Universidade Federal de Viçosa
INF 222 - Computação Experimental - 2023/2
Trabalho Prático 3
Estudo comparativo de métodos para a resolução do problema Dutch National Flag

Integrantes: Danilo Freitas Vieira - 108201
Yuri Cardoso Bragine - 108199

Introdução: Este trabalho aborda o problema conhecido como Dutch National Flag, propondo uma análise comparativa de diferentes fatores e métodos de resolução. O problema consiste em ordenar um array de modo que todos os números negativos estejam antes dos zeros, e todos os números positivos venham depois dos zeros. Para a análise comparativa, são considerados diversos fatores, incluindo o algoritmo de ordenação utilizado, o tamanho do vetor, a linguagem de programação empregada, o estado inicial do array e a uniformidade dos valores.

Metodologia: Neste estudo, adotamos uma abordagem que compreende a criação de algoritmos para abordar a questão proposta, utilizando como base o algoritmo de Dijkstra e o MergeSort. O processo inclui experimentação com a variação de fatores, coleta de dados, análises comparativas e a verificação da conformidade com os resultados esperados. Essa metodologia pode ser desdobrada nas seguintes etapas:

Implementação dos Algoritmos de Ordenação: Nesse estágio, elaborou-se um programa capaz de criar um array de números decimais (double), cujo tamanho e nível de aleatoriedade são ajustáveis. Além disso, é possível escolher quantidades distintas de números negativos, zeros e positivos. Para realizar a ordenação, empregou-se o algoritmo MergeSort, juntamente com o código desenvolvido por Dijkstra para a resolução do Dutch National Flag.

Coleta de dados: Na etapa de coleta de dados, que envolve a realização dos experimentos e o registro dos tempos de execução, empregamos as linguagens de programação C++ e Python. Em ambas, foram utilizadas bibliotecas específicas, como a time em Python e a chrono em C++, responsáveis por realizar medições precisas dos tempos de execução em cada contexto e experimento.

Estudos Comparativos: Na fase de comparação, adotamos o método do Projeto Fatorial 2^k, 2^(k - p), e também 2^kr, conforme estudado na disciplina de

Computação Experimental. Esse enfoque foi aplicado para determinar as influências individuais de cada fator no desempenho do algoritmo.

Análise de Resultados: Concluindo o estudo comparativo, procedemos à interpretação dos resultados, avaliando se as hipóteses iniciais formuladas sobre os fatores foram confirmadas. Essa análise se baseou na porção de variação identificada para cada fator. Apresentamos as conclusões de maneira clara e objetiva, destacando eventuais não conformidades e as implicações dos algoritmos estudados no problema.

Estudo Comparativo:

Fatorial 2^4: Este teste é crucial para obter uma compreensão fundamental da interação entre os fatores unidirecionais, onde cada um é representado por valores mínimos e máximos.

Desde o início, é possível antecipar que o método de ordenação será um dos fatores mais impactantes no tempo de execução, especialmente porque o algoritmo MergeSort, embora notavelmente rápido, ordenará todos os números de um vetor muito extenso de maneira crescente. Em contraste, o algoritmo proposto por Dijkstra organiza o vetor sem considerar a ordenação dentro do intervalo de números negativos e positivos, permitindo uma ordenação como: −1.41, −9.8, 0, 0, 0, 0, 3.14, 0.577, 1.618, 2.718.

Quando se trata do mesmo método, é previsível que o tamanho do vetor seja o fator de maior impacto.

Para a representação na tabela, cada fator foi designado por uma letra para facilitar a compreensão. "M" refere-se ao método (Dijkstra ou MergeSort), "T" ao tamanho do array (500.000 ou 1 milhão), "E" ao estado inicial do array (completamente aleatório ou quase ordenado), e "Q" à quantidade de valores (quantidades iguais entre negativos, positivos e zeros ou diferentes). Isso resulta em 4 fatores, gerando 16x16 = 256 combinações na tabela, onde o símbolo "1" representa o nível máximo do fator e "-1" o mínimo.

Dessa forma temos:

Sinal	-1	+1
Método	Dijkstra	MergeSort
Tamanho do array	500 mil	1 milhão
Estado inicial	Completamente aleatório	Quase ordenado

Quantidade de valores	Não uniforme	Uniforme

A seguir, estão os resultados obtidos com a linguagem Python para este problema:

Matriz de Combinacoes:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	I	T1(s)
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.7315
1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	1	3.2501
1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	1	7.5924
1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	7.3177
1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	4.3113
1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	1	4.5804
1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	8.6823
1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	1	9.3032
-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	1	0.1459
-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	1	0.1689
-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	0.2896
-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1	1	0.3169
-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	1	0.1445
-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	0.1909
-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	1	0.309
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.4009

Na coluna **T1(s)** estão os tempos medidos nos experimentos de cada linha. Nessa etapa já é possível fazer uma análise preliminar de que quando o fator M é 1, ou seja o método é MergeSort os tempos são significativamente maiores do que quando o fator é -1, ao utilizar o algoritmo previsto por dijkstra.

A matriz a seguir é dada pela seguinte expressão aplicada em cada fator ou interação de fatores:

$$\sum_{i=1}^{2^k} S_{ij}\bar{y}_i$$

onde Sij é o sinal na tabela e $\bar{y}_i$ é o valor **T1(s)** do experimento correspondente,ou seja, o tempo gasto no experimento i.

Matriz das somas:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	I
46.8023	-5.1095	-17.6885	-0.3225	-4.8615	-16.3561	0.0547	1.2481	1.7341	0.6083	1.0825	1.5581	0.5087	-0.1863	-0.1039	50.7355

A matriz a seguir é dada pela seguinte expressão aplicada em cada fator ou interação de fatores:

$$\frac{1}{2^k}\sum_{i=1}^{2^k} S_{ij}\bar{y}_i$$

onde Sij é o sinal na tabela e $\bar{y}_i$ é o valor **T1(s)** do experimento correspondente,ou seja, o tempo gasto no experimento i.

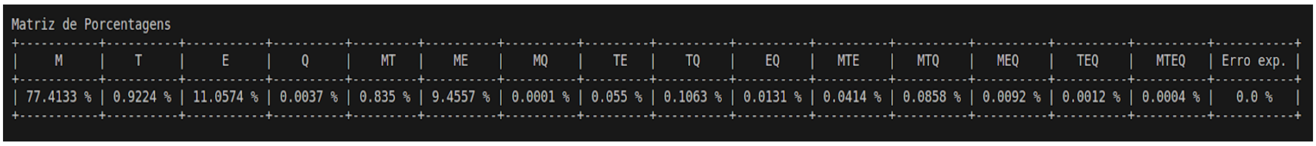
Matriz dos Qs:

qM	qT	qE	qQ	qMT	qME	qMQ	qTE	qTQ	qEQ	qMTE	qMTQ	qMEQ	qTEQ	qMTEQ	qI
2.9251	-0.3193	-1.1055	-0.0202	-0.3038	-1.0223	0.0034	0.078	0.1084	0.038	0.0677	0.0974	0.0318	-0.0116	-0.0065	3.171

A matriz a seguir é dada pela seguinte expressão:

$$M = \frac{2^4 \cdot qM^2}{SST}$$

onde M representa o percentual de variação referente a cada fator ou interação entre fatores e qM representa o efeito (q) de cada fator ou interação entre fatores.



Matriz de Porcentagens

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	Erro exp.
77.4133 %	0.9224 %	11.0574 %	0.0037 %	0.835 %	9.4557 %	0.0001 %	0.055 %	0.1063 %	0.0131 %	0.0414 %	0.0858 %	0.0092 %	0.0012 %	0.0004 %	0.0 %

A análise da última tabela revela claramente que o estudo aponta o fator "Método" como tendo o maior impacto no tempo do algoritmo, seguido pelo estado inicial do array.

Fatorial 2^(5−2): Este método assemelha-se ao anterior, mas reduz o número de experimentos para apenas 8x8 = 64 combinações. Essa abordagem é altamente útil para projetos que envolvem numerosos fatores inter-relacionados. No entanto, não é possível isolar os efeitos individuais de cada fator, uma vez que se trata de uma análise de efeitos combinados.

Os sinais referentes às linguagens de programação são os seguintes:

Sinal	-1	+1
Linguagem de programação	C++	Python

Para elaborar a tabela de sinais deste projeto, escolhemos criar a tabela de sinais para o fatorial completo dos três fatores principais: Método, Tamanho e Estado Inicial. Substituindo TE (interação entre Tamanho e Estado Inicial) por L (Linguagem) e MTE (interação entre Método, Tamanho e Estado Inicial) por Q (Quantidade), obtemos a seguinte tabela:

M	T	E	MT	ME	L	Q	Tempo(s)
1	1	1	1	1	1	1	Python(8.3498 s)
1	1	-1	1	-1	-1	-1	C++(0.50076 s)
1	-1	1	-1	1	-1	-1	C++(0.212589 s)
1	-1	-1	-1	-1	1	1	Python(4.7148 s)
-1	1	1	-1	-1	1	-1	Python(0.3713 s)
-1	1	-1	-1	1	-1	1	C++(0.011233 s)

-1	-1	1	1	-1	-1	1	C++(0.005616 s)
-1	-1	-1	1	1	1	-1	Python(0.2256 s)

Tabela das somas:

M	T	E	MT	ME	L	Q	I
13.1642	4.0744	3.4869	3.7719	3.2067	12.9304	11.7703	14.3917

Tabela dos Efeitos (Q's):

qM	qT	qE	qMT	qME	qL	qQ	qI
1.6455	0.5093	0.4359	0.4715	0.4008	1.6163	1.4712	1.7990

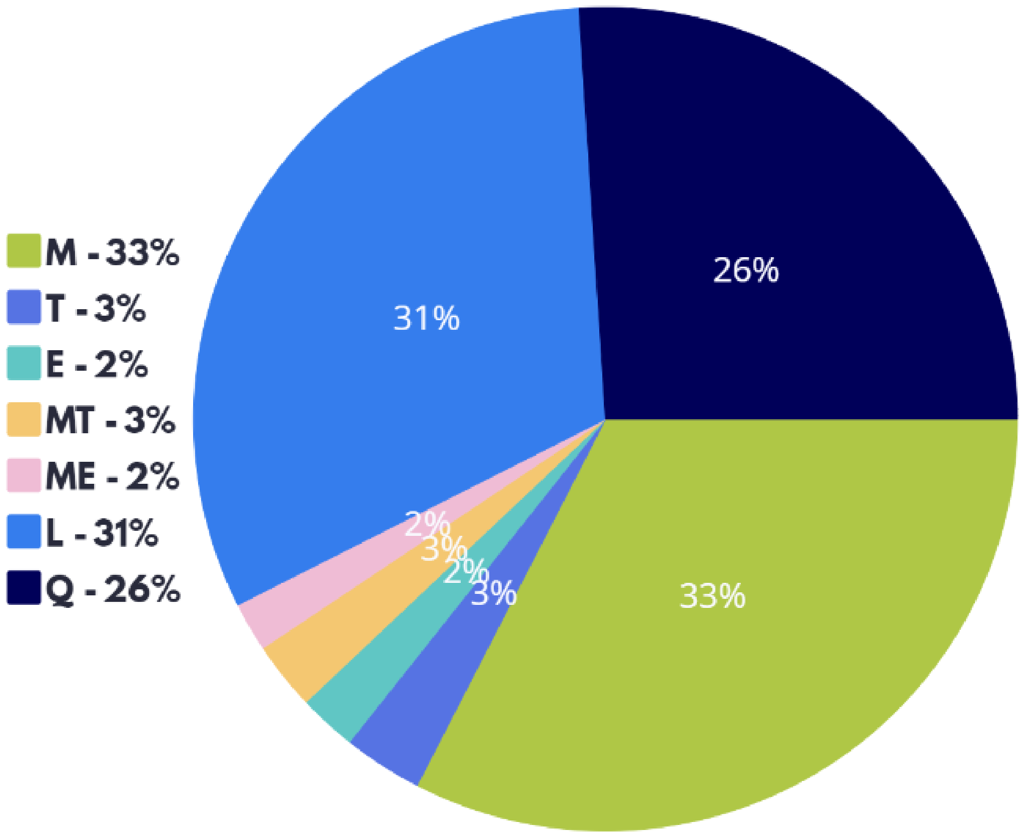
Tabela das porções de variação:
Primeiramente, devemos calcular o SSTotal = 2^3(2.7076 + 0.2593 + 0.1900 + 0.2223 + 0.1606 + 2.6124 + 2.1644) = 66.5328.

A seguir, desdobramos as contribuições individuais:

- SSM = 8 * 2.7076 = 21.6608
- SST = 8 * 0.2593 = 2.0744
- SSE = 8 * 0.1900 = 1.5200
- SSMT = 8 * 0.2223 = 1.7784
- SSME = 8 * 0.1606 = 1.2848
- SSL = 8 * 2.6124 = 20.8992
- SSQ = 8 * 2.1644 = 17.3152

M	T	E	MT	ME	L	Q
32.55 %	3.11%	2.28 %	2.67 %	1.93%	31.41 %	26.02 %

O gráfico a seguir demonstra melhor as porções de variação que cada fator ou interação possui:



Com base nesses dados, é evidente que os fatores Método, Linguagem e Uniformidade dos valores exercem as maiores influências no tempo de execução do algoritmo. Uma descoberta intrigante foi constatar que o método de ordenação possui um impacto quase equivalente ao da linguagem de programação. Isso destaca a importância de fazer escolhas criteriosas em relação à tecnologia ao desenvolver projetos sensíveis ao tempo de execução.

A partir do primeiro estudo com o projeto fatorial 2^k, foi possível identificar quais interações têm um impacto menor no tempo. Portanto, essas interações foram substituídas por dois outros fatores neste estudo de 2^(k - p).

Fatorial (2^4)r:
Neste projeto experimental, replicamos o primeiro experimento em um total de r vezes. Definimos r como 5 e, a partir disso, calculamos os erros e sua porcentagem. Além disso, determinamos os intervalos de confiança para os efeitos identificados.

Tabela de observações para cada experimento Ti e média entre eles:
Esta tabela apresenta os tempos das 5 replicações, medidos a partir de 16 experimentos distintos, juntamente com a média calculada entre eles.

Tabela de sinais:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	I	T1(s)	T2(s)	T3(s)	T4(s)	T5(s)	Média
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.6313	3.544	3.5314	3.4849	3.4832	3.535
1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	1	3.1919	3.3336	3.4072	3.1958	3.3496	3.2956
1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	1	7.5399	7.3467	7.2967	7.2196	7.2744	7.3355
1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	6.8441	7.7487	7.095	6.9755	7.0578	7.1442
1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	4.2333	4.1066	4.1288	4.2825	4.9789	4.346
1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	1	5.5684	4.9133	5.2438	4.8893	4.8898	5.1009
1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	10.0664	10.5349	9.1355	9.4739	9.9887	9.8399
1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	1	10.1957	11.2348	9.8469	10.8137	11.0244	10.6231
-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	1	0.1926	0.1834	0.1772	0.1848	0.18	0.1836
-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	1	0.1863	0.1918	0.1861	0.1991	0.2063	0.1939
-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	0.3618	0.3571	0.3627	0.3624	0.3674	0.3623
-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1	1	0.3973	0.326	0.3378	0.3219	0.4042	0.3574
-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	1	0.1728	0.1904	0.1887	0.1711	0.1996	0.1845
-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	0.2308	0.216	0.2368	0.2134	0.2118	0.2218
-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	1	0.3101	0.3041	0.3119	0.3029	0.3103	0.3078
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.4056	0.4299	0.4876	0.5339	0.5049	0.4724

A matriz a seguir é dada pela seguinte expressão aplicada em cada fator ou interação de fatores:

$$\sum_{i=1}^{2^k} S_{ij}\bar{y}_i$$

onde Sij é o sinal na tabela e $\bar{y}_i$ é o valor **Média** do experimento correspondente,ou seja, a média do tempo gasto no experimento i com as 5 repetições.

Somas:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	I
48.9365	-8.6889	-19.3813	-1.3147	-8.5103	-17.9491	-0.9001	3.3987	2.1653	0.1885	3.3353	1.7723	-0.0357	-0.1227	0.1623	53.5039

Matriz dos efeitos (Q's)

A matriz a seguir é dada pela seguinte expressão aplicada em cada fator ou interação de fatores:

$$\frac{1}{2^k}\sum_{i=1}^{2^k} S_{ij}\bar{y}_i$$

onde Sij é o sinal na tabela e $\bar{y}_i$ é o valor **Média** do experimento correspondente,ou seja, a média do tempo gasto no experimento i com as 5 repetições.

Matriz dos Qs:

qM	qT	qE	qQ	qMT	qME	qMQ	qTE	qTQ	qEQ	qMTE	qMTQ	qMEQ	qTEQ	qMTEQ	qI
3.0585	-0.5431	-1.2113	-0.0822	-0.5319	-1.1218	-0.0563	0.2124	0.1353	0.0118	0.2085	0.1108	-0.0022	-0.0077	0.0101	3.344

Erros experimentais:

O cálculo do erro experimental é feito a partir da diferença entre o valor coletado e a média dos tempos em cada linha, de forma genérica, temos:

$$e_{ij} = y_{ij} - \bar{y}_i$$

No nosso experimento foi utilizado r = 5, portanto temos uma tabela de 5x16 para as observações:

Observações					Média	Erros		
$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	$y_{1,5}$	$\bar{y}_1$	$\bar{y}_1 - y_{1,1}$	...	$\bar{y}_1 - y_{1,5}$
$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	$y_{2,5}$	$\bar{y}_2$	$\bar{y}_2 - y_{2,1}$	...	$\bar{y}_2 - y_{2,5}$
...	...	...	...	...	...	...	...	...
$y_{16,1}$	$y_{16,2}$	$y_{16,3}$	$y_{16,4}$	$y_{16,5}$	$\bar{y}_{16}$	$\bar{y}_{16} - y_{16,1}$	...	$\bar{y}_{16} - y_{16,5}$

Abaixo, os dados obtidos:

Matriz dos erros experimentais:

T1(s)	T2(s)	T3(s)	T4(s)	T5(s)	Média	E(1)	E(2)	E(3)	E(4)	E(5)
3.6313	3.544	3.5314	3.4849	3.4832	3.535	0.0963	0.009	-0.0036	-0.0501	-0.0518
3.1919	3.3336	3.4072	3.1958	3.3496	3.2956	-0.1037	0.038	0.1116	-0.0998	0.054
7.5399	7.3467	7.2967	7.2196	7.2744	7.3355	0.2044	0.0112	-0.0388	-0.1159	-0.0611
6.8441	7.7487	7.095	6.9755	7.0578	7.1442	-0.3001	0.6045	-0.0492	-0.1687	-0.0864
4.2333	4.1066	4.1288	4.2825	4.9789	4.346	-0.1127	-0.2394	-0.2172	-0.0635	0.6329
5.5684	4.9133	5.2438	4.8893	4.8898	5.1009	0.4675	-0.1876	0.1429	-0.2116	-0.2111
10.0664	10.5349	9.1355	9.4739	9.9887	9.8399	0.2265	0.695	-0.7044	-0.366	0.1488
10.1957	11.2348	9.8469	10.8137	11.0244	10.6231	-0.4274	0.6117	-0.7762	0.1906	0.4013
0.1926	0.1834	0.1772	0.1848	0.18	0.1836	0.009	-0.0002	-0.0064	0.0012	-0.0036
0.1863	0.1918	0.1861	0.1991	0.2063	0.1939	-0.0076	-0.0021	-0.0078	0.0052	0.0124
0.3618	0.3571	0.3627	0.3624	0.3674	0.3623	-0.0005	-0.0052	0.0004	0.0001	0.0051
0.3973	0.326	0.3378	0.3219	0.4042	0.3574	0.0399	-0.0314	-0.0196	-0.0355	0.0468
0.1728	0.1904	0.1887	0.1711	0.1996	0.1845	-0.0117	0.0059	0.0042	-0.0134	0.0151
0.2308	0.216	0.2368	0.2134	0.2118	0.2218	0.009	-0.0058	0.015	-0.0084	-0.01
0.3101	0.3041	0.3119	0.3029	0.3103	0.3078	0.0023	-0.0037	0.0041	-0.0049	0.0025
0.4056	0.4299	0.4876	0.5339	0.5049	0.4724	-0.0668	-0.0425	0.0152	0.0615	0.0325

Porção de variação de cada fator e percentual de erro experimental:
Esses valores são derivados da divisão entre a soma dos quadrados dos efeitos (q's) e o SST. No caso dos erros, a computação é realizada dividindo a soma dos quadrados dos erros experimentais pelo SST. Por exemplo, no cálculo de M(%), temos:

$$M = \frac{2^4 \cdot qM^2}{SST}$$

Para o percentual de erro experimental fazemos: SSE/SST (%)
Sendo SSE o somatório dos valores da matriz de erros ao quadrado, ou seja:

$$\sum_{i=1}^{2^4} \sum_{j=1}^{r} e_{ij}^2$$

Matriz das porções de variação:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	Erro exp.
71.7321 %	2.2618 %	11.2512 %	0.0518 %	2.1695 %	9.65 %	0.0243 %	0.3459 %	0.1404 %	0.0011 %	0.3334 %	0.0941 %	0.0 %	0.0004 %	0.0008 %	1.9432 %

De acordo com os dados da tabela acima é possível perceber que de fato o estudo indica que o fator método é o mais impactante no quesito tempo de execução seguido do estado inicial do array (fator que indica quão desordenado o array está).

Intervalos de confiança para as médias dos efeitos encontrados:

Faremos os intervalos de confiança para os experimentos 1 e 16 (na ordem em que aparecem na tabela) para um número m = 5 de replicações.
Sabemos que k = 4, r = 5, m = 5 e:

$$Se = \sqrt{\frac{\text{SSE}}{2^k(r-1)}}$$

Pelos cálculos, sabemos que SSE = 4.05450902, assim:

$$Se = \sqrt{4.055/64}$$
$$Se = 0.251$$

Ainda:

$$Sym = Se * \sqrt{5/(r * 2^4) + 1/m}$$

Logo, para m = 5 e r = 5, chegamos em:

$$Sym = Se * \sqrt{5/(5 * 16) + 1/5}$$
$$Sym = Se * \sqrt{5/80 + 1/5}$$
$$Sym = Se * \sqrt{21/80}$$
$$Sym = Se * 0.51$$
$$Sym = 0.251 * 0.51 = 0.12801$$

Para o cálculo dos intervalos de confiança para as médias dos efeitos dos experimentos, usamos:

$$\bar{y} - t\left[1 - \frac{\alpha}{2}\right] \cdot Sym \leq Y \leq \bar{y} + t\left[1 - \frac{\alpha}{2}\right] \cdot Sym$$

Para o primeiro experimento, os valores dos fatores são
XM = 1, XT= 1, XE = 1 e XQ = 1

Logo, todas as interações também têm valores positivos, então temos:

Y1 = qM + qT + qE
+ qQ + qMT + qME + qMQ + qTE + qTQ
+qEQ +qMTE + qMTQ + qMEQ + qTEQ + qMTEQ + qI = **3.535**

Assim, para o primeiro experimento, o intervalo de confiança de 90% é dado por:

$$\bar{y} - t\left[1 - \frac{\alpha}{2}\right] \cdot Sym \leq Y \leq \bar{y} + t\left[1 - \frac{\alpha}{2}\right] \cdot Sym$$

Buscando **t(0.95, 64)** na tabela t de Student, encontramos t aproximadamente igual a **1.669,** logo:

$$3.535 - 0.12801 * 1.669 \leq Y1 \leq 3.535 + 0.12801 * 1.669$$

Portanto, para o experimento de número 1, o intervalo de confiança de 90% para a média dos efeitos é:

$$3.321 \leq Y1 \leq 3.748$$

Para o experimento de número 16:
XM = -1, XT=-1, XE=-1, XQ = -1

Assim,

Y16 = - qM - qT - qE - qQ
+ qMT + qME + qMQ + qTE + qTQ +qEQ
- qMTE - qMTQ - qMEQ - qTEQ + qMTEQ + qI = **0.4724**

Seguindo a mesma regra:

$$\bar{y} - t\left[1 - \frac{\alpha}{2}\right] \cdot Sym \leq Y \leq \bar{y} + t\left[1 - \frac{\alpha}{2}\right] \cdot Sym$$

Já sabemos que **t(0.95,64)** = 1.669, portanto:

$$0.4724 - 0.12801 * 1.669 \leq Y16 \leq 0.4724 + 0.12801 * 1.669$$

Assim, para o experimento de número 16, o intervalo de confiança de 90% para a média dos efeitos é dado por:

$$0.262 \leq Y16 \leq 0.686$$

Construindo as m replicações:

Fazendo as m=5 replicações para cada experimento, temos o seguinte:

Matriz com as 5 replicações dos experimentos 1 e 16:

M	T	E	Q	MT	ME	MQ	TE	TQ	EQ	MTE	MTQ	MEQ	TEQ	MTEQ	I	T1(s)	T2(s)	T3(s)	T4(s)	T5(s)	Média
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4.01	3.5548	3.477	3.4897	3.5181	3.6099
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.5311	3.4079	3.4238	3.4501	3.4338	3.4493
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.445	3.4461	3.4638	3.4198	3.4204	3.439
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.5437	3.4299	3.4629	3.4883	3.5039	3.4857
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3.4623	3.4156	3.4337	3.4153	3.5018	3.4457
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.3961	0.3827	0.3865	0.3955	0.389	0.39
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.4045	0.403	0.3981	0.3881	0.3938	0.3975
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.4009	0.3919	0.408	0.4084	0.3903	0.3999
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.3926	0.398	0.4124	0.4165	0.4017	0.4043
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	0.3985	0.4027	0.3974	0.3985	0.3895	0.3973

Note que as 5 primeiras linhas da tabela são referentes ao experimento 1 e as 5 últimas referentes ao experimento 16.

Note ainda que para as 5 replicações do experimento 1, **todas as médias ficaram no intervalo de confiança previsto** e o mesmo ocorre para todas as replicações do experimento 16.

Conclusão geral do estudo:

Este estudo contribuiu significativamente para a revisão e aprofundamento das concepções dos participantes sobre os fatores preponderantes na elaboração de programas. Além disso, promoveu o desenvolvimento de competências estatísticas e experimentais práticas entre os envolvidos.

A análise cuidadosa dos dados permitiu a desconstrução de suposições anteriores, proporcionando uma visão mais fundamentada sobre a influência dos diversos fatores no processo de construção de programas. Este processo de desconstrução e aprendizado não apenas esclareceu conceitos, mas também incentivou a aplicação prática de métodos estatísticos e experimentais, enriquecendo a compreensão dos participantes sobre a matéria. O resultado final foi uma sólida base de conhecimento e habilidades analíticas.

Divisão de Tarefas:

Construção do Código (feito no Google Meet, majoritariamente por Danilo, com compartilhamento de tela)

Elaboração do Relatório (Feito no Google Docs em conjunto, majoritariamente por Yuri)