INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO – 2015/16

5 de Janeiro de 2016 **Segundo Teste** **Duração: 2h30**

I [10 valores]

No quadro de um estudo sobre transpiração em olival foram medidos rigorosamente no terreno valores diários de transpiração (variável `Tr`, em mm/dia) em dias escolhidos ao acaso ao longo de 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:

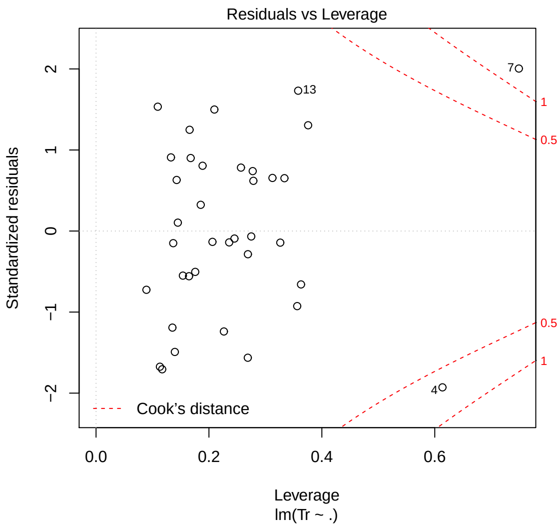
	Tr	BLUE	GREEN	RED	NIR	SWIR1	SWIR2	NDVI	TIR		média	desv.padrão
Tr	1.00	0.86	0.89	0.89	0.47	0.86	0.87	-0.83	0.83	Tr	1.3360	0.90107068
BLUE	0.86	1.00	0.98	0.97	0.60	0.91	0.89	-0.86	0.81	BLUE	0.0770	0.01762375
GREEN	0.89	0.98	1.00	0.98	0.69	0.94	0.92	-0.83	0.86	GREEN	0.1073	0.02440693
RED	0.89	0.97	0.98	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 duma 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. Com base no modelo completo foi produzido o seguinte gráfico de diagnósticos. Descreva o gráfico e discuta as principais conclusões da sua análise.



5. 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 neste submodelo a distância de Cook da observação 7, sabendo que o seu resíduo usual é $e_7 = -0.1333$ e que o seu efeito alavanca é $h_{7,7} = 0.39051$. Comente, tendo também em atenção o gráfico do ponto 4, criado a partir do modelo completo.

II [6 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 $\alpha = 0.05$ (caso necessite de vários testes, descreva um 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.

III [4 valores]

1. Seja dada uma matriz rectangular $\mathbf{X}_{n\times m}$.
   (a) Mostre que a projecção ortogonal de qualquer vector $\vec{z} = \mathbf{X}\vec{a} \in \mathcal{C}(\mathbf{X})$ sobre o próprio subespaço $\mathcal{C}(\mathbf{X})$ deixa o vector $\vec{z}$ invariante.
   (b) Mostre que a matriz de projecção ortogonal sobre $\mathcal{C}(\mathbf{X})$ 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.