

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2023-24

22 janeiro 2024 Segunda Chamada de Exame Duração: 3h00 (+15min)

I [7 valores]

Num estudo sobre características comerciais dos bolbos de variedades tradicionais de alho, efectuaram-se 39 medições independentes, em plantas tomadas ao acaso, das seguintes características: peso do bolbo (variável **peso**, em *g*), diâmetro do bolbo (variável **diametro**, em *cm*); número de dentes por bolbo (variável **ndentes**); espessura do dente (variável **espessura**, em *cm*); largura do dente (variável **largura**, em *cm*); altura do dente (variável **altura**, em *cm*). Eis os seguintes indicadores, bem como a matriz de correlações para todos os pares de variáveis.

Variável	Mínimo	Média	Máximo	Variância
peso	5.339	10.876	16.659	9.8094
diametro	2.488	3.007	4.210	0.1767
ndentes	5.167	8.895	12.839	1.3309
espessura	0.750	1.098	1.347	0.0174
largura	0.710	1.094	1.938	0.0815
altura	1.583	2.269	2.567	0.0346

```
> round(cor(alho),4)
           peso diametro ndentes espessura largura altura
peso      1.0000  0.8416  0.2193  0.8038  0.6689  0.7126
diametro  0.8416  1.0000  0.1530  0.5591  0.9025  0.4326
ndentes   0.2193  0.1530  1.0000 -0.1676 -0.1211  0.0712
espessura 0.8038  0.5591 -0.1676  1.0000  0.4835  0.7484
largura   0.6689  0.9025 -0.1211  0.4835  1.0000  0.2653
altura    0.7126  0.4326  0.0712  0.7484  0.2653  1.0000
```

Um dos técnicos responsável pelo estudo decidiu ajustar o modelo de regressão linear simples do diâmetro do bolbo (variável **diametro**) sobre o peso do bolbo (variável **peso**). Eis os resultados obtidos:

```
> alho1.lm<-lm(diametro~peso , data=alho)
> summary(alho1.lm)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.77840    0.13477  13.196 1.45e-15
peso          0.11296    0.01192    A    1.94e-11
---
Residual standard error: B on 37 degrees of freedom
Multiple R-squared:  0.7083, Adjusted R-squared:  0.7004
F-statistic: C on D and E Degrees of freedom, p-value: 1.939e-11
```

1. Descreva em pormenor o modelo linear ajustado.
2. Complete o *output* do R, justificando como obtém os cinco valores em falta (indicados por letras).
3. Utilize um teste de hipóteses sobre o declive da recta populacional para validar a seguinte afirmação: *não existe uma relação linear significativa entre peso e diâmetro dos bolbos de alho.*

4. Para a observação com maior efeito alavanca calcule o valor ajustado do diâmetro.
5. Qual a média dos valores ajustados do diâmetro? Justifique, demonstrando o que é a média dos valores ajustados de y numa regressão linear simples.
6. Indique um intervalo de predição a 95% associado ao diâmetro de um bolbo com 12 gramas. Interprete o resultado obtido.

II [6 valores]

Para o mesmo conjunto de dados descritos no grupo I, foi seguidamente ajustado um modelo de regressão linear múltipla para modelar o diâmetro do bolbo (variável `diâmetro`) a partir das restantes variáveis. Eis os resultados obtidos com o ajustamento deste modelo, bem como a matriz de (co-)variâncias estimadas dos estimadores dos parâmetros do modelo:

```
> alho3.lm<-lm(diâmetro~. , data=alho)
> summary(alho3.lm)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.35114      0.48733   2.773  0.00907
peso          0.06523      0.01809   3.607  0.00101
ndentes      0.03725      0.02343   1.590  0.12139
espessura    -0.43296      0.32610  -1.328  0.19340
largura       0.96275         ?         ?    2.71e-10
altura       0.01666      0.16270   0.102  0.91907

Residual standard error: 0.1089 on 33 degrees of freedom
Multiple R-squared:  0.9417, Adjusted R-squared:  0.9329
F-statistic: ? on ? and ? DF
```

```
> round(vcov(alho3.lm), 4)
              (Intercept)      peso ndentes espessura largura  altura
(Intercept)    0.2375  0.0076 -0.0090  -0.0959 -0.0355 -0.0420
peso            0.0076  0.0003 -0.0003  -0.0041 -0.0015 -0.0010
ndentes        -0.0090 -0.0003  0.0005   0.0049  0.0014  0.0002
espessura      -0.0959 -0.0041  0.0049   0.1063  0.0109 -0.0144
largura        -0.0355 -0.0015  0.0014   0.0109  0.0117  0.0064
altura         -0.0420 -0.0010  0.0002  -0.0144  0.0064  0.0265
```

1. Prove que, dado o modelo de regressão linear múltipla, a distribuição das observações \vec{Y} da variável resposta é:

$$\vec{Y} \sim \mathcal{N}_n(\mathbf{X}\vec{\beta}, \sigma^2 \mathbf{I}_n).$$

2. Comparativamente ao modelo de regressão linear simples ajustado no grupo I, explique (sem fazer cálculos) quais as alterações nas 3 somas de quadrados (SQT, SQR e SQRE) agora obtidas com o ajustamento do modelo de regressão linear múltipla.
3. Interprete o valor obtido do coeficiente de determinação e descreva em pormenor o teste F de ajustamento global.
4. Teste, com um nível de significância $\alpha = 0.05$, se quando a largura do dente aumenta 1 cm (mantendo os valores dos restantes preditores constantes) o diâmetro do bolbo aumenta, em média, mais de 1 cm.

5. Responda, justificando, se a seguinte afirmação é falsa ou verdadeira: *Atendendo aos valores dos p-values dos testes t de hipóteses $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ ($j = 1, \dots, 5$), aplicando o algoritmo de exclusão sequencial, podemos garantir que o submodelo final fica apenas com as variáveis predictoras peso e largura.*
6. Teste formalmente, através de um Teste F parcial de comparação de um modelo com o seu submodelo, se a qualidade do ajustamento do modelo completo (com 5 preditores) e do submodelo sem o preditor altura difere significativamente ($\alpha=0.05$).

III [7 valores]

1. Num estudo sobre 5 variedades tradicionais de tomate (designadas V22, V26, V2B, V6 e V9), foi instalado um ensaio com um delineamento experimental em blocos completos casualizados, com 4 blocos (designados B1, B2, B3 e B4). Isto é, cada bloco foi dividido em 5 parcelas, sendo a cada uma dessas parcelas associada, aleatoriamente, uma diferente variedade. Em cada parcela avaliou-se o peso do fruto (variável peso, em g/fruto). As médias amostrais obtidas são indicadas seguidamente:

```
> model.tables(tomate.aov, type="mean")
Tables of means
Grand mean
222.887
  variedade
variedade
  V22  V26  V2B  V6  V9
374.2 281.1  76.1 162.9 220.2
  bloco
bloco
  B1  B2  B3  B4
197.76 220.72 239.38 233.68
```

- (a) Descreva em pormenor o modelo ANOVA adequado ao problema sob estudo.
- (b) Seguidamente apresenta-se a tabela ANOVA relativa ao estudo efectuado:

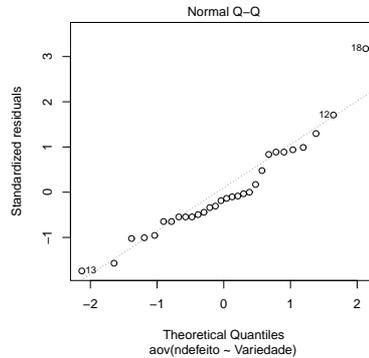
```
> tomate.aov<-aov(peso~variedade+bloco, data=tomate)
> summary(tomate.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
variedade	?	205769	51442	?	3.26e-07
bloco	?	5123	?	1.546	0.253
Residuals	?	13255	1105		

Complete a tabela, indicando como obtém cada um dos valores em falta.

- (c) Qual é a variância amostral dos pesos dos frutos na totalidade das observações?
 - (d) A variedade de tomate afecta o peso do fruto? Justifique formalmente através de um teste de hipóteses (ao nível $\alpha=0.05$).
 - (e) Calcule o valor ajustado do peso do fruto para a observação feita na variedade V22 no bloco B1.
2. Um outro ensaio com outras 5 variedades tradicionais de tomate (designadas VA, VB, VC, VD e VE) foi instalado de acordo com um delineamento experimental totalmente casualizado, com 6 repetições de cada variedade. Em cada unidade experimental avaliou-se o número de frutos com defeito numa amostra de 100 frutos.

- (a) O responsável pelo estudo começou por fazer o ajustamento de um modelo ANOVA a 1 factor de efeitos fixos. Na análise dos resíduos para validação dos pressupostos do modelo, obteve o seguinte gráfico:



Explique e comente o gráfico apresentado.

- (b) Seguidamente, o responsável pelo estudo optou por uma análise de variância de tipo não paramétrico. Os dados da experiência, as respetivas ordenações por variedade (entre parênteses), bem como a soma de cada uma dessas ordenações (Soma(ord): R_i , para $i = 1, \dots, 5$), encontram-se seguidamente descritos:

	Número de frutos com defeito (Ordenação)				
	Variedade				
	VA	VB	VC	VD	VE
	16 (7)	60 (24)	14 (5)	8 (1)	24 (10)
	25 (11)	63 (26)	21 (8)	10 (2)	29 (14)
	26 (12)	64 (27)	28 (13)	12 (3)	32 (17)
	36 (19)	65 (28)	30 (15)	13 (4)	33 (18)
	41 (20)	78 (29)	31 (16)	15 (6)	42 (21)
	44 (23)	86 (30)	62 (25)	22 (9)	43 (22)
Soma(Ord)	(92)	(164)	(82)	(25)	(102)

De acordo com o delineamento experimental do ensaio, o que conclui sobre o estudo efectuado ao nível $\alpha=0.05$? Descreva em pormenor o teste realizado.

Notas:

- Estatística do teste de Kruskal Wallis: sob H_0 ,

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) \sim \chi_{k-1}^2 ;$$

- Estatística do teste de Friedman: sob H_0 ,

$$S = \frac{12}{ab(a+1)} \sum_{i=1}^a R_i^2 - 3b(a+1) \sim \chi_{a-1}^2 ;$$

- $\chi_{0.05(4)}^2 = 9.488$; $\chi_{0.05(5)}^2 = 11.070$; $\chi_{0.05(6)}^2 = 12.592$; $\chi_{0.05(7)}^2 = 14.067$.