

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2022-23
Primeira Chamada de Exame

6 Janeiro 2023

Duração: 3h

I [7 valores]

Num estudo de ampelografia realizado na variedade de videira Arinto, efectuaram-se 112 medições foliares independentes, em folhas selecionadas ao acaso, das seguintes características: comprimento da nervura principal (variável NP, em cm); comprimento da segunda nervura lateral principal (variável NP2, em cm); distância do ponto peciolar ao fundo do seio lateral superior (variável LS, em cm); distância do ponto peciolar ao fundo do seio lateral inferior (variável LI, em cm); comprimento da primeira nervura lateral com origem sobre a segunda nervura lateral principal (variável NL12, em cm). Eis a média amostral das variáveis estudadas, bem como as matrizes de (co-)variâncias e de correlações relativas aos dados:

Variável	Média
NP	12.04
LS	5.858
LI	5.200
NP2	8.250
NL12	5.633

```
> var(arintofolhas)
      NP      LS      LI      NP2      NL12
NP  0.25867380 0.10639599 0.06039359 0.15904242 0.08924807
LS  0.10639599 0.18286326 0.07089944 0.07013263 0.03667590
LI  0.06039359 0.07089944 0.11877072 0.05656135 0.04057893
NP2 0.15904242 0.07013263 0.05656135 0.14014332 0.09281780
NL12 0.08924807 0.03667590 0.04057893 0.09281780 0.10830030
```

```
> cor(arintofolhas)
      NP      LS      LI      NP2      NL12
NP  1.0000000 0.4891993 0.3445561 0.8353155 0.5332217
LS  0.4891993 1.0000000 0.4810884 0.4380974 0.2606170
LI  0.3445561 0.4810884 1.0000000 0.4384084 0.3577924
NP2 0.8353155 0.4380974 0.4384084 1.0000000 0.7534080
NL12 0.5332217 0.2606170 0.3577924 0.7534080 1.0000000
```

1. Qual será o melhor modelo de regressão linear simples para modelar o comprimento da segunda nervura lateral principal (variável NP2)? Justifique a sua resposta.
2. Um analista decidiu ajustar um modelo de regressão linear simples para modelar o comprimento da segunda nervura lateral principal (variável NP2) a partir da variável NP. Eis os resultados obtidos:

```
Call:
lm(formula = NP2 ~ NP, data = arintofolhas)
```

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  ????      0.46480   1.829   0.0701
NP           ????      0.03858  15.935  <2e-16
```

```
---
Residual standard error: 0.2067 on 110 degrees of freedom
Multiple R-squared:  ???? , Adjusted R-squared:  0.695
F-statistic: 253.9 on 1 and 110 DF, p-value: < 2.2e-16
```

- (a) Determine a equação da recta ajustada. Calcule e interprete o valor do coeficiente de determinação.
- (b) Calcule os valores da Soma dos Quadrados dos Resíduos (SQRE) e da Soma dos Quadrados da Regressão (SQR).

- (c) Construa um intervalo a 95% de confiança para o declive da recta de regressão populacional. Interprete o seu significado.
3. Mostre que, numa Regressão Linear Simples, baseada em n pares de observações $\{(x_i, y_i)\}_{i=1}^n$,
- a média dos valores ajustados de y é a média dos valores observados de y ;
 - a média dos resíduos é nula.
4. Uma *curva exponencial* é uma curva de equação $y = ce^{dx}$ (com $y > 0$; $c > 0$). Explique como lineariza este modelo. Indique a relação linear obtida, explicitando os seus parâmetros.

II [6 valores]

Foi seguidamente ajustado um modelo de regressão linear múltipla para modelar o comprimento da segunda nervura lateral principal (variável NP2) 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:

Call:

```
lm(formula = NP2 ~ ., data = arintofolhas)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.071629	0.366578	-0.195	0.845
NP	0.431024	0.037578	11.470	< 2e-16
LS	0.001853	0.042553	0.044	0.965
LI	0.097137	0.050278	1.932	0.056
NL12	0.464819	0.054003	8.607	7.08e-14

Residual standard error: 0.1539 on 107 degrees of freedom

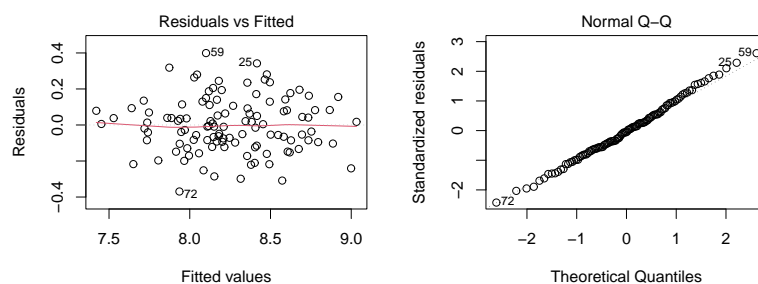
Multiple R-squared: 0.8371, Adjusted R-squared: 0.831

F-statistic: 137.5 on 4 and 107 DF, p-value: < 2.2e-16

```
> vcov(arinto.lm)
```

	(Intercept)	NP	LS	LI	NL12
(Intercept)	1.343793e-01	-7.906155e-03	3.994325e-05	-4.327602e-03	-0.0029717481
NP	-7.906155e-03	1.412105e-03	-6.231720e-04	-2.357934e-05	-0.0009438146
LS	3.994325e-05	-6.231720e-04	1.810766e-03	-8.371657e-04	0.0002140041
LI	-4.327602e-03	-2.357934e-05	-8.371657e-04	2.527846e-03	-0.0006442188
NL12	-2.971748e-03	-9.438146e-04	2.140041e-04	-6.442188e-04	0.0029163387

- Com um nível de significância $\alpha = 0.05$, o modelo ajustado difere significativamente do modelo nulo? Justifique a sua resposta.
- Comente os seguintes gráficos.



3. Será admissível considerar que os correspondentes valores populacionais dos coeficientes dos preditores NP e NL12 são iguais? Efectue um teste de hipóteses adequado para avaliar esta hipótese.
4. Para o modelo ajustado, caracterize a matriz \mathbf{X} do modelo.
5. Dado o Modelo de Regressão Linear Múltipla, e sabendo que o vector de estimadores dos parâmetros é dado por $\vec{\beta} = (\mathbf{X}^t\mathbf{X})^{-1} \mathbf{X}^t\vec{Y}$, prove que a matriz de (co-)variâncias do vector de estimadores é dada por $V[\vec{\beta}] = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}$.
6. Um algoritmo de exclusão sequencial produziu o submodelo final com os seguintes resultados:

```
Call:
lm(formula = NP2 ~ NP + NL12, data = arintofolhas)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.12558     0.35857   0.350   0.727
NP           0.44593     0.03433  12.989 < 2e-16
NL12        0.48956     0.05306   9.227 2.48e-15
---
Residual standard error: 0.1556 on 109 degrees of freedom
Multiple R-squared:  0.8303, Adjusted R-squared:  0.8272
F-statistic: 266.7 on 2 and 109 DF,  p-value: < 2.2e-16
```

- (a) Indique, justificando, qual foi a ordem de exclusão das variáveis LI e LS.
- (b) Teste formalmente, com nível de significância $\alpha = 0.05$, se o ajustamento deste submodelo difere significativamente do modelo original com 4 preditores. Comente o resultado obtido.

III [7 valores]

1. Num estudo sobre melhoramento de variedades tradicionais de alho foram estudadas no mesmo local 11 variedades, designadas VA, VB, VC, VD, VE, VF, VG, VH, VI, VJ e VL, em 3 anos, designados A2007, A2008 e A2009. No ensaio realizado em cada ano, foram aleatoriamente associadas 6 parcelas a cada variedade e avaliou-se o peso do bolbo por parcela (g/bolbo). As médias amostrais por combinação variedade:ano são indicadas seguidamente:

```
variedade:ano
      ano
variedade A2007 A2008 A2009
VA 15.52 23.34 17.39
VB 22.18 28.45 19.94
VC 15.63 20.86 16.92
VD 20.50 23.38 18.20
VE 18.18 21.18 16.92
VF 16.65 18.09 12.55
VG 21.48 23.47 17.23
VH 17.29 18.79 11.80
VI 24.67 34.16 19.69
VJ 19.85 20.07 22.48
VL 17.83 35.79 38.23
```

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

```

> summary(alho.aov)
              Df  Sum Sq Mean Sq  F value  Pr(>F)
variedade     10   3577    ???      11.63  5.25e-15
ano           ???   1183    591.3     19.22  3.13e-08
variedade:ano 20   2024    101.2     ???    1.31e-05
Residuals     ???   5075    30.8

```

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

- (c) Quais os tipos de efeitos que devem ser considerados significativos ao nível $\alpha=0.05$? (caso necessite de vários testes, descreva um em pormenor e enuncie as conclusões para os restantes).
- (d) O maior peso médio do bolbo foi observado na variedade VL no ano 2009. Este peso médio amostral do bolbo é significativamente diferente do peso médio do bolbo observado nesse ano de 2009 para as restantes variedades, ao nível $\alpha=0.05$?
2. O agrónomo responsável pelo programa de melhoramento descrito na pergunta 1 decidiu avaliar no ano de 2009 o número de dentes por bolbo nas variedades VB, VH, VJ e VL. Dada a natureza dos dados, para estudar se a variedade afecta o número de dentes por bolbo, decidiu fazer o Teste de Kruskal Wallis. A ordenação do número de dentes médio por bolbo em cada uma das 6 parcelas avaliadas para as 4 variedades encontra-se seguidamente indicada:

Var	Ord	Var	Ord	Var	Ord	Var	Ord
VB	9	VH	17	VJ	24	VL	16
VB	2	VH	8	VJ	23	VL	19
VB	4	VH	10	VJ	18	VL	14
VB	6	VH	3	VJ	12	VL	22
VB	5	VH	1	VJ	21	VL	20
VB	7	VH	11	VJ	13	VL	15
Soma(Ord)	33		50		111		106

O que conclui sobre o estudo efectuado ao nível $\alpha=0.05$? Descreva em pormenor o teste realizado.

Notas:

- (a) Estatística do teste de Kruskal Wallis: sob H_0 ,

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

- (b) $\chi_{0.05(3)}^2 = 7.815$.