

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

I [14 valores]

Um estudo sobre características de qualidade do mosto na variedade de videira Grenache envolveu a avaliação de várias variáveis em 343 amostras de bagos independentes: peso do bago (variável **pesobago**, em g), volume do mosto do bago (variável **volumebago**, em ml), pH, **brix** (em graus brix), **acidez** (em g/l de ácido tartárico), teor de antocianas (variável **antocianas**, em mg/l) e teor de fenóis total (variável **fenois**, em mg/l). Eis as matrizes de (co-)variâncias e de correlações para todos os pares de variáveis.

```
> round(var(grenache[,-1]),4)
      pesobago volumebago      pH      brix    acidez antocianas      fenois
pesobago   398.9785  301.2972  0.1369 -0.5852   1.7685 -412.4644 -1903.6865
volumebago 301.2972  232.5795  0.0179 -2.0306   1.5403 -365.3285 -1555.2572
pH         0.1369     0.0179  0.0155  0.1147 -0.0264   1.9022   4.0189
brix       -0.5852   -2.0306  0.1147  2.8548 -0.3830   78.8894  146.0464
acidez      1.7685    1.5403 -0.0264 -0.3830   0.3282 -21.0275  -32.5844
antocianas -412.4644 -365.3285  1.9022  78.8894 -21.0275  5278.3866  8755.2304
fenois     -1903.6865 -1555.2572  4.0189 146.0464 -32.5844  8755.2304 26042.9144
> round(cor(grenache[,-1]),4)
      pesobago volumebago      pH      brix    acidez antocianas      fenois
pesobago   1.0000     0.9891  0.0551 -0.0173   0.1545 -0.2842 -0.5906
volumebago  0.9891    1.0000  0.0095 -0.0788   0.1763 -0.3297 -0.6319
pH         0.0551     0.0095  1.0000  0.5460 -0.3703   0.2105  0.2002
brix       -0.0173   -0.0788  0.5460  1.0000 -0.3956   0.6427  0.5356
acidez      0.1545    0.1763 -0.3703 -0.3956  1.0000 -0.5052 -0.3524
antocianas -0.2842   -0.3297  0.2105  0.6427 -0.5052   1.0000  0.7467
fenois     -0.5906   -0.6319  0.2002  0.5356 -0.3524   0.7467  1.0000
```

1. Um dos técnicos que realizou este estudo decidiu ajustar o modelo de regressão linear múltipla do teor de fenóis total (variável **fenois**, em mg/l) sobre as restantes variáveis observadas. Eis os resultados obtidos com o ajustamento desse modelo, bem como a matriz de (co-)variâncias estimadas dos estimadores dos parâmetros do modelo:

```
> grenachelm1<-lm(fenois~., data=grenache[,-1])
> summary(grenachelm1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 343.24998 165.60438  2.073 0.03896
pesobago     1.04711  1.66193     A  0.52909
volumebago  -6.27885  2.19074  -2.866 0.00442
pH          17.49293 45.92207  0.381 0.70350
brix        18.24684  4.20701  4.337 1.91e-05
acidez      17.50120  9.44038  1.854 0.06463
antocianas  1.09665  0.09469 11.582 < 2e-16
---
Residual standard error: 81.37 on 336 degrees of freedom
Multiple R-squared:  0.7502, Adjusted R-squared:  B
F-statistic: 168.2 on 6 and 336 DF,  p-value: < 2.2e-16
```

```
> round(vcov(grenachelm1),4)
            (Intercept) pesobago volumebago      pH      brix    acidez antocianas
(Intercept)           C  48.4083  -69.1278 -6389.8795  62.2829 -953.4617 -4.3049
pesobago          48.4083  2.7620  -3.6040 -10.9613 -1.3122 -1.3763 -0.0155
volumebago       -69.1278 -3.6040   4.7993  14.4345  1.5040  1.7139  0.0297
pH             -6389.8795 -10.9613  14.4345 2108.8364 -96.0831 135.8672  1.3598
brix            62.2829 -1.3122   1.5040 -96.0831 17.6989 -2.2694 -0.2374
acidez          -953.4617 -1.3763   1.7139 135.8672 -2.2694  89.1208  0.3511
antocianas      -4.3049 -0.0155   0.0297  1.3598 -0.2374  0.3511  0.0090
```

- (a) No *output* dos comandos *summary* e *vcov* do R identifique e calcule, justificando, os valores em falta A, B, e C .
- (b) Comente o valor do coeficiente de determinação. Ao nível de significância $\alpha=0.05$, o modelo ajustado difere significativamente do modelo nulo? Justifique a sua resposta.
- (c) Calcule um intervalo a 95% de confiança para o coeficiente associado ao preditor *antocianas*. Explique o seu significado.
- (d) Será admissível afirmar que quando o volume do mosto do bago (*volumebago*) aumenta 1ml, mantendo os restantes preditores constantes, o teor de fenóis diminui, em média, mais de 5mg/l? Responda através do teste de hipóteses adequado explicando, em pormenor, todos os seus passos.
- (e) Efetuando um algoritmo de exclusão sequencial com base em testes T aos parâmetros do modelo (para $\alpha=0.10$), qual a primeira variável preditora a sair do modelo? Justifique a sua resposta.
- (f) A aplicação do algoritmo de exclusão sequencial com base no critério de Informação de Akaike, obteve o seguinte resultado:

```
> step(grenachelm1)
Start: AIC=3024.64
fenois ~ pesobago + volumebago + pH + brix + acidez + antocianas

          Df Sum of Sq    RSS    AIC
- pH      1     961 2225578 3022.8
- pesobago 1     2628 2227246 3023.0
<none>            2224617 3024.6
- acidez   1     22755 2247372 3026.1
- volumebago 1     54387 2279004 3030.9
- brix     1     124550 2349168 3041.3
- antocianas 1     888076 3112694 3137.8

Step: AIC=3022.78
fenois ~ pesobago + volumebago + brix + acidez + antocianas

          Df Sum of Sq    RSS    AIC
- pesobago 1     3170 2228748 3021.3
<none>            2225578 3022.8
- acidez   1     22088 2247666 3024.2
- volumebago 1     57668 2283246 3029.6
- brix     1     180253 2405831 3047.5
- antocianas 1     964199 3189777 3144.2

Step: AIC=3021.27
fenois ~ volumebago + brix + acidez + antocianas

          Df Sum of Sq    RSS    AIC
<none>            2228748 3021.3
- acidez   1     22902 2251651 3022.8
- brix     1     214506 2443254 3050.8
- antocianas 1     973720 3202468 3143.6
- volumebago 1     1652170 3880918 3209.5

Call:
lm(formula = fenois ~ volumebago + brix + acidez + antocianas,
    data = grenache[, -1])
```

Coefficients:

(Intercept)	volumebago	brix	acidez	antocianas
389.862	-4.914	19.806	16.656	1.089

- Sem fazer quaisquer cálculos, diga qual é o valor do critério de informação de Akaike (AIC) para o modelo completo.
- Com base no critério (AIC), qual foi a ordem de exclusão das variáveis que saíram do modelo? Justifique.
- Dado um Modelo de Regressão Linear Múltipla com p preditores, e um seu submodelo com apenas $k < p$ preditores, a estatística do teste para comparar modelo completo e

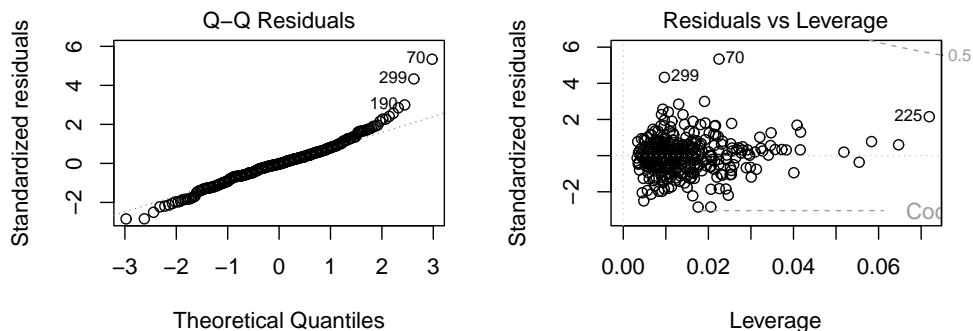
submodelo (estatística do teste F parcial) é dada por:

$$F = \frac{\frac{SQRE_S - SQRE_C}{p-k}}{\frac{SQRE_C}{n-(p+1)}},$$

onde $SQRE_C$ e $SQRE_S$ indicam as Somas de Quadrados Residuais, respectivamente, do modelo completo e do submodelo. Prove que esta estatística pode ser escrita numa forma alternativa, envolvendo os Coeficientes de Determinação do modelo completo (R_C^2) e do submodelo (R_S^2), isto é:

$$F = \frac{n - (p + 1)}{p - k} \cdot \frac{R_C^2 - R_S^2}{1 - R_C^2}.$$

- iv. A qualidade do ajustamento do modelo completo e do seu submodelo final difere significativamente ($\alpha = 0.05$)? Descreva, em pormenor, o teste efectuado para responder à questão.
- v. Analise os gráficos de resíduos abaixo, obtidos com o ajustamento do submodelo final.



- 2. No estudo anterior, existem observações referentes a genótipos de quatro regiões de origem da casta: Saragoça (Espanha), Sardenha (Itália), Toledo (Espanha) e Vaucluse (França). O técnico responsável pelo estudo considerou importante compreender a relação entre o teor de fenóis totais e o volume do mosto do bago nessas diferentes regiões de origem, decidindo, para tal, ajustar uma análise de covariância. A seguir, apresentam-se os resultados obtidos com essa abordagem:

```
> grenachelm3<-lm(fenois~volumebago*origem, data=grenache)
> summary(grenachelm3)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1321.3900   79.2017 16.684 < 2e-16  
volumebago   -6.9941    0.9120 -7.669 1.9e-13  
origemSardenha 91.8421  124.8782  0.735  0.463    
origemToledo -74.8751  111.4065 -0.672  0.502    
origemVaucluse 13.2992  100.0947  0.133  0.894    
volumebago:origemSardenha -0.3055   1.4123 -0.216  0.829    
volumebago:origemToledo     1.1915   1.3017  0.915  0.361    
volumebago:origemVaucluse -0.1942   1.1704 -0.166  0.868    
```

Residual standard error: 123.8 on 335 degrees of freedom
Multiple R-squared: 0.4238, Adjusted R-squared: 0.4118
F-statistic: 35.2 on 7 and 335 DF, p-value: < 2.2e-16

```
> grenachelm4<-lm(fenois~volumebago, data=grenache)
```

```
> anova(grenachelm4,grenachelm3)
Analysis of Variance Table
Model 1: fenois ~ volumebago
Model 2: fenois ~ volumebago * origem
  Res.Df   RSS Df Sum of Sq    F Pr(>F)
1     341 5349880
2     335 5131729  6    218151 2.3735 0.02931
```

- (a) Escreva a equação da recta ajustada referente à região de origem de "Sardenha".
- (b) Os declives das rectas de regressão para as origens "Sardenha" e "Saragoça" são significativamente diferentes ao nível $\alpha = 0.01$? Justifique a sua resposta, referindo apenas as hipóteses e conclusões do teste efetuado.
- (c) 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 $\alpha = 0.01$.
3. Um outro técnico sugeriu ajustar uma regressão linear simples para prever o logaritmo do teor de fenóis total a partir do logaritmo do volume do mosto do bago. O resultado do ajustamento foi o seguinte:
- ```
> grenachelm5<-lm(log(fenois)~log(volumebago), data=grenache)
> summary(grenachelm5)

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.30706 0.23636 43.61 <2e-16
log(volumebago) -0.84051 0.05346 -15.72 <2e-16

Residual standard error: 0.1714 on 341 degrees of freedom
Multiple R-squared: 0.4202, Adjusted R-squared: 0.4185
F-statistic: 247.2 on 1 and 341 DF, p-value: < 2.2e-16
```
- (a) Qual o valor do coeficiente de correlação entre as variáveis transformadas? Justifique.
- (b) Deduza a relação não linear entre o teor de fenóis total e o volume do mosto do bago que corresponde à regressão linear ajustada. Escreva a equação da curva ajustada.
- (c) Calcule a Soma de Quadrados Total referente ao modelo ajustado com variáveis transformadas.
4. Considere um modelo de regressão linear simples, ajustado com  $n$  pares de observações  $\{(x_i, y_i)\}_{i=1}^n$ .
- (a) Mostre que a média dos valores de  $y$  ajustados é igual à média dos valores observados  $y_i$ .
- (b) Considere o vector dos estimadores dos parâmetros da recta,  $\vec{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ . Diga quais são as expressões dos elementos da matriz de (co-)variâncias de  $\vec{\hat{\beta}}$ .

**II [6 valores]**

1. Considere um modelo ANOVA a um factor, com 3 níveis e 3 observações da variável resposta  $Y$  em cada nível do factor. Caracterize a matriz do modelo  $\mathbf{X}$ .
2. Num estudo sobre o efeito da área foliar na produção de girassol, foi efectuado uma experiência em que se avaliou o rendimento (kg/ha) para 3 níveis de desfolhação (controlo, sem desfolhação - D0, 33% de desfolhação - D33, 66% de desfolhação - D66), em 5 locais (L1, L2, L3, L4 e L5). Em cada local, assumindo condições homogéneas, o técnico responsável pelo estudo aplicou cada um dos 3 níveis desfolhação aleatoriamente a 3 parcelas, cada uma com 20 plantas.
- (a) Indique quais são as unidades experimentais do ensaio.
- (b) Esta experiência apresenta pseudo-repetições? Em caso afirmativo, identifique-as.
- (c) Quantas repetições existem em cada local?
- (d) Qual é o tipo de delineamento experimental conduzido em cada local?
- (e) Quantos fatores estão em estudo?
- (f) Descreva, em pormenor, o modelo de Análise de Variância considerado adequado para abordar o problema descrito.

- (g) Independentemente das suas respostas às alíneas anteriores, o técnico responsável pelo estudo decidiu conduzir a análise descrita seguidamente:

```
> girassol.aov<-aov(rend~local*desfol, data=girassol)
> summary(girassol.aov)
 Df Sum Sq Mean Sq F value Pr(>F)
local A 9037473 2259368 B 2.03e-05
desfol 2 2904191 C 6.710 0.0039
local:desfol 8 15717121 1964640 9.078 3.04e-06
Residuals 30 6492383 216413

> model.tables(girassol.aov,type="means")
Tables of means
...
local:desfol
 desfol
local D0 D33 D66
 L1 3955 1558 1420
 L2 2342 2369 1749
 L3 2854 3990 2176
 L4 1973 1829 2383
 L5 1515 1565 1810
```

- Complete a tabela ANOVA, indicando como obtém cada um dos valores em falta A, B e C.
- Pode afirmar-se que existem efeitos de interação local-tipo de desfolhação? Formalize e efectue o teste  $F$  adequado ao problema, ao nível  $\alpha=0.05$ .
- Para o estudo efetuado, são seguidamente apresentados dois intervalos de Tukey a 95% de confiança:

```
> TukeyHSD(girassol.aov)
 Tukey multiple comparisons of means
 95% family-wise confidence level
 Fit: aov(formula = rend ~ local * desfol, data = girassol)
 \$'local:desfol'
 diff lwr upr
L2:D0-L1:D0 -1613.33333 -3013.0314 -213.63522
L3:D0-L1:D0 -1101.33333 -2501.0314 298.36478
...

```

Interprete os resultados apresentados (não necessita de fazer quaisquer cálculos).

- O técnico responsável pela exploração agrícola do local L3 optou também por estudar o efeito da área foliar no número médio de sementes por planta por parcela para os 3 níveis de desfolhação (D0, D33 e D66). Ao realizar uma análise de variância clássica, o técnico não ficou satisfeito quando analisou os gráficos dos resíduos para validação dos pressupostos do modelo, adoptando de seguida uma abordagem não paramétrica. O número médio de sementes por planta por parcela e as respetivas ordenações por tipo de desfolhação (entre parênteses), bem como a soma de cada uma dessas ordenações (Soma(ord):  $R_i$ , para  $i = 1, \dots, 3$ ), encontram-se seguidamente descritos:

|           | DO         | D33        | D66       |
|-----------|------------|------------|-----------|
|           | 738.1 (3)  | 1137.2(7)  | 629.1 (1) |
|           | 949.9 (5)  | 1150.3 (8) | 630.3 (2) |
|           | 1054.1 (6) | 1166.5 (9) | 853.4 (4) |
| Soma(Ord) | (14)       | (24)       | (7)       |

Indique, justificando, qual é o teste não paramétrico que, neste contexto, deve ser realizado. Enuncie as respetivas hipóteses e, admitindo válida a distribuição assintótica da estatística do teste, indique que conclusão retira sobre o estudo efetuado ao nível de significância  $\alpha=0.05$ .

( $\chi^2_{0.05(1)} = 3.841$ ;  $\chi^2_{0.05(2)} = 5.991$ ;  $\chi^2_{0.05(3)} = 7.815$ ;  $\chi^2_{0.05(4)} = 9.488$ ).