

# ANÁLISE DE DADOS

Ano Lectivo 2018/2019

## Exercícios

1. Perguntou-se a cada um dos 80 estudantes de um determinado curso, qual o seu grau de satisfação relativamente ao curso que frequenta. Obtiveram-se os seguintes resultados:

NS	MB	B	S	NS	NS	SP	SP	S	NS
NS	B	NS	NS	SP	B	B	MB	B	MB
SP	NS	NS	MB	SP	B	NS	B	S	MB
SP	S	SP	SP	NS	NS	SP	S	MB	SP
MB	S	B	MB	NS	S	S	S	B	NS
SP	S	B	NS	S	S	SP	B	MB	SP
B	B	MB	NS	B	S	NS	NS	NS	MB
B	S	MB	S	MB	NS	MB	SP	SP	S

NS-"Não Satisfaz"; SP-"Satisfaz Pouco"; S-"Satisfaz"; B- "Bom"; MB- "Muito Bom".

Faça uma representação gráfica adequada para os dados e indique uma outra que seja igualmente adequada. Calcule as características amostrais que achar convenientes. Substitua as categorias consideradas anteriormente, respectivamente por 1, 2, 3, 4 e 5. Calcule agora as características amostrais que achar convenientes. De que tipo é a variável que está a estudar?

2. O número de pares de sapatos de senhora vendidos por determinado estabelecimento comercial durante o último ano foi o apresentado no seguinte quadro:

Tamanho dos sapatos	35	35,5	36	36,5	37	37,5	38	38,5	39	39,5	40	40,5	41
Nº de pares vendidos	30	40	50	150	300	600	950	820	750	440	250	150	40

Represente graficamente os dados e encontre as características que achar convenientes.

3. Num estudo sobre a incidência de certa doença numa população de insectos, um grupo de biólogos registou ao longo de um ano o número de insectos contaminados em cada amostra de 5 insectos, tendo para tal recolhido 200 amostras. Os resultados obtidos foram os seguintes:

Nº de insectos contaminados	0	1	2	3	4	5
Nº de amostras	17	53	68	44	16	2

- a) Construa a tabela de frequências e represente graficamente os dados através de um diagrama de barras.  
b) Determine as seguintes características amostrais: moda, mediana, média, desvio-padrão e  $Q_{1/3}$ .
4. Considere a seguinte amostra ordenada de tempos de vida de uma espécie de insectos:

0.10	0.21	0.45	0.63	1.22	1.40	1.88	2.48	3.54	5.42
0.19	0.30	0.55	0.80	1.31	1.57	2.10	2.77	4.60	6.20

- a) Construa o histograma para uma amplitude de classe,  $h$ , conveniente.  
b) Determine a média e a mediana amostrais.
5. Os dados seguintes (que já se encontram ordenados) referem-se ao tempo de vida (em anos) de 50 doentes padecendo de certa doença rara:

0.8	2.5	9.7	23.5	33.2
0.9	2.6	13.5	23.6	36.6
1.0	2.6	13.5	23.7	36.7
1.1	3.2	14.4	27.1	38.0
1.1	3.5	15.5	27.6	40.2
1.7	4.8	16.2	28.1	45.0
1.9	6.3	18.2	29.7	45.1
2.0	6.9	18.2	30.9	61.7
2.0	7.6	20.7	31.2	66.4
2.4	9.0	21.8	31.7	67.4

- a) Construa um histograma dos dados, após ter escolhido uma amplitude de classe conveniente.  
b) Determine as seguintes características amostrais: média, mediana e variância.

## ANÁLISE DE DADOS

Ano Lectivo 2018/2019

### Exercícios

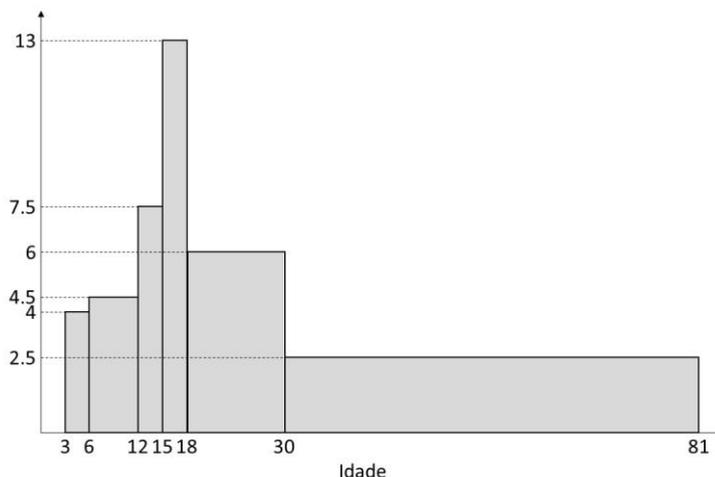
6. Fizeram-se 50 medições do comprimento (em cm) de determinada espécie animal no estado adulto, obtendo-se os seguintes valores:

10.15	10.16	10.23	10.25	10.3	10.33	10.35	10.35	10.36	10.38
10.39	10.41	10.42	10.43	10.45	10.45	10.47	10.50	10.51	10.52
10.53	10.54	10.55	10.55	10.55	10.56	10.57	10.58	10.58	10.59
10.62	10.65	10.65	10.65	10.71	10.72	10.73	10.74	10.75	10.76
10.77	10.78	10.79	10.81	10.85	10.87	10.94	10.97	10.98	11.10

- a) Determine as seguintes características: amplitude amostral, mediana e quantis  $Q_{1/3}$  e  $Q_{2/3}$ .  
b) Faça uma representação gráfica dos dados que ache conveniente.
7. Mediu-se a percentagem de gordura no leite, de 120 vacas pertencentes a uma exploração leiteira Açoriana. A informação recolhida encontra-se resumida na tabela seguinte:

Classes	Freq. abs.
[3.47, 3.69[	5
[3.69, 3.91[	18
[3.91, 4.13[	37
[4.13, 4.35[	30
[4.35, 4.57[	16
[4.57, 4.79[	9
[4.79, 5.01[	5

- a) Sugira um valor aproximado para a mediana. Justifique a escolha.  
b) Calcule um valor aproximado para a média.  
c) Qual é a representação gráfica adequada para os dados?
8. Para se proceder ao planeamento da rede escolar de determinada região, fez-se um levantamento das idades dos indivíduos dessa região e resumiu-se a informação no seguinte histograma:



Atendendo a que 4% da população está em idade pré-escolar, isto é, tem idade compreendida entre os 3 e os 6 anos:

- a) Determine a percentagem de indivíduos com idades:  
i) entre os 6 e os 12 anos; ii) entre os 12 e os 18 anos; iii) superiores a 30 anos.  
b) Existem mais jovens com idades entre os 6 e os 12 anos ou entre os 12 e os 15 anos?  
c) Diga em que classes etárias se encontram os 1º e 3º quartis. Justifique.
9. Os dados seguintes dizem respeito à dureza de 30 moldes de alumínio.

53.0	70.2	84.3	55.3	78.5	63.5	71.4	53.4	82.5	67.3
69.5	73.0	55.7	85.8	95.4	51.1	74.4	54.1	77.8	52.4
69.1	53.5	64.3	82.7	55.7	70.5	87.5	50.7	72.3	59.5

Faça a representação desta amostra em *Boxplot*.

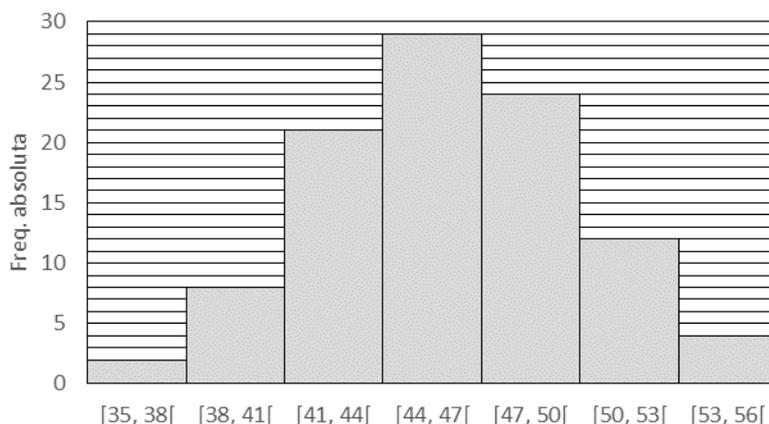
**Exercícios**

**10.** Os resíduos provenientes do ajustamento de um modelo a uma colecção de 20 pontos são:

-1.4	1.2	0.5	-0.3	-0.8	0.4	-1.3	0.5	0.7	-0.2
0.1	2.3	0.1	-0.8	-2.6	0.7	-0.9	-0.4	0.2	1.3

Faça a representação em *Boxplot* e verifique se esta colecção tem alguns *outliers*.

**11.** Num viveiro dos Serviços Florestais, está-se a ensaiar um novo tipo de pinheiro (PN), quanto ao crescimento, no nosso clima. Passados dois meses sobre o lançamento à terra das sementes, mediu-se a altura atingida pelos pinheiros, tendo-se recolhido uma amostra de dimensão 100, a partir da qual se construiu o seguinte histograma (a unidade de medida é mm):



- Qual a percentagem de pinheiros com tamanho inferior a 44 mm?
- Calcule valores aproximados para a média e desvio padrão do tamanho dos pinheiros, passados os dois meses.
- Pensa-se que o pinheiro habitual (PH) tem um crescimento muito mais lento que esta nova espécie ensaiada, admitindo-se até que o crescimento do PH seja metade do do PN. Por outro lado, pensa-se que se se utilizar um fertilizante adequado, o PN cresce mais 10 mm do que se não se utilizar o fertilizante. Tendo em consideração o histograma apresentado pela amostra de PN, esboce histogramas que representem uma amostra de PH e outra amostra de PN com fertilizante. Justifique os esboços apresentados.

**12.** Dividiu-se uma propriedade em porções com área de 10m<sup>2</sup> cada uma. A amostra seguinte representa o número de plantas doentes, registadas em 40 porções escolhidas aleatoriamente:

0	5	11	11	11	12	12	17	18	21
22	29	33	35	35	39	42	42	44	44
48	48	50	61	62	73	81	83	83	88
91	91	91	94	94	111	117	141	147	210

- Calcule as seguintes características amostrais: média, mediana, variância e os quartis.
- Investigue a existência de possíveis *outliers*.
- Represente graficamente os dados.

**13.** No estudo do comportamento do caudal de um rio, obtiveram-se os seguintes dados dos caudais anuais:

4.84	5.06	5.20	5.28	5.35	5.68	5.69	5.71	5.71	5.76	5.99	6.07
6.39	6.53	6.56	6.57	6.59	6.68	6.82	6.84	6.85	6.91	6.96	7.03
7.04	7.09	7.24	7.27	7.27	7.29	7.46	7.49	7.52	7.62	7.63	7.83
8.04	8.05	8.15	8.24	8.28	8.44	8.90	9.10				

- Calcule a média, a mediana, e os quartis.
- Investigue a existência de possíveis *outliers*.
- Represente graficamente os dados e sugira uma distribuição que lhe pareça adequada.

## Exercícios

14. Considere de novo o exercício 11.

- Admitindo que a média e a variância da amostra recolhida de PN foram respectivamente 45.5 mm e 15.21 mm<sup>2</sup>, obtenha um intervalo de 95% de confiança para a altura média dos pinheiros ao fim de 2 meses.
- Estudos feitos sobre a nova espécie de pinheiro, permitiram concluir que o crescimento da planta é tal que, ao fim de 2 meses, a probabilidade do seu tamanho ser superior ou igual a 44 mm é igual a 0,75. Perante a amostra obtida, haverá indícios de que, nas nossas terras, a probabilidade desse crescimento ser superior ou igual a 44 mm, é inferior a 0,75?

15. a) Com a saída do novo Código da Estrada, tornou-se obrigatório o uso do cinto de segurança atrás. Com o objectivo de saber se esta norma estava a ser cumprida interrogaram-se 20 passageiros, 15 dos quais afirmaram que quando viajavam atrás usavam regularmente cinto de segurança. Poderemos concluir, a partir dos dados apresentados, que a proporção de passageiros que, viajando atrás, usam regularmente cintos de segurança é superior a 70%?

b) Resolva novamente a questão anterior, supondo agora que em 300 inquiridos, 230 responderam afirmativamente.

c) Encontre um intervalo de 95% de confiança para a verdadeira proporção de passageiros que viajando no banco de trás, usam regularmente cinto de segurança.

16. a) Com o objectivo de estudar a incidência do vírus da Hepatite B na população médica e paramédica de um determinado hospital, foi feito um rasteio a 100 empregados desse hospital que contactam usualmente com sangue e seus derivados. Encontraram-se 23 cujo resultado deu positivo. Podemos concluir que a proporção de contagiados é inferior a 25%?

b) Encontre um intervalo de 99% de confiança para a verdadeira proporção de contagiados com o vírus da hepatite B, na população em estudo naquele hospital.

17. Discute-se a possibilidade da instalação de um pólo de uma universidade numa certa região. Interessados na instalação do referido pólo, os presidentes das câmaras dessa região emitiram um comunicado conjunto afirmando que mais de 70% dos jovens pretendem ingressar no ensino superior, ao contrário do que se tem vindo a verificar nos últimos anos, em que esta percentagem é menor. Com o objectivo de testar esta afirmação, perguntou-se a jovens estudantes do ensino secundário se pretendiam ingressar na universidade, tendo-se obtido os seguintes resultados:

S	S	N	S	N	N	S	S	S	S
S	S	S	S	S	S	S	N	S	N

Tire conclusões.

18. Considere de novo o exercício 12.

- Teste se existe evidência para afirmar que a mediana do número de plantas doentes por canteiro, é diferente de 50.
- Haverá evidência para afirmar que a proporção de canteiros com mais de 100 plantas doentes é inferior a 25%?
- Determine um intervalo de 95% de confiança para a probabilidade de um canteiro ter mais de 100 plantas doentes.

19. Considere uma v.a. X com função massa de probabilidade

x	1	2	3	4	5	6
Sob $H_0$	1/6	1/6	1/6	1/6	1/6	1/6
Sob $H_1$	2/15	1/6	1/5	1/5	1/6	2/15

Recolheu-se uma amostra de dimensão 1. Se  $x \in \{3, 4\}$ , rejeita-se  $H_0$ .

- Determine o nível de significância do teste.
- Obtenha a função potência e determine a probabilidade de cometer o erro de tipo II.

## Exercícios

20. Um fabricante de lâmpadas afirma que 90% das lâmpadas produzidas na sua fábrica têm tempo de vida de pelo menos 2000 horas. Um representante da Associação de Defesa dos Consumidores recolheu aleatoriamente uma amostra de 15 lâmpadas e decidiu que rejeitava a afirmação do fabricante se nela encontrasse pelo menos três lâmpadas que falhassem antes das 2000 horas.
- a) Qual poderá ter sido o nível de significância usado na formulação do teste?  
 b) Determine a probabilidade de cometer um erro de tipo II, se a verdadeira proporção  $p$  de lâmpadas defeituosas for 0.2.

21. Duas moedas de diferentes tamanhos foram lançadas 100 vezes tendo-se registado os resultados. Os resultados possíveis são Ff, Fc, Cc e Cf (as letras maiúsculas dão o resultado da moeda maior) e estes ocorreram 19, 21, 33 e 27 vezes respectivamente. Investigue se, face aos resultados obtidos, podemos concluir que as moedas são equilibradas.

22. De acordo com a teoria de Mendel, os membros da 2ª geração de um cruzamento apresentam cor vermelha, azul ou branca com probabilidades iguais respectivamente a 9/16, 3/16 e 4/16. Em 145 exemplares observados, os resultados foram os seguintes:

vermelho	azul	branco
72	35	38

Serão estes resultados compatíveis com a teoria de Mendel, ao nível de significância de 1%?

23. É convicção popular que os nascimentos ocorrem "depois da 9ª Lua nova". Apresenta-se em baixo uma tabela onde se registaram o número de nascimentos nos 7 dias seguintes a cada fase da lua, de crianças seleccionadas ao acaso entre as nascidas numa determinada maternidade, em 1995:

Lua nova	Crescente	Lua cheia	Minguante
72	61	68	61

Com base nos dados, teste ao nível de significância de 5% se a convicção popular não tem fundamento, sendo os nascimentos distribuídos regularmente ao longo das diversas fases da Lua.

24. Numa amostra aleatória de 200 famílias, cada uma com quatro filhos, registou-se o número de raparigas em cada família, tendo-se obtido os seguintes resultados:

Nº de filhas:	0	1	2	3	4
Nº de famílias:	5	32	65	75	23

Teste o ajustamento de uma distribuição binomial a estes dados.

25. Num estudo sobre o gorgulho do feijão, introduziram-se larvas nos feijões, que lhes serviram de alimento. As crisálidas saíram através de um buraco feito no feijão e, como tal, o número de buracos por feijão indica-nos o número de insectos adultos que saíram. Na tabela abaixo são apresentados os resultados de uma amostra de 100 feijões:

Nº de gorgulhos saídos por feijão:	0	1	2	3	≥4
Nº de feijões:	60	22	10	5	3

Será o número de gorgulhos por feijão uma v.a. com distribuição de *Poisson*?

26. O médico responsável pelo gabinete médico de uma fábrica registou o nº de acidentes, por mês, verificados nessa fábrica, durante os últimos 10 anos:

Nº acidentes/mês	0	1	2	3	4	5	6	7	≥8
Nº meses	2	10	15	30	28	15	10	6	4

Relativamente aos dados anteriores:

- a) Determine valores aproximados para a média e para a variância amostral.  
 b) Faça o resumo de 5 números e investigue a existência de possíveis *outliers*.

## Exercícios

- c) Faça uma representação gráfica conveniente.  
 d) Pensa-se que o nº de acidentes por mês nessa fábrica, é uma v.a. com distribuição de *Poisson* de valor médio 4. Verifique se existem razões que nos levem a duvidar desta suposição.  
 e) Verifique se existem razões para afirmar que em mais de 15% dos meses se verificam 6 ou mais acidentes.

27. Com o objectivo de testar se uma dada moeda é equilibrada, essa moeda é lançada até se obter cara pela 1ª vez. Repetiu-se a experiência 150 vezes, tendo-se obtido os resultados seguintes:

Nº de lançamentos necessários até à 1ª cara (inclusivé)	1	2	3	4	5 ou mais
Frequência	60	48	22	11	9

Que pode concluir ao nível de significância  $\alpha = 5\%$ ?

28. Lançou-se um dado 600 vezes, tendo-se obtido os seguintes resultados:

Face	1	2	3	4	5	6
Frequência	78	106	68	129	122	97

- a) Investigue da existência de possíveis candidatos a *outliers*.  
 b) Verifique se o dado é ou não equilibrado.  
 c) A probabilidade de sair face ímpar é significativamente diferente da de sair face par?

29. Considere o exercício 6. Tendo em consideração a representação gráfica obtida na alínea b), sugira uma distribuição para a população subjacente aos dados e verifique se a sua escolha é adequada.

30. Considere os seguintes dados, que dizem respeito ao peso de 37 crianças de uma determinada classe etária:

18.2	17.4	17.6	16.7	17.1	20.1	17.9	16.8	19.6	18.4	17.7	19.3
18.4	18.6	17.8	16.9	20.6	19.8	18.7	17.5	17.8	18.3	18.9	19.6
20.6	18.7	18.3	18.8	19.6	18.6	19.9	20.7	19.6	18.9	20.8	19.6
											20.4

- a) Sugira uma distribuição de probabilidade que lhe pareça ajustar-se à população subjacente aos dados. Justifique a sua escolha.  
 b) Utilizando um teste adequado, teste a adaptabilidade do modelo sugerido em a).

31. O sr. X não consegue chegar a horas ao emprego. Todos os dias marca o ponto depois da hora estipulada para a sua entrada. No final do mês, juntamente com uma repreensão escrita, recebeu uma folha com um registo dos seus atrasos (em minutos):

0.01	2.66	3.30	3.77	4.47	5.13	7.56
8.79	10.26	14.36	15.29	19.64	21.45	28.41

- a) Investigue a existência de possíveis *outliers* na amostra.  
 b) O sr. X acha injusta a repreensão já que, segundo diz, desde que trabalha naquela empresa, mais de 50% das vezes o atraso é inferior a 5 minutos. Com base nos dados anteriores verifique se existe evidência suficiente para dar razão ao sr. X.  
 c) Sugira uma distribuição de probabilidade que lhe pareça adequada aos dados. Justifique.  
 d) Utilize o teste de ajustamento que lhe pareça conveniente para testar a adaptabilidade do modelo sugerido em c). Para a população subjacente à amostra considere que o valor médio é 10 e o desvio padrão é 10.

32. Num estudo sobre o desenvolvimento de uma certa ave marinha, fez-se a medição da largura das patas de 12 destas aves, tendo-se obtido os seguintes valores:

4.9	5.2	6.3	5.8	7.3	3.8	5.4	4.7	6.1	7.1	4.6	6.0
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Admite-se que a distribuição da largura das patas desta ave é Normal. Verifique se os valores apresentados são coerentes com a hipótese formulada.

## Exercícios

33. Com o objectivo de estudar a diferença de proporções de peças defeituosas produzidas em duas linhas de montagem, recolheu-se duas amostras de 200 peças cada, tendo-se verificado que o número de peças defeituosas foi de 15 e 27, respectivamente nas linhas de montagem 1 e 2.
- Construa um intervalo de 96% de confiança para a diferença de proporções de peças defeituosas nas duas linhas de montagem.
  - Haverá evidência para concluir que a proporção de peças defeituosas produzidas pela linha de montagem 2 é superior à da 1?
34. Considere os seguintes dados referentes ao número de alunos que beneficiaram de abonos e protecção social numa dada escola primária:

	1983	1984	1985	1986	1987	1988	1989
Total de alunos	346	321	358	406	392	402	415
Alunos c/ prot. social	42	38	45	36	31	32	31

- Construa um intervalo de 95% de confiança para a diferença de proporções  $p_1 - p_2$  de crianças carenciadas, relativas aos anos de 1983 e 1989, respectivamente.
  - Teste, ao nível de 5%, a igualdade das referidas proporções.
  - Poder-se-á concluir, a partir dos dados apresentados, que o nível económico da população escolar na zona onde está inserida esta escola melhorou significativamente de 1983 para 1989?
  - Será razoável fazer um estudo semelhante ao anterior para comparar os valores referentes a 1983 e a 1984? Justifique a sua resposta.
35. Numa experiência concebida para testar a sensibilidade relativa de dois explosivos, A e B, usados na exploração mineira, fizeram-se rebentar 33 cargas do explosivo A e 20 cargas do explosivo B numa câmara de gás e tomou-se nota se cada explosão ocasionava ignição de gás ou não. Os resultados foram como se seguem:

	Ignição	Não ignição	Totais
explosivo A	12	21	33
explosivo B	6	14	20
	18	35	53

Sejam  $P_A$  e  $P_B$  as probabilidades de ignição usando respectivamente A e B.

- Teste a hipótese nula  $P_A = P_B$ .
  - Se não rejeitou a hipótese especificada em a) determine um intervalo de 95% de confiança para  $P = P_A = P_B$ ; se rejeitou a hipótese determine um intervalo de 95% de confiança para  $P_A$
36. Num estudo sobre a obesidade realizado com 150 homens e 200 mulheres, com idades compreendidas entre os 20 e os 50 anos, verificou-se que 21 homens e 48 mulheres eram obesos. Indicarão os dados anteriores a existência de uma diferença significativa entre as duas populações, no que diz respeito à obesidade?
37. A Rodoviária Nacional afirma que a mediana do tempo de espera numa determinada paragem é no máximo de 20 minutos. Os tempos de espera de 12 passageiros escolhidos aleatoriamente foram os seguintes:

25 ; 15 ; 19 ; 16 ; 21 ; 24 ; 18 ; 18 ; 24 ; 28 ; 15 ; 11

Verifique se existem razões para duvidar da afirmação, ao nível de significância de 10 %.

38. Num estudo sobre o abuso do consumo de drogas, numa determinada área suburbana, os investigadores verificaram que a mediana do QI dos consumidores presos, com 16 ou mais anos de idade, era de 107. O investigador pretende saber se noutra área e para a mesma população (indivíduos com idade  $\geq 16$  anos, consumidores de drogas e presos) e com base na seguinte amostra, terá razões para depreender que o QI é diferente de 107. Conclua com base nos dados:

99; 100; 90; 94; 135; 108; 107; 111; 119; 104; 127; 109; 117; 105; 125

**Exercícios**

- 39.** O departamento de pesquisa de uma empresa produtora de medicamentos, realizou uma experiência para verificar se a ingestão de um determinado produto aumenta o tempo de reacção a um estímulo. Para tal, seleccionou aleatoriamente 12 indivíduos e registou o tempo de reacção de cada um a esse estímulo, antes e depois de tomar o medicamento. Os resultados foram os seguintes:

Antes:	0.75	0.82	1.04	0.77	0.92	1.11	0.69	0.84	0.91	0.98	0.83	0.75
Depois:	0.84	0.78	1.15	0.81	0.95	1.08	0.82	0.96	0.95	0.83	0.91	0.81

Analise os dados e tire conclusões. Se tivesse admitido a normalidade, como procederia?

- 40.** Um grupo de 20 indivíduos hipertensos foi submetido, durante 30 dias, a um regime de dieta sem sal. Apresentam-se a seguir os valores da pressão sistólica para esses indivíduos:

Género	M	M	M	F	F	M	F	M	M	F
Antes	17	17.7	17.9	18.1	18.1	18.2	18.3	18.4	18.4	18.5
Depois	15.6	16.6	16.9	15.6	16.0	15.5	16.5	17.2	15.0	17.5
Género	F	F	M	M	F	M	M	F	F	F
Antes	18.5	18.6	18.7	18.8	18.9	19.2	19.3	19.5	19.8	20.1
Depois	15.9	16.2	17.5	15.8	17.2	17.3	17.8	16.0	16.9	17.5

Um especialista afirma que após um mês em regime de dieta sem sal, pelo menos 30% dos indivíduos apresenta uma diminuição da pressão sistólica superior a 10%. Averigüe se existem razões para duvidar da afirmação do especialista.

- 41.** Uma amostra de 15 doentes sofrendo de asma, participou numa experiência para estudar o efeito de um novo tratamento na função pulmonar. Uma das medidas registadas foi o volume de expiração (em litros) em 1 segundo, antes e depois do tratamento. Os resultados obtidos foram:

Doente	1	2	3	4	5	6	7	8
Antes	1.69	2.77	1.00	1.66	3.00	.85	1.42	2.82
Depois	1.69	2.22	3.07	3.35	3.00	2.74	3.61	5.14
Doente	9	10	11	12	13	14	15	
Antes	2.58	1.84	1.89	1.91	1.75	2.46	2.35	
Depois	2.44	4.17	2.42	2.94	3.04	4.62	4.42	

Com base nos dados poder-se-á concluir que o tratamento é eficaz no que diz respeito ao aumento da capacidade de expiração?

- 42.** Um casal de jogadores de bowling e estatísticos efectuou 4 jogos e obteve os seguintes resultados:

Pontuação do marido:	122	132	115	160
Pontuação da mulher:	147	120	183	154

- a) Determine a distribuição de probabilidade da estatística de *Wilcoxon*,  $T_n^+$ , para  $n = 4$ .  
 b) Use o resultado da alínea anterior para testar se a probabilidade de a pontuação da mulher ser superior à do marido é maior do que 0.5.

- 43.** Pretende-se investigar se o consumo de tabaco aumenta a actividade cardíaca. Para isso, registou-se o nº de batimentos cardíacos por minuto em 20 indivíduos, antes e depois de fumarem um cigarro de determinada marca. Os resultados obtidos foram os seguintes:

Antes	70	74	69	69	68	70	70	69	66	70
Depois	69	74	72	70	71	71	71	72	68	72
Antes	73	72	68	71	72	68	71	72	67	69
Depois	75	72	72	71	71	67	72	73	69	70

Teste, ao nível de significância  $\alpha = 1\%$ , se a probabilidade do nº de batimentos cardíacos depois de fumar ser superior ao nº de batimentos cardíacos antes de fumar é maior do que 0.5:

- a) Utilizando o teste dos sinais.  
 b) Utilizando o teste de *Wilcoxon*.

**Exercícios**

- 44.** Foram registados os valores da pressão diastólica de 38 homens seleccionados aleatoriamente, tendo-se obtido os seguintes dados que já se encontram ordenados:

64	65	68	70	71	72	73	75
76	77	78	78	79	80	80	81
82	82	83	83	84	85	85	86
86	87	88	88	89	89	90	91
92	93	93	94	95	97		

- a) Pensa-se que 25% dos homens tem uma pressão diastólica que não ultrapassa 78 mm Hg. Verifique se, perante a amostra anterior, existem razões para duvidar da hipótese anterior.
- b) Utilize um processo rápido para averiguar a simetria aproximada dos dados. Teste a hipótese de que a mediana é superior a 84.
- 45.** Foi recolhida a seguinte amostra aleatória, proveniente de uma população com mediana nula:

-1.85	-1.19	-.96	-.63	-.53	-.47	-.29	-.04	.05	.06
.19	.24	.32	.54	.79	1.02	1.28	1.39	1.49	2.06

Que pode concluir acerca da simetria da distribuição subjacente à população?

- 46.** Foi conduzida uma experiência com indivíduos adultos, padecendo de um certo tipo de atraso, em que se pretendia que respondessem a 62 questões. Cinco indivíduos receberam um tratamento novo, e outros cinco indivíduos, constituindo o grupo de controlo, receberam o tratamento convencional. Na tabela seguinte apresentam-se as percentagens de respostas correctas, para os dois grupos de indivíduos. Indicarão os dados evidência suficiente para concluir que o tratamento novo aumenta a percentagem de respostas correctas?

Tratamento novo	73	42	90	58	62
Grupo de controlo	50	23	68	40	45

- 47.** Considere o seguinte procedimento para testar a homogeneidade de duas populações independentes  $X$  e  $Y$ : considera-se uma amostra  $X_1, X_2, \dots, X_m$  proveniente da população  $X$ , com função distribuição  $F(\cdot)$  e outra amostra  $Y_1, Y_2, \dots, Y_n$  proveniente da população  $Y$ , com função distribuição  $G(\cdot)$ . A estatística de teste  $T$ , é a variável aleatória que se obtém somando as ordens dos elementos  $Y_i$ 's, quando se considera a amostra conjunta ordenada.

- a) Determine, sob a validade da hipótese nula

$$H_0: F(x) = G(x), \text{ para todo o } x,$$

a distribuição exacta da estatística de teste  $T$ , para o caso em que  $m = 3$  e  $n = 2$ .

- b) No tratamento de determinada doença, o doente recebe um de dois medicamentos, sendo posteriormente registado o nº de dias que a febre leva a baixar:

Medicamento A:	4.3	6.0	7.5
Medicamento B:	4.5	5.0	

Há evidência significativa de que o medicamento  $B$  é melhor do que o  $A$ ? Utilize o teste delineado na alínea a) ou, caso não tenha resolvido aquela alínea, utilize outro procedimento estatístico adequado.

- 48.** Pretende comparar-se duas culturas laboratoriais  $A$  e  $B$  quanto ao número de bactérias que nelas se desenvolveram. De cada uma das culturas foi retirada uma amostra, registando-se em seguida o número de bactérias por unidade de volume. Os resultados obtidos foram os seguintes:

Cultura A:	32	29	34	47	33	27		
Cultura B:	38	36	31	42	35	40	39	30

Utilize o teste de Mann-Whitney-Wilcoxon para testar, ao nível de significância  $\alpha = 5\%$ , se a contagem de bactérias nas duas populações é, ou não, a mesma.

## ANÁLISE DE DADOS

Ano Lectivo 2018/2019

### Exercícios

49. Um investigador planeou a seguinte experiência para estudar o efeito de uma inalação prolongada ao óxido de cádmio. Considerou 15 cobaias que expôs à inalação do óxido e pegou em 10 animais semelhantes, que constituíram o grupo de controlo. A variável medida foi o nível de hemoglobina, tendo-se obtido os seguintes resultados:

Animais expostos			Animais não expostos	
14.4	16.6	15.7	17.4	16.0
14.2	15.9	16.7	16.2	16.9
13.8	15.6	13.7	17.1	15.0
16.5	14.1	15.3	17.5	16.3
14.1	15.3	14.0	15.0	16.8

Teste se o nível de hemoglobina nas duas populações é, ou não, o mesmo.

50. Num estudo sobre anestésicos, verificou-se que a pressão diastólica de anestesiados com halotano ( $X$ ) e a pressão diastólica de anestesiados com morfina ( $Y$ ) foi:

x:	63.4	65.6	73.2	71.6	72.1	70.9	71.3
y:	61.6	73.9	75.2	60.9	73.6	70.2	

Teste se existe evidência suficiente para afirmar que a pressão diastólica de anestesiados com halotano é inferior à pressão diastólica de anestesiados com morfina.

51. Os cinco planetas conhecidos na Antiguidade podem ser divididos em dois grupos:

Grupo 1: Planetas mais distantes do Sol do que a Terra (Marte, Júpiter e Saturno)

Grupo 2: Planetas mais próximos do Sol do que a Terra (Mercúrio e Vénus)

Tomando como unidade a densidade da Terra, os astrónomos fizeram as seguintes medições das densidades planetárias:

Planeta	Grupo 2			Grupo 1		
	Mercúrio	Vénus	Terra	Marte	Júpiter	Saturno
Densidade	0.68	0.94	1	0.71	0.24	0.12

- a) Determine a distribuição de probabilidade da estatística de *Mann-Whitney* para  $m = 2$  e  $n = 3$ . (Sugestão: Construa uma tabela de duas entradas com os valores que  $W_{2,3}$  pode tomar para cada par de ordens possíveis das duas observações da amostra de dimensão 2.)
- b) Utilize a alínea anterior para testar a hipótese da igualdade de distribuições da densidade dos planetas do grupo 1 e do grupo 2.

52. Foram registados os tempos de vida (em centenas de horas) de um certo tipo de lâmpadas, produzidas por duas fábricas  $A$  e  $B$ , tendo-se obtido os seguintes resultados:

A:	3.7	2.8	7.1	8.4	6.2	2.7				
B:	8.9	6.7	9.1	7.4	8.1	6.8	8.2	6.4	7.7	6.9

Teste, ao nível de significância de 5%, se as lâmpadas da fábrica  $B$  apresentam um tempo de vida superior ao das lâmpadas produzidas pela fábrica  $A$ .

53. Num estudo sobre a avaliação escolar, foram inquiridos 21 estudantes (10 rapazes e 11 raparigas) sobre a importância que atribuem à componente de avaliação contínua, indicada numa escala de 1 a 10. Os resultados obtidos foram os seguintes:

Género	Resposta	Género	Resposta	Género	Resposta
M	1	F	6	M	8
M	2	F	6	M	9
F	3	F	6	M	9
F	3	F	6	M	9
M	4	F	7	F	10
M	5	M	7	F	10
F	6	M	8	F	10

Verifique se existem razões para duvidar que este aspecto da avaliação seja igualmente importante nos dois grupos.

## ANÁLISE DE DADOS

Ano Lectivo 2018/2019

### Exercícios

54. Seleccionaram-se, aleatoriamente, 15 baterias de um lote produzido na fábrica A e outras tantas de um lote produzido na fábrica B. As 30 baterias foram postas em funcionamento simultaneamente. A sequência seguinte mostra a ordem pela qual falharam as 30 baterias colocadas em teste:

A B B B A B A A B B B B A B A B B B B A A B A A A B A A A A

Determine se há evidência suficiente para concluir que o tempo de vida das baterias da fábrica A é superior ao tempo de vida das baterias da fábrica B. Use  $\alpha = .05$ .

55. Num inquérito sobre a resistência individual ao esforço físico, submeteram-se 3 grupos de indivíduos a 3 exercícios diferentes (bicicleta, passadeira rolante e prova dos degraus), medindo-se o tempo (em minutos) até ao consumo máximo do oxigénio. Os resultados foram os seguintes:

<u>Bicicleta:</u>	7.5	8.7	9.2	9.8	10.9	11.1	11.2	12.8	13.5	15.2
<u>Passadeira:</u>	8.7	13.2	13.8	14.7	15.5	16.2	16.2	17.8	18.1	20.0
<u>Degraus:</u>	15.9	16.1	16.2	17.8	18.9	19.3	20.1	21.0	22.6	24.6

Haverá diferença entre os grupos?

56. Foram testados três insecticidas de diferentes marcas A, B, e C e registou-se o número de insectos mortos por cada um deles.

Insecticida A:	72	65	67	75	62
Insecticida B:	67	75	62	68	
Insecticida C:	68	67	61	75	

Utilizando um teste conveniente verifique se, ao nível de significância  $\alpha = 1\%$ , se existe evidência para concluir que os três insecticidas não têm o mesmo efeito mortífero.

57. Num estudo destinado a comparar o peso de indivíduos com diferentes alturas, foram seleccionados aleatoriamente indivíduos pertencentes a cada uma das seguintes classes e registaram-se os seus pesos:

Altura (cm)	Peso (Kg)
[150,160)	50, 54, 52, 55, 52, 51
[160,170)	55, 60, 62, 64, 61, 60
[170,180)	65, 68, 67, 66, 67
[180,190)	72, 75, 80, 79, 82, 81, 74

Teste se a distribuição do peso é idêntica nas 4 classes consideradas.

58. Três tipos diferentes de magnetron (componente principal dos fornos micro-ondas) foram sujeitos a condições de "stress", com o objectivo de testar o tempo até à falha. Os resultados seguintes indicam o nº de horas de funcionamento de várias componentes, de cada um dos tipos, colocadas em teste:

Tipo A	Tipo B	Tipo C
36	49	71
48	33	31
5	60	140
67	2	59
53	55	42

Teste a homogeneidade dos três tipos de magnetron, no que diz respeito ao tempo até à falha.

59. Realizou-se um estudo destinado a comparar três analgésicos A, B e C no que respeita ao tempo de eficácia. Assim, seleccionaram-se aleatoriamente 3 grupos de indivíduos sofrendo de cefaleias, aos quais foi administrado um dos três analgésicos, tendo-se obtido os seguintes resultados (em minutos):

A:	18.7	19.2	23.4	27.2	29.1
B:	19.7	24.6	28.2	33.2	
C:	19.6	19.8	21.4	22.1	

Teste a hipótese de os três analgésicos terem tempos de eficácia idênticos.

**Exercícios**

- 60.** Realizou-se uma experiência laboratorial com 3 grupos de cobaias, para estudar o efeito de dois medicamentos no tempo de reacção a um certo estímulo. Antes da aplicação do estímulo, os animais no grupo I foram tratados com o medicamento A, os do grupo II foram tratados com o medicamento B e os animais do grupo III receberam um placebo. Os tempos de reacção (em segundos) foram os seguintes:

Grupo I:	17	20	40	31	35
Grupo II :	8	7	9	8	
Grupo III:	2	5	4	3	

Poder-se-á concluir que existe diferença significativa entre os três grupos, no que diz respeito aos tempos de reacção?

- 61.** Para testar a independência entre as alturas de marido e mulher, seleccionaram-se aleatoriamente 12 casais e registaram-se as respectivas alturas (em metros):

Marido	1.57	1.85	1.73	1.88	1.75	1.70	1.98	1.78	1.71	1.69	1.67	1.72
Mulher	1.60	1.53	1.68	1.73	1.65	1.63	1.83	1.70	1.61	1.55	1.57	1.50

Teste a hipótese referida, utilizando o coeficiente de correlação ordinal de *Spearman*.

- 62.** Um psicólogo recolheu uma amostra de 12 pais e filhos e obteve os seguintes resultados relativamente aos respectivos quocientes de inteligência (Q.I.):

Pai	105	120	148	94	110	168	116	132	143	128	124	135
Filho	115	132	154	100	108	149	98	125	153	130	135	160

Conclua se existe dependência entre o Q.I. de pai e filho utilizando o coeficiente de correlação ordinal de *Spearman*.

- 63.** A tabela seguinte apresenta as classificações (do melhor (1), para o pior (8)) de 8 marcas de vinho branco feitas por 2 provadores:

Vinho	A	B	C	D	E	F	G	H
Provador 1	2	3	1	4	6	7	8	5
Provador 2	2	3	1	5	4	6	7	8

Verifique se há diferença entre os vinhos (existe associação entre as classificações).

- 64.** Pediu-se a 100 mulheres e a 100 homens que usassem uma nova pasta dentífrica, dizendo se gostavam ou não do seu sabor. 32 homens e 26 mulheres afirmaram que não gostavam. Acha que este facto indica uma diferença de preferência entre homens e mulheres?

- 65.** Foi dado um teste de Matemática a 150 alunos escolhidos aleatoriamente. A classificação podia ser uma de três: A, B ou C. 40 crianças obtiveram classificação A, 80 classificação B e as restantes C. Das crianças que obtiveram A, 35% eram canhotas, das que obtiveram B, 25% eram canhotas e nas que obtiveram C a percentagem de canhotas era 10%. Tendo em consideração os dados anteriores, poderá concluir da existência de associação entre a nota a Matemática e o facto de um aluno ser, ou não, canhoto?

- 66.** Lançaram-se 3 dados, cada um 360 vezes, tendo-se verificado os seguintes resultados:

Face	dado 1	dado 2	dado 3
1	50	62	38
2	48	55	60
3	69	61	64
4	45	54	58
5	71	78	73
6	77	50	67
Total	360	360	360

Será que os dados são “iguais”?

## ANÁLISE DE DADOS

Ano Lectivo 2018/2019

### Exercícios

67. Diga, considerando  $\alpha = 5\%$ , se o nível de absentismo e o género são características independentes, considerando os seguintes dados obtidos em certa zona industrial:

Género	Nível de absentismo		
	nulo	reduzido	forte
F	40	17	3
M	35	11	4

68. Pretende estudar-se se há relação entre a poluição atmosférica e as doenças pulmonares. Para tal, escolheram-se duas cidades com problemas de poluição e uma área rural pouco poluída. Estudaram-se 400 adultos em cada área, tendo-se obtido os seguintes resultados:

Área	nº indivíduos com doenças pulmonares	nº indivíduos sem doenças pulmonares
Cidade A	34	366
Cidade B	42	358
Campo	21	379

- a) Haverá diferença entre as 3 áreas, quanto à incidência de doenças pulmonares?  
b) Como procederia se, a partir destes dados, pretendesse comparar o meio citadino com o campo, relativamente à ocorrência de doenças pulmonares?
69. Pretende-se estudar o efeito de dois tipos de fertilizantes para plantas de interior. Cada um dos fertilizantes foi aplicado em 150 plantas durante um mês, registando-se os resultados obtidos:

Fertilizante	sem alteração	melhoria ligeira	melhoria significativa
A	15	60	75
B	18	42	90

Haverá evidência para afirmar que os dois fertilizantes produzem efeitos diferentes?

70. A relação entre o tabaco e o cancro no pulmão é uma suspeita de longa data, que preocupa todos os epidemiologistas. Numa campanha de prevenção contra o tabagismo, pretende-se elucidar a população quanto ao risco de contrair cancro do pulmão. Assim, realizou-se uma sondagem a 172 indivíduos que conduziu aos seguintes resultados:

	# Cancros no pulmão	# Indivíduos saudáveis
Não fumadores	3	14
Fumadores moderados	40	63
Fumadores viciados	43	9

Utilize um teste apropriado para análise destes dados e comente os resultados obtidos.

**FIM**