

# **Metodologias Experimentais em Informática**

## **Meta 3 - Teste das hipóteses**

22 de dezembro de 2021

### **1 - Introdução**

No âmbito da unidade curricular de Metodologias Experimentais em Informática, foi-nos proposto desenhar uma experiência com o intuito de avaliar o desempenho de um programa que gere o escalonamento de exames no Departamento de Engenharia Informática. O objetivo principal é a exploração do método experimental apresentado, seguindo os passos sugeridos, testando, observando e realizando uma análise dos resultados, de forma a retirarmos conclusões sobre este programa.

O programa em causa é-nos fornecido pelo docente em duas versões que resolvem o mesmo problema, apresentando implementações diferentes, acompanhadas por um gerador de dados de entrada que permite obter os dados simulados necessários para executar os programas. Na primeira meta deste projeto começámos por realizar uma análise exploratória dos dados (Exploratory Data Analysis), desenhando experiências de forma a testar os dois programas e gerar dados cuja análise se encontra no documento gerado durante a realização da primeira meta. Na segunda etapa deste projeto avançámos com o pré-registo de duas hipóteses que formulamos com base nos dados observados na primeira fase do trabalho. Estas hipóteses serão agora testadas nesta terceira e última meta de forma a tirarmos conclusões sobre os programas inicialmente dados.

O presente documento deverá descrever o nosso processo com detalhe suficiente para permitir a reprodução do mesmo, como solicitado no enunciado.

De notar que é imprescindível a leitura dos outros dois documentos gerados durante a elaboração deste trabalho prático, pois são mencionadas particularidades que podem fazer referência a temas discutidos anteriormente.

### **2 - Recolha dos dados**

À semelhança da Meta 1 deste trabalho prático, os dados foram gerados recorrendo à mesma máquina cujas especificações se encontram no documento correspondente a essa meta. Todo o software utilizado foi o mesmo

também de modo a não introduzir variáveis extra no nosso processo de obtenção de resultados. Ao contrário da primeira meta onde era importante analisarmos o máximo de dados à nossa disposição, neste caso apenas precisamos de realizar os testes de forma a obter os dados que nos eram relevantes, ou seja, para cada hipótese apenas nos interessava obter os resultados e valores relevantes a essa hipótese.

No caso da primeira, apenas necessitamos de registar o valor da probabilidade em cada patamar (1%, 20%, 40% e 60%), o tempo de execução total do programa e qual código estava a ser utilizado (1 ou 2). Para a segunda hipótese registamos o valor do número de exames em cada patamar (2, 20 e 40), o tempo de execução total do programa e qual código estava a ser utilizado.

## 3 - Hipóteses

Como mencionado anteriormente, após a realização da análise exploratória dos dados, registámos duas hipóteses que iremos testar e sobre as quais pretendemos obter uma conclusão correta do ponto de vista da estatística.

De seguida, as hipóteses:

### 3.1 - Hipótese 1

O tempo de execução do *code1.c* é pelo menos 5% superior ao do *code2.c* para valores de **Probabilidade (p)** de 1% até 60% e com um valor fixo de **Número de Exames** igual a 20.

### 3.2 - Hipótese 2

A diferença entre o tempo de execução do *code1.c* e *code2.c* nunca é superior a 5%, fixando a **Probabilidade (p)** em 20% e com um de **Número de Exames** a variar entre 2 e 40.

## 4 - Teste das Hipóteses

Em ambas as hipóteses pretendemos testar o **tempo de execução** dos programas, mais especificamente como é que se comparam estes tempos entre os dois programas (*code1.c* e *code2.c*), e também a interação com a variação de **uma das variáveis independentes** dos programas. Na Hipótese 1 fixamos o valor do **Número de Exames** no valor 20 e testamos, ambos os códigos, com

valores de **Probabilidade (p)** de 1% até 60%. Na Hipótese 2 fixamos o valor de **Probabilidade (p)** em 20% e variamos o **Número de Exames** entre 2 e 40. Como estamos a testar a influência de 2 variáveis (código e respetiva variável independente) no tempo de execução decidimos recorrer ao Two -Way ANOVA. Este teste estatístico permite-nos tirar conclusões sobre as médias do tempo de execução de cada programa e como a interação com as variáveis independentes correspondentes a cada hipótese afeta o tempo de execução de cada código. A hipótese nula do Two -Way ANOVA afirma que as médias da variável que estamos a observar (tempo de execução do programa) são iguais, deste modo, se no fim do teste **não rejeitarmos** a hipótese nula, para um determinado nível de significância, podemos concluir que o tempo de execução de cada código é, em média, igual e, portanto podemos aceitar ou rejeitar as hipóteses apresentadas no **Ponto 2**.

Também tínhamos ao nosso dispor a possibilidade de utilizar os T-Tests ou One-Way ANOVA mas após alguma discussão entre nós e até algumas conversas com o docente decidimos que, para o nosso caso, tendo em conta as variáveis que íamos alterando, faria mais sentido e seria mais fácil aplicar o Two -Way ANOVA de forma a chegar às conclusões.

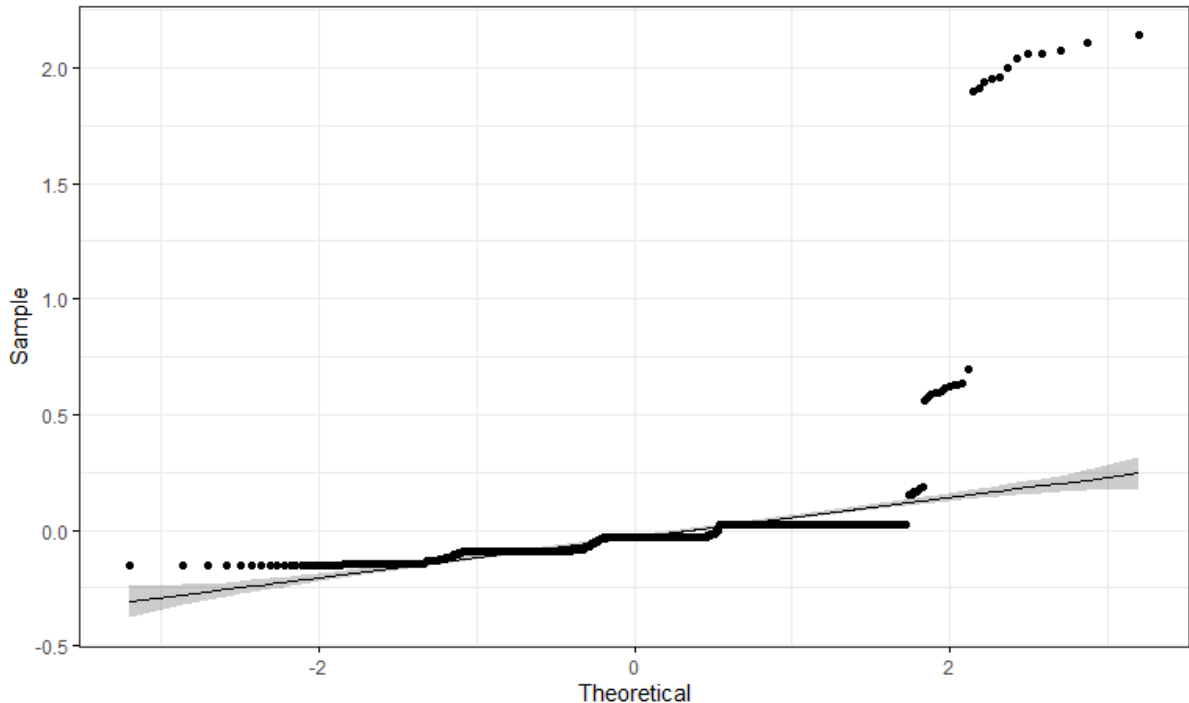
Realizamos os testes necessários recorrendo a *scripts* semelhantes aos utilizados na primeira meta e seguindo os parâmetros estabelecidos pelas nossas hipóteses, o que gerou ficheiros com a informação sobre as diferentes variáveis independentes (dependendo da hipótese) e a variável dependente que pretendemos observar (tempo de execução do programa), devidamente organizadas.

## 4.1 - Hipótese 1

### 4.1.1 - Verificação das suposições

Após a obtenção dos dados necessários para testar a **Hipótese 1** passámos à verificação das suposições (*assumptions*) deste tipo de teste estatístico. Primeiramente podemos afirmar que a recolha dos dados é feita corretamente e que todas as experiências são independentes umas das outras. O tipo de dados é totalmente numérico, logo essa suposição também se verifica. No caso da verificação se os dados obtidos seguem uma distribuição normal pensámos no Teorema do Limite Central onde podemos afirmar que as  $n$  medições utilizadas para calcular as médias dos tempos de execução são, de facto, independentes e sendo que realizámos 90 testes para cada uma das combinações (respeitando  $n \geq 30$ ), podemos dizer que os dados seguem uma

distribuição normal com média  $\mu$  e desvio padrão  $\sigma/\sqrt{n}$ . Também podemos observar no gráfico da **Figura 1** que a esmagadora maioria dos dados obtidos seguem a distribuição normal.



**Figura 1.** Gráfico Q-Q Plot dos dados resultantes dos testes da Hipótese 1

Deste modo, considerámos que, para esta experiência, estes dados estariam válidos para serem utilizados.

#### 4.1.2 - Two -Way ANOVA

Ao aplicar o teste Two -Way ANOVA obtivemos a seguinte tabela da **Figura 1** no programa R.

```
> hip1 = read.table("hip1.txt", header=TRUE)
> aov.out = aov(time_in_seconds~factor(code)*factor(probability_pairs), data=hip1)
> summary(aov.out)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(code)	1	0.00	0.0003	0.004	0.952
factor(probability_pairs)	3	3.62	1.2052	15.156	1.43e-09 ***
factor(code):factor(probability_pairs)	3	0.00	0.0001	0.001	1.000
Residuals	712	56.62	0.0795		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

**Figura 2.** Resultado Two -Way ANOVA -> Hipótese 1

A partir da tabela obtida podemos verificar que no fator “code” (que corresponde aos códigos *code1.c* e *code2.c*) o valor de  $Pr(>F)$  obtido é de 0.952 que é superior a nível de significância de 95% que considerámos ( $\alpha = 0.05 < 0.952$ ). Este resultado leva-nos a **aceitar** a hipótese nula deste teste de Two-Way ANOVA para o fator “code”, ou seja, para um  $\alpha = 0.05$  podemos dizer que **não existem diferenças entre as médias de tempo de execução de cada código**. O resultado que vemos na segunda linha da tabela leva-nos a rejeitar claramente a hipótese nula ( $\alpha = 0.05 > 1.43 \times 10^{-9}$ ) visto que a alteração do fator “*probability\_pairs*” (Probabilidade (p)) tem uma influência clara no tempo final de execução do programa. Na linha correspondente à interação entre estes dois fatores podemos notar que não rejeitamos a hipótese nula, ou seja, o efeito do valor da variável “*probability\_pairs*” no tempo de execução do programa não depende do fator “code”.

#### 4.1.3 - Conclusões sobre a Hipótese 1

A partir dos resultados obtidos em [3.1.2](#) podemos, após alguma interpretação de resultados, tirar conclusões sobre a **Hipótese 1** originalmente definida no segundo documento deste trabalho (Meta 2) e apontada novamente em [2.1](#). Esta hipótese afirma que existe uma diferença significativa no tempo de execução dos dois programas (*code1.c* e *code2.c*) sendo que o *code1.c* apresenta um tempo de execução superior, por pelo menos 5%, em relação ao *code2.c*, isto para os valores e variações das variáveis independentes registadas. A partir do Two-Way ANOVA podemos concluir que, para o nível de significância definido, podemos considerar que as médias dos tempos de execução dos dois programas são iguais, mesmo com a variação do valor da Probabilidade (p). Ao concluirmos que as médias são iguais podemos também concluir que **rejeitamos a Hipótese 1** pois, se são iguais, isso implica que não existe uma diferença superior a 5% entre os dois códigos.

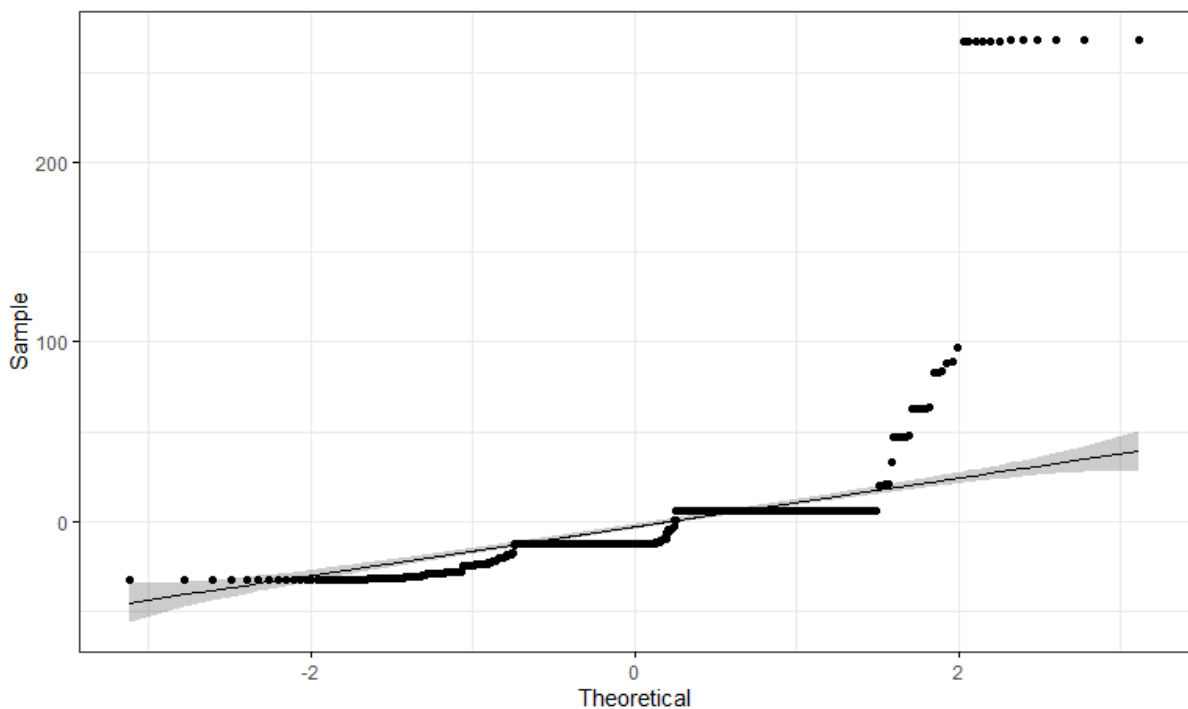
### 4.2 - Testes da Hipótese 2

#### 4.2.1 - Verificação das suposições

Após a obtenção dos dados necessários para testar a **Hipótese 2** passámos à verificação das suposições (*assumptions*) deste tipo de teste.

Novamente, podemos afirmar que a recolha dos dados é feita corretamente e que todas as experiências são independentes umas das outras. O tipo de dados é totalmente numérico, logo essa suposição também se verifica, tal como na

primeira hipótese. No caso da verificação se os dados obtidos seguem uma distribuição normal pensamos no Teorema do Limite Central onde podemos afirmar que as  $n$  medições utilizadas para calcular as médias dos tempos de execução são, de facto, independentes e sendo que realizámos 90 testes para cada uma das combinações (respeitando  $n \geq 30$ ), podemos dizer que os dados seguem uma distribuição normal com média  $\mu$  e desvio padrão  $\sigma/\sqrt{n}$ . Também podemos observar no gráfico da **Figura 3** que a esmagadora maioria dos dados obtidos seguem a distribuição normal.



**Figura 3.** Gráfico Q-Q Plot dos dados resultantes dos testes da Hipótese 2

Deste modo, considerámos que, para esta experiência, estes dados estariam válidos para serem utilizados.

#### 4.2.2 - Two -Way ANOVA

Ao aplicar o teste Two -Way ANOVA obtivemos a seguinte tabela da **Figura 1** no programa R.

```

> hip2 = read.table("hip2.txt", header=TRUE)
> aov.out = aov(time_in_seconds~factor(code)*factor(n_exams), data=hip2)
> summary(aov.out)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(code)	1	2	2	0.001	0.977
factor(n_exams)	2	173246	86623	44.623	<2e-16 ***
factor(code):factor(n_exams)	2	3	2	0.001	0.999
Residuals	534	1036608	1941		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

```

**Figura 4.** Resultado Two -Way ANOVA -> Hipótese 2

A partir da tabela obtida podemos verificar que no fator “code” (que corresponde aos códigos *code1.c* e *code2.c*) o valor de  $Pr(>F)$  obtido é de 0.977 que é superior a nível de significância de 95% que considerámos ( $\alpha = 0.05 < 0.977$ ). Este resultado leva-nos a **aceitar** a hipótese nula deste teste de Two -Way ANOVA para o fator “code”, ou seja, para um  $\alpha = 0.05$  podemos dizer que **não existem diferenças entre as médias de tempo de execução de cada código**. O resultado que vemos na segunda linha da tabela leva-nos a rejeitar claramente a hipótese nula ( $\alpha = 0.05 > 2 \times 10^{-16}$ ) visto que a alteração do fator “n\_exams” (Número de Exames) tem uma influência clara no tempo final de execução do programa. Na linha correspondente à interação entre estes dois fatores podemos notar que não rejeitamos a hipótese nula, ou seja, o efeito do valor da variável “n\_exams” no tempo de execução do programa não depende do fator “code”.

### 4.2.3 - Conclusões sobre a Hipótese 2

A partir dos resultados obtidos em [3.2.2](#) podemos, após alguma interpretação de resultados, tirar conclusões sobre a **Hipótese 2** originalmente definida no segundo documento deste trabalho (Meta 2) e apontada novamente em [2.2](#). Esta hipótese afirma que a diferença no tempo de execução dos dois programas (*code1.c* e *code2.c*) nunca é superior a 5% para os valores e variações das variáveis independentes registadas. A partir do Two -Way ANOVA podemos concluir que, para o nível de significância definido, podemos considerar que as médias dos tempos de execução dos dois programas são iguais, mesmo com a variação do valor do Número de Exames. Ao concluirmos que as médias são iguais podemos também concluir que **aceitamos a Hipótese 2** pois, se são iguais, isso implica que a diferença entre o tempo de execução dos dois códigos nunca é superior a 5%.

## 5 - Trabalho Futuro

Futuramente consideramos que seria importante formular outro tipo de hipóteses mais complexas e que, através da aplicação de testes estatísticos, levassem a conclusões diferentes e úteis acerca da realidade deste tipo de programas de modo a chegar a decisões concretas acerca da utilização dos mesmos.

## 6 - Notas Finais

Antes de terminar o presente documento não poderíamos deixar de realizar uma apreciação global sobre o nosso trabalho durante a realização deste projeto e apontar algumas dificuldades e certas situações que achamos relevantes. Em primeiro lugar consideramos que este tipo de projeto e esta área das ciências da tecnologia é de extrema importância pois, no decorrer da nossa vida profissional como futuros engenheiros da área da informática, iremo-nos deparar com inúmeras situações em que será importante a aplicação dos conteúdos e temas lecionados durante as aulas da unidade curricular de Metodologias Experimentais em Informática de forma a tirar conclusões sobre produtos de software, estudos de mercado, questionários a populações, entre outros.

No que toca ao trabalho desenvolvido ao longo destas três fases do trabalho prático achamos que existiram alturas em que existiu alguma hesitação da nossa parte na realização do projeto (experiências, testes, etc.) pela falta de conhecimento dos temas necessários, falta essa que levou a momentos de maiores dúvidas que procurámos aprofundar e esclarecer junto do docente.

Mais concretamente sobre esta terceira e última meta, achamos que as hipóteses, ao serem simples, tiraram um pouco de conteúdo ao nosso trabalho pois vimo-nos na situação em que seria possível usar o mesmo tipo de teste e tirar conclusões de forma bastante fácil devido à natureza das hipóteses registadas. Esta facilidade na aceitação ou rejeição das hipóteses após poucos testes impediu-nos de explorar outros tipos de testes como, por exemplo, os testes não paramétricos e outros tipos de testes estatísticos falados nas aulas.

Gonçalo São Marcos  
gmarcos@student.dei.uc.pt  
2018284198  
PL2

José Esperança  
esperanca@student.dei.uc.pt  
2018278596  
PL2