

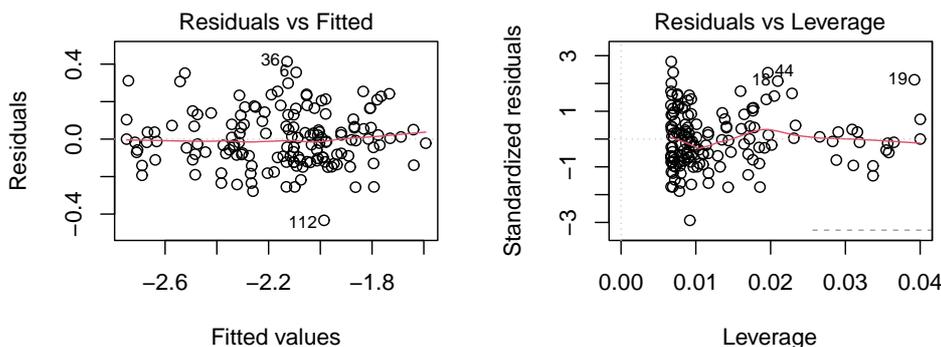
I [13 valores]

1. De acordo com a teoria da dinâmica de fluidos, a condutância ( $g$ ) à transferência de calor entre a superfície de uma folha e a camada de ar que a envolve é bem descrita por um modelo potência do tipo  $g = \alpha u^\beta$ , em que  $u$  é a velocidade do vento a que a folha está exposta. Num estudo com folhas de sobreiro foi avaliada a condutância  $g$  (em  $\text{m s}^{-1}$ ) para diferentes velocidades do vento  $u$  (em  $\text{m s}^{-1}$ ).

- (a) Mostre que a relação teórica entre  $g$  e  $u$  pode ser linearizada.
- (b) Numa primeira análise, ajustou-se um modelo linear às variáveis transformadas (designando por  $x$  a transformação de  $u$  e  $y$  a transformação de  $g$ ) tendo-se obtido os seguintes resultados:

```
> mod1<-lm(y ~ x)
> summary(mod1)
Call:
lm(formula = y ~ x)
(...)
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.13212      0.01218  -174.99  <2e-16 ***
x              0.45001      0.02045   22.01  <2e-16 ***
---
Residual standard error: 0.1489 on 148 degrees of freedom
Multiple R-squared:  0.766, Adjusted R-squared:  0.7644
F-statistic: 484.5 on 1 and 148 DF,  p-value: < 2.2e-16
```

- i. Teste formalmente a qualidade do ajustamento global do modelo, comentando o valor do coeficiente de determinação.
- ii. De acordo com as condições em que foi realizada a experiência, a variação média de  $y$  deve ser metade da variação de  $x$ . Construa um intervalo a 95% de confiança para verificar se os dados obtidos estão de acordo com esta hipótese.
- iii. Alguns dos gráficos dos resíduos obtidos com o ajustamento deste modelo apresentam-se seguidamente:



A. Analise os 2 gráficos e discuta as suas implicações para o estudo efetuado.

B. Calcule o valor *aproximado* da distância de Cook associada à observação 19 (referenciada no gráfico da direita). Trata-se de uma observação influente? Justifique a sua resposta.

- (c) Os dados da experiência foram obtidos utilizando folhas de três diferentes dimensões (fator `dimensao` com os níveis `pequena`, `media` e `grande`). Com o objectivo de estudar a influência do tamanho das folhas, foi ajustado um novo modelo para as variáveis  $x$  e  $y$  que prevê a possibilidade de, a cada nível do fator `dimensao`, corresponder uma reta de regressão diferente. Os resultados obtidos foram os seguintes:

```
> mod2<-lm(y ~ x * dimensao)
> summary(mod2)
Call:
lm(formula = y ~ x * dimensao)
(...)
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.21434    0.01811  -122.277 < 2e-16 ***
x                0.46153    0.02822   16.352 < 2e-16 ***
dimensaomedia   0.05621    0.02542    2.211  0.0286 *
dimensaopequena 0.19170    0.02544    7.536 4.95e-12 ***
x:dimensaomedia -0.02431    0.04002   -0.607  0.5446
x:dimensaopequena -0.02543    0.04506   -0.564  0.5735
---
Residual standard error: 0.1259 on 144 degrees of freedom
Multiple R-squared:  0.8372, Adjusted R-squared:  0.8316
F-statistic: 148.1 on 5 and 144 DF, p-value: < 2.2e-16
```

- i. Escreva a equação linear ajustada para cada tamanho de folha.
- ii. Diga se é possível considerar que as rectas populacionais associadas às folhas de dimensão grande e pequena têm igual declive. Justifique a sua resposta com um teste de hipóteses adequado ( $\alpha=0.05$ ).

2. A análise química de 39 mostos da casta Moscatel Graúdo registou valores de seis características associadas ao aroma, juntamente com o `rendimento` (em kg/planta), grau álcool provável (variável `grau`) e `acidez` (g/l de ácido tartárico) da respectiva produção. Eis as médias e variâncias amostrais das variáveis observadas, bem como a respectiva matriz de correlações:

|          | media     | variancia    |  | linalol | nerol  | geraniol | hexenol | hexanal | 2fenil. | rendim. | grau   | acidez |
|----------|-----------|--------------|--|---------|--------|----------|---------|---------|---------|---------|--------|--------|
| 0.477818 | 0.0278082 | linalol      |  | 1.000   | -0.125 | 0.463    | 0.340   | 0.060   | 0.099   | -0.294  | 0.445  | 0.027  |
| 0.057077 | 0.0006659 | nerol        |  | -0.125  | 1.000  | 0.318    | -0.201  | -0.165  | 0.050   | 0.227   | -0.326 | -0.111 |
| 0.383831 | 0.0042863 | geraniol     |  | 0.463   | 0.318  | 1.000    | 0.293   | 0.011   | 0.058   | -0.144  | 0.023  | -0.018 |
| 0.317028 | 0.0080113 | hexenol      |  | 0.340   | -0.201 | 0.293    | 1.000   | 0.112   | 0.029   | -0.117  | 0.380  | -0.170 |
| 0.117162 | 0.0061367 | hexanal      |  | 0.060   | -0.165 | 0.011    | 0.112   | 1.000   | 0.692   | 0.017   | 0.047  | 0.079  |
| 0.124846 | 0.0048344 | 2feniletanol |  | 0.099   | 0.050  | 0.058    | 0.029   | 0.692   | 1.000   | 0.075   | -0.028 | 0.068  |
| 8.610233 | 2.2003850 | rendimento   |  | -0.294  | 0.227  | -0.144   | -0.117  | 0.017   | 0.075   | 1.000   | -0.293 | -0.048 |
| 8.757108 | 0.1416529 | grau         |  | 0.445   | -0.326 | 0.023    | 0.380   | 0.047   | -0.028  | -0.293  | 1.000  | -0.092 |
| 4.230297 | 0.0886225 | acidez       |  | 0.027   | -0.111 | -0.018   | -0.170  | 0.079   | 0.068   | -0.048  | -0.092 | 1.000  |

Dada a dificuldade em analisar as características aromáticas, seria útil poder prever pelo menos algumas a partir das restantes variáveis observadas. Foi sugerido tentar prever o teor de `hexanal` (**Atenção:** não confundir com outra variável de nome semelhante) a partir de outras variáveis.

- (a) Diga qual o melhor modelo de regressão linear simples para prever o teor de `hexanal`. Para a regressão linear simples que escolher determine o coeficiente de determinação.

- (b) Seguidamente foi ajustada uma regressão linear múltipla para tentar prever o teor de hexanal a partir dos restantes 8 preditores, tendo sido obtidos os seguintes resultados:

```
> summary(lm(hexanal ~ . , data=moscatel))
```

Coefficients:

|              | Estimate   | Std. Error | t value | Pr(> t ) |
|--------------|------------|------------|---------|----------|
| (Intercept)  | -0.0330805 | 0.3544070  | -0.093  | 0.926    |
| linalol      | -0.0436050 | 0.0787920  | -0.553  | 0.584    |
| nerol        | -0.6497245 | 0.4631393  | -1.403  | 0.171    |
| geraniol     | 0.0784129  | 0.2000816  | 0.392   | 0.698    |
| hexenol      | 0.0517223  | 0.1307570  | 0.396   | 0.695    |
| 2feniletanol | 0.7943703  | 0.1435978  | 5.532   | 5.19e-06 |
| rendimento   | 0.0004273  | 0.0072603  | 0.059   | 0.953    |
| grau         | 0.0038168  | 0.0327531  | 0.117   | 0.908    |
| acidez       | 0.0060019  | 0.0343731  | 0.175   | 0.863    |

---

Residual standard error: 0.06056 on 30 degrees of freedom

Multiple R-squared: 0.5281, Adjusted R-squared: ???

F-statistic: 4.197 on 8 and 30 DF

- Para o modelo de regressão linear múltipla ajustado, calcule e comente o valor do  $R^2$  modificado.
  - A qualidade do ajustamento deste modelo de regressão linear múltipla e do modelo de regressão linear simples da alínea 2a) difere significativamente ( $\alpha=0.05$ )? Descreva, em pormenor, o teste efectuado para responder à questão.
  - Num algoritmo de exclusão sequencial com base em testes T aos parâmetros do modelo (para  $\alpha=0.10$ ), qual a primeiro preditor a excluir do modelo? Justifique a sua resposta.
  - Um algoritmo de exclusão sequencial optou por um submodelo com apenas dois preditores: **nerol** e **2feniletanol**. Justifique que o coeficiente de determinação deste submodelo de dois preditores está entre 0.4788 e 0.5281.
- (c) Dado o modelo de regressão linear, a distribuição do vector  $\vec{Y}$ , de dimensão  $n$ , da variável resposta é  $\vec{Y} \sim \mathcal{N}_n(\mathbf{X}\vec{\beta}, \sigma^2 \mathbf{I}_n)$  e o vector dos resíduos é dado por  $\vec{E} = (\mathbf{I}_n - \mathbf{H})\vec{Y}$ , sendo  $\mathbf{H}$  a matriz de projecção ortogonal sobre o subespaço gerado pelas colunas da matriz  $\mathbf{X}$  e  $\mathbf{I}_n$  a matriz identidade de ordem  $n$ . Prove que o valor esperado do vector dos resíduos é o vector nulo,  $E[\vec{E}] = \vec{0}$ .

## II [7 valores]

- Num estudo sobre o efeito da diminuição da área foliar na produção de girassol, foi efectuado uma experiência em que se avaliou o número médio de sementes produzidas por planta para quatro níveis de desfolhação (control-0%, treat1-33%, treat2-66% e treat3-100%), em várias fases de desenvolvimento da cultura, desde a pré-floração (stage 1) até à maturidade fisiológica (stage 5). Foram seleccionadas 60 parcelas e a cada combinação nível de desfolhação - fase de desenvolvimento foram atribuídas aleatoriamente 3 parcelas. A média das 60 observações, as médias de nível e de célula são as seguintes:

```
> model.tables(girassol.aov,type="means")
```

Tables of means

Grand mean

769.35

```

desfolhacao
desfolhacao
control  treat1  treat2  treat3
  889.3   940.9   670.4   576.8

fase
fase
stage1 stage2 stage3 stage4 stage5
  720.0  648.8  780.3  914.8  782.9

desfolhacao:fase
      fase
desfolhacao stage1 stage2 stage3 stage4 stage5
  control   959.7  808.0  980.3  942.0  756.7
  treat1    885.7  953.0 1132.0  984.0  749.7
  treat2    649.0  634.3  703.7  821.7  543.3
  treat3    385.7  199.7  305.3  911.3 1082.0

```

- (a) Identifique qual o tipo de delineamento experimental utilizado neste estudo e explicitamente de forma completa o modelo que lhe corresponde.
- (b) Seguidamente apresenta-se a tabela ANOVA relativa ao estudo efectuado:

```

> girassol.aov<-aov(numseed~desfolhacao*fase, data=girassol)
> summary(girassol.aov)

```

|                  | Df | Sum Sq  | Mean Sq | F value | Pr(>F)   |
|------------------|----|---------|---------|---------|----------|
| desfolhacao      | 3  | 1360209 | 453403  | C       | 0.000564 |
| fase             | 4  | B       | 115277  | 1.831   | 0.141924 |
| desfolhacao:fase | A  | 1878441 | 156537  | 2.486   | 0.015440 |
| Residuals        | 40 | 2518821 | 62971   |         |          |

---

- i. Complete a tabela, indicando como obtém cada um dos valores em falta A, B e C.
  - ii. Teste formalmente a existência de efeitos de interacção. Comente a sua conclusão para  $\alpha=0.01$  e  $\alpha=0.05$ .
  - iii. Na última fase de maturidade fisiológica (*stage5*), o número médio de sementes por planta difere entre o nível 0% (*control*) e o nível 100% (*treat3*) de desfolha? Realize um teste de Tukey para responder a esta questão ao nível de significância  $\alpha=0.05$ .
2. Pretendeu-se estudar a afinidade da enxertia da casta Moscatel Galego com 2 porta-enxertos (1103P e R99). Para tal, em condições homogêneas e para cada um dos porta-enxertos, foram plantados aleatoriamente no viveiro 10 talhões com 20 plantas enxertadas. Ao levantamento do viveiro foi avaliada a percentagem de pegamento em cada talhão.
- (a) Indique as unidades experimentais do ensaio.
  - (b) Esta experiência tem pseudo-repetições?
  - (c) Quantas repetições existem por porta-enxerto?
  - (d) O técnico que realizou este estudo não ficou satisfeito quando analisou os gráficos dos resíduos para validação dos pressupostos do modelo. Dada a natureza dos dados, optou por uma abordagem não paramétrica. A soma das ordenações das percentagens de pegamento obtidas nos 10 talhões para cada um dos porta-enxertos ( $R_i$ ) foi a seguinte:  $R_{1103P} = 108$ ,  $R_{R99} = 102$ .  
 Descreva, em pormenor, o teste não paramétrico que neste contexto deverá ser realizado. O que conclui sobre o estudo efectuado ao nível  $\alpha=0.05$ ?  
 $(\chi_{0.05(1)}^2 = 3.841; \chi_{0.05(2)}^2 = 5.991; \chi_{0.05(3)}^2 = 7.815; \chi_{0.05(4)}^2 = 9.488)$ .