INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO – 2018-19

2 de Novembro de 2018 PRIMEIRO TESTE Duração: 2h30

I [3,5 valores]

Num ensaio sobre pereiras Rocha, foi testado o sistema de condução Tatura. Foi observado o número de gomos florais em 80 pereiras, tendo-se obtido as contagens abaixo indicadas. Pretende-se saber se é possível considerar que o número de gomos por árvore segue uma lei Poisson, com valor esperado 7.

No. de gomos	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. de árvores	0	1	3	9	6	15	4	8	7	6	4	6	3	4	3	1

1. É possível usar a tabela de contagens acima indicada para efectuar um teste de ajustamento à distribuição Poisson, baseado na estatística de Pearson? Justifique a sua resposta.
2. Agruparam-se algumas categorias na tabela acima, criando uma classe de árvores com 3 ou menos gomos e uma classe de árvores com 11 ou mais gomos. Com base nesta tabela, obteve-se um valor calculado da estatística de Pearson de $X^2_{calc} = 28.122$.
   (a) Descreva pormenorizadamente o teste de hipóteses, para um nível de significância $\alpha = 0.05$, e admitindo válida a distribuição assintótica da estatística.
   (b) Calcule a parcela da estatística de Pearson correspondente à classe agrupada com 11 ou mais gomos. Comente.
   (c) Comente a seguinte frase: "*O teste efectuado permite concluir que a distribuição de Poisson não é adequada para descrever o número de gomos por árvore*".

II [12 valores]

Um estudo sobre videiras da casta Aragonez, realizado na Régua, visou estudar a possibilidade de estimar o peso total das uvas a partir do número de cachos. Para tal, foram demarcadas 180 parcelas com igual número de plantas e, para cada parcela, foi contado o número de cachos (variável `ncachos`) e medido, em kg, o peso total das uvas (variável `pesototal`). Eis alguns indicadores:

Variável	Mínimo	Máximo	Média	Variância
`ncachos`	4	84	42.75556	285.64941
`pesototal`	0.650	21.100	7.909722	14.26665

Foi ajustada uma recta de regressão para modelar o peso total das uvas a partir do número de cachos, e obtidos os seguintes resultados:

```
Call: lm(formula = pesototal ~ ncachos, data = elsa1t1)
[...]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.61660    0.49737    1.24    0.217
ncachos      0.17058        ???   15.76   <2e-16
---
Residual standard error: 2.447 on 178 degrees of freedom
Multiple R-squared:  0.5826,   Adjusted R-squared:  0.5802
F-statistic: ??? on 1 and 178 DF,  p-value: ???
```

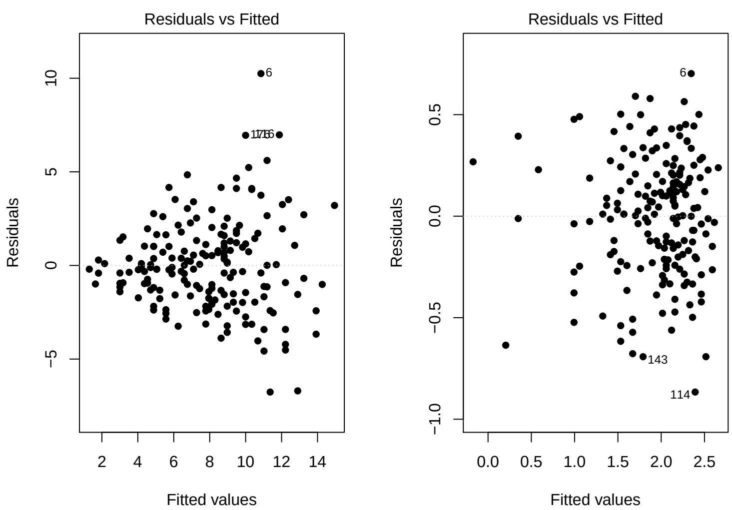
1. Calcule as três Somas de Quadrados associadas a esta regressão.
2. Discuta, de forma pormenorizada, a qualidade do modelo de regressão linear ajustado. Comente.
3. Teste a hipótese de que, a cada cacho de uvas adicional, corresponda um aumento médio do peso total das uvas inferior a 200 gramas. Não dê o benefício da dúvida à hipótese.
4. Construa um intervalo a 95% de confiança para o peso esperado das uvas em parcelas com 50 cachos. Interprete o resultado obtido.
5. Foi ajustado um novo modelo linear, relacionando os logaritmos das duas variáveis referidas. Obtiveram-se os seguintes resultados:

```
Call: lm(formula = log(pesototal) ~ log(ncachos), data = elsa1t1)
[...]
Coefficients:
             Estimate  Std.Error t value Pr(>|t|)
(Intercept)  -1.46223    0.16271  -8.987 3.65e-16
log(ncachos)  0.93036    0.04414  21.080  < 2e-16
---
Residual standard error: 0.3015 on 178 degrees of freedom
Multiple R-squared:  0.714,   Adjusted R-squared:  0.7124
F-statistic: 444.4 on 1 and 178 DF,  p-value: < 2.2e-16
```

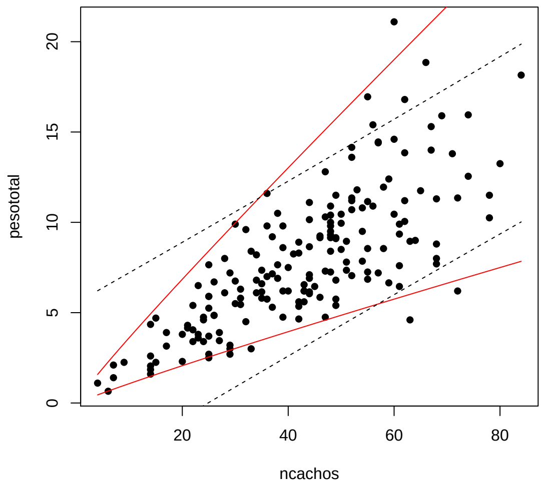
   (a) O maior resíduo, em valor absoluto, corresponde à observação 114, uma parcela com 63 cachos e um peso total de uvas de $4.6kg$. Calcule o valor desse resíduo e indique a proporção da Soma de Quadrados Residual correspondente a esta observação. Qual o peso total de uvas (em kg) ajustado pelo modelo, correspondente a esta observação?

   (b) Deduza a fórmula resultante deste ajustamento para, sem logaritmizações, prever o peso total das uvas a partir do número de cachos numa parcela. Justifique brevemente, usando um teste de hipoteses adequado, se é admissível afirmar que as taxas de variação relativas do peso total (y) e do número de cachos (x) são iguais.

6. Eis os gráficos de resíduos usuais contra valores ajustados de y, à esquerda para o modelo inicialmente ajustado, e à direita para o modelo com a log-transformação das variáveis. Comente os gráficos e as suas implicações para o uso de cada modelo.



7. Por cima da nuvem de pontos de `pesototal` (eixo vertical) e `ncachos` (eixo horizontal) foram traçados dois pares de bandas de precisão a 95%: (i) obtido com o modelo linear nas variáveis originais, ajustado no início deste grupo; e (ii) obtido a partir do modelo com as transformações logarítmicas (ajustado na alínea 5), após a sua conversão para as escalas originais de `pesototal` e `ncachos`. Identifique, justificando, cada par de bandas. Comente.



III [4,5 valores]

1. Seja dado o modelo de regressão linear simples, em contexto inferencial, com n pares de observações $\{(x_i, Y_i)\}_{i=1}^n$.

   (a) Deduza, justificando, a distribuição de probabilidades de cada Y_i.

   (b) Deduza, justificando, a expressão de $V[\hat{\beta}_1]$, a variância do estimador do declive da recta populacional.

   (c) Admitindo conhecida a distribuição do estimador do declive, $\hat{\beta}_1$, mostre que $\hat{\beta}_0$ é um estimador centrado da ordenada na origem da recta populacional.

   (d) Mostre que o Coeficiente de Determinação amostral verifica a relação $R^2 = \frac{F_{calc}}{F_{calc}+(n-2)}$, onde F_{calc} é o valor calculado da estatística do teste de ajustamento global do modelo. Qual a relação aproximada entre o valor calculado da estatística F e o tamanho da amostra, para um Coeficiente de Determinação de 0.90?

2. A relação de Michaelis-Menten entre duas variáveis y e x é da forma $y = \frac{x}{c+dx}$. Mostre que resulta de admitir que y é função de x, satisfazendo a equação diferencial $y'(x) = c\left(\frac{y(x)}{x}\right)^2$.