ANÁLISE DE DADOS COM MEDIDAS REPETIDAS

Prof. César Gonçalves de Lima (FZEA/USP)

cegdlima@usp.br

1. INTRODUÇÃO

Os **planejamentos com medidas repetidas** envolvem a realização de duas ou mais observações em cada unidade experimental (ou parcela), como por exemplo:

- **Planejamentos do tipo *split-plot*:** Surgiram na experimentação agronômica onde um único nível de um fator (ou tratamento) é aplicado a uma parcela relativamente grande de terra (*whole plot*) e todos os níveis de um segundo fator são aplicados as subparcelas (*split-plots*) dessa parcela maior. Os tratamentos primários são distribuídos às parcelas de acordo com um delineamento especificado (DIC, DBC, DQL etc.) e os tratamentos secundários são distribuídos *aleatoriamente* às subparcelas.

- **Planejamentos do tipo *cross-over*:** onde cada uma das unidades experimentais recebe uma seqüência de tratamentos. Estes planejamentos são comuns em estudos que envolvem vacas em lactação.

- **Planejamentos longitudinais:** Envolvem a observação de uma ou mais variáveis respostas em uma mesma unidade experimental em diversas ocasiões ou condições de avaliação (tempo, diferentes distâncias de uma origem etc). Como as medidas são *repetidas de modo sistemático*, espera-se que exista uma *correlação não nula* entre as medidas e uma *heterocedasticidade das variâncias* nas diversas ocasiões.

  Nos planejamentos longitudinais:

  - As *variáveis respostas* podem ser *contínuas* (ganho de peso, conversão alimentar etc.) ou *discretas* (número de perfilhos, presença de algum sintoma etc.);
  - As *unidades experimentais* (canteiros, baias com um ou mais animais, vasos, touceiras etc.) podem formar *grupos* ou *subpopulações* segundo um ou mais tratamentos ou fatores.
  - Cada unidade experimental pode gerar diversas *unidades de observação* e cada um desses conjuntos de unidades de observação pode ser entendido como um *perfil individual de respostas*.
  - A cada tratamento (ou grupo) está associado um *perfil médio de respostas* que deve evidenciar o efeito do tratamento e o seu comportamento *ao longo do tempo.*

- Os *dados longitudinais* são chamados de *regulares* (em relação ao tempo) se o intervalo entre duas medidas consecutivas quaisquer for constante ao longo do estudo e de *balanceados* (com relação ao tempo) se as observações forem feitas nos mesmos instantes de tempo em todas as unidades experimentais. A estrutura de dados é dita *completa* se não apresentar observações perdidas.

Quadro 1. Estrutura básica dos dados longitudinais balanceados e completos

Grupo ou Tratamento	**Unidade Experimental**	**Condições de Avaliação**			
		1	**2**	**…**	**t**
1	1	y_{111}	y_{112}	$\cdots$	y_{11t}
	2	y_{121}	y_{122}	$\cdots$	y_{12t}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_1	y_{1n_11}	y_{1n_12}	$\cdots$	y_{1n_1t}
2	1	y_{211}	y_{212}	$\cdots$	y_{21t}
	2	y_{221}	y_{222}	$\cdots$	y_{22t}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_2	y_{2n_21}	y_{2n_22}	$\cdots$	y_{2n_2t}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
g	1	y_{g11}	y_{g12}	$\cdots$	y_{g1t}
	2	y_{g21}	y_{g22}	$\cdots$	y_{g2t}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_g	y_{gn_g1}	y_{gn_g2}	$\cdots$	y_{gn_gt}

onde y_{ijk} é o valor da variável resposta da j-ésima unidade experimental dentro do i-ésimo tratamento, sob a k-ésima condição de avaliação (tempo, por exemplo), para i = 1, ..., g, j = 1,..., n_i, k = 1, ..., t e $N = \sum_{i=1}^{g} n_i$.

A cada unidade experimental (ij) está associado um vetor $\mathbf{y}_{ij}^{t} = [y_{ij1}, y_{ij2}, \ldots, y_{ijt}]$, de dimensão t, denominado ***perfil individual de respostas***, cujos componentes são os valores observados da variável resposta nas t ocasiões de avaliação.

Exemplo 1. Consideraremos para análise os dados de um experimento de corte, planejado de acordo com um *delineamento casualizado em blocos* (4 repetições), com três capins (Coastcross, Florona e Tifton 85). As medidas de MS (t/ha) foram feitas em 13 períodos de 4 semanas (6 períodos de verão e 7 períodos de inverno). As parcelas tinham tamanho 4 x 4m e as colheitas foram feitas mecanicamente.

Ano	Verão			Inverno							Verão		
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	x	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x	x	x	x	x

A idéia inicial de analisar esses dados utilizando as técnicas multi e univariadas usuais, comparando as produções dos capins em cada ano, dentro de cada ano comparar as estações e dentro das estações comparar as 6 ou 7 medidas realizadas, não evoluiu por restrição do modelo multivariado de análise, que pressupõe que o número de unidades experimentais (N = 4x3 = 12, neste caso) seja superior ao número de medidas repetidas (t = 13 + 13 = 26, neste caso).

Com o intuito de apresentarmos algumas técnicas de análise, calculamos as produções totais dos capins, por ano e por estação (t = 4 medidas repetidas), cujos dados estão apresentados na Tabela 1.

Tabela 1. Totais de MS (t/ha) dos capins Coastcross, Florona e Tifton 85, por estação e por ano

Capim	Bloco	Ano 1		Ano 2	
		Verão	Inverno	Verão	Inverno
Coastcross	1	13,16	5,85	10,70	4,07
	2	12,48	7,40	12,67	6,08
	3	14,74	6,70	12,51	5,48
	4	14,28	6,86	13,14	6,35
Florona	1	12,28	6,51	12,85	4,20
	2	13,06	6,25	12,49	4,76
	3	12,83	6,82	13,56	4,51
	4	14,01	7,65	13,56	5,02
Tifton 85	1	14,72	5,59	13,44	3,55
	2	15,28	4,90	14,14	3,71
	3	18,00	6,64	15,50	4,66
	4	19,20	8,50	13,84	4,77

Os gráficos apresentados nas Figuras 1 e 2 correspondem aos perfis individuais e médios de produções de MS, respectivamente.

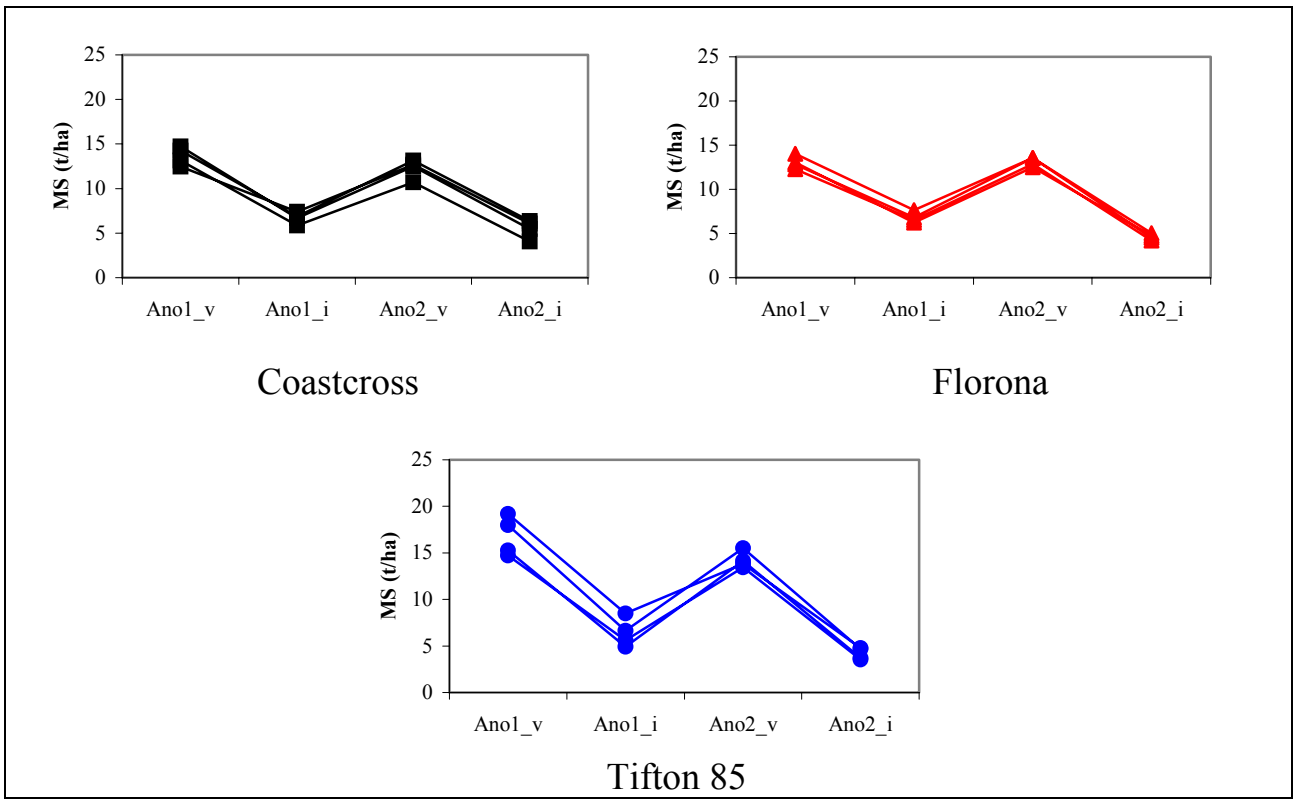


Figura 1. Produção total de MS (t/ha) por ano e por estação (verão e inverno) dos capins Coastcross, Florona e Tifton 85.

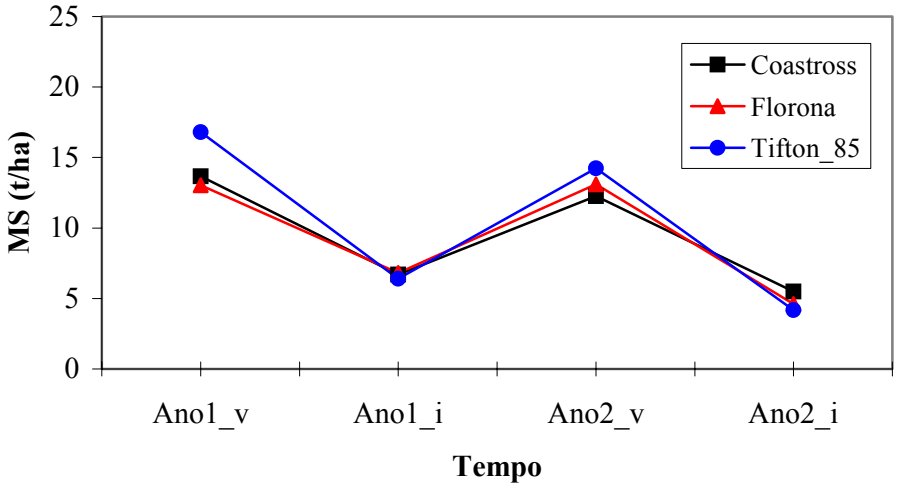


Figura 2. Produção média de MS (t/ha) dos capins Coastcross, Florona e Tifton 85, por ano e por estação.

Antes de entrarmos na discussão da análise de dados com medidas repetidas, cabe uma reflexão sobre o planejamento de experimentos, a escolha dos fatores (efeitos fixos ou aleatórios?) e dos níveis desses fatores. No Apêndice são apresentadas algumas considerações sobre este assunto.

2. ANÁLISE ESTATÍSTICA

As técnicas clássicas de análise de dados com medidas repetidas (dados longitudinais), geralmente, são dirigidas para o caso de dados *completos* e *balanceados* em relação ao tempo. Dentre as técnicas disponíveis, destacam-se: a Análise de Perfis (uni e multivariada) e a Análise de Curvas de Crescimento.

2.1. ANÁLISE DE PERFIS

É realizada com o objetivo de testar hipóteses sobre os perfis médios de respostas dos tratamentos, isto é, sobre os valores médios da variável resposta nas diferentes condições de observação (tempo). Basicamente, visa responder às perguntas:

i) Os perfis médios de resposta dos diferentes tratamentos são **paralelos**? (i.e., *a interação entre tratamento e tempo é nula?*)

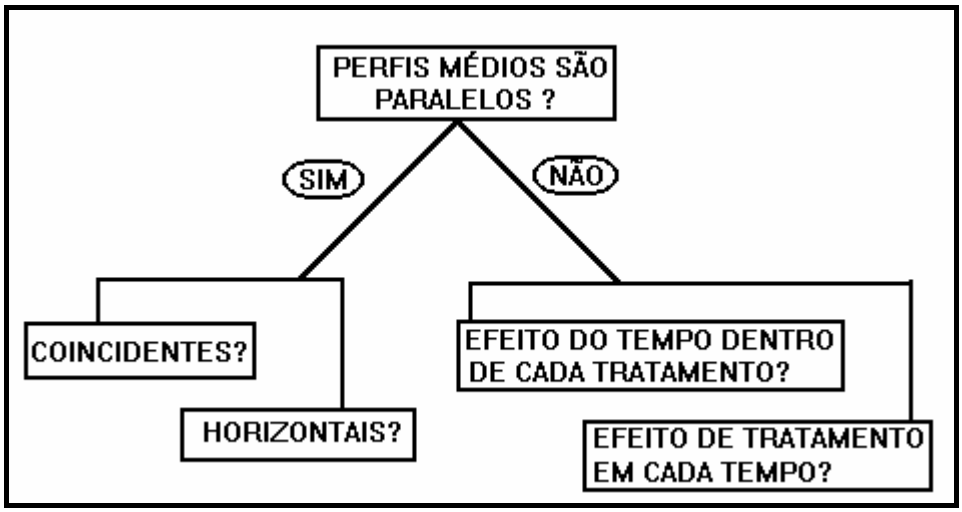
ii) Se os perfis são paralelos, eles são **coincidentes**? (i.e., *o efeito de tratamento é nulo?*)

iii) Se os perfis são paralelos, ele são **horizontais**? (i.e., *o efeito do tempo é nulo?*)

iv) Se os perfis não são paralelos, o efeito do tempo é nulo em cada um dos tratamentos?

v) Se os perfis não são paralelos, o efeito de tratamento é nulo em cada um dos tempos?

O esquema seguinte sintetiza as perguntas a serem respondidas através da análise de perfis:



Para maiores detalhes ver: STEEL & TORRIE (1980); AUBIN (1984); SINGER & ANDRADE (1986); MORRISON (1990); MILLIKEN & JOHNSON (1992) etc.

A análise de perfis pode ser feita utilizando-se *técnicas univariadas* ou *multivariadas* e a escolha por uma dessas técnicas depende das suposições que podemos admitir como verdadeiras para o conjunto de dados em estudo. Uma ferramenta muito útil nesses estudos é o PROC GLM do SAS.

2.1.1. ANÁLISE UNIVARIADA DE PERFIS

O modelo para a análise univariada de perfis (MILLIKEN & JOHNSON, 1984, cap. 26) é escrito como:

$$y_{ijk} = \mu + \alpha_i + \gamma_{ij} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk} \tag{1}$$

para $i = 1, ..., g$, $j = 1, ..., n_i$, $k = 1, ..., t$, onde

μ : constante comum a todas as observações,

α_i : efeito do i-ésimo tratamento,

γ_{ij} : erro associado às parcelas ,

β_k : efeito do k-ésimo tempo,

$(\alpha\beta)_{ik}$: efeito da interação do i-ésimo tratamento e k-ésimo tempo, e

ε_{ijk} : erro associado à observação y_{ijk}.

Por suposição, a constante μ e os efeitos α_i, β_k e $(\alpha\beta)_{ik}$ são considerados *fixos* e os erros γ_{ij} e ε_{ijk} são considerados *aleatórios*, normais e independentemente distribuídos com médias nulas e variâncias comuns σ_γ^2 e σ_ε^2, respectivamente.

Sob o modelo (1), tem-se que $E(y_{ijk}) = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik}$ e

$$Var(\mathbf{y}_{ij}) = \mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

onde $\sigma^2 = \sigma_\gamma^2 + \sigma_\varepsilon^2$ e $\rho = \sigma_\gamma^2/(\sigma_\gamma^2 + \sigma_\varepsilon^2)$, σ_γ^2 é a variância associada às parcelas e σ_ε^2 é a variância associada às subparcelas. Neste caso, diz-se que a matriz de covariâncias é do tipo *uniforme*, ou segue o *padrão de uniformidade*, ou tem a forma de *simetria composta* (a variância das respostas em qualquer um dos tempos é igual a $\sigma_\gamma^2 + \sigma_\varepsilon^2$ e a covariância entre dois tempos quaisquer é igual a σ_γ^2).

A análise feita através do modelo (1) corresponde à análise de um *experimento em parcelas subdivididas* ("split-plot") onde a causa de variação *entre indivíduos* (tratamento) é agrupada separadamente daquelas que fazem parte da variação *intra-indivíduos* (tempo e interação tratamento x tempo).

Apesar das facilidades de obtenção e de interpretação dos resultados dos testes das hipóteses, esta abordagem *não é recomendada para a análise de dados com medidas re-*

petidas, pois considerando o modo *sistemático* como são feitas as observações (ao longo do tempo, por exemplo) nas mesmas unidades experimentais, não se espera que a matriz de covariâncias, **Σ,** seja do tipo uniforme.

HUYNH & FELDT (1970) e ROUANET & LÉPINE (1970) mostraram que uma condição *suficiente* e *necessária* para que as estatísticas dos testes de hipóteses envolvendo as comparações *intra-indivíduos* tenham distribuição F exata é que a matriz Σ satisfaça a ***condição de esfericidade* ou *circularidade***, ou seja, que seus elementos $\sigma_{kk'}$ para k e k'=1, 2,...,t, satisfaçam a

$$\sigma_{kk'} = \begin{cases} a_k + a_{k'} + \lambda, & se\, k = k' \\ a_k + a_{k'} & \mathbf{,\ se\ k \neq k'} \end{cases}$$

com $\lambda > 0$, a_k e $a_{k'}$, constantes; ou ainda que a matriz de covariâncias Σ satisfaça a

$$\mathbf{P}^t \boldsymbol{\Sigma}\, \mathbf{P} = \lambda \mathbf{I}_{(t-1)},$$

onde **P** é uma matriz de contrastes ortogonais de dimensão t x (t – 1). ANDRADE & SINGER (1986) afirmaram que "apesar dessa estrutura não ser tão restritiva quanto à estrutura uniforme, ela também não é totalmente adequada para representar a estrutura de covariâncias de dados longitudinais".

MORRISON (1990) descreveu o teste de MAUCHLY (1940) para a validade da condição de esfericidade da matriz de covariâncias, Σ, cuja estatística é

$$\mathbf{W} = \frac{(t-1)^{t-1} \left|\mathbf{P}^t \mathbf{S} \mathbf{P}\right|}{\left\{tr(\mathbf{P}^t \mathbf{S} \mathbf{P})\right\}^{t-1}}$$

onde **S** é a matriz de covariâncias amostrais, de dimensões txt e *tr* é o operador traço. Sob a hipótese de que a matriz de covariâncias satisfaz à condição de esfericidade, a estatística:

$$\chi^2 = -\left[\nu - \frac{2\,\mathrm{t}^2 - 3t + 3}{6(t-1)}\right] \ln(\mathbf{W}),$$

tem distribuição quiquadrado com $f = \frac{1}{2} t(t-1) - 1$ graus de liberdade, quando ν (número de graus de liberdade do resíduo) é grande. CROWDER & HAND (1990) consideraram este teste "o mais popular para verificar a condição de circularidade" e alertaram para o fato de que, como todos os testes que envolvem variâncias e covariâncias, este teste é bastante sensível à não normalidade dos dados.

Para verificar a condição de esfericidade da matriz Σ pode-se utilizar o teste desenvolvido por MAUCHLY (1940), que está disponível no PROC GLM do SAS, com o comando **`repeated`** e a opção **`printe`**.

A análise de dados de um experimento em parcelas subdivididas pode ser implementada usando o PROC GLM, utilizando-se o comando **`random`** (dados no formato univariado) ou o comando **`repeated`** (dados no formato multivariado).

EXEMPLO: De início, a partir dos dados da Tabela 1, criamos os arquivos de dados **`MSTotal_multi`** e **`MSTotal_uni`**, que correspondem aos formatos multivariado (as medidas repetidas aparecem nas colunas) e univariado (as medidas repetidas são indexadas pelo fator Tempo), utilizando os seguintes comandos:

```
data MSTotal_multi (keep = Capim Bloco Ano1_v Ano1_i Ano2_v Ano2_i)
     MSTotal_uni (keep = Capim Bloco Tempo MStotal);
 input Capim$ 1-9 Bloco Ano1_v Ano1_i Ano2_v Ano2_i;
 output MSTotal_multi;
 MSTotal = Ano1_v; Tempo = 'Ano1_v'; output MSTotal_uni;
 MSTotal = Ano1_i; Tempo = 'Ano1_i'; output MSTotal_uni;
 MSTotal = Ano2_v; Tempo = 'Ano2_v'; output MSTotal_uni;
 MSTotal = Ano2_i; Tempo = 'Ano2_i'; output MSTotal_uni;
 cards;
Coastross  1   13.16  5.85  10.70  4.07
Coastross  2   12.48  7.40  12.67  6.08
Coastross  3   14.74  6.70  12.51  5.48
Coastross  4   14.28  6.86  13.14  6.35
Florona    1   12.28  6.51  12.85  4.20
Florona    2   13.06  6.25  12.49  4.76
Florona    3   12.83  6.82  13.56  4.51
Florona    4   14.01  7.65  13.56  5.02
Tifton_85  1   14.72  5.59  13.44  3.55
Tifton_85  2   15.28  4.90  14.14  3.71
Tifton_85  3   18.00  6.64  15.50  4.66
Tifton_85  4   19.20  8.50  13.84  4.77
;
```

A análise univariada de perfis pode ser implementada pelos comandos:

```
proc glm data=MSTotal_uni;
 class Bloco Capim Tempo;
 model MSTotal = Bloco Capim Bloco*Capim Tempo Capim*Tempo / ss3;
 random bloco bloco*Capim / test;
run;
```

Quadro 1. Análise univariada dos perfis (parcelas subdivididas) dos dados de MS (t/ha), utilizando o PROC GLM com o comando **`random`**

```
                                                                    (a)
Class Level Information

Class          Levels    Values
Bloco               4    1 2 3 4
Capim               3    Coastross Florona Tifton_85
Tempo               4    Ano1_i Ano1_v Ano2_i Ano2_v
```

(b)

```
Number of observations    48
Dependent Variable: MSTotal
                                          Sum of
Source                DF         Squares     Mean Square    F Value    Pr > F
Model                 20     898.1619417      44.9080971      85.44    <.0001
Error                 27      14.1911562       0.5255984
Corrected Total       47     912.3530979

R-Square     Coeff Var      Root MSE    MSTotal Mean
0.984446      7.415586      0.724982        9.776458

Source                 DF     Type III SS     Mean Square    F Value    Pr > F
Bloco                   3      20.3022063       6.7674021      12.88    <.0001
Capim                   2       9.5454042       4.7727021       9.08    0.0010
Bloco*Capim             6       5.1811125       0.8635188       1.64    0.1739
Tempo                   3     828.4560729     276.1520243     525.41    <.0001
Capim*Tempo             6      34.6771458       5.7795243      11.00    <.0001
```

(c)

```
Source           Type III Expected Mean Square
Bloco            Var(Error) + 4 Var(Bloco*Capim) + 12 Var(Bloco)
Capim            Var(Error) + 4 Var(Bloco*Capim) + Q(Capim,Capim*Tempo)
Bloco*Capim      Var(Error) + 4 Var(Bloco*Capim)
Tempo            Var(Error) + Q(Tempo,Capim*Tempo)
Capim*Tempo      Var(Error) + Q(Capim*Tempo)
```

(d)

```
Tests of Hypotheses for Mixed Model Analysis of Variance
Dependent Variable: MSTotal

  Source               DF     Type III SS     Mean Square    F Value    Pr > F
  Bloco                 3       20.302206        6.767402       7.84    0.0169
* Capim                 2        9.545404        4.772702       5.53    0.0435
  Error                 6        5.181113        0.863519
Error: MS(Bloco*Capim)
* This test assumes one or more other fixed effects are zero.


  Source               DF     Type III SS     Mean Square    F Value    Pr > F
  Bloco*Capim           6        5.181113        0.863519       1.64    0.1739
* Tempo                 3      828.456073      276.152024     525.41    <.0001
  Capim*Tempo           6       34.677146        5.779524      11.00    <.0001
  Error: MS(Error)     27       14.191156        0.525598
* This test assumes one or more other fixed effects are zero.
```

(a) Indica o nome, o número e os níveis das variáveis classificatórias (Bloco, Capim e Tempo), além no número de observações usadas na análise.

(b) Apresenta o quadro de ANOVA, mas o teste para Capim utiliza *erroneamente* o QMError como denominador da estatística F.

(c) Apresenta as *E*(QM) de todas as fontes de variação do modelo, indicando que os testes para Bloco e Capim devem utilizar QM(Bloco*Capim) como denominador da estatística F.

(d) Neste novo quadro de ANOVA todos os testes são feitos corretamente. Observe que o valor (*correto!*) da estatística F para Capim é F = 5,53 (p = 0,0435) e não F = 9,08 (p = 0,0010), como apresentado no quadro (1).

Comentário: Esta técnica de análise deve ser usada com cautela, já que pressupõe que as variâncias das medidas feitas em cada uma das ocasiões sejam idênticas e que a correlação entre as medidas de quaisquer duas dessas ocasiões também seja iguais.

2.1.2. ANÁLISE MULTIVARIADA DE PERFIS

O modelo usado para Análise Multivariada de Perfis (geralmente) é parametrizado através das médias de caselas e tem a vantagem de proporcionar uma grande facilidade de interpretação. Esse modelo pode ser representado matricialmente na forma usual da Análise Multivariada de Variância (MANOVA), isto é,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

onde

$\mathbf{Y}_{(Nxt)} = [\mathbf{y}_{11},\ldots, \mathbf{y}_{gn_g}]'$ é uma matriz de dados, como a representada no **Quadro 1;**

$\mathbf{y}_{ij} = [y_{ij1}, y_{ij2}, ..., y_{ijt}]'$ é o perfil de respostas da unidade experimental (i,j);

$\mathbf{X}_{(Nxg)} = \begin{bmatrix} 1_{n_1} & 0 & \cdots & 0 \\ 0 & 1_{n_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1_{n_g} \end{bmatrix}$ é a matriz de especificação do modelo, onde $\mathbf{1}_{n_i}$ é um vetor com n_i uns;

$\boldsymbol{\beta}_{(gxt)} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1t} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2t} \\ \vdots & \vdots & & \vdots \\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gt} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}'_1 \\ \boldsymbol{\mu}'_2 \\ \vdots \\ \boldsymbol{\mu}'_g \end{bmatrix}$ é a matriz de parâmetros; onde μ_{ik} representa a média das parcelas submetidas ao i-ésimo tratamento no k-ésimo tempo e $\boldsymbol{\mu}_i'$ representa o perfil médio de respostas do i-ésimo tratamento, e

$\boldsymbol{\varepsilon}_{(Nxt)} = [\boldsymbol{\varepsilon}_{11}\ \boldsymbol{\varepsilon}_{12}\ \ldots\ \boldsymbol{\varepsilon}_{gn_g}]'$ é a matriz de erros, onde $\boldsymbol{\varepsilon}_{ij} = [\varepsilon_{ij1}\ \varepsilon_{ij2}\ \ldots\ \varepsilon_{ijt}]'$.

Para efeito de inferência, supõe-se que os N perfis $\mathbf{y}_{ij} \overset{iid}{\sim} N_t(\mathbf{X}\boldsymbol{\beta}; \boldsymbol{\Sigma})$, onde

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1t} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{1t} \\ \vdots & \vdots & & \vdots \\ \sigma_{1t} & \sigma_{2t} & \cdots & \sigma_t^2 \end{bmatrix}$$

é uma matriz de variâncias e covariâncias chamada *não estruturada* ou *completamente parametrizada* e tem **t(t+1)/2** parâmetros. Note que se precisa admitir que todos os grupos (tratamentos) devem ter a mesma matriz de variâncias e covariâncias, Σ.

Qualquer hipótese sobre os parâmetros pode ser expressa na forma linear geral:

$$\boldsymbol{H}\mathbf{:}\ \mathbf{C}\beta\mathbf{U} = \mathbf{0},$$

onde **C** (cxg) e **U** (txu) são matrizes de constantes conhecidas e de postos *c* e *u*, respectivamente. Vale observar que a matriz **C** é responsável por comparações entre os tratamentos (*linhas da matriz* β) e a matriz **U**, por comparações entre as ocasiões de observação (*colunas da matriz* β). Por exemplo, a hipótese de *paralelismo dos perfis médios de respostas* pode ser expressa como:

$$\boldsymbol{H}_{01}\mathbf{:} \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(t-1)} - \mu_{1t} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(t-1)} - \mu_{2t} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{g1} - \mu_{g2} \\ \mu_{g2} - \mu_{g3} \\ \vdots \\ \mu_{g(t-1)} - \mu_{gt} \end{bmatrix}$$

e é equivalente à hipótese de não existência da interação entre tratamentos e tempo. Na forma da hipótese linear geral utilizam-se as matrizes (*que não são únicas!*):

$$\mathbf{C}_{((g-1)\ x\ g)} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \quad \text{e} \quad \mathbf{U}_{(t\ x\ (t-1))} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & -1 \end{bmatrix}$$

Quando a hipótese de paralelismo dos perfis não é aceita, MORRISON (1990) sugeriu que se teste a igualdade das respostas médias dos tratamentos *separadamente* para cada um dos t tempos, através de t análises de variâncias univariadas. SINGER & ANDRADE (1986) sugeriram que, neste caso, seja feito um estudo da *natureza da interação* através de um exame da estrutura dos tratamentos.

O teste da hipótese linear geral ***H:*** **CβU = 0** pode ser feito através das estatísticas: *Lambda de Wilks, Traço de Pillai, Traço de Hotelling-Lawley e Maior Raiz Característica de Roy*.

Essas estatísticas são funções das raízes características da matriz $\mathbf{HE}^{-1}$, onde **H** é a *matriz de somas de quadrados e produtos cruzados devido à hipótese nula* e **E** é a *matriz de somas de quadrados e produtos cruzados devido ao erro*. Detalhes sobre essas estatísticas, bem como tabelas apropriadas, podem ser encontradas em TIMM (1980) ou MORRISON (1990), dentre outros. Aproximações assintóticas através de distribuições quiquadrado ou F-Snedecor são consideradas em SEBER (1984), dentre outros.

A análise multivariada de perfis pode ser facilmente implementada utilizando-se os pacotes estatísticos mais comuns, como o BMDP, MINITAB, SAS e SPSS, através de *procedimentos de medidas repetidas* ou de procedimentos de MANOVA, especificando-se corretamente as matrizes **C** e **U**. No PROC GLM do SAS, as hipóte-ses de interesse são escritas na forma ***H:*** **LBM = 0** e a especificação das matrizes **L** e **M** (que correspondem às matrizes **C** e **U**, respectivamente) é feita nos comandos *contrast* e *manova*.

É importante salientar que a aplicação da Análise Multivariada de Perfis apresenta algumas restrições:

i) Só pode ser usada quando N > t (número de unidades experimentais maior que o número de ocasiões);

ii) Necessidade de *perfis de dados completos* (na perda de uma ou mais observações para um mesmo indivíduo, todo o perfil de respostas deste indivíduo é excluído da análise);

iii) O pequeno poder dos testes;

iv) As diferentes estatísticas de teste podem levar a conclusões diferentes (*Dica*: assumir como verdadeiro, o resultado mais comum entre as estatísticas) .

A forma mais simples de implementar a análise multivariada de perfis no PROC GLM consiste em utilizar o comando `repeated` e o conjunto de dados no *formato multivariado* (veremos um exemplo de aplicação na próxima seção)

2.1.3 SOLUÇÃO UNIVARIADA APROXIMADA

Quando a *condição de esfericidade da matriz de covariâncias* **Σ** *não está satisfeita*, BOX (1954) e GEISSER & GREENHOUSE (1958) propuseram o uso de *soluções univariadas aproximadas*, que envolvem a correção do número de graus de liberdade das estatísticas dos testes que envolvem as comparações *intra-indivíduos* por um fator multiplicativo de correção **ε**.

GREENHOUSE & GEISSER (1959) e HUYHN & FELDT (1976) propuseram estimadores para este fator de correção, que são baseados na matriz de covariâncias amostrais. As estimativas G-G e H-F do fator de correção estão disponíveis no PROC GLM com o uso do comando **repeated** (dados do formato multivariado) e a opção **printe**.

Vale observar que, para *qualquer que seja a estrutura de* Σ, as estatísticas de testes envolvendo as comparações *entre indivíduos* (tratamentos, p.ex.) *sempre* terão distribuição F exata.

EXEMPLO. A análise de perfis (uni e multivariada) dos dados do Exemplo 1 será feita utilizando-se o PROC GLM com o comando **repeated**.

```
proc glm data=MSTotal_Multi;
 class Bloco Capim;
 model Ano1_v Ano1_i Ano2_v Ano2_i = Bloco Capim / nouni;
 manova / printe;
 repeated Tempo 4 polynomial / summary printe;
 lsmeans Capim / stderr;
run;
```

Quadro 2. Análise univariada dos perfis dos dados de MS (t/ha), utilizando o PROC GLM com o comando **repeated**

```
                                                                                (a)
Class Level Information

Class          Levels    Values
Bloco               4    1 2 3 4
Capim               3    Coastross Florona Tifton_85
Number of observations    12


                                                                                (b)
Multivariate Analysis of Variance

                            E = Error SSCP Matrix

                  Ano1_v            Ano1_i            Ano2_v            Ano2_i
Ano1_v      5.7452666667      2.3280333333           -1.7763      0.0593833333
Ano1_i      2.3280333333      4.5996666667           -0.1544      1.2722666667
Ano2_v           -1.7763           -0.1544      2.8037333333      1.3448833333
Ano2_i      0.0593833333      1.2722666667      1.3448833333           1.42805
                                                                                (c)
           Repeated Measures Level Information

Dependent Variable      Ano1_v   Ano1_i   Ano2_v   Ano2_i
    Level of Tempo           1        2        3        4

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 6         Ano1_v         Ano1_i         Ano2_v         Ano2_i

Ano1_v       1.000000       0.452867      -0.442581       0.020732
                              0.3075         0.3200         0.9648
```

Ano1_i	0.452867	1.000000	-0.042995	0.496413
	0.3075		0.9271	0.2571
Ano2_v	-0.442581	-0.042995	1.000000	0.672116
	0.3200	0.9271		0.0981
Ano2_i	0.020732	0.496413	0.672116	1.000000
	0.9648	0.2571	0.0981	

(d)

Sphericity Tests

Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	5	0.1368233	9.3928077	0.0944
Orthogonal Components	5	0.1368233	9.3928077	0.0944

(e)

Manova Test Criteria and Exact F Statistics for the **Hypothesis of no Tempo Effect**
H = Type III SSCP Matrix for tempo E = Error SSCP Matrix
S=1 M=0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.0006143	2169.10	3	4	<.0001
Pillai's Trace	0.9993857	2169.10	3	4	<.0001
Hotelling-Lawley Trace	1626.8270947	2169.10	3	4	<.0001
Roy's Greatest Root	1626.8270947	2169.10	3	4	<.0001

(f)

Manova Test Criteria and F Approximations for the Hypothesis of **no Tempo*Bloco Effect** H = Type III SSCP Matrix for tempo*Bloco E = Error SSCP Matrix
S=3 M=-0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.17774075	1.14	9	9.8856	0.4207
Pillai's Trace	1.11757755	1.19	9	18	0.3597
Hotelling-Lawley Trace	3.09433323	0.92	9	8	0.5543
Roy's Greatest Root	2.56797771	5.14	3	6	0.0428

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

(g)

Manova Test Criteria and F Approximations for the Hypothesis of **no Tempo*Capim Effect** H = Type III SSCP Matrix for tempo*Capim E = Error SSCP Matrix
S=2 M=0 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.00619834	15.60	6	8	0.0005
Pillai's Trace	1.68346284	8.86	6	10	0.0016
Hotelling-Lawley Trace	49.06804373	24.53	6	6	0.0006
Roy's Greatest Root	46.68470554	77.81	3	5	0.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

(h)

Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Bloco	3	20.30220625	6.76740208	7.84	0.0169
Capim	2	9.54540417	4.77270208	5.53	0.0435
Error	6	5.18111250	0.86351875		

(i)

Repeated Measures Analysis of Variance

Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F	Adj Pr > F G - G	Adj Pr > F H - F
Tempo	3	828.4560729	276.1520243	529.05	<.0001	<.0001	<.0001
Tempo*Bloco	9	4.7955521	0.5328391	1.02	0.4602	0.4553	0.4602
Tempo*Capim	6	34.6771458	5.7795243	11.07	<.0001	0.0012	<.0001
Error(Tempo)	18	9.3956042	0.5219780				

Greenhouse-Geisser Epsilon 0.5712
Huynh-Feldt Epsilon 1.4435

(j)

Analysis of Variance of Contrast Variables

Tempo_N represents the nth degree polynomial contrast for Tempo

Contrast Variable: Tempo_1

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Mean	1	308.0627004	308.0627004	384.02	<.0001	
Bloco	3	3.1115812	1.0371937	1.29	0.3597	
Capim	2	16.4317008	8.2158504	10.24	0.0116	**(*)**
Error	6	4.8132425	0.8022071			

Contrast Variable: Tempo_2

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Mean	1	0.98326875	0.98326875	2.93	0.1377	
Bloco	3	0.44993958	0.14997986	0.45	0.7283	
Capim	2	4.25498750	2.12749375	6.34	0.0331	**(*)**
Error	6	2.01222917	0.33537153			

Contrast Variable: Tempo_3

Source	DF	Type III SS	Mean Square	F Value	Pr > F	
Mean	1	519.4101038	519.4101038	1212.57	<.0001	
Bloco	3	1.2340313	0.4113438	0.96	0.4699	
Capim	2	13.9904575	6.9952287	16.33	0.0037	**(*)**
Error	6	2.5701325	0.4283554			

(k)

Least Squares Means

Capim	Ano1_v LSMEAN	Standard Error	Pr > |t|
Coastross	13.6650000	0.4892710	<.0001
Florona	13.0450000	0.4892710	<.0001
Tifton_85	16.8000000	0.4892710	<.0001

Capim	Ano1_i LSMEAN	Standard Error	Pr > |t|
Coastross	6.70250000	0.43778166	<.0001
Florona	6.80750000	0.43778166	<.0001
Tifton_85	6.40750000	0.43778166	<.0001

Capim	Ano2_v LSMEAN	Standard Error	Pr > |t|
Coastross	12.2550000	0.3417927	<.0001
Florona	13.1150000	0.3417927	<.0001
Tifton_85	14.2300000	0.3417927	<.0001

Capim	Ano2_i LSMEAN	Standard Error	Pr > |t|
Coastross	5.49500000	0.24393049	<.0001
Florona	4.62250000	0.24393049	<.0001
Tifton_85	4.17250000	0.24393049	<.0001

(a) Indica o nome, o número e os níveis das variáveis classificatórias, além no número de observações usadas na análise.

(b) Estimativa da matriz de somas de quadrados e duplos produtos residuais.

(c) Apresenta o fator que define as ocasiões (Tempo) e seus níveis, a correlação entre as ocasiões e o nível descritivo do teste H_0: $\rho(y_{i\bullet}; y_{j\bullet}) = 0$.

(d) Mostra o resultado do teste de que (H_0:) a matriz de covariâncias Σ satisfaz a condição de esfericidade. Neste caso, essa hipótese nula não é rejeitada ($p = 0{,}0944$), indicando que a análise de perfis pode ser feita utilizando uma *abordagem univariada*.

(e) O efeito do tempo resultou significativo ($p < 0{,}0001$)

(f) A interação entre os fatores tempo e bloco resultou não significativa ($p > 0{,}05$)

(g) A interação entre os fatores tempo e sexo resultou significativa ($p < 0{,}01$) indicando que os *perfis médios de respostas não são paralelos*.

(h) Apresenta os resultados dos testes (univariados) feitos nas parcelas. O efeito de capim foi significativo ($p = 0{,}0435$).

(i) Apresenta os resultados dos testes univariados (exato e aproximado) feitos nas sub-parcelas. Foram significativos os efeitos da interação Tempo*Capim ($p < 0{,}0001$) e do Tempo ($p < 0{,}0001$). Também apresenta as estimativas dos fatores de correção G-G e H-F.

(j) Mostra os resultados dos testes de tendência das respostas médias, indicando que um polinômio de terceiro grau explica bem o comportamento das médias ao longo do tempo.

(k) Apresenta as médias de mínimos quadrados (e respectivos erros padrões) dos três capins nas quatro ocasiões.

Para compararmos a média de Tifton 85 com a média dos capins Coastcross e Florona e as médias dos capins Coastcross e Florona, no verão do primeiro ano, usamos os comandos:

```
contrast 'Ano1_v: Tifton 85 vs. (Coastcross e Florona)' Capim -1 -1 2;
contrast 'Ano1_v: Coastcross vs. Florona' Capim 1 -1 0;
manova H = Capim M = (1 0 0 0);
```

Resultando em:

```
                                                                    (1)
MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Ano1_v:
Tifton 85 vs. (Coastcross e Florona) Effect on the Variables Defined by the M Matrix
Transformation
H = Contrast SSCP Matrix for Ano1_v: Tifton 85 vs. (Coastcross e Florona)
E = Error SSCP Matrix
```

```
                                  S=1    M=-0.5    N=2

Statistic                       Value    F Value    Num DF    Den DF    Pr > F
Wilks' Lambda               0.15364414      33.05         1         6    0.0012
Pillai's Trace              0.84635586      33.05         1         6    0.0012
Hotelling-Lawley Trace      5.50854617      33.05         1         6    0.0012
Roy's Greatest Root         5.50854617      33.05         1         6    0.0012
                                                                                (m)
MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Ano1_v:
Coastcross vs Florona Effect on the Variables Defined by the M Matrix Transformation
H = Contrast SSCP Matrix for Ano1_v: Coastcross vs Florona E = Error SSCP Matrix
                                  S=1    M=-0.5    N=2

Statistic                       Value    F Value    Num DF    Den DF    Pr > F
Wilks' Lambda               0.88197849       0.80         1         6    0.4047
Pillai's Trace              0.11802151       0.80         1         6    0.4047
Hotelling-Lawley Trace      0.13381450       0.80         1         6    0.4047
Roy's Greatest Root         0.13381450       0.80         1         6    0.4047
```

(l) A produção média de Tifton 85 foi superior (p = 0,0012) à produção média de Coastcross e Florona, no verão do primeiro ano.

(m) As produções médias de Coastcross e Florona podem ser consideradas iguais (p = 0,4047), no verão do primeiro ano.

EXEMPLO: A análise dos dados do Exemplo 1 pode ser feita adotando-se um modelo de parcelas sub-sub-divididas (*split-split-plot*), se admitirmos que as medidas foram repetidas em dois níveis: primeiramente nos anos (medida feita nas sub-parcelas) e depois das estações (medida feita nas sub-sub-parcelas). Neste caso, devemos utilizar a *formatação univariada* dos dados e substituir o fator Tempo pelos fatores Ano e Estação. Assim, as observações indexadas pelo nível Tempo = Ano1_v, passarão a ser indexadas por Ano = 1 e Estação = verão.

Os comandos do PROC GLM para a referida análise são os seguintes:

```
proc glm data=MS_Total;
 class Bloco Capim Ano Estacao;
 model MS_Total = Bloco Capim Bloco*Capim Ano Capim*Ano Bloco*Capim*Ano
                  Estacao Capim*Estacao Capim*Ano*Estacao / ss3;
 random Bloco Bloco*Capim Bloco*Capim*Ano / test;
run;
```

Quadro 3. Análise dos dados de MS (t/ha) baseado num modelo em parcelas sub-sub-divididas, utilizando o PROC GLM com o comando **`repeated`**

```
                                                                              (a)
Class          Levels    Values
Bloco               4    1 2 3 4
Capim               3    Coastross Florona Tifton_85
Ano                 2    1 2
Estacao             2    inverno verão
Number of observations    48
```

(b)

Dependent Variable: MS_total

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	906.8202104	31.2696624	101.73	<.0001
Error	18	5.5328875	0.3073826		
Corrected Total	47	912.3530979			

R-Square	Coeff Var	Root MSE	MS_total Mean
0.993936	5.670980	0.554421	9.776458

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Bloco	3	20.3022063	6.7674021	22.02	<.0001
Capim	2	9.5454042	4.7727021	15.53	0.0001
Bloco*Capim	6	5.1811125	0.8635188	2.81	0.0415
Ano	1	30.3213021	30.3213021	98.64	<.0001
Capim*Ano	2	4.0912542	2.0456271	6.65	0.0069
Bloco*Capim*Ano	9	8.6582687	0.9620299	3.13	0.0188
Estacao	1	797.1515021	797.1515021	2593.35	<.0001
Capim*Estacao	2	26.3309042	13.1654521	42.83	<.0001
Ano*Estacao	1	0.9832688	0.9832688	3.20	0.0905
Capim*Ano*Estacao	2	4.2549875	2.1274937	6.92	0.0059

(c)

Source	Type III Expected Mean Square
Bloco	Var(Error) + 2 Var(Bloco*Capim*Ano) + 4 Var(Bloco*Capim) + 12 Var(Bloco)
Capim	Var(Error) + 2 Var(Bloco*Capim*Ano) + 4 Var(Bloco*Capim) + Q(Capim,Capim*Ano,Capim*Estacao,Capim*Ano*Estacao)
Bloco*Capim	Var(Error) + 2 Var(Bloco*Capim*Ano) + 4 Var(Bloco*Capim)
Ano	Var(Error) + 2 Var(Bloco*Capim*Ano) + Q(Ano,Capim*Ano,Ano*Estacao,Capim*Ano*Estacao)
Capim*Ano	Var(Error) + 2 Var(Bloco*Capim*Ano) + Q(Capim*Ano,Capim*Ano*Estacao)
Bloco*Capim*Ano	Var(Error) + 2 Var(Bloco*Capim*Ano)
Estacao	Var(Error) + Q(Estacao,Capim*Estacao,Ano*Estacao,Capim*Ano*Estacao)
Capim*Estacao	Var(Error) + Q(Capim*Estacao,Capim*Ano*Estacao)
Ano*Estacao	Var(Error) + Q(Ano*Estacao,Capim*Ano*Estacao)
Capim*Ano*Estacao	Var(Error) + Q(Capim*Ano*Estacao

(d)

Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: MS_total

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	Bloco	3	20.302206	6.767402	7.84	0.0169
*	Capim	2	9.545404	4.772702	5.53	0.0435
	Error	6	5.181113	0.863519		

Error: MS(Bloco*Capim)

** This test assumes one or more other fixed effects are zero.*

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	Bloco*Capim	6	5.181113	0.863519	0.90	0.5355
*	Ano	1	30.321302	30.321302	31.52	0.0003
*	Capim*Ano	2	4.091254	2.045627	2.13	0.1753
	Error	9	8.658269	0.962030		

Error: MS(Bloco*Capim*Ano)

** This test assumes one or more other fixed effects are zero.*

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
	Bloco*Capim*Ano	9	8.658269	0.962030	3.13	0.0188
*	Estacao	1	797.151502	797.151502	2593.35	<.0001
*	Capim*Estacao	2	26.330904	13.165452	42.83	<.0001
*	Ano*Estacao	1	0.983269	0.983269	3.20	0.0905
	Capim*Ano*Estacao	2	4.254987	2.127494	6.92	0.0059
	Error: MS(Error)	18	5.532888	0.307383		

* This test assumes one or more other fixed effects are zero.

Nota-se em (c) que, além da interação tripla (Capim*Ano*Estação) também resultaram significativas, a interação dupla Capim*Estação e os efeitos principais de Capim, Ano e Estação.

RESUMINDO:

1) A análise de perfis de dados com medidas repetidas deve ser feita utilizando-se preferencialmente, a técnica multivariada, que admite uma matriz de covariâncias não estruturada.
2) *Se a condição de esfericidade da matriz de covariâncias for satisfeita* (ou seja, se o teste de Mauchly resultar não significativo), podemos analisar os dados utilizando uma técnica univariada, como a Análise Univariada dos Perfis, admitindo um modelo de análise de parcelas subdivididas.
3) *Se a condição de esfericidade da matriz de covariâncias não for satisfeita* (se o teste de Mauchly resultar significativo), o ideal é realizar uma Análise Multivariada de Perfis. Uma alternativa univariada consiste em corrigir os graus de liberdade das estatísticas dos testes envolvendo as comparações *intra-indivíduos* (dentro das subparcelas) realizando uma *análise univariada aproximada.*
4) Uma abordagem mais atual consiste em analisar os dados através de *Modelos de Efeitos Aleatórios.* Esses modelos permitem a modelagem da matriz de covariâncias com um número bem menor de parâmetros que a dos modelos multivariados gerais e podem ser usados quando os dados não são balanceados em relação ao tempo. Nesse caso, pode-se utilizar o PROC MIXED do SAS na análise dos dados (mais detalhes na seção 2.4).

2.3. ANÁLISE DE CURVAS DE CRESCIMENTO

Segundo SINGER & ANDRADE (1986), o modelo para análise de curvas (polinomiais) de crescimento proposto por POTTHOFF & ROY (1964) pode ser considerado um paradigma nessa área. Esse modelo corresponde a uma generalização do modelo usual de análise de variância multivariada (MANOVA), no sentido de permitir a expressão dos perfis médios de respostas através de uma forma polinomial. O *modelo multivariado de curvas de crescimento* pode ser escrito como

$$\mathbf{Y_0} = \mathbf{X}\,\boldsymbol{\beta}\,\mathbf{G} + \boldsymbol{\varepsilon} \qquad (3)$$

onde:

$\mathbf{Y_0}$ (nxt) é a matriz das observações;

$\mathbf{X}$ (nxg) é a matriz do delineamento, com $n > g$, constituída de 0's e 1's, de forma a associar cada unidade experimental ao respectivo grupo;

$\boldsymbol{\beta}$ (gxr) é a matriz de parâmetros desconhecidos das curvas polinomiais, cujos elementos das linhas são os coeficientes dos polinômios de grau $r - 1$;

G (rxt) é uma matriz de especificação, de posto completo $r \leq t$, que associa os níveis do fator que define as ocasiões de avaliação (tempo) ao polinômio desejado, e

$\boldsymbol{\varepsilon}$ (nxt) é a matriz de componentes aleatórios. Supõe-se que suas linhas, $\boldsymbol{\varepsilon}_{ij}$, sejam não correlacionadas e que $\boldsymbol{\varepsilon}_{ij} \sim N_p(\mathbf{0}; \boldsymbol{\Sigma})$, onde $\boldsymbol{\Sigma}$ (txt) é uma matriz de covariâncias não estruturada comum a todas as linhas.

POTTHOFF & ROY (1964); KHATRI (1966) e GRIZZLE & ALLEN (1969) desenvolveram técnicas de estimação e testes de hipóteses para os parâmetros das curvas de crescimento, considerando dados são *completos* e *balanceados* em relação ao tempo e sem impor qualquer estrutura à matriz de covariâncias das respostas. SINGER (1977) descreveu um procedimento para a análise de curvas de crescimento para dados balanceados utilizando um programa convencional de Análise de Variância Multivariada (MANOVA), que foi utilizado por LIMA (1980) na análise de dados de crescimento em peso de frangos de corte e de poedeiras [*ver Exemplo 2 no Apêndice*].

Algumas restrições ao uso do modelo de Potthoff & Roy:

- As curvas polinomiais ajustadas aos diversos grupos devem ser do mesmo grau, apesar de os seus coeficientes não precisarem ser iguais.
- Os dados precisam ser *completos.*
- Existem dificuldades de estimação e de eficiência das estimativas dos parâmetros das curvas (grande número de parâmetros associados à matriz $\boldsymbol{\Sigma}$).
- *A implementação da análise é trabalhosa*, pois não existe um pacote estatístico que faça essa análise diretamente (Alternativa: usar o PROC MIXED ??!!!)

Para maiores detalhes: POTTHOFF & ROY (1964), KHATRI (1966), GRIZZLE & ALLEN (1969) e SINGER (1977), dentre outros.

2.4. ANÁLISE DE DADOS UTILIZANDO MODELOS DE EFEITOS ALEATÓRIOS

LAIRD & WARE (1982), FAIRCLOUGH & HELMS (1986), LAIRD *et al*. (1987) e LINDSTROM & BATES (1988) dentre outros, estudaram uma classe de estruturas de covariâncias induzidas através da especificação de ***Modelos (Mistos) de Efeitos Aleatórios***, que proporcionam uma maior versatilidade na aplicação da técnica de *análise de dados com medidas repetidas* (*análise de perfis e de curvas de crescimento*), pois:

i) permitem a modelagem da matriz de covariâncias com um número menor de parâmetros que a dos modelos multivariados gerais;

ii) podem ser usados quando os dados não são balanceados em relação ao tempo;

iii) possibilitam o ajuste de curvas polinomiais de graus diferentes para cada grupo de unidades experimentais.

O *Modelo de Efeitos Aleatórios* foi proposto por LAIRD & WARE (1982) e baseia-se no trabalho de HARVILLE (1977). Para um indivíduo (ij), o modelo pode ser escrito como:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij} \tag{4}$$

para i = 1, 2, ..., g e j = 1, 2, ... n_i onde

$\mathbf{y}_{ij}$ (tx1) é o perfil de respostas do indivíduo (ij);

$\mathbf{X}_{ij}$ (txr) é uma matriz de posto $r < t$, conhecida e de especificação, associada ao vetor $\boldsymbol{\beta}_i$ (rx1) de parâmetros sub-populacionais desconhecidos;

$\mathbf{Z}_{ij}$ (txq) é uma matriz conhecida e de especificação, de posto coluna completo, associada ao vetor de efeitos aleatórios $\mathbf{b}_{ij}$ (qx1) de diferenças individuais em torno dos valores populacionais;

$\boldsymbol{\varepsilon}_{ij}$ (tx1) é um vetor de erros aleatórios.

Supõe-se que $\mathbf{b}_{ij}$ e $\boldsymbol{\varepsilon}_{ij}$ são independentes, que $\boldsymbol{\varepsilon}_{ij} \sim N_t(\mathbf{0}, \mathbf{R}_{ij})$ e que $\mathbf{b}_{ij} \sim N_q(\mathbf{0}, \mathbf{G})$, para i = 1, 2, ..., g e j = 1, 2, ..., b, onde $\mathbf{R}_{ij}$(txt) e $\mathbf{G}$(qxq) são matrizes de covariâncias associadas às t *ocasiões de avaliação* e aos q efeitos aleatórios, respectivamente.

As matrizes de especificação $\mathbf{X}_{ij}$ e $\mathbf{Z}_{ij}$ podem ser diferentes e variar entre unidades experimentais, estendendo o modelo para o caso de dados não balanceados em relação ao tempo. As matrizes $\mathbf{Z}_{ij}$ podem conter quaisquer covariáveis que afetem diferentemente as unidades experimentais.

A forma de especificação das matrizes $\mathbf{X}_{ij}$ é similar àquela utilizada nos modelos de regressão. Suas colunas podem estar associadas:

i) aos fatores que definem a estrutura das subpopulações (tratamentos);
ii) ao fator tempo, identificando, por exemplo, a forma da curva a ser ajustada e
iii) a covariáveis, cujos efeitos na resposta média deseja-se pesquisar.

O modelo **(4)** pode ser formulado em dois estágios, evidenciando a identificação das características individuais e populacionais. No primeiro estágio, para cada unidade experimental (ij), tem-se:

$$\mathbf{y}_{ij} \mid \mathbf{b}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\varepsilon}_{ij} \sim \mathbf{N}(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{b}_{ij}; \mathbf{R}_{ij}) \tag{5}$$

onde $\mathbf{R}_{ij}$ é conhecida como *matriz de dispersão condicional* e está associada ao *erro condicional* $\boldsymbol{\varepsilon}_{ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta} - \mathbf{Z}_{ij}\mathbf{b}_{ij}$. Diversas estruturas de dispersão do erro condicional podem ser consideradas, como a completamente parametrizada e as associadas a séries temporais. Quando $\mathbf{R}_{ij} = \sigma^2\mathbf{I}_{(t)}$, o modelo é conhecido como *modelo com independência condicional* e reflete a independência e a homocedasticidade das observações intra-indivíduos.

No segundo estágio, assume-se que $\mathbf{b}_{ij} \sim N(\mathbf{0}; \mathbf{G})$ é independente de $\boldsymbol{\varepsilon}_{ij}$ obtendo-se o *modelo marginal* (ou *não condicional*)

$$\mathbf{y}_{ij} \sim N(\mathbf{X}_{ij}\boldsymbol{\beta};\ \mathbf{Z}_{ij}\mathbf{G}\,\mathbf{Z}_{ij}^{t} + \mathbf{R}_{ij}) \qquad \textbf{(6)}$$

onde a matriz $\mathbf{V}_{ij} = \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}^{t} + \mathbf{R}_{ij}$ é chamada *matriz de dispersão marginal* e está associada ao *erro marginal* $\mathbf{e}_{ij} = \mathbf{y}_{ij} - \mathbf{X}_{ij}\boldsymbol{\beta}$. Quando $\mathbf{Z}_{ij} = \mathbf{1}_t$, $\mathbf{G}=\sigma_0^2$ e $\mathbf{R}_{ij} = \sigma^2\mathbf{I}_{(t)}$, que resulta em $\mathbf{V}_{ij} = \sigma_0^2\mathbf{1}_t\mathbf{1}_t^t + \sigma^2\mathbf{I}_{(t)}$, o modelo é chamado de *modelo de simetria composta*.

Como no modelo linear usual, $E(\mathbf{y}_{ij}) = \mathbf{X}_{ij}\boldsymbol{\beta}_i$ é modelada pelos efeitos fixos, $\boldsymbol{\beta}_i$, e a extensão proporcionada pelos modelos de efeitos aleatórios é que

$$Var(\mathbf{y}_{ij}) = \mathbf{V}_{ij} = \mathbf{V}_{ij}(\boldsymbol{\theta}) = \mathbf{Z}_{ij}\mathbf{G}\mathbf{Z}_{ij}^{t} + \mathbf{R}_{ij} \qquad \textbf{(7)}$$

onde $\boldsymbol{\theta}$ é um vetor (k x 1) com os parâmetros de covariâncias, desconhecidos. A matriz $\mathbf{V}_{ij}$ pode ser modelada com um número menor de parâmetros que t(t+1)/2, impondo-se diferentes estruturas para as matrizes $\mathbf{Z}_{ij}$, $\mathbf{G}$ e $\mathbf{R}_{ij}$. Este método de estruturar a matriz de covariâncias $\mathbf{V}_{ij}$ tem como atrativos a possibilidade de:

(i) englobar as abordagens uni e multivariada, que são comumente utilizadas na análise de dados longitudinais, pois a matriz de covariâncias com a estrutura $\mathbf{V}_{ij} = (\sigma_\varepsilon^2)\mathbf{I}_t + (\sigma_\gamma^2)\mathbf{1}_t\mathbf{1}_t^t$ corresponde à abordagem univariada, enquanto que a matriz $\mathbf{V}_{ij}$ com a estrutura completamente parametrizada, corresponde à abordagem multivariada;

(ii) lidar com dados perdidos, por causa da facilidade de construir a verossimilhança somente dos dados observados, e

(iii) usar estruturas relacionadas com séries temporais ou estruturas mais complexas, em adição às estruturas uniforme e não estruturada.

A estimação dos parâmetros de um modelo de efeitos aleatórios é baseada na *verossimilhança* dos dados e quando os dados não são normalmente distribuídos, diversos trabalhos envolvendo *modelos lineares generalizados* podem ser consultados, como os de STIRATELLI et al. (1984); ZEGER et al. (1985); LIANG & ZEGER (1986); ZEGER (1988) e JØRGENSEN et al. (1991). Os interessados em mais referências sobre esses modelos devem consultar WARE (1985), JØRGENSEN et al. (1985) e ZEGER & LIANG (1986).

Para a totalidade das observações, o modelo (12) pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \qquad (6)$$

onde: $\mathbf{y}$ (Nt x 1), $\boldsymbol{\beta}$ (gt x 1), $\mathbf{b}$ (Nq x 1) e $\boldsymbol{\varepsilon}$ (Nt x 1) são construídos empilhando-se os vetores $\mathbf{y}_{ij}$, $\boldsymbol{\beta}_i$, $\mathbf{b}_{ij}$ e $\boldsymbol{\varepsilon}_{ij}$, respectivamente; $\mathbf{X} = \text{diag}[\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_g]$ (Nt x gr), onde $\mathbf{X}_i$ é obtida empilhando-se as n_i matrizes $\mathbf{X}_{ij}$; $\mathbf{Z} = \text{diag}[\mathbf{Z}_{11}, \mathbf{Z}_{12}, \cdots, \mathbf{Z}_{gn_g}]$, de dimensão ($Nt$ x Nq).

Como **b** e $\boldsymbol{\varepsilon}$ têm distribuições normais com médias nulas e $Var\begin{bmatrix}\mathbf{b}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$, a variância de **y** é dada por:

$$\mathbf{V} = \mathbf{ZWZ'} + \mathbf{R} \tag{7}$$

onde $\mathbf{W} = \text{diag}[\mathbf{G}, \mathbf{G}, ..., \mathbf{G}]$ e $\mathbf{R} = \text{diag}[\mathbf{R}_{11}, \mathbf{R}_{12}, \cdots, \mathbf{R}_{gn_g}]$

Quando essas matrizes **W** e **R** são conhecidas, as estimativas para $\boldsymbol{\beta}$ e **b** podem ser escritas como $\hat{\boldsymbol{\beta}} = \left[\mathbf{X'V}^{-1}\mathbf{X}\right]^{-1}\mathbf{X'V}^{-1}\mathbf{y}$ e $\hat{\mathbf{b}} = \mathbf{WZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, respectivamente. Neste caso, $\hat{\boldsymbol{\beta}}$ é o <u>melhor estimador não viesado</u> (*BLUE*) de $\boldsymbol{\beta}$ e $\hat{\mathbf{b}}$ é o <u>melhor preditor não viesado</u> (*BLUP*) de **b**. A matrizes de covariâncias de $\hat{\boldsymbol{\beta}}$ é

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2[\mathbf{X'}(\mathbf{ZWZ'}+\mathbf{R})^{-1}\mathbf{X}]^{-1}$$

Entretanto, na maioria das vezes, as matrizes **W** e **R** são desconhecidas e uma abordagem através da teoria da *máxima verossimilhança* (MV) ou *máxima verossimilhança restrita* (MVR) pode ser usada para obter as estimativas dos parâmetros $\boldsymbol{\theta}$ (ver SUYAMA, 1995, dentre outros). Obtidas essas estimativas, os parâmetros $\boldsymbol{\beta}$ e **b** são estimados resolvendo-se o *sistema de equações do modelo misto* (HENDERSON, 1984) através de algoritmos iterativos:

$$\begin{bmatrix}\mathbf{X}^t\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}^t\hat{\mathbf{R}}^{-1}\mathbf{Z}\\ \mathbf{Z}^t\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}^t\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{W}}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{b}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^t\hat{\mathbf{R}}^{-1}\mathbf{y}\\ \mathbf{Z}^t\hat{\mathbf{R}}^{-1}\mathbf{y}\end{bmatrix} \tag{8}$$

ALGUMAS ESTRUTURAS DE COVARIÂNCIAS INTERESSANTES

Como já foi mencionado anteriormente, a matriz de dispersão marginal $\mathbf{V}_{ij} = \mathbf{V}_{ij}(\boldsymbol{\theta})$, com $\boldsymbol{\theta}$ (k x 1), pode ser *modelada* através das matrizes $\mathbf{Z}_{ij}$, **G** e $\mathbf{R}_{ij}$. Assumindo que o número de parâmetros de covariâncias desconhecidos é $k = k_R + k_G$, onde k_R e k_G correspondem aos números de parâmetros de covariâncias associados às matrizes $\mathbf{R}_{ij}$ e **G**, respectivamente.

Apesar de ser usual supor que $\mathbf{R}_{ij} = \mathbf{R}$ (mesma estrutura para todos os indivíduos) seja uma matriz de dispersão não estruturada com $k_R = t(t+1)/2$ parâmetros, nada impede que a estrutura de dependência entre as observações repetidas na mesma unidade experimental seja adequadamente modelada através de um número mais restrito de parâmetros. Dentre as possíveis estruturas para **R**, podemos listar:

(a) $\mathbf{R} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1t} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2t} \\ \vdots & \vdots & & \vdots \\ \sigma_{t1} & \sigma_{t2} & \cdots & \sigma_t^2 \end{bmatrix}$, com $k_R = t(t+1)/2$ parâmetros.

É a estrutura chamada *completamente parametrizada* ou *não estruturada*, sendo usada na análise multivariado de perfis.

(b) $\mathbf{R} = \sigma^2 \mathbf{I}_{(t)} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$, com $k_R = 1$ parâmetro.

É própria para modelar a situação em que a discrepância entre uma observação e o seu valor esperado é devida apenas a um erro de medida, que é independente das medidas feitas em outras observações, ou seja, as medidas feitas ao longo do tempo não são correlacionadas. É a estrutura usada no modelo linear tradicional, que pressupõe homocedasticidade das variâncias.

(c) $\mathbf{R} = \sigma_\gamma^2 1_t 1_t^t + \sigma_\varepsilon^2 \mathbf{I}_{(t)} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\varepsilon^2 + \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \vdots & & \vdots \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\gamma^2 \end{bmatrix}$, com $k_R = 2$ parâmetros.

Supõe que a variância das respostas em qualquer tempo é igual a $\sigma_\gamma^2 + \sigma_\varepsilon^2$ e que a covariância entre dois tempos quaisquer é constante e igual a σ_γ^2. É a estrutura comum de ensaios em parcelas subdivididas.

(d) $\mathbf{R} = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_t^2\right) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_t^2 \end{bmatrix}$, com $k_R = t$ parâmetros

É uma estrutura mais geral que a estrutura (a), quando supõe que a variância do erro de medida pode ser diferente para diferentes ocasiões.

(e) $\mathbf{R} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho^{t-1} \\ \rho & 1 & \cdots & \rho^{t-2} \\ \vdots & \vdots & & \vdots \\ \rho^{t-1} & \rho^{t-2} & \cdots & 1 \end{bmatrix}$, com $0 \le \rho \le 1$ e $k_R = 2$ parâmetros.

Reflete uma estrutura auto-regressiva de primeira ordem – AR(1).

Apesar de ser comum assumir que a matriz **G** (associada aos efeitos aleatórios) seja uma matriz de dispersão não estruturada com $k_G = q(q+1)/2$ parâmetros, outras estruturas de dependência entre os efeitos aleatórios podem ser utilizadas para modelá-la com um número menor de parâmetros.

TESTES DE HIPÓTESES

Os testes estatísticos usados na seleção de modelos incluem o ***teste de Wald***, que serve para avaliar a significância dos efeitos fixos e os ***testes da razão de verossimilhanças***, que são usados para selecionar tanto os efeitos fixos quanto os aleatórios em modelos *encaixados* (modelo completo vs. modelo reduzido) ou *hierárquicos*.

Observe-se que os resultados desses testes devem ser encarados sob um *ponto de vista exploratório* e não sob uma ótica inferencial, pois as estatísticas relacionadas aos testes, têm propriedades assintóticas (ANDRADE & SINGER, 1986).

A *estatística de Wald* para testar H_0: $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, onde **C** ($c$ x gr) é uma matriz de constantes conhecidas e de posto c ($c \leq gr$) é escrita como

$$\mathbf{Q_c} = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}\hat{\mathbf{V}}\mathbf{ar}(\hat{\boldsymbol{\beta}})\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}) \tag{9}$$

onde $\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}})$ é uma estimativa da matriz de covariâncias de $\hat{\boldsymbol{\beta}}$. Sob $\mathbf{H}_0$ a estatística $\mathbf{Q_c}$ tem distribuição assintótica quiquadrado com c graus de liberdade. Dividindo-se $\mathbf{Q_c}$ por c, obtém-se uma outra estatística que tem distribuição **F** com c e Nt–posto(**X**) graus de liberdade.

Quando os *modelos são encaixados* e a estimação dos parâmetros é feita através do método MV (máxima verossimilhança), nós usamos o ***teste da razão de verossimilhança*** para testar a hipótese nula, *de que o modelo mais simples é o adequado*. Essa estatística é escrita como:

$$\mathbf{Q_1} = \mathbf{2}\left[l(\hat{\boldsymbol{\beta}}^G, \hat{\boldsymbol{\theta}}^G) - l(\hat{\boldsymbol{\beta}}^S, \hat{\boldsymbol{\theta}}^S)\right] \tag{10}$$

onde $\hat{\boldsymbol{\beta}}^G$ e $\hat{\boldsymbol{\theta}}^G$ são os valores de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ que maximizam a verossimilhança sob o modelo mais geral (G) e $\hat{\boldsymbol{\beta}}^S$ e $\hat{\boldsymbol{\theta}}^S$, sob o modelo mais simples (S).

Sob H_0, $\mathbf{Q_1}$ tem distribuição assintótica quiquadrado com ($p^G + k^G - p^S - k^S$) graus de liberdade, onde (p^G e k^G) e (p^S e k^S) correspondem ao número de parâmetros de $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ sob os modelos mais geral e mais simples, respectivamente.

Quando a estimação dos parâmetros utilizar o método da MVR (*máxima verossimilhança restrita*), somente serão consideradas hipóteses relativas à adequabilidade de uma estrutura de covariâncias (SCHLUCHTER, 1992) se os modelos a serem comparados tiverem a mesma estrutura de médias. Neste caso, a estatística da razão de verossimilhança é dada por:

$$\mathbf{Q_2} = 2[\boldsymbol{l}_{\mathrm{R}}(\hat{\theta}_R^G) - \boldsymbol{l}_{\mathrm{R}}(\hat{\theta}_R^S)] \tag{11}$$

onde $\hat{\theta}_R^G$ e $\hat{\theta}_R^S$ são os valores de θ que maximizam a verossimilhança restrita sob os modelos mais geral e mais simples, respectivamente.

Sob a hipótese de que o modelo com a estrutura de covariâncias mais simples é adequado, $\mathbf{Q_2}$ tem distribuição assintótica quiquadrado com $(k^G - k^S)$ graus de liberdade, onde k^G e k^S representam o número de elementos estimados de θ sob os modelos mais geral e mais simples, respectivamente.

Quando os modelos ajustados *não são encaixados*, a identificação do modelo mais adequado aos dados pode ser feita utilizando alguns critérios, dentre os quais citamos:

- ***Critério de Informação de Akaike*** **(AIC)**, que é definido como

$$\mathbf{AIC} = \boldsymbol{l}(\hat{\theta}) - \boldsymbol{d} \tag{12}$$

  onde $\boldsymbol{l}$ é o valor do logaritmo da verossimilhança (restrita ou não) e $\boldsymbol{d}$ é o número de parâmetros de covariâncias estimados. Por este critério, o *melhor modelo* será aquele que apresentar o *maior AIC*.

- ***Critério Bayesiano de Schwarz*** **(BIC)**, que é definido como

$$\mathbf{BIC} = l(\hat{\theta}) - \boldsymbol{d}\log(n^*) \tag{13}$$

  onde $n^* = Nt$ para MV e $n^* = Nt - k$. Também por este critério, o *melhor modelo* será aquele que apresentar o *maior SBC*.

Vale notar que o critério BIC penaliza os modelos com maior número de parâmetros e que nem sempre esses critérios (AIC e BIC) concordam em indicar o modelo melhor. Alguns autores indicaram, após extensivos estudos de simulação, que a performance do AIC é melhor que do BIC na tentativa de identificar o melhor modelo.

Importante: A partir da versão 8 do PROC MIXED, as fórmulas de cálculo dessas estatísticas foram alteradas, (ver Tabela 41.2, retirado do manual) de tal forma que os *menores valores* passaram indicar os *melhores modelos* (*smaller is better*).

Table 41.2. Information Criteria

Criteria	Larger-is-better	Smaller-is-better	Reference
AIC	$\ell - d$	$-2\ell + 2d$	Akaike (1974)
HQIC	$\ell - d \log\log n$	$-2\ell + 2d \log\log n$	Hannan and Quinn (1979)
BIC	$\ell - d/2 \log n$	$-2\ell + d \log n$	Schwarz (1978)
CAIC	$\ell - d(\log n + 1)/2$	$-2\ell + d(\log n + 1)$	Bozdogan (1987)

O PROC MIXED DO SAS

O PROC MIXED ajusta uma variedade de modelos lineares mistos a dados, permitindo usar tais modelos ajustados para fazer inferências estatísticas sobre os dados. É uma poderosa ferramenta de análise de dados com medidas repetidas e disponibiliza um grande número de estruturas de covariâncias, que podem ser utilizadas nas especificações das matrizes **G** e **R**.

A especificação de um modelo misto no PROC MIXED é feita utilizando-se os seguintes comandos básicos:

```
proc mixed <opções>;
 class variáveis;
 model dependente = <efeitos fixos>  </opções>;
 random efeitos aleatórios em G </opções>;
 repeated <efeitos aleatórios em R > </opções>;
 contrast 'nome' <valores para o efeito fixo ...>
          <|valores para o efeito aleatório>, ... </opções>;
 estimate 'nome' < valores para o efeito fixo ...>
          <| valores para o efeito aleatório>, ... </opções>;
 lsmeans efeitos fixos </opções>;
```

onde os itens entre os sinais **< >** são opcionais.

O comando **proc mixed** invoca o procedimento de análise. Suas opções definem o critério de convergência e o número de iterações, o nome do arquivo de dados utilizado, o método de estimação dos parâmetros de covariância e a impressão da matriz de covariâncias assintóticas dos parâmetros de covariância, das equações do modelo misto e da solução dessas equações.

O comando **`class`** declara as variáveis qualitativas (classificatórias) que criarão variáveis indicadoras na matriz de delineamento e o comando **`model`** especifica a variável dependente e os efeitos fixos que definem a matriz $\mathbf{X}_{ij}$ do modelo misto. Basicamente, as opções do comando **`model`** definem a exclusão do intercepto do modelo (no caso de se considerarem interceptos diferentes para os níveis de um fator e não de desvios em relação a um intercepto geral); o método de cálculo do número de graus de liberdade do denominador e a impressão de uma tabela com os valores observados, ajustados e resíduos, de uma solução para os efeitos fixos e da estatística de Wald etc..

O comando `random` serve para definir a porção $\mathbf{Z}_{ij}\mathbf{b}_{ij}$ do modelo misto (4) e permite modelar a variação *entre* as unidades experimentais (matriz G), tais como as parcelas de um experimento em parcelas subdivididas. Já o comando `repeated` é usado para modelar a variação *dentro* das unidades experimentais. Quando analisamos conjuntos de dados com medidas repetidas, usamos o comando `repeated` para modelar a estrutura de covariâncias entre as ocasiões (matriz R), que são medidas feitas repetidamente em cada indivíduo. Os efeitos aleatórios podem ser variáveis classificatórias ou contínuas, enquanto que os efeitos repetidos devem ser variáveis classificatórias (listadas no comando `class`).

Uma das opções mais importantes dos comandos `random` e `repeated` é a opção `type` = **`estrutura`** que especifica a estrutura das matrizes **G** e **R**, respectivamente. Algumas opções de estruturas já foram mencionadas anteriormente. Duas outras opções importantes do comando `repeated` são `subject` **= `efeito`** e `group` **= `efeito`**. O efeito **`subject`** indica quem é o sujeito sobre o qual são feitas as medidas repetidas, ou seja, define o mecanismo para a bloco-diagonalização das matrizes **W** e **R** e assume independência completa entre os sujeitos.

A opção `group` **= efeito** define a heterogeneidade nas estruturas de covariâncias **G** e **R** de tal modo que todas as observações que têm o mesmo nível deste efeito têm os mêsmos parâmetros de covariâncias e, cada novo nível produz um novo conjunto de parâmetros de covariância, com a mesma estrutura do grupo original.

Outras opções definem a impressão de estimativas das matrizes **G** e **R** e uma solução para os efeitos aleatórios. Na Tabela 41.3 do manual do PROC MIXED estão indicadas algumas estruturas de covariância disponíveis para as matrizes **R** e **G**.

Os comandos `contrast` e `estimate` são usados para calcular estatísticas correspondentes a contrastes de médias para vários espaços de inferência e podem aparecer múltiplas vezes num programa. A opção `chisq` do comando `contrast` indica que a estatística de Wald também será impressa.

O comando `lsmeans` define a impressão das médias de mínimos quadrados generalizados dos efeitos fixos, bem como os seus erros padrões, que *são ajustados para os efeitos aleatórios presentes no modelo*. Uma de suas opções mais interessantes é a opção `slice` **= efeitos**, que é usada para desdobrar os efeitos da interação entre dois ou mais fatores, como no PROC GLM. A opção `pdiff` provoca a impressão das diferenças entre as médias. A opção `adjust` = **`teste`** serve para indicar a estatística que será usada no ajuste do nível descritivo nas comparações múltiplas das médias (Bonferroni, Dunnett, Scheffe, Sidak, Tukey etc.).

Table 41.3. Covariance Structures

Structure	Description	Parms	(i,j)th element
ANTE(1)	Ante-Dependence	$2t-1$	$\sigma_i\sigma_j\prod_{k=i}^{j-1}\rho_k$
AR(1)	Autoregressive(1)	2	$\sigma^2\rho^{\lvert i-j\rvert}$
ARH(1)	Heterogeneous AR(1)	$t+1$	$\sigma_i\sigma_j\rho^{\lvert i-j\rvert}$
ARMA(1,1)	ARMA(1,1)	3	$\sigma^2[\gamma\rho^{\lvert i-j\rvert-1}1(i\neq j)+1(i=j)]$
CS	Compound Symmetry	2	$\sigma_1^2+\sigma^2 1(i=j)$
CSH	Heterogeneous CS	$t+1$	$\sigma_i\sigma_j[\rho 1(i\neq j)+1(i=j)]$
FA(q)	Factor Analytic	$\frac{q}{2}(2t-q+1)+t$	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}+\sigma_i^2 1(i=j)$
FA0(q)	No Diagonal FA	$\frac{q}{2}(2t-q+1)$	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}$
FA1(q)	Equal Diagonal FA	$\frac{q}{2}(2t-q+1)+1$	$\Sigma_{k=1}^{\min(i,j,q)}\lambda_{ik}\lambda_{jk}+\sigma^2 1(i=j)$
HF	Huynh-Feldt	$t+1$	$(\sigma_i^2+\sigma_j^2)/2+\lambda 1(i\neq j)$
LIN(q)	General Linear	q	$\Sigma_{k=1}^{q}\theta_k\mathbf{A}_{ij}$
TOEP	Toeplitz	t	$\sigma_{\lvert i-j\rvert+1}$
TOEP(q)	Banded Toeplitz	q	$\sigma_{\lvert i-j\rvert+1}1(\lvert i-j\rvert<q)$
TOEPH	Heterogeneous TOEP	$2t-1$	$\sigma_i\sigma_j\rho_{\lvert i-j\rvert}$
TOEPH(q)	Banded Hetero TOEP	$t+q-1$	$\sigma_i\sigma_j\rho_{\lvert i-j\rvert}1(\lvert i-j\rvert<q)$
UN	Unstructured	$t(t+1)/2$	σ_{ij}
UN(q)	Banded	$\frac{q}{2}(2t-q+1)$	$\sigma_{ij}1(\lvert i-j\rvert<q)$
UNR	Unstructured Corrs	$t(t+1)/2$	$\sigma_i\sigma_j\rho_{\max(i,j)\min(i,j)}$
UNR(q)	Banded Correlations	$\frac{q}{2}(2t-q+1)$	$\sigma_i\sigma_j\rho_{\max(i,j)\min(i,j)}$
UN@AR(1)	Direct Product AR(1)	$t_1(t_1+1)/2+1$	$\sigma_{i_1j_1}\rho^{\lvert i_2-j_2\rvert}$
UN@CS	Direct Product CS	$t_1(t_1+1)/2+1$	$\sigma_{i_1j_1}(1-\sigma^2 1(i_2\neq j_2)), 0\leq\sigma^2\leq 1$
UN@UN	Direct Product UN	$t_1(t_1+1)/2+$ $t_2(t_2+1)/2-1$	$\sigma_{1,i_1j_1}\sigma_{2,i_2j_2}$
VC	Variance Components	q	$\sigma_k^2 1(i=j)$ and i corresponds to kth effect

EXEMPLO: Análise univariada (parcelas subdivididas) dos dados do Exemplo 1 utilizando o *proc mixed*, com o comando **random**.

```
proc mixed data=MSTotal_uni;
 class Bloco Capim Tempo;
 model MSTotal = Bloco Capim Tempo Capim*Tempo / ddfm=satterth;
 random bloco*Capim;
run;
```

Quadro 4. Análise univariada (parcelas subdivididas) dos dados de MS (t/ha), utilizando o PROC MIXED com o comando **random**

```
                                                                    (a)
The Mixed Procedure

           Model Information
Data Set                     WORK.MSTOTAL_UNI
Dependent Variable           MSTotal
Covariance Structure         Variance Components
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Satterthwaite
```

```
Class Level Information
Class    Levels    Values
Bloco         4    1 2 3 4
Capim         3    Coastross Florona Tifton_85
Tempo         4    Ano1_i Ano1_v Ano2_i Ano2_v

Dimensions
Covariance Parameters             2
Columns in X                     24
Columns in Z                     12
Subjects                          1
Max Obs Per Subject              48
Observations Used                48
Observations Not Used             0
Total Observations               48

                     Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1        98.77595730
        1              1        98.10657995      0.00000000

                   Convergence criteria met.


                                                                    (b)
Covariance Parameter Estimates

Cov Parm        Estimate
Bloco*Capim      0.08448
Residual          0.5256
                                                                    (c)
Fit Statistics

-2 Res Log Likelihood            98.1
AIC (smaller is better)         102.1
AICC (smaller is better)        102.5
BIC (smaller is better)         103.1


                                                                    (d)
Type 3 Tests of Fixed Effects

                Num     Den
Effect           DF      DF    F Value    Pr > F
Bloco             3       6       7.84    0.0169
Capim             2       6       5.53    0.0435
Tempo             3      27     525.41    <.0001
Capim*Tempo       6      27      11.00    <.0001
```

(a) Traz informações sobre o modelo (estrutura de covariâncias, método de estimação, método de cálculo dos graus de liberdade), sobre as variáveis classificatórias (nome de fatores, número e nome dos níveis), sobre a dimensão do modelo (número de parâmetros fixos e aleatórios, número de observações usadas) e sobre o processo iterativo.

(b) Traz as estimativas dos parâmetros de covariâncias

(c) Traz informações sobre estatísticas relacionadas ao ajuste do modelo, como os valores dos critérios AIC e BIC.

(d) Traz o quadro de ANOVA com o valor das estatísticas F para os efeitos fixos do modelo. Note que não aparecem os valores das SQ's e QM's dos efeitos, mas os resultados são idênticos àqueles obtidos no Quadro 1(c).

EXEMPLO: Uma forma alternativa de executarmos essa mesma análise consiste em trocar o comando **random** pelo comando **repeated**. A sintaxe dos comandos é a seguinte:

```
proc mixed data = MSTotal_uni;
 class Bloco Capim Tempo;
 model MSTotal = Bloco Capim Tempo Capim*Tempo / ddfm=satterth;
 repeated / subject = Bloco(Capim) type=CS;
 lsmeans Capim Tempo Capim*Tempo;
run;
```

Quadro 5. Análise univariada (parcelas subdivididas) dos dados de MS (t/ha), utilizando o PROC MIXED com o comando **repeated.**

(a)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Bloco	3	6	7.84	0.0169
Capim	2	6	5.53	0.0435
Tempo	3	27	525.41	<.0001
Capim*Tempo	6	27	11.00	<.0001

(b)

Least Squares Means

Effect	Tempo	Capim	Estimate	Standard Error	DF	t Value	Pr > |t|
Capim		Coastross	9.5294	0.2323	6	41.02	<.0001
Capim		Florona	9.3975	0.2323	6	40.45	<.0001
Capim		Tifton_85	10.4025	0.2323	6	44.78	<.0001
Tempo	Ano1_i		6.6392	0.2255	27.5	29.44	<.0001
Tempo	Ano1_v		14.5033	0.2255	27.5	64.32	<.0001
Tempo	Ano2_i		4.7633	0.2255	27.5	21.13	<.0001
Tempo	Ano2_v		13.2000	0.2255	27.5	58.54	<.0001
Tempo*Capim	Ano1_i	Coastross	6.7025	0.3905	27.5	17.16	<.0001
Tempo*Capim	Ano1_i	Florona	6.8075	0.3905	27.5	17.43	<.0001
Tempo*Capim	Ano1_i	Tifton_85	6.4075	0.3905	27.5	16.41	<.0001
Tempo*Capim	Ano1_v	Coastross	13.6650	0.3905	27.5	34.99	<.0001
Tempo*Capim	Ano1_v	Florona	13.0450	0.3905	27.5	33.40	<.0001
Tempo*Capim	Ano1_v	Tifton_85	16.8000	0.3905	27.5	43.02	<.0001
Tempo*Capim	Ano2_i	Coastross	5.4950	0.3905	27.5	14.07	<.0001
Tempo*Capim	Ano2_i	Florona	4.6225	0.3905	27.5	11.84	<.0001
Tempo*Capim	Ano2_i	Tifton_85	4.1725	0.3905	27.5	10.68	<.0001
Tempo*Capim	Ano2_v	Coastross	12.2550	0.3905	27.5	31.38	<.0001
Tempo*Capim	Ano2_v	Florona	13.1150	0.3905	27.5	33.58	<.0001
Tempo*Capim	Ano2_v	Tifton_85	14.2300	0.3905	27.5	36.44	<.0001

No Quadro 5 você pode perceber em (a), que os valores das estatísticas F e os seus respectivos valores-*p* são idênticos aos obtidos anteriormente com o PROC GLM no Quadro 1 e com o PROC MIXED no Quadro 3. Em (b) estão apresentadas as médias dos níveis de todos os fatores envolvidos na análise.

O desdobramento da interação para comparar o efeito dos capins em cada um dos tempos é feito através do comando:

```
lsmeans Capim*Tempo / slice=Tempo diff;
```

Resultando em:

(Continuação do Quadro 5)

(c)

Differences of Least Squares Means

Effect	Tempo	Capim	_Tempo	_Capim	Estimate	Standard Error	DF	t Value	Pr > |t|
Tempo*Capim	Ano1_i	Coastross	Ano1_i	Florona	-0.1050	0.5523	27.5	-0.19	0.8506
Tempo*Capim	Ano1_i	Coastross	Ano1_i	Tifton_85	0.2950	0.5523	27.5	0.53	0.5975
Tempo*Capim	Ano1_i	Florona	Ano1_i	Tifton_85	0.4000	0.5523	27.5	0.72	0.4750
Tempo*Capim	Ano1_v	Coastross	Ano1_v	Florona	0.6200	0.5523	27.5	1.12	0.2713
Tempo*Capim	Ano1_v	Coastross	Ano1_v	Tifton_85	-3.1350	0.5523	27.5	-5.68	<.0001
Tempo*Capim	Ano1_v	Florona	Ano1_v	Tifton_85	-3.7550	0.5523	27.5	-6.80	<.0001
Tempo*Capim	Ano2_i	Coastross	Ano2_i	Florona	0.8725	0.5523	27.5	1.58	0.1256
Tempo*Capim	Ano2_i	Coastross	Ano2_i	Tifton_85	1.3225	0.5523	27.5	2.39	0.0237
Tempo*Capim	Ano2_i	Florona	Ano2_i	Tifton_85	0.4500	0.5523	27.5	0.81	0.4222
Tempo*Capim	Ano2_v	Coastross	Ano2_v	Florona	-0.8600	0.5523	27.5	-1.56	0.1309
Tempo*Capim	Ano2_v	Coastross	Ano2_v	Tifton_85	-1.9750	0.5523	27.5	-3.58	0.0013
Tempo*Capim	Ano2_v	Florona	Ano2_v	Tifton_85	-1.1150	0.5523	27.5	-2.02	0.0533

Tests of Effect Slices

Effect	Tempo	Num DF	Den DF	F Value	Pr > F
Capim*Tempo	Ano1_i	2	27.5	0.28	0.7564
Capim*Tempo	Ano1_v	2	27.5	26.57	<.0001
Capim*Tempo	Ano2_i	2	27.5	2.96	0.0682
Capim*Tempo	Ano2_v	2	27.5	6.43	0.0051

(c) Podemos perceber pelos testes F e seus respectivos valores-*p*, que as médias dos capins podem ser consideradas diferentes ($p < 0,05$) somente no verão de ambos os anos. Nessas ocasiões, o capim Tifton 85 apresentou produção superior à dos outros dois capins. No inverno dos dois anos as produções foram semelhantes.

A análise multivariada dos perfis pode ser feita especificando a estrutura UN no comando **repeated**

```
repeated / subject = Bloco(Capim) type=UN;
```

COMPARANDO ESTRUTURAS DE COVARIÂNCIAS

Podemos identificar a melhor estrutura de covariâncias, ou seja, a estrutura que melhor explica as variâncias e covariâncias entre as medidas repetidas nas quatro ocasiões, através do teste da razão de verossimilhança (11) e dos critérios de Akaike (12) e o de Schwarz (13),

A Tabela 2 apresenta, para cada estrutura testada, os valores dos critérios AIC e BIC, o valor de *–2log(verossimilhança restrita)*, o número de parâmetros de cada estrutura de covariâncias e as estatísticas quiquadrado para o teste H_0: a estrutura mais simples é mais indicada que a estrutura UN.

Tabela 2. Informações sobre diversas estruturas de covariâncias (Exemplo1)

Estrutura	-2LogVeross	Parms	Qcalc	gl	valor-*p*	AIC	BIC
UN	83.1	10	-	-	-	103.1	107.9
UN(1)	94.7	4	11.6	6	0.0715	102.7	104.7
CSH	93.1	5	10.0	5	0.0753	103.1	105.5
CS	98.1	2	15.0	8	0.0592	102.1	103.1
VC	98.8	1	15.7	9	0.0734	100.8	101.3

Ao nível de significância $\alpha = 5\%$, a hipótese H_0 não foi rejeitada para nenhuma das estruturas testadas, UN(1), CSH, CS e VC, indicando que qualquer uma delas pode substituir a estrutura UN. Admitindo que a estrutura VC (variâncias iguais e covariâncias nulas) é bem simples e pode ser utilizada, a análise passa a ser feira como se o esquema entre os fatores Capim e Tempo fosse fatorial (!!!???)

Como exemplo: a comparação entre as estruturas CSH (variâncias diferentes e covariâncias iguais) e CS (variâncias iguais e covariâncias iguais) resulta em $Q_{calc} = 15.0 - 10.0 = 5.0$, com $5 - 2 = 3$ graus de liberdade. Neste caso, o valor-*p* é igual a 0,1718 e indica que a estrutura mais simples (CS) é a mais indicada (ou a mais adequada).

Utilizando os critérios AIC e BIC para escolher a estrutura mais adequada, nós procuramos a estrutura com os *menores* valores para ambos os critérios. A estrutura escolhida é (mais uma vez) a VC, que apresentou AIC = 100,8 e BIC = 101,3.

EFEITOS DA ESTRUTURA DE COVARIÂNCIAS NAS ESTIMATIVAS E TESTES DE HIPÓTESES

A escolha da estrutura de covariâncias altera a estimativa do erro padrão das médias dos tratamentos (efeitos fixos) nas diversas ocasiões e pode alterar até o valor dessas médias. Por conseguinte, as conclusões dos testes de hipóteses da ANOVA também podem ser alteradas.

Na Tabela 3 podemos perceber que não ocorreram divergências nos resultados dos testes para Tempo e a interação Capim*Tempo, mas apareceram algumas divergências nos resultados do teste para Capim quando, por exemplo, comparamos os níveis descritivos associados às estruturas VC (0.0013) e CSH (0.0749) que levariam a conclusões diferentes.

Tabela 3. Níveis descritivos (valores-*p*) dos testes para os efeitos fixos, admitindo diferentes estruturas de covariâncias

Causa de Variação	UN	UN(1)	CSH	CS	VC
Capim	0.0393	0.0019	0.0749	0.0435	0.0013
Tempo	<.0001	<.0001	<.0001	<.0001	<.0001
Capim*Tempo	<.0001	<.0001	<.0001	<.0001	<.0001

ANÁLISE EM PARCELAS SUB-SUB-DIVIDIDAS

EXEMPLO: Podemos usar o PROC MIXED para analisar os dados de MS (t/ha) usando um modelo em parcelas sub-sub-divididas, como apresentado no Quadro 3, com o PROC GLM. Para tanto usamos os seguintes comandos:

```
proc mixed data=MS_Total covtest;
 class Bloco Capim Ano Estacao;
 model MS_Total = Capim Ano Capim*Ano
                  Estacao Capim*Estacao Ano*Estacao Capim*Ano*Estacao;
 repeated / subject=Bloco(capim) type=cs;
 random Bloco Bloco*Capim Bloco*Capim*Ano;
run;
```

Quadro 6. Análise dos dados de MS (t/ha) baseado num modelo em parcelas sub-sub-divididas, utilizando o PROC MIXED

```
                                                                    (a)
Covariance Parameter Estimates
                                            Standard      Z
Cov Parm          Subject       Estimate      Error  Value   Pr > Z

Bloco                             0.4920     0.4623   1.06   0.1436
Bloco*Capim                     0.002598     0.1687   0.02   0.4939
Bloco*Capim*Ano                   0.3273     0.2325   1.41   0.0796
CS                Bloco(Capim)  -0.02723   0.009076  -3.00   0.0027
Residual                          0.3074     0.1025   3.00   0.0013
```

(b)

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Capim	2	6	5.53	0.0435
Ano	1	9	31.52	0.0003
Capim*Ano	2	9	2.13	0.1753
Estacao	1	18	2593.34	<.0001
Capim*Estacao	2	18	42.83	<.0001
Ano*Estacao	1	18	3.20	0.0905
Capim*Ano*Estacao	2	18	6.92	0.0059

(a) Apresenta as estimativas das variâncias dos efeitos aleatórios e daquelas associadas aos parâmetros da estrutura CS, além do resultado do teste que o efeito pode ser considerado nulo.

(b) No quadro de análise de variância com os fatores de efeitos considerados fixos, nós podemos observar que os resultados (Pr > F) são idênticos àqueles obtidos no Quadro 3(d).

3. TEMPO É UM FATOR DE EFEITO FIXO OU ALEATÓRIO?

Efeitos fixos: são os *"efeitos de um fator cujos níveis estão todos presentes no estudo"*; alguém com perspectivas menos ortodoxas pode dizer que são *"efeitos de um fator que é o objetivo do estudo"*.

Exemplo: Em um estudo onde as mudanças da resposta no tempo são importantes, então estação, data de colheita ou ano, podem ser considerados de efeitos fixos. É algo parecido com os *blocos* num *delineamento casualizado em blocos*, quando os blocos são diferentes tipos de solos e a resposta nos diferentes tipos de solo é de interesse do pesquisador. Nesses casos, bloco deve ser considerado de efeito fixo. Se as repetições no tempo servem para avaliarmos os efeitos ambientais (em diferentes anos, por exemplo), ou se estamos interessados em avaliar como a resposta é afetada pela duração do tempo no qual o tratamento foi aplicado, então nós podemos definir ano (tempo) como um fator de efeito fixo.

Efeitos aleatórios: são *"os efeitos de um fator cujos níveis no estudo constituem uma amostra dos possíveis níveis"* ou *"efeitos de um fator que é usado para obter repetição"*.

Exemplo: As repetições ocorrem no espaço (isto é, repetições de um experimento em um certo ano) e no tempo (repetindo um estudo em diferentes estações ou ano). Se a repetição no tempo é simplesmente para ganhar amplitude na inferência, então estação ou ano podem ser considerados como de efeitos aleatórios.

Resumindo: escolher se o tempo é um fator de efeito fixo ou aleatório é uma função dos objetivos do experimento e do motivo que levou o estudo a ser repetido nas diferentes estações ou anos.

Uma desvantagem de considerar o fator ano como de efeito aleatório é que não podemos determinar os efeitos de ano ou da interação tratamento x ano. O Dr Littell sugere que:

- SE ano for considerado de efeito aleatório, que iniciemos analisando os dados "por ano" e não "entre anos", como nós fazemos quando ano é considerado fixo (neste último caso, nós testamos a interação tratamento x ano).
- SE ano é de efeito aleatório e a resposta ao tratamento é consistente entre os anos, então podemos combinar os dados, analisando os dados "entre anos".
- SE ano é um fator de efeito fixo e o efeito dos tratamentos está próximo da significância (p = 0,12, por exemplo) em dois ou mais anos, quando fazemos a análise "entre anos" nós teremos um efeito significativo de tratamento. Outra desvantagem da análise "entre anos", com o fator ano considerado de efeito aleatório, é que em alguns casos, o teste F para tratamento tem um valor-p maior que nas análises "por ano", por causa das mudanças nos termos de erro (denominador) do teste F.

4. ALGUMAS REFERÊNCIAS IMPORTANTES:

ANDRADE, D.F.; SINGER J.M. **Análise de dados longitudinais.** IN: VII Simpósio Nacional de Probabilidade e Estatística - SINAPE, Campinas, SP. São Paulo: Associação Brasileira de Estatística, 1986, 106p.

ANDREONI, S. **Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo**. São Paulo. 1989. 142p. Dissertação (Mestrado) - IME-USP.

AUBIN, E.C.Q. **Análise de experimentos com medidas repetidas**. São Paulo, 1984. 164p. Dissertação (Mestrado) - IME-USP.

CROWDER, M.J.; HAND, J. **Analysis of repeated measures.** London: Chapman & Hall, 1990, 257p.

DIGGLE, P.J.; LIANG, K.Y.; ZEGER, S.L. **Analysis of Longitudinal Data**. Oxford: Clarendon, 1994, 253p.

GRIZZLE, J.E.; ALLEN, D.M. Analysis of growth and dose-response curves. **Biometrics**, v.25, p.357-81, 1969.

JENNRICH, R.I.; SCHLUCHTER, M.D. Unbalanced repeated measures models with structured covariance matrices. **Biometrics**, v.42, p.805-20, 1986.

KHATTREE, R. & NAIK, D.N. **Applied Multivariate Statistics with SAS software.** Cary, North Carolina: SAS Institute Inc., 1995.

KSHIRSAGAR, A.M.; SMITH, W.B. **Growth curves**. New York: Marcel Dekker, 1995, 359p.

LIMA, C.G. **Análise de Dados Longitudinais provenientes de Experimentos em Blocos Casualizados**. Piracicaba, 1996. 126p. Tese (doutorado). ESALQ/ USP.

LITTELL, R.C.; MILLIKEM, G.A.; STROUP, W.W. & WOLFINGER, R.D. **SAS System for Mixed Models**, Cary, NC: SAS Institute Inc., 1996, 633 pp.

MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of Messy Data**. v.1.: Designed Experiments. New York: Chapman & Hall, 1992.

MORRISON, D.F. **Multivariate Statistical Methods.** 3.ed. New York: McGraw-Hill, 1990, 415p.

POTTHOFF, R.F.; ROY, S.N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrika**, v.51, p.313-26, 1964.

RUTTER, C.M.; ELASHOFF, RM. Analysis of longitudinal data: random coefficient regression modelling. **Statistics in Medicine**, v.13, p.1211-31, 1994.

SAS Technical Report P-229. *SAS/STAT Software: Changes and Enhancements Release 6.07*. Chapter 16: The MIXED Procedure. Cary, North Carolina: SAS Institute Inc., 1992.

SCHLUCHTER, M.D. Unbalanced Repeated Measures Models with Structured Covariance Matrices. IN: **BMDP statistical software manual**, 1990 software release, v.2, eds. Los Angeles: BMDP Statistical Software Inc.,1990.

SINGER, J.M. **Análise de curvas de crescimento.** São Paulo, 1977. 113p. Dissertação (Mestrado). IME-USP.

STEEL, R.G.D.; TORRIE, J.H. **Principles and Procedures of Statistics - A Biometrical Approach**. 2 ed., Singapura: McGraw-Hill, Inc., 1980, 633p.

SUYAMA, E. **Identificação de um modelo de efeitos aleatórios**. São Paulo, 1995. 129p. Tese (Doutorado) - IME-USP.

WOLFINGER,R. Covariance structure selection in general mixed models. **Commun. in Statistics - Simulation**, v.22 n.4, p.1079-1106, 1993.

TIMM, N.H. Multivariate analysis of variance of repeated measures. **Handbook of Statistics, Analysis of Variance**, ed. P.R. Krishnaiah, v.1, p.41-87. New York, North Holland, 1980.

APÊNDICE 1: ESCOLHA DOS FATORES E DE SEUS RESPECTIVOS NÍVEIS[1]

Fator é a variável independente em estudo, por exemplo: tempo de estocagem; temperatura de estocagem; tipo de cultura lática para fermentação do vegetal; tipo de polímero de que é feito o filme da embalagem; tipo de agente antioxidante etc.

Níveis são os valores ou as categorias do fator, ou seja: 7, 14, 21 e 28 dias de estocagem; 5ºC, 10ºC e 15°C para armazenagem etc. As combinações dos níveis dos fatores em estudo constituem as *condições experimentais*, os chamados *tratamentos*.

Na experimentação, geralmente, o objetivo é observar de que maneira uma ou mais condições experimentais imposta ao material interfere no comportamento da variável resposta (variável dependente), importante dentro do contexto da pesquisa.

No planejamento de um experimento é fundamental conhecer os fatores que afetam as variáveis independentes e dependentes de interesse. Esses fatores podem ser agrupados em dois tipos:

1) *Fatores de tratamentos*: aqueles que o pesquisador tem interesse em verificar a sua influência sobre as variáveis respostas;
2) *Fatores que restringem a casualização*: aqueles que exigem a formação de grupos de unidades experimentais homogêneas para aplicação dos tratamentos; são fatores que permitem que as conclusões tiradas sejam livres de efeitos que se sabe são importantes e que podem ser controlados, mas cujo estudo não é objeto da pesquisa. São fatores de *blocagem* das unidades experimentais homogêneas entre elas.

   A necessidade de inclusão de fatores de restrição da casualização é uma das premissas básicas do planejamento e delineamento de experimentos. De fato a aplicação do(s) fator(es) de restrição na casualização refere-se ao princípio do controle local em experimentação, isto é, no caso do ambiente ou material experimental ser heterogêneo. Em experimentos sensoriais, os julgadores representam um fator de restrição na casualização, havendo também uma limitação no número de amostras avaliadas por sessão. Outro exemplo seria um experimento com leite de vaca de diferentes raças. Não há interesse em testar o efeito de raça, mas ele precisa ser separado, controlado no experimento. O pesquisador sabe que o leite tem composição diferente entre as raças, o que pode ser importante para o efeito da condição experimental (tratamento) em questão. Então raça será um fator de restrição da casualização dos tratamentos e, assim constituirá os blocos casualizados.

Os níveis dos fatores podem ser *qualitativos* (categorias) ou *quantitativos*. Os níveis qualitativos podem ser apenas nominais (as categorias) ou ainda, em alguns casos, ordinais.

---

[1] *In:* PLANEJAMENTO ESTATÍSTICO DE EXPERIMENTOS CIENTÍFICOS
2005 José Benício Paes Chaves, Ph.D. - Professor Titular – DTA/UFV
http://www.dta.ufv.br/artigos/planestat.htm

Os *níveis qualitativos nominais* são completamente independentes, não estruturados e não guardam nenhuma relação entre eles, por exemplo: espécies de bactérias em culturas láticas, tipos de antioxidante, marcas de produto no mercado, variedades ou cultivar de espécies vegetais, tipos de polímeros em material de embalagens etc.

Já os *níveis qualitativos ordinais* representam ordens e não são completamente independentes, por exemplo: baixa, média, alta e muito alta intensidade. O que é comum nestes casos é o aspecto de qualidade, não havendo intermediários aparentes entre os diversos níveis e sim alguma independência entre eles.

Para os *níveis de fatores quantitativos*, se a expressão for em intervalo, sendo os níveis apenas valores que representam esses intervalos, dependendo da metodologia de análise empregada, podem ser realizadas interpolações. Os níveis dos fatores quantitativos são estruturados e completamente dependentes entre eles.

MODELOS DE EXPERIMENTOS

Na experimentação, dependendo da forma com que os níveis dos fatores são selecionados pelo pesquisador, pode ser *fator de efeito fixo* (experimentos no modelo I) ou *fator de efeito aleatório* (experimentos no modelo II). Também pode ocorrer, no mesmo experimento, fatores de efeitos fixos e de efeitos aleatórios, os chamados *modelos mistos*.

Se o pesquisador escolhe, com base em algum critério, quais os níveis dos fatores a terem seus efeitos testados na pesquisa, então se têm um experimento no modelo I (*efeitos fixos*). Neste caso, o pesquisador estará interessado em testar os efeitos fixos dos níveis dos fatores e as conclusões são restritas aos tratamentos (níveis) testados, ou seja, as inferências se restringem aos tratamentos pesquisados.

Por outro lado, se os níveis (tratamentos) a serem testados na pesquisa forem selecionados de uma população de forma completamente aleatória, por meio de algum mecanismo de sorteio, sem interferência direta do pesquisador, então se tem um experimento no modelo II (efeitos aleatórios). Nesta situação, o pesquisador estará interessado nos efeitos aleatórios dos fatores, estimando seus componentes de variância, sendo as conclusões extensivas à população de onde os tratamentos foram sorteados, isto é, as inferências são válidas para toda a população de tratamentos.

Se ocorrerem fatores de efeitos fixos e fatores de efeitos aleatórios no mesmo experimento, tem-se um modelo misto.

É comum trabalhar com mais de um fator ao mesmo tempo no experimento. Assim, os fatores podem ser classificados de acordo com a relação existente entre eles:

- **Classificação cruzada:** quando os níveis dos fatores são cruzados, isto é, no conjunto do experimento, todos os níveis de um fator combinam com todos os níveis do outro fator, podendo ser estudados e testados, todos os efeitos das interações entre os fatores. Por exemplo, sejam dois fatores designados pelas letras A e B, cada um com três níveis (0, 1 e 2) a serem trabalhados no experimento. As condições experimentais ou tratamentos disponíveis na pesquisa serão nove, quais sejam: A0B0, A0B1, A0B2, A1B0,

A1B1, A1B2, A2B0, A2B1 e A2B2. Daí, a possibilidade de se estudar e testar o efeito da interação, pois ao se repetir os mesmos níveis de um fator em cada um dos níveis do outro fator, torna-se possível estudar o comportamento individualizado de um fator quando o outro muda de nível. Na classificação cruzada há uma relação de independência entre as combinações de níveis dos fatores. Isto leva aos chamados experimentos fatoriais.

- **Classificação hierárquica:** quando o nível i do fator A no nível j do fator B não for o mesmo nível i do fator A em outro nível do fator B; isto é, quando se muda o nível de um fator (fator de agrupamento principal) mudam-se também os níveis do outro fator (fator de sub-agrupamento). Nestes casos há uma relação de hierarquia entre os fatores, ou seja, há um fator principal e um outro fator que se encontra como que "aninhado" dentro deste fator principal.

  O fator aninhado ou de sub-agrupamento não pode ser considerado sem que se indique em qual nível do fator principal ele está aninhado. Neste caso, não é possível testar efeitos de interações entre os fatores, uma vez que ao se mudar o nível de um fator mudam-se também os níveis do outro fator, impossibilitando-se a avaliação deste tipo de efeito.

  O entendimento de interações e de aninhamento é importante para se compreender as inferências na análise de dados experimentais que envolvem mais de um fator.

- **Classificação mista:** ocorre quando no mesmo conjunto de um experimento aparecem fatores cruzados e hierárquicos. Por fim, enfatiza-se que, em um experimento, a escolha dos fatores e dos seus níveis é basicamente um problema do pesquisador. Ele é o profissional que conhece o "estado da arte" em sua área de conhecimento e, sabe o quanto (nível) de que (fator) deve ser pesquisado para se conhecer o seu efeito (do fator) sobre uma ou várias variáveis respostas (o que se mede).

APÊNDICE 2

Exemplo 2: Trinta e dois frangos de corte da linhagem Hubbard (13 fêmeas e 19 machos) foram alojados em dois boxes, separados por sexo e alimentados com a mesma ração comercial. As aves foram identificadas por um anel de alumínio numerado colocado em sua asa direita. Cada ave foi pesada semanalmente, durante um período de sete semanas, sendo as avaliações feitas sempre nos mesmos horários e dias da semana. Os pesos individuais das aves estão apresentados na Tabela 1, os pesos médios das aves e respectivos desvios padrões estão apresentados na Tabela 2 e os perfis individuais dos pesos, na Figura 1.

Tabela 1. Pesos corporais, em gramas, de frangos de corte por sexo, durante as sete primeiras semanas de idade

Fêmeas	Semana							Machos	Semana						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
1	122	291	500	712	1041	1430	1760	**1**	147	365	702	974	1293	1850	2220
2	129	314	551	830	1096	1485	1820	**2**	126	331	624	784	1128	1567	1900
3	133	308	563	857	1085	1422	1660	**3**	141	327	594	852	1029	1463	1820
4	135	348	584	854	1109	1493	1760	**4**	94	262	547	812	1090	1562	1850
5	110	286	556	782	1105	1538	1870	**5**	119	311	588	864	1184	1681	2100
6	130	302	518	740	1009	1337	1630	**6**	114	315	613	845	1180	1571	1950
7	133	336	630	831	1108	1514	1760	**7**	127	310	589	836	1248	1702	2120
8	138	337	618	937	1144	1570	1820	**8**	111	306	604	921	1297	1880	2270
9	153	352	637	830	1052	1464	1820	**9**	141	347	654	953	1365	1845	2180
10	138	332	484	767	1132	1548	1870	**10**	153	356	691	1014	1457	1897	2380
11	137	329	576	844	1127	1391	1660	**11**	136	351	664	906	1235	1735	2060
12	133	298	464	670	988	1387	1720	**12**	132	335	670	1000	1411	1831	2130
13	142	345	598	844	1172	1570	1860	**13**	125	289	577	830	1232	1700	2110
								14	123	323	611	864	1014	1449	1760
								15	126	320	596	872	1267	1735	2150
								16	137	334	610	904	1264	1624	1900
								17	152	353	654	964	1302	1744	2150
								18	139	349	635	948	1338	1773	2180
								19	118	277	591	870	1256	1738	2050

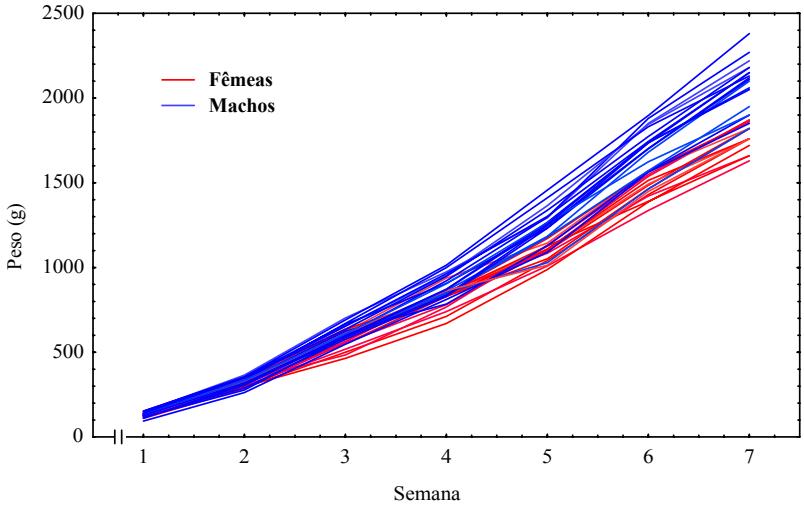


Figura 1. Perfis individuais de peso das aves

Tabela 2. Pesos médios corporais (em gramas) das aves e respectivos desvios padrões (d.p.), por sexo.

Semana	Fêmeas		Machos	
	Média	**d.p.**	**Média**	**d.p.**
1	133,31	2,80	129,53	3,42
2	321,38	6,29	324,26	6,35
3	559,92	15,39	621,79	9,43
4	807,54	19,69	895,42	15,06
5	1089,85	14,88	1241,58	27,06
6	1473,00	20,88	1702,47	30,67
7	1770,00	22,95	2067,37	37,61

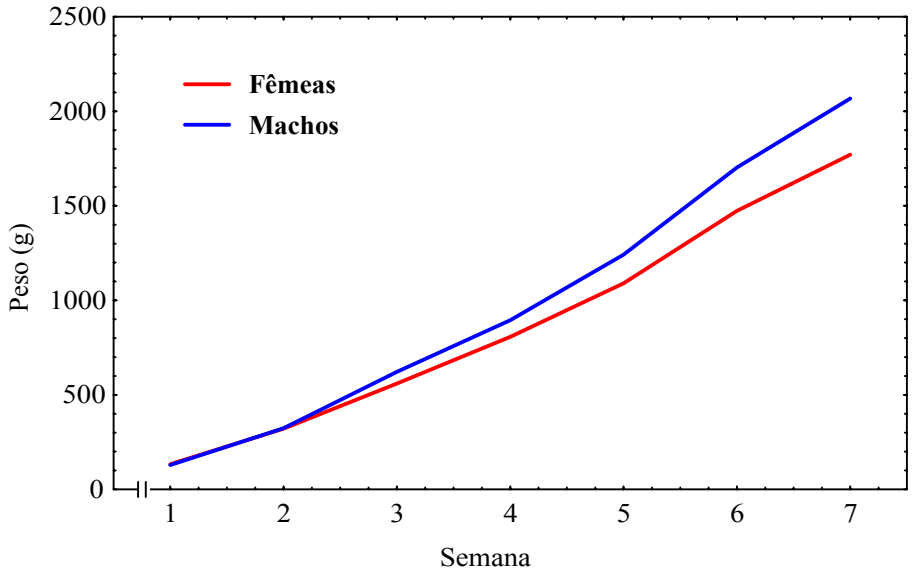


Figura 2. Perfis médios de peso das aves

Comentários:

- Na Tabela 2 e Figura 1 pode-se perceber um aumento na variabilidade dos pesos dos machos e das fêmeas com o aumento da idade das aves.
- Na Figura 2 pode-se perceber que a partir da segunda semana de vida, os machos são mais pesados que as fêmeas e a diferença de pesos aumenta com o aumento da idade das aves. O comportamento dos pesos médios ao longo do tempo pode ser bem explicado por uma reta ou um polinômio do segundo grau.

Os códigos para realização da análise dos dados utilizando o *proc glm* e o *proc mixed* são apresentados a seguir:

```
data frangoU (keep = indiv sexo idade y)
     frangoM (keep = sexo p1 p2 p3 p4 p5 p6 p7);
 input indiv sexo $ p1 p2 p3 p4 p5 p6 p7;
 output frangoM;
 y=p1; idade=1; output frangoU;
 y=p2; idade=2; output frangoU;
 y=p3; idade=3; output frangoU;
 y=p4; idade=4; output frangoU;
 y=p5; idade=5; output frangoU;
 y=p6; idade=6; output frangoU;
 y=p7; idade=7; output frangoU;
 cards;
01  F  122  291  500   712  1041  1430  1760
02  F  129  314  551   830  1096  1485  1820
03  F  133  308  563   857  1085  1422  1660
04  F  135  348  584   854  1109  1493  1760
05  F  110  286  556   782  1105  1538  1870
06  F  130  302  518   740  1009  1337  1630
07  F  133  336  630   831  1108  1514  1760
08  F  138  337  618   937  1144  1570  1820
09  F  153  352  637   830  1052  1464  1820
10  F  138  332  484   767  1132  1548  1870
11  F  137  329  576   844  1127  1391  1660
12  F  133  298  464   670   988  1387  1720
13  F  142  345  598   844  1172  1570  1860
14  M  147  365  702   974  1293  1850  2220
15  M  126  331  624   784  1128  1567  1900
16  M  141  327  594   852  1029  1463  1820
17  M   94  262  547   812  1090  1562  1850
18  M  119  311  588   864  1184  1681  2100
19  M  114  315  613   845  1180  1571  1950
20  M  127  310  589   836  1248  1702  2120
21  M  111  306  604   921  1297  1880  2270
22  M  141  347  654   953  1365  1845  2180
23  M  153  356  691  1014  1457  1897  2380
24  M  136  351  664   906  1235  1735  2060
25  M  132  335  670  1000  1411  1831  2130
26  M  125  289  577   830  1232  1700  2110
27  M  123  323  611   864  1014  1449  1760
28  M  126  320  596   872  1267  1735  2150
29  M  137  334  610   904  1264  1624  1900
30  M  152  353  654   964  1302  1744  2150
31  M  139  349  635   948  1338  1773  2180
32  M  118  277  591   870  1256  1738  2050
;
proc glm data=frangoM;
 title 'Exemplo 2';
 class sexo;
 model p1-p7 = sexo / nouni;
 repeated idade 7 (1 2 3 4 5 6 7) polynomial / printe summary;

proc mixed data=frangoU info;
 class sexo idade;
 model y = sexo idade sexo*idade;
 repeated / type=UN subject=indiv r;

proc mixed data=frangoU info;
 class sexo ;
 model y = sexo sexo*idade sexo*idade*idade / htype=1 solution noint;
 repeated / type=UN subject=indiv r;
 contrast 'Compara grau 2:' sexo*idade*idade  1 -1;
 contrast 'Compara grau 1:' sexo*idade  1 -1;
 contrast 'Compara grau 0:' sexo  1 -1;
run;
```