



Relatório

Maximum Flow Problem - Hypothesis testing

1. Introdução

Este trabalho foi desenvolvido no âmbito da disciplina de Metodologias Experimentais em Informática, com o objetivo de estudar o desempenho e eficiência de certos algoritmos (Dinic, EK e MPM), que procuram resolver o *Maximum Flow Problem*. Este relatório encontra-se direcionado ao tema abordado na terceira meta, *Hypothesis testing*, sendo uma continuação dos relatórios anteriores, *Exploratory Data Analysis and Linear Regression* e *Pre-registration of Hypotheses*.

Tendo em conta os resultados já obtidos, constata-se que o algoritmo com melhor *performance* é o Dinic, com o MPM a ser ligeiramente pior. Quanto ao EK, este é o que apresenta piores valores de tempo, muitos deles atingindo o tempo limite de execução.

Neste relatório, pretende-se testar as diversas hipóteses já formuladas na Meta 2, acerca de diferentes aspetos evidenciados na Meta 1, de maneira a verificar a sua veracidade. Para testar as hipóteses, foram realizadas várias experiências, determinando se os pressupostos se verificam e usando testes paramétricos e não paramétricos quando aplicável.

2. Hipóteses

As hipóteses são criadas recorrendo ao par Hipótese Nula, H_0 , e Hipótese Alternativa, H_1 , que constituem formulações opostas.

Para verificar a sua validade, é necessário recorrer a dois valores:

- nível de significância, α - *threshold* usado para determinar se o resultado do teste é significativo ou não, constituindo a probabilidade de rejeitar a hipótese nula quando ela é verdadeira; para aplicação nos testes das hipóteses, este valor foi definido como 5%.

- *p-value* – probabilidade de se obter uma estatística de teste pelo menos tão extrema quanto a observada numa amostra, usada para determinar se a hipótese nula deve ser rejeitada ou não.

Deste modo, se o *p-value* for menor que o nível de significância, os resultados são considerados significantes e a hipótese nula é rejeitada. Pelo contrário, se o *p-value* for maior que o nível de significância, os resultados não são considerados significantes e a hipótese nula não é rejeitada.

2.1. Hipótese 1 - A capacidade máxima de cada arco não influencia os tempos de execução dos algoritmos

Esta hipótese foi levantada através da análise realizada na meta 1, onde ao visualizar dos gráficos podemos observar que a capacidade máxima de cada arco não influenciava o tempo de execução dos algoritmos. A hipótese é formulada então da seguinte forma:

H_0 : A capacidade máxima de cada arco não tem influência nos tempos de execução dos algoritmos.

H_1 : A capacidade máxima de cada arco tem influência nos tempos de execução dos algoritmos.

$$\begin{cases} H_0: \forall(i, j) : \mu_i = \mu_j, i, j \in \{50, 100, 150, 200\} \\ H_1: \exists(i, j) : \mu_i \neq \mu_j, i, j \in \{50, 100, 150, 200\} \end{cases}$$

Considerando um grau de confiança de 95% e antes de analisar o resultado do teste 2-Way ANOVA, é necessário verificar alguns pressupostos deste teste estatístico tais como: Normalidade dos resíduos e Homogeneidade dos resíduos.

Normalidade dos resíduos

De forma a verificar este pressuposto foram utilizados dois métodos diferentes. O primeiro foi através da análise visual do qq-plot, Figura 1.

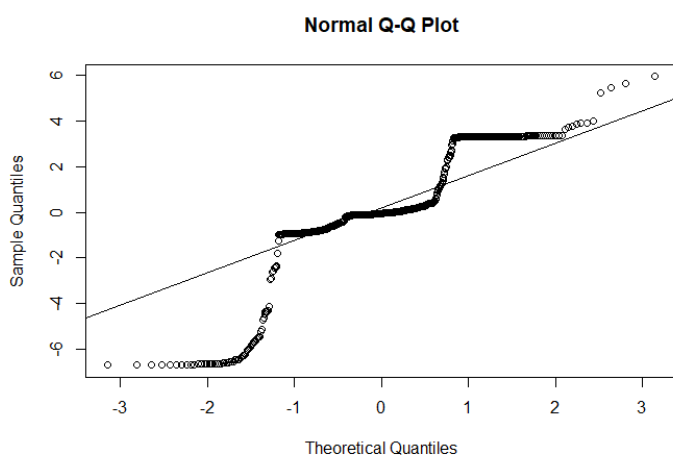


Figura 1 - Normal Q-Q plot

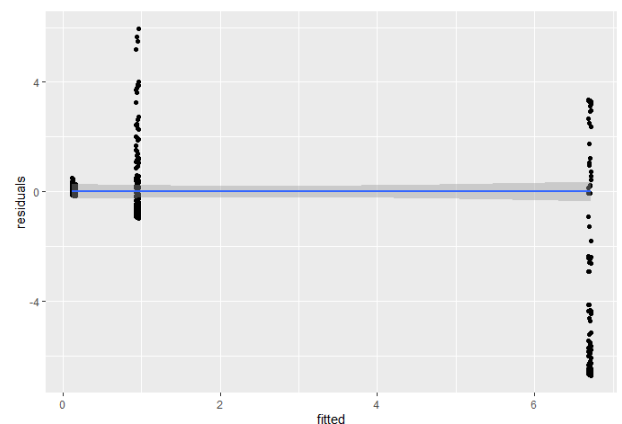


Figura 2 - Gráfico de Resíduos

Através da Figura 1, podemos constatar que os resíduos não seguem uma distribuição normal. Para confirmar esta análise recorreremos ao teste de Shapiro-Wilk.

O resultado deste teste foi então:

$$W = 0.64457, p\text{-value} < 2.2e-16$$

O p-value é menor que o nível de significância considerado $\alpha = 0.05$ o que confirma que os resíduos não seguem uma distribuição normal.

Homogeneidade dos resíduos

A homogeneidade dos resíduos pode ser verificada através da análise do gráfico de resíduos, Figura 2.

Ao analisar a Figura 2 podemos afirmar que a distribuição dos resíduos não é a mesma ao longo do eixo dos x, podendo então concluir-se que não existe homogeneidade dos resíduos.

Uma vez que dois dos pressupostos para realizar o teste 2-Way ANOVA falharam, então foi necessário recorrer a um teste não paramétrico de forma a ser possível avaliar a hipótese considerada. O teste utilizado foi então o teste de Kruskal-Wallis.

Os resultados obtidos foram:

$$\text{chi-squared} = 0.027188, df = 3, p\text{-value} = 0.9988$$

Para o nível de significância considerado ($\alpha=0.05$) e analisando o p-value podemos concluir que não existe diferenças nos tempos perante a capacidade máxima de cada arco, pois $p\text{-value} > \alpha$. Então não se rejeita a hipótese nula (H_0).

2.2. Hipótese 2 - O algoritmo EK é consideravelmente mais lento que os outros (Dinik e MPM) para quaisquer que sejam os valores das variáveis independentes

Esta hipótese serve para mostrar que, independentemente dos valores atribuídos às variáveis independentes, o tempo de execução do algoritmo *EK* é sempre mais lento do que o dos algoritmos *Dinik* e *MPM*, sendo que é o único que atinge o tempo máximo execução dentro dos testes realizados. Para provar a veracidade desta hipótese, tentou-se mostrar que a hipótese nula formulada é rejeitada. A hipótese é então apresentada desta forma:

H_0 : O algoritmo EK não é consideravelmente mais lento que os outros (Dinik e MPM) para quaisquer que sejam os valores das variáveis independentes.

H_1 : O algoritmo EK é consideravelmente mais lento que os outros (Dinik e MPM) para quaisquer que sejam os valores das variáveis independentes.

$$\begin{cases} H_0 : \mu_{EK} - \mu_{Dinik} = 0 \wedge \mu_{EK} - \mu_{MPM} = 0 \\ H_1 : \mu_{EK} - \mu_{Dinik} \neq 0 \vee \mu_{EK} - \mu_{MPM} \neq 0 \end{cases}$$

Considerando três grupos de amostras (Dinik, MPM, EK) utilizou-se o *t-test*. Uma vez que os inputs para os três algoritmos são os mesmos, constatou-se que os dados eram dependentes e pareados.

Antes de proceder ao *t-test pareado*, tivemos de verificar alguns pressupostos. De forma a realizá-lo, para cada par de algoritmos (EK-Dinik, EK-MPM), calculamos a diferença das amostras.

Normalidade dos resíduos

Como foi referido anteriormente é necessário verificar alguns pressupostos em antes de continuar para o t-test. Um desses pressupostos é a normalidade dos resíduos que, observando as Figuras 3 e 4, podemos especular que não seguem uma distribuição normal. Para confirmar foi então realizado o teste Shapiro Wilk.

- Para o par EK-Dinic:

$$W = 0.7211, p\text{-value} < 2.2e-16$$

- Para o par EK-MPM:

$$W = 0.83333, p\text{-value} < 7.144e-14$$

Analisando os resultados obtidos para os pares de algoritmos, podemos afirmar que os resíduos não seguem uma distribuição normal, pois, tendo em conta o p-value com o nível de significância, vemos que $p\text{-value} > \alpha$.

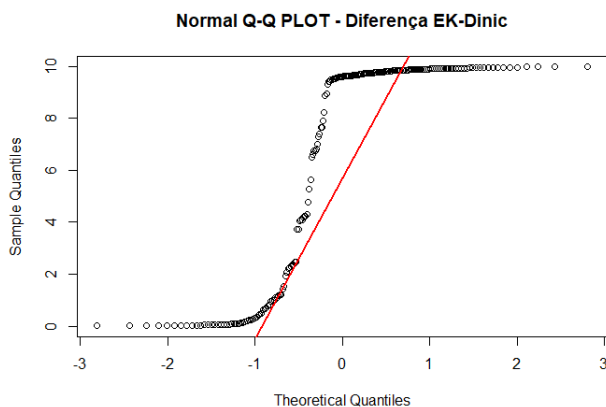


Figura 3 - Normal Q-Q plot

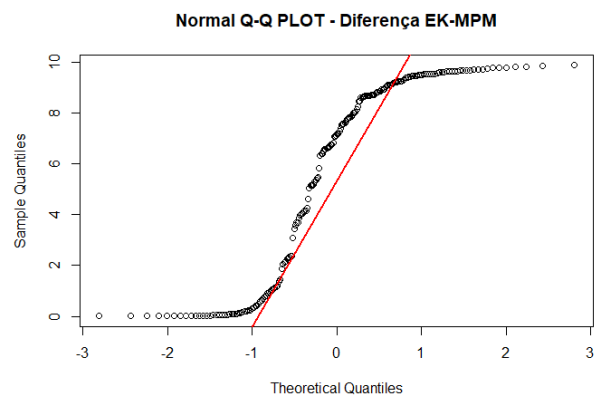


Figura 4 - Normal Q-Q plot

Outro pressuposto que fomos avaliar foi a homogeneidade dos resíduos.

Homogeneidade dos resíduos

Observando a Figura 5, podemos confirmar que os resíduos não seguem o pressuposto de homogeneidade, uma vez que a sua distribuição ao longo do eixo dos x é diferente e não constante.

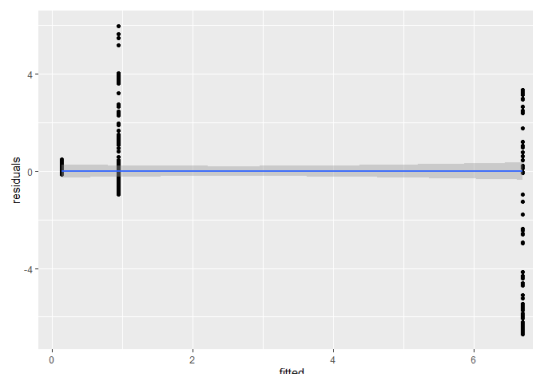


Figura 5 - Gráfico Resíduos

Tendo em conta que os pressupostos falharam, foi necessário recorrer a um teste não paramétrico, tendo-se utilizado o teste Wilcoxon.

Os resultados obtidos foram:

- Para o par EK-Dinic:

$$V = 20100, p\text{-value} < 2.2e-16$$

- Para o par EK-MPM:

$$V = 20100, p\text{-value} < 2.2e-16$$

Considerando o valor de α (0.05), concluímos que a performance do algoritmo EK é significativamente diferente da performance dos algoritmos Dinic e MPM, sendo que é mais lenta. Visto que $p\text{-value} < \alpha$, rejeitamos assim a hipótese nula (H_0), confirmando a nossa análise.

2.3. Hipótese 3 - A probabilidade de gerar arco não influencia os tempos de execução do algoritmo Dinic

Esta hipótese tenta provar que não existe relação entre a probabilidade de formar arco e o tempo de execução do algoritmo Dinic. A hipótese é formulada da seguinte forma:

H_0 : A probabilidade de gerar arco não influencia os tempos de execução do algoritmo Dinic.

H_1 : A probabilidade de gerar arco influencia os tempos de execução do algoritmo Dinic

$$\begin{cases} H_0 : \forall(i, j) : \mu_i = \mu_j, i, j \in \{0.2, 0.4, 0.6, 0.8, 1\} \\ H_1 : \exists(i, j) : \mu_i \neq \mu_j, i, j \in \{0.2, 0.4, 0.6, 0.8, 1\} \end{cases}$$

Como na hipótese anterior, foi considerado um grau de confiança de 95%. Antes de analisar o resultado do teste ANOVA, é necessário verificar alguns pressupostos deste teste estatístico tais como: Normalidade dos resíduos e Homogeneidade dos resíduos.

Normalidade dos resíduos

Como já foi referido anteriormente, de modo a verificar este pressuposto foram utilizados dois métodos.

O primeiro foi através da análise do Normal Q-Q Plot, Figura 3. Analisando a Figura 3 podemos então constatar que os resíduos não aparentam seguir uma distribuição normal.

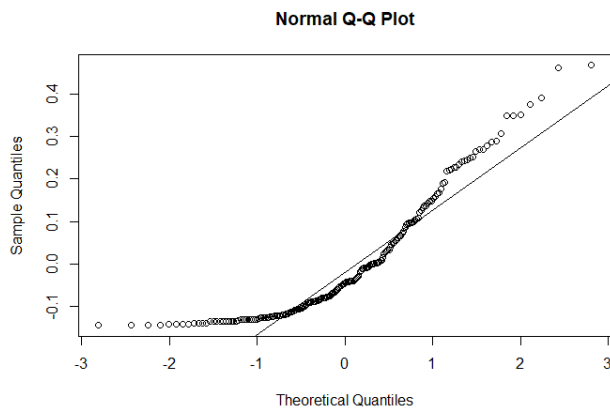


Figura 6 - Normal Q-Q plot

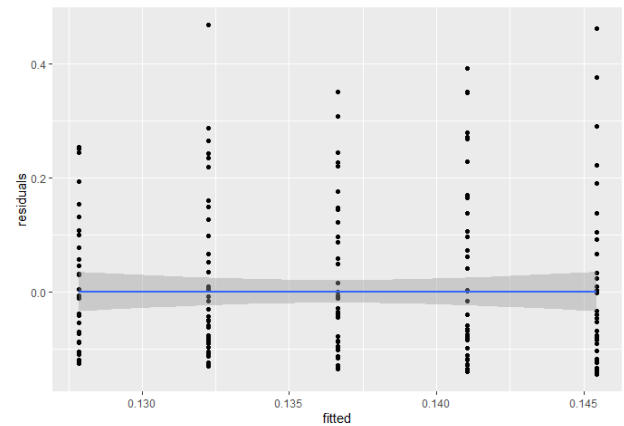


Figura 7 - Gráfico de resíduos

De modo a confirmar esta observação, foi efetuado o teste de Shapiro-Wilk, obtendo assim o seguinte resultado:

$$W = 0.85374, p\text{-value} < 6.556e-13$$

O p-value é menor que o nível de significância considerado, $\alpha = 0.05$, o que confirma que os resíduos não seguem uma distribuição normal.

Homogeneidade dos resíduos

Observando a Figura 4 podemos constatar que a distribuição dos resíduos é aproximadamente a mesma ao longo do eixo dos x, mas de forma a confirmar esta análise foi realizado o teste Bartlett.

Obtemos os seguintes resultados:

$$K\text{-squared} = 5.0067, df=4, p\text{-value} = 0.2866$$

Analisando os resultados podemos então concluir que os resíduos respeitam a homogeneidade, pois $p\text{-value} > \alpha$, sendo $\alpha = 0.05$.

Como na hipótese anterior, um dos pressupostos do teste ANOVA foi violado sendo, deste modo, necessário recorrer a um teste não paramétrico.

O teste não paramétrico escolhido foi o Kruskal-Wallis, obtendo os seguintes resultados:

$$\chi^2 = 2.2627, df= 4, p\text{-value} = 0.6876$$

Perante estes resultados, podemos afirmar que a probabilidade de formar arco não influencia o tempo de execução, pois $p\text{-value} > \alpha$, não se rejeitando H_0 .

2.4. Hipótese 4 - A probabilidade de gerar arco influencia o número de limites de tempo excedido no algoritmo EK

Esta hipótese pretende provar que a probabilidade de gerar arco influencia o número de TLE's no algoritmo EK. Posto isto, a hipótese foi formulada da seguinte forma:

H_0 : A probabilidade de gerar arco não influencia o número de limites de tempo excedido no algoritmo EK

H_1 : A probabilidade de gerar arco influencia o número de limites de tempo excedido no algoritmo EK

$$\begin{cases} H_0 : \forall(i, j) : \#TLE_i = \#TLE_j, i, j \in \{0.2, 0.4, 0.6, 0.8, 1\} \\ H_1 : \exists(i, j) : \#TLE_i \neq \#TLE_j, i, j \in \{0.2, 0.4, 0.6, 0.8, 1\} \end{cases}$$

Para provar esta hipótese, foi inicialmente utilizado o teste ANOVA com um intervalo de confiança de 95%. Como é esperado, a validade dos pressupostos foi verificada.

Normalidade dos resíduos

Analisando a Figura 8, não conseguimos retirar uma conclusão acertada sobre a distribuição dos resíduos. Para fazer uma análise mais concreta, foi realizado um teste de Shapiro-Wilk.

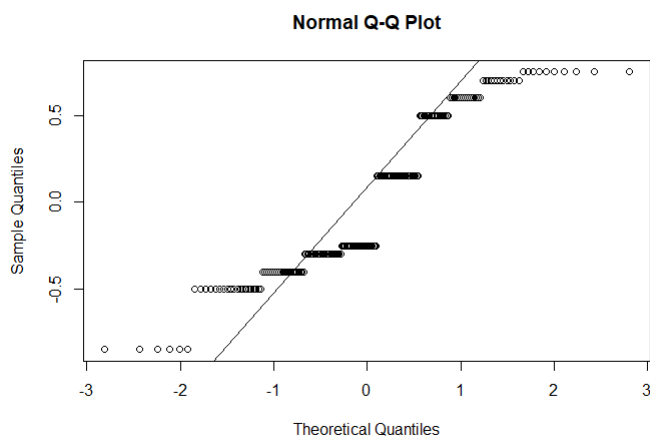


Figura 8 - Normal Q-Q plot

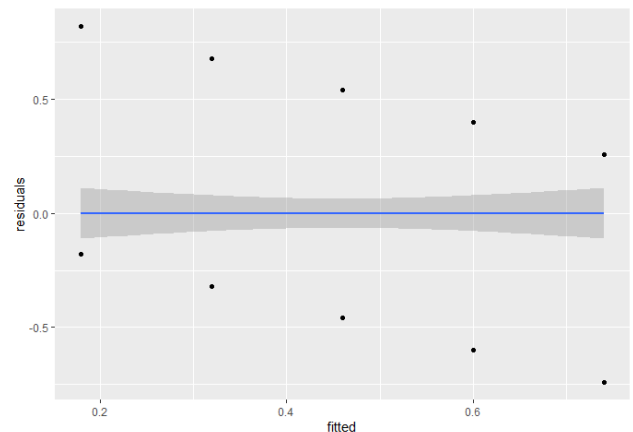


Figura 9 - Gráfico de Resíduos

Os resultados obtidos nesse teste foram:

$$W = 0.6341, p\text{-value} < 2.2e-16$$

Considerado o nível de significância $\alpha = 0.05$ e $p\text{-value} < 2.2e-16$, podemos então confirmar a análise de que os resíduos seguem uma distribuição normal, pois $p\text{-value} < \alpha$.

Homogeneidade dos resíduos

Observando a Figura 9, podemos constatar que os resíduos seguem o pressuposto da homogeneidade. Para ter a certeza desta análise foi realizado o teste de Bartlett, onde se obteve os seguintes resultados:

$$K\text{-squared} = 5.2087, df=4, p\text{-value} = 0.2665$$

Com os resultados do teste de Bartlett foi possível confirmar a análise feita anteriormente à Figura 9, pois o $p\text{-value} > \alpha$.

Como os dados se encontram distribuídos de forma homogênea em torno da média zero, é possível afirmar que o pressuposto da esfericidade é cumprido.

Uma vez que os três pressupostos se verificam é então possível avançar com o teste ANOVA.

```
      Df Sum Sq Mean Sq F value    Pr(>F)
factor(prob)  4    9.08   2.2700    10.9 5.35e-08 ***
Residuals    195   40.60   0.2082
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 10 - Resultados teste ANOVA

Analisando os resultados obtidos, podemos então concluir que a probabilidade de gerar arco influencia o número de TLE's no algoritmo EK, pois $p\text{-value} < \alpha$, refutando a hipótese H_0 .

7. Conclusões

Após realizada a testagem das hipóteses, pode-se então concluir que a capacidade máxima de cada arco não influencia o tempo de execução dos algoritmos. Foi também ainda possível constatar que de facto a probabilidade de gerar arco não influencia o tempo de execução do algoritmo Dinic. A hipótese específica que apontava para que o algoritmo EK fosse consideravelmente mais lento que os outros dois (Dinic, MPM) estava correta.

Observando os resultados, apesar de considerada uma amostra relativamente elevada de resultados por algoritmo, foi possível notar alguma aleatoriedade na distribuição dos dados. Para contrariar esta aleatoriedade, poderíamos aumentar o número de amostras por teste e aumentar o tempo limite por execução dos algoritmos. Finalmente, seria interessante testar as hipóteses para uma maior gama de probabilidade e número de vértices, para que as conclusões sejam de um maior nível de confiança.

8. Referências

- [1] BEVANS, Rebecca. ANOVA in R | A Complete Step-by-Step Guide with Examples. [S. l.], 17 nov. 2022. Disponível em: <https://www.scribbr.com/statistics/anova-in-r/>. Acesso em: 22 dez. 2022.
- [2] , Zach. How to Test for Normality in R (4 Methods). [S. l.], 29 set. 2021. Disponível em: <https://www.statology.org/test-for-normality-in-r/>. Acesso em: 22 dez. 2022.
- [3] R DOCUMENTATION. Wilcox.test: Wilcoxon Rank Sum and Signed Rank Tests. [S. l.], 2022. Disponível em: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/wilcox.test>. Acesso em: 22 dez. 2022.
- [4] R DOCUMENTATION. Kruskal.test: Kruskal-Wallis Rank Sum Test. [S. l.], 2022. Disponível em: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kruskal.test>. Acesso em: 22 dez. 2022.