

Modelos Lineares

Teste - 31/01/2020

Duração: 2h30m

Mestrado em Estatística

Departamento de Matemática

Nome: _____ Número: _____

Grupo I

1. Considere os seguintes dados:

```
> data(iris)
```

Para saber informação sobre os dados:

```
> help(iris)
```

Nota: Em todas as questões que utilizar o programa R apresentar os comandos usados.

Trata-se de observações em quatro variáveis morfológicas: largura e comprimento de pétalas e sépalas de 3 espécies de lírios.

- (a) Ajuste um modelo de regressão linear múltipla da largura das pétalas (*Petal.Width*) sobre as restantes quatro variáveis preditoras. Identifique as observações outliers, high-leverage points e observações influentes.

Nas alíneas seguintes, considere o modelo ajustado na alínea (a).

- (b) Determine um intervalo a 95% de confiança para a diminuição esperada na largura das pétalas (*Petal.Width*), por aumento de uma unidade no comprimento das sépalas (*Sepal.Length*).
- (c) Determine um intervalo a 99% de confiança para o valor esperado da largura das pétalas para um lírio da espécie "setosa" com *Sepal.Length*=6, *Sepal.Width*=3 e *Petal.Length*=5.
- (d) Teste se o efeito da largura das sépalas (*Sepal.Width*) é igual ao efeito do comprimento das pétalas (*Petal.Length*) sobre a largura das pétalas (*Petal.Width*).
- (e) Ajuste agora um modelo de regressão linear múltipla da largura das pétalas (*Petal.Width*) sobre as restantes três variáveis quantitativas. Teste se este modelo difere ou não significativamente do modelo da alínea (a) quanto ao ajustamento aos dados.

2. Na tentativa de explicar a afluência de público a uma determinada sala de cinema foi utilizado o modelo:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 F_i + \epsilon_i, i = 1, \dots, 10$$

com y_i o número total de espetadores que assistiu ao filme i , x_i o preço do bilhete para o filme i (em euros), $F_i = 1$ se o filme i é novo e 0 caso contrário.

Y	100	50	70	50	80	25	40	50	100	55
x	6	5	5	5	6	4	5.5	6	8	6
F	1	1	1	0	0	1	0	0	1	1

- (a) Obtenha as estimativas dos parâmetros do modelo. Interprete os resultados.
- (b) Teste o ajustamento global do modelo.
- (c) Estime a variância dos erros.
- (d) Um dos técnicos envolvidos neste estudo afirmou que a única variável relevante para a explicação da afluência de público a este cinema é o preço do bilhete. Concorda com esta afirmação? Justifique a sua resposta aplicando um teste de hipóteses adequado.

Grupo II

- 1 Consideremos o modelo $Y_i = \beta_1 x_i + \epsilon$ $i = 1, \dots, n$, que verifica os pressupostos usuais do modelo de regressão linear. Encontre o estimador de mínimos quadrados para β_1 .
- 2 Usando a definição calcule $Cov(\alpha X + \beta Y, \beta Z + \alpha W)$

GRUPO III

Num estudo duma espécie de 149 árvores de fruto pretende-se prever a produção total (variável produção, em kg/árvore), a partir de potenciais variáveis preditoras: o diâmetro das árvores (variável diâmetro), a altura da árvore (variável altura); e os números de frutos contados em três datas diferentes: em Junho (variável nfrJun); em Setembro (variável nfrSet) e na colheita (variável nfrColh). Obteve-se a seguinte informação:

Variável	diametro	altura	nfrJun	nfrSet	nfrColh	producao
Mínimo	14.13	1.499	1.333	1.067	0.433	0.0450
Máximo	23.47	2.117	16.600	14.000	9.900	1.2730
Média	18.82	1.842	7.364	6.156	4.025	0.5221
Desvio Padrão	1.8569648	0.1161166	3.7105911	3.0958839	2.3597379	0.2975462

Matriz das correlações entre as variáveis.

	diametro	altura	nfrJun	nfrSet	nfrColh	producao
diametro	1.0000000	0.7207190	0.5783022	0.5950834	0.5834737	0.5936943
altura	0.7207190	1.0000000	0.4343267	0.4310161	0.3950239	0.3863919
nfrJun	0.5783022	0.4343267	1.0000000	0.9895452	0.9554149	0.9444425
nfrSet	0.5950834	0.4310161	0.9895452	1.0000000	0.9716697	0.9616189
nfrColh	0.5834737	0.3950239	0.9554149	0.9716697	1.0000000	0.9853879
producao	0.5936943	0.3863919	0.9444425	0.9616189	0.9853879	1.0000000

Considere o modelo de regressão linear simples da variável produção versus o número de frutos contados na colheita.

- Calcule a equação da reta de regressão ajustada.
- Calcule o valor do coeficiente de determinação correspondente.
- Calcule a estimativa da variância dos erros aleatórios do modelo.

Considere agora o modelo de Regressão Linear Múltipla, que permitisse prever a produção antes da data da colheita. Eis os resultados de um modelo de regressão linear com os quatro preditores disponíveis no final de Setembro:

```

Call:
lm(formula = producao ~ diametro + altura + nfrJun + nfrSet , data = macieira2)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.086704   0.109663   0.791   0.4305
diametro     0.014061   0.005772   2.436   0.0161
altura      -0.210009   0.082022  -2.560   0.0115
nfrJun       -0.021418   0.012389  -1.729   0.0860
nfrSet        0.116200   0.015066   7.713    ???
---
Residual standard error: 0.07935 on ??? degrees of freedom
Multiple R-squared: 0.9308, Adjusted R-squared: 0.9289
F-statistic: ??? on ?? and ??? DF, p-value: < 2.2e-16

```

- d) Complete os resultados obtidos, indicando os valores em falta (???) . Justifique as suas respostas.
- e) Compare, através de um Teste de Hipóteses adequado, o modelo completo ajustado neste ponto e o submodelo com apenas os preditores diâmetro e nfrJun, ao qual corresponde um Coeficiente de Determinação de valor 0.8954.