

INSTITUTO SUPERIOR DE AGRONOMIA  
ESTATÍSTICA E DELINEAMENTO – 2015/16

5 de Janeiro de 2016

Primeira Chamada de Exame

Duração: 3h30

I [2,5 valores]

Num estudo sobre a susceptibilidade de variedades de oliveira a fungos causadores de gafa, inocularam-se azeitonas de 4 variedades de oliveira (Galega, Cobrançosa, Negrinha e Picual) com um isolado de *Colletotrichum acutatum* (PT170). Após 24 horas, escolheram-se ao acaso 600 esporos presentes nos frutos de cada variedade, tendo sido registados o número de esporos não germinados, o número de esporos germinados sem formação de apressórios e o número de esporos germinados com formação de apressórios. Os resultados obtidos são indicados na tabela seguinte.

	Não germinados	Germ. sem apressórios	Germ. com apressório	Total
Galega	197	303	100	600
Cobrançosa	267	216	117	600
Negrinha	419	128	53	600
Picual	391	189	20	600
Total	1274	836	290	2400

1. Pretende-se saber se cada possível resultado da germinação tem, em todas as variedades, igual probabilidade. Qual o tipo de teste adequado? Justifique, e estime o número esperado de observações em cada entrada da tabela, caso seja verdade a hipótese nula correspondente.
2. Descreva o teste e as suas conclusões ( $\alpha = 0.05$ ), sabendo que o valor calculado da estatística do teste é  $X_{\text{calc}}^2 = 259.7168$ . Em particular, diga, justificando, se considera válida a distribuição assintótica da estatística do teste.
3. A parcela que mais contribui para o valor calculado da estatística de teste corresponde à variedade Galega e resultado “não germinado”. Calcule e comente o valor desta parcela.

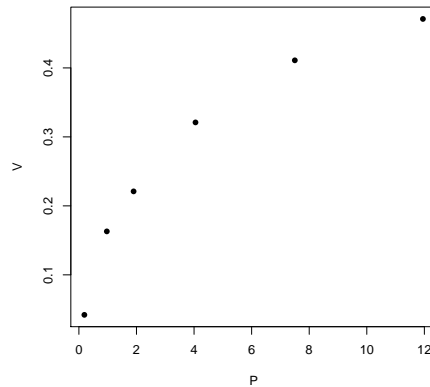
II [3 valores]

O volume  $V$  de um gás adsorvido num sólido, em situação de equilíbrio, pode relacionar-se com a pressão aplicada  $P$  através da isotérmica de Langmuir cuja equação, para uma adsorção não dissociativa, é  $V = V_m \frac{KP}{KP + 1}$ . As constantes  $V_m$  (chamada volume da monocamada, ou seja, volume máximo adsorvido) e  $K$  (chamada constante de equilíbrio de adsorção/desorção) variam consoante o contexto da experiência.

1. Mostre que as transformações  $x = \frac{1}{P}$  e  $y = \frac{1}{V}$  linearizam a equação, isto é, que  $\frac{1}{V}$  pode ser escrita como uma função linear de  $\frac{1}{P}$ .
2. Num estudo sobre a adsorção de  $H_2$  em cobre em pó, a  $25^\circ\text{C}$ , obtiveram-se os seguintes dados e nuvem de pontos:

$P$ (Torr)	$V$ ( $cm^3$ )
0.19	0.042
0.97	0.163
1.90	0.221
4.05	0.321
7.50	0.411
11.95	0.471

(sendo 1 Torr = 133.3 Pa)



Eis os resultados de ajustar uma recta de regressão aos dados transformados  $x = \frac{1}{P}$  e  $y = \frac{1}{V}$ , bem como alguns indicadores relativos às variáveis transformadas  $x$  e  $y$ :

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.98780	0.11187	17.77	5.89e-05	$\bar{y} = 7.02348$
x	4.14782	0.05077	81.69	1.35e-07	$Cov_{xy} = 16.82589$
Residual standard error: 0.2287 on 4 degrees of freedom					
Multiple R-squared: 0.9994, Adjusted R-squared: 0.9993					

- Determine os valores estimados das constantes  $V_m$  e  $K$ , bem como a equação da curva de Langmuir ajustada à nuvem de pontos acima indicada.
- Calcule um intervalo de predição (a 95%) para o recíproco do volume ( $\frac{1}{V}$ ) correspondente a uma observação efectuada à pressão  $P = 10$  Torr. Calcule o correspondente intervalo de predição para o volume (em  $cm^3$ ). Comente.

### III [6,5 valores]

No quadro de um estudo sobre transpiração em olival foram medidos rigorosamente no terreno valores diários de transpiração (variável  $Tr$ , em mm/dia) ao longo de aproximadamente 3 anos. Numa tentativa de relacionar transpiração com variáveis obtidas por observação da terra por satélite, foi ajustado um modelo de regressão linear cujos preditores são reflectâncias à superfície em várias bandas de comprimento de onda (variáveis BLUE, GREEN, RED, NIR, SWIR1 e SWIR2, adimensionais, com valores entre 0 e 1) para sensores Landsat, o índice de vegetação NDVI (adimensional, com valores entre -1 e 1) derivado das bandas RED e NIR, e a temperatura (variável TIR, em graus centígrados) derivada da banda térmica dos sensores. Eis a matriz de correlações das observações e as médias e desvios padrão de cada variável:

	Tr	BLUE	GREEN	RED	NIR	SWIR1	SWIR2	NDVI	TIR		média	desv. padrão	
Tr	1.00	0.86	0.89	0.89	0.47	0.86	0.87	-0.83	0.83		Tr	1.3360	0.90107068
BLUE	0.86	1.00	0.98	0.97	0.60	0.91	0.89	-0.86	0.81		BLUE	0.0770	0.01762375
GREEN	0.89	0.98	1.00	0.98	0.69	0.94	0.92	-0.83	0.86		GREEN	0.1073	0.02440693
RED	0.89	0.97	0.98	1.00	0.62	0.97	0.95	-0.89	0.90		RED	0.1264	0.03308019
NIR	0.47	0.60	0.69	0.62	1.00	0.65	0.58	-0.22	0.51		NIR	0.2793	0.02962072
SWIR1	0.86	0.91	0.94	0.97	0.65	1.00	0.99	-0.85	0.92		SWIR1	0.2872	0.06304776
SWIR2	0.87	0.89	0.92	0.95	0.58	0.99	1.00	-0.87	0.90		SWIR2	0.2105	0.06055228
NDVI	-0.83	-0.86	-0.83	-0.89	-0.22	-0.85	-0.87	1.00	-0.86		NDVI	0.3853	0.09287009
TIR	0.83	0.81	0.86	0.90	0.51	0.92	0.90	-0.86	1.00		TIR	25.8878	8.93238028

Eis o ajustamento dum regressão linear múltipla de Tr sobre as restantes variáveis:

	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	-5.92560	2.05224	-2.887	0.00741
BLUE	-14.11886	20.26174	-0.697	0.49166
GREEN	64.48073	21.24687	3.035	0.00515
RED	31.97623	20.36151	1.570	0.12755
NIR	-36.21879	10.34842	-3.500	0.00158
SWIR1	-7.61723	14.60201	-0.522	0.60601
SWIR2	10.21927	9.83247	1.039	0.30754
NDVI	16.55497	5.46086	3.032	0.00519
TIR	0.04466	0.02403	1.858	0.07365

Residual standard error: 0.3179 on 28 DF  
Multiple R-squared: 0.9032, Adjusted R-squared: 0.8755  
F-statistic: 32.64 on 8 and 28 DF, p-value: 3.262e-12

1. Com base em quantas observações foi ajustado o modelo? Justifique.
2. Como interpretar, no contexto do problema, o valor da estimativa  $-14.11886$  associada à variável BLUE, tendo em conta que essa variável só toma valores no intervalo  $]0.04, 0.12[$ ?
3. Teste formalmente ( $\alpha = 0.05$ ) se se pode concluir que um aumento da temperatura está associado a um aumento da transpiração, para valores iguais nas restantes variáveis. Exija o ónus da prova para esta afirmação.
4. Foi escolhido um submodelo do modelo original, apenas com quatro preditores: GREEN, NIR, NDVI e TIR. Foram obtidos os seguintes resultados:

Coefficients:

	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	-3.64659	1.21674	-2.997	0.00523 **
GREEN	66.82066	13.29883	5.025	1.85e-05 ***
NIR	-24.96892	6.95825	-3.588	0.00110 **
NDVI	9.04922	3.29300	2.748	0.00977 **
TIR	0.05009	0.01710	2.930	0.00620 **

---  
Residual standard error: 0.3384 on 32 degrees of freedom  
Multiple R-squared: 0.8746, Adjusted R-squared: 0.859  
F-statistic: 55.82 on 4 and 32 DF, p-value: 5.574e-14

- (a) Teste formalmente ( $\alpha = 0.05$ ) se este submodelo tem um ajustamento significativamente diferente do modelo completo. Comente as suas conclusões.
- (b) Com base na informação disponível, discuta possíveis razões pelas quais o facto das variáveis BLUE e RED terem sido excluídas deste submodelo tenha pouco efeito sobre o valor do coeficiente de determinação.
- (c) Calcule o resíduo da sétima observação, sabendo que a sua transpiração observada foi  $y_7 = 0.8437172$  e os seus valores nos quatro preditores destes submodelo foram GREEN=0.0791, NIR=0.1839, NDVI=0.3399 e TIR=17.05.

#### IV [4 valores]

No âmbito de um estudo realizado em Mangualde sobre o rendimento de diferentes genótipos da casta de videira Alfrocheiro, foram ensaiados cada um de  $a$  genótipos nos anos 1994, 1996, 1997, 1999 e 2000. Foi efectuada uma Análise de Variância para estudar os efeitos dos factores **genotipo** e **ano** sobre a variável resposta **rendimento** (medida em kg/planta) em cada uma das  $n$  parcelas observadas, tendo sido obtida a seguinte tabela resumo:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
genotipo	11	15.31	1.39	2.945	0.000904
ano	4	208.25	52.06	???	???
genotipo:ano	44	22.29	0.51	1.072	0.354088
Residuals	420	198.44	???		

1. Descreva pormenorizadamente o tipo de delineamento e o modelo ANOVA correspondente, sabendo que o delineamento utilizado era equilibrado. Em particular, indique com base na informação disponível e justificando, qual o número de diferentes genótipos que foram usados no estudo e qual o número de observações efectuadas, em cada ano, para cada genótipo.
2. Qual o valor estimado da variância dos erros aleatórios do modelo, e quais as suas unidades de medida?
3. Utilize testes  $F$  para indicar quais os tipos de efeitos que devem ser considerados significativos ao nível  $\alpha=0.05$  (descreva um teste em pormenor e sintetize os restantes).
4. Um analista decidiu que bastava analisar estes dados com uma ANOVA a um factor, o factor **genotipo**. Construa, justificando, a tabela ANOVA obtida pelo referido analista. As conclusões do teste aos efeitos do factor **genotipo** obtidas pelo analista são coerentes com as obtidas inicialmente? Discuta os seus resultados.

#### V [4 valores]

1. Seja dada uma matriz rectangular  $\mathbf{X}_{n \times m}$ . Mostre que a matriz de projecção ortogonal sobre o subespaço das colunas de  $\mathbf{X}$ ,  $\mathcal{C}(\mathbf{X})$ , é simétrica e idempotente.
2. Considere o modelo ANOVA para um delineamento a um factor, equilibrado. Tratando-se dum modelo linear, é possível calcular o respectivo coeficiente de determinação  $R^2$ , embora não seja usual fazê-lo em modelos ANOVA.
  - (a) Indique condições equivalentes ao valor extremo  $R^2=0$ , envolvendo as médias amostrais de nível e da totalidade das observações. Interprete essa situação em termos do teste  $F$  da ANOVA.
  - (b) Indique condições equivalentes ao valor extremo  $R^2=1$ , envolvendo as variâncias amostrais de nível. Interprete essa situação em termos do teste  $F$  da ANOVA.
3. Considere o contexto duma regressão linear simples.
  - (a) Prove que a Soma de Quadrados da Regressão é dada por  $SQR = b_1^2 (n-1) s_x^2$ , onde  $b_1$  é o declive da recta ajustada e  $s_x^2$  é a variância amostral das observações do preditor.
  - (b) Prove que a Soma de Quadrados da Regressão é dada por  $SQR = r_{xy}^2 (n-1) s_y^2$ , onde  $r_{xy}$  é o coeficiente de correlação entre as duas variáveis do modelo e  $s_y^2$  é a variância amostral das observações da variável resposta.