

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO EXPERIMENTAL – 2025-26
26 de janeiro de 2026 Segunda Chamada de Exame Duração: 3h (+15 min.)

I [14 valores]

Num estudo sobre variedades tradicionais de alho, observaram-se três características em 150 parcelas independentes: peso do bolbo (variável **pesobolbo**, em g), peso do dente (**pesodente**, em g) e o número de dentes por bolbo **ndentesbolbo**. Nas análises usou-se a média de cada característica obtida por parcela. Eis alguns indicadores estatísticos, bem como as matrizes de (co-)variâncias e de correlações para todos os pares de variáveis.

```
> summary(alho[, -1])
  pesobolbo      pesodente      ndentesbolbo
Min.   : 4.42   Min.   :0.590   Min.   : 5.800
1st Qu.:18.37   1st Qu.:1.825   1st Qu.: 9.425
Median :26.49   Median :2.625   Median :10.305
Mean   :27.12   Mean   :2.686   Mean   :10.642
3rd Qu.:34.14   3rd Qu.:3.370   3rd Qu.:11.145
Max.   :77.51   Max.   :8.140   Max.   :32.700
> round(var(alho[, -1]), 4)
      pesobolbo pesodente ndentesbolbo
pesobolbo  156.9295  11.9652  15.4679
pesodente   11.9652   1.3656  -0.3121
ndentesbolbo 15.4679 -0.3121   9.0649
> round(cor(alho[, -1]), 4)
      pesobolbo pesodente ndentesbolbo
pesobolbo  1.0000  0.8173  0.4101
pesodente   0.8173  1.0000 -0.0887
ndentesbolbo 0.4101 -0.0887  1.0000
```

1. Um dos técnicos que realizou este estudo decidiu ajustar o modelo de regressão linear do peso do dente (variável **pesodente**, em g) sobre as restantes variáveis observadas. Eis os resultados obtidos com o ajustamento desse modelo:

```
> alho1<-lm(pesodente~pesobolbo+ndentesbolbo, data=alho)
> summary(alho1)
Call:
lm(formula = pesodente ~ pesobolbo + ndentesbolbo, data = alho)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.194789    0.123224   17.81  <2e-16
pesobolbo     0.095742      A      33.34  <2e-16
ndentesbolbo -0.197802    0.011950  -16.55  <2e-16
---
Residual standard error: B on 147 degrees of freedom
Multiple R-squared:  0.8841, Adjusted R-squared:  0.8825
F-statistic: 560.7 on 2 and 147 DF, p-value: < 2.2e-16
```

- (a) Quantos preditores e quantos parâmetros tem o modelo? Identifique-os.
- (b) No *output* dos comandos do R identifique e calcule, justificando, os valores em falta A e B.
- (c) Interprete o significado do coeficiente associado ao preditor **ndentesbolbo**.
- (d) Será admissível afirmar que, em média, existe uma diminuição no peso do dente quando o número de dentes no bolbo aumenta? Responda através do teste de hipóteses adequado explicando, em pormenor, todos os seus passos.
- (e) Prove que a estatística do teste F de ajustamento global é a estatística do teste F Parcial quando o submodelo é o modelo nulo.

(f) Para bolbos com peso de 50 g e 20 dentes, calcule:

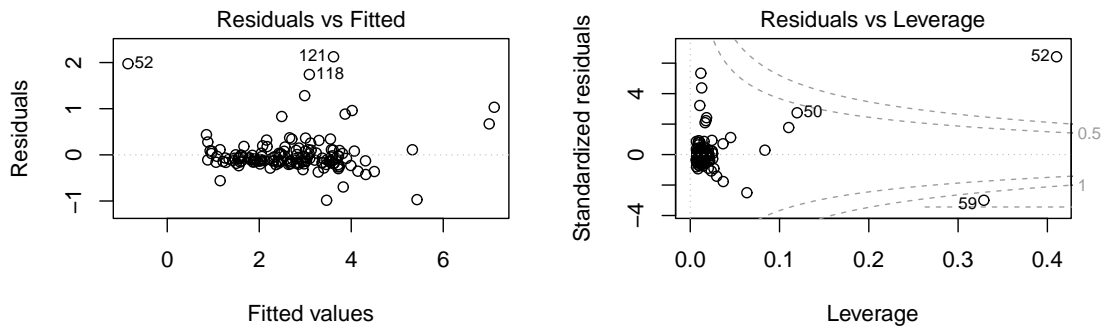
- i. O valor estimado do peso do dente.
- ii. Um intervalo de confiança (95%) para o valor esperado do peso do dente. Para responder a esta questão tenha em conta o seguinte resultado obtido no R:

```
> predict(alho1, new=data.frame(pesobolbo=50, ndentesbolbo=20), int="conf")
      fit      lwr      upr
1     ???     ???    3.241125
```

- (g) É possível concluir que os modelos com e sem o preditor `ndentesbolbo` têm ajustamento significativamente diferente? Responda à questão utilizando um teste F parcial.
- (h) Qual o coeficiente de determinação do submodelo resultante da exclusão da variável `ndentesbolbo`?
- (i) Prove que, dado o modelo de regressão linear múltipla, a distribuição das observações \vec{Y} da variável resposta é:

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

(j) Alguns dos gráficos de resíduos do modelo apresentam-se seguidamente:



- i. Sem fazer quaisquer cálculos, diga qual é o valor aproximado do resíduo usual e do efeito alavanca da observação 52.
 - ii. Para a observação 52 calcule a distância de Cook. Trata-se de uma observação influente?
 - iii. Analise os gráficos apresentados.
2. No estudo anterior, as observações foram efetuadas em cinco locais diferentes: Braga, Elvas, Esposende, Faro e Vila do Conde. O técnico responsável pelo estudo considerou importante compreender a relação entre o peso do dente e o peso do bolbo nesses diferentes locais, decidindo, para tal, ajustar uma análise de covariância. A seguir, apresentam-se os resultados obtidos com essa abordagem:

```
> alho2<-lm(pesodente~pesobolbo*Local, data=alho)
> summary(alho2)
```

Call:

```
lm(formula = pesodente ~ pesobolbo * Local, data = alho)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.97825	-0.19985	-0.02946	0.14023	2.76406

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.1768434	0.3718699	3.165	0.00191
pesobolbo	0.0650181	0.0097958	6.637	6.49e-10
LocalElvas	-0.7565812	0.6133661	-1.233	0.21946
LocalEsposende	0.2942538	0.4878346	0.603	0.54736
LocalFaro	-0.7242777	0.5174387	-1.400	0.16380
LocalVilaConde	-0.3996510	0.5064529	-0.789	0.43138
pesobolbo:LocalElvas	0.0094821	0.0222768	0.426	0.67102

```

pesobolbo:LocalEsposende -0.0135776  0.0127694  -1.063  0.28948
pesobolbo:LocalFaro      -0.0007299  0.0261444  -0.028  0.97777
pesobolbo:LocalVilaConde  0.0233769  0.0161654   1.446  0.15038
---
```

```

Residual standard error: 0.6033 on 140 degrees of freedom
Multiple R-squared:  0.7496, Adjusted R-squared:  0.7335
F-statistic: 46.56 on 9 and 140 DF,  p-value: < 2.2e-16
```

```
> alho3<-lm(pesodente~pesobolbo, data=alho)
```

```
> anova(alho3,alho2)
Analysis of Variance Table
```

```

Model 1: pesodente ~ pesobolbo
Model 2: pesodente ~ pesobolbo * Local
  Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1     148 67.541
2     140 50.952   8     16.589 5.6976 2.842e-06
---
```

- (a) Escreva a equação da recta ajustada referente a Braga.
- (b) Tendo em conta a informação disponibilizada e sem fazer cálculos, responda se o modelo de covariância ajustado e o modelo de recta única diferem significativamente ao nível $\alpha=0.05$.

3. Um outro técnico sugeriu ajustar a regressão linear simples apresentada seguidamente:

```

> alho4<-lm(log(pesodente)~log(pesobolbo), data=alho)
> summary(alho4)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.7122     0.1247  -13.73  <2e-16
log(pesobolbo)  0.8157     0.0386   21.13  <2e-16
---
Residual standard error: 0.2235 on 148 degrees of freedom
Multiple R-squared:  0.7511, Adjusted R-squared:  0.7494
F-statistic: 446.5 on 1 and 148 DF,  p-value: < 2.2e-16
```

- (a) Comente o valor do coeficiente de determinação. Ao nível de significância $\alpha=0.05$, o modelo ajustado difere significativamente do modelo nulo? Responda através de um teste T .
- (b) Escreva a equação da recta ajustada e deduza a relação não linear existente entre as variáveis não transformadas.

II [6 valores]

1. Num estudo sobre a tolerância ao calor e à seca em clones de videira da casta Arinto, foi medida a temperatura da superfície foliar (variável SLT, em °C) no mês de julho, num dia sem vento, entre as 12h e as 14h, em quatro clones (AR1, AR2, AR3 e AR4). O ensaio foi conduzido segundo um delineamento experimental em blocos completos casualizados, com seis blocos. Em cada unidade experimental, composta por três plantas, foi medida a SLT em 30 folhas, tendo sido utilizada a média da SLT dessas 30 folhas para as análises subsequentes.

- (a) Esta experiência apresenta pseudo-repetições? Em caso afirmativo, identifique-as.
- (b) Quantas repetições tem o ensaio?
- (c) Quantos fatores estão em estudo?
- (d) Descreva, em pormenor, o modelo de Análise de Variância considerado adequado para abordar o problema descrito.

- (e) O técnico responsável pelo estudo decidiu conduzir a análise descrita seguidamente:

```
> arinto.aov<-aov(SLT~Clone+Bloco, data=arinto)
> summary(arinto.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Clone	3	A	28.246	5.183	0.01175
Bloco	B	130.58	26.116	C	0.00813
Residuals	15	81.74	5.449		

```
---
```

```
> model.tables(arinto.aov,type="means")
Tables of means
Grand mean
```

25.45639

```
Clone
Clone
AR1      AR2      AR3      AR4
28.551  25.023  24.767  23.485

Bloco
Bloco
      B1      B2      B3      B4      B5      B6
26.083 21.553 27.784 27.848 26.289 23.181
```

- Complete a tabela ANOVA, indicando como obtém cada um dos valores em falta A, B e C.
 - As médias amostrais da SLT de cada clone sugerem que há efeitos do clone. A fim de responder a esta questão, efectue o teste adequado, explicando, em pormenor, todos os seus passos (ao nível $\alpha=0.05$).
 - Calcule a SLT ajustada para o clone AR1 no Bloco 2.
 - Um técnico que apenas conhecia ANOVAs a um factor decidiu estudar os efeitos do clone, ignorando os efeitos do bloco, utilizando os mesmos dados. Construa a tabela-resumo da ANOVA resultante e indique a que conclusão chegaria, ao nível de significância $\alpha = 0.05$. Concorda com esta opção metodológica? Justifique a sua resposta, apresentando as razões que a sustentam.
2. O técnico responsável pela análise em blocos completos casualizados, não ficou satisfeito quando analisou os gráficos dos resíduos para validação dos pressupostos do modelo, adoptando de seguida uma abordagem não paramétrica. Da análise conduzida, obteve os seguintes resultados:

```
> ?.test(SLT ~ Clone | Bloco, data = arinto)
---
```

data: SLT and Clone and Bloco
?chi-squared = 11.4, df = 3, p-value = 0.009748

- Indique, justificando, qual é o teste não paramétrico adequado ao estudo e que foi utilizado na análise efetuada no R.
- Para o cálculo da estatística do teste, foram usados os valores observados da variável resposta? Justifique.
- Enuncie as hipóteses relativas ao teste efetuado e, admitindo válida a distribuição assintótica da estatística do teste, indique que conclusão retira sobre o estudo efetuado ao nível de significância $\alpha = 0.05$.