INSTITUTO SUPERIOR DE AGRONOMIA

ESTATÍSTICA E DELINEAMENTO – 2008/09

ALGUNS EXERCÍCIOS (COM RESOLUÇÕES)

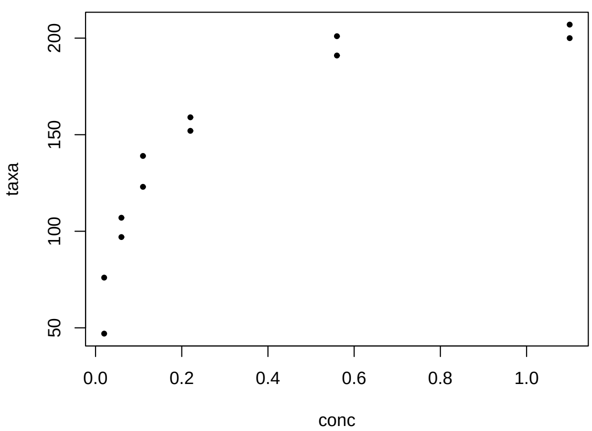
I

Num estudo sobre reacções enzimáticas, procura-se analisar a "velocidade" da reacção em células tratadas com Puromicina. Para diferentes concentrações do substrato (variável *conc*), medidas em partes por milhão (ppm), registou-se o número de emissões radioactivas por minuto, e a partir destas calculou-se a taxa inicial ou "velocidade" da reacção, em contagens/minuto/minuto (variável *taxa*). Os resultados obtidos foram:

													Média	Var.
conc	0.02	0.02	0.06	0.06	0.11	0.11	0.22	0.22	0.56	0.56	1.10	1.10	0.345	0.1589
taxa	76	47	97	107	123	139	159	152	191	201	207	200	141.5833	2805.356

A relação entre taxas da reacção e concentrações do substrato é representada no gráfico à direita. Admite-se que o modelo de Michaelis-Menten é adequado à descrição da relação referida, e decide-se usar este modelo com a seguinte parametrização (onde y representa a *taxa* e x a concentração *conc*),

$$y = \frac{ax}{b+x} \qquad (a > 0, b > 0 \quad \text{e} \quad x > 0)\ .$$



1. Mostre que o modelo referido pode ser linearizado, indicando a relação linearizada e as transformações de variáveis necessárias.

2. Eis o resultado de ajustar uma regressão linear entre as variáveis transformadas:

```
Call:  lm(formula = 1/taxa ~ I(1/conc))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0051072  0.0007040   7.255 2.74e-05 ***
I(1/conc)   0.0002472  0.0000321    ??   1.64e-05 ***
---
Residual standard error: 0.001892 on 10 degrees of freedom
Multiple R-Squared: 0.8557,     Adjusted R-squared: 0.8413
F-statistic:  ?? on ?? and ?? DF,  p-value: 1.642e-05
```

(a) Complete a tabela, preenchendo os valores omissos (valor da estatística t para o declive da recta, valor da estatística F e respectivos graus de liberdade), indicando como os obtem.

(b) Teste formalmente o ajustamento global do modelo linearizado. Comente.

(c) Interprete o significado da estimativa do coeficiente que surge na segunda linha (isto é, do valor 0.0002472) em termos da relação entre as variáveis transformadas.

(d) Construa um intervalo a 95% de confiança para a ordenada na origem da relação linear entre as variáveis transformadas. À luz desse intervalo, diga se é de admitir que *na relação entre as variáveis originais* (modelo Michaelis-Menten), se pode admitir que a assíntota horizontal à direita é dada pela recta $y = 210$. Comente.

II

Um estudo sobre macieiras envolveu o acompanhamento de $n = 48$ árvores ao longo de 15 anos. Para cada árvore foram medidas as seguintes variáveis:

$Y1$ – perímetro do tronco, aos 4 anos de vida (em dm);
$Y2$ – crescimento anual acumulado ao fim de 4 anos (em m);
$Y3$ – perímetro do tronco, aos 15 anos de vida (em dm);
$Y4$ – peso da árvore (acima do solo) aos 15 anos de vida (em libras).

Alguns dos dados obtidos, e a matriz de correlações das quatro variáveis numéricas, foram:

```
   Pe   Y1    Y2   Y3    Y4
1   1 1.11 2.569 3.58 0.760
2   1 1.19 2.928 3.75 0.821
3   1 1.09 2.865 3.93 0.928
4   1 1.25 3.844 3.94 1.009
5   1 1.11 3.027 3.60 0.766
...........................
47  6 1.13 3.064 3.63 0.707
48  6 1.11 2.469 3.95 0.952
```

```
          Y1        Y2        Y3        Y4
Y1 1.0000000 0.8808781 0.4380177 0.3304323
Y2 0.8808781 1.0000000 0.5152646 0.4514731
Y3 0.4380177 0.5152646 1.0000000 0.9456171
Y4 0.3304323 0.4514731 0.9456171 1.0000000
```

Uma vez que a medição rigorosa da variável $Y4$ implica a destruição da árvore, procurou-se saber se é possível prever a variável $Y4$ a partir das restantes variáveis. Ajustou-se um Modelo Linear aos dados, tendo-se obtido:

```
Call:  lm(formula = Y4 ~ Y1 + Y2 + Y3)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.81432    0.26250  -3.102  0.00335
Y1          -0.75868    0.31649  -2.397  0.01042
Y2           0.08028    0.05152   1.558  0.12630
Y3           0.61469    0.03400  18.081  < 2e-16

Residual standard error: 0.09391 on 44 degrees of freedom
Multiple R-Squared: 0.908,      Adjusted R-squared: 0.9017
F-statistic: 144.7 on 3 and 44 DF,  p-value: < 2.2e-16
```

1. Que significado pode atribuir ao valor estimado $b_3 = 0.61469$?

2. Teste formalmente se é admissível considerar que o coeficiente da variável $Y1$ na relação linear é nulo. Comente as suas conclusões.

3. Indique qual o melhor modelo com apenas duas variáveis preditoras, e diga, justificando, se o seu ajustamento difere significativamente do ajustamento do modelo acima estudado.

4. Independentemente da sua resposta na alínea anterior, teste formalmente se o modelo inicialmente ajustado difere significativamente de uma Regressão Linear Simples de $Y4$ sobre $Y3$.

5. A variável $Y4$ foi medida utilizando o sistema de pesos anglo-saxónico. Sabendo que uma libra corresponde a cerca de 0.4536 kg diga, justificando, se haveria ou não alterações nas seguintes quantidades (e, se sim, quais) resultantes de transformar as medições de $Y4$ em kg:

   (a) a matriz $\mathbf{H}$ de projecções ortogonais associada ao ajustamento do modelo;

   (b) o vector $\hat{Y}4$ dos valores ajustados da variável resposta;

   (c) o coeficiente de determinação R^2.

III

1. Descreva completamente o Modelo de Regressão Linear Simples e mostre que, dado o Modelo, o estimador de Mínimos Quadrados da ordenada na origem, $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x}$, tem distribuição

$$\hat{\alpha} \quad \cap \quad \mathcal{N}\left(\alpha \, , \, \sigma \cdot \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}} \right) .$$

2. Tendo em conta as interpretações geométricas das Somas de Quadrados e do Coeficiente de Determinação, no espaço $\mathbb{R}^n$, interprete geometricamente a condição associada a uma rejeição da hipótese nula num teste F de ajustamento global numa regressão linear: rejeitar H_0 se $F_{calc} > f_{\gamma\,(p,n-(p+1))}$.