INSTITUTO SUPERIOR DE AGRONOMIA

**ESTATÍSTICA E DELINEAMENTO – 2015/16**

**2 de Novembro de 2015 PRIMEIRO TESTE Duração: 2h30**

**I** [4 valores]

Foram escolhidos ao acaso 175 locais nas margens de rios portugueses e recolhida informação relativa à vegetação e a variáveis ambientais em cada local. Os locais foram classificados em 3 grupos, definidos com base na composição da vegetação, e em quatro classes definidas pela percentagem de rocha nas margens de cada local. Pretende-se analisar se a classificação baseada na composição de vegetação está relacionada com a classificação baseada na percentagem de rocha. Eis a tabela de contingência resultante:

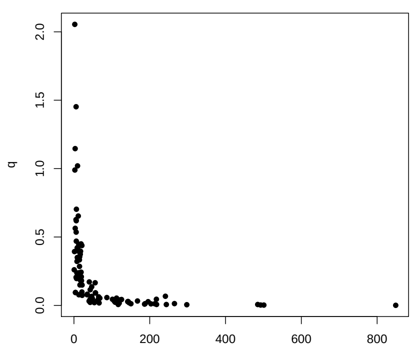
| | [0, 10] | ]10, 20] | ]20, 40] | ]40, 80] | Soma |
|---|---|---|---|---|---|
| Grupo 1 | 64 | 14 | 14 | 8 | 100 |
| Grupo 2 | 16 | 6 | 10 | 14 | 46 |
| Grupo 3 | 9 | 3 | 6 | 11 | 29 |
| Soma | 89 | 23 | 30 | 33 | 175 |

1. Descreva pormenorizadamente as hipóteses a testar, a estatística de teste e sua distribuição assintótica, bem como a região crítica, adequadas à questão que foi colocada.
2. Verifique, com o mínimo de contas possível, se é válido admitir a distribuição assintótica da estatística do teste.
3. Sabendo que o valor calculado da estatística do teste é $X^2_{\text{calc}} = 24.675$, qual a resposta à questão colocada ($\alpha = 0.05$)? O *p-value* correspondente é inferior a 0.001? Justifique.
4. Avalie a contribuição para o resultado obtido dos locais na classe com percentagem de rocha entre 40 % e 80 %. Comente.

**II** [11 valores]

Num estudo de comportamento da fotossíntese da videira face à intensidade luminosa disponível no ambiente, foram medidas curvas de resposta à luz. Com base nestas curvas, determinou-se a intensidade luminosa quando a fotossíntese é nula (chamada ponto de compensação para a luz), registada na variável de nome `LCP` (em $\mu$mol quanta $m^{-2}\ s^{-1}$), e também o rendimento quântico da fotossíntese, na variável `q` (adimensional). Pretende-se modelar o rendimento quântico da fotossíntese (`q`) a partir dos pontos de compensação para a luz (`LCP`). Eis os principais indicadores e a nuvem de pontos correspondente aos dados do estudo:

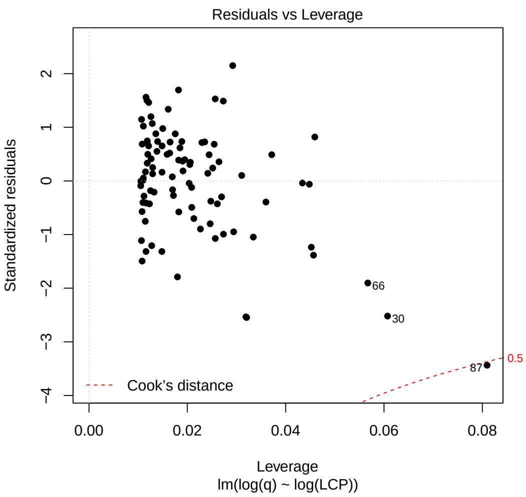
```
      LCP                  q
Min.   :    0.7336   Min.   :0.00063
1st Qu.:   11.2448   1st Qu.:0.03095
Median :   40.5356   Median :0.09067
Mean   :   86.3829   Mean   :0.22806
3rd Qu.:  118.0194   3rd Qu.:0.33589
Max.   :  849.0622   Max.   :2.05449
Var.   :17121.1223   Var.   :0.10856
```



Após alguma exploração, decidiu-se ajustar um modelo de regressão linear simples *sobre os dados logaritmizados* (logaritmos naturais). Eis alguns indicadores e resultados da regressão:

```
     log(LCP)          log(q)
 Mean   : 3.5152   Mean   :-2.454
 Var.   : 2.2098   Var.   : 2.552

Call: lm(formula = log(q)~log(LCP), data=fotossintese)
Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
(Intercept)  0.87106   0.20157    4.321  3.88e-05
log(LCP)    -0.94603   0.05286  -17.898  < 2e-16
---
Residual standard error: 0.7618 on 93 DF
Multiple R-squared:  0.775
F-statistic: 320.3 on 1 and 93 DF, p-value: < 2.2e-16
```




1. Com base em quantas observações foi ajustado o modelo? Justifique a sua resposta.
2. Qual o coeficiente de correlação linear para a nuvem de pontos *dos dados logaritmizados*?
3. Discuta formalmente a qualidade do modelo ajustado.
4. Deduza o tipo de relação de fundo *entre as variáveis originalmente medidas* a que corresponde o modelo estudado e indique a equação da curva ajustada à nuvem de pontos em `q` e `LCP`.
5. Teste, ao nível de significância de 5%, se é admissível que as variáveis `q` e `LCP` sejam inversamente proporcionais.
6. Construa um intervalo de predição (a 95%) para o rendimento quântico da fotossíntese (isto é, para `q`), duma observação associada a um ponto de compensação de valor `LCP` $= 100$. Comente.
7. Comente o diagrama no topo desta página. Em particular, identifique o valor da variável `LCP` para a observação com maior efeito alavanca.

## III [5 valores]

Considere o modelo de regressão linear simples (modelo RLS), e os habituais estimadores $\hat{\beta}_1$ e $\hat{\beta}_0$ dos parâmetros da recta.

1. Escreva (sem deduzir) a expressão de $\hat{\sigma}_{\hat{\beta}_0}$, o estimador do desvio padrão de $\hat{\beta}_0$.
2. Sabendo que $\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \cap t_{n-2}$, indique os passos necessários para construir um intervalo a $(1-\alpha)\times 100\%$ de confiança para a ordenada na origem $\beta_0$ da recta de regressão populacional.
3. Sabendo que $\text{Cov}[\overline{Y},\hat{\beta}_1] = 0$, justifique que $\text{Cov}[\hat{\beta}_1,\hat{\beta}_0] = -\frac{\overline{x}\,\sigma^2}{(n-1)s_x^2}$. Em que caso é que os estimadores dos parâmetros da recta podem ser independentes?
4. Usando a alínea anterior e sabendo que $V[\hat{\beta}_1] = \frac{\sigma^2}{(n-1)s_x^2}$, deduza a expressão constante do formulário para a variância $V[\hat{\beta}_0]$ do estimador da ordenada na origem populacional.