

INSTITUTO SUPERIOR DE AGRONOMIA
ESTATÍSTICA E DELINEAMENTO – 2014/15

3 de Novembro de 2014

PRIMEIRO TESTE

Duração: 2h00

I [4 valores]

Pretende-se determinar se a proporção de frutos com defeito, em 4 diferentes variedades tradicionais de tomate, é igual. Foram escolhidos para avaliação 120 frutos na variedade 41A, 244 na variedade 40C, 219 na variedade 15A e 181 na variedade 35. Os resultados são dados na seguinte tabela.

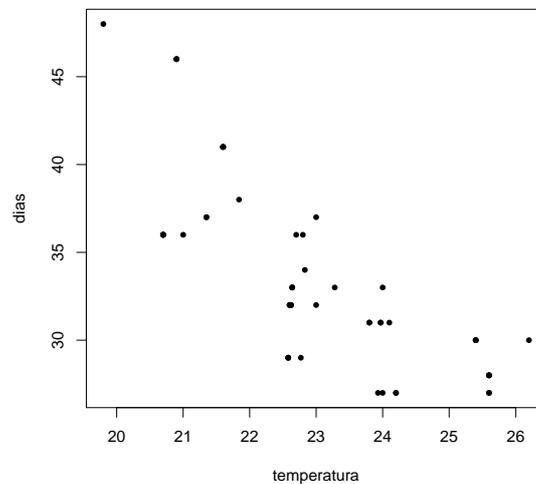
	frutos defeituosos	frutos com valor comercial	total
variedade 41A	69	51	120
variedade 40C	113	131	244
variedade 15A	77	142	219
variedade 35	135	46	181

1. Descreva pormenorizadamente um teste baseado na estatística de Pearson, adequado à questão que foi colocada, sabendo que o valor calculado da estatística é $X^2_{calc} = 66.3941$ e admitindo a validade das condições de Cochran. Comente as conclusões.
2. Qual a contribuição da variedade 35 para o valor calculado da estatística? Comente.
3. Justifique a validade da distribuição assintótica da estatística do teste, acima utilizada.

II [11 valores]

Num estudo no âmbito do controlo de pragas, pretende-se determinar se existe uma relação entre a temperatura ambiente (variável **temperatura**, em graus centígrados) e o número de dias que decorre entre a postura de ovos por um predador e a emergência de novos adultos (variável **dias**). Foi utilizada uma sala de criação do predador, que permite controlar a temperatura. No final do estudo foram calculados alguns indicadores simples sobre os resultados e foi construída a nuvem de pontos das observações, como indicado de seguida.

temperatura	dias
Min. :19.80	Min. :27.00
1st Qu.:21.60	1st Qu.:29.75
Median :22.64	Median :32.00
Mean :22.96	Mean :33.48
3rd Qu.:24.00	3rd Qu.:36.00
Max. :26.20	Max. :48.00
Var. :2.7046	Var. :28.4360



1. Ajustou-se uma regressão linear de dias sobre temperatura, com os seguintes resultados.

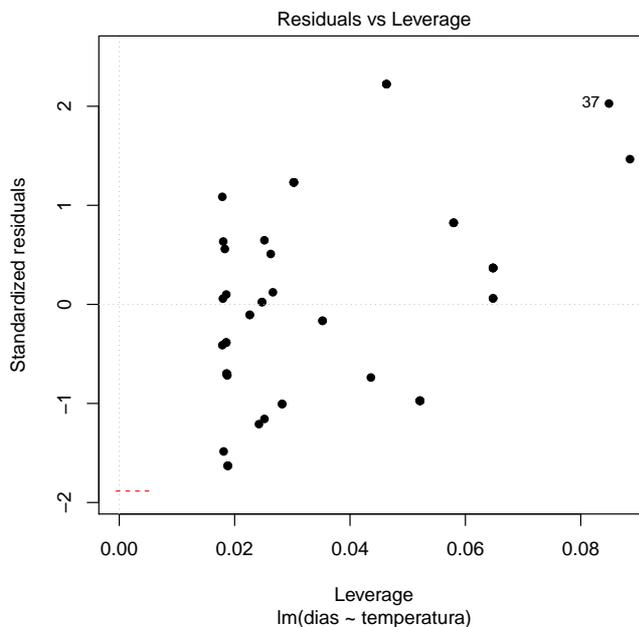
```
> summary(lm(dias ~ temperatura, data=elisabete2))
[...]
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	91.5285	6.3582	14.395	< 2e-16
temperatura	-2.5284	0.2763	-9.152	1.43e-12

Residual standard error: 3.369 on 54 degrees of freedom
Multiple R-squared: 0.6080, Adjusted R-squared: 0.6008
F-statistic: ??? on 1 and 54 DF

- (a) Diga, justificando, qual o valor do coeficiente de correlação amostral entre as variáveis dias e temperatura.
- (b) Discuta a qualidade do modelo ajustado. Em particular, interprete o valor do coeficiente de determinação, e efectue o teste F de ajustamento global do modelo, indicando como obtém o valor da estatística. Discuta as suas conclusões.
- (c) Indique uma estimativa da variância dos erros aleatórios do modelo. Indique as respectivas unidades de medida.
- (d) Calcule um intervalo de predição (95%) para o número de dias entre postura do ovo e emergência dum novo adulto, associado a uma observação com temperatura 22.9°C .
- (e) Com base no seguinte gráfico de resíduos internamente estandardizados vs. valores dos efeitos alavanca, determine um valor aproximado da distância de Cook da observação 37 (assinalada no gráfico). Comente esse valor, tendo em conta o significado das distâncias de Cook.



2. Procurando um modelo melhor, ajustou-se seguidamente uma regressão linear *dos logaritmos (naturais) de dias sobre os logaritmos (naturais) de temperaturas*, com os seguintes resultados.

```

> summary(elisabete2.lm2)
Call: lm(formula = log(dias) ~ log(temperatura), data = elisabete2)
[...]
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.8404     0.5371  16.460 < 2e-16
log(temperatura) -1.7058     0.1715  -9.947 8.25e-14
---
Residual standard error: 0.09072 on 54 degrees of freedom
Multiple R-squared:  0.6469,    Adjusted R-squared:  0.6404
F-statistic: 98.94 on 1 and 54 DF,  p-value: 8.253e-14

```

- (a) Deduza a relação *não linear* de fundo entre dias e temperatura, correspondente ao modelo agora ajustado. Escreva a equação não linear, resultante do ajustamento, que relaciona as variáveis originais.
- (b) É admissível afirmar que o número de dias entre postura e emergência é inversamente proporcional ao quadrado da temperatura? Responda usando um teste de hipóteses adequado ($\alpha = 0.05$).
- (c) Determine um intervalo de predição (95%) para uma observação do número de dias entre postura do ovo e emergência de novo adulto, associado a uma temperatura de $22.9^{\circ}C$, sabendo que o logaritmo desta temperatura coincide com a média dos logaritmos das temperaturas observadas. Compare com o intervalo obtido na alínea 1d) e comente.

III [5 valores]

Foi provado nas aulas teóricas que, dado o Modelo de Regressão Linear Simples, os resíduos usuais E_i têm a seguinte distribuição: $E_i \cap \mathcal{N}(0, \sigma^2(1-h_{ii}))$, sendo $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n-1)s_x^2}$. Partindo deste resultado, responda às seguintes questões.

(NOTA: na resposta a cada alínea pode usar também os resultados de alíneas anteriores.)

1. Mostre que a variância dum resíduo é o valor esperado do seu quadrado: $V[E_i] = E[E_i^2]$.
2. Mostre que $E[SQRE] = (n-2)\sigma^2$, sendo $SQRE$ a soma de quadrados dos resíduos.
3. Justifique que o Quadrado Médio Residual, $QMRE$, é um estimador centrado da variância σ^2 dos erros aleatórios do modelo.
4. Sabendo que $Cov[E_i, \hat{Y}_i] = 0$, mostre que a variância do i -ésimo valor ajustado de Y é dada por $V[\hat{Y}_i] = \sigma^2 h_{ii}$.
5. Mostre que a distribuição do i -ésimo valor ajustado é $\hat{Y}_i \cap \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2 h_{ii})$.