

Licenciatura em Engenharia Informática

Métodos Estatísticos

Exercícios

Testes de Hipóteses Não Paramétricos

Departamento de Matemática



7 Testes de Hipóteses Não Paramétricos

7.1 Teste de Ajustamento

7.1.1 Teste de Ajustamento do Qui-Quadrado

Exercício 7.1 Foram inquiridos 995 indivíduos sobre a preferência em relação a 5 marcas de comprimidos para a gripe, tendo-se obtido os resultados apresentados na tabela seguinte:

Marca	Número de consumidores
1	190
2	210
3	180
4	205
5	210

Verifique, para um nível de significância de 5%, se os dados se ajustam a uma distribuição Uniforme Discreta.

Exercício 7.2 O recenseamento de 320 famílias com 5 filhos forneceu os seguintes resultados:

Rapazes	0	1	2	3	4	5
Famílias	8	40	88	110	56	18

Verifique, para um nível de significância de 1%, se estes resultados são compatíveis com a hipótese do número de rapazes por família ser uma variável aleatória com distribuição binomial, admitindo a equiprobabilidade dos sexos.

Exercício 7.3 A procura diária de um certo produto foi, em 40 dias escolhidos ao acaso, a seguinte:

Número de unidades	Número de dias
0	6
1	14
2	10
3	7
4	2
5	1

Será que tais observações foram extraídas de uma população com distribuição de Poisson, isto é, será de admitir que tal procura segue uma distribuição de Poisson? Considere um nível de significância de 5%.

Exercício 7.4 O Departamento de Defesa de determinado país supõe que a distribuição de probabilidade do número de avarias em determinado equipamento de um submarino, durante a realização de uma dada tarefa, segue uma distribuição de Poisson. Os dados relativos a 500 dessas tarefas foram os seguintes:

Número de falhas/tarefa	0	1	2	3	4
Número de tarefas	185	180	95	30	10

Analise, para um nível de significância de 5%, a hipótese do citado Departamento.

Exercício 7.5 Aceita-se com frequência que o número de gralhas por página de um livro, segue uma distribuição de Poisson. Através de uma amostra de 400 páginas de um livro obtiveram-se os dados seguintes:

Número de gralhas	0	1	2	3	4	5	6	7	8
Número de páginas	14	48	101	98	87	37	12	2	1

Verifique se a hipótese habitualmente aceite é válida para um nível de significância de 5%.

Exercício 7.6 Numa fábrica de motores elétricos, o gerente de produção precisa avaliar o problema de ruído excessivo do motor. Uma das possíveis causas está associada com variações no diâmetro do eixo. Assim, o responsável de produção mediu o diâmetro do eixo de 200 motores (em milésimos de milímetros) e obteve os resultados apresentados no ficheiro “motores.xlsx”

1. Faça uma análise descritiva dos dados e sugira uma distribuição adequada a este conjunto de dados.
2. Recorra ao teste de ajustamento do Qui-Quadrado para testar a distribuição sugerida e considere um nível de significância de 5%.

Exercício 7.7 Considere o ficheiro “População.xlsx”. Recorra ao teste de ajustamento do Qui-Quadrado e considere um nível de significância de 5% para testar as seguintes afirmações:

1. A variável “altura” pode ser modelada por uma distribuição Normal.
2. A variável “peso” pode ser modelada por uma distribuição Normal.
3. A variável “tempo (em minutos) que a medicação demora a fazer efeito” pode ser modelada por uma distribuição Exponencial.

7.1.2 Teste de Ajustamento de Kolmogorov-Smirnov

Exercício 7.8 Num determinado Instituto Politécnico do país efetuou-se um contrato com uma determinada empresa que ficou responsável pelo abastecimento da carne que compunha as refeições na cantina dessa escola. O contrato refere uma média de 290 gramas de carne por refeição, por estudante. No entanto, alguns alunos queixaram-se acerca da comida, em particular acerca da quantidade de carne servida por refeição. Os alunos falaram com o cozinheiro chefe, que lhes disse que a quantidade de carne servida por refeição a cada estudante tinha aproximadamente distribuição normal de média 290 gramas com um desvio padrão de 56 gramas. Após esta conversa com o cozinheiro, alguns alunos concordaram em recolher as suas refeições ao longo de vários dias, resultando assim uma amostra de 10 refeições, que foram levadas para um laboratório afim de serem pesados os pedaços de carne nelas contidos. Os dados obtidos são os seguintes:

198; 254; 262; 272; 275; 278; 285; 287; 287; 292.

Ao nível de significância de 5%, há evidência para rejeitar a hipótese de que o cozinheiro seguia as regras que afirmou em relação à quantidade de carne servida?

Exercício 7.9 Os níveis de glicose no sangue (mg/100ml) numa amostra aleatória de 18 indivíduos, aparentemente saudáveis são os seguintes:

75; 92; 80; 80; 84; 72; 84; 77; 81;
77; 75; 81; 80; 92; 72; 77; 78; 76.

Pretende testar-se, para um nível de significância de 5%, se os dados se referem a uma distribuição normal de média 80 mg e desvio padrão 6.91 mg.

Exercício 7.10 Suponha que os dados seguintes representam os erros de ajustamento (resíduos), obtidos através de um modelo linear:

-2.7; -1.1; -0.5; 0.4; 1.8; -2.1; -1.0; -0.3; 0.8; 2.0;
-1.9; -0.9; 0.1; 1.2; 2.6; -1.7; -0.7; 0.4; 1.3; 2.9.

Verifique se os resíduos se distribuem normalmente, com média nula e variância 2.5 para um nível de significância de 5%.

Exercício 7.11 No âmbito de um programa de reabilitação, registaram-se os QI's de 30 jovens dependentes do consumo de drogas:

95; 100; 91; 106; 109; 110; 98; 104; 97; 100;
107; 119; 92; 106; 103; 106; 105; 112; 101; 91;
105; 102; 101; 110; 101; 95; 102; 104; 107; 118.

Para um nível de significância de 0.05, poder-se-á concluir que os dados provêm de uma população normal com média de QI de 105 e desvio padrão de 10? Determine o valor-p.

7.2 Teste de Independência do Qui-Quadrado

Exercício 7.12 Considere a seguinte tabela de dados em que 1237 indivíduos adultos são classificados segundo a pressão sanguínea (mm Hg) e o nível de colesterol (mg /dl):

		Pressão sanguínea		
		< 127	127 a 166	> 166
Nível de colesterol	< 200	117	168	22
	200 a 260	204	418	63
	> 260	67	145	33

Verifique, para um nível de significância de 5%, se existe independência entre a pressão sanguínea e o nível de colesterol.

Exercício 7.13 Um supermercado quer testar ao nível de significância de 5% a hipótese do modo de pagamento dos clientes nesse estabelecimento ser independente do período do dia em que fazem as compras. Existem quatro modos de efetuar os pagamentos (por cheque, dinheiro, cartão de débito e cartão de crédito). A tabela seguinte apresenta os resultados obtidos numa amostra de 4000 clientes:

		Período do dia		
		Manhã	Tarde	Noite
Modo de Pagamento	Cheque	50	100	50
	Dinheiro	175	375	100
	Cartão de débito	500	1025	475
	Cartão de crédito	275	500	375

Exercício 7.14 Duzentas donas de casa foram inquiridas sobre a frequência com que veem televisão, e também sobre as suas preferências relativamente aos detergentes A, B ou C. A informação obtida resultou na seguinte tabela de contingência:

		Frequência com que vê TV		
		Nunca	Ocasionalmente	Frequentemente
Tipo de detergente	A	10	25	40
	B	10	40	31
	C	5	9	30

Verifique, para um nível de significância de 0.05, se a marca de detergente que as donas de casa consomem é independente da frequência com que veem televisão.

Exercício 7.15 A tabela seguinte apresenta os resultados de uma experiência destinada a investigar o efeito da vacinação de animais de laboratório contra determinada doença:

	Contraíram a doença	Não contraíram a doença
Vacinados	9	42
Não vacinados	17	28

Utilizando um nível de significância de 0.01, verifique a independência dos efeitos da vacinação sobre a doença.

Exercício 7.16 A tabela seguinte apresenta os resultados de uma pesquisa destinada a determinar se a idade do motorista tem influência no número de acidentes:

		Idade				
		18-30	31-43	44-56	57-69	70-82
Número de acidentes	0	748	821	786	720	672
	1	74	60	51	66	50
	2	31	25	22	16	15
	> 2	9	10	6	5	7

Verifique para um nível de significância de 0.05, a existência de independência entre estes dois fatores.

Exercício 7.17 Foi efetuado um estudo para verificar a preferência pela leitura de um determinado jornal e o nível de ensino dos leitores. Os resultados obtidos foram:

		Jornal		
		A	B	C
Nível de ensino	Básico	10	8	12
	Secundário	80	172	135
	Superior	200	250	233

Verifique para um nível de significância de 0.10, a existência de independência entre estes dois fatores.

Exercício 7.18 Considere o ficheiro “População.xlsx”. Recorra ao teste de independência do Qui-Quadrado e considere um nível de significância de 5% para testar as seguintes afirmações:

1. Há independência entre os fatores: indivíduo ter sido infetado por VNO (vírus do Nilo Ocidental) e o país em que vive.
2. Há independência entre os fatores: indivíduo ter sido infetado por VNO (vírus do Nilo Ocidental) e viver perto de um lago.
3. Há independência entre os fatores: indivíduo ter sido infetado por VNO (vírus do Nilo Ocidental) e o seu género.

7.3 Teste à Igualdade de Duas Distribuições

7.3.1 Teste de Wilcoxon

Exercício 7.19 Os valores registados na tabela seguinte referem-se a uma amostra de 9 indivíduos, classificados de acordo com as respostas a um questionário efetuado antes e depois da população a que pertencem ter recebido um conjunto de instruções sobre regras de higiene alimentar. As pontuações aumentam com a melhoria dos hábitos alimentares.

Indivíduo	1	2	3	4	5	6	7	8	9
Antes	21	33	30	34	26	25	15	20	21
Depois	21	36	26	35	31	25	17	20	21

Para um nível de significância de 5%, conclua se se registam progressos nos hábitos alimentares da população.

Exercício 7.20 Num estudo sobre nutrição, pretendem-se avaliar determinados tipos de dieta, bem como as perdas de peso associadas a cada dieta. Para tal realizou-se a experiência, em que a um conjunto de 10 pessoas que seguiu um determinado tipo de dieta, analisou-se o peso antes e depois do plano de dieta. Os pesos (em kg) foram os seguintes:

Pessoa	1	2	3	4	5	6	7	8	9	10
Antes	82.7	73.2	84.1	84.1	81.6	78.9	85.6	80.2	84.5	73.8
Depois	74.5	73.2	79.1	85.6	81.6	76.9	81.5	80.2	86.9	73.8

Para um nível de significância de 5%, será que existem diferenças significativas no peso antes e depois do plano de dieta?

Exercício 7.21 Num estudo sobre alcoolismo um grupo de investigadores pretendia conhecer o efeito inibidor de um novo medicamento (DHT) e por isso recolheu os dados seguintes referentes a níveis de testosterona (em nmol/lb) em 12 homens saudáveis antes e depois do tratamento com DHT:

Homem	1	2	3	4	5	6	7	8	9	10	11	12
Antes	21.5	23.0	21.0	21.8	22.8	14.7	21.0	23.4	20.0	29.5	23.5	20.5
Depois	9.4	17.2	13.0	6.4	4.8	4.5	10.7	15.6	12.5	7.7	10.8	9.2

Para um nível de significância de 1%, poder-se-á concluir com base nestes dados, que o tratamento reduz a concentração de testosterona em homens saudáveis?

Exercício 7.22 Uma amostra de 15 doentes asmáticos participou numa experiência que pretendia estudar o efeito de um novo tratamento na função pulmonar. Entre outros registos recolheram-se valores do volume expiratório forçado (VEF) por segundo (em l/s), antes e depois do tratamento:

Doente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Antes	1.69	2.77	1.00	1.66	3.00	0.85	1.42	2.82	2.58	1.84	1.89	1.91	1.75	2.46	2.35
Depois	1.69	2.22	3.07	3.35	3.00	2.74	3.61	5.14	2.44	4.17	2.42	2.94	3.04	4.62	4.42

Para um nível de significância de 5%, poder-se-á concluir com base nestes dados, que o tratamento aumenta o VEF?

7.3.2 Teste de Mann-Whitney

Exercício 7.23 Num estudo sobre nutrição, pretendem-se avaliar determinados tipos de dieta, bem como as perdas de peso associadas a cada dieta. Para tal realizou-se a experiência, em que a cada uma de 10 pessoas com excesso de peso atribuiu-se aleatoriamente um de dois planos de dieta. As perdas de peso (em kg) foram as seguintes:

Dieta 1	2.7	3.2	4.1	4.5	3.8
Dieta 2	4.5	3.6	4.1	3.9	3.9

Para um nível de significância de 5%, pode admitir que as duas dietas são realmente distintas?

Exercício 7.24 É medida a altura de determinada espécie de árvores com igual número de anos em duas localidades diferentes, A e B, pois há suspeitas de que a altura das espécies de árvores da localidade A seja diferente da altura das espécies de árvores da localidade B. Foram recolhidas algumas amostras de modo aleatório, registando-se o seguinte:

localidade A	8	14	11	9	10	12	
localidade B	7	5	10	6	10	11	8

Para um nível de significância de 5%, pretende-se verificar se as alturas são realmente distintas.

Exercício 7.25 *Lebranchu et al.* conduziu um estudo em que observou 9 pacientes com Imunodeficiência Variável Comum (IVC), cujos resultados foram confrontados com os de um grupo de controle constituído por 12 indivíduos aparentemente normais. Entre os dados recolhidos encontravam-se os seguintes, referentes ao número de células CD4+ por mm³ de sangue periférico:

Pacientes IVC	623	437	370	300	330	527	290	730	1000			
Grupo de controle	710	1260	717	590	930	995	630	977	530	710	1275	825

Será possível concluir, com base nestes dados e para $\alpha = 0.01$, que os pacientes IVC têm um nível reduzido de células CD4+ ?

Exercício 7.26 Um investigador propôs-se estudar as modificações ocorridas nas vias respiratórias de um grupo de doentes asmáticos após inalação de um antigénio. Um grupo de controle constituído por indivíduos não asmáticos, supostamente saudáveis foi sujeito ao mesmo procedimento. Na tabela seguinte apresentam-se dados referentes à percentagem de antigénio presente em recolhas resultantes de lavagem bronco-alveolar:

Não asmático	70	55	63	68	73	77	67			
Asmático	64	25	70	35	43	49	62	56	43	66

Poder-se-á concluir, com base nestes dados e para $\alpha = 0.05$, que o grupo de doentes asmáticos tem uma percentagem inferior de antigénio nas recolhas efetuadas?

Soluções

7.1 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que os dados se ajustam a uma distribuição Unifome Discreta.

7.2 Com base na amostra e para um nível de significância de 1%, não se rejeita H_0 , ou seja, há evidência estatística que estes resultados são compatíveis com uma distribuição Binomial.

7.3 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que a procura segue uma distribuição de Poisson.

7.4 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que o número de avarias em determinado equipamento de um submarino segue uma distribuição de Poisson.

7.5 Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, não há evidência estatística que o número de gralhas por página de um livro segue uma distribuição de Poisson.

7.6 1) Distribuição Normal **2)** Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, não há evidência estatística que o diâmetro do eixo dos motores segue uma distribuição de Normal.

7.7 1) Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, não há evidência estatística que a altura segue uma distribuição de Normal. **2)** Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, não há evidência estatística que o peso segue uma distribuição de Normal. **3)** Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que o tempo que a medicação demora a fazer efeito segue uma distribuição de Exponencial.

7.8 Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que o cozinheiro não segue as regras que afirmou em relação à quantidade de carne servida.

7.9 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que os dados se referem a uma distribuição normal de média 80 mg e desvio padrão 6.91 mg.

7.10 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que os resíduos se distribuem normalmente com média nula e variância 2.5.

7.11 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que os dados provêm de uma população normal com média de QI de 105 e desvio padrão de 10.

7.12 Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que não existe independência entre a pressão sanguínea e o nível de colesterol.

7.13 Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que não existe independência entre modo de pagamento dos clientes nesse estabelecimento e o período do dia em que fazem as compras.

7.14 Com base na amostra e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que não existe independência entre a frequência com que veem televisão e as preferências relativamente aos detergentes.

7.15 Com base na amostra e para um nível de significância de 1%, não se rejeita H_0 , ou seja, há evidência estatística que existe independência entre contrair ou não a doença e ser ou não vacinado.

7.16 Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que existe independência entre a idade do

motorista e o número de acidentes.

7.17 Com base na amostra e para um nível de significância de 10%, rejeita-se H_0 , ou seja, há evidência estatística que não existe independência entre a preferência por determinado jornal e o nível de ensino dos leitores.

7.18 **1)** Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que existe independência os indivíduos terem sido infetados por VNO e o país em que vivem. **2)** Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que existe independência entre os indivíduos terem sido infetados por VNO e viverem perto de um lago. **3)** Com base na amostra e para um nível de significância de 5%, não se rejeita H_0 , ou seja, há evidência estatística que existe independência entre os indivíduos terem sido infetados por VNO e o seu género.

7.19 Com base nas amostras e para um nível de significância de 5%, não se rejeita H_0 , ou seja, não há evidência estatística que tenham ocorrido progressos nos hábitos alimentares da população.

7.20 Com base nas amostras e para um nível de significância de 5%, não se rejeita H_0 , ou seja, não há evidência estatística que existam diferenças significativas no peso antes e depois do plano de dieta.

7.21 Com base nas amostras e para um nível de significância de 1%, rejeita-se H_0 , ou seja, há evidência estatística que o tratamento reduz a concentração de testosterona em homens saudáveis.

7.22 Com base nas amostras e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que o tratamento aumenta o VEF.

7.23 Com base nas amostras e para um nível de significância de 5%, não se rejeita H_0 , ou seja, não há evidência estatística que as duas dietas sejam distintas.

7.24 Com base nas amostras e para um nível de significância de 5%, não se rejeita H_0 , ou seja, não há evidência estatística que as alturas sejam distintas.

7.25 Com base nas amostras e para um nível de significância de 1%, rejeita-se H_0 , ou seja, há evidência estatística que os pacientes IVC têm um nível reduzido de células CD4+.

7.26 Com base nas amostras e para um nível de significância de 5%, rejeita-se H_0 , ou seja, há evidência estatística que o grupo de doentes asmáticos tem uma percentagem inferior de antigénio.