

Modelos Matemáticos e Aplicações

Módulo 2: Modelos lineares mistos
Alguns casos particulares: aplicações

2016-2017

Elsa Gonçalves
ISA/UL

Caso particular 1

modelo aleatório, equilibrado com matrizes G e R diagonais

Modelo ANOVA a

1 factor de efeitos aleatórios, equilibrado

$$Y_{ij} = \mu + u_i + e_{ij}$$

para $i = 1, \dots, a, j = 1, \dots, b, n = ab$.

Y_{ij} é a j -ésima observação no nível i do factor A ;

μ é a média populacional;

u_i é o efeito do nível i do factor A (factor de efeitos aleatórios);

e_{ij} é o erro aleatório associado à observação Y_{ij} .

- $u_i, i. i. d., \mathcal{N} \left(0, \sigma^2_u \right), \forall i$
- $e_{ij}, i. i. d., \mathcal{N} \left(0, \sigma^2_e \right), \forall ij$

Quadro resumo ANOVA a **1 factor de efeitos aleatórios**

	G.L.	S.Q.	QM	E[QM]	F
Factor A	$a - 1$	$SQA = \sum_{i=1}^a b (\bar{y}_i - \bar{y}_{..})^2$	$QMA = \frac{SQA}{a - 1}$	$b\sigma_u^2 + \sigma_e^2$	$\frac{QMA}{QMRE}$
Resíduos	$a(b - 1)$	$SQRE = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_i)^2$	$QMRE = \frac{SQRE}{a(b - 1)}$	σ_e^2	
TOTAL	$ab - 1$	$SQT = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$			

- Estimadores para as componentes de variância deduzidos a partir do quadro de análise de variância

$$E[SQA] = E\left[\sum_{i=1}^a b (\bar{Y}_i - \bar{Y}_{..})^2\right] = (a-1)(b\sigma_u^2 + \sigma_e^2)$$

$$E[QMA] = \frac{E[SQA]}{(a-1)} = b\sigma_u^2 + \sigma_e^2$$

$$E[SQRE] = E\left[\sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_i)^2\right] = a(b-1)\sigma_e^2$$

$$E[QMRE] = \frac{E[SQRE]}{a(b-1)} = \sigma_e^2$$

Igualando os valores observados aos esperados, fica-se com

$$SQA = (a-1)(b\hat{\sigma}_u^2 + \hat{\sigma}_e^2)$$

$$SQRE = a(b-1)\hat{\sigma}_e^2$$

Chegando-se a

$$\hat{\sigma}_e^2 = \frac{SQRE}{a(b-1)} = QMRE \quad \hat{\sigma}_u^2 = \left(\frac{SQA}{a-1} - \hat{\sigma}_e^2\right) / b = \frac{QMA - QMRE}{b}$$

Os estimadores de máxima verosimilhança restrita para as componentes de variância são:

$$\left\{ \begin{array}{l} \hat{\sigma}_u^2 = \frac{QMA - QMRE}{b}, \quad \text{se } QMA \geq QMRE, \\ \hat{\sigma}_u^2 = 0, \quad \text{caso contrário} \end{array} \right.$$

$$\left\{ \begin{array}{l} \hat{\sigma}_e^2 = QMRE, \quad \text{se } QMA \geq QMRE, \\ \hat{\sigma}_e^2 = \frac{SQT}{ab-1}, \quad \text{caso contrário} \end{array} \right.$$

Os estimadores de máxima verosimilhança para as componentes de variância são:

$$\left\{ \begin{array}{l} \hat{\sigma}_u^2 = \frac{\left(1 - \frac{1}{a}\right)QMA - QMRE}{b}, \quad \text{se } \left(1 - \frac{1}{a}\right)QMA \geq QMRE, \\ \hat{\sigma}_u^2 = 0, \quad \text{caso contrário} \end{array} \right.$$

$$\left\{ \begin{array}{l} \hat{\sigma}_e^2 = QMRE, \quad \text{se } \left(1 - \frac{1}{a}\right)QMA \geq QMRE, \\ \hat{\sigma}_e^2 = \frac{SQT}{ab}, \quad \text{caso contrário,} \end{array} \right.$$

Testes de razão de verosimilhanças à componente de variância σ_u^2

- **Hipóteses:** $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$
- **A estatística do teste** de razão de verosimilhanças restritas é dada por

$$\Lambda = 2(l_{R_1} - l_{R_0}) \sim \chi^2_\nu$$

sendo l_{R_1} a log-verosimilhança restrita associada ao modelo completo e l_{R_0} a log-verosimilhança restrita associada ao modelo reduzido (submodelo).

Satisfazendo as condições de regularidade, a estatística do teste de razão de verosimilhanças, sob H_0 , segue assintoticamente uma **distribuição χ^2 com número de graus de liberdade (ν) igual à diferença entre o número de parâmetros de covariância dos dois modelos**. Uma das condições de regularidade sobre a qual a validade do teste atrás descrito assenta é que H_0 não esteja na fronteira do espaço parâmetro. Ora, as componentes de variância não são estritamente positivas, o que significa que estão na fronteira do espaço parâmetro. Em tais condições, esta aproximação falha, tendo consequências na inferência relativa às componentes de variância. Nestas condições, sob H_0 , caracterizou-se a distribuição assintótica da estatística do teste de razão de verosimilhanças como uma mistura de distribuições χ^2 . Foi provado que, ao testar uma única componente de variância, a distribuição assintótica da estatística Λ é uma mistura 50:50 de χ_0^2 e χ_1^2 . É prática comum efectuar o teste nas condições usuais e dividir o *p-value* obtido por 2.

- **Nível de significância do teste:** α
- **Região crítica (Região de rejeição):** unilateral direita

$$\text{Rejeitar } H_0 \text{ se } \Lambda_{calc} > \chi^2_{\alpha(\nu)}$$

Neste caso particular, podemos testar a componente de variância através de um teste de razão de verossimilhanças, ou através do seguinte teste:

Teste de hipóteses à componente de variância associada ao factor A (aos efeitos aleatórios do factor A)

- **Hipóteses:** $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$
- **Estatística do teste:** $F = \frac{QMA}{QMRE} \cap \mathcal{F}_{(a-1, a(b-1))}$, sob H_0
- **Nível de significância do teste:** α
- **Região critica (Região de rejeição):** unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(a-1, a(b-1))}$

Nota:

$$\frac{SQRE}{\sigma_e^2} \sim \chi^2_{a(b-1)}$$

$$\frac{SQA}{b\sigma_u^2 + \sigma_e^2} \sim \chi^2_{a-1}$$



Variáveis aleatórias independentes,

$$\frac{\frac{QMA}{b\sigma_u^2 + \sigma_e^2}}{\frac{QMRE}{\sigma_e^2}} \sim \mathcal{F}_{a-1, a(b-1)}$$

Exemplo: para um modelo a 1 factor de efeitos aleatórios, equilibrado (factor com a níveis, b observações por nível), o melhor preditor empírico linear não enviesado de u_i (para o nível i) é:

$$EBLUP(u_i) = \frac{b\hat{\sigma}_u^2}{b\hat{\sigma}_u^2 + \hat{\sigma}_e^2} (\bar{Y}_{i.} - \bar{Y}_{..})$$

Aplicação 1 (exercício 1)

Num estudo de selecção da videira, com vista a estudar a variabilidade genética do rendimento da casta Touriga Nacional, foi instalado um ensaio de campo em Vila Nova de Fozcoa, com uma amostra aleatória de genótipos da casta (196 genótipos). No campo, cada genótipo foi disposto aleatoriamente em 5 parcelas (ensaio com 5 repetições). Em 1994 foram obtidos os dados de rendimento (kg/planta) que se encontram disponíveis na *data.frame* *touriga*.

- a)** Descreva o modelo que lhe parece ser adequado para estudar a variabilidade genética do rendimento da casta.
- b)** Utilize o comando *aov* do R para obter a tabela de análise de variância do respectivo modelo.
 - (i) Estime as componentes de variância envolvidas no modelo descrito na alínea a).
 - (ii) Com base na informação disponível, efectue um teste à existência de variabilidade genética do rendimento da casta (utilize o nível de significância de 0.05).
 - (iii) Sabendo que $\bar{Y}_{..} = 1.196$ kg/planta e $\bar{Y}_{c0101.} = 1.6044$ kg/planta, qual o melhor preditor empírico linear não enviesado do efeito genotípico do rendimento do clone c0101?

Aplicação 1 (cont.)

- c) Ajuste o modelo anteriormente descrito com o método de estimação de máxima verosimilhança restrita, recorrendo
- (i) ao comando *lme* do pacote “nlme”; ao comando *lmer* do pacote “lme4”; ao comando *varComp* do pacote “varComp”.
 - (ii) Aplique o comando *summary* aos três objectos criados anteriormente e identifique as estimativas de máxima verosimilhança restrita das componentes de variância. Compare os resultados agora obtidos com os da alínea b(i).
 - (iii) Efectue um teste à existência de variabilidade genética do rendimento (teste de razão de verosimilhanças).
 - (iv) Qual o valor ajustado do rendimento do clone c0101 na repetição 2?
 - (v) Explore os comandos *ranef* e *fitted* do “nlme” e “lme4”.

Aplicação 1 (cont.)

d) Na verdade, o ensaio de Touriga Nacional atrás descrito foi instalado de acordo com um delineamento experimental em blocos casualizados completos (5 blocos).

(i) Ajuste um novo modelo que incorpore o efeito do bloco (considere também que é um factor de efeitos aleatórios).

Utilize o pacote *lme4*.

(ii) Efectue testes de hipóteses às componentes de variância associadas aos factores clone e bloco.

(iii) Calcule o AIC e o BIC dos dois modelos ajustados (o da alínea a) e o da alínea d)) e seleccione o melhor modelo com base nesses critérios.

NOTA: No final, ajuste o modelo com 2 factores de efeitos aleatórios, utilizando a função *lme* do pacote *nlme* e a função *varComp* do pacote *varComp*.

Caso particular 2

Modelo linear misto (1 factor de efeitos fixos, 1 factor de efeitos aleatórios, sem e com interacção), equilibrado com matrizes G e R diagonais

Caso particular: modelo misto
1 factor de efeitos fixos
1 factor de efeitos aleatórios, equilibrado

$$Y_{ijk} = \mu_1 + \beta_i + u_j + e_{ijk}$$

para $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc$, com $\beta_1 = 0$.

Y_{ijk} é a k -ésima observação no nível i do factor A e nível j do factor B ;

μ_1 é a média populacional no nível 1 do factor A ;

β_i é o efeito do nível i do factor A (acréscimo relativamente a μ_1), **fixo**;

u_j é o efeito do nível j do factor B , **aleatório**;

e_{ijk} é o erro aleatório associado à observação Y_{ijk} .

- $u_j, i. i. d., \mathcal{N} \left(0, \sigma^2_u \right), \forall j$
- $e_{ijk}, i. i. d., \mathcal{N} \left(0, \sigma^2_e \right), \forall ijk$

Quadro ANOVA RESUMO: modelo linear misto, **1 factor de efeitos fixos, 1 factor de efeitos aleatórios**

	G.L.	S.Q.	QM	E[QM]	F
Factor A	$a - 1$	SQA	QMA	$\frac{bc}{a - 1} \sum_{i=1}^a (\beta_i - \bar{\beta})^2 + \sigma_e^2$	$\frac{QMA}{QMRE}$
Factor B	$b - 1$	SQB	QMB	$\sigma_e^2 + ac\sigma_u^2$	$\frac{QMB}{QMRE}$
Resíduos	$n - (a + b - 1)$	$SQRE$	$QMRE$	σ_e^2	
TOTAL	$n - 1$	SQT			

- Estimadores ANOVA para as componentes de variância

$$E[SQB] = (b - 1)(ac\sigma_u^2 + \sigma_e^2)$$

$$E[QMB] = \frac{E[SQB]}{(b-1)} = ac\sigma_u^2 + \sigma_e^2$$

$$E[SQRE] = n - (a + b - 1)\sigma_e^2$$

$$E[QMRE] = \frac{E[SQRE]}{n - (a + b - 1)} = \sigma_e^2$$

Chegando-se a

$$\hat{\sigma}_e^2 = \frac{SQRE}{n - (a + b - 1)} = QMRE$$

$$\hat{\sigma}_u^2 = \frac{QMB - QMRE}{ac}$$

- Os estimadores de máxima verosimilhança para as componentes de variância ($\hat{\sigma}_u^2 \geq 0$)

$$\hat{\sigma}_e^2 = \left[1 - \frac{a-1}{b(ac-1)} \right] QMRE,$$

$$\hat{\sigma}_u^2 = \frac{SQB/b - \hat{\sigma}_e^2}{ac}$$

- Os estimadores de máxima verosimilhança restrita para as componentes de variância ($\hat{\sigma}_u^2 \geq 0$)

$$\hat{\sigma}_e^2 = \frac{SQRE}{n - (a + b - 1)} = QMRE$$

$$\hat{\sigma}_u^2 = \frac{QMB - QMRE}{ac}$$

Neste caso particular, podemos testar a componente de variância através de um teste de razão de verosimilhanças (atrás descrito), ou através do seguinte teste:

Teste de hipóteses à componente de variância associada ao factor B

- **Hipóteses:** $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$
- **Estatística do teste:** $F = \frac{QMB}{QMRE} \cap F_{(b-1, n-(a+b-1))}$, sob H_0
- **Nível de significância do teste:** α
- **Região crítica (Região de rejeição):** unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(b-1, n-(a+b-1))}$

Testes de hipóteses a combinações lineares dos efeitos do modelo linear misto

($L^T \begin{bmatrix} \beta \\ u \end{bmatrix}$), sendo L uma matriz não aleatória

- **Hipóteses:** $H_0: L^T \begin{bmatrix} \beta \\ u \end{bmatrix} = \mathbf{0}$ vs. $H_1: \text{existe pelo menos um elemento do vector } L^T \begin{bmatrix} \beta \\ u \end{bmatrix} \text{ que é } \neq 0$

- **Estatística do teste:** $F = \frac{\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}^T L(L^T \hat{C} L)^{-1} L^T \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}}{\text{rank}(L)} \sim \mathcal{F}_{v_1, v_2}$, sob H_0

- **Nível de significância do teste:** α

- **Região crítica (Região de rejeição):** unilateral direita

$$\text{Rejeitar } H_0 \text{ se } F_{\text{calc}} > f_{\alpha(v_1, v_2)}$$

Em geral, F tem apenas aproximadamente distribuição \mathcal{F} e os seus graus de liberdade são estimados. $v_1 = \text{rank}(L)$ (característica da matriz L), v_2 graus de liberdade são estimados (aproximação de *Satterthwaite*, *Kenward and Roger* ou outras). Só em alguns casos de dados equilibrados (e em casos especiais de dados não equilibrados), sendo os vectores u e e vectores de variáveis aleatórias *i.i.d.*, é que tal não acontece. Nesses casos particulares, $v_2 = n - r(\mathbf{W})$, sendo $r(\mathbf{W})$ a característica da matriz \mathbf{W} que contém as colunas da matriz \mathbf{X} e da matriz \mathbf{Z} .

Neste caso particular: modelo misto
1 factor de efeitos fixos
1 factor de efeitos aleatórios, equilibrado

Teste de hipóteses aos efeitos do factor A (efeitos fixos)

- Hipóteses: $H_0: \beta_i = 0, \forall i=2, \dots, a$ vs $H_1: \exists i=2, \dots, a$ tal que $\beta_i \neq 0$
- Estatística do teste: $F = \frac{QMA}{QMRE} \cap F_{(a-1, n-(a+b-1))}$, sob H_0
- Nível de significância do teste: α
- Região crítica (Região de rejeição): unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(a-1, n-(a+b-1))}$

Aplicação 2 (exercício 3)

Considere os dados disponíveis na *data.frame* terrenos. Nesse estudo pretende-se comparar o rendimento obtido com quatro variedades de trigo. Identificaram-se 13 terrenos com características de solos diferentes. Assuma que esses terrenos são uma amostra aleatória da população de terrenos onde essas variedades serão cultivadas.

- a) Ajuste no R (por exemplo, com o *package nlme*, função *lme*) o modelo apropriado para o caso em estudo.
- b) Efectue o teste de hipóteses aos efeitos fixos do modelo (efeitos associados às variedades). Para o cálculo da estatística do teste recorde os testes de hipóteses a combinações lineares dos efeitos fixos do modelo linear misto dados nas aulas teóricas. Considere a matriz das covariâncias estimadas dos estimadores dos efeitos fixos (comando `vcov(terrenolme1)`), defina a matriz L , crie o vector com as estimativas dos efeitos fixos e , com a ajuda do R, calcule a estatística do teste. Para as suas conclusões utilize o nível de significância de 0,05. No final, execute o comando *anova(terrenolme1)*.
- c) Existe decréscimo de rendimento da variedade B comparativamente à variedade A (para $\alpha = 0,05$)?

Caso particular: modelo misto

1 factor de efeitos fixos

1 factor de efeitos aleatórios, com interacção, equilibrado

$$Y_{ijk} = \mu_1 + \beta_i + u_j + (\beta u)_{ij} + e_{ijk}$$

para $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc$, com $\beta_1 = 0$.

Y_{ijk} é a k -ésima observação no nível i do factor A e nível j do factor B ;

μ_1 é a média populacional no nível 1 do factor A ;

β_i é o efeito do nível i do factor A (acréscimo relativamente a μ_1), **fixo**;

u_j é o efeito do nível j do factor B , **aleatório**;

$(\beta u)_{ij}$ é o efeitos de interacção do nível i do factor A com o nível j do factor B , **aleatório**;

e_{ijk} é o o erro aleatório associado à observação Y_{ijk} .

- $u_j, i. i. d., \mathcal{N} \left(0, \sigma^2_u \right), \forall j$
- $(\beta u)_{ij}, i. i. d., \mathcal{N} \left(0, \sigma^2_{\beta u} \right), \forall ij$
- $e_{ijk}, i. i. d., \mathcal{N} \left(0, \sigma^2_e \right), \forall ijk$

**QUADRO ANOVA RESUMO: modelo linear misto, 1 factor de efeitos fixos
1 factor de efeitos aleatórios, com interacção**

	G.L.	S.Q.	QM	E[QM]	F
Factor A	$a - 1$	SQA	QMA	$\frac{bc}{a - 1} \sum_{i=1}^a (\beta_i - \bar{\beta})^2 + \sigma_e^2 + c\sigma_{\beta u}^2$	$\frac{QMA}{QMAB}$
Factor B	$b - 1$	SQB	QMB	$\sigma_e^2 + c\sigma_{\beta u}^2 + ca\sigma_u^2$	$\frac{QMB}{QMAB}$
Interacção	$(a - 1)(b - 1)$	$SQAB$	$QMAB$	$\sigma_e^2 + c\sigma_{\beta u}^2$	$\frac{QMAB}{QMRE}$
Resíduos	$ab(c - 1)$	$SQRE$	$QMRE$	σ_e^2	
TOTAL	$n - 1$	SQT			

- Estimadores ANOVA para as componentes de variância

$$E[SQB] = (b - 1) \left(ac\sigma_u^2 + c\sigma_{\beta u}^2 + \sigma_e^2 \right)$$

$$E[QMB] = \frac{E[SQB]}{(b-1)} = ac\sigma_u^2 + c\sigma_{\beta u}^2 + \sigma_e^2$$

$$E[SQAB] = (a - 1)(b - 1) \left(c\sigma_{\beta u}^2 + \sigma_e^2 \right)$$

$$E[QMAB] = \frac{E[SQAB]}{(a-1)(b-1)} = c\sigma_{\beta u}^2 + \sigma_e^2$$

$$E[SQRE] = ab(c - 1)\sigma_e^2$$

$$E[QMRE] = \frac{E[SQRE]}{ab(c - 1)} = \sigma_e^2$$

chegando-se a

$$\hat{\sigma}_e^2 = QMRE$$

$$\hat{\sigma}_{\beta u}^2 = \frac{QMAB - QMRE}{c}$$

$$\hat{\sigma}_u^2 = \frac{QMB - QMAB}{ac}$$

- Os estimadores de máxima verosimilhança para as componentes de variância ($\hat{\sigma}_u^2 \geq 0, \hat{\sigma}_{\beta u}^2 \geq 0$)

$$\hat{\sigma}_e^2 = QMRE$$

$$\hat{\sigma}_u^2 = \frac{(1 - \frac{1}{b})(QMB - QMAB)}{ac}$$

$$\hat{\sigma}_{\beta u}^2 = \frac{(1 - \frac{1}{b})QMAB - QMRE}{c}$$

- Os estimadores de máxima verosimilhança restrita para as componentes de variância ($\hat{\sigma}_u^2 \geq 0, \hat{\sigma}_{\beta u}^2 \geq 0$)

$$\hat{\sigma}_e^2 = QMRE$$

$$\hat{\sigma}_{\beta u}^2 = \frac{QMAB - QMRE}{c}$$

$$\hat{\sigma}_u^2 = \frac{QMB - QMAB}{ac}$$

Neste caso particular, podemos testar a componente de variância através de um teste de razão de verosimilhanças (atrás descrito), ou através do seguinte teste:

Teste de hipóteses à componente de variância associada à interacção

- **Hipóteses:** $H_0: \sigma_{\beta u}^2 = 0$ vs $H_1: \sigma_{\beta u}^2 > 0$
- **Estatística do teste:** $F = \frac{QMAB}{QMRE} \cap \mathcal{F}_{((a-1)(b-1), ab(c-1))}$, sob H_0
- **Nível de significância do teste:** α
- **Região crítica (Região de rejeição):** unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha((a-1)(b-1), ab(c-1))}$

Neste caso particular, podemos testar a componente de variância através de um teste de razão de verosimilhanças (atrás descrito), ou através do seguinte teste:

Teste de hipóteses à componente de variância associada ao factor B

- **Hipóteses:** $H_0: \sigma_u^2 = 0$ vs $H_1: \sigma_u^2 > 0$
- **Estatística do teste:** $F = \frac{Q_{MB}}{Q_{MAB}} \cap \mathcal{F}_{(b-1, (a-1)(b-1))}$, sob H_0
- **Nível de significância do teste:** α
- **Região crítica (Região de rejeição):** unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(b-1, (a-1)(b-1))}$

Neste caso particular: modelo misto
1 factor de efeitos fixos
1 factor de efeitos aleatórios, com interacção, equilibrado

Teste de hipóteses aos efeitos do factor A (efeitos fixos)

- **Hipóteses:** $H_0: \beta_i = 0, \forall i=2, \dots, a$ vs $H_1: \exists i=2, \dots, a$ tal que $\beta_i \neq 0$
- **Estatística do teste:** $F = \frac{Q_{MA}}{Q_{MAB}} \cap F_{(a-1, (a-1)(b-1))}$, sob H_0
- **Nível de significância do teste:** α
- **Região crítica (Região de rejeição):** unilateral direita

Rejeitar H_0 se $F_{calc} > f_{\alpha(a-1, (a-1)(b-1))}$

Nota

- Para modelos aleatórios ou mistos mais complexos não existem testes estatísticos exactos para certos efeitos do modelo (o numerador e o denominador da estatística F são combinações lineares de quadrados médios). Fazem-se, nesse casos, testes F aproximados. Um dos métodos clássicos mais utilizado para essa aproximação é o chamado método de *Satterthwaite (1941)*. No entanto, existem outros métodos aplicados em modelos mistos mais complexos frequentemente referidos na bibliografia e correntemente usados nos *softwares*, são exemplos os métodos de Giesbrecht and Burns (1985) e de Kenward and Roger (1997).

Aplicação 3 (exercício 4)

Na *library nlme e lme4* do R está disponível a *data.frame Machines*. Estes dados referem-se a uma experiência referente à comparação de três marcas de máquinas utilizadas num processo industrial. Seis trabalhadores foram escolhidos aleatoriamente entre os funcionários de uma fábrica para operar cada máquina três vezes. A variável resposta é uma pontuação global da produtividade tendo em conta o número e a qualidade das componentes produzidas. No R, com a função *lmer* do pacote *lme4*, efectue os comandos necessários de forma a responder às seguintes questões.

- a) Descreva, em pormenor, o modelo que lhe parece ser adequado para o estudo acima descrito.
- b) Execute os comandos *plot.design (Machines)* e *interaction.plot (Machine, Worker, score)*.
- c) Quais os valores das estimativas de máxima verosimilhança restrita das componentes de variância do modelo?
- d) Os valores das estimativas das componentes de variância obtidas pelo método de máxima verosimilhança seriam superiores ou inferiores aos valores das estimativas enunciados na alínea anterior?

Aplicação 3 (cont.)

e) Efectue o teste de hipóteses adequado para averiguar se existe variabilidade relativamente à interacção trabalhador×máquina? Enuncie detalhadamente o teste de hipóteses que está a efectuar.

f) Efectue o teste de hipóteses adequado para averiguar se existe variabilidade relativamente ao trabalhador? Enuncie detalhadamente o teste de hipóteses que está a efectuar. Utilize um nível de significância de 0,01.

g) Quais os valores das estimativas dos efeitos fixos do modelo? Explique o seu significado.

h) Efectue um teste de hipóteses adequado para avaliar se existem efeitos principais associados às marcas de máquinas. Enuncie detalhadamente o teste de hipóteses que está a efectuar. Utilize um nível de significância de 0,01.₃₂

Aplicação 3 (cont.)

i) Execute os comandos e comente os resultados.

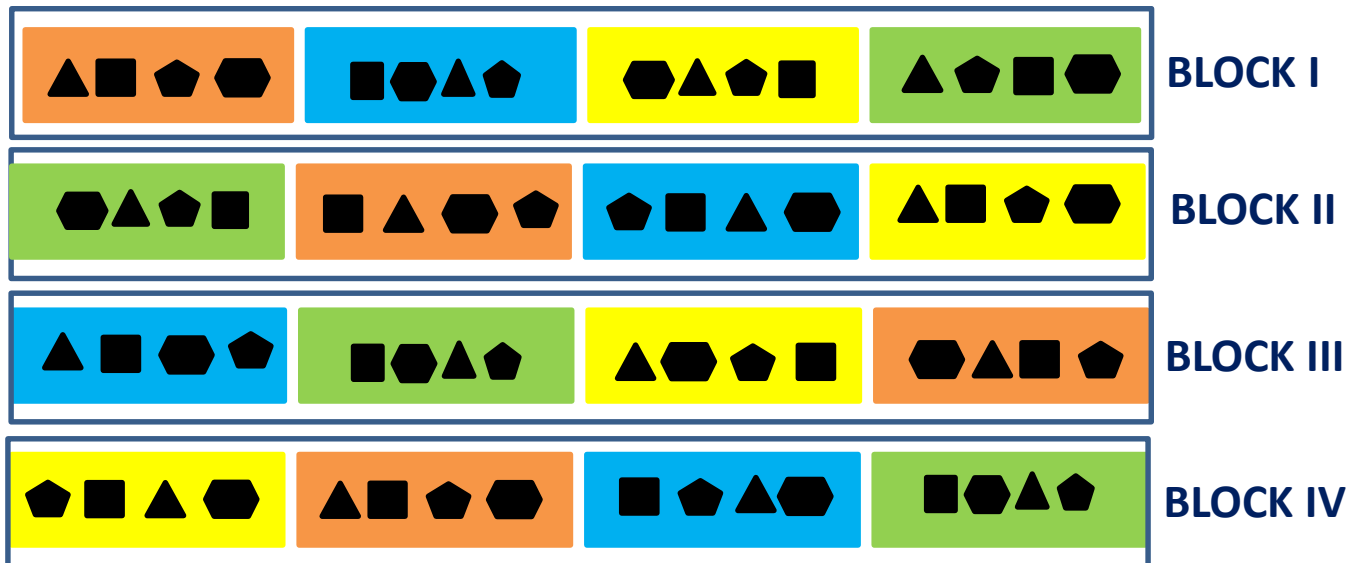
```
plot(machines1r)  
residuos<-resid(machines1r)  
qqnorm(residuos)  
eblupsworker<-ranef(machines1r)$Worker  
qqnorm(eblupsworker[,1])  
eblupsinteracao<-  
ranef(machines1r)$`Worker:Machine`  
qqnorm(eblupsinteracao[,1])
```

Caso particular 3

Modelo linear misto para análise de dados provenientes de um delineamento *split-plot* em blocos completos casualizados.

Delineamento em parcelas divididas em blocos completos casualizados (*split-plot design on a RCB*)

As diferentes cores representam os níveis do factor A, os diferentes símbolos representam os níveis do factor B, cada linha representa 1 bloco. Existem 4 blocos (I-IV), cada um com 4 níveis do factor A (cores, grandes parcelas, *whole plots*). Cada grande parcela é posteriormente dividida, sendo aleatoriamente atribuídos os 4 níveis do factor B (pequenas parcelas, *sub-plots*)



Consideremos o caso em que os fatores A e B admitem-se de efeitos fixos. O modelo adequado para a análise de dados provenientes de um delineamento deste tipo pode ser descrito como:

$$Y_{ijk} = \mu_{11} + \alpha_i + u_j + (\alpha u)_{ij} + \beta_k + (\alpha\beta)_{ik} + (\beta u)_{kj} + e_{ijk}$$

com $i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, n = abc,$

e com $\alpha_1 = 0, \beta_1 = 0, (\alpha\beta)_{1k} = 0, \forall k, (\alpha\beta)_{i1} = 0, \forall i.$

Sendo:

Y_{ijk} , a observação no nível i do factor A (atribuído à grande parcela, *whole-plot*), no bloco j e no nível k do factor B (atribuído à parcela dividida, *sub-plot* ou *split-plot*);

μ_{11} , a média populacional na célula (1,1), ou seja, na combinação do nível 1 factor A com o nível 1 do factor B;

α_i , o efeito do nível i do factor A (acrécimo), atribuído à grande parcela (*fixo*);

u_j , o efeito do bloco j (*aleatório*);

$(\alpha u)_{ij}$, o efeito da interacção do nível i do factor A com o bloco j , designado correntemente como *whole-plot error* (*aleatório*);

β_k , o efeito do nível k do factor B (acrécimo), atribuído à parcela dividida (*fixo*);

$(\alpha\beta)_{ik}$, o efeito da interacção do nível i do factor A com o nível k do factor B (acrécimo) (*fixo*);

$(\beta u)_{kj}$, o efeito da interacção do nível k do factor B com o bloco j (*aleatório*);

e_{ijk} , o erro aleatório associado à observação Y_{ijk} .

A análise mais comum incorpora o efeito $(\beta u)_{jk}$ no erro aleatório e_{ijk} , tendo-se no e_{ijk} a parte $(\beta u)_{jk}$ e a parte $(\alpha\beta u)_{ijk}$ (designada correntemente como *within plot error*).

Optando-se por esta última opção, admite-se que:

$$u_j, i. i. d., \mathcal{N}\left(0, \sigma^2_u\right), \forall j; (\alpha u)_{ij}, i. i. d., \mathcal{N}\left(0, \sigma^2_{\alpha u}\right), \forall ij;$$

$$e_{ijk}, i. i. d., \mathcal{N}\left(0, \sigma^2_e\right), \forall ijk; Cov(u_j, (\alpha u)_{ij}) = 0; Cov(u_j, e_{ijk}) = 0;$$

$$Cov\left((\alpha u)_{ij}, e_{ijk}\right) = 0.$$

O quadro de análise de variância para o modelo descrito, equilibrado, apresenta-se seguidamente.

	G.L.	S.Q.	QM	E[QM]	F
Factor A	$a - 1$	SQA	QMA	$c\sigma_{\alpha u}^2 + \sigma_e^2 + bc \frac{\sum_{i=1}^a (\alpha_i - \bar{\alpha})^2}{a - 1}$	$\frac{QMA}{QMError}$
Bloco	$b - 1$	$SQBL$	$QMBL$	$ac\sigma_u^2 c\sigma_{\alpha u}^2 + \sigma_e^2$	
Interacção FactorA×Bloco (<i>Whole-plot error</i>)	$(a - 1)(b - 1)$	$SQWError$	$QMWError$	$c\sigma_{\alpha u}^2 + \sigma_e^2$	
Factor B	$c - 1$	SQB	QMB	$\sigma_e^2 + ab \frac{\sum_{k=1}^c (\beta_k - \bar{\beta})^2}{c - 1}$	$\frac{QMB}{QMRE}$
Interacção FactorA×FactorB	$(a - 1)(c - 1)$	$SQAB$	$QMAB$	$\sigma_e^2 + b \frac{\sum_{i=1}^a \sum_{k=1}^c (\alpha\beta_{ik} - \bar{\alpha}\bar{\beta})^2}{(a - 1)(c - 1)}$	$\frac{QMAB}{QMRE}$
Resíduos (<i>Within plot error</i>)	$a(b - 1)(c - 1)$	$SQRE$	$QMRE$	σ_e^2	

Aplicação 4 (exercício 6)

No *package* “nlme” do R, existe um conjunto de dados (designado por “Alfalfa”).

```
> head(Alfalfa)
```

Grouped Data: Yield ~ Date | Block/Variety

Variety Date Block Yield

1	Ladak	None	1	2.17
2	Ladak	S1	1	1.58
3	Ladak	S20	1	2.29
4	Ladak	O7	1	2.23
5	Ladak	None	2	1.88
6	Ladak	S1	2	1.26

...

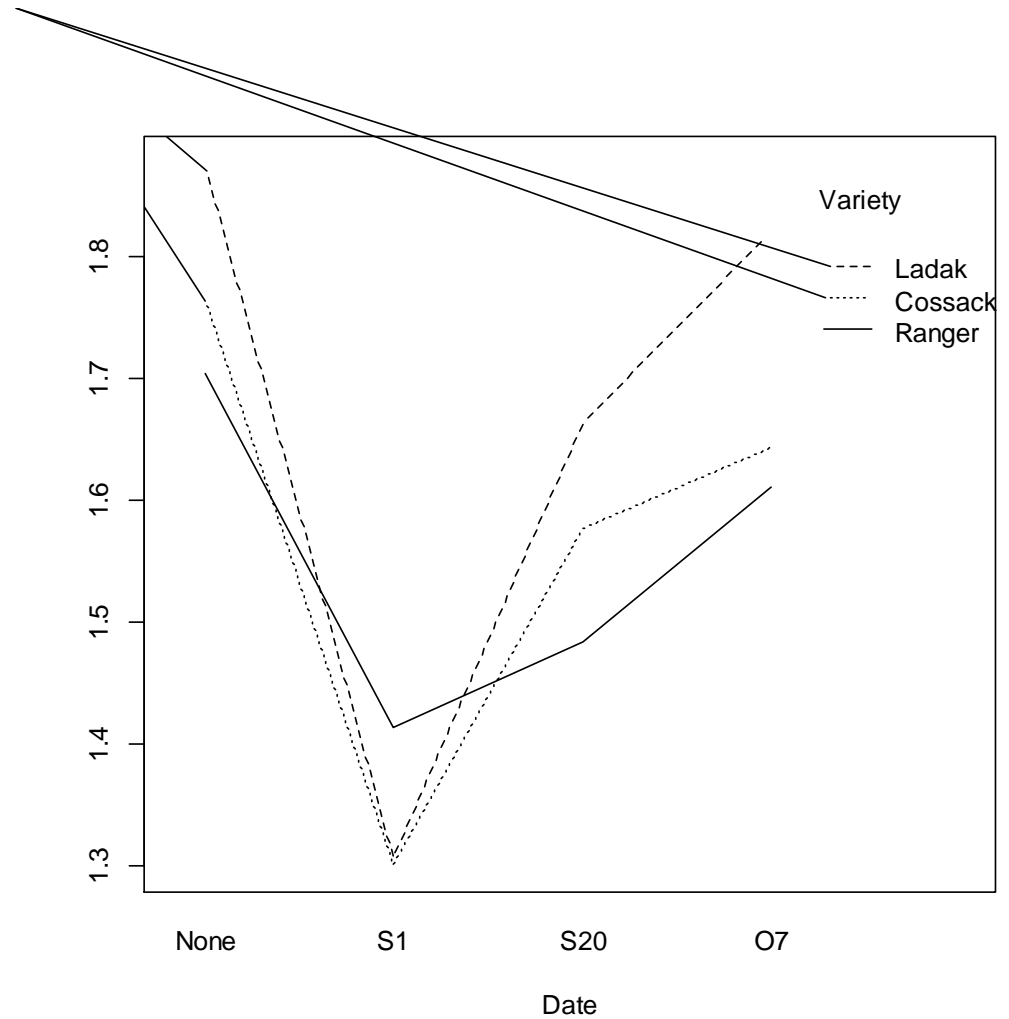
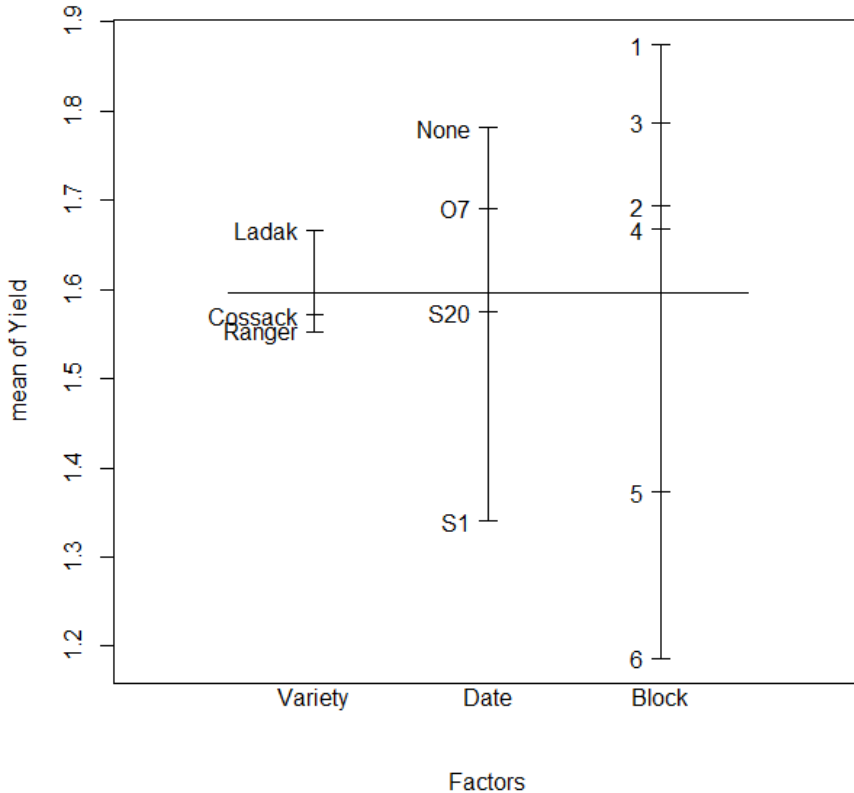
Aplicação 4 (cont.)

Estes dados foram descritos por Snedecor & Cochran (1980) como exemplo de um delineamento em parcelas divididas (*split-plot design*). Pretende-se estudar se o rendimento (T/acre) da luzerna (*Medicago sativa*) é afectado pela variedade e pela data do terceiro corte. Existem, portanto, dois factores em estudo: factor variedade de luzerna, com 3 níveis (*Cossac*, *Ladak* e *Ranger*) e factor data do corte, com 4 níveis (*none*—sem corte, *S1*— Sep1; *S20* — Sep20; and *O7* — Oct7). As unidades experimentais foram organizadas em 6 blocos, cada bloco com 3 grandes parcelas (*whole plots*), às quais foram aleatoriamente atribuídas as 3 variedades, e cada grande parcela foi dividida em 4 pequenas parcelas (*split plots* ou *subplots*), às quais foram aleatoriamente atribuídas as datas do corte.

a) Descreva o modelo que lhe parece adequado para responder ao objectivo do estudo.

Aplicação 4 (cont.)

b) Execute o comando `plot.design (Alfalfa)` e `interaction.plot (Date, Variety, Yield)`. Comente.



Aplicação 4 (cont.)

- c) Ajuste o modelo descrito na alínea a) no R recorrendo ao comando *lmer* do pacote “*lme4*”.
- d) Efectue os testes de hipóteses que acha adequados para responder aos objectivos do estudo.
- e) Compare os resultados obtidos nas alíneas anteriores com o resultados obtidos com o comando
“*aov(Yield~Date*Variety+Error(Block*Variety), data=Alfalfa)*”.

Algumas considerações

Nem sempre os factores A e B são de efeitos fixos. Seguidamente apresentam-se as tabelas de análise de variância para modelos equilibrados que consideram:

- (1) os efeitos dos factores A e B como aleatórios;
- (2) os efeitos do factor A como fixos e os efeitos do factor B como aleatórios.

(1) Quadro de análise de variância para um modelo com factor A e factor B de efeitos aleatórios, equilibrado:

	G.L.	QM	E[QM]	F
Factor A	$a - 1$	QMA	$bc\sigma_{\alpha}^2 + c\sigma_{\alpha u}^2 + b\sigma_{\alpha\beta}^2 + \sigma_e^2$	$\frac{QMA+QMRE}{QMWError+QMAB}^*$
Bloco	$b - 1$	$QMBL$	$ac\sigma_u^2 + c\sigma_{\alpha u}^2 + \sigma_e^2$	
Interacção FactorA×Bloco (<i>Whole-plot error</i>)	$(a - 1)(b - 1)$	$QMWError$	$c\sigma_{\alpha u}^2 + \sigma_e^2$	
Factor B	$c - 1$	QMB	$ab\sigma_{\beta}^2 + b\sigma_{\alpha\beta}^2 + \sigma_e^2$	$\frac{QMB}{QMAB}$
Interacção FactorA×FactorB	$(a - 1)(c - 1)$	$QMAB$	$b\sigma_{\alpha\beta}^2 + \sigma_e^2$	$\frac{QMAB}{QMRE}$
Resíduos	$a(b - 1)(c - 1)$	$QMRE$	σ_e^2	

*graus de liberdade aproximados. Por exemplo, pelo método de *Satterthwaite* tem-se:

$$v_1 = \frac{(QMA+QMRE)^2}{\frac{(QMA)^2}{a-1} + \frac{(QMRE)^2}{a(b-1)(c-1)}}, v_2 = \frac{(QMWError+QMAB)^2}{\frac{(QMWError)^2}{(a-1)(b-1)} + \frac{(QMAB)^2}{(a-1)(c-1)}}$$

(2) Quadro de análise de variância para um modelo com factor A de efeitos fixos e factor B de efeitos aleatórios, equilibrado:

	G.L.	QM	E[QM]	F
Factor A	$a - 1$	QMA	$c\sigma_{\alpha u}^2 + b \frac{a}{a-1} \sigma_{\alpha\beta}^2 + \sigma_e^2 + bc \frac{\sum_{i=1}^a (\alpha_i - \bar{\alpha})^2}{a-1}$	$\frac{QMA+QMRE}{QMWError+QMAB} *$
Bloco	$b - 1$	$QMBL$	$ac\sigma_u^2 + c\sigma_{\alpha u}^2 + \sigma_e^2$	
Interacção FactorA×Bloco (<i>Whole-plot error</i>)	$(a - 1)(b - 1)$	$QMWError$	$c\sigma_{\alpha u}^2 + \sigma_e^2$	
Factor B	$c - 1$	QMB	$ab\sigma_{\beta}^2 + \sigma_e^2$	$\frac{QMB}{QMRE}$
Interacção FactorA×FactorB	$(a - 1)(c - 1)$	$QMAB$	$b \frac{a}{a-1} \sigma_{\alpha\beta}^2 + \sigma_e^2$	$\frac{QMAB}{QMRE}$
Resíduos	$a(b - 1)(c - 1)$	$QMRE$	σ_e^2	

*graus de liberdade aproximados. Por exemplo, pelo método de *Satterthwaite* tem-se:

$$v_1 = \frac{(QMA+QMRE)^2}{\frac{(QMA)^2}{a-1} + \frac{(QMRE)^2}{a(b-1)(c-1)}}, v_2 = \frac{(QMWError+QMAB)^2}{\frac{(QMWError)^2}{(a-1)(b-1)} + \frac{(QMAB)^2}{(a-1)(c-1)}}$$

Caso particular 4

Segue-se um exemplo de aplicação do modelo linear misto com variáveis preditoras categóricas e numéricas (análise de covariância) e em que as observações são feitas no mesmo indivíduo ao longo do tempo*.

*As matrizes de correlação utilizadas para este tipo de análise são inspiradas nas séries temporais e na estatística espacial. Para o seu aprofundamento seriam necessárias umas bases teóricas sobre séries temporais e estatística espacial, o que não faz parte desta UC. Assim sendo, vamos apenas exemplificar a sua aplicação, para que fique registado que são instrumentos actualmente muito utilizados no âmbito dos modelos mistos.

Aplicação 5 (exercício 8)

No R está disponível o conjunto de dados *BodyWeight* (Pinheiro e Bates, 2000), referente ao peso do corpo de ratos (em gramas), ao longo de 64 dias. Mais precisamente, cada rato é pesado no dia 1 e a cada sete dias até ao dia 64, com uma pesagem adicional no dia 44. Há três grupos de ratos, cada grupo sujeito a uma dieta diferente.

```
> head(BodyWeight)
```

```
Grouped Data: weight ~ Time | Rat
```

```
  weight Time Rat Diet
```

```
1   240   1  1  1
```

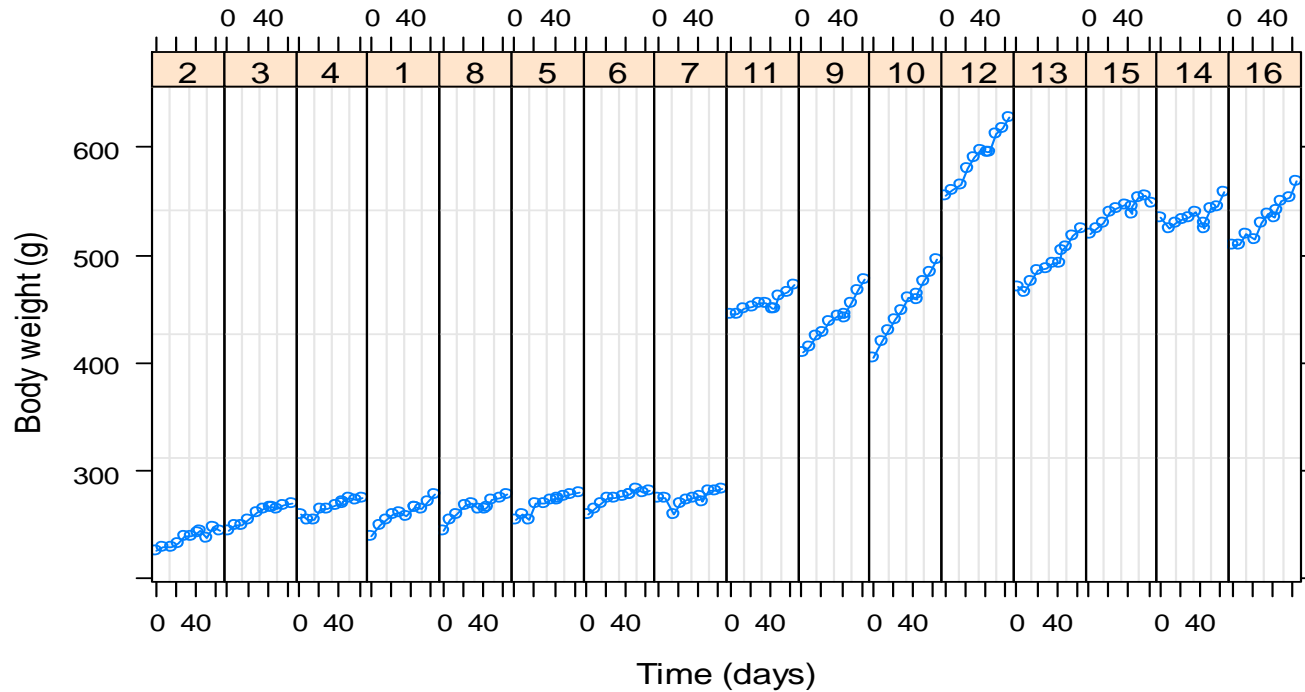
```
2   250   8  1  1
```

```
3   255  15  1  1
```

```
4   260  22  1  1
```

Aplicação 5 (cont.)

a) Execute o comando `plot(BodyWeight)` e comente.



Aplicação 5 (cont.)

b) Ajuste o modelo que lhe parecer adequado com a função *lme* do package *nlme*. Execute os comandos *summary*, *anova*, *ranef* e *fitted*. Explique como se obtém cada valor ajustado.

c) Uma vez no mesmo individuo são realizadas várias pesagens ao longo do tempo, o pressuposto de independência entre observações é posto em causa. Um vez que as observações não são igualmente espaçadas no tempo (note-se que se fez uma observação adicional no dia 44), é mais adequado considerar uma estrutura de correlação espacial para os erros aleatórios. No package *nlme* existem várias estruturas de correlação que podem ser aplicáveis ao caso em estudo, por exemplo, *corExp*, *corGaus*, *corSpher*. Efectue os comandos:

```
bodyw2.lme<-update(bodyw1.lme, corr=corExp(form=~Time))
```

```
bodyw3.lme<-update(bodyw1.lme, corr=corGaus(form=~Time))
```

```
bodyw4.lme<-update(bodyw1.lme, corr=corSpher(form=~Time)).
```

Com base no AIC e no BIC qual é a estrutura de correlação mais adequada?

d) O modelo seleccionado em c) difere significativamente do modelo ajustado em b)?