

Instituto Superior de Agronomia  
Estatística – 3ª chamada – Exame Época Especial (2020/2021)

Duração: 3 horas

6 de Setembro de 2021

Apresente os cálculos necessários para responder às questões apresentadas.

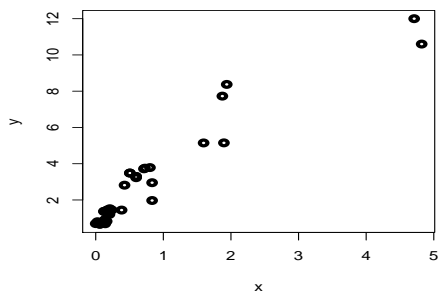
1. (4.5 val.) Os ácidos gordos (AG) ómega-3 são nutrientes importantes que não são produzidos pelos vertebrados, pelo que têm de ser obtidos através da sua alimentação. Com o objetivo de quantificar a resposta do teor deste nutriente no tecido adiposo de suínos, suplementados com fontes ricas em AG ómega-3, foram analisados 34 ensaios. Os dados recolhidos foram a quantidade de AG ómega-3 na dieta ( $x$ , em g/100 g de matéria seca) e a concentração de AG ómega-3 no tecido adiposo ( $y$ , em g/100 g de tecido), os quais foram introduzidos na aplicação R. Para responder às questões colocadas considere os seguintes comandos e respectivos resultados.

```
> sort(y)
[1] 0.665 0.665 0.689 0.696 0.800 0.825 0.872 0.929 1.205 1.312 1.370
...
[24] 3.485 3.485 3.719 3.754 3.787 5.146 5.150 7.728 8.374 10.600 12.000

> mean(x)      > median(x)      > sd(x)          > mean(y)      > median(y)      > sd(y)
[1] 0.7796      [1] 0.4045      [1] 1.148228     [1] 3.0428      [1] 1.744         [1] 2.838798

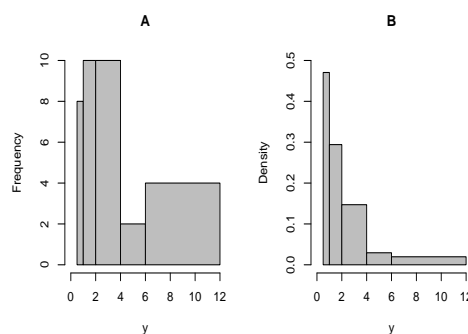
> cor(x,y)
[1] 0.9471066

> plot(x,y,lwd=4)
```



```
> hist(y,breaks=c(0.5,1,2,4,6,12),plot=FALSE)
$breaks
[1] 0.5 1.0 2.0 4.0 6.0 12.0
$counts
[1] 8 10 10 2 4
$midpoints
[1] 0.75 1.50 3.00 5.00 9.00
$density
[1] 0.47058824 0.29411765 0.14705882 0.02941176 0.01960784
(...)
```

- a) Classifique, justificando, as variáveis em estudo.
- b) Considere as observações da **variável  $y$** .
- Construa um *boxplot* para estes dados indicando a designação das observações e das medidas de localização e dispersão em destaque no gráfico.
  - Qual o valor da média e da variância das observações de concentração de AG ómega-3 no tecido adiposo se os dados fossem registados em g/1000 g de tecido?
  - Indique, justificando, qual dos histogramas em baixo corresponde ao agrupamento de dados efectuado no comando `hist` apresentado.



(v.s.f.f.)

- c) Considere agora as **duas variáveis observadas nos ensaios,  $x$  e  $y$** .
- Parece-lhe adequada a utilização de uma regressão linear para descrever a relação entre a quantidade de AG ómega-3 na dieta e a concentração de AG ómega-3 no tecido adiposo? Justifique.
  - Independentemente da resposta dada na alínea anterior, escreva a equação da recta de regressão dos mínimos quadrados de  $y$  sobre  $x$ .
  - Interprete, no contexto do problema, os coeficientes da recta de regressão determinada na alínea anterior.

2. (4 val.) Uma máquina produz peças para um equipamento de rega que devem ter comprimento de acordo com a norma especificada pelo mercado. Suponha que o desvio (em cm) do comprimento especificado que uma peça pode apresentar é uma v.a.  $X$ , com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} 1 + x & -1 \leq x \leq 0, \\ 1 - x & 0 < x \leq 1, \\ 0 & \text{restantes valores de } x. \end{cases}$$

- Determine o valor médio e a variância de  $X$ .
- Determine a função de distribuição cumulativa de  $X$ .
- Será o 3.<sup>o</sup> quartil superior a 0.5? Justifique.
- O lucro, em unidades monetárias (u.m.) obtido pela venda de cada peça é uma variável aleatória  $Y$ , definida pela seguinte relação  $Y = 2 - 2X^2$ .  
Qual a probabilidade de o lucro da venda de uma peça ser inferior a 1 u.m.?

3. (5 val.) O nemátodo da madeira de pinheiro (*Bursaphelenchus xylophilus*) é um animal microscópico que ataca algumas espécies de pinheiro em Portugal que, tipicamente, causa declínio aos pinheiros atacados. Estudos recentes mostram que, na região de Setúbal, 15% dos pinheiros bravos adultos que apresentam declínio estão de facto atacados por este nemátodo, e que 95% dos pinheiros bravos adultos atacados exibem sinais de declínio.

A probabilidade de um pinheiro bravo adulto nessa região exibir sinais de declínio é de 10%.

- Escolhido, ao acaso, um pinheiro bravo adulto nessa região, determine a probabilidade de esse pinheiro
  - Apresentar sinais de declínio e não estar atacado por este nemátodo;
  - Estar atacado por este nemátodo.
- Os pinheiros em declínio são por vezes cortados e estilhaçados para evitar a propagação do nemátodo da madeira do pinheiro. Tendo sido cortados 1000 pinheiros bravos adultos, escolhidos aleatoriamente entre os que apresentavam sinais de declínio na região de Setúbal, considere a v.a.  $X$  que designa o número de pinheiros bravos adultos atacados por este nemátodo nesses 1000 pinheiros. Determine:
  - A distribuição da v.a.  $X$ . Justifique.
  - A probabilidade, aproximada, de haver menos de 120 pinheiros atacados nesses 1000 pinheiros.
- Na região do Oeste foram analisados 100 pinheiros bravos adultos escolhidos ao acaso de entre os que apresentavam sinais de declínio. Observou-se que desses, 20 tinham de facto sido atacados por este nemátodo. Pode concluir-se, com um nível de significância de 5%, que a proporção de ataque daquele nemátodo em pinheiros bravos adultos que apresentam sinais de declínio é superior na região do Oeste do que na região de Setúbal? Responda à questão de forma completa.

4. (2.5 val.) Seja  $X$  uma variável aleatória com distribuição normal com parâmetros  $\mu$  e  $\sigma$ .
- Calcule a probabilidade de  $X$  tomar valores entre  $\mu - \sigma/4$  e  $\mu + \sigma/4$ .
  - Os acontecimentos  $A = \{X : X < \mu - \sigma/4\}$  e  $B = \{X : X < \mu\}$  são independentes? Justifique.
  - Prove que  $Z = \frac{X - \mu}{\sigma}$  tem distribuição normal reduzida.

5. (4 val.) Dos tempos obtidos em provas disputadas, ao longo das suas carreiras, por cada uma de duas atletas de uma dada modalidade foram extraídas amostras aleatórias, cujos valores (em segundos) foram:

Atleta A: 16.2 15.7 15.9 15.4 16.0 14.8 15.4 15.2 14.9 16.1

Atleta B: 16.3 16.0 15.9 16.6 14.7 15.9 16.0 15.5 16.4 16.7

**Nota: Para responder a esta pergunta utilize resultados que lhe são apresentados no Anexo. Note que alguns comandos podem não ser adequados.**

- Para o valor médio do tempo realizado pela atleta A ao longo da sua carreira:
  - Indique um estimador centrado deste parâmetro, explicitando-o.
  - Calcule, justificando adequadamente, um intervalo a 95% de confiança para aquele parâmetro.
- Estas duas atletas apresentam-se a um determinado campeonato. Nos bastidores do campeonato aposta-se na candidata A como sendo a que irá ganhar a prova, por ser a mais rápida das duas. Considera que os dados obtidos sustentam o sentido dessa aposta? Explícite e valide as hipóteses necessárias à sua resposta.

## ANEXO

```
> AtletaA<-c( 16.2, 15.7, 15.9, 15.4, 16.0, 14.8, 15.4, 15.2, 14.9,16.1)
> AtletaB<-c( 16.3, 16.0, 15.9, 16.6, 14.7, 15.9, 16.0, 15.5, 16.4, 16.7)
>
> shapiro.test(AtletaA)                                > shapiro.test(AtletaB)

      Shapiro-Wilk normality test                        Shapiro-Wilk normality test
data:  AtletaA                                         data:  AtletaB
W = 0.93609, p-value = 0.5104                          W = 0.9066, p-value = 0.2584

> shapiro.test(AtletaA-AtletaB)                        > var(AtletaA)
      Shapiro-Wilk normality test                       [1] 0.2471111
data:  AtletaA - AtletaB
W = 0.92153, p-value = 0.3699

> var.test(AtletaA,AtletaB)

      F test to compare two variances
data:  AtletaA and AtletaB
F = 0.7268, num df = 9, denom df = 9, p-value = 0.6422
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1805262 2.9260820
sample estimates:
ratio of variances
 0.7267974
```

```
> t.test(AtletaA,AtletaB,alternative="less",var.equal=TRUE)
```

Two Sample t-test

```
data: AtletaA and AtletaB
t = -1.8159, df = 18, p-value = 0.04304
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.01982987
sample estimates:
mean of x mean of y
 15.56    16.00
```

```
> t.test(AtletaA,AtletaB,var.equal=TRUE)
```

Two Sample t-test

```
data: AtletaA and AtletaB
t = -1.8159, df = 18, p-value = 0.08608
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.94906131  0.06906131
sample estimates:
mean of x mean of y
 15.56    16.00
```

```
> t.test(AtletaA,AtletaB)
```

Welch Two Sample t-test

```
data: AtletaA and AtletaB
t = -1.8159, df = 17.56, p-value = 0.0865
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.94997622  0.06997622
sample estimates:
mean of x mean of y
 15.56    16.00
```

```
> t.test(AtletaA,AtletaB,alternative="less",paired=TRUE)
```

Paired t-test

```
data: AtletaA and AtletaB
t = -1.7696, df = 9, p-value = 0.05529
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 0.01578657
sample estimates:
mean of the differences
 -0.44
```