

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2014/15
PRIMEIRA CHAMADA de EXAME

14 Janeiro 2015

Duração: 3h30

I [2,5 valores]

Um estudo analisou a adaptabilidade de pinheiros mansos de seis diversas proveniências (Marrocos, Alcácer do Sal, Itália, Turquia, Viseu e Espanha) nas condições edafo-climáticas de Alcácer do Sal. O estudo consistiu em plantar 150 árvores de cada proveniência e, ao fim de dois anos, contar o número de árvores mortas e sobreviventes, a fim de avaliar se as probabilidades de mortalidade são idênticas para as várias proveniências. Eis os resultados obtidos.

	mortas	sobreviventes
Marrocos	30	120
Alcácer, Portugal	36	114
Itália	23	127
Turquia	67	83
Viseu, Portugal	21	129
Espanha	36	114
Total	213	687

1. Descreva formalmente um teste χ^2 adequado para responder ao objectivo do estudo. Sabendo que o valor da estatística de Pearson calculada foi $X_{calc}^2 = 51.2762$, indique e discuta as suas conclusões. Discuta a validade da distribuição assintótica da estatística do teste.
2. Identifique a proveniência que mais contribuiu para o valor calculado da estatística do teste, e calcule a respectiva contribuição para o valor de X_{calc}^2 indicado na alínea anterior. Comente.

II [9 valores]

A análise química de 39 recolhas do mosto da casta Moscatel Graúdo registou valores de seis características associadas ao aroma, juntamente com o rendimento (em kg/planta), grau álcool provável (variável grau) e acidez (g/l de ácido tartárico) da respectiva produção. Eis a média e variância amostrais das variáveis observadas, bem como a respectiva matriz de correlações.

media	variância		linalol	nerol	geraniol	hexenol	hexanal	2fenil.	rendim.	grau	acidez
0.477818	0.0278082	linalol	1.000	-0.125	0.463	0.340	0.060	0.099	-0.294	0.445	0.027
0.057077	0.0006659	nerol	-0.125	1.000	0.318	-0.201	-0.165	0.050	0.227	-0.326	-0.111
0.383831	0.0042863	geraniol	0.463	0.318	1.000	0.293	0.011	0.058	-0.144	0.023	-0.018
0.317028	0.0080113	hexenol	0.340	-0.201	0.293	1.000	0.112	0.029	-0.117	0.380	-0.170
0.117162	0.0061367	hexanal	0.060	-0.165	0.011	0.112	1.000	0.692	0.017	0.047	0.079
0.124846	0.0048344	2feniletanol	0.099	0.050	0.058	0.029	0.692	1.000	0.075	-0.028	0.068
8.610233	2.2003850	rendimento	-0.294	0.227	-0.144	-0.117	0.017	0.075	1.000	-0.293	-0.048
8.757108	0.1416529	grau	0.445	-0.326	0.023	0.380	0.047	-0.028	-0.293	1.000	-0.092
4.230297	0.0886225	acidez	0.027	-0.111	-0.018	-0.170	0.079	0.068	-0.048	-0.092	1.000

Dada a dificuldade em analisar as características aromáticas, seria útil poder prever pelo menos algumas a partir das restantes variáveis observadas. Foi ajustada uma regressão linear múltipla para tentar prever os teores de hexanal (**Atenção:** não confundir com outra variável de nome semelhante!) a partir dos restantes 8 preditores, tendo sido obtidos os seguintes resultados:

```
> summary(lm(hexanal ~ . , data=moscatel2014))
```

Coefficients:

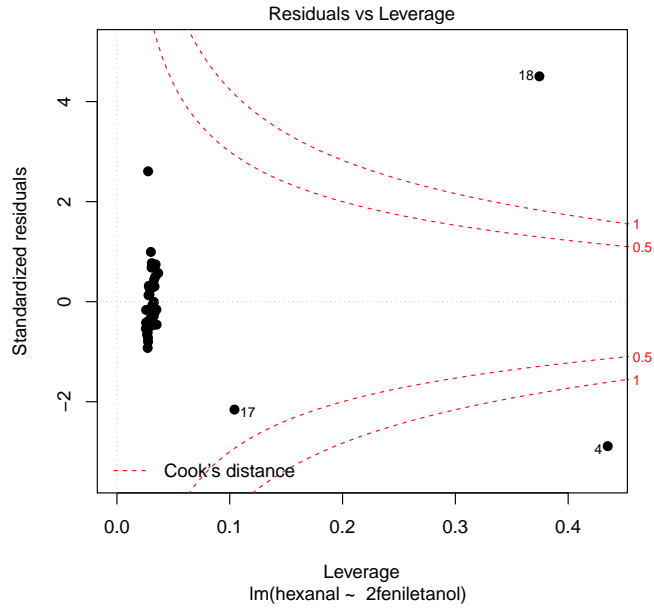
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0330805	0.3544070	-0.093	0.926
linalol	-0.0436050	0.0787920	-0.553	0.584
nerol	-0.6497245	0.4631393	-1.403	0.171
geraniol	0.0784129	0.2000816	0.392	0.698
hexenol	0.0517223	0.1307570	0.396	0.695
2feniletanol	0.7943703	0.1435978	5.532	5.19e-06
rendimento	0.0004273	0.0072603	0.059	0.953
grau	0.0038168	0.0327531	0.117	0.908
acidez	0.0060019	0.0343731	0.175	0.863

Residual standard error: 0.06056 on 30 degrees of freedom

Multiple R-squared: 0.5281, Adjusted R-squared: 0.4023

F-statistic: 4.197 on 8 and 30 DF, p-value: ???

1. Interprete e comente o valor do coeficiente de determinação.
2. Efectue o teste de ajustamento global ($\alpha = 0.05$) e comente as suas conclusões.
3. Será possível afirmar que o coeficiente da variável **grau** de álcool provável, na equação de regressão populacional, é inferior a 0.005, não dando o benefício da dúvida a esta afirmação?
4. Um algoritmo de exclusão sequencial baseado no Critério de Informação de Akaike (AIC) optou por um submodelo com apenas dois preditores: **nerol** e **2feniletanol**.
 - (a) Com base na informação disponível até aqui, justifique que o coeficiente de determinação deste submodelo de dois preditores está entre 0.4788 e 0.5281.
 - (b) Sabendo que o AIC deste submodelo com dois preditores é -222.23 , calcule o respectivo valor de R^2 .
 - (c) Calcule o valor do R^2 modificado deste submodelo. Compare e comente os valores dos R^2 e R^2 modificados, quer deste submodelo, quer do modelo original. **Nota:** Caso não tenha resolvido a alínea anterior, considere que o R^2 do submodelo é 0.52.
5. Foi ajustada uma regressão linear simples da variável **hexanal** sobre o preditor **2feniletanol**.
 - (a) Determine, justificando, a equação da respectiva recta ajustada.
 - (b) Discuta o seguinte gráfico dos resíduos (internamente) estandardizados contra os valores do efeito alavanca.



- (c) Procurando compreender o gráfico mostrado no ponto anterior, calculou-se o coeficiente de correlação entre as variáveis `hexanal` e `2feniletanol`, mas omitindo as duas observações com a maior distância de Cook, tendo-se obtido o valor $r = 0.21908$. Discuta as implicações que retira do conjunto da informação apresentada sobre esta regressão linear simples.

III [4,5 valores]

Num estudo de árvores de diferentes variedades de oliveiras, decidiu-se analisar as dimensões de folhas adultas, a fim de avaliar a sua eventual relação com o escoamento da pluviosidade ao longo do tronco. Foram analisadas 6 variedades de oliveira: azeiteira, blanqueta, cordovil, maçanilha, negrita e picual. Duas árvores de cada variedade haviam já sido estudadas no que respeita ao escoamento da pluviosidade e em cada uma delas foram amostradas 10 folhas ao acaso. Efectuou-se uma análise de variância para analisar a existência de efeitos, quer de variedades, quer das árvores de cada variedade, nas larguras das folhas. As larguras médias global, por variedade e por árvore de cada variedade, obtidas foram as seguintes.

```
> model.tables(aov(largura~variedade/arvore,data=folhas),type="means")
Tables of means
Grand mean
11.78236

variedade
azeiteira blanqueta cordovil macanilha negrinha picual
  11.793   14.450   11.005   12.031   10.224   11.191

variedade:arvore
      arvore
variedade  1      2
azeiteira 12.865 10.721
blanqueta 14.187 14.713
cordovil  11.559 10.451
```

```

macanilha 12.043 12.020
negrinha  10.348 10.100
picual    11.179 11.203

```

Foi ajustado um modelo ANOVA com os seguintes resultados.

```

> summary(aov(largura~variedade/arvore,data=folhas))
              Df Sum Sq Mean Sq F value    Pr(>F)
variedade      5 211.21   42.24  14.494     ???
variedade:arvore 6  30.81    5.13   1.762    0.114
Residuals     108 314.75    2.91

```

1. Descreva detalhadamente o modelo ANOVA ajustado e discuta a opção tomada para lidar com os dois factores presentes na experiência.
2. Existem efeitos de variedade significativos ao nível $\alpha = 0.05$? Comente a sua conclusão.
3. Com base num teste de Tukey, para que variedades é que podem ser consideradas significativamente diferentes as larguras médias das folhas do par de árvores observadas ($\alpha = 0.05$)? Comente as suas conclusões, tendo também em conta as conclusões dum teste F aos efeitos de árvore (para o mesmo valor de α).
4. Decidiu-se usar o teste de Bartlett para estudar a homogeneidade de variâncias das folhas de cada árvore observada. Descreva brevemente o teste ($\alpha = 0.05$) e comente as conclusões a que conduzem os resultados obtidos, sabendo que o valor calculado da estatística é $K_{calc}^2 = 18.1765$.

IV [4 valores]

Considere o modelo linear, *sem preditores* (o Modelo Nulo), dado pela equação $Y_i = \beta_0 + \epsilon_i, \forall i = 1, \dots, n$, e pelos pressupostos $\epsilon_i \cap \mathcal{N}(0, \sigma^2), \forall i$, e independência dos $\{\epsilon_i\}_{i=1}^n$.

Usando a notação matricial na formulação do modelo, a matriz do modelo \mathbf{X} terá uma única coluna, composta por uns, ou seja, $\mathbf{X} = \mathbf{1}_n$.

1. Mostre que a matriz de projecção ortogonal sobre $\mathcal{C}(\mathbf{X})$, o espaço das colunas de \mathbf{X} , será, neste caso, dada por $\mathbf{H} = \frac{1}{n}\mathbf{J}$, onde \mathbf{J} é a matriz $n \times n$ cujos elementos são todos 1.
2. Identifique os valores ajustados da variável resposta, \hat{y}_i , produzidos por este modelo nulo.
3. A partir da fórmula geral para o vector $\hat{\boldsymbol{\beta}}$ dos estimadores dos parâmetros dum modelo linear, determine o estimador de mínimos quadrados de β_0 .
4. Determine os valores de SQR e $SQRE$ neste modelo. Comente.
5. Utilize os resultados da alínea anterior para mostrar que a estatística do teste F parcial comparando um modelo de regressão linear com p preditores e o submodelo Nulo (sem preditores) é igual à estatística do teste F de ajustamento global do modelo completo.