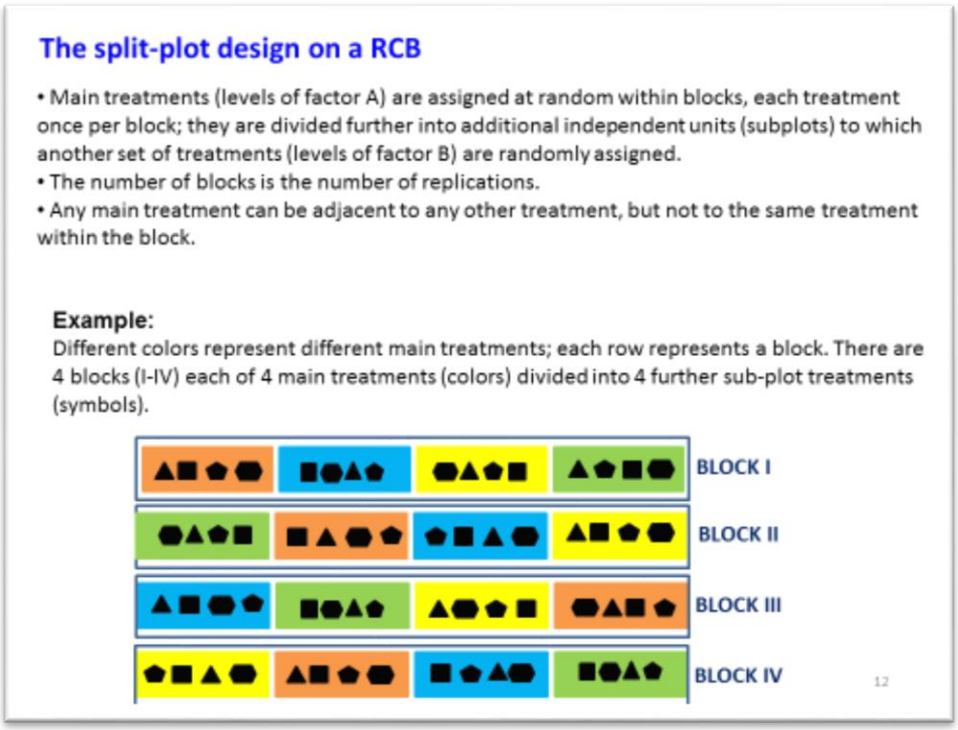
MODELOS MATEMÁTICOS E APLICAÇÕES 2015/2016

EXERCÍCIOS DE MODELOS LINEARES MISTOS – PARTE II

Uma das aplicações mais clássicas de modelos lineares mistos aplica-se à análise de dados provenientes de delineamentos em parcelas divididas (*split-plot designs*). Relembre o exemplo das aulas teóricas.



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 equilibrado 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* (acréscimo), 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 (acréscimo), 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 (acréscimo) (*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:

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

O Quadro de análise de variância para o modelo descrito apresenta-se seguidamente.

	G.L.	**S.Q.**	**QM**	**E[QM]**
Factor A	$a-1$	SQA	QMA	$c{\sigma_{\alpha u}}^2 + {\sigma_e}^2 + bc\dfrac{\sum_{i=1}^{a}(\alpha_i - \bar{\alpha}_.)^2}{a-1}$
Bloco	$b-1$	SQBL	QMBL	$ac{\sigma_u}^2 \; c{\sigma_{\alpha u}}^2 + {\sigma_e}^2$
Interacção FactorA×Bloco *(Whole-plot error)*	$(a-1)(b-1)$	SQWError	QMWError	$c{\sigma_{\alpha u}}^2 + {\sigma_e}^2$
Factor B	$c-1$	SQB	QMB	${\sigma_e}^2 + ab\dfrac{\sum_{k=1}^{c}\left(\beta_k - \bar{\beta}_.\right)^2}{c-1}$
Interacção FactorA×FactorB	$(a-1)(c-1)$	$SQAB$	$QMAB$	${\sigma_e}^2 + b\dfrac{\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\alpha\beta_{ik} - \overline{\alpha\beta}_{..}\right)^2}{(a-1)(c-1)}$
Resíduals	$a(b-1)(c-1)$	$SQRE$	$QMRE$	${\sigma_e}^2$

1. Descreva para este caso os testes de hipóteses aos efeitos da interacção A×B e aos efeitos principais dos dois factores (hipóteses, estatística do teste e região critica).

2. 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
…
```

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.
b) Execute o comando *plot.design (Alfalfa).*
c) Ajuste o modelo descrito no R recorrendo ao comando *lme* do pacote "*nlme*".
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)".*

3. No package "*nlme*" do R, existe um conjunto de dados (designado por "*Oats*").

```
> head(Oats)
Grouped Data: yield ~ nitro | Block
  Block     Variety nitro yield
1     I     Victory   0.0   111
2     I     Victory   0.2   130
3     I     Victory   0.4   157
4     I     Victory   0.6   174
5     I Golden Rain   0.0   117
6     I Golden Rain   0.2   114
…
```

Estes dados foram introduzidos por Yates (1935) como o exemplo de um delineamento em parcelas divididas. Pretende-se estudar se o rendimento (bushels/acre) da aveia é afectado pela variedade e concentrações de nitrogénio. Existem, portanto, dois factores em estudo: factor variedade de aveia, com 3 níveis (*Golden Rain*, *Marvellous* e *Victory*) e factor concentração de nitrogénio (cwt/acre), com 4 níveis (0.0, 0.2, 0.4 e 0.6). As unidades experimentais foram organizadas em 6 blocos, cada bloco com 3 grandes parcelas, cada uma dividida em 4 parcelas. As variedades de aveia foram aleatoriamente atribuídas às grandes parcelas e as concentrações de nitrogénio foram aleatoriamente atribuídas às parcelas divididas. Admite-se variedade e concentração de nitrogénio como factores de efeitos fixos e o bloco como factor de efeitos aleatórios.

a) Dada a natureza da variável resposta concentração de nitrogénio, esta pode ser vista como uma variável numérica (quantitativa). Ajuste o modelo que lhe parecer adequado no R com a função *lme*. O que conclui sobre a influência da variedade e da concentração de nitrogénio sobre o rendimento?
b) Repita a análise, mas agora considerando a concentração de nitrogénio como um factor (no R utilize *factor(nitro)* na descrição do modelo).

NOTA: 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 e (2) os efeitos do factor A como fixos e os efeitos do factor B como aleatórios.

(1) O Quadro de análise de variância para factor A e factor B de efeitos aleatórios

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

(2) O Quadro de análise de variância para factor A de efeitos fixos e factor B de efeitos aleatórios

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

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

$\nu_1 = \frac{(QMA+QMRE)^2}{\frac{(QMA)^2}{a-1}+\frac{(QMRE)^2}{a(b-1)(c-1)}}$, $\nu_2 = \frac{(\mathrm{QMWError}+QMAB)^2}{\frac{(\mathrm{QMWError})^2}{(a-1)(b-1)}+\frac{(QMAB)^2}{(a-1)(c-1)}}$