

Licenciatura em Engenharia Informática – DEI/ISEP

**Análise de Dados em Informática 2018/2019**

Aula Teórico-Prática

**Ficha Teórico-Prática 3**

**Testes de Hipóteses**

**Objetivos:**

- Testar o valor hipotético de um parâmetro
- Familiarização com a ferramenta R no suporte aos testes de hipóteses paramétricos;
- Análise e discussão de resultados.

**Exercícios**

1. Escolheram-se aleatoriamente 15 computadores portáteis de uma determinada marca, e obtiveram-se as seguintes medidas para as suas espessuras (em mm):

30	30	30	30	31	32	32	32	32	33	33
34	34	34	35							

Considerando que a espessura do computador é uma variável aleatória normal, teste a hipótese  $H_0: \mu = 32,5$  contra  $H_1: \mu \neq 32,5$  (admita que  $\alpha = 0,05$ ).

2. Uma loja *online* indicou no seu *website* que a entrega é realizada, em média, em 5 dias. Um cliente habitual efetuou uma reclamação, afirmando que o tempo médio de entrega foi superior ao valor indicado pela loja. Para averiguar se o cliente tem razão, foram analisadas aleatoriamente 36 compras efetuadas no respetivo *website*, tendo-se registado os tempos de entrega, em dias, abaixo indicados:

5	4	4	5	5	5	6	5	4	4	3	4	4	5	5	7	6	5
6	4	6	5	5	6	6	6	4	4	5	5	5	3	6	3	6	5

Considerando um nível de significância de 1%:

- Formule a hipótese nula e a respetiva hipótese alternativa para o problema. Trata-se de um teste unilateral ou bilateral?
- Indique o valor observado da estatística teste.
- O que se pode concluir relativamente à reclamação do cliente?

3. Numa empresa, o departamento de controlo da qualidade quer efetuar alguns testes sobre o peso de um determinado modelo de portátil, cujo peso, segundo as especificações de fabrico, é de 2,5 kg. Com o objetivo de verificar se o peso efetivo de cada portátil ultrapassa o indicado nas suas especificações foram selecionados, aleatoriamente, 16 portáteis do referido modelo. Os pesos, em gramas são os indicados na tabela seguinte.

2550	2550	2450	2560	2520	2530	2530	2500
2490	2510	2520	2520	2530	2510	2550	2550

Considerando que o peso dos computadores segue uma distribuição normal:

- Formule a hipótese nula e a respetiva hipótese alternativa para o problema. Trata-se de um teste unilateral ou bilateral?
  - Para um nível de significância de 10%, indique o valor observado da estatística teste. É possível concluir que o peso de cada portátil indicado nas especificações de fabrico é de facto o peso correto?
4. Registaram-se as velocidades máximas (em dpi) de 24 ratos de computadores de uma determinada marca, divididos em 2 grupos: com e sem fio. Os resultados foram:

<b>Com fio</b>	2300	2000	1800	2000	2400	2200	2000	1800	1900	2100	2200	2400
<b>Sem fio</b>	2400	2200	1800	1900	1800	1900	2100	2050	2200	2000	1900	2000

- Indique se as amostras são independentes ou emparelhadas.
  - Para um nível de significância de 1%, assumindo que a velocidade é uma variável aleatória normal, teste a hipótese de igualdade das velocidades médias, nos ratos com e sem fio.
5. Considera-se que os indivíduos canhotos têm mais força na mão esquerda do que na mão direita. Para testar esta hipótese foram registadas as forças (em N), da mão direita e da mão esquerda, de 6 pessoas canhotas:

Força (em N)	Pessoa					
	1	2	3	4	5	6
mão esquerda	140	90	125	130	95	121
mão direita	138	89	126	128	92	122

- Indique se as amostras são independentes ou emparelhadas.
- Assumindo a normalidade dos dados, diga se estes apoiam a hipótese, com um nível de significância de 5%.

6. O diretor de um hotel resolveu investir numa obra durante o ano de 2014, com o objetivo de modernizar o empreendimento, melhorar a qualidade dos seus equipamentos, e assim atrair um maior número de clientes. Para verificar se esta remodelação teve um efeito positivo, foram registadas as taxas mensais de ocupação, em %, em 2013 (antes das obras) e em 2015 (depois das obras). Os dados obtidos são apresentados na tabela seguinte.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Taxas médias de ocupação, em %												
Antes das obras (2013)	20	35	40	55	60	75	95	100	90	80	45	25
Depois das obras (2015)	25	30	45	75	80	100	100	100	100	85	65	30

Admitindo que os dados da amostra provêm de uma distribuição normal:

- Indique se as amostras são independentes ou emparelhadas.
  - Formule a hipótese nula e a respetiva hipótese alternativa para o problema. Trata-se de um teste unilateral ou bilateral?
  - Usando um nível de significância de 0,05, indique o p-valor e interprete.
  - Que podemos concluir acerca do investimento realizado?
7. Um engenheiro informático pretende averiguar se existe uma diferença significativa entre a duração das baterias de computadores com as mesmas características mas de marcas distintas. Para tal, 12 computadores da marca A e 12 computadores da marca B foram selecionados aleatoriamente e registou-se para cada um a duração da bateria, em horas, como podemos observar na tabela abaixo.

<b>Marca A</b>	6,3	5,2	6,0	6,1	6,5	5,6	5,8	6,0	5,9	5,8	5,9	6,2
<b>Marca B</b>	4,8	6,7	7,1	5,0	6,2	6,1	6,0	5,9	7,0	4,5	5,3	6,2

Com um grau de confiança de 90% e assumindo que a duração da bateria é uma variável aleatória normal, o que podemos concluir?

8. No *website* de uma universidade, a seguinte notícia é publicada: “mais de 85% dos estudantes usam o Windows como sistema operativo”. Para averiguar a veracidade desta notícia, 200 estudantes foram escolhidos aleatoriamente e registou-se o sistema operativo usado. Os dados obtidos encontram-se no ficheiro *Ex 8\_dados\_TP6e7\_TH.xlsx*.
- Formule a hipótese nula e a respetiva hipótese alternativa para o problema. Trata-se de um teste unilateral ou bilateral?
  - Para um nível de significância de 5%, que podemos concluir acerca da notícia publicada?

9. Pretende-se realizar um estudo sobre o uso das redes sociais nos jovens portugueses entre os 20 e os 25 anos. Para tal, foram escolhidos aleatoriamente 150 jovens e para cada um registou-se: o sexo, o tempo diário dispensado nas redes sociais em horas e a rede social usada (*Facebook*, *Twitter*, *Instagram*, *Linkedin*). Os dados recolhidos encontram-se no ficheiro *Ex 9\_dados\_TP6e7\_TH.xlsx*.

Considerando um grau de confiança de 95%:

- Construa um intervalo de confiança para o tempo médio diário passado nas redes sociais pelos jovens do sexo masculino.
  - Verifique se os tempos diários médios passados nas redes sociais diferem significativamente entre os homens e as mulheres.
  - Podemos concluir que mais de metade dos jovens utiliza o *Facebook*?
10. Uma empresa do ramo químico produz diferentes tipos de produtos, mas somente pode fabricar um de cada vez. O gestor de produção tem que decidir quanto ao problema de sequenciamento de 5 tarefas numa única máquina. Para cada tarefa  $j$  ( $j=1, \dots, 5$ ), seja  $p_j$  o tempo de processamento,  $d_j$  a data de entrega e  $w_j$  a penalização no caso da tarefa  $j$  se atrasar. O objetivo consiste em encontrar uma sequência que minimize a soma dos atrasos pesados  $1 \mid \mid \sum w_j T_j$ . O número de soluções admissíveis para este problema corresponde a  $5! = 120$ .

Instâncias	Ótimo	ACO	PSO	ABC
Wa40_1	913	1003	913	913
Wa40_2	1225	1781	1225	1225
Wa40_3	537	1557	537	537
Wa40_4	2094	3020	2094	2094
Wa40_5	990	1180	990	990
Wa40_6	6955	17008	7614	6955
Wa40_7	6324	10116	6324	6324
Wa40_8	6865	10853	7048	6865
Wa40_9	16225	23467	16289	16225
Wa40_10	9737	16800	9741	9737
Wa40_11	17465	38563	20334	17465
Wa40_12	19312	33729	19961	19312
Wa40_13	29256	49262	30812	29256
Wa40_14	14377	28531	14497	14377
Wa40_15	26914	46192	27718	26914

Foram consideradas para o estudo computacional instâncias do problema de sequenciamento de Máquina Única (WT - "Weighted Tardiness") com 40 tarefas, 50 e 100 tarefas. Foram retirados os resultados das diferentes MetaHeurísticas (MH), ACO, PSO e ABC, em apenas uma corrida, pretendendo-se analisar o seu desempenho. A tabela descreve um excerto dos dados obtidos para as instâncias com 40 tarefas (*Ficheiro de Dados SingleMachine-Técnicas de Otimização.csv da ficha TP1*).

- Construa um intervalo de confiança para o desempenho médio da Meta-Heurística PSO.
- Verifique se os desempenhos das MH's: PSO e ABC diferem significativamente entre si.

11. Os sistemas electrónicos, tais como as redes de computadores, os sistemas de gestão e os sistemas de segurança são ferramentas essenciais para a garantia da continuidade e eficiência dos processos de negócio e fazer com que as organizações funcionem de forma eficaz. Um técnico de informática pretende analisar o desempenho de uma fonte de alimentação ininterrupta, também conhecida pelo acrónimo UPS (sigla em inglês de *Uninterruptible Power Supply*). Uma UPS é um sistema de alimentação secundário de energia eléctrica que fornece energia de emergência e permite a estabilização da alimentação eléctrica (reduzindo os efeitos dos picos de tensão) aos equipamentos electrónicos a este ligado. Para tal foi medida a tensão (volts) à saída da fonte de alimentação em vários momentos ao longo do dia, tendo-se registado os valores constantes da tabela em baixo. Considera-se que a fonte está em perfeitas condições se a tensão for à saída próxima de 220 volts e a variabilidade for de  $\pm 10$  volts.

214	220	228	245	226	229	233	218	230	239	212	220	225	250	240	219	244	230
237	241	234	214	225	242	235	236	233	218	256	225	223	231	236	241	228	244

- Construa um intervalo de confiança a 95% para a tensão média à saída da UPS.
  - Teste a hipótese de a tensão média ser 220 volts (determine o p-valor associado). Compare com o resultado obtido na alínea anterior.
  - A fonte encontra-se em perfeitas condições de funcionamento?
  - Assumindo uma distribuição normal para a tensão, determine um intervalo de confiança a 95% para o desvio-padrão. O que conclui?
12. Na tabela seguinte encontram-se os quocientes entre o custo final e o custo inicialmente previsto dos projetos de I&D realizados em 4 grandes empresas.

Empresa	Custo final/Custo previsto					
A	1.0	0.8	1.9	1.1	2.7	
B	1.7	2.5	3.0	2.2	3.7	1.9
C	1.0	1.3	3.2	1.4	1.3	2.0
D	3.8	2.8	1.9	3.0	2.5	

Admitindo a normalidade dos dados, pretende-se investigar se o fator "Empresa" tem efeito sobre o agravamento dos projetos. Considere um nível de significância ( $\alpha = 0.05$ ).