

Anexo I – Programa de Bioestatística



BIOESTATÍSTICA

PROGRAMA 2015/16

1. Introdução à Bioestatística e Análise Exploratória de Dados

- Estatística na Biologia.
- Tipos de dados.
 - i. Tipos de variáveis;
 - ii. Escalas de medida.
- Análise Exploratória de Dados.
 - i. Medição de variáveis;
 - ii. Organização de dados em tabelas e gráficos;
 - iii. Medidas de localização, dispersão e forma.

2. Distribuições de Probabilidade

- Variável aleatória, função de distribuição.
- Variáveis aleatórias discretas e contínuas: função massa de probabilidade e função densidade de probabilidade.
- Parâmetros de uma distribuição: valor esperado, variância, momentos, moda, mediana e quantis.
- Distribuições com especial importância: binomial, Poisson, normal.

3. Estatística Inferencial

- Amostragem: noções gerais de amostragem.
 - i. Distribuições de amostragem;
 - ii. Teorema Limite Central.
- Estimação pontual: estimador e estimativa.
- Estimação intervalar: noções gerais sobre intervalos de confiança.
 - i. Intervalos de confiança em populações normais: médias e variâncias;
 - ii. Intervalo de confiança para a proporção.
- Testes de hipóteses paramétricos.
 - i. Noções gerais sobre teste de hipóteses: tipos de hipóteses, erro de 1ª e de 2ª espécie, potência do teste e valor p;
 - ii. Testes de hipóteses paramétricos em populações normais:
 - para a média,



BIOESTATÍSTICA

PROGRAMA 2015/16

- para a variância,
- para a comparação de médias,
- para a comparação de variâncias.
- iii. Teste para a proporção e teste para a comparação de proporções.
- o Validação de pressupostos.
 - i. QQ-plots;
 - ii. Testes de ajustamento à uma distribuição normal: teste de Kolmogorov-Smirnov com correção de Lilliefors, teste de Shapiro Wilk;
 - iii. Teste para a igualdade de variâncias.

4. Análise de Variância

- o Planeamento de experiências com um factor.
 - i. Modelo de efeitos fixos: Pressupostos do modelo, Interpretação dos resultados, Comparações múltiplas;
 - ii. Planeamento de experiências com dois factores, com dois ou mais níveis cada um (Blocos e Observações repetidas)

5. Correlação e Regressão

- o Análise da correlação.
 - i. Medidas de associação variáveis não nominais;
 - Correlação de Pearson,
 - Correlação de Spearman.
 - ii. Diagramas de dispersão;
 - iii. Transformações de dados;
 - iv. Testes para coeficientes de correlação.
- o Regressão linear.
 - i. Apresentação e interpretação do modelo;
 - ii. Estimação pontual dos parâmetros da recta de regressão;
 - iii. Predição de uma observação futura.



BIOESTATÍSTICA

PROGRAMA 2015/16

- iv. Inferência sobre os parâmetros do modelo: intervalos de confiança e testes de hipótese;
- v. O coeficiente de determinação como indicador da qualidade do ajustamento;
- vi. Validação de pressupostos.

6. Análise de dados qualitativos

- i. Tabelas de contingência;
- ii. Teste do χ^2 para a independência;
- iii. Teste de ajustamento do χ^2 .

Aveiro, 4 de fevereiro de 2016

A responsável pela Unidade Curricular,

Anexo II – Teste aplicado aos alunos



Mini Teste 1

Q1 Considere o seguinte conjunto de dados de nível concentração sérica (em g / ml) de Gentamicina no sangue recolhido a partir de uma amostra casual de 9 ovelhas:

concentração sérica (g/ml)	33	23	31	32	34	25	34	27	25
----------------------------	----	----	----	----	----	----	----	----	----

Tem-se que:

- ☐ a moda é 2 pois é a frequência que se observa mais vezes (níveis 25 e 34).
- ☐ a amplitude interquartil é igual a 3.
- ☐ mais de 75% das ovelhas apresentam concentração sérica de Gentamicina superior a 33.5.
- ☐ pelo menos 50% das ovelhas apresentam concentração sérica de Gentamicina no sangue acima da média.

Q2 Analisaram-se 200 amostras de água recolhidas numa certa pateira tendo sido registado a concentração de nutrientes. Dos dados conclui-se que:

$$\begin{aligned}\text{Percentil de ordem 25} &= 0.4gr/cm^3. \\ 3^o\text{Quartil} = \text{Média} &= 0.5gr/cm^3\end{aligned}$$

Relativamente a este conjunto de dados tem-se que:

- ☐ o comprimento da caixa de bigodes (sem os ditos "bigodes") é igual a 0.5.
- ☐ a se observarem, na caixa de bigodes, níveis atípicos de concentração de nutrientes eles corresponderão a amostras com níveis de concentração superiores a 0.55.
- ☐ a percentagem de observações superiores ou iguais à média é não inferior a 25%.
- ☐ a mediana dos dados será necessariamente um valor superior a 0.4 e inferior a 0.5.

Q3 Considere um estudo para avaliar a relação entre a obesidade e a procriação em ratos da linhagem *wistar*. Nesse estudo analisou-se o número de ratinhos gerados por ninhada (variável 1), o peso da progenitora (variável 2), a idade (em dias) da progenitora (variável 3), o estado de stress (com os níveis baixo, médio e alto) (variável 4), entre outras variáveis.

- ☐ A variável 2 é classificada como quantitativa, contínua, numa escala positiva.
- ☐ A variável 3 é classificada como qualitativa, discreta, numa escala numérica.
- ☐ A variável 4 é classificada como nominal numa escala discreta.
- ☐ A variável 1 é classificada como quantitativa, discreta, numa escala de razão.



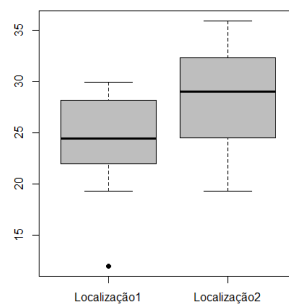
Q4 Num estudo ecológico realizado em rios portugueses mediu-se, durante 45 dias não consecutivos casualmente seleccionados, a concentração de estrôncio (mg/ml) no curso de água do rio Vouga. Os dados foram organizados numa tabela de frequências:

concentração	36	37	38	39	40	41	42	45
Nº de dias	3	11	10	8	9	1	2	1

Relativamente à caixa de bigodes associada a este conjunto de dados:

- ☐ não existem observações atípicas e o "bigode" que se observa do lado direito tem comprimento igual a 5.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 3.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 2.
- ☐ existe uma observação atípica e o "bigode" que se observa do lado direito tem comprimento igual a 4.5.

Q5 É sabido que a qualidade da água não se mantém constante com o tempo e varia de zona para zona. Em duas zonas em estudo (Localização 1 e Localização 2) foram recolhidas várias amostras e medido o índice de salinidade da água em cada amostra. Com base nos valores observados foram obtidas no R as seguintes caixas de bigodes comparativas dos índices de salinidade amostrados nas duas zonas:



Comente as caixas de bigodes, interpretando no contexto do problema o que observa quanto à localização central, dispersão e assimetria da distribuição do índice de salinidade entre as duas zonas;

☐0 ☐1 ☐2 ☐3 ☐4



Mini Teste 2

Q6 A probabilidade de uma ovelha adulta apresentar concentração sérica de Gentamicina elevada no sangue é 0.309. Qual a probabilidade de, numa amostra de 10 ovelhas selecionadas ao acaso, 6 ou mais apresentarem concentração sérica de Gentamicina elevada no sangue?

- ☐ $\approx 6 \times 0.309$
☐ ≈ 0.054
☐ ≈ 0.987
☐ ≈ 0.042

Q7 O nível Y de concentração de um dado composto é bem modelado por uma distribuição de probabilidade $N(\mu = 30, \sigma^2 = 4)$. Então

- ☐ $E(2Y) = 60$ e $Var(2Y) = 8$
☐ $E(2Y) = 30$ e $Var(2Y) = 2$
☐ $E(2Y) = 60$ e $Var(2Y) = 16$
☐ $E(2Y) = 30$ e $Var(2Y) = 4$

Q8 A abundância (X) da espécie *Columbia livia* (pombo-comum) por $10m^2$ numa certa zona da cidade de Aveiro segue uma distribuição de Poisson de parâmetro 4.5. Logo, nessa zona

- ☐ esperam-se encontrar 9 pombos-comum por cada $20m^2$ sendo que $P(X = 0) \approx 0.0111$.
☐ esperam-se encontrar 9 pombos-comum por cada $100m^2$ sendo que $Var(X) = 4.5$.
☐ será pouco provável encontrar mais de 10 pombos-comum por m^2 sendo o valor dessa probabilidade ≈ 0.007 .
☐ com área total de $500m^2$ esperam um total de 500 pombos-comum.

Q9 Num estudo para avaliar características de ratos da linhagem *wistar* sabe-se que o peso X ao desmame de uma cria é bem modelado por uma distribuição normal de média 40g e desvio padrão igual a 5g. A probabilidade de uma ninhada de 5 crias pesar ao desmame mais 220g é

- ☐ ≈ 0.44
☐ ≈ 0.21
☐ ≈ 0.04
☐ ≈ 0.20

Q10 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso de água do rio Vouga, tem média igual a 40mg/ml e desvio padrão igual a 3.2mg/ml. Nestas condições, em 100 unidades causais de 1ml cada, recolhidas no rio Vouga, a probabilidade da média amostral das 100 unidades variar entre 39.0 e 40.5 é aproximadamente igual a

- ☐ 0.32
☐ 0.77
☐ 0.94
☐ 0.06



Q11 Considere as seguintes afirmações:

- A. O 1º quartil de uma distribuição $N(\mu = 3, \sigma^2 = 4)$ é aproximadamente 1.65 .
- B. Uma dada amostra fornece $[10, 20]$ como intervalo a 90% de confiança para μ de uma população normal. Se aumentássemos a confiança para 95%, então o valor do seu limite inferior será superior a 10.

Tem-se

- ☐ A é falsa e B é verdadeira.
- ☐ A é verdadeira e B é falsa.
- ☐ A é verdadeira e B é verdadeira.
- ☐ A é falsa e B é falsa.

Q12 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso da água do rio, é bem modelado por uma distribuição normal. Com o objetivo de comparar a concentração de estrôncio no rio Vouga e no rio Tejo foram recolhidas 50 amostras de cada rio. Das amostras do rio Vouga registou-se uma média de 38mg/ml com um desvio padrão corrigido de 1.2mg/ml. Da amostra do rio Tejo registou-se uma média de 39.3 mg/ml com um desvio padrão corrigido de 1mg/ml. Um intervalo de confiança a 95% para a diferença de concentrações médias de estrôncio entre os dois rios é:

- ☐ $[-1.873, -0.8672]$
- ☐ $[-1.733, -0.8670]$
- ☐ $[-1.738, -0.8616]$
- ☐ $[-1.739, -0.8614]$

Q13 Um estudo avaliou a obesidade e a procriação de gatos domésticos. Numa amostra de 110 ninhadas de gatos domésticos, verificou-se que 70% das crias nascidas vivas tinham peso abaixo do peso desejado. Assim, uma estimativa intervalar 90% de confiança para a proporção de crias nascidas vivas com peso abaixo do desejado é:

- ☐ $[0.614, 0.786]$.
- ☐ $[0.561, 0.712]$.
- ☐ $[0.628, 0.772]$.
- ☐ $[0.543, 0.857]$.



Mini Teste 3

Q14 Os comprimentos das caudas de ratos do campo são bem modelados por uma distribuição normal de variância $\sigma^2 = 0.25$ (parâmetro populacional). Quantos ratos do campos devem ser selecionado para garantir que o intervalo, a 95% confiança para comprimento médio da cauda, construído à custa dessa amostra tem amplitude igual a 0.5?

- ☐ 24
- ☐ 15
- ☐ 5
- ☐ 40

Q15 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml) no curso da água do rio Tejo é bem modelada uma distribuição normal de variância desconhecida σ^2 . Foram recolhidas 51 amostras do rio Tejo tendo-se registado uma média de 39.3 mg/ml de concentração de estrôncio com um desvio padrão corrigido de 1.2mg/ml. Tendo em conta os seguintes resultados obtidos do R e com base na amostra, um intervalo a 95% de confiança para σ^2 é:

```
> qchisq(0.95,50)      > qchisq(0.05,50)      > qchisq(0.95,49)      > qchisq(0.05,49)
[1] 67.50481            [1] 34.76425            [1] 66.33865            [1] 33.93031
> qchisq(0.975,50)     > qchisq(0.025,50)     > qchisq(0.975,49)     > qchisq(0.025,49)
[1] 71.4202             [1] 32.35736            [1] 70.22241            [1] 31.55492
```

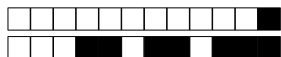
- ☐ [1.025, 2.282]
- ☐ [0.988, 2.180]
- ☐ [1.008, 2.225]
- ☐ [1.005, 2.236]

Q16 Com o objetivo de investigar a diversidade e abundância de aves em ambiente urbano, foram realizados censos de avifauna por pontos durante a época de nidificação em várias zonas da cidades de Aveiro. Numa dessas zonas, das 560 aves omnívoras observadas, contabilizaram-se 302 da espécie *Columbia livia* (pombo-comum). Os dados permitem então concluir, ao nível de significância 10%, que a percentagem de pombos-comuns naquela zona de Aveiro é significativamente superior a 50% já que a amostra conduziu ao valor p, do teste de hipótese em causa,

- ☐ ≈ 0.031
- ☐ ≈ 0.539
- ☐ ≈ 0.000
- ☐ ≈ 0.063

Q17 Num estudo pretende-se estudar características de ratos da linhagem *wistar*. Uma dessas características corresponde ao seu peso ao desmame, a qual se assume seguir uma distribuição normal. A experiência contou com 50 ninhadas, com pelo menos 3 crias, observadas em laboratório. De cada ninhada selecionou-se uma cria ao acaso e registou-se o seu peso ao desmame. Com base na amostra dos 50 pesos obteve-se o intervalo [36, 43] como intervalo a 95% de confiança para o peso médio de uma cria ao desmame. Nestas circunstâncias, o estudo permite concluir que o peso médio de uma cria ao desmame é significativamente

- ☐ superior ou igual a 36, ao nível de confiança de 95%
- ☐ inferior ou igual a 42, ao nível de significância de 5%
- ☐ diferente de 42, ao nível de significância de 5%
- ☐ diferente de 35, ao nível de significância de 5%



Q18 Durante três meses submeteram-se 50 golden retriever a um tipo de dieta, tendo-se observado que a média das diferenças (inicial – final) do índice de massa corporal canino (IMCC) pré e pós-dieta é de 1.8 e o desvio padrão corrigido das diferenças é 0.3. Assuma que os dados são normalmente distribuídos. Averigüe, ao nível de significância de 1%, se o IMCC médio pré-dieta é significativamente superior ao IMCC médio pós-dieta.

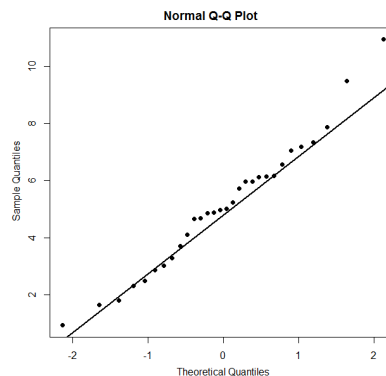
- ☐ Como valor p é inferior a 0.01, decide-se pela rejeição de H_0
- ☐ Como valor p é superior a 0.01, decide-se pela não rejeição de H_0
- ☐ Como valor p é inferior a 0.01, conclui-se que IMCC médio pré-dieta não é significativamente superior ao IMCC médio pós-dieta
- ☐ Como valor p é superior a 0.01, conclui-se que IMCC médio pré-dieta é significativamente superior ao IMCC médio pós-dieta

Q19 Foi recolhida uma amostra de valores correspondentes aos comprimentos dos corpos de 30 camarões de aquacultura selecionados casualmente. Assumindo que os dados são normais foi efetuado um teste- t para averiguar se o comprimento médio de corpos de camarões de aquacultura é significativamente inferior a 5.50. Sabe-se que o valor observado da estatística de teste T é $t_{obs} = -1.3$. Indique o valor p obtido. Tem-se

- ☐ valor $p = P(T \leq 5.50)/2 \approx 0.4999$
- ☐ valor $p = P(T \leq t_{obs}) \approx 0.10192$
- ☐ valor $p = P(T \leq 5.50) \approx 0.999997$
- ☐ valor $p = 2 * P(T \leq t_{obs}) \approx 0.20384$

Q20 Foi recolhida uma amostra de valores correspondentes aos comprimentos dos corpos de 30 camarões de aquacultura selecionados casualmente. Com esses valores foi construído o seguinte QQ-plot. Das seguintes afirmações, selecione a que não está correta.

- ☐ O QQ-plot da normal é uma ferramenta gráfica que permite averiguar se os dados provêm de uma população normal.
- ☐ Um QQ-plot da normal não é um teste de hipóteses associado a um dado nível de significância.
- ☐ É de crer que os dados provêm de uma população normal pois todos os pontos no QQ-plot se encontram relativamente próximos da reta.
- ☐ É de crer que os dados provêm de um modelo definido por uma reta.





Q21 De estudos ecológicos realizados em rios portugueses sabe-se que a concentração de estrôncio (mg/ml), no curso da água dos rios Sado e Guadiana, é bem modelado por distribuições normais. Foram recolhidas 50 amostras de cada rio. Das amostras do rio Sado registou-se uma concentração média de estrôncio de 38.1mg/ml com um desvio padrão corrigido de 1.2mg/ml. Da amostra do rio Guadiana registou-se uma média de 37 mg/ml com um desvio padrão corrigido de 1.1mg/ml. Averigue, ao nível de significância de 1% se a concentração média de estrôncio no rio Sado é superior à do rio Guadiana. Na sua resposta deverá: especificar as hipóteses H_0 e H_1 (0.7), averiguar a homogeneidade de variâncias (0.5), indicar o valor observado da estatística de teste (0.3), indicar o valor p do teste (0.4), decidir e concluir (0.6).

☐ 0 ☐ 0.7 ☐ 0.5 ☐ 0.3 ☐ 0.4 ☐ 0.6 ☐ 2.5



Mini Teste 4

Q22 Para averiguar o efeito da idade nos níveis de concentração sérica de Gentamicina no sangue, em ovelhas sujeitas àquele antibiótico, planeou-se a seguinte experiência envolvendo 18 ovelhas: 6 ovelhas do grupo etário 1 (idade entre 2 e 4 anos), 6 ovelhas do grupo etário 2 (idade entre 5 e 7 anos), e 6 ovelhas do grupo etário 3 (idade superior a 8 anos). A cada ovelha foi administrado o referido antibiótico na dose de 10 mg/kg de peso corporal e 2h depois foram recolhidos os níveis de concentração sérica (em g/ml) de Gentamicina no sangue. Com vista a efectuar uma ANOVA paramétrica foi construído o seguinte quadro resumo (incompleto):

Fonte de variação	Soma dos quadrados	Graus de liberdade	Media dos quadrados	Valor observado da estatística de teste	valor p
ENTRE GRUPOS	(a)	(c)	188.6	(g)	(h)
DENTRO DOS GRUPOS	265	(d)	(f)		
TOTAL	(b)	(e)			

Nas condições dadas tem-se que:

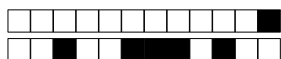
- ☐ (b) = 830.8 ; (c) = 3 ; (g) \approx 18.9.
- ☐ (b) = 359.3 ; (d) = 15 ; (g) \approx 10.7.
- ☐ (a) = 94.3 ; (f) = 17.7 ; (h) \approx 0.0013.
- ☐ (a) = 377.2 ; (d) = 15 ; (h) \approx 0.0013.

Q23 O planeamento considerado na questão **Q22** corresponde a uma

- ☐ ANOVA de efeitos fixos com 1 fator com 6 grupos.
- ☐ ANOVA de efeitos fixos a 1 fator e medidas repetidas.
- ☐ ANOVA de efeitos aleatórios com 1 fator.
- ☐ ANOVA de efeitos fixos com 1 fator com 3 grupos.

Q24 Para que a ANOVA paramétrica considerada na questão **Q22** seja válida, um dos pressupostos que deverá ser verificado é:

- ☐ a existência de homogeneidade de médias entre os grupos etários.
- ☐ que os dados dentro de cada grupo provenham de uma distribuição normal.
- ☐ a existência de homogeneidade de variância entre as ovelhas.
- ☐ a existência de igualdade de médias entre os grupos etários.



Q25 Foi aplicado o teste de Tukey de comparação múltipla aos dados considerados na questão **Q22**. Com recurso ao R foi obtido o seguinte resultado:

```
Tukey multiple comparisons of means
95% family-wise confidence level

            diff            lwr            upr            p adj
grupo etario 2-grupo etario 1 -8.666667 -14.969956 -2.363377 0.0073786
grupo etario 3-grupo etario 1 -10.500000 -16.803290 -4.196710 0.0016308
grupo etario 3-grupo etario 2  -1.833333  -8.136623  4.469956 0.7350194
```

Os resultados

- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue não são significativamente diferentes entre os grupos etários 1 e 2.
- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue são significativamente diferentes entre os grupos etários 2 e 3.
- ☐ não são válidos pois não é possível aplicar o teste de Tukey a este conjunto de dados.
- ☐ permitem concluir que, ao 5% de significância, os níveis médios de concentração sérica de Gentamicina no sangue são significativamente diferentes entre os grupos etários 1 e 3.

Q26 São conhecidos 3 tipos diferentes de habitat de porco preto ibérico (*Sus scrofa mediterraneus*). Pretende-se investigar se o habitat afeta o peso médio (μ) dos porcos pretos ibéricos de uma determinada faixa etária. Para tal foram recolhidos, em cada habitat, os pesos de 5 porcos pretos ibéricos.

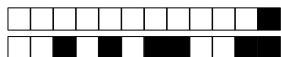
$h1$	$h2$	$h3$
45	72	65
53	75	61
57	85	51
48	81	55
60	74	63

Assuma válidos os pressupostos da ANOVA paramétrica. Neste planeamento, avaliar a questão de investigação corresponde a testar:

- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = 0$ vs $H_1 : \mu_{hi} \neq 0$ para algum habitat hi .
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = \mu$ vs $H_1 : \mu_{hi} \neq 0$ para algum habitat hi .
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = \mu$ vs $H_1 : \mu_{hi} \neq \mu_{hj}$ para algum par de habitats hi, hj ($i \neq j$).
- ☐ $H_0 : \mu_{h1} = \mu_{h2} = \mu_{h3} = 0$ vs $H_1 : \mu_{hi} \neq \mu_{hj}$ para algum par de habitats hi, hj ($i \neq j$).

Q27 Realizando o teste estatístico referido na alínea anterior, com base no conjunto de dados indicado, tem-se que o valor observado da estatística de teste é

- ☐ ≈ 5.8195
- ☐ ≈ 5.827
- ☐ ≈ 0.00006
- ☐ ≈ 24.47



Q28 Um investigador pretende averiguar se a concentração média de cálcio no sangue depende da realização um dado tratamento hormonal e do sexo. Para tal, foram selecionados casualmente 20 machos e 20 fêmeas de uma população de mamíferos. Dos 40 mamíferos foram selecionados casualmente 20 (10 machos e 10 fêmeas) aos quais foi administrado o tratamento hormonal. Aos restantes (10 machos e 10 fêmeas) foi administrado um tratamento placebo. Ao fim de um mês foi registado a concentração de cálcio dos 40 mamíferos em observação. Assuma a validade dos pressupostos da ANOVA paramétrica. Com recurso ao R obteve-se o seguinte resultado:

```
          Df Sum Sq Mean Sq F value Pr(>F)
tratamento  1  0.1796  0.17956    7.110  0.0114
sexo         1  0.1323  0.13225    5.237  0.0281
tratamento:sexo  1  0.0314  0.03136    1.242  0.2725
Residuals   36  0.9092  0.02526
```

Assim, podemos concluir que, ao nível de significância de 5%, o tratamento hormonal afeta significativamente a concentração média de cálcio no sangue de um mamífero já que o valor p associado ao teste é

- ☐ 0.0314
- ☐ 0.2725
- ☐ 0.1796
- ☐ 0.0114

Q29 Relativamente à alínea anterior o valor 0.2725 que surge nos resultados obtidos do R significa que, ao nível de significância de 5%,

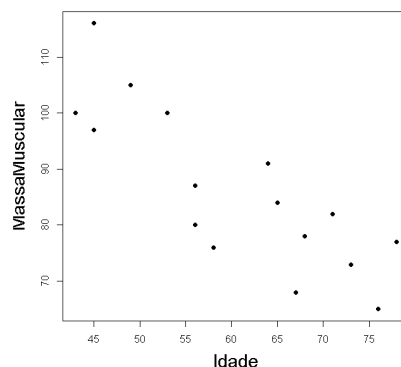
- ☐ existe uma concentração média de cálcio significativamente diferente entre o sexo e o tratamento hormonal.
- ☐ não existe interação significativa entre o tratamento hormonal e o sexo na concentração média de cálcio no sangue.
- ☐ não existe evidência estatística do sexo afetar significativamente o tratamento hormonal.
- ☐ não se rejeita a hipótese nula de igualdade de médias.



Mini Teste 5

Q30 Para averiguar a existência de relação entre a massa muscular de um adulto com a sua idade, um nutricionista recolheu uma amostra de 18 indivíduos, com idade entre 40 e 79 anos, e observou em cada um deles a idade (X) e a massa muscular (Y). Com os dados recolhidos obteve o seguinte diagrama de dispersão. Logo, é possível concluir que

- ☐ um adulto mais velho tenderá a ter menor massa muscular.
- ☐ existe uma associação do tipo linear entre os 18 indivíduos e cada uma das variáveis (X e Y).
- ☐ o valor do coeficiente de correlação amostral de Spearman entre X e Y deve ser muito próximo de zero.
- ☐ o coeficiente de correlação amostral de Pearson entre X e Y deve corresponder a um valor negativo indicando que os adultos mais novos tendem a ter menor massa muscular.



Q31 Tomando a amostra considerada na questão **Q30**, e usando o coeficiente de correlação de Pearson, foi testado se as duas variáveis X (idade) e Y (massa muscular) estão correlacionadas. Foi obtido o valor p do teste igual a 1.5×10^{-5} . Podemos então concluir que

- ☐ a idade e a massa muscular estão relacionados significativamente por uma função linear.
- ☐ a massa muscular e a idade estão relacionadas por uma constante igual a 1.5×10^{-5} .
- ☐ o coeficiente de correlação de Pearson entre a idade e a massa muscular não é significativamente diferente de zero.
- ☐ existe uma relação de independência estatisticamente significativa entre a idade e a massa muscular.

Q32

O nutricionista propôs-se realizar uma análise de resíduos para verificar os pressupostos de realização de uma análise de regressão linear simples. Assim, usando os resíduos (erros) deverá verificar que

- ☐ o declive da reta de regressão a ajustar aos dados é nulo.
- ☐ os dados não provêm de uma população normalmente distribuída.
- ☐ os erros têm variância constante.
- ☐ os erros têm média positiva.

Q33 Um investigador pretende avaliar se uma dada balança está bem calibrada. Considerou 9 pesagens conhecidas (X) e determinou o respetivo peso atribuído pela balança (Y). Os dados foram:

X	2.0	2.0	2.0	4.0	4.0	4.0	6.0	6.0	6.0
Y	2.1	1.8	1.9	4.5	4.2	4.0	6.2	6.0	6.5

A equação da reta de regressão estimada com base nos dados é

- ☐ $y = 0.202 + 0.919x$
- ☐ $y = -0.167 + 1.075x$.
- ☐ $y = 0.919 + 0.202x$.
- ☐ $y = 1.075 - 0.167x$.



Q34 Tomando os dados considerados na questão **Q33**, o investigador verificou que a relação entre o peso real e o peso atribuído pela balança é bem modelada por uma regressão linear simples já que o valor obtido para o coeficiente de determinação é elevado e aproximadamente igual a:

- ☐ 1.075.
- ☐ 0.988.
- ☐ 0.994.
- ☐ 23.72.

Q35 Tomando os dados considerados na questão **Q33**, foi realizado o teste ao declive do modelo de regressão para avaliar se é positivo, ao nível de significância de 5%. Tal corresponde a ter:

- (A) $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$
- (B) Valor observado da estatística de teste aproximadamente igual a 0.222
- (C) Valor p do teste aproximadamente igual a 0.00000003

Podemos dizer que as afirmações

- ☐ (B) e (C) estão certas mas (A) está errada. O declive é significativamente positivo.
- ☐ (A) e (B) estão certas mas (C) está errada. O declive não é significativamente positivo.
- ☐ (A), (B) e (C) estão certas. O declive não é significativamente positivo.
- ☐ (A) e (C) estão certas mas (B) está errada. O declive é significativamente positivo.

Q36 Tomando os dados considerados na questão **Q33**, qual a alteração esperada no peso indicado pela balança quando o verdadeiro peso sofre um aumento 1.5 unidades?

- ☐ sofre uma diminuição de aproximadamente 0.167 unidades
- ☐ sofre um aumento de aproximadamente 1.446 unidades
- ☐ sofre um aumento de aproximadamente 1.075 unidades
- ☐ sofre um aumento de aproximadamente 1.613 unidades

Q37 Do seguinte quadro pretende-se avaliar se a escolha de um sabor de gelado depende do género.

	Chocolate	Baunilha	Morango
Masculino	100	120	60
Feminino	350	200	90

É verdade que:

- ☐ Se as variáveis forem independentes, é esperado observar 100 pessoas do sexo masculino e que preferem o sabor de chocolate.
- ☐ A escolha do sabor de gelado não depende do sexo, considerando 5% de significância.
- ☐ O valor observado da estatística de teste é aproximadamente 28.4.
- ☐ Existem 4 graus de liberdade no teste.