

I [6 valores]

- Um estudo sobre a variedade de macieira "**Bravo** de Esmolfe", realizado um ano após a transplantação das árvores para o terreno, visa encontrar um modelo para prever o número médio de frutos à colheita (variável `nfrColh`) a partir de outras 2 variáveis. Dispõem-se de valores médios de todas as variáveis para 66 parcelas plantados ao acaso num ensaio. Foi ajustado o modelo de regressão linear múltipla do número médio de frutos à colheita (variável `nfrColh`) sobre a altura da árvore (variável `altura`, em m) e o número médio de frutos contados em Setembro (variável `nfrSet`):

```
> macieirabravo2.lm<-lm(nfrColh~altura+nfrSet, data=macieirabravo)
> summary(macieirabravo2.lm)
```

Coefficients:

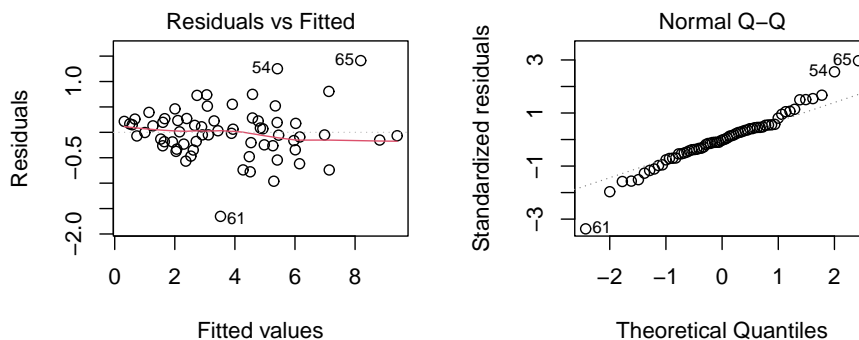
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.41493	0.95357	1.484	0.1428
altura	-1.10790	0.54136	-2.047	0.0449
nfrSet	0.76444	0.02365	32.322	<2e-16

Residual standard error: 0.4968 on 63 degrees of freedom

Multiple R-squared: 0.9493, Adjusted R-squared: 0.9477

F-statistic: 589.8 on 2 and 63 DF, p-value: < 2.2e-16

Alguns dos gráficos dos resíduos obtidos com o ajustamento deste modelo apresentam-se seguidamente:



- Analise os 2 gráficos e discuta as suas implicações para o estudo efetuado.
 - Calcule o valor exacto do resíduo usual associado à observação 61 (valores das variáveis na observação 61: $altura_{61} = 1.917$, $nfrSet_{61} = 5.533$, $nfrColh_{61} = 1.867$).
- No estudo efetuado, as árvores das 66 parcelas são de diferentes regiões, estando 22 parcelas plantadas com material originário de cada uma de 3 regiões de origem: Trás-os-Montes (TM), Beira Interior (BI) e Beira Litoral (BL). Assim, o técnico, considerando válidos os pressupostos do modelo e olhando para o resultado do ajustamento do modelo ao nível $\alpha = 0.01$, decidiu ajustar um modelo de Análise de Covariância à totalidade das $n = 66$ observações, tendo como variável resposta o número médio de frutos à colheita (variável `nfrColh`), variável preditora numérica o número médio de frutos contados em Setembro (variável `nfrSet`) e variável preditora categórica a `origem` (factor, com 3 níveis). Eis os resultados do ajustamento deste modelo:

```

> macieirabravo3.lm<-lm(nfrColh~nfrSet*origem, data=macieirabravo)
> summary(macieirabravo3.lm)
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.16111    0.22461  -0.717  0.4760
nfrSet          0.64392    0.04845  13.290 <2e-16
origemBL       -0.29445    0.34268  -0.859  0.3936
origemTM       -0.70381    0.36783  -1.913  0.0605
nfrSet:origemBL 0.10240    0.06123   1.672  0.0996
nfrSet:origemTM 0.15256    0.06316   2.416  0.0188
---
Residual standard error: 0.4987 on 60 degrees of freedom
Multiple R-squared:  0.9513, Adjusted R-squared:  0.9473
F-statistic: 234.6 on 5 and 60 DF,  p-value: < 2.2e-16

```

- (a) Interprete as estimativas dos parâmetros do modelo e escreva a equação da recta ajustada referente a cada região de origem de "**Bravo** de Esmolfe".
- (b) Os declives das rectas de regressão para as origens Trás-os-Montes (*TM*) e Beira Interior (*BI*) são significativamente diferentes ao nível $\alpha=0.05$? Justifique a sua resposta com o teste de hipóteses adequado, indicando todos os seus passos.

II [14 valores]

1. No estudo sobre a variedade de macieira "**Bravo** de Esmolfe" anteriormente descrito foram estudadas no mesmo local as 3 regiões de origem das árvores plantadas nas parcelas, designadas *TM*, *BI* e *BL*. No ensaio realizado, foram aleatoriamente associadas 22 parcelas a cada região de origem e avaliou-se o número médio de frutos à colheita em cada parcela. A média amostral de todas as observações e as médias amostrais por região de origem são indicadas seguidamente:

```

> model.tables(macieirabravo.aov, type="means")
Tables of means
Grand mean

3.727788

  origem
origem
  BI    BL    TM
2.468 4.248 4.467

```

- (a) Descreva o delineamento experimental adotado para o ensaio e, em pormenor, o modelo ANOVA adequado ao problema sob estudo.
- (b) Seguidamente apresenta-se a tabela ANOVA relativa ao estudo efectuado:

```

> summary(macieirabravo.aov)
                Df Sum Sq Mean Sq F value Pr(>F)
origem           2    A      26.441    C      0.00258
Residuals       B 253.83   4.029

```

Complete a tabela, indicando como obtém cada um dos valores em falta A, B e C.

- (c) Pode afirmar-se que o número médio de frutos à colheita é igual para todas as regiões de origem? Formalize e efectue o teste *F* adequado ao problema, ao nível $\alpha=0.05$.

- (d) Quais as estimativas dos parâmetros do modelo?
- (e) O número médio de frutos à colheita em parcelas com origem Beira Litoral é igual ao número médio de frutos à colheita em parcelas com origem Beira Interior? Justifique a sua resposta com o teste de Tukey ($\alpha=0.05$).
- (f) Prove que, no contexto da ANOVA descrita neste exercício, o resíduo da observação Y_{ij} é dado pela sua diferença em relação à média amostral de nível

$$E_{ij} = Y_{ij} - \bar{Y}_i.$$

- (g) Prove que no contexto da ANOVA descrita neste exercício, o Quadrado Médio Residual é a média (simples) das k variâncias de nível S_i^2 da variável resposta Y :

$$QMRE = \frac{1}{k} \sum_{i=1}^k S_i^2$$

2. 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, adoptou de seguida uma abordagem não paramétrica. A soma das ordenações do número médio de frutos à colheita das 22 parcelas avaliadas para cada uma das 3 regiões de origem (R_i) foi a seguinte: $R_{BI} = 474$, $R_{BL} = 854$, $R_{TM} = 883$.

Descreva em pormenor o teste não paramétrico que neste contexto deverá ser realizado. O que conclui sobre o estudo efectuado ao nível $\alpha=0.05$?

$$(\chi_{0.05(1)}^2 = 3.841; \chi_{0.05(2)}^2 = 5.991; \chi_{0.05(3)}^2 = 7.815; \chi_{0.05(4)}^2 = 9.488).$$

3. Ainda neste estudo, pretendeu-se determinar o conteúdo de vitamina C na polpa da maçã. Para tal, ao acaso, separaram-se 2 frutos de cada uma das 22 parcelas de cada região de origem, tendo sido levados para o laboratório.
- (a) Indique as unidades experimentais do ensaio.
- (b) Esta experiência tem pseudo-repetições?
- (c) Quantas repetições existem por região de origem?