

Ficha de Trabalho 5

Modelos Lineares e Aplicações

Exercício 1:

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 $\hat{\beta}_3 = 0.61469$?
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 $Y3$.

Exercício 2:

Os dados seguintes foram recolhidos com o objectivo de estudar a relação entre o desgaste de uma peça (y) e as variáveis x_1 (viscosidade do óleo) e x_2 (carga), todas em unidades convenientes:

y	x_1	x_2
193	1.6	85.1
230	15.5	81.6
172	22.0	105.8
91	43.0	120.1
113	33.0	135.7
125	40.0	111.5

- (a) Obtenha a equação de regressão estimada para o modelo

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \dots, 6 \quad \text{com } \varepsilon_i \underset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

- (b) Vários modelos foram ajustados aos dados, alguns deles incluindo uma nova variável $x_3 = x_1 \times x_2$, tendo-se obtido os resultados seguintes:

Modelo	Variáveis no modelo	SSE
I	x_1	3872
II	x_2	2720
III	x_1, x_2	1950
IV	x_1, x_3	1225
V	x_2, x_3	1721
VI	x_1, x_2, x_3	1123

Qual é, em sua opinião, o melhor modelo ?

(1.5 val.)

- (c) Construa a tabela ANOVA e o teste F correspondente para o modelo que escolheu na alínea (b). O que conclui?

(2.0 val.)