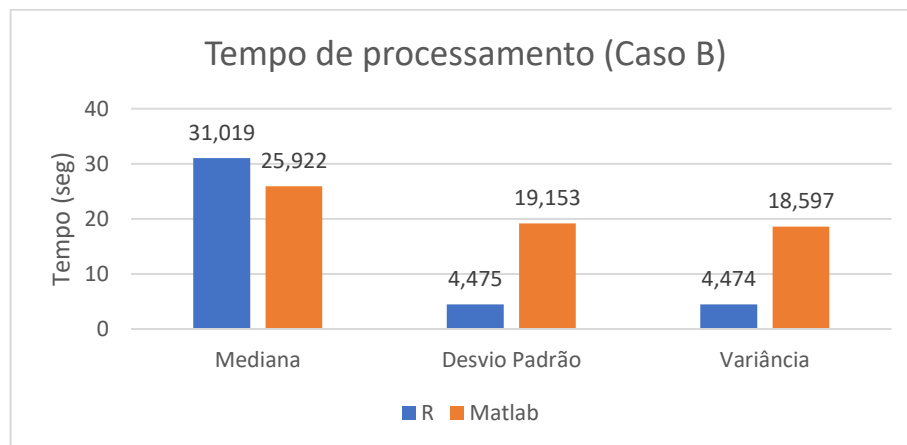
Figura 4 – Tempo de processamento das funções Mediana, Desvio Padrão e Variância para o Caso A



Fonte: da autora (2021).

A Figura 5 compara os resultados encontrados para o Caso B (um bilhão de dados). Para a função de mediana, o software R teve um tempo de processamento 19,66% mais lento se comparado ao Matlab, para a função de desvio padrão, o software R teve um tempo de processamento 76,64% mais rápido e para a função de variância, o software R teve um tempo de processamento 75,94% mais rápido.

Figura 5 – Tempo de processamento das funções Mediana, Desvio Padrão e Variância para o Caso B



Fonte: da autora (2021).

Para todas as seis funções analisadas, tanto para o caso A quanto no caso B, quatro das funções foram mais rápidas no Matlab e duas das funções foram mais rápidas no software R. Também se percebeu que com o aumento das quantidades de dados, maiores foram as diferenças entre os softwares, ou seja, com o aumento da quantidade de dados, maior foi a diferença entre o tempo de processamento das funções.

4. Considerações finais

O presente trabalho comparou os softwares Matlab e R. Ambos os softwares são linguagens de programação, utilizados para criação de rotinas matemáticas, são de fácil utilização e possuem recursos de interface gráfica. Após foram comparadas as desvantagens, em que o Matlab tem um alto custo de aquisição e o software R não é tão rápido se comparado a outras linguagens.

Após foram estudados quais os recursos disponíveis e mais utilizados dentro da estatística para analisar um grupo de dados. A estatística descritiva é a divisão que faz esta

análise de dados, e nela temos as medidas de posição (por exemplo o cálculo de média, mediana, valor máximo e valor mínimo) e as medidas de dispersão (por exemplo desvio padrão e variância).

Na sequência foi criado rotinas nas duas linguagens para realizar essas seis técnicas para duas populações de dados. Durante a execução dos testes, deparou-se com um problema de uso extremo da memória RAM, em que o software Matlab fez o gerenciamento do uso da memória de forma automática, enquanto o software R interrompeu a execução e precisou de adaptações no roteiro de cálculo para expandir o uso da memória.

Conforme explicado, este bloqueio de memória tem lados positivos e negativos. O software realizar o gerenciamento sozinho faz com que ele utilize toda a memória RAM, fazendo com que o usuário perceba este uso excessivo e não consiga realizar outras tarefas de modo normal (muita demora na para utilizar outros programas). Porém, o software não gerenciar sozinho o uso da memória faz com que a rotina seja interrompida e com que o desenvolvedor do roteiro tenha que ter conhecimentos extras de como liberar este uso.

Quatro das seis técnicas analisadas (média, mediana, valor máximo e valor mínimo) apresentaram valores de tempo execução inferiores no software Matlab, e com o aumento da população, as diferenças entre o tempo de processamento aumentou. Já nas outras duas técnicas (desvio padrão e variância), o tempo de processamento foi menor no software R, sendo que a diferença entre os softwares também cresce com o aumento da população.

Espera-se que este trabalho sirva de base para trabalhos futuros. Recomenda-se que sejam estudados outras funções estatísticas e outros tamanhos de população de dados para melhor comparar as linguagens e confirmar as tendências encontradas neste estudo.

Referências

BARROS, Anna Carolina *et al.* **Análise de séries temporais em R: curso introdutório**. Rio de Janeiro: Elsevier: FGV IBRE, 2018.

BECKER, João L. **Estatística básica: transformando dados em informação**. Porto Alegre: Bookman, 2015.

CHAPMAN, Stephen. J. **Programação em MATLAB para engenheiros**. 5 ed. São Paulo: Cengage Learning, 2016.

COSTA NETO, Pedro L. de O. **Estatística**. 3 ed. São Paulo: Editora Edgard Blücher Ltda, 2002.

FARIA, Pedro D.; PARGA, **João P. F. A. INTRODUÇÃO À LINGUAGEM R: Seus fundamentos e sua prática**. 1ed. Belo Horizonte: [s.n.], 2020. Disponível em: <https://pedro-faria.netlify.app/pt/publication/book/introducao_linguagem_r/>

GILAT, Amos. **MATLAB com aplicações em engenharia**. 4 ed. Porto Alegre: Bookman, 2012.

OLIVEIRA, P. F. de; GUERRA, S.; MCDONNELL, R. **Ciência de Dados com R Introdução**. Brasília: Editora IBPAD, 2018.

SILVIA, Juliane S. F. da; GRAMS, Ana L. B.; SILVEIRA, J. F. **Estatística**. Porto Alegre: SAGAH, 2018.

VIRGILLITO, Salvatore B. **Estatística aplicada**. 1 ed. São Paulo: Saraiva, 2017.