

Licenciatura em Engenharia Informática – DEI/ISEP
Análise de Dados em Informática 2023/2024

Ficha Teórico-Prática

Testes de Hipóteses Paramétricos

Objetivos:

- Familiarização com a ferramenta R no suporte aos testes de hipóteses paramétricos;
- Análise e discussão de resultados.

Exercícios Propostos

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 o nível de significância é 5%, $\alpha=0.05$).

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%:

- a) 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.
 - b) O que se pode concluir relativamente à reclamação do cliente?
3. Registaram-se as velocidades máximas (em dpi) de 24 ratos de computador, de uma determinada marca, divididos em 2 grupos: **com fio** e **sem fio**. Os resultados foram:

Com fio	2300	2000	2800	2000	2400	2200	2000	1800	1900	2100	2200	2400
Sem fio	2400	2200	1800	1900	1800	1900	2100	2050	2200	2000	1900	2000

- a) Indique se as amostras são independentes ou emparelhadas.
- b) 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.
4. Um ensaio de rotura à compressão efetuado sobre 12 provetes cúbicos de betão conduziu aos seguintes valores da tensão de rotura (kgf / cm^2).

263 254 261 236 228 253 249 262 250 252 257 258

Admita que a variável em estudo segue uma distribuição normal.

- a) Um engenheiro pretende saber se a tensão esperada de rotura não é inferior a $255 \text{ kgf} / \text{cm}^2$. Que evidência fornecem os dados acerca desta questão se se admitir um nível de significância menor ou igual a 5?
- b) Sabendo que o valor característico da tensão de rotura se define como o valor da variável que tem uma probabilidade de 95 de ser excedido, calcule uma estimativa do valor característico da tensão de rotura do betão em estudo.
5. Para testar um novo medicamento para tratar a perturbação de hiperatividade com défice de atenção (PDAH) escolheu-se uma amostra aleatória de 24 crianças com PDAH. Cada criança foi medicada com o novo medicamento (D60) e com um placebo (D0). Após cada tratamento, cada criança respondeu a 60 perguntas. O número de respostas corretas encontram-se guardadas no ficheiro **Data_C.csv**. Supondo que o número de respostas corretas é uma variável com distribuição normal, há razões para supormos que o novo tratamento aumenta o número médio de respostas certas? Use um índice de significância de 1%.
6. Dois grupos de 36 estudantes foram selecionados ao acaso para participarem numa experiência que consiste em aprender o significado de palavras numa língua que não conhecem. Durante 30 minutos os estudantes tentaram aprender o maior número de palavras. No grupo I os estudantes trabalharam isoladamente. No grupo II os estudantes trabalharam aos pares procurando certificar-se mutuamente que iam aprendendo as palavras. Em seguida foi efetuado um teste para determinar o número de palavras aprendidas por cada aluno. Os resultados podem ser encontrados em **Data_A.csv**. Averigue se o segundo método de aprendizagem pode considerar-se significativamente melhor do que o primeiro método.
7. Efetuaram-se ensaios em 10 e 8 amostras aleatórias de 1kg petróleo bruto, provenientes de furos de dois campos, para se determinar a quantidade (em gramas) de enxofre. Os resultados obtidos encontram-se indicados na seguinte tabela:

Campo A:	111	114	105	112	107	109	112	110	110	106
Campo B:	109	103	101	105	106	108	110	104		

Supondo que a quantidade de enxofre é uma variável com distribuição normal, poderá afirmar que, em média, a quantidade de enxofre por quilograma do campo A é:

- a) superior à do campo B? b) inferior à do campo B?
8. O diretor de um hotel resolveu investir numas obras 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?
 - Indique o p-valor e interprete (use um nível de significância $\alpha=0.05$).
 - Que podemos concluir acerca do investimento realizado?
9. Uma empresa tem 4 fornecedores de placas de fibra de carbono. A divisão de controlo pretende averiguar se a qualidade das placas é igual em todos os fornecedores. Para o efeito selecionaram-se aleatoriamente 10 placas de cada fornecedor, e, efetuaram testes à resistência das placas. Os resultados dos testes encontram-se no ficheiro **Data_D.csv**. Há razões para admitir que há diferenças significativas na qualidade das placas nos 4 fornecedores ?
10. 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. Admitindo a normalidade dos dados, pretende-se investigar se o fator “Empresa” tem efeito sobre o agravamento dos projetos. Considere um nível de significância ($\alpha=0.05$).

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	

Exercícios de Consolidação

- Segundo um estudo de 2019, Portugal está entre os países da OCDE com maior percentagem de bebés de baixo peso à nascença. No ficheiro **nutri.csv** estão os dados recolhidos no âmbito de um projeto com dados de recém-nascidos em várias maternidades do país num determinado ano, selecionados de forma aleatória do total de bebés nascidos nesse ano. Verifique:

- a) para um nível de significância de 0.01, se a média do peso dos recém-nascidos é inferior à média do peso dos recém-nascidos em 2011 que foi de 3400 g.
 - b) se os dados concordam com a hipótese de que as meninas (sexo=2) são em média mais leves do que os rapazes (sexo=1) à nascença.
2. No âmbito de um projeto com o objetivo de investigar o impacto do exercício físico regular nos níveis de triglicéridos, foi medida a concentração de triglicéridos (mmol/L) no soro sanguíneo de 34 voluntários do sexo masculino, antes e depois da participação num programa de exercício físico de 10 semanas.
- Recorrendo a um teste adequado, verifique se existe evidência estatística ao nível de 1%, com base nos dados que se encontram no ficheiro **tri.csv**, de que o exercício físico regular reduz o nível de triglicéridos no sangue.
3. Num ensaio clínico pretende-se estudar a forma de administração do medicamento antiviral X a pacientes com gripe. Os pacientes foram distribuídos aleatoriamente por 3 grupos: os que receberam X inalado (grupo 1), os que receberam X inalado e intranasal (grupo 2) e os que receberam placebo (grupo 3). O intervalo de tempo (dias) até ao alívio dos principais sintomas de gripe foi medido e os dados encontram-se no ficheiro **antiviral.csv**.
- a) Estabeleça a hipótese nula e alternativa apropriadas ao problema.
 - b) Represente os dados graficamente e teste a hipótese nula com o teste que considere apropriado.
4. Para cada uma das seguintes situações suponha que se está a testar $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$. Indique se existe evidência estatística significativa de H_1 nas seguintes situações:
- a) Valor $p = 0.078$ para um nível de significância $= 0.10$. Há evidência de H_1 pois $p < \alpha$
 - b) Valor $p = 0.055$ para um nível de significância $= 0.05$. Não há evidência de H_1 pois $p > \alpha$
 - c) $t(\text{observado}) = 3.75$ com 19 graus de liberdade para $\alpha = 0.01$