



CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS - UFES / DCFM

A METODOLOGIA DE MODELOS MISTOS APLICADA À ANÁLISE DE DADOS COM MEDIDAS REPETIDAS

Prof. Dr. Adair José Regazzi

Pesquisador Bolsista DCR-A

Apoio: UFES/CNPq/FAPES

adairreg@ufv.br

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
DEPARTAMENTO DE CIÊNCIAS FLORESTAIS E DA MADEIRA

**A METODOLOGIA DE MODELOS MISTOS
APLICADA À ANÁLISE DE DADOS COM
MEDIDAS REPETIDAS**

Prof. Adair José Regazzi

Prof. Gilson Fernandes da Silva

Jerônimo Monteiro – ES

Agosto de 2018

Assuntos abordados: 1ª PARTE

I – PRELIMINARES

1.1. OPERAÇÕES MATRICIAIS BÁSICAS

- Transposta ($'$)
- Multiplicação de matrizes
- Determinante de uma matriz ${}_n A_n$
- Inversa ($^{-1}$)
- Inversa generalizada ($^{-}; +; \ell$)

- Posto de uma matriz
- Vetor
- Vetor normalizado
- Autovalores e autovetores

1.2. ESTATÍSTICA EXPERIMENTAL

1.2.1. Alguns conceitos

1.2.2. Exemplo 1: Blocos completos casualizados (DBC)

1.2.3. Exemplo 2: DBC/Split-Plot

1.3. DISTRIBUIÇÃO NORMAL MULTIVARIADA

1.4. ANÁLISE DE VARIÂNCIA MULTIVARIADA E TESTES DE HIPÓTESES

Assuntos abordados: 2ª PARTE

II – ANÁLISE DE DADOS COM MEDIDAS REPETIDAS

2.1. Definições

2.2. Objetivos

2.3. Métodos de análise estatística

2.3.1. **Análise Univariada de Perfis**

- Parcelas subdivididas
- Parcelas subdivididas com correções para os números de graus de liberdade

2.3.2. Análise Multivariada de Perfis

- Perfis paralelos
- Perfis coincidentes
- Perfis horizontais

2.3.3. Metodologia de Modelos Mistos

- Modelo estatístico
- Distribuições conjunta, marginal e condicional

- Estimando $\tilde{\beta}$ e predizendo \tilde{y}
- Seleção da estrutura de covariância
 - (i) Teste da razão de verossimilhanças
 - (ii) Critérios de informação
- Inferência sobre os efeitos fixos e aleatórios

2.4. Exemplos de Aplicação

2.4.1. Análises utilizando o PROC

GLM do SAS

- Análise com o modelo de parcelas subdivididas
- Análise univariada de perfis
- Análise multivariada de perfis
- Análise de tendência de variação

2.4.2. Análises utilizando o PROC MIXED do SAS

- Modelando a estrutura de covariância
- Escolha da estrutura de covariância
- Efeito da estrutura de covariância nos testes e estimativas
- Modelando o tempo como uma variável de regressão (covariável)

2.5. EXTENSÕES:

- ▶ **Modelos Mistos Generalizados**
- ▶ **Modelos Mistos Não Lineares**
- ▶ **Modelos Mistos sob enfoque Bayesiano**

2.6. REFERÊNCIAS

1ª PARTE

I – PRELIMINARES

1.1. OPERAÇÕES MATRICIAIS BÁSICAS

- Transposta (')

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

$$A' = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

• Multiplicação de matrizes

Duas matrizes podem ser multiplicadas, se e somente se, o número de colunas da primeira for igual ao número de linhas da segunda matriz.

Exemplo:

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (2 \times 1 + 3 \times 1 + 1 \times 0) & (2 \times 1 + 3 \times 0 + 1 \times 1) \\ (1 \times 1 + 2 \times 1 + 3 \times 0) & (1 \times 1 + 2 \times 0 + 3 \times 1) \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$$

- Determinante de uma matriz $n \times n$ A_n

O determinante da matriz $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

é dado por: $\text{Det}(A) = |A| = a_{11}a_{22} - a_{12}a_{21}$

O determinante da matriz $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

pode ser obtido assim: $\text{Det}(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & | & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & | & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & | & a_{31} & a_{32} \end{vmatrix}$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33})$$

► Expansão de um determinante pelos elementos de uma linha i :

$$\det(A) = |A| = \sum_{j=1}^n a_{ij} (-1)^{i+j} |M_{ij}|, \text{ para qualquer } i,$$

onde $|M_{ij}|$ é o menor de a_{ij} , que é o determinante da matriz obtida de A , suprimindo-se a i -ésima linha e a j -ésima coluna.

► Expansão de um determinante pelos elementos de uma coluna j :

$$\det(A) = |A| = \sum_{i=1}^n a_{ij} (-1)^{i+j} |M_{ij}|, \text{ para qualquer } j,$$

onde $|M_{ij}|$ é o menor de a_{ij} , que é o determinante da matriz obtida de A , suprimindo-se a i -ésima linha e a j -ésima coluna.

- Inversa (-1)

**Um método para determinar a inversa de uma matriz A
(A não singular)**

$$A^{-1} = \frac{(\text{Cofatores de } A)'}{\text{Det}(A)}$$

em que:

- a) $\text{Det}(A)$ é o determinante da matriz A ;
- b) Cofatores de A é a matriz dos cofatores de A .

O cofator do elemento a_{ij} é $c_{ij} = (-1)^{i+j} D_{ij}$, onde D_{ij} é o determinante da matriz resultante após a eliminação da linha i e da coluna j da matriz original A .

Exemplo:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \\ 0 & 5 & 3 \end{bmatrix} \text{ e } B = \begin{bmatrix} 2 & -1 & 0 \\ -3 & 3 & -1 \\ 5 & -5 & 2 \end{bmatrix} = A^{-1}$$

B é a inversa de A, pois $A^{-1}A = AA^{-1} = I$

$$BA = \begin{bmatrix} 2 & -1 & 0 \\ -3 & 3 & -1 \\ 5 & -5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \\ 0 & 5 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

- Inversa generalizada ($-; +; \ell$)

Uma solução para um sistema consistente de equações $Ax = g$ pode ser expresso em termos de uma inversa generalizada de A . Neste texto, dada a matriz ${}_m A_n$ então sua inversa generalizada ${}_n A_m^G$ terá os seguintes nomes e notações:

- Inversa condicional de $A = A^-$
- Inversa de Moore-Penrose de $A = A^+$
- Inversa de mínimos quadrados de $A = A^\ell$

Inversa condicional de $A(A^-)$

Definição: Dada uma matriz ${}_m A_n$, então toda matriz ${}_n A_m^-$ que satisfaz a condição

$$A A^- A = A$$

é definida como uma inversa condicional de A .

→ Uma inversa condicional só é única quando A for não-singular. Neste caso, $A^- = A^{-1}$.

Exemplo: Obter uma A^- para $A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$.

Uma A^- pode ser: $A^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$

Inversa de Moore-Penrose de A (A^+)

Definição: Dada a matriz ${}_m A_n$ de posto k , então a matriz ${}_n A_m^+$ de posto k , que satisfaz às quatro condições seguintes, é definida como a inversa de Moore-Penrose de A .

$$\text{i) } AA^+A = A$$

$$\text{ii) } A^+AA^+ = A^+$$

$$\text{iii) } AA^+ = (AA^+)' \quad (\text{simétrica})$$

$$\text{iv) } A^+A = (A^+A)' \quad (\text{simétrica})$$

Para obtenção de A^+ , fatora-se a matriz ${}_m A_n$ na forma ${}_m A_n = {}_m B_{k \times k} C_n$, onde $\text{posto}(A) = k > 0$, e usa-se a expressão:

$$A^+ = C'(CC')^{-1}(B'B)^{-1}B'$$

Exemplo: Dada a matriz

$$A = \begin{bmatrix} 2 & 6 & 4 & 2 \\ 4 & 15 & 14 & 7 \\ 2 & 9 & 10 & 5 \end{bmatrix}, \quad \text{obter a}$$

inversa de Moore-Penrose de A.

$$A^+ = \frac{1}{78} \begin{bmatrix} 14 & 3 & -11 \\ 19 & 5 & -14 \\ -18 & -2 & 16 \\ -9 & -1 & 8 \end{bmatrix}$$

Sem dúvida, satisfaz às quatro condições da definição de A^+ .

Inversa de mínimos quadrados de A (A^ℓ)

Definição: Dada a matriz ${}_m A_n$, então uma matriz ${}_n A_m^\ell$ que satisfaz às duas condições seguintes, é definida como inversa de mínimos quadrados de A.

i) $AA^\ell A = A$

ii) $(AA^\ell)' = AA^\ell$ (simétrica)

Obtenção de uma A^ℓ : $A^\ell = (A'A)^- A'$

Exemplo: Seja $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, obter uma A^ℓ .

$$A^\ell = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

- **Posto de uma matriz**

O posto (“rank”) de uma matriz $m \times n$ é igual ao número de linhas ou colunas linearmente independentes, ou ainda, é igual a ordem do maior determinante não nulo da matriz.

Exemplo: $A = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 3 & 1 & 4 & 2 \\ 5 & 2 & 9 & 2 \end{bmatrix}$

Posto (A) = 2, pois $l_3 = 2 l_2 - l_1$ e $c_3 = 7 c_2 - c_1$

$$c_4 = 2 c_1 - 4 c_2$$

• Vetor e vetor normalizado

A norma Euclidiana de um vetor real $\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$

é definida como:

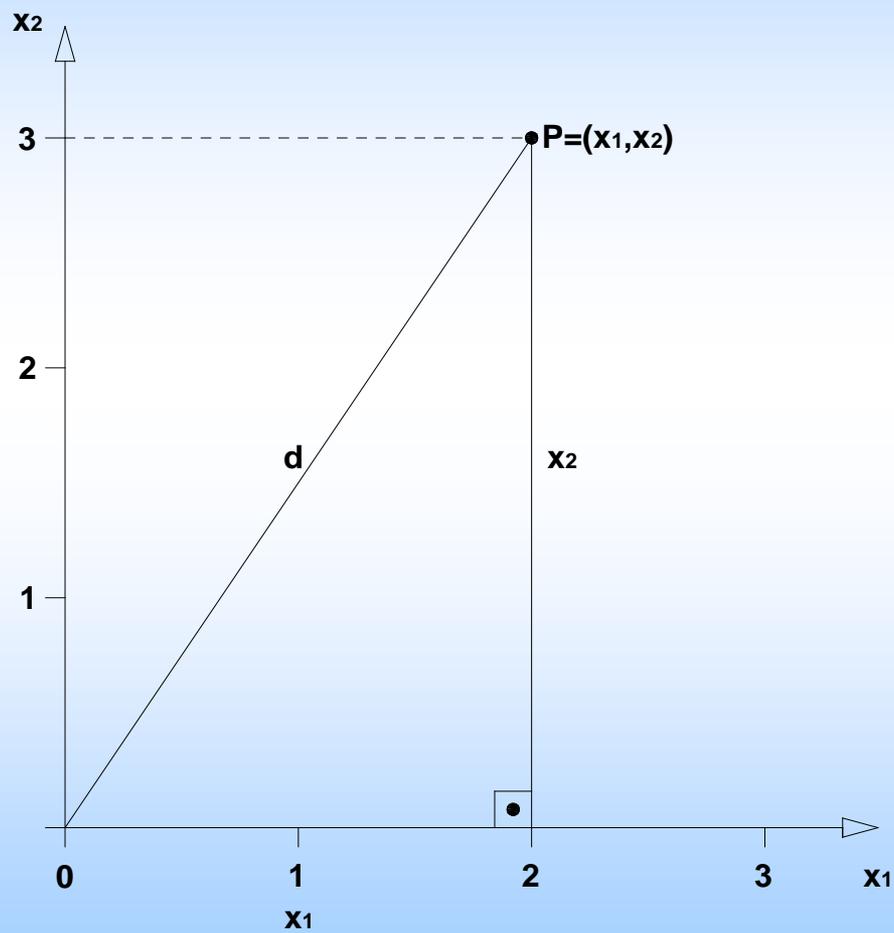
$$\text{norma de } \tilde{x} = \|\tilde{x}\| = \sqrt{\tilde{x}'\tilde{x}} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}$$

Exemplo: Para $\tilde{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, sua norma é:

$$\|\tilde{x}\| = \sqrt{2^2 + 3^2} = \sqrt{4 + 9} = \sqrt{13}$$

Vetor normalizado: $\tilde{v} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}$. Note que: $\tilde{v}'\tilde{v} = 1$

$\therefore d = \sqrt{13}$ é o comprimento do vetor \vec{x} .



• Autovalores e autovetores

Definição: Seja a equação

$$A \underline{\tilde{x}} = \lambda \underline{\tilde{x}}$$

em que, A é uma matriz quadrada de ordem n , $\underline{\tilde{x}}$ é um vetor $n \times 1$ não-nulo e λ um escalar tal que $\lambda \in \mathbb{R}$. O número real λ tal que $A \underline{\tilde{x}} = \lambda \underline{\tilde{x}}$ é chamado autovalor de A associado ao autovetor $\underline{\tilde{x}}$. Os autovalores são também denominados valores próprios ou valores característicos, enquanto os autovetores são também denominados vetores próprios ou vetores característicos.

Exemplo: Considere a matriz

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

Determinação dos autovalores λ_1 e λ_2

A equação característica da matriz A é:

$$|A - \lambda I| = 0 \rightarrow \begin{vmatrix} 3 - \lambda & -\sqrt{2} \\ -\sqrt{2} & 2 - \lambda \end{vmatrix} = 0$$

$$\lambda^2 - 5\lambda + 4 = 0 \rightarrow \lambda = \frac{5 \pm \sqrt{25 - 16}}{2} = \frac{5 \pm 3}{2} \rightarrow \lambda_1 = 4 \text{ e } \lambda_2 = 1 \text{ (autovalores)}$$

Determinação dos autovetores:

Um autovetor $\tilde{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$ associado a $\lambda_1 = 4$, é dado por:

$$(A - \lambda_1 I) \tilde{x}_1 = \tilde{0} \quad \rightarrow \quad A \tilde{x}_1 = \lambda_1 \tilde{x}_1$$

$$\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 4 \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} \quad \dots \quad \begin{cases} x_{11} + \sqrt{2} x_{21} = 0 \\ \sqrt{2} x_{11} + 2 x_{21} = 0 \end{cases}$$

A matriz do sistema homogêneo de equações lineares é singular sendo o sistema consistente e indeterminado, isto é, admite infinitas soluções. Para obter uma solução não-nula, podemos abandonar uma das equações e dar a uma das incógnitas um valor arbitrário não-nulo. Por exemplo, abandonar a segunda equação e fazer $x_{21} = -1$.

Assim virá:

$$x_{11} - \sqrt{2} = 0 \quad \Rightarrow \quad x_{11} = \sqrt{2}$$

Logo, $\tilde{x}_1 = \begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix}$ é um autovetor associado a $\lambda_1 = 4$.

Cabe ressaltar que sempre haverá duas soluções, uma obtida da outra multiplicada por -1.

Um autovetor $\tilde{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$ associado a $\lambda_2 = 1$, pode

ser: $\tilde{x}_2 = \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$.

A forma normalizada dos vetores \tilde{x}_1 e \tilde{x}_2 são dadas por \tilde{v}_1 e \tilde{v}_2 , respectivamente, como a seguir:

$$\tilde{v}_1 = \frac{1}{\|\tilde{x}_1\|} \tilde{x}_1 = \frac{1}{\sqrt{x_{11}^2 + x_{21}^2}} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

$$\tilde{v}_1 = \frac{1}{\sqrt{(\sqrt{2})^2 + (-1)^2}} \begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}$$

$$\tilde{v}_2 = \frac{1}{\|\tilde{x}_2\|} \tilde{x}_2 = \frac{1}{\sqrt{x_{12}^2 + x_{22}^2}} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

$$\tilde{v}_2 = \frac{1}{\sqrt{1^2 + (\sqrt{2})^2}} \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ \sqrt{2}/\sqrt{3} \end{bmatrix}$$

Note que $\tilde{v}'_1 \tilde{v}_1 = 1$, $\tilde{v}'_2 \tilde{v}_2 = 1$ e $\tilde{v}'_1 \tilde{v}_2 = 0$, isto é, \tilde{v}_1 e \tilde{v}_2 são

vetores ortonormais (normalizados e ortogonais).

1.2. ESTATÍSTICA EXPERIMENTAL

1.2.1. Alguns conceitos

- **Valor-p (“p-value”)**

Chama-se valor-p (ou nível crítico ou ainda probabilidade de significância) ao menor nível de significância α , para o qual o resultado observado seria declarado significativo, isto é, para o qual rejeitaríamos a hipótese que está sendo considerada.

O valor-p associado ao resultado de um teste é usualmente muito mais informativo do que uma simples afirmação sobre se uma dada hipótese deve ou não ser rejeitada a um determinado nível de significância.

- **Componentes de Variância:** são as variâncias associadas aos efeitos aleatórios de um modelo estatístico.

Importância: → Genética Quantitativa, Teste F, Amostragem, etc.

- **Modelos: fixo, aleatório e misto**

Sejam I tratamentos repetidos J vezes num ensaio em blocos completos casualizados (DBC).

O modelo é $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$.

Modelo fixo:

Se todos os parâmetros do modelo, com exceção do erro, forem de efeitos fixos, o modelo é chamado fixo.

Neste caso o interesse é estimar e testar hipóteses de funções lineares dos efeitos.

Modelo aleatório:

Se todos os efeitos do modelo, com exceção da média, forem aleatórios, o modelo é chamado aleatório.

Neste caso o interesse é estimar componentes de variância e fazer previsões.

Modelo Misto:

Se o modelo envolver parâmetros de efeitos fixos e também efeitos aleatórios, excetuando-se nestas considerações a média e o erro, o modelo é chamado misto.

• Somas de Quadrados Utilizando a Notação de Redução R – GLM/SAS

Modelo Estatístico:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \text{com } i = 1, 2, \dots, I \text{ (tratamentos)}$$
$$j = 1, 2, \dots, J \text{ (blocos)}$$

- Se $n=IJ$ observações \Rightarrow Blocos Completos (DBC).
- Se pelo menos um tratamento não estiver presente em todos os blocos \Rightarrow Blocos Incompletos (BI).
- Média ajustada para o i -ésimo tratamento (LSMEANS):

$$\bar{Y}_{i.} = \mu^0 + \alpha_i^0 + \frac{1}{J} \sum_{j=1}^J \beta_j^0$$

MODEL $y = \text{blocos trat}$;

FV	SQTipo I	SQTipo III
Blocos	$R(\beta / \mu)$	$R(\beta / \mu, \alpha)$
Tratamentos	$R(\alpha / \mu, \beta)$	$R(\alpha / \mu, \beta)$

Decomposição Ortogonal

Decomposição não ortogonal

MODEL $y = \text{trat blocos}$;

FV	SQTipo I	SQTipo III
Tratamentos	$R(\alpha / \mu)$	$R(\alpha / \mu, \beta)$
Blocos	$R(\beta / \mu, \alpha)$	$R(\beta / \mu, \alpha)$

Decomposição Ortogonal

Decomposição não ortogonal

Obs.: Dados Balanceados \Rightarrow SQTipo I = SQTipo III

1.2.2. Exemplo 1: Blocos completos casualizados (DBC)

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \text{ com } i=1, 2, \dots, I \text{ e } j=1, 2, \dots, J$$

Modelo Fixo: $\sum_{i=1}^I \alpha_i = 0$ e $\sum_{j=1}^J \beta_j = 0$

Tabela 1.1- Esquema da ANOVA com as E(QM) e o teste F

FV	GL	QM	E(QM)	F*
Blocos	J-1	QMB	$\sigma^2 + I\Phi(\beta)$	
Tratamentos	I-1	QMT	$\sigma^2 + J\Phi(\alpha)$	QMT/QMR
Resíduo	(I-1)(J-1)	QMR	σ^2	

* $H_0 : \alpha_i = 0 \quad \forall i$

$$\Phi(\beta) = \frac{\sum_{j=1}^J \beta_j^2}{J-1}, \quad \Phi(\alpha) = \frac{\sum_{i=1}^I \alpha_i^2}{I-1} \quad \text{e} \quad \hat{V}(\hat{m}_i) = \frac{QMR}{J}, \text{ com } (I-1)(J-1) \text{ gl.}$$

Modelo Aleatório:

Tabela 1.2 - Esquema da ANOVA com as E(QM) e o teste F

FV	GL	QM	E(QM)	F*
Blocos	J-1	QMB	$\sigma^2 + I\sigma_\beta^2$	
Tratamentos	I-1	QMT	$\sigma^2 + J\sigma_\alpha^2$	QMT / QMR
Resíduo	(I-1)(J-1)	QMR	σ^2	

* $H_0 : \sigma_\alpha^2 = 0$

$$\hat{\sigma}^2 = \text{QMR}$$

$$\hat{\sigma}_\alpha^2 = \frac{\text{QMT} - \text{QMR}}{J}$$

$$\hat{\sigma}_\beta^2 = \frac{\text{QMB} - \text{QMR}}{I}$$

Modelo Misto:

Tabela 1.3 - Esquema da ANOVA com as E(QM)

FV	GL	QM	E(QM)	
			α_i fixo*	α_i aleatório
Blocos	J-1	QMB	$\sigma^2 + I\sigma_\beta^2$	$\sigma^2 + I\Phi(\beta)$
Tratamentos	I-1	QMT	$\sigma^2 + J\Phi(\alpha)$	$\sigma^2 + J\sigma_\alpha^2$
Resíduo	(I-1)(J-1)	QMR	σ^2	σ^2

$$* \hat{V}(\hat{m}_i) = \frac{\hat{\sigma}^2 + \hat{\sigma}_\beta^2}{J} = \frac{1}{IJ} [(I-1)QMR + QMB], \text{ com } n' \text{ gl. } n' = \frac{[(I-1)QMR + QMB]^2}{\frac{[(I-1)QMR]^2}{(I-1)(J-1)} + \frac{(QMB)^2}{J-1}}$$

Obs: Dados balanceados → E(QM) → HICKS (1973)

```

data DBC;
input bloco trat $ y;
cards;
1 A 67.0
1 B 71.9
1 C 72.2
2 A 67.5
2 B 68.8
2 C 66.4
3 A 76.0
3 B 82.6
3 C 74.5
4 A 72.7
4 B 78.1
4 C 67.3
5 A 73.1
5 B 74.2
5 C 73.2
6 A 65.8
6 B 70.8
6 C 68.7
7 A 75.6
7 B 84.9
7 C 69.0
;

```

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$



```

proc mixed data=DBC;
class bloco trat;
model y = trat / ddfm=satterth;
random bloco;
Estimate 'media trat a' intercept 1 trat 1 0 0 / cl;
lsmeans trat / diff;
run;

```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trat	2	12	6.36	0.0131

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
media trat a	71.1000	1.7655	11.6	40.27	<.0001	0.05	67.2388	74.9612

Least Squares Means

Effect	trat	Estimate	Standard Error	DF	t Value	Pr > t
trat	A	71.1000	1.7655	11.6	40.27	<.0001
trat	B	75.9000	1.7655	11.6	42.99	<.0001
trat	C	70.1857	1.7655	11.6	39.75	<.0001

Differences of Least Squares Means

Effect	trat	_trat	Estimate	Standard Error	DF	t Value	Pr > t
trat	A	B	-4.8000	1.7214	12	-2.79	0.0164
trat	A	C	0.9143	1.7214	12	0.53	0.6050
trat	B	C	5.7143	1.7214	12	3.32	0.0061

```
proc glm data=DBC;  
class bloco trat;  
model y = bloco trat;  
random bloco;  
lsmeans trat / stderr pdiff;  
run;
```

► **MODELO MISTO → GLM → CUIDADO?????**

The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	400.1904762	50.0238095	4.82	0.0076
Error	12	124.4590476	10.3715873		
Corrected Total	20	524.6495238			

R-Square	Coeff Var	Root MSE	y Mean
0.762777	4.448490	3.220495	72.39524

Source	DF	Type I SS	Mean Square	F Value	Pr > F
bloco	6	268.2895238	44.7149206	4.31	0.0151
trat	2	131.9009524	65.9504762	6.36	0.0131

Source	DF	Type III SS	Mean Square	F Value	Pr > F
bloco	6	268.2895238	44.7149206	4.31	0.0151
trat	2	131.9009524	65.9504762	6.36	0.0131

The GLM Procedure
Least Squares Means

trat	y LSMEAN	Standard Error	Pr > t	LSMEAN Number
A	71.1000000	1.2172327	<.0001	1
B	75.9000000	1.2172327	<.0001	2
C	70.1857143	1.2172327	<.0001	3

Least Squares Means for effect trat
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: y

i/j	1	2	3
1		0.0164	0.6050
2	0.0164		0.0061
3	0.6050	0.0061	

1.2.3. Exemplo 2: DBC/Split-Plot

Os experimentos clássicos em parcelas subdivididas (“Split plot”) se caracterizam pela sua estruturação através de **tratamentos principais** (ou tratamentos A) nas parcelas, e estas, por sua vez, são constituídas de **tratamentos secundários** (ou tratamentos B), que são as subparcelas.

Por exemplo: Fator A (Fórmulas de Adubação) em três níveis - A_1 , A_2 e A_3 , distribuídos em 4 blocos completos casualizados. Fator B (Clones de Eucalipto) em dois níveis - B_1 e B_2 , sobreposto, dividindo-se cada parcela com A em 2 subunidades, e casualizando os níveis de B nestas 2 subunidades.

Bloco1:

A_2B_2	A_1B_1	A_3B_2
A_2B_1	A_1B_2	A_3B_1

Bloco 2:

A_1B_1	A_3B_2	A_2B_2
A_1B_2	A_3B_1	A_2B_1

Bloco 3:

A_2B_1	A_3B_2	A_1B_1
A_2B_2	A_3B_1	A_1B_2

Bloco 4:

A_3B_2	A_1B_1	A_2B_2
A_3B_1	A_1B_2	A_2B_1

Figura 1 – Croqui da distribuição dos níveis dos fatores A e B, em um experimento em blocos completos casualizados com os tratamentos dispostos no esquema de parcelas subdivididas

Fato: Se o fator B for avaliação na mesma unidade experimental ao longo do tempo não há aleatorização desse fator. Como analisar? Veremos na 2ª PARTE da apresentação.

O modelo estatístico normal clássico desse delineamento é:

$$y_{ijk} = \underbrace{\mu + a_i + b_j + ab_{ij}}_{\text{FIXOS}} + \underbrace{r_k + \delta_{ik} + \varepsilon_{ijk}}_{\text{ALEATÓRIOS}} \quad (1)$$

em que:

$i = 1, 2, \dots, a$ é o índice para níveis do fator A (tratamento das parcelas);

$j = 1, 2, \dots, b$ é o índice para níveis do fator B (tratamento das subparcelas);

$k = 1, 2, \dots, r$ é o índice para blocos;

Sejam as seguintes restrições e suposições:

$$\sum_{i=1}^a a_i = \sum_{j=1}^b b_j = \sum_{i=1}^a \sum_{j=1}^b ab_{ij} = \sum_{j=1}^b \sum_{i=1}^a ab_{ij} = 0$$

$$r_k \text{ IID } N(0, \sigma_r^2), \delta_{ik} \text{ IID } N(0, \sigma_\delta^2) \text{ e } \varepsilon_{ijk} \text{ IID } N(0, \sigma^2).$$

Os diferentes efeitos aleatórios r_k , δ_{ik} e ε_{ijk} são independentes entre si.

Tabela 1.4 - Esquema da análise de variância e teste F, considerando o delineamento em blocos completos casualizados com os tratamentos no esquema de parcelas subdivididas

FV	GL	QM	E(QM)*	F
Blocos	r-1	QMblocc.	$\sigma^2 + b\sigma_{\delta}^2 + ab\sigma_r^2$	-
Fator A	a-1	QMA	$\sigma^2 + b\sigma_{\delta}^2 + rb\phi(a)$	$F_1 = QMA/QMR(a)$
Resíduo(a)	(r-1)(a-1)	QMR(a)	$\sigma^2 + b\sigma_{\delta}^2$	
Fator B	b-1	QMB	$\sigma^2 + ra\phi(b)$	$F_2 = QMB/QMR(b)$
A x B	(a-1)(b-1)	QMAxB	$\sigma^2 + r\phi(ab)$	$F_3 = QMAxB/QMR(b)$
Resíduo(b)	a(r-1)(b-1)	QMR(b)	σ^2	
Total	n-1=abr-1			

* **r**, blocos; **a**, níveis de A; e **b**, níveis de B.

$$\phi(a) = \frac{\sum_{i=1}^a a_i^2}{a-1}, \quad \phi(b) = \frac{\sum_{j=1}^b b_j^2}{b-1}, \quad \phi(ab) = \frac{\sum_{i=1}^a \sum_{j=1}^b ab_{ij}^2}{(a-1)(b-1)}$$

As hipóteses associadas aos testes F apresentados na Tabela 1.4 são:

a) $F_1 = QMA/QMR(a)$, testa $H_0: a_1 = a_2 = \dots = a_a = 0$ vs. H_a : não H_0 .

b) $F_2 = QMB/QMR(b)$, testa $H_0: b_1 = b_2 = \dots = b_b = 0$ vs. H_a : não H_0 .

c) $F_3 = QMAXB/QMR(b)$, testa $H_0: ab_{11} = ab_{12} = \dots = ab_{ab} = 0$ vs. H_a : não H_0 .

Sob as respectivas hipóteses nulas associadas e também sob as suposições impostas ao modelo dado anteriormente, as estatísticas F_3 , F_2 e F_1 da Tabela 1.4 seguem distribuição exata F central com $(a-1)(b-1)$ e $a(r-1)(b-1)$, $(b-1)$ e $a(r-1)(b-1)$, e $(a-1)$ e $(r-1)(a-1)$ graus de liberdade, respectivamente.

A variância para um contraste entre médias de dois tratamentos das parcelas (fator A) é estimada por:

$$\hat{V}(\bar{y}_{i..} - \bar{y}_{i'..}) = \frac{2}{rb} \text{QMR}(a)$$

A variância para um contraste entre médias de dois tratamentos das subparcelas (fator B) é estimada por:

$$\hat{V}(\bar{y}_{.j.} - \bar{y}_{.j'..}) = \frac{2}{ra} \text{QMR}(b)$$

A variância para um contraste entre médias de dois tratamentos das subparcelas dentro de um tratamento de parcela é estimada por:

$$\hat{V}(\bar{y}_{ij.} - \bar{y}_{ij'..}) = \frac{2}{r} \text{QMR}(b)$$

A variância para um contraste entre médias de dois tratamentos das parcelas dentro de um tratamento de subparcela é estimada por:

$$\hat{V}(\bar{y}_{ij.} - \bar{y}_{i'j.}) = \frac{2}{r} \left[\frac{\text{QMR}(a) + (b-1)\text{QMR}(b)}{b} \right] = \frac{2}{r} \text{QMR}_{\text{comb}}$$

sendo QMR_{comb} um resíduo combinado. O número de graus de liberdade (n') associado ao QMR_{comb} é dado pela aproximação de Satterthwaite (1946):

$$n' = \frac{[\text{QMR}(a) + (b-1)\text{QMR}(b)]^2}{\frac{[\text{QMR}(a)]^2}{n_a} + \frac{[(b-1)\text{QMR}(b)]^2}{n_b}}$$

em que n_a e n_b são os graus de liberdade do resíduo(a) e resíduo(b), respectivamente; as variâncias e os graus de liberdade são necessários para construção de intervalos de confiança ou para testar hipóteses. Em um possível desdobramento da interação para estudar o fator B dentro de A, a estatística F tem como denominador o QMR(b). Já no estudo do fator A dentro de B, o denominador apropriado é o resíduo combinado (QMR_{comb}).

ILUSTRAÇÃO DO MÉTODO

Vamos considerar um experimento em parcelas subdivididas com 3 tratamentos nas parcelas (Fator A com 3 níveis, sendo 3 fórmulas de adubação) e 2 nas subparcelas (Fator B com 2 níveis, sendo 2 clones de eucalipto), com 4 repetições em blocos completos casualizados, cujos resultados experimentais sobre o volume de madeira ($\text{m}^3/\text{subparcela}$) estão apresentados na Tabela 1.5.

Tabela 1.5 – Dados sobre o volume de madeira (m³/subparcela)*

Blocos	A ₁		A ₂		A ₃		Totais
	B ₁	B ₂	B ₁	B ₂	B ₁	B ₂	
1	56	41	50	36	39	35	257
2	30	25	36	28	33	30	182
3	32	24	31	27	15	19	148
4	30	25	35	30	17	18	155
Totais	148	115	152	121	104	102	742

* Adaptado de Stroup (1989)

- Delineamento experimental: blocos completos casualizados.
- Trata-se de experimento em **parcelas subdivididas** com três tratamentos nas parcelas (Fator A com três níveis) e dois nas subparcelas (Fator B com dois níveis), com quatro repetições em blocos completos casualizados.

Tabela 1.6 – Valores médios do volume de madeira (m³/subparcela)*

Fator A \ Fator B	A ₁	A ₂	A ₃	Médias
B ₁	37,00	38,00	26,00	33,67
B ₂	28,75	30,25	25,50	28,17
Médias	32,875	34,125	25,750	30,92

*Fator A – três fórmulas de adubação; Fator B – dois clones de eucalipto

Tabela 1.7 – Resultado da análise de variância (ANOVA)

FV	GL	SQ	QM	F	Pr>F
Blocos	3	1243,5000			
Fator A	2	326,5833	163,2916	4,07 ^{ns}	0,0764
Resíduo (a)	6	240,7500	40,1250		
(Parcelas)	(11)	(1810,8333)			
Fator B	1	181,5000	181,5000	19,39 ^{**}	0,0017
Interação AxB	2	75,2500	37,6250	4,02 ^{ns}	0,0566
Resíduo (b)	9	84,2500	9,3611		
Total	23	2151,8333			

** p<0,01; n.s. p>0,05

Valores de F tabelado:

- Fator A

$$2 \text{ e } 6\text{gl} \begin{cases} 5\% = 5,14 \\ 1\% = 10,92 \end{cases}$$

- Fator B

$$1 \text{ e } 9\text{gl} \begin{cases} 5\% = 5,12 \\ 1\% = 10,56 \end{cases}$$

- Interação AxB

$$2 \text{ e } 9\text{gl} \begin{cases} 5\% = 4,26 \\ 1\% = 8,02 \end{cases}$$

Coeficientes de variação:

Nos experimentos em parcelas subdivididas, temos dois coeficientes de variação:

(a) Em nível de parcelas:

$$CV(a) = \frac{\sqrt{QMR(a)}}{\bar{Y}} \cdot 100 \qquad \bar{Y} = \frac{\sum_{i,j,k} y_{ijk}}{n} = \frac{742}{24} = 30,92$$

$$CV(a) = \frac{\sqrt{40,1250}}{30,92} \cdot 100 = 20,49\%$$

(b) Em nível de subparcelas:

$$CV(b) = \frac{\sqrt{QMR(b)}}{\bar{Y}} \cdot 100$$

$$CV(b) = \frac{\sqrt{9,3611}}{30,92} \cdot 100 = 9,89\%$$

No nosso exemplo, apesar de a interação ter sido não significativa ($p > 0,05$), vamos mostrar como proceder aos desdobramentos.

1° Desdobramento: B/A

Tabela 1.8 – Estudo do efeito do fator B em cada nível do fator A

FV	GL	SQ	QM	F	Pr>F
B/A ₁	1	136,1250	136,1250	14,54 ^{**}	0,0041
B/A ₂	1	120,1250	120,1250	12,83 ^{**}	0,0059
B/A ₃	1	0,5000	0,5000	0,05 ^{n.s.}	0,8224
Resíduo (b)	9	84,2500	9,3611		

** $p < 0,01$; n.s. $p > 0,05$

Valores de F tabelado: $F_{1\%} (1; 9) = 10,56$

$F_{5\%} (1; 9) = 5,12$

2° Desdobramento: A/B

Tabela 1.9 – Estudo do efeito do fator A em cada nível do fator B

FV	GL	SQ	QM	F	Pr>F
A/B ₁	2	354,6667	177,3333	7,17 [*]	0,0142
A/B ₂	2	47,1667	23,5833	0,95 ^{n.s}	0,4220
Resíduo Combinado	8,81	-	24,7430		

*p<0,05; n.s. p>0,05

Valores de F tabelado (obtidos pelo SAS):

$$F_{1\%} (2; 8,81) = 8,13 \text{ e } F_{5\%} (2; 8,81) = 4,29$$

Pela tabela de F, considerando aproximadamente 9 graus de liberdade associado ao resíduo combinado, a conclusão seria a mesma, pois, $F_{1\%} (2; 9) = 8,02$ e $F_{5\%} (2; 9) = 4,26$.

Comparações de médias pelo teste de Tukey:

(a) Entre duas médias de tratamentos principais

(Fator A):

$$A_1 = 32,875 \text{ a}$$

$$A_2 = 34,125 \text{ a}$$

$$A_3 = 25,750 \text{ a}$$

$$\hat{Y}_1 = \hat{m}_i - \hat{m}_{i'} = \bar{y}_{i..} - \bar{y}_{i'..} \quad (i \neq i')$$

A variância do contraste \hat{Y}_1 é estimada por:

$$\hat{V}(\hat{Y}_1) = \frac{2}{br} \text{QMR}(a)$$

$$q_{(5\%, 3, 6)} = 4,34$$

$$\Delta_1 = q_{(\alpha, a, glRes(a))} \sqrt{\frac{1}{2} \hat{V}(\hat{Y}_1)} = q_1 \sqrt{\frac{\text{QMR}(a)}{br}}$$

$$\Delta_1 = 4,34 \sqrt{\frac{40,1250}{8}} = 9,72$$

(b) Entre duas médias de tratamentos secundários
(Fator B):

$$\hat{Y}_2 = \hat{m}_j - \hat{m}_{j'} = \bar{y}_{.j} - \bar{y}_{.j'} \quad (j \neq j')$$

A variância do contraste \hat{Y}_2 é estimada por:

$$\hat{V}(\hat{Y}_2) = \frac{2}{ar} \text{QMR}(b)$$

O teste de Tukey nos leva a:

$$\Delta_2 = q_{(\alpha, b, g | \text{Res}(b))} \sqrt{\frac{1}{2} \hat{V}(\hat{Y}_2)} = q_2 \sqrt{\frac{\text{QMR}(b)}{ar}}$$

No exemplo temos as seguintes médias de tratamentos secundários (Fator B):

$$B_1 = 33,67 \text{ a} \quad a > b \text{ pelo teste F (} p < 0,05 \text{)}$$

$$B_2 = 28,17 \text{ b}$$

No exemplo, como temos apenas dois níveis do fator B, qualquer PPCM não deve ser aplicado, pois, neste caso, o teste F é conclusivo (Tabela 1.7). Vamos calcular Δ_2 apenas para ilustrar o método.

$$q_{(5\%, 2, 9)} = 3,20$$

$$\Delta_2 = 3,20 \sqrt{\frac{9,3611}{12}} = 2,83$$

(c) Entre duas médias de tratamentos secundários (Fator B), num mesmo tratamento principal (Fator A)

$$\hat{Y}_3 = \hat{m}_{ij} - \hat{m}_{ij'} = \bar{y}_{ij.} - \bar{y}_{ij'}. \quad (j \neq j')$$

A variância do contraste \hat{Y}_3 é estimada por:

$$\hat{V}(\hat{Y}_3) = \frac{2}{r} \text{QMR}(b)$$

O teste de Tukey nos leva a:

$$\Delta_3 = q_{(\alpha, b, glRes(b))} \sqrt{\frac{1}{2} \hat{V}(\hat{Y}_3)} = q_3 \sqrt{\frac{\text{QMR}(b)}{r}}$$

$$q_{(5\%, 2, 9)} = 3,20$$

$$\Delta_3 = 3,20 \cdot \sqrt{\frac{9,3611}{4}} = 4,89$$

Como o fator B possui apenas dois níveis, o teste de Tukey é equivalente ao teste F (Tabela 1.8), proporcionando as mesmas conclusões.

As médias com o resultado do teste F são apresentadas na Tabela 1.10.

Tabela 1.10 – Valores médios do volume de madeira (m³/parcela), com os resultados do teste F (coluna)*

Fator A \ Fator B		Fator A		
		A ₁	A ₂	A ₃
Fator B	B ₁	37,00 A	38,00 A	26,00 A
	B ₂	28,75 B	30,25 B	25,50 A

* Na coluna: A>B pelo teste F (P<0,05).

(d) Entre duas médias de tratamentos principais (Fator A), num mesmo tratamento secundário (Fator B):

$$\hat{Y}_4 = \hat{m}_{ij} - \hat{m}_{i'j} = \bar{y}_{ij.} - \bar{y}_{i'j.} \quad (i \neq i')$$

A variância do contraste \hat{Y}_4 é estimada por:

$$\hat{V}(\hat{Y}_4) = \frac{2}{r} \text{QMR}_{\text{comb}}, \text{ com } n' \text{ graus de liberdade.}$$

A d.m.s., pelo teste de Tukey em nível de significância α é:

$$\Delta_4 = q_{(\alpha, a, n')} \sqrt{\frac{1}{2} \hat{V}(\hat{Y}_4)} = q_4 \sqrt{\frac{\text{QMR}_{\text{comb}}}{r}}$$

Para $\alpha = 5\% \Rightarrow q_{(5\%; 3; 8,81)}$. Vamos usar aproxima-

$q_{(5\%; 3; 9)} = 3,95$. Então,

$$\Delta_4 = 3,95 \sqrt{\frac{24,7430}{4}} = 9,82$$

→ **Resultados dos testes:** B/A (Teste F)

A/B (Teste de Tukey)

Tabela 1.11 – Valores médios do volume de madeira (m³/parcela), com os resultados dos testes F (coluna) e Tukey (linha)*

Fator A \ Fator B	A ₁	A ₂	A ₃
B ₁	37,00 Aa	38,00 Aa	26,00 Ab
B ₂	28,75 Ba	30,25 Ba	25,50 Aa

* Na coluna: A>B pelo teste F (P<0,05). As médias seguidas de pelo menos uma letra minúscula em cada linha, não diferem significativamente pelo teste de Tukey (P>0,05).

```
data psplot;  
input bloco a b y @@;  
datalines;
```

```
1 1 1 56 1 1 2 41  
1 2 1 50 1 2 2 36  
1 3 1 39 1 3 2 35  
2 1 1 30 2 1 2 25  
2 2 1 36 2 2 2 28  
2 3 1 33 2 3 2 30  
3 1 1 32 3 1 2 24  
3 2 1 31 3 2 2 27  
3 3 1 15 3 3 2 19  
4 1 1 30 4 1 2 25  
4 2 1 35 4 2 2 30  
4 3 1 17 4 3 2 18
```

GLM

```
;  
proc glm data=psplot;  
class a b bloco;  
model y = bloco a a*bloco b a*b;  
random bloco a*bloco;  
test h=a e =a*bloco;  
means a / tukey e=a*bloco;  
means b / tukey;  
lsmeans a*b / slice=a; /* B/A: usa o resíduo correto: Resíduo(b) */  
lsmeans a*b / slice=b; /* A/B: usa o resíduo errado. Usou Resíduo(b).  
Tem que usar o resíduo combinado */  
run; /* ESTIMATE, CONTRAST, etc. */
```

```
data psplot;  
input bloco a b y @@;  
datalines;
```

```
1 1 1 56 1 1 2 41  
1 2 1 50 1 2 2 36  
1 3 1 39 1 3 2 35  
2 1 1 30 2 1 2 25  
2 2 1 36 2 2 2 28  
2 3 1 33 2 3 2 30  
3 1 1 32 3 1 2 24  
3 2 1 31 3 2 2 27  
3 3 1 15 3 3 2 19  
4 1 1 30 4 1 2 25  
4 2 1 35 4 2 2 30  
4 3 1 17 4 3 2 18  
;
```

```
proc mixed data=psplot;
```

```
class a b bloco;
```

```
model y = a b a*b / ddfm=satterth solution;
```

```
random bloco a*bloco /solution G V;
```

```
lsmeans a*b / slice=a diff;
```

```
lsmeans a*b / slice=b diff;
```

```
run; /* ESTIMATE, CONTRAST, etc. */
```

MIXED

No caso de modelos mistos o **PROC GLM** pode não computar as variâncias corretas para certas funções estimáveis e assim produzir testes incorretos. Isto ocorre principalmente nos casos envolvendo combinações lineares de quadrados médios. Esta é a principal razão para não usar o **PROC GLM** para análise de experimentos em parcelas subdivididas.

1.3. DISTRIBUIÇÃO NORMAL MULTIVARIADA

Definição

Seja a seguinte notação: $\exp(u) = e^u$

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_p \end{bmatrix} \quad \text{e} \quad \underset{\sim}{Y}' = [Y_1, Y_2, \dots, Y_p]$$

Sejam Y_1, Y_2, \dots, Y_p , **p** variáveis aleatórias do tipo contínuo.

A f.d.p. do vetor aleatório $\underline{Y}' = [Y_1, Y_2, \dots, Y_p]$

com distribuição normal multivariada ou multinormal é:

$$f_{Y_1, Y_2, \dots, Y_p}(y_1, y_2, \dots, y_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{y} - \underline{\mu})' \Sigma^{-1} (\underline{y} - \underline{\mu}) \right]$$

onde $\underline{\mu}' = [\mu_1, \mu_2, \dots, \mu_p]$ é um vetor de médias e Σ é

uma covariância real e simétrica positiva

definida e $-\infty < y_i < \infty$ para $i = 1, 2, \dots, p$.

Seja agora $\underline{Y} \sim N_p \left(\underline{y} : \underline{\mu}, \underline{\Sigma} \right)$ e a partição

$$\underline{Y} = \begin{bmatrix} \underline{Y}_r \\ \underline{Y}_s \end{bmatrix} \text{ (com } r + s = p); \underline{\mu} = \begin{bmatrix} \underline{\mu}_r \\ \underline{\mu}_s \end{bmatrix}; \underline{\Sigma} = \begin{bmatrix} \underline{\Sigma}_{rr} & \underline{\Sigma}_{rs} \\ \underline{\Sigma}_{sr} & \underline{\Sigma}_{ss} \end{bmatrix}$$

A distribuição marginal do vetor \underline{y}_r é dada por

$\underline{Y}_r \sim N_r \left(\underline{y}_r : \underline{\mu}_r; \underline{\Sigma}_{rr} \right)$, cuja f.d.p. será:

$$f_{\underline{Y}_r} \left(\underline{y}_r \right) = \frac{1}{(2\pi)^{r/2} |\underline{\Sigma}_{rr}|^{1/2}} \exp \left[-\frac{1}{2} \left(\underline{y}_r - \underline{\mu}_r \right)' \underline{\Sigma}_{rr}^{-1} \left(\underline{y}_r - \underline{\mu}_r \right) \right].$$

A distribuição condicional de \underline{Y}_r dado $\underline{Y}_s = \underline{y}_s$ é

dada por $\underline{Y}_r / \underline{Y}_s \sim N_r \left[\underline{y}_r : \underline{\mu}_{r,s}; \underline{\Sigma}_{rr.s} \right]$, cuja f.d.p. será:

$$f_{\underline{Y}_r / \underline{Y}_s} \left(\underline{y}_r / \underline{y}_s \right) = \frac{1}{(2\pi)^{r/2} \left| \underline{\Sigma}_{rr.s} \right|^{1/2}} \cdot \exp \left[-\frac{1}{2} \left(\underline{y}_r - \underline{\mu}_{r,s} \right)' \underline{\Sigma}_{rr.s}^{-1} \left(\underline{y}_r - \underline{\mu}_{r,s} \right) \right],$$

em que,
$$\begin{cases} \underline{\mu}_{r,s} = \underline{\mu}_r + \underline{\Sigma}_{rs} \underline{\Sigma}_{ss}^{-1} \left(\underline{y}_s - \underline{\mu}_s \right) \rightarrow \text{Média condicional} \\ \underline{\Sigma}_{rr.s} = \underline{\Sigma}_{rr} - \underline{\Sigma}_{rs} \underline{\Sigma}_{ss}^{-1} \underline{\Sigma}_{sr} \rightarrow \text{Variância condicional} \\ r + s = p \end{cases}$$

1.4. ANÁLISE DE VARIÂNCIA MULTIVARIADA E TESTES DE HIPÓTESES

Modelo estatístico:

O modelo estatístico para um delineamento em blocos completos casualizados com **b** blocos e **k** tratamentos, em que são medidas **p** variáveis, é dado a seguir:

$$y_{ijr} = \mu_r + t_{ir} + b_{jr} + e_{ijr}$$

em que:

$$i = 1, 2, \dots, k$$

$$j = 1, 2, \dots, b$$

$$r = 1, 2, \dots, p$$

$$\sum_{i=1}^k t_{ir} = 0 \text{ para todo } r$$

Na forma matricial, esse modelo pode ser assim escrito:

$$Y = XB + \varepsilon$$

$$Y = \begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11p} \\ y_{121} & y_{122} & \cdots & y_{12p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{1b1} & y_{1b2} & \cdots & y_{1bp} \\ \hline y_{211} & y_{212} & \cdots & y_{21p} \\ y_{221} & y_{222} & \cdots & y_{22p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{2b1} & y_{2b2} & \cdots & y_{2bp} \\ \hline \cdots & \cdots & \cdots & \cdots \\ \hline y_{k11} & y_{k12} & \cdots & y_{k1p} \\ y_{k21} & y_{k22} & \cdots & y_{k2p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{kb1} & y_{kb2} & \cdots & y_{kbp} \end{bmatrix} = \begin{bmatrix} \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_p \end{bmatrix}$$

$$X = \begin{bmatrix}
 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\
 \dots & \dots \\
 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\
 \hline
 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\
 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\
 \dots & \dots \\
 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\
 \hline
 \dots & \dots \\
 \hline
 1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\
 1 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\
 \dots & \dots \\
 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1
 \end{bmatrix}$$

$$B = \begin{bmatrix}
 \mu_1 & \mu_2 & \dots & \mu_p \\
 \hline
 t_{11} & t_{12} & \dots & t_{1p} \\
 t_{21} & t_{22} & \dots & t_{2p} \\
 \dots & \dots & \dots & \dots \\
 t_{k1} & t_{k2} & \dots & t_{kp} \\
 \hline
 b_{11} & b_{12} & \dots & b_{1p} \\
 b_{21} & b_{22} & \dots & b_{2p} \\
 \dots & \dots & \dots & \dots \\
 b_{b1} & b_{b2} & \dots & b_{bp}
 \end{bmatrix} = \begin{bmatrix}
 \beta_{\sim 1} & \beta_{\sim 2} & \dots & \beta_{\sim p}
 \end{bmatrix}$$

$$\begin{aligned}
\varepsilon = & \begin{bmatrix}
e_{111} & e_{112} & \cdots & e_{11p} \\
e_{121} & e_{122} & \cdots & e_{12p} \\
\cdots & \cdots & \cdots & \cdots \\
e_{1b1} & e_{1b2} & \cdots & e_{1bp} \\
\hline
e_{211} & e_{212} & \cdots & e_{21p} \\
e_{221} & e_{222} & \cdots & e_{22p} \\
\cdots & \cdots & \cdots & \cdots \\
e_{2b1} & e_{2b2} & \cdots & e_{2bp} \\
\hline
\cdots & \cdots & \cdots & \cdots \\
\hline
e_{k11} & e_{k12} & \cdots & e_{k1p} \\
e_{k21} & e_{k22} & \cdots & e_{k2p} \\
\cdots & \cdots & \cdots & \cdots \\
e_{kb1} & e_{kb2} & \cdots & e_{kbp}
\end{bmatrix}
= \begin{bmatrix}
\tilde{e}_{11} \\
\tilde{e}_{12} \\
\cdots \\
\tilde{e}_{1b} \\
\hline
\tilde{e}_{21} \\
\tilde{e}_{22} \\
\cdots \\
\tilde{e}_{2b} \\
\hline
\cdots \\
\hline
\tilde{e}_{k1} \\
\tilde{e}_{k2} \\
\cdots \\
\tilde{e}_{kb}
\end{bmatrix}
= \begin{bmatrix}
\tilde{e}_1 & \tilde{e}_2 & \cdots & \tilde{e}_p
\end{bmatrix}
\end{aligned}$$

Matrizes de Sistemas de Equações Normais

$$X'XB^0 = X'Y$$

OU

$$X'X \begin{bmatrix} \beta_{\tilde{1}}^0 & \beta_{\tilde{2}}^0 & \cdots & \beta_{\tilde{p}}^0 \end{bmatrix} = X' \begin{bmatrix} y_{\tilde{1}} & y_{\tilde{2}} & \cdots & y_{\tilde{p}} \end{bmatrix} \text{ em que:}$$

$$B^0 = (X'X)^{-1} X'Y = \begin{bmatrix} \beta_{\tilde{1}}^0 & \beta_{\tilde{2}}^0 & \cdots & \beta_{\tilde{p}}^0 \end{bmatrix} \quad (\text{solução geral})$$

$$\beta_{\tilde{r}}^{0'} = \begin{bmatrix} \mu_r^0 & t_{1r}^0 & \cdots & t_{kr}^0 & b_{1r}^0 & \cdots & b_{br}^0 \end{bmatrix};$$

$$X'Y = \begin{bmatrix} G_1 & G_2 & \cdots & G_p \\ \hline T_{11} & T_{12} & \cdots & T_{1p} \\ T_{21} & T_{22} & \cdots & T_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ T_{k1} & T_{k2} & \cdots & T_{kp} \\ \hline B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ B_{b1} & B_{b2} & \cdots & B_{bp} \end{bmatrix} = \begin{bmatrix} X'y_{\tilde{1}} & X'y_{\tilde{2}} & \cdots & X'y_{\tilde{p}} \end{bmatrix}; e$$

$$G_r = \sum_{i=1}^k \sum_{j=1}^b y_{ijr}, \quad T_{ir} = \sum_{j=1}^b y_{ijr} \quad e \quad B_{jr} = \sum_{i=1}^k y_{ijr} .$$

Decomposição das Somas de Quadrados e de Produtos Totais

Uma vez que $Y = XB + \varepsilon$, pelo método dos mínimos quadrados, ter-se-á:

$$\hat{\varepsilon}'\hat{\varepsilon} = Y'Y - \hat{B}'X'Y$$

isto é, uma matriz de somas de quadrados e de produtos residuais, que será denotada por **E**, com $n_e = n - \text{posto}(X)$ graus de liberdade. No modelo em questão, $n_e = kb - (k+b-1) = kb - k - b + 1 = (k - 1)(b - 1)$.

$$\hat{B}'X'Y = \begin{bmatrix} \hat{\beta}_1' \\ \hat{\beta}_2' \\ \dots \\ \hat{\beta}_p' \end{bmatrix} \begin{bmatrix} X'y_1 & X'y_2 & \dots & X'y_p \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\beta}_1' X'y_1 & \hat{\beta}_1' X'y_2 & \dots & \hat{\beta}_1' X'y_p \\ \hat{\beta}_2' X'y_1 & \hat{\beta}_2' X'y_2 & \dots & \hat{\beta}_2' X'y_p \\ \dots & \dots & \dots & \dots \\ \hat{\beta}_p' X'y_1 & \hat{\beta}_p' X'y_2 & \dots & \hat{\beta}_p' X'y_p \end{bmatrix}$$

em que: $\hat{B}'X'Y$ é uma matriz de somas de quadrados e de produtos de parâmetros com $\text{posto}(X)$ graus de liberdade.

De forma semelhante ao modelo univariado, as somas de quadrados são dadas por:

$$\text{SQTotal}_r = \text{SQBlocos}_r + \text{SQTratamentos}_r + \text{SQResíduo}_r, \text{ em que:}$$

$$\text{SQTotal}_r = \sum_{i=1}^k \sum_{j=1}^b y_{ijr}^2 - \frac{G_r^2}{kb},$$

$$\text{SQBlocos}_r = \frac{1}{k} \sum_{j=1}^b B_{jr}^2 - \frac{G_r^2}{kb},$$

$$\text{SQTratamentos}_r = \frac{1}{b} \sum_{i=1}^k T_{ir}^2 - \frac{G_r^2}{kb}.$$

As somas de produtos são dadas por:

$$SPTotal_{r,s} = SPBlocos_{r,s} + SPTratamentos_{r,s} + SPResíduo_{r,s},$$

$$SPTotal_{r,s} = \sum_{i=1}^k \sum_{j=1}^b y_{ijr} y_{ijs} - \frac{G_r G_s}{kb}$$

$$SPBlocos_{r,s} = \frac{1}{k} \sum_{j=1}^b B_{jr} B_{js} - \frac{G_r G_s}{kb}$$

$$SPTratamentos_{r,s} = \frac{1}{b} \sum_{i=1}^k T_{ir} T_{is} - \frac{G_r G_s}{kb}$$

logo, $\mathbf{A} = \mathbf{B} + \mathbf{H} + \mathbf{E}$, em que \mathbf{A} , \mathbf{B} , \mathbf{H} e \mathbf{E} são matrizes $p \times p$, de somas de quadrados e de produtos, sendo \mathbf{A} referente a totais, \mathbf{B} a blocos, \mathbf{H} a tratamentos e \mathbf{E} ao resíduo.

Na **MANOVA**, a hipótese de real interesse a ser testada é a de igualdade dos vetores de médias de tratamentos, isto é, $H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_k$, ou ainda:

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \dots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \dots \\ \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{k1} \\ \mu_{k2} \\ \dots \\ \mu_{kp} \end{bmatrix} \quad \text{ou}$$

$$H_0 : \begin{bmatrix} t_{11} \\ t_{12} \\ \dots \\ t_{1p} \end{bmatrix} = \begin{bmatrix} t_{21} \\ t_{22} \\ \dots \\ t_{2p} \end{bmatrix} = \dots = \begin{bmatrix} t_{k1} \\ t_{k2} \\ \dots \\ t_{kp} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad \text{(modelo restrito)}$$

No modelo linear geral multivariado, a hipótese H_0 pode ser expressa da seguinte maneira:

$$H_0 : C'BW = \phi \quad (C' \text{ é a L do SAS})$$

Uma vez que na MANOVA, para testar $H_0 : \underline{t}_1 = \underline{t}_2 = \dots = \underline{t}_k = \underline{0}$, W é uma matriz identidade de ordem \mathbf{p} , pode-se escrever:

$$H_0 : C'B = \phi$$

C' é uma matriz de posto linha completo, constituída de funções estimáveis.

Uma, delas de ordem $(k-1) \times (1+k+b)$, pode ser:

$$C' = \left[\begin{array}{c|cccc|cccc} 0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots \\ 0 & 1 & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 0 \end{array} \right]$$

A matriz de somas de quadrados e de produtos devida à hipótese testada é a matriz H de tratamentos, que matricialmente é dada por:

$$B^0 = (X'X)^{-1} X'Y \quad \rightarrow \quad H = [C'B^0]' [C'(X'X)^{-1}C]^{-1} [C'B^0]$$

A matriz **A** de somas de quadrados e de produtos totais com $n-1$ graus de liberdade, é dada por:

$$A = Y' \left[I_{(n)} - \frac{1}{n} \mathbf{1} \mathbf{1}' \right] Y$$

$$A = Y'Y - \frac{1}{n} Y' \mathbf{1} \mathbf{1}' Y, \quad \mathbf{1} \text{ é um vetor de } 1\text{'s } n \times 1$$

A matriz **E** de somas de quadrados e produtos residual com $n - \text{posto}[X] = n_e$ graus de liberdade, é dada por:

$$E = Y' \left[I_{(n)} - X(X'X)^{-1} X' \right] Y = Y'Y - \hat{B}' X' Y$$

Tabela 1.12 - Esquema da análise de variância multivariada para um experimento com um fator em blocos completos casualizados

Causas de variação	G.L.*	Matrizes de somas de quadrados e de produtos
Blocos	$b - 1$	B
Tratamentos	$k - 1$	H
Resíduo	$n_e = (b-1)(k-1)$	E
Total	$bk-1$	A

* $n = bk$ e $n_e = n - \text{posto}(X)$.

Modelo Estatístico (DIC):

$$y_{ijl} = \mu_l + t_{ij} + e_{ijl}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, r; \quad l = 1, 2, \dots, p$$

Tabela 1.13 - Esquema da análise de variância multivariada para um experimento com um fator inteiramente casualizado

Causas de variação	G.L.*	Matrizes de somas de quadrados e produtos
Tratamentos	$k - 1$	H
Resíduo	$n_e = n - k$	E
Total	$n - 1$	A

* $n = kr$ e $n_e = n - \text{posto}(X)$.

A seguir, são apresentados quatro testes de hipóteses:

- i) **Lambda de Wilks** (Teste de Wilks);
- ii) **Traço de Pillai** (Teste de Pillai);
- iii) **Traço de Hotelling-Lawley** (Teste de Hotelling-Lawley); e
- iv) **Raiz máxima de Roy** (Teste de Roy).

Testes de Hipóteses:

$$H_0 : C'BW = \phi$$

com C' de posto linha completo, e ainda:

$p = \text{posto}(W)$ – é importante ressaltar que p aqui não é necessariamente igual ao número de variáveis.

$$q = \text{posto}(C')$$

$$w = \max [\text{posto}(W), \text{posto}(C')]$$

$$H = [C'B^0W]' [C'(X'X)^- C]^{-1} [C'B^0W]$$

$$H = W' [C'B^0]' [C'(X'X)^- C]^{-1} [C'B^0] W$$

$$E = W' [Y'Y - B^0'X'Y] W$$

Define-se ainda:

$$s = \min [\text{posto}(W), \text{posto}(C')] = \min(p, q)$$

$$m' = \frac{1}{2} (|p - q| - 1)$$

$$n' = \frac{1}{2} (n_e - p - 1)$$

Teste de Wilks

A estatística do teste é:

$$\Lambda = \frac{\det(\mathbf{E})}{\det(\mathbf{H} + \mathbf{E})} = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|}$$

O valor de Λ tabelado para o teste de Wilks é função de α , p , q e n_e .

$$\Lambda_{\text{Tabelado}} \rightarrow \Lambda_{(\alpha; p; q; n_e)}$$

Regra decisória: rejeita-se H_0 em nível de significância α se $\Lambda_{\text{calculado}} < \Lambda_{\text{Tabelado}}$. Caso contrário, não se rejeita H_0 .

A transformação de Λ em F se faz por meio das fórmulas a seguir:

Caso A: $p=2$, q e n_e quaisquer:

$$F_0 = \frac{n_e - 1}{q} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F[2q; 2(n_e - 1)]$$

Caso B: $q=1$, p e n_e quaisquer:

$$F_0 = \frac{n_e - p + 1}{p} \cdot \frac{1 - \Lambda}{\Lambda} \sim F[p; n_e - p + 1]$$

Caso C: $q=2$, p e n_e quaisquer:

$$F_0 = \frac{n_e - p + 1}{p} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F[2p; 2(n_e - p + 1)]$$

Caso geral (HARRIS, 1975):

$$F_0 = \left[\left(\frac{1}{\Lambda} \right)^{1/b} - 1 \right] \frac{ab - c}{pq} \sim F(pq; ab - c)$$

ou

$$F_0 = \left[\frac{1 - \Lambda^{1/b}}{\Lambda^{1/b}} \right] \frac{ab - c}{pq} \sim F(pq; ab - c)$$

em que:

$$a = n_e - \frac{1}{2}(p - q + 1)$$

$$c = \frac{1}{2}(pq - 2)$$

$$b = \sqrt{\frac{p^2 q^2 - 4}{p^2 + q^2 - 5}}, \text{ se } (p^2 + q^2 - 5) > 0, \text{ ou } 1, \text{ caso contrário.}$$

A distribuição é exata se $\min(p, q) \leq 2$ (RAO, 1973, p. 556).

Teste de Pillai

Define-se a estatística:

$$V = \text{traço} \left[H(H + E)^{-1} \right]$$

V se relaciona com a distribuição F , aproximadamente, por:

$$F_0 = \frac{2n'+s+1}{2m'+s+1} \cdot \frac{V}{s-V} \sim F(n_1; n_2),$$

em que $n_1 = s(2m'+s+1)$ e $n_2 = s(2n'+s+1)$.

Regra decisória: rejeita-se H_0 em nível de significância α se $F_0 > F_\alpha(n_1; n_2)$.

Teste de Hotelling-Lawley

Define-se a estatística:

$$U = \text{traço}(E^{-1}H)$$

U se relaciona com a distribuição F, aproximadamente por (RENCHER, 2002):

1º caso:

$$F_1 = \frac{2(sn'+1)}{s^2(2m'+s+1)} \cdot U \sim F[s(2m'+s+1), 2(sn'+1)]$$

2º caso:

$$F_2 = \frac{U}{c} \sim F(a, b) ,$$

em que:

$$a = pq \quad b = 4 + \frac{a+2}{D-1} \quad c = \frac{a(b-2)}{b(n_e - p - 1)} \quad D = \frac{(n_e + q - p - 1)(n_e - 1)}{(n_e - p - 3)(n_e - p)}$$

3º caso:

$$F_3 = \frac{[s(n_e - q - 1) + 1]}{spq} \cdot U \sim F[pq, s(n_e - q - 1)]$$

Fato: Versões mais antigas do SAS usavam o 1º caso. A versão 9.1 usa o 2º caso para $n' > 0$ e o 1º caso para $n' \leq 0$.

Regra decisória: rejeita-se H_0 em nível de significância α se $F_{\text{calculado}} > F_{\text{Tabelado}}$. Caso contrário, não se rejeita H_0 .

Teste de Roy

Para o teste de Roy, inicialmente determinam-se os autovalores da matriz $E^{-1}H$, isto é:

$$|E^{-1}H - \lambda I| = 0 \quad \text{ou} \quad |H - \lambda E| = 0$$

Sendo λ_{\max} o maior autovalor de $E^{-1}H$, define-se a estatística: $\theta_0 = \frac{\lambda_{\max}}{1 + \lambda_{\max}}$

λ_{\max} se relaciona com a distribuição F , aproximadamente, por:

$$F_0 = \lambda_{\max} \cdot \frac{(n_e - w + q)}{w} \sim F(w; n_e - w + q)$$

Regra decisória: rejeita-se H_0 em nível de significância α se $F_0 > F_\alpha(w; n_e - w + q)$.

Tabelas incluídas nos livros de Harris (1975) e Morrison (1976), entre outros.

$$\theta_{\text{Tabelado}} \rightarrow \theta_{(\alpha; s; m'; n')}$$

Regra decisória: rejeita-se H_0 em nível de significância α se $\theta_0 > \theta_{\alpha}(s; m'; n')$. Caso contrário, não se rejeita H_0 .

→ Para todos os quatro testes apresentados, é preciso ter $n_e \geq p$. Quando $s = \min[\text{posto}(W), \text{posto}(C')] = 1$, todos os quatro testes estatísticos dão resultados equivalentes.

2ª PARTE

II – ANÁLISE DE DADOS COM MEDIDAS REPETIDAS

2.1. Definições

→ **MEDIDAS REPETIDAS:** O termo medidas repetidas é usado para designar medidas feitas ou na mesma unidade experimental ou mesmo indivíduo em mais de uma ocasião (DIGGLE, 1988; CROWDER; HAND, 1990).

Observações dentro das unidades de investigação não foram ou não podem ter sido atribuídas aleatoriamente para as diferentes condições de avaliação (tempo: medida semanalmente, mensalmente, etc.; espaço: medida em várias faixas de profundidade do solo).

2.2. Objetivos

(i) comparação das diferentes subpopulações quanto ao padrão de variação das respectivas distribuições de respostas ao longo das diferentes condições de avaliação, isto é, verificar se os **perfis são paralelos**.

$H_0^{(1)}$: a interação entre tratamentos e tempo é nula.

(ii) comparação das diferentes subpopulações quanto às respectivas distribuições médias (em relação às diversas condições de avaliação) de respostas, isto é, verificar se os **perfis são coincidentes**.

$H_0^{(2)}$: dado que os perfis são paralelos, o efeito de tratamento é nulo.

Uma forma alternativa para a hipótese $H_0^{(2)}$ que independe do conhecimento a priori da validade de $H_0^{(1)}$ é dada por:

$H_0^{(2)*}$: Igualdade dos vetores de médias de tratamentos.

(iii) comparação das diferentes condições de avaliação quanto às respectivas distribuições médias (em relação às diversas subpopulações) de respostas, isto é, verificar se os **perfis são horizontais**.

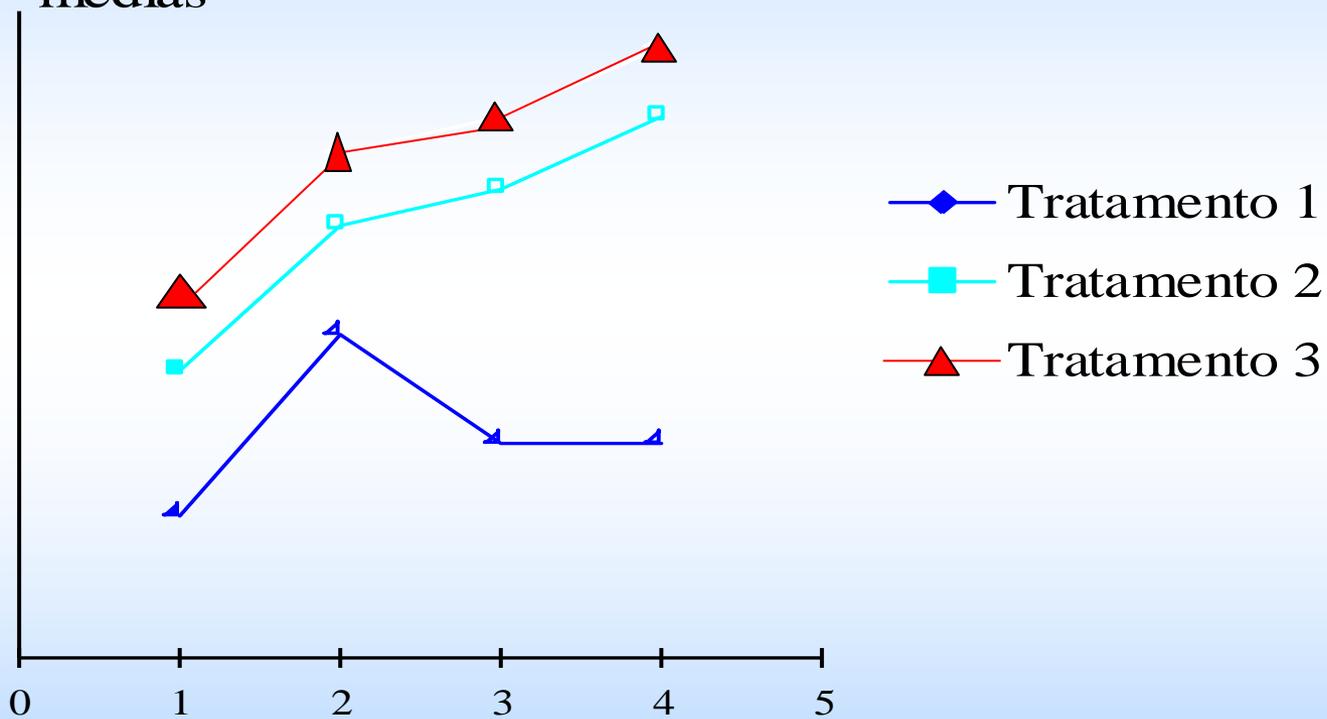
$H_0^{(3)}$: dado que os perfis são paralelos, o efeito do tempo é nulo.

Uma forma alternativa para a hipótese $H_0^{(3)}$ que independe do conhecimento a priori da validade de $H_0^{(1)}$ é dada por:

$H_0^{(3)*}$: Igualdade dos vetores de médias de tempos.

(iv) ajuste de modelos para explicar a variação das respostas médias como função do fator que define as condições de avaliação.

Respostas
médias



Tratamento 1

Tratamento 2

Tratamento 3

Condições de avaliação

2.3. Métodos de Análise Estatística

- (i) Análise de variância univariada como se fosse um experimento em parcelas subdivididas (Split Plot), com tratamentos como o fator da parcela e o tempo como o fator da subparcela.
- (ii) Análises univariada e multivariada para transformações lineares dos dados, tais como médias, diferenças entre respostas de diferentes pontos de tempo, inclinações e outras tendências em curvas de regressão.
- (iii) Metodologia de Modelos Mistos com estruturas paramétricas especiais para as matrizes de covariâncias.

Dos 3 métodos considerados:

(i) A análise de variância univariada ignora a estrutura de covariância o que pode resultar em conclusões incorretas.

(ii) Análises univariada e multivariada para transformações lineares dos dados, evitam a estrutura de covariância o que pode resultar em análises ineficientes, equivalente a desperdícios dos dados.

(iii) Metodologia de Modelos Mistos que permite eficientemente conduzir a questão diretamente na modelagem da estrutura de covariância.

2.3.1. ANÁLISE UNIVARIADA DE PERFIS

(a) Parcelas Subdivididas

Modelo de experimentos em parcelas subdivididas (Split-Plot) e DIC:

Consideram-se tratamentos e condições de avaliação como dois fatores cruzados fixos e as unidades experimentais como um fator aleatório hierárquico a tratamentos.

$$y_{ijk} = \mu + \alpha_i + \delta_{j(i)} + \beta_k + \alpha\beta_{ik} + \varepsilon_{ijk}$$

$$i = 1, 2, \dots, g; j = 1, 2, \dots, n_i; k = 1, 2, \dots, p$$

$$\sum_{i=1}^g n_i = n$$

- $\mu \rightarrow$ constante comum a todas as observações;
- $\alpha_i \rightarrow$ representa o efeito do i -ésimo tratamento;
- $\beta_k \rightarrow$ representa o efeito da k -ésima condição de avaliação (tempo);
- $\alpha\beta_{ik} \rightarrow$ representa o efeito de interação entre o i -ésimo tratamento e a k -ésima condição de avaliação;
- $\delta_{j(i)} \rightarrow$ representa o efeito aleatório da j -ésima unidade experimental dentro do i -ésimo tratamento;
- $\varepsilon_{ijk} \rightarrow$ representa o erro aleatório da observação y_{ijk}

$\delta_{j(i)} \sim N(0, \sigma_{\delta}^2)$ representa a variação entre sujeitos ("subjects");

$\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon}^2)$ representa a variação dentro de sujeitos;

$$E(y_{ijk}) = \mu + \alpha_i + \beta_k + \alpha\beta_{ik}$$

$$\text{Var}(y_{ijk}) = \text{Var}(\delta_{j(i)} + \varepsilon_{ijk}) = \sigma_{\delta}^2 + \sigma_{\varepsilon}^2 = \sigma^2;$$

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_{\delta}^2, \text{ para } k \neq k'.$$

A correlação entre quaisquer duas medidas no mesmo sujeito é igual a $\rho = \sigma_{\delta}^2 / (\sigma_{\delta}^2 + \sigma_{\varepsilon}^2) \Rightarrow \sigma_{\delta}^2 = \rho\sigma^2$.

$\text{Var}(\tilde{Y}) = V$. Esta matriz é bloco diagonal, com cada bloco correspondendo a um sujeito. Cada bloco é uma matriz Σ de ordem $p \times p$ com σ^2 na diagonal e $\rho\sigma^2$ fora da diagonal o que corresponde a estrutura de covariância denominada Simetria Composta (CS).

Tabela 2.1 - ANOVA para o Modelo de Parcelas Subdivididas no Tempo

Fontes de Variação	GL	SQ	QM	F*
Tratamentos	$g - 1$	SQ_1	QM_1	F_1
Resíduo (a)	$n - g$	SQ_2	QM_2	
Tempos	$p - 1$	SQ_3	QM_3	F_2
TratamentosxTempos	$(g - 1)(p - 1)$	SQ_4	QM_4	F_3
Resíduo (b)	$(n - g)(p - 1)$	SQ_5	QM_5	
Total	$np - 1$	SQ_6		

* $F_1=QM_1/QM_2$; $F_2=QM_3/QM_5$; $F_3=QM_4/QM_5$

Temos n unidades de investigação, sendo cada uma medida p vezes.

$N = np$ é o número total de observações.

- **Estrutura de Covariância Uniforme**

$$\Sigma_{(p \times p)} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ & 1 & \rho & \cdots & \rho \\ & & 1 & \cdots & \rho \\ & & & \ddots & \vdots \\ \text{sim.} & & & & 1 \end{bmatrix}$$

onde $\sigma^2 = \sigma_{\delta}^2 + \sigma_{\varepsilon}^2$ é uma variância das respostas em qualquer um dos tempos e a covariância entre dois tempos quaisquer é σ_{δ}^2 .

- Todos os pares de observações sobre a mesma unidade experimental têm igual correlação:

$$\rho = \sigma_{\delta}^2 / (\sigma_{\delta}^2 + \sigma_{\varepsilon}^2)$$

- A matriz Σ com estrutura de covariância uniforme é chamada de Simetria Composta (“Compound Symmetry”).
- No PROC MIXED dos SAS, o modelo com esta estrutura é especificado como TYPE = CS. Condição suficiente para que F_2 e F_3 da Tabela 2.1 tenham distribuição F exata.
- Estrutura extremamente restritiva e muitas vezes não realista.
- Huynh e Feldt (1970) → Condição necessária e suficiente para a validade dos testes F_2 e F_3 da Tabela 2.1 sejam válidos: **condição H-F**.

Segundo Huynh e Feldt (1970) uma condição necessária e suficiente para que os testes F (F_2 e F_3 da Tabela 2.1.) sejam exatos é que a matriz Σ satisfaça a condição de esfericidade, ou seja:

$$\sigma_{ij} = \frac{(\sigma_i^2 + \sigma_j^2)}{2} - \lambda, \text{ para todo } i \neq j, \text{ e } \lambda \text{ uma constante positiva.}$$

Por muitos anos, a condição de simetria composta foi requerida para garantir a validade da estatística F em ANOVA de medidas repetidas.

Teste de Esfericidade: Mauchly(1940)

- Testa se a matriz de covariâncias tem uma estrutura do tipo H-F, ou seja, se ela satisfaz à condição H-F.
- Pressuposição não atendida → correção para os números de graus de liberdade de F_2 e F_3 .

(b) Parcelas subdivididas com correções para os números de graus de liberdade

F_2 e F_3 não têm distribuição F Central Exata

☞ Correção G-G (Greenhouse & Geisser, 1959)

Estimador: $\hat{\epsilon}_{GG}$

☞ Correção H-F (Huynh & Feldt, 1976)

Estimador: $\tilde{\epsilon}_{HF}$

2.3.2. ANÁLISE MULTIVARIADA DE PERFIS

- Modelo na Forma Matricial é $Y = X\beta + \varepsilon$

onde

$$\underset{(n \times p)}{Y} = \underbrace{\begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11p} \\ y_{121} & y_{122} & \cdots & y_{12p} \\ \vdots & \vdots & & \vdots \\ y_{1n_1 1} & y_{1n_1 2} & \cdots & y_{1n_1 p} \\ \vdots & \vdots & & \vdots \\ y_{g11} & y_{g12} & \cdots & y_{g1p} \\ \vdots & \vdots & & \vdots \\ y_{gn_g 1} & y_{gn_g 2} & \cdots & y_{gn_g p} \end{bmatrix}}_{\text{matriz dos dados}} = \underbrace{\begin{bmatrix} y'_{\sim 11} \\ y'_{\sim 12} \\ \vdots \\ y'_{\sim 1n_1} \\ \vdots \\ y'_{\sim g1} \\ \vdots \\ y'_{\sim gn_g} \end{bmatrix}}_{\text{perfil de resposta da unidade experimental (ij)}} ; y'_{\sim 11} \text{ é um vetor}$$

$$X_{(n \times g)} = \begin{bmatrix} \mathbf{1}_{\sim n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{\sim n_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{\sim n_g} \end{bmatrix}, \quad \mathbf{1}_{\sim n_i} \text{ é um vetor de 1's com } i=1, 2, \dots, g$$

$X_{(n \times g)}$ é uma matriz de especificação (planejamento) e

$$\beta_{(g \times p)} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & & \vdots \\ \mu_{g1} & \mu_{g1} & \cdots & \mu_{gp} \end{bmatrix} = \begin{bmatrix} \mu'_{\sim 1} \\ \mu'_{\sim 2} \\ \vdots \\ \mu'_{\sim g} \end{bmatrix} \quad \mu'_{\sim i} \text{ é um vetor}$$

- $\beta_{(g \times p)}$ é a matriz de parâmetros e μ_{ik} representa a média das unidades submetidas ao i -ésimo tratamento no h -ésimo tempo e $\mu'_{\sim i}$ representa o perfil médio de respostas do i -ésimo tratamento e $\varepsilon_{(n \times p)} = \begin{bmatrix} \varepsilon_{\sim 11} & \varepsilon_{\sim 12} & \cdots & \varepsilon_{\sim gn_g} \end{bmatrix}'$ é

é a matriz de erros onde $\varepsilon'_{\sim ij} = [\varepsilon_{ij1} \quad \varepsilon_{ij2} \quad \cdots \quad \varepsilon_{ijp}]$.

(1xp)

Para efeito de inferência, na Análise Multivariada de Perfis, supomos que os perfis de resposta $Y_{\sim ij}$ obedecem as distribuições normais p-variadas e que as matrizes de covariâncias correspondentes são todas iguais e seguem a forma geral a seguir:

$$\Sigma_{(p \times p)} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ & \sigma_2^2 & \cdots & \sigma_{2p} \\ & & \ddots & \vdots \\ \text{sim.} & & & \sigma_p^2 \end{bmatrix}$$

A eficiência dos estimadores pode ser afetada em função do elevado n° de parâmetros e de observações incompletas.

Em termos dos parâmetros do modelo, as hipóteses podem ser expressas por:

(a) Perfis Paralelos (Não existe interação entre tratamentos e tempos)

$$H_0^{(1)} : \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1(p-1)} - \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} - \mu_{22} \\ \mu_{22} - \mu_{23} \\ \vdots \\ \mu_{2(p-1)} - \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} - \mu_{g2} \\ \mu_{g2} - \mu_{g3} \\ \vdots \\ \mu_{g(p-1)} - \mu_{gp} \end{bmatrix}$$

Na forma da hipótese linear geral $H_0^{(1)} : C_1' \beta W_1 = \phi$, tem-se:

$$C_1' = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

É de ordem $[(g-1) \times g]$ e posto = $g-1$.

$$W_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{bmatrix}$$

É de ordem $p \times (p-1)$ e posto = $p-1$.

(b) Perfis Coincidentes (Não existe efeito de tratamentos)

$$H_0^{(2)} : \bar{\mu}_1 = \bar{\mu}_2 = \dots = \bar{\mu}_g$$

$$\text{ou ainda } H_0^{(2)} : \sum_{k=1}^p \mu_{1k} = \sum_{k=1}^p \mu_{2k} = \dots = \sum_{k=1}^p \mu_{gk}$$

Na forma da hipótese linear geral , $H_0^{(2)} : C_2' \beta W_2 = \phi$, tem-se:

$$C_2' = C_1' \text{ com posto } (C_2') = \text{posto } (C_1') = g - 1;$$

$W_2 = \underline{1}_p$, em que $\underline{1}_p$ é um vetor coluna de uns de dimensão p, sendo p o número de medidas (p tempos); e $\text{posto } (W_2) = 1$.

Este teste é idêntico ao teste F para tratamentos da parcela, considerando o esquema de parcelas subdivididas. Pelo programa SAS (2004), essa hipótese é testada a partir de uma análise de variância univariada, na qual se utiliza a variável:

$$Y_{ij}^* = \frac{1}{\sqrt{p}} (Y_{ij1} + Y_{ij2} + \dots + Y_{ijp}) = \frac{1}{\sqrt{p}} Y_{ij}.$$

em que p é o número de medidas.

Uma forma alternativa para a hipótese $H_0^{(2)}$ que independe do conhecimento a priori da validade de $H_0^{(1)}$ é dada por $H_0^{(2)*}$.

Agora, $H_0^{(2)*}$ é a hipótese de igualdade dos vetores de médias de tratamentos (como na MANOVA), isto é,

$$H_0^{(2)*} = H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_g$$

ou ainda:

$$H_0^{(2)*} : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \dots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \dots \\ \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{g1} \\ \mu_{g2} \\ \dots \\ \mu_{gp} \end{bmatrix} .$$

$$H_0^{(2)*} : C' \beta W = \phi;$$

$c' = c'_1$, como definido em $H_0^{(1)}$;

$w = I_p$, I_p é uma matriz identidade de ordem p ;
 $\text{posto}(W) = p$ (neste caso, p coincide com o número de medidas). Logo, $H_0^{(2)*} : c' \beta W = c' \beta = \phi$.

(c) Perfis Horizontais (Não existe efeito de tempos)

$$H_0^{(3)} : \bar{\mu}_{.1} = \bar{\mu}_{.2} = \cdots = \bar{\mu}_{.p}$$

$$\text{ou ainda } H_0^{(3)} : \sum_{i=1}^g \mu_{i1} = \sum_{i=1}^g \mu_{i2} = \cdots = \sum_{i=1}^g \mu_{ip}$$

Na forma da hipótese linear geral

$$H_0^{(3)} : C_3' \beta W_3 = \phi, \text{ tem-se:}$$

$C_3' = 1_g'$ é um vetor linha de uns de dimensão g (g tratamentos) com posto $(C_3') = 1$;

$W_3 = W_1$, com posto $(W_3) = \text{posto}(W_1) = p-1$.

Uma forma alternativa para a hipótese $H_0^{(3)}$ que independe do conhecimento a priori da validade de $H_0^{(1)}$ é dada por $H_0^{(3)*}$.

Agora, $H_0^{(3)*}$ é a hipótese de igualdade dos vetores de médias de tempos (medidas), isto é, $H_0^{(3)*} = H_0 : \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \dots = \underset{\sim}{\mu}_p$, ou ainda:

$$H_0^{(3)*} : \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \dots \\ \mu_{g1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \dots \\ \mu_{g2} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{1p} \\ \mu_{2p} \\ \dots \\ \mu_{gp} \end{bmatrix}.$$

$$H_0^{(3)*} : C' \beta W = \phi;$$

$C' = I_g$ é uma matriz identidade de ordem g (g tratamentos) e posto $(C') = g$;

$W = W_1$, como definido em $H_0^{(1)}$.

Convém notar que as representações das matrizes C e W não são únicas.

Uma abordagem considerando dados provenientes de um delineamento em Blocos Completos Causalizados é apresentada no livro:

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J.
Modelos biométricos aplicados ao melhoramento genético. v.2, 3. ed. rev. e ampl. – Viçosa: Ed. UFV, 2014. 668p.

Testes para a hipótese linear geral: $H_0 : C' \beta W = \phi$.

Em geral as estatísticas de teste correspondente são funções dos autovalores da matriz $E^{-1}H$.

H é a matriz de somas de quadrados e produtos cruzados devido à hipótese nula, dada por:

$$H = [C' \beta^0 W]' [C' (X' X)^{-1} C]^{-1} [C' \beta^0 W] = W' [C' \beta^0]' [C' (X' X)^{-1} C]^{-1} [C' \beta^0] W$$

E é a matriz de somas de quadrados e produtos cruzados devido ao erro, dada por:

$$E = W' [Y' Y - \beta^0' X' Y] W$$

em que $\beta^0 = (X' X)^- X' Y$ é a matriz das soluções de mínimos quadrados de $X' X \beta^0 = X' Y$. $(X' X)^-$ é uma inversa condicional (generalizada) de $X' X$.

Para X de posto coluna completo, virá:

$(X' X)^- = (X' X)^{-1}$ e $\beta^0 = \hat{\beta} = (X' X)^{-1} X' Y$ é o estimador de mínimos quadrados ou de máxima verossimilhança de β .

Critérios de Teste

Seja $\theta_i = \lambda_i (1 + \lambda_i)^{-1}$ e λ_i o i -ésimo autovalor de $E^{-1}H$, e \underline{s} o número de autovalores não nulos.

- Lambda de Wilks (Teste de Wilks)

$$\Lambda = \frac{\det(E)}{\det(H + E)} = \frac{|E|}{|H + E|} = \prod_{i=1}^s (1 - \theta_i) = \prod_{i=1}^s \frac{1}{1 + \lambda_i}$$

- Traço de Pillai (Teste de Pillai)

$$V = \text{traço}[H(H + E)^{-1}] = \sum_{i=1}^s \theta_i = \sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i}$$

- Traço de Hotelling-Lawley (Teste de Hotelling-Lawley)

$$U = \text{traço}[E^{-1}H] = \sum_{i=1}^s \theta_i (1 - \theta_i)^{-1} = \sum_{i=1}^s \lambda_i$$

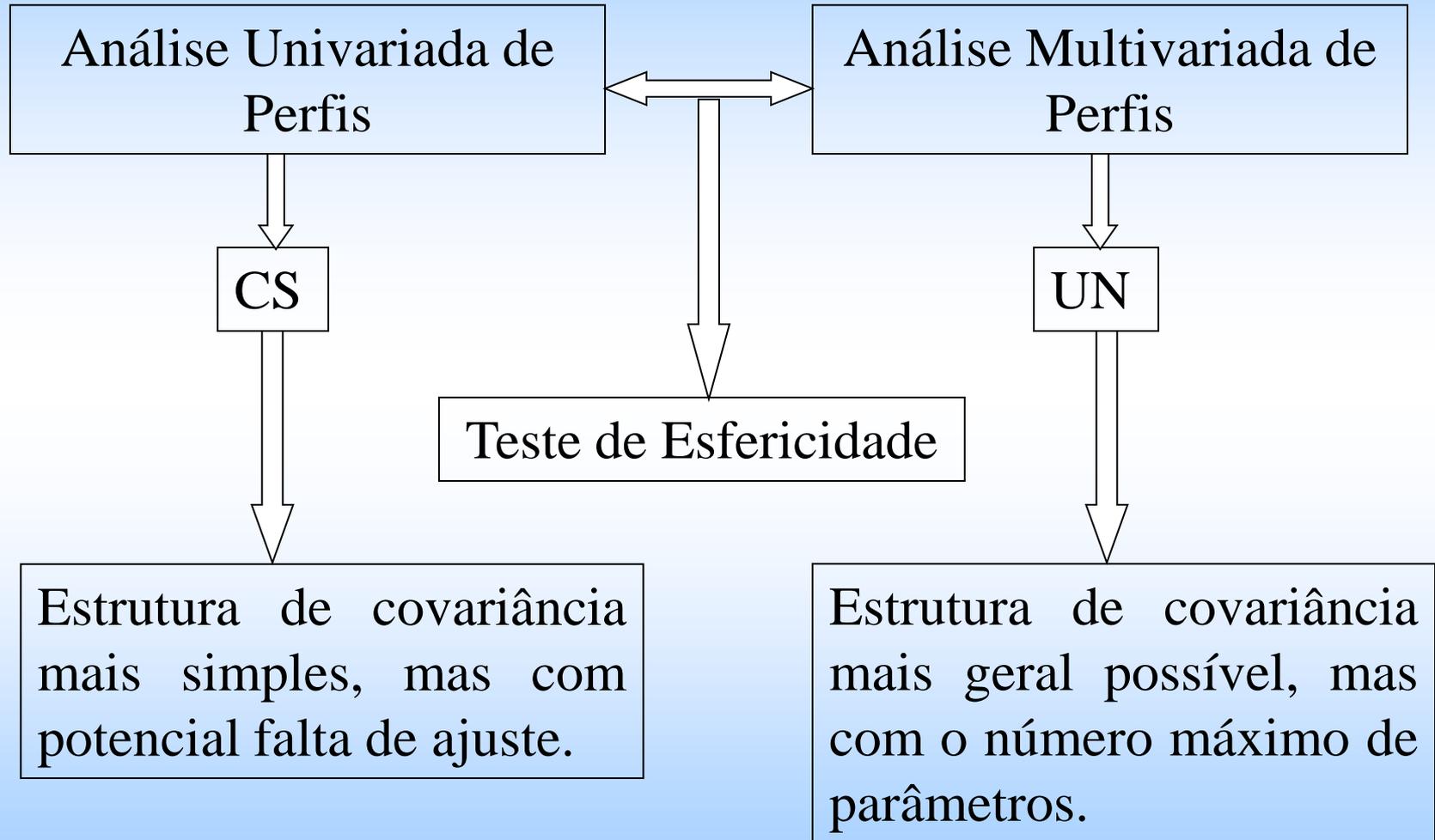
- Raiz Máxima de Roy (Teste de Roy)

Raiz máxima de Roy=maior autovalor de $E^{-1}H = \lambda_{\max}$

A estatística do teste é dada por $\theta_0 = \frac{\lambda_{\max}}{1 + \lambda_{\max}}$.

As estatísticas dos quatro testes se relacionam com a distribuição F aproximadamente.

Dois Extremos



Busca-se geralmente uma estrutura intermediária mais realista e esta estrutura é possível com a metodologia de modelos mistos.

A Metodologia de Modelos Mistos, constitui-se num procedimento bastante difundido para a avaliação de dados de medidas repetidas (dados longitudinais), constituindo uma generalização do modelo linear geral, separando os efeitos fixos dos aleatórios.

2.3.3. Metodologia de Modelos Mistos (MMM)

👉 Modelo estatístico:

$$\underset{\sim}{Y} = \underset{\sim}{X}\beta + \underset{\sim}{Z}\underset{\sim}{v} + \underset{\sim}{\varepsilon}$$

Vetor de Respostas

Matriz de Planejamento

Vetor de erros não observáveis

Vetor de efeitos aleatórios

Vetor de parâmetros de efeitos fixos

onde:

$$E \begin{bmatrix} \underset{\sim}{v} \\ \underset{\sim}{\varepsilon} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} \underset{\sim}{v} \\ \underset{\sim}{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \phi \\ \phi & \mathbf{R} \end{bmatrix}$$

$$E(\underset{\sim}{Y}) = \underset{\sim}{X}\beta$$

$$\text{Var}(\underset{\sim}{Y}) = \mathbf{ZGZ}' + \mathbf{R} = \mathbf{V}$$

$\underset{\sim}{Y}$ → respostas como variáveis aleatórias

$\underset{\sim}{y}$ → valores observados

Assumindo que, $\underset{\sim}{Y}$, $\underset{\sim}{V}$ e $\underset{\sim}{\varepsilon}$ tenham distribuição normal multivariada, vem:

$$\begin{bmatrix} \underset{\sim}{Y} \\ \underset{\sim}{V} \\ \underset{\sim}{\varepsilon} \end{bmatrix} \sim \text{NMV} \left\{ \begin{bmatrix} \underset{\sim}{X\beta} \\ \underset{\sim}{0} \\ \underset{\sim}{0} \end{bmatrix}, \begin{bmatrix} \underset{\sim}{ZGZ'+R} & \underset{\sim}{ZG} & \underset{\sim}{R} \\ \underset{\sim}{GZ'} & \underset{\sim}{G} & \underset{\sim}{\phi} \\ \underset{\sim}{R} & \underset{\sim}{\phi} & \underset{\sim}{R} \end{bmatrix} \right\}$$

• Distribuições Conjunta, Marginal e Condicional

- Distribuição Marginal de $\underline{y} \sim N(\underline{X}\underline{\beta}, \underline{V})$

$$\underline{V} = \underline{Z}\underline{G}\underline{Z}' + \underline{R}$$

$$f(\underline{y}) = \frac{1}{(2\pi)^{n/2} |\underline{V}|^{1/2}} \exp \left\{ -\frac{1}{2} \left[\left(\underline{y} - \underline{X}\underline{\beta} \right)' \underline{V}^{-1} \left(\underline{y} - \underline{X}\underline{\beta} \right) \right] \right\}$$

n é o tamanho da amostra

- Distribuição condicional de $\underset{\sim}{Y} | \underset{\sim}{v} \sim N(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{v}, R)$:

$$f(\underset{\sim}{y} | \underset{\sim}{v}) = \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \cdot \exp \left\{ -\frac{1}{2} \left[\left(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v} \right)' R^{-1} \left(\underset{\sim}{y} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v} \right) \right] \right\}$$

- Distribuição conjunta de \underline{y} e \underline{v} ou \underline{v} e $\underline{\varepsilon}$:

$$L = f(\underline{y}, \underline{v}) = f(\underline{y}/\underline{v}) \cdot f(\underline{v}) \quad \text{Fato: } f(\underline{y}, \underline{v}) = f(\underline{v}, \underline{\varepsilon}) = f(\underline{\varepsilon}/\underline{v}) \cdot f(\underline{v})$$

$$L = \frac{\exp \left\{ -\frac{1}{2} \left[\left(\underline{y} - X\beta - Z\underline{v} \right)' R^{-1} \left(\underline{y} - X\beta - Z\underline{v} \right) + \underline{v}' G^{-1} \underline{v} \right] \right\}}{(2\pi)^{\frac{1}{2}(n+q)} |R|^{1/2} |G|^{1/2}}$$

q . é o número de elementos em \underline{v}

• **Estimando $\underline{\beta}$ e predizendo \underline{v}** : as equações de modelos mistos (EMM) – Henderson (1950, 1984)

Procede-se à maximização de $f(\underline{y}, \underline{\varepsilon})$.

(Observação: log=logaritmo neperiano).

$$\Psi = \log L = -\frac{1}{2}(n + q.)\log(2\pi) - \frac{1}{2}(\log|R| + \log|G|)$$

$$-\frac{1}{2}(\underline{y}'R^{-1}\underline{y} - 2\underline{y}'R^{-1}X\underline{\beta} - 2\underline{y}'R^{-1}Z\underline{v} + 2\underline{\beta}'X'R^{-1}Z\underline{v}$$

$$+ \underline{\beta}'X'R^{-1}X\underline{\beta} + \underline{v}'Z'R^{-1}Z\underline{v} + \underline{v}'G^{-1}\underline{v})$$

$$\begin{bmatrix} \frac{\partial \Psi}{\partial \underline{\beta}} \\ \frac{\partial \Psi}{\partial \underline{v}} \end{bmatrix} = \begin{bmatrix} -X'R^{-1}\underline{y} + X'R^{-1}X\underline{\beta}^0 + X'R^{-1}Z\underline{\hat{v}} \\ -Z'R^{-1}\underline{y} + Z'R^{-1}X\underline{\beta}^0 + Z'R^{-1}Z\underline{\hat{v}} + G^{-1}\underline{\hat{v}} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}$$

Assim, as equações de modelos mistos (EMM) e as soluções são dadas a seguir:

$$\begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{v} \end{bmatrix} = \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix}$$

$$\begin{bmatrix} \beta^0 \\ \hat{v} \end{bmatrix} = \begin{bmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X' R^{-1} y \\ Z' R^{-1} y \end{bmatrix}$$

$$\begin{cases} \beta^0 = (X' V^{-1} X)^{-1} X' V^{-1} y & \text{(Solução GLS)} \\ \hat{v} = G Z' V^{-1} \left(y - X (X' V^{-1} X)^{-1} X' V^{-1} y \right) = G Z' V^{-1} \left(y - X \beta^0 \right) \end{cases}$$

As soluções podem também ser obtidas como a seguir:

$$\underline{\hat{\beta}}^0 = \left\{ \underline{X}' \left[\underline{R}^{-1} - \underline{R}^{-1} \underline{Z} (\underline{Z}' \underline{R}^{-1} \underline{Z} + \underline{G}^{-1})^{-1} \underline{Z}' \underline{R}^{-1} \right] \underline{X} \right\}^{-1} \cdot \underline{X}' \left[\underline{R}^{-1} - \underline{R}^{-1} \underline{Z} (\underline{Z}' \underline{R}^{-1} \underline{Z} + \underline{G}^{-1})^{-1} \underline{Z}' \underline{R}^{-1} \right] \underline{y}$$

$$\underline{\hat{y}} = (\underline{Z}' \underline{R}^{-1} \underline{Z} + \underline{G}^{-1})^{-1} \underline{Z}' \underline{R}^{-1} (\underline{y} - \underline{X} \underline{\hat{\beta}}^0)$$

Se G e R são conhecidas:

► **Funções Estimáveis:** não dependem dos efeitos aleatórios. Se $K' \beta = T' E(\tilde{Y}) = T' X \beta$ para algum T , então $K' \beta$ é estimável.

$K' \beta^0$ é o BLUE ("Best Linear Unbiased Estimator") de $K' \beta$

O "melhor" significa mínimo erro quadrático médio.

► Funções de Predição:

\hat{y} das equações de modelos mistos (EMM) é o BLUP (“Best Linear Unbiased Predictor”) de y .

Combinações lineares de efeitos fixos e aleatórios, $K' \beta + M' y$, com $K' \beta$ estimável, são chamadas funções de predição.

$$K' \beta^0 + M' \hat{y} \text{ é o BLUP de } K' \beta + M' y$$

A matriz de covariâncias de $\begin{bmatrix} \beta^0 \\ \hat{\tilde{v}} - \tilde{v} \end{bmatrix}$ é:

$$C = \begin{bmatrix} \Omega & -\Omega X' V^{-1} Z G \\ -G Z' V^{-1} X \Omega & Q + G Z' V^{-1} X \Omega X' V^{-1} Z G \end{bmatrix}$$

$$\Omega = (X' V^{-1} X)^{-1}$$

$$Q = (Z' R^{-1} Z + G^{-1})^{-1}$$

Assim: $\text{Var} \begin{pmatrix} \beta^0 \\ \hat{\tilde{v}} - \tilde{v} \end{pmatrix} = (X' V^{-1} X)^{-1}$

$$\text{Var} \begin{pmatrix} \hat{\tilde{v}} - \tilde{v} \end{pmatrix} = Q + G Z' V^{-1} X \Omega X' V^{-1} Z G$$

G e R desconhecidas: Usamos nas expressões anteriores as estimativas \hat{G} e $\hat{R} \rightarrow$ EBLUE e EBLUP.

Estimação dos Componentes de Variância

Métodos: Máxima Verossimilhança (ML) e Máxima Verossimilhança Restrita/Residual (REML)

-2 log da função de verossimilhança $L = f(\underline{y})$ é:

$$\text{ML} : -2 \log L = \log |V| + \underline{r}' V^{-1} \underline{r} + n \log(2\pi)$$

$$\text{REML} : -2 \log L_R = \log |V| + \log |X' V^{-1} X| + \underline{r}' V^{-1} \underline{r} + (n - k) \log(2\pi)$$

em que $\underline{r} = \underline{y} - X(X' V^{-1} X)^{-1} X' V^{-1} \underline{y}$ e k é o posto de X.

O PROC MIXED minimiza as funções ML ou REML usando o algoritmo de Newton-Raphson.

Com o PROC MIXED, na análise de dados com medidas repetidas, é possível especificar várias estruturas para a matriz de covariâncias Σ , e assim escolher a mais adequada.

Componentes de Variância (VC):

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ & \sigma^2 & 0 & 0 \\ & & \sigma^2 & 0 \\ \text{(sim.)} & & & \sigma^2 \end{bmatrix}$$

ou

$$VC(AB) \rightarrow \Sigma = \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 \\ & \sigma_A^2 & 0 & 0 \\ & & \sigma_B^2 & 0 \\ \text{(sim.)} & & & \sigma_B^2 \end{bmatrix}$$

Simetria Composta (CS):

$$\Sigma = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ & & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \text{(sim.)} & & & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

Auto-Regressiva de 1ª Ordem (AR(1)):

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ \text{(sim.)} & & & 1 \end{bmatrix}$$

Sem Estrutura (UN):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \\ \text{(sim.)} & & & \end{bmatrix}$$

Toeplitz (TOEP):

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ & \sigma^2 & \sigma_1 & \sigma_2 \\ & & \sigma^2 & \sigma_1 \\ \text{(sim.)} & & & \sigma^2 \end{bmatrix}$$

Huynh-Feldt (HF):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ & & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \text{(sim.)} & & & \sigma_4^2 \end{bmatrix}$$

onde $\lambda > 0$ e λ é a diferença entre a média das variâncias e a média das covariâncias.

Simetria Composta Heterogênea (CSH):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho & \sigma_1\sigma_4\rho \\ & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \text{(sim.)} & & & \sigma_4^2 \end{bmatrix}$$

Auto-Regressiva de 1ª Ordem Heterogênea (ARH(1)):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho^2 & \sigma_1\sigma_4\rho^3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho & \sigma_2\sigma_4\rho^2 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho \\ \text{(sim.)} & & & \sigma_4^2 \end{bmatrix}$$

Antedependência de 1ª Ordem (ANTE(1)):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ \text{(sim.)} & & & \sigma_4^2 \end{bmatrix}$$

Toeplitz Heterogênea (TOEPH):

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\ & & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \text{(sim.)} & & & \sigma_4^2 \end{bmatrix}$$

Auto-Regressiva de 1ª Ordem Médias Móveis (ARMA(1,1)):

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ & 1 & \gamma & \gamma\rho \\ & & 1 & \gamma \\ \text{(sim.)} & & & 1 \end{bmatrix}$$

- **SELEÇÃO DA ESTRUTURA DE COVARIÂNCIA**

Modelos ajustados pelo REML (mesma X)

- (i) **Teste da Razão de Verossimilhanças**

O teste é baseado na diferença entre $-2 \times \log$ da função de verossimilhança residual (restrita) para as duas estruturas de covariâncias, que tem aproximadamente uma distribuição qui-quadrado (χ^2) com graus de liberdade igual à diferença entre os números de parâmetros nas duas estruturas.

(ii) Critérios de Informação (menor é melhor)

- **AIC – Critério de Informação de Akaike**

(Akaike, 1974)

$$AIC = -2\ell + 2d$$

ℓ = \log_e da função de verossimilhança residual (restrita)

d = número de parâmetros de covariância estimados no modelo adotado (d é a dimensão do modelo)

- **BIC – Critério de Informação Bayesiano de Schwarz** (Schwarz, 1978)

$$\text{BIC} = -2\ell + d \log_e(n')$$

$n' = n$ = número de observações (ML), ou

$n' = n - s$ (REML e $s = \text{posto}(X)$), ou

$n' =$ número de sujeitos (“subjects” > 1) – Medidas Repetidas

- **AICC – Critério de Informação de Akaike corrigido**

O AIC corrigido (AICC) é uma versão do AIC que é ajustado para os efeitos dos parâmetros estimados (Burnham e Anderson, 2002) dado por:

$$AICC = -2\ell + 2dn^*/(n^*-d-1)$$

n^* = número total de observações para estimação por máxima verossimilhança (ML) $\rightarrow (n^* = N)$;

$n^* = N - s$ para estimação por máxima verossimilhança restrita, em que $s = \text{posto}(X)$.

Se $n^* < d+2$, toma-se $n^* = d+2$.

REML $\Rightarrow d = q$ o número efetivo de parâmetros de covariância.

ML $\Rightarrow d = q+s$.

Em muitas situações não é totalmente claro como se determina o melhor d e n .

O modelo que minimiza AIC, AICC ou BIC é o escolhido. Quando AIC, AICC ou BIC são próximos, geralmente é escolhido o modelo mais simples com o objetivo de usar um modelo parcimonioso.

- **Inferência sobre os Efeitos Fixos e Aleatórios**

Com o PROC MIXED estimam-se os componentes de variância para os efeitos aleatórios do modelo de acordo com o método especificado, bem como as funções estimáveis dos Tipos I-III conforme as opções E-E3 especificadas no MODEL e o teste F usando as somas de quadrados dos tipos I-III para os efeitos fixos.

Considere combinações lineares estimáveis da seguinte forma:

$$L \begin{bmatrix} \beta \\ \sim \\ \underline{v} \\ \sim \end{bmatrix} \quad (\text{Funções Preditíveis})$$

A estimabilidade requerida (Searle, 1971) se aplica somente para a porção $\underline{\beta}$ de L , porque qualquer combinação linear de \underline{v} é estimável. Tipicamente, inferência em efeitos fixos é o foco, e neste caso, a porção \underline{v} de L é suposta conter todos os elementos nulos.

$$H_0 = L \begin{bmatrix} \beta \\ \sim \\ v \\ \sim \end{bmatrix} = 0$$

Quando L consiste em uma única linha (Posto(L)=1), uma estatística t pode ser construída como segue:

$$t = \frac{L \begin{bmatrix} \beta^0 \\ \sim \\ \hat{v} \\ \sim \end{bmatrix}}{\sqrt{L\hat{C}L'}} \sim t_{\hat{\omega}}$$

em que:

$$\hat{C} = \begin{bmatrix} \hat{\Omega} & -\hat{\Omega}X'\hat{V}^{-1}ZG \\ -GZ'\hat{V}^{-1}X\hat{\Omega} & \hat{Q} + GZ'\hat{V}^{-1}X\hat{\Omega}X'\hat{V}^{-1}ZG \end{bmatrix}$$

$$\hat{\Omega} = (X'\hat{V}^{-1}X)^{-}$$

$$\hat{Q} = (Z'\hat{R}^{-1}Z + \hat{G}^{-1})^{-1}$$

$\hat{\omega}$ = graus de liberdade aproximado

A variância de certas funções pode envolver combinações lineares de componentes de variância. Na instrução MODEL do PROC MIXED podemos usar, dentre outros:

► DDFM=SATTERTH (Satterthwaite, 1946)

ou

► DDFM=KENWARDROGER (ou DDFM=KR)
(Kenward e Roger, 1997)

• Intervalos de confiança em nível de confiança $1-\alpha$

$$\text{IC} \left\{ L \begin{bmatrix} \beta \\ \sim \\ \mathbf{v} \\ \sim \end{bmatrix} \right\}_{1-\alpha} : L \begin{bmatrix} \beta^0 \\ \sim \\ \hat{\mathbf{v}} \\ \sim \end{bmatrix} \pm t_{\hat{\omega}, \alpha/2} \sqrt{L\hat{C}L'}$$

em que

$t_{\hat{\omega}, \alpha/2}$ é o percentil $(1-\alpha/2)100\%$ da distribuição $t_{\hat{\omega}}$.

Quando o posto(L) é maior que 1, a seguinte estatística F geral deve ser considerada:

$$F = \frac{\begin{bmatrix} \beta^0 \\ \sim \\ \hat{v} \\ \sim \end{bmatrix}' L' (L \hat{C} L')^{-1} L \begin{bmatrix} \beta^0 \\ \sim \\ \hat{v} \\ \sim \end{bmatrix}}{r}$$

em que $r = \text{posto}(L)$. Analogamente ao t, em geral, F tem uma distribuição aproximada F com r graus de liberdade no numerador e $\hat{\omega}$ graus de liberdade no denominador.

2.4. EXEMPLOS DE APLICAÇÃO

Littell, Freund e Spector (1991)

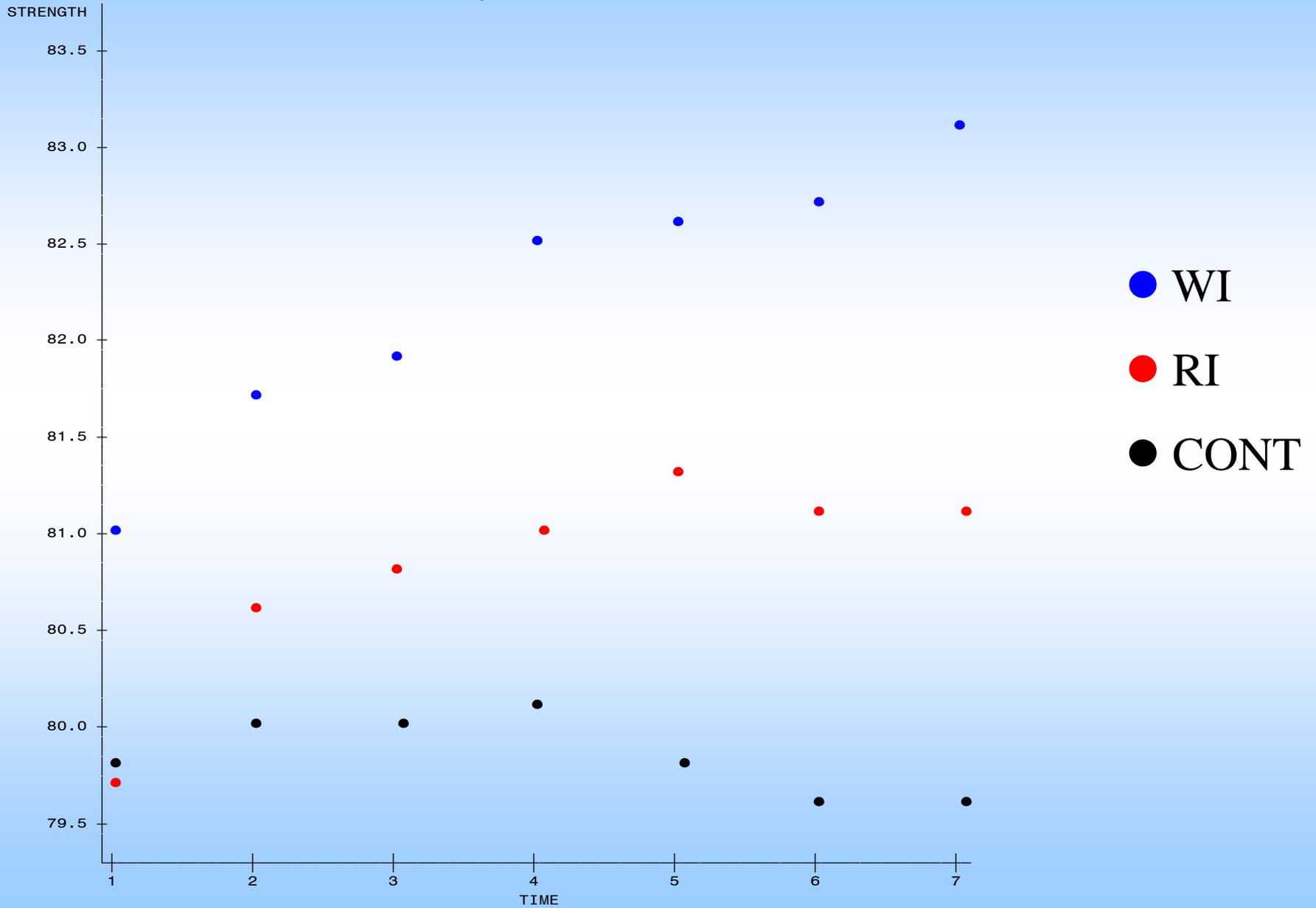
Sujeitos em um estudo de exercícios de terapia foram destinados para um dos três programas de halterofilismo. No primeiro programa (RI), o número de repetições dos exercícios aumentou a medida que os sujeitos se tornaram mais fortes. No segundo programa (WI), a quantidade de peso aumentou a medida que os sujeitos se tornaram mais fortes. No terceiro programa (CONT), os sujeitos não participaram do halterofilismo. A força dos sujeitos foi medida em dias alternados, por duas semanas após o início do estudo.

Dados:

SUBJ	PROGRAM	S1	S2	S3	S4	S5	S6	S7
1	CONT	85	85	86	85	87	86	87
2	CONT	80	79	79	78	78	79	78
3	CONT	78	77	77	77	76	76	77
4	CONT	84	84	85	84	83	84	85
...								
20	CONT	78	79	80	81	80	79	80
1	RI	79	79	79	80	80	78	80
2	RI	83	83	85	85	86	87	87
3	RI	81	83	82	82	83	83	82
4	RI	81	81	81	82	82	83	81
...								
16	RI	84	85	85	85	85	83	82
1	WI	84	85	84	83	83	83	84
2	WI	74	75	75	76	75	76	76
3	WI	83	84	82	81	83	83	82
4	WI	86	87	87	87	87	87	86
...								
21	WI	80	81	80	81	81	82	83

PERFIS

Plot of STRENGTH*TIME. Symbol is value of PROGRAM.

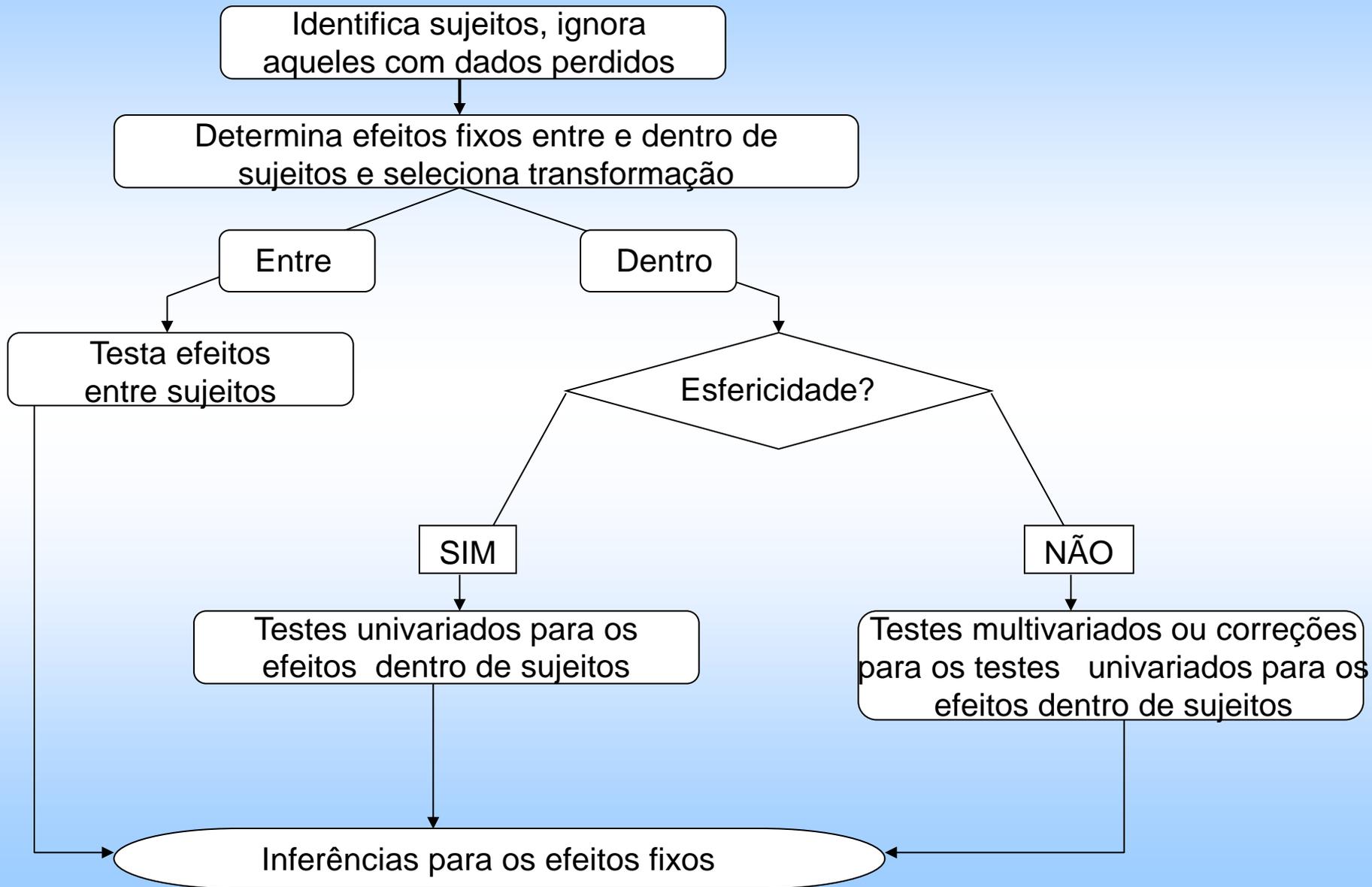


2.4.1. Análises utilizando o GLM do SAS

- Análise com o Modelo de Parcelas Subdivididas
- Análise Multivariada de Perfis
- Análise Univariada de Perfis
- Análise de Tendências de Variação

Wolfinger & Chang (1999)

Esquema da análise de medidas repetidas no PROC GLM do SAS



PROGRAMA 1

Análise com o Modelo de Parcelas Subdivididas

```
options nodate nonumber;
data multiv (keep=parcela program rep s1 s2 s3 s4 s5 s6 s7)
    univ (keep=parcela program rep time y);
input parcela program $ rep s1 s2 s3 s4 s5 s6 s7;
output multiv;
y=s1; time=1; output univ;
y=s2; time=2; output univ;
y=s3; time=3; output univ;
y=s4; time=4; output univ;
y=s5; time=5; output univ;
y=s6; time=6; output univ;
y=s7; time=7; output univ;
cards;
```

01	CONT	1	85	85	86	85	87	86	87
02	CONT	2	80	79	79	78	78	79	78
03	CONT	3	78	77	77	77	76	76	77
04	CONT	4	84	84	85	84	83	84	85
05	CONT	5	80	81	80	80	79	79	80
06	CONT	6	76	78	77	78	78	77	74
07	CONT	7	79	79	80	79	80	79	81
08	CONT	8	76	76	76	75	75	74	74
09	CONT	9	77	78	78	80	80	81	80
10	CONT	10	79	79	79	79	77	78	79
11	CONT	11	81	81	80	80	80	81	82
12	CONT	12	77	76	77	78	77	77	77
13	CONT	13	82	83	83	83	84	83	83
14	CONT	14	84	84	83	82	81	79	78
15	CONT	15	79	81	81	82	82	82	80
16	CONT	16	79	79	78	77	77	78	78

17	CONT	17	83	82	83	85	84	83	82
18	CONT	18	78	78	79	79	78	77	77
19	CONT	19	80	80	79	79	80	80	80
20	CONT	20	78	79	80	81	80	79	80
21	RI	1	79	79	79	80	80	78	80
22	RI	2	83	83	85	85	86	87	87
23	RI	3	81	83	82	82	83	83	82
24	RI	4	81	81	81	82	82	83	81
25	RI	5	80	81	82	82	82	84	86
26	RI	6	76	76	76	76	76	76	75
27	RI	7	81	84	83	83	85	85	85
28	RI	8	77	78	79	79	81	82	81
29	RI	9	84	85	87	89	88	85	86
30	RI	10	74	75	78	78	79	78	78
31	RI	11	76	77	77	77	77	76	76
32	RI	12	84	84	86	85	86	86	86

33	RI	13	79	80	79	80	80	82	82
34	RI	14	78	78	77	76	75	75	76
35	RI	15	78	80	77	77	75	75	75
36	RI	16	84	85	85	85	85	83	82
37	WI	1	84	85	84	83	83	83	84
38	WI	2	74	75	75	76	75	76	76
39	WI	3	83	84	82	81	83	83	82
40	WI	4	86	87	87	87	87	87	86
41	WI	5	82	83	84	85	84	85	86
42	WI	6	79	80	79	79	80	79	80
43	WI	7	79	79	79	81	81	83	83
44	WI	8	87	89	91	90	91	92	92
45	WI	9	81	81	81	82	82	83	83
46	WI	10	82	82	82	84	86	85	87
47	WI	11	79	79	80	81	81	81	81
48	WI	12	79	80	81	82	83	82	82
49	WI	13	83	84	84	84	84	83	83
50	WI	14	81	81	82	84	83	82	85
51	WI	15	78	78	79	79	78	79	79
52	WI	16	83	82	82	84	84	83	84
53	WI	17	80	79	79	81	80	80	80
54	WI	18	80	82	82	82	81	81	81
55	WI	19	85	86	87	86	86	86	86
56	WI	20	77	78	80	81	82	82	82
57	WI	21	80	81	80	81	81	82	83

i

```
proc glm data=univ;  
class program rep time;  
model y=program rep(program) time program*time;  
test h=program e=rep(program);  
lsmeans program*time;  
run;
```

ESTIMATE

CONTRAST

DIFF, etc.

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	74	4210.052864	56.892606	47.53	<.0001
Error (RES (b))	324	387.786735	1.196873		
Corrected Total	398	4597.839599			

R-Square Coeff Var Root MSE y Mean
0.915659 1.350972 1.094017 80.97995

Source	DF	Type I SS	Mean Square	F Value	Pr > F
program	2	419.435262	209.717631	175.22	<.0001
rep(program)	54	3694.690051	68.420186	57.17	<.0001
time	6	52.927318	8.821220	7.37	<.0001
program*time	12	43.000233	3.583353	2.99	0.0005

Source	DF	Type III SS	Mean Square	F Value	Pr > F
program	2	419.435262	209.717631	175.22	<.0001
rep(program) (RES (a))	54	3694.690051	68.420186	57.17	<.0001
time	6	53.354264	8.892377	7.43	<.0001
program*time	12	43.000233	3.583353	2.99	0.0005

Tests of Hypotheses Using the Type III MS for rep(program) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
program	2	419.4352623	209.7176311	3.07	0.0548

Least Squares Means

PROGRAM	S1	S2	S3	S4	S5	S6	S7
CONT	79.75	79.95	80.00	80.05	79.80	79.60	79.60
RI	79.68	80.56	80.81	81.00	81.25	81.12	81.12
WI	81.04	81.66	81.90	82.52	82.61	82.71	83.09

E = matriz de somas de quadrados e de produtos cruzados do resíduo.

Programas: MANOVA, DIC com um fator (Program), etc.

E = Error SS&CP Matrix

	S1	S2	S3	S4	S5	S6	S7
S1	474.13	472.89	484.15	442.72	468.63	443.91	454.52
S2	472.89	511.55	511.02	462.71	496.88	471.47	469.14
S3	484.15	511.02	578.24	536.04	575.98	543.80	551.56
S4	442.72	462.71	536.04	544.18	572.39	534.54	542.35
S5	468.63	496.88	575.98	572.39	653.15	612.61	613.66
S6	443.91	471.47	543.80	534.54	612.61	634.83	629.12
S7	454.52	469.14	551.56	542.35	613.66	629.12	686.35

Matriz de Correlações:

DF = 54

	S1	S2	S3	S4	S5	S6	S7
S1	1.000000 0.0001	0.960210 0.0001	0.924649 0.0001	0.871580 0.0001	0.842113 0.0001	0.809118 0.0001	0.796771 0.0001
S2	0.960210 0.0001	1.000000 0.0001	0.939585 0.0001	0.876990 0.0001	0.859610 0.0001	0.827336 0.0001	0.791740 0.0001
S3	0.924649 0.0001	0.939585 0.0001	1.000000 0.0001	0.955591 0.0001	0.937237 0.0001	0.897542 0.0001	0.875517 0.0001
S4	0.871580 0.0001	0.876990 0.0001	0.955591 0.0001	1.000000 0.0001	0.960087 0.0001	0.909447 0.0001	0.887424 0.0001
S5	0.842113 0.0001	0.859610 0.0001	0.937237 0.0001	0.960087 0.0001	1.000000 0.0001	0.951369 0.0001	0.916529 0.0001
S6	0.809118 0.0001	0.827336 0.0001	0.897542 0.0001	0.909447 0.0001	0.951369 0.0001	1.000000 0.0001	0.953077 0.0001
S7	0.796771 0.0001	0.791740 0.0001	0.875517 0.0001	0.887424 0.0001	0.916529 0.0001	0.953077 0.0001	1.000000 0.0001

PROGRAMA 2

(Usando Análise Multivariada)

```
proc glm data=multiv;  
class program;  
model s1 s2 s3 s4 s5 s6 s7 = program / nouni;  
repeated time 7(1 2 3 4 5 6 7) polynomial / printe printh  
printm summary;  
run;
```

Fornece: (i) Resultados do Programa 1; (ii) Testes multivariados para os fatores intraindivíduos (time: perfis horizontais; time*program: perfis paralelos); (iii) Teste de esfericidade; (iv) Correções G-G e H-F; (v) Testes sobre tendências.

Testes Multivariados para os Fatores Intraindivíduos

- Para o fator Tempo (Perfis Horizontais)

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no time Effect

H = Type III SSCP Matrix for **time**

E = Error SSCP Matrix

S=1 M=2 N=23.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.55848168	6.46	6	49	<.0001
Pillai's Trace	0.44151832	6.46	6	49	<.0001
Hotelling-Lawley Trace	0.79056903	6.46	6	49	<.0001
Roy's Greatest Root	0.79056903	6.46	6	49	<.0001

- Para o fator TempoxProgram (Perfis Paralelos)

The GLM Procedure
Repeated Measures Analysis of Variance

MANOVA Test Criteria and F Approximations for the Hypothesis of no time*program Effect

H = Type III SSCP Matrix for **time*program**

E = Error SSCP Matrix

S=2 M=1.5 N=23.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.73167437	1.38	12	98	0.1880
Pillai's Trace	0.28188936	1.37	12	100	0.1943
Hotelling-Lawley Trace	0.34819029	1.40	12	73.199	0.1847
Roy's Greatest Root	0.28259027	2.35	6	50	0.0442

Teste de Esfericidade

("Orthogonal Components")

- Teste para verificar se a matriz de covariâncias pode ser considerada do tipo Huynh-Feldt.

The GLM Procedure

Repeated Measures Analysis of Variance

Sphericity Tests

Variables	DF	Mauchly's		Pr > ChiSq
		Criterion	Chi-Square	
Transformed Variates	20	0.0403737	166.18471	<.0001
Orthogonal Components	20	0.0403737	166.18471	<.0001

- Testes Univariados e correções para os números de graus de liberdade para os efeitos intraindivíduos

The GLM Procedure
 Repeated Measures Analysis of Variance
 Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
program	2	419.435262	209.717631	3.07	0.0548
Error	54	3694.690051	68.420186		

The GLM Procedure
 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	H - F
time	6	53.3542637	8.8923773	7.43	<.0001	0.0003	0.0002
time*program	12	43.0002327	3.5833527	2.99	0.0005	0.0130	0.0104
Error(time)	324	387.7867347	1.1968726				
			Greenhouse-Geisser Epsilon			0.4233	
			Huynh-Feldt Epsilon			0.4624	

PROGRAMA 3

(MANOVA)

- Testa a hipótese de igualdade dos vetores de médias de tratamentos (Program) → **Perfis coincidentes**

```
proc glm data=multiv;
```

```
class program;
```

```
model s1 s2 s3 s4 s5 s6 s7 = program / nouni;
```

```
manova h=program / printh printe short;
```

```
run;
```

The GLM Procedure

Multivariate Analysis of Variance

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall program Effect

H = Type III SSCP Matrix for **program**

E = Error SSCP Matrix

S=2 M=2 N=23

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.67261322	1.50	14	96	0.1242
Pillai's Trace	0.35155260	1.49	14	98	0.1279
Hotelling-Lawley Trace	0.45081028	1.52	14	73.513	0.1243
Roy's Greatest Root	0.34738552	2.43	7	49	0.0320

RESUMO DO GLM

EFEITO	PARCELA SUBDIVIDIDA	UNIVARIADA DE PERFIS		
			G-G	H-F
Program	0,0548	0,0548	-	-
Time	0,0001	0,0001	0,0003	0,0002
Program*Time	0,0005	0,0005	0,0130	0,0104

EFEITO	MULTIVARIADA DE PERFIS			
	WILKS	PILLAI	H-L	ROY
Time	0,0001	0,0001	0,0001	0,0001
Program*Time	0,1880	0,1943	0,1827	0,0442
Igualdade dos vetores de médias: Program	0,1242	0,1279	0,1243	0,0320

● Análise de Tendências de Variação

Analysis of Variance of Contrast Variables

TIME.N represents the nth degree polynomial contrast for TIME

Contrast Variable: TIME.1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	40.51445294	40.51445294	9.85	0.0028
PROGRAM	2	40.39136233	20.19568117	4.91	0.0110
Error	54	222.18758503	4.11458491		

Contrast Variable: TIME.2

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	10.57713133	10.57713133	8.64	0.0048
PROGRAM	2	1.42410491	0.71205245	0.58	0.5626
Error	54	66.13395692	1.22470291		

Contrast Variable: TIME.3

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	1.31320035	1.31320035	2.55	0.1163
PROGRAM	2	0.03999060	0.01999530	0.04	0.9620
Error	54	27.83720238	0.51550375		

Analysis of Variance of Contrast Variables

TIME.N represents the nth degree polynomial contrast for TIME

Contrast Variable: TIME.4

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	0.08132173	0.08132173	0.12	0.7329
PROGRAM	2	0.53719750	0.26859875	0.39	0.6798
Error	54	37.31675557	0.69105103		

Contrast Variable: TIME.5

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	0.69325639	0.69325639	2.19	0.1451
PROGRAM	2	0.20663265	0.10331633	0.33	0.7234
Error	54	17.12670068	0.31716112		

Contrast Variable: TIME.6

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MEAN	1	0.17490098	0.17490098	0.55	0.4617
PROGRAM	2	0.40094473	0.20047237	0.63	0.5365
Error	54	17.18453412	0.31823211		

Além do comando **POLYNOMIAL**, o **REPEATED** do GLM possibilita outros tipos de estudo:

→ CONTRAST (k)

→ MEAN

→ HELMERT

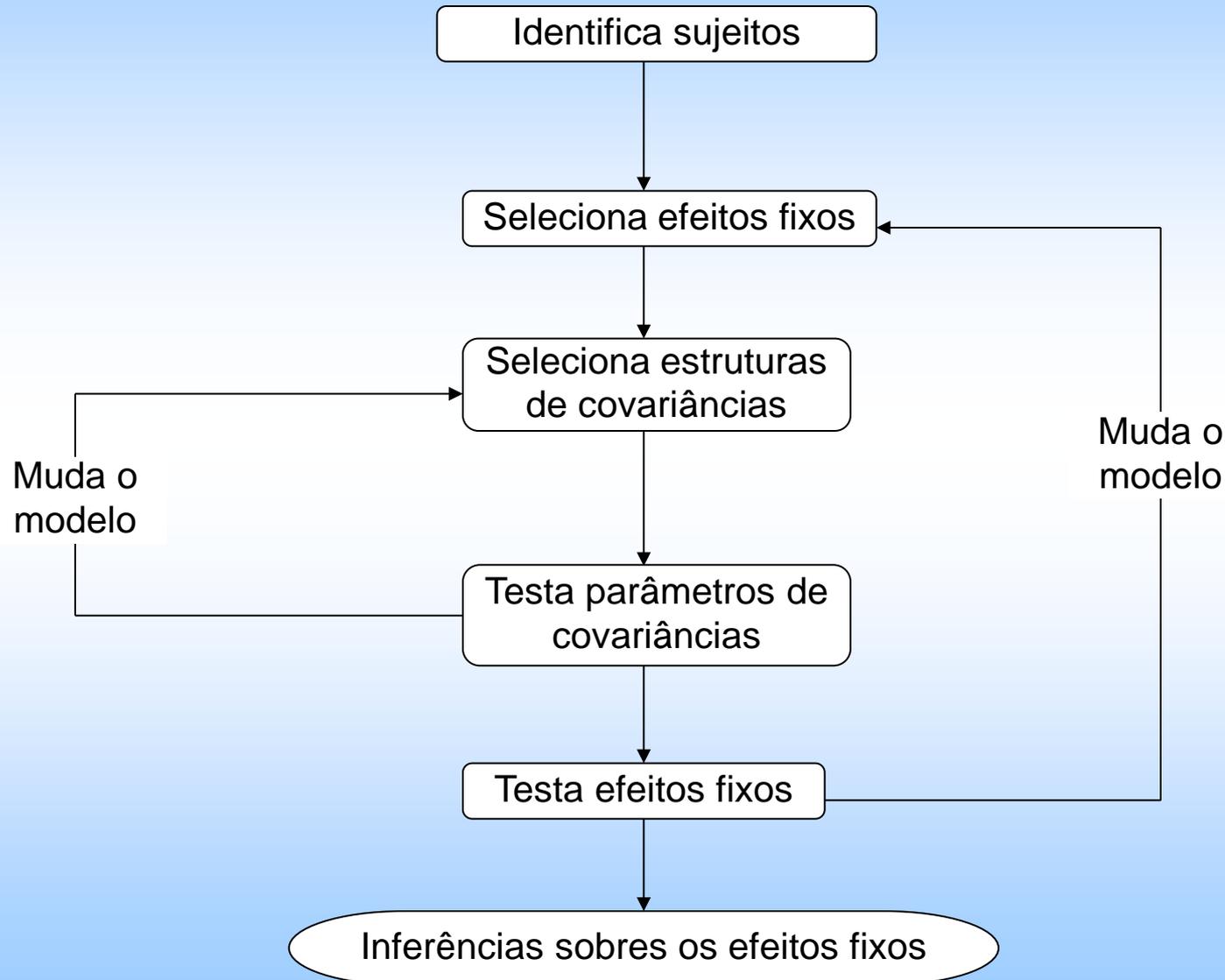
→ PROFILE

2.4.2. Análises utilizando o Proc MIXED do SAS

- Modelando a Estrutura de Covariância:
- Escolha da Estrutura de Covariância
- Efeito da Estrutura de Covariância nos Testes e Estimativas
- Modelando o TEMPO como uma variável de Regressão (Covariável)

Wolfinger & Chang (1999)

Esquema da análise de medidas repetidas no proc MIXED do SAS



Estrutura	Parâ- metros	AIC (Akaike)	BIC (Schwarz)	AICC (AIC c.)
Auto-Reg. de 1ª Ordem.: AR(1)	2	1270,8	1274,9	1270,8
Auto-Reg. Het. de 1ª O.: ARH(1)	8	1270,9	1287,2	1271,3
Auto-Reg. de 1ª Ordem Médias Móveis: ARMA(1,1)	3	1271,3	1277,4	1271,3
Toeplitz Heterogênea: TOEPH	8	1274,7	1301,2	1275,7
Toeplitz: TOEP	7	1275,1	1289,4	1275,4
Antedependência de 1ª Ordem: ANTE(1)	13	1277,6	1304,2	1278,6
Sem Estrutura: UN	28	1290,9	1348,1	1295,5
Simetria Composta Heter.: CSH	8	1423,1	1439,4	1423,5
Huynh-Feldt: HF	8	1424,2	1440,6	1424,6
Simetria Composta: CS	2	1424,8	1428,9	1424,9
Componentes de Variância: VC	1	2035,9	2037,9	2035,9

PROGRAMA M1: Sem Estrutura (UN)

```
proc mixed data=univ;  
class program rep time;  
model y=program time program*time;  
repeated time / subject=rep(program) type=un r rcorr;  
run;
```

Estrutura UN

Matrizes de covariâncias e de correlações:

R Matrix for SUBJ(PROGRAM) 1 CONT

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7
1	8.78	8.75	8.96	8.19	8.67	8.22	8.41
2	8.75	9.47	9.46	8.56	9.20	8.73	8.68
3	8.96	9.46	10.70	9.92	10.66	10.07	10.21
4	8.19	8.56	9.92	10.07	10.59	9.89	10.04
5	8.67	9.20	10.66	10.59	12.09	11.34	11.36
6	8.22	8.73	10.07	9.89	11.34	11.75	11.65
7	8.41	8.68	10.21	10.04	11.36	11.65	12.71

R Correlation Matrix for SUBJ(PROGRAM) 1 CONT

Row	COL1	COL2	COL3	COL4	COL5	COL6	COL7
1	1.00	0.96	0.92	0.87	0.84	0.80	0.79
2	0.96	1.00	0.93	0.87	0.85	0.82	0.79
3	0.92	0.93	1.00	0.95	0.93	0.89	0.87
4	0.87	0.87	0.95	1.00	0.96	0.90	0.88
5	0.84	0.85	0.93	0.96	1.00	0.95	0.91
6	0.80	0.82	0.89	0.90	0.95	1.00	0.95
7	0.79	0.79	0.87	0.88	0.91	0.95	1.00

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood	1234.9
AIC (smaller is better)	1290.9
AICC (smaller is better)	1295.5
BIC (smaller is better)	1348.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
27	798.99	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	2	54	3.07	0.0548
time	6	54	7.12	<.0001
program*time	12	54	1.57	0.1297

Comparando Resultado de Duas Estruturas de Covariâncias

PROGRAMA M2: Simetria Composta (CS)

```
proc mixed data=univ;  
class program rep time;  
model y=program time program*time;  
repeated time / subject=rep(program) type=cs r rcorr;  
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	rep(program)	9.6033
Residual		1.1969

Fit Statistics

-2 Res Log Likelihood	1420.8
AIC (smaller is better)	1424.8
AICC (smaller is better)	1424.9
BIC (smaller is better)	1428.9

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	613.06	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	2	54	3.07	0.0548
time	6	324	7.43	<.0001
program*time	12	324	2.99	0.0005

Teste da Razão de Verossimilhança para UN vs. CS no Mixed

Estrutura	Parâmetros	-2 Res Log Verossimilhança
Sem Estrutura: UN	28	1234,9
Simetria Composta: CS	2	1420,8
Diferença	26	185,9
Qui-quadrado ($\alpha=0,00001$)	68,771	Valor-p bem menor que 0,00001

PROGRAMA M3: Huynh-Feldt (HF)

```
proc mixed data=univ;  
class program rep time;  
model y=program time program*time;  
repeated time / subject=rep(program) type=hf r rcorr;  
run;
```

Fit Statistics

-2 Res Log Likelihood	1408.2
AIC (smaller is better)	1424.2
AICC (smaller is better)	1424.6
BIC (smaller is better)	1440.6

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
7	625.67	<.0001

The Mixed Procedure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	2	54	2.46	0.0953
time	6	324	7.43	<.0001
program*time	12	324	2.99	0.0005

- **CONTRAST**
- **ESTIMATE**

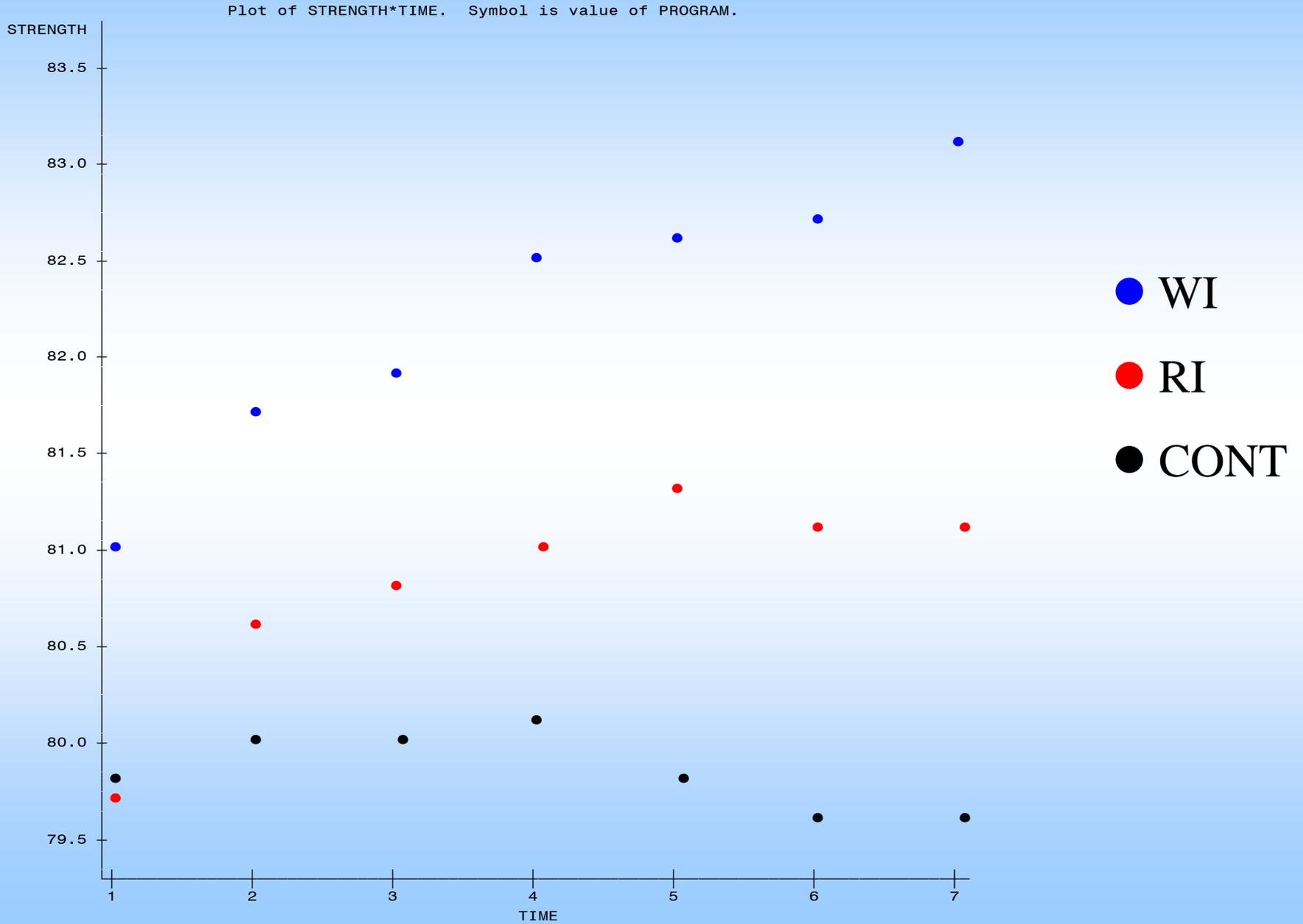
Teste da Razão de Verossimilhança para verificação da condição de H-F no Mixed

Estrutura	Parâmetros	-2 Res Log Verossimilhança
Sem Estrutura: UN	28	1234,9
Huynh-Feldt: HF	8	1408,2
Diferença	20	173,3
Qui-quadrado ($\alpha=0,00001$)	58,976	Valor-p bem menor que 0,00001

Efeito da Estrutura de Covariância nos Testes e Estimativas

Estrutura	CS	AR(1)	UN
PROGRAM	0,0548	0,0528	0,0548
TIME	<0,0001	0,0003	<0,0001
PROGRAM*TIME	0,0005	0,3007	0,1297

PERFIS



Modelando o TEMPO como uma variável de Regressão

PROGRAMA M4:

```
proc mixed data=univ;  
class program rep;  
model y=program time time*program time*time  
time*time*program / htype=1;  
repeated / type=ar(1) subject=rep(program);  
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	rep(program)	0.9523
Residual		10.7585

Fit Statistics

-2 Res Log Likelihood	1281.0
AIC (smaller is better)	1285.0
AICC (smaller is better)	1285.1
BIC (smaller is better)	1289.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	797.49	<.0001

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	2	54	3.10	0.0530
time	1	336	12.69	0.0004
time*program	2	336	4.75	0.0093
time*time	1	336	7.18	0.0077
time*time*program	2	336	0.88	0.4167

Ajustando as Equações de Regressão para cada PROGRAM

PROGRAMA M5:

```
proc mixed data=univ;  
class program rep;  
model y=program time*program time*time*program / noint s  
htype=1;  
repeated / type=ar(1) subject=rep(program);  
run;
```

Ajustando as Equações de Regressão para cada PROGRAM

Solution for Fixed Effects

Effect	program	Estimate	Standard Error	DF	t Value	Pr > t
program	CONT	79.5708	0.7972	54	99.82	<.0001
program	RI	78.9054	0.8913	54	88.53	<.0001
program	WI	80.4928	0.7780	54	103.47	<.0001
time*program	CONT	0.2092	0.2353	336	0.89	0.3746
time*program	RI	0.8606	0.2630	336	3.27	0.0012
time*program	WI	0.5861	0.2296	336	2.55	0.0111
time*time*program	CONT	-0.02930	0.02731	336	-1.07	0.2842
time*time*program	RI	-0.07767	0.03054	336	-2.54	0.0114
time*time*program	WI	-0.03063	0.02666	336	-1.15	0.2514

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
program	3	54	12910.8	<.0001
time*program	3	336	7.39	<.0001
time*time*program	3	336	2.98	0.0316

Equações Ajustadas

- CONT:

$$\hat{Y} = 79,5708 + 0,2092*TIME - 0,02930*TIME*TIME$$

- RI:

$$\hat{Y} = 78,9054 + 0,8606*TIME - 0,07767*TIME*TIME$$

- WI:

$$\hat{Y} = 80,4928 + 0,5861*TIME - 0.03063*TIME*TIME$$

Outro Exemplo de Análise utilizando o PROC MIXED do SAS

Dados de um Experimento no Delineamento
Inteiramente Casualizado:

- ▶ 3 CLONES e 6 repetições (18 parcelas)
- ▶ Avaliação feita ao longo de 5 anos
- ▶ Y= Volume de madeira de eucalipto (m^3)

Dados: Volume de madeira de eucalipto (m³):
3 clones (A, B e C) e 6 repetições

	Ano 2008	Ano 2009	Ano 2010	Ano 2011	Ano 2012
1 A 1	1.89888	4.82983	6.52776	8.12235	9.83442
2 A 2	0.95203	3.00719	5.55510	7.51161	9.26246
3 A 3	0.85062	3.11800	5.50205	7.65503	9.35150
4 A 4	0.54591	2.23165	4.20842	5.99528	6.94800
5 A 5	0.59523	2.28909	4.12230	5.78299	6.96796
6 A 6	0.78231	2.64666	4.41512	5.95988	7.59593
7 B 1	0.81954	2.77965	5.24232	7.35190	9.40889
8 B 2	0.94547	3.04266	5.82373	7.47648	9.17779
9 B 3	0.95504	2.95438	5.19410	6.54054	8.78722
10 B 4	1.43249	4.04425	6.01496	6.95222	8.23683
11 B 5	1.11420	3.93039	6.16654	7.80244	10.03936
12 B 6	0.92283	2.64115	5.07950	6.22205	7.59171
13 C 1	1.30652	3.21875	4.91990	6.07924	6.97748
14 C 2	2.03611	4.95572	6.83481	8.12368	9.55007
15 C 3	1.71894	4.90129	7.02010	8.34912	9.96506
16 C 4	1.10707	3.05426	4.61908	5.42976	7.28891
17 C 5	1.28574	3.27673	4.92921	6.21577	7.36372
18 C 6	1.56088	4.29780	5.89195	7.87488	9.69615

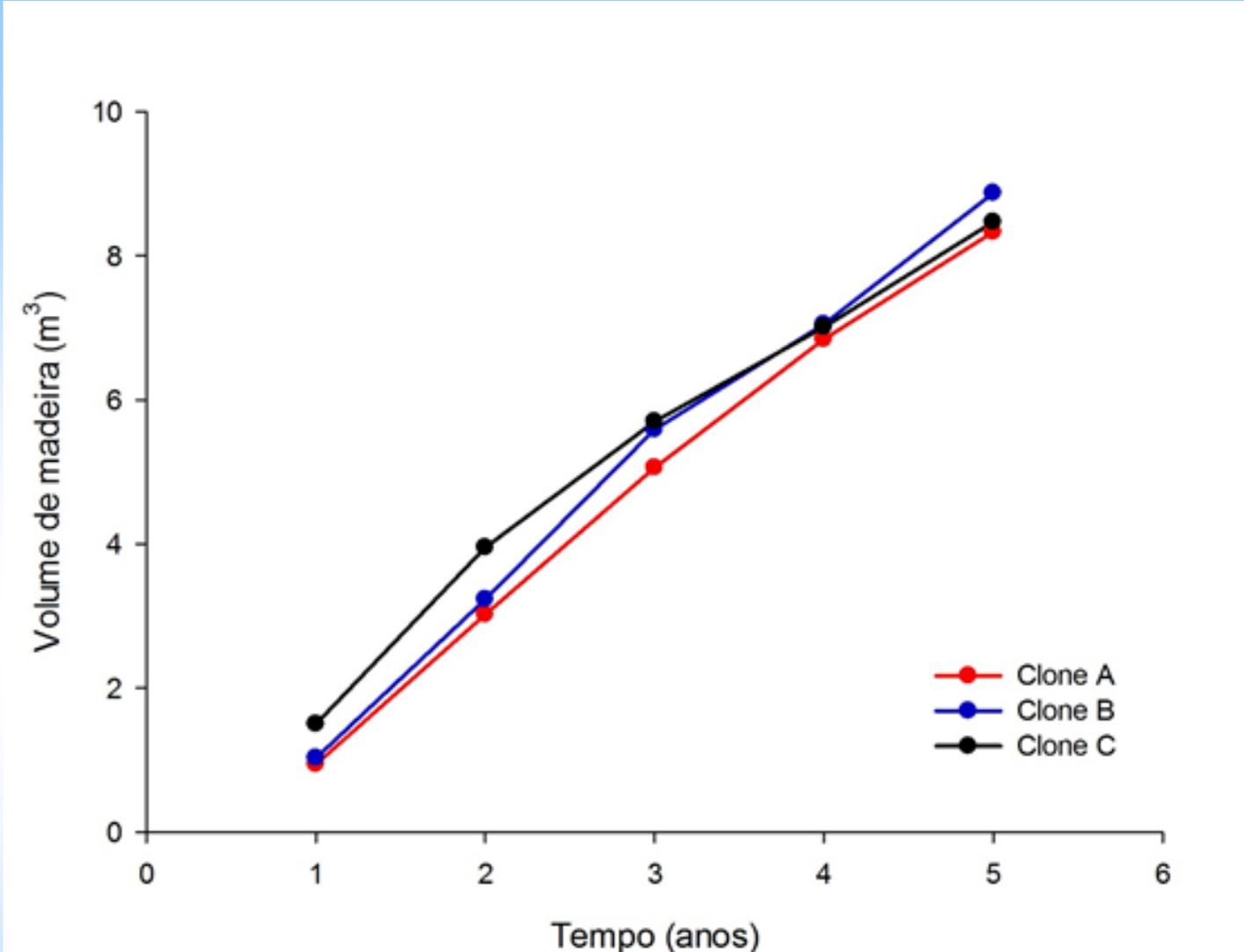


Figura 2.1- Efeito do tempo sobre o volume de madeira produzido por três clones de eucalipto

PROGRAMA M1: Sem Estrutura (UN)

```
options nodate nonumber;  
data multiv (keep=parcela clone rep a1 a2 a3 a4 a5)  
    univ (keep=parcela clone rep time y);  
input parcela clone $ rep a1 a2 a3 a4 a5;  
output multiv;  
y=a1; time=1; output univ;  
y=a2; time=2; output univ;  
y=a3; time=3; output univ;  
y=a4; time=4; output univ;  
y=a5; time=5; output univ;  
cards;
```

1	A	1	1.89888	4.82983	6.52776	8.12235	9.83442
2	A	2	0.95203	3.00719	5.55510	7.51161	9.26246
3	A	3	0.85062	3.11800	5.50205	7.65503	9.35150
4	A	4	0.54591	2.23165	4.20842	5.99528	6.94800
5	A	5	0.59523	2.28909	4.12230	5.78299	6.96796
6	A	6	0.78231	2.64666	4.41512	5.95988	7.59593
7	B	1	0.81954	2.77965	5.24232	7.35190	9.40889
8	B	2	0.94547	3.04266	5.82373	7.47648	9.17779
9	B	3	0.95504	2.95438	5.19410	6.54054	8.78722
10	B	4	1.43249	4.04425	6.01496	6.95222	8.23683
11	B	5	1.11420	3.93039	6.16654	7.80244	10.03936
12	B	6	0.92283	2.64115	5.07950	6.22205	7.59171
13	C	1	1.30652	3.21875	4.91990	6.07924	6.97748
14	C	2	2.03611	4.95572	6.83481	8.12368	9.55007
15	C	3	1.71894	4.90129	7.02010	8.34912	9.96506
16	C	4	1.10707	3.05426	4.61908	5.42976	7.28891
17	C	5	1.28574	3.27673	4.92921	6.21577	7.36372
18	C	6	1.56088	4.29780	5.89195	7.87488	9.69615

i

```
proc mixed data=univ;  
class clone rep time;  
model y=clone time clone*time;  
repeated time / subject=rep(clone) type=un r rcorr;  
run;
```

Estrutura UN

Matrizes de covariâncias e de correlações:

Estimated R Matrix for rep(clone) 1 A

Row	Col1	Col2	Col3	Col4	Col5
1	0.1361	0.2890	0.2761	0.2644	0.2874
2	0.2890	0.6815	0.6723	0.6896	0.7852
3	0.2761	0.6723	0.7439	0.8072	0.9168
4	0.2644	0.6896	0.8072	0.9956	1.1585
5	0.2874	0.7852	0.9168	1.1585	1.4684

Estimated R Correlation Matrix for rep(clone) 1 A

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.9492	0.8678	0.7184	0.6429
2	0.9492	1.0000	0.9442	0.8372	0.7849
3	0.8678	0.9442	1.0000	0.9379	0.8772
4	0.7184	0.8372	0.9379	1.0000	0.9581
5	0.6429	0.7849	0.8772	0.9581	1.0000

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood	59.0
AIC (smaller is better)	89.0
AICC (smaller is better)	97.2
BIC (smaller is better)	102.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
14	164.42	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
clone	2	15	0.57	0.5746
time	4	15	333.06	<.0001
clone*time	8	15	3.78	0.0128

Comparando Resultado de Duas Estruturas de Covariâncias

PROGRAMA M2: Simetria Composta (CS)

```
proc mixed data=univ;  
class clone rep time;  
model y=clone time clone*time;  
repeated time / subject=rep(clone) type=cs r rcorr;  
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	rep(clone)	0.6147
Residual		0.1905

Fit Statistics

-2 Res Log Likelihood	158.0
AIC (smaller is better)	162.0
AICC (smaller is better)	162.1
BIC (smaller is better)	163.7

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	65.50	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
clone	2	15	0.57	0.5746
time	4	60	803.46	<.0001
clone*time	8	60	1.81	0.0933

Teste da Razão de Verossimilhança para UN vs. CS no Mixed

Estrutura	Parâmetros	-2 Res Log Verossimilhança
Sem Estrutura: UN	15	59,0
Simetria Composta: CS	2	158,0
Diferença	13	99,0**

Qui-quadrado ($\alpha=1\%$; 13 gl.) = 27,69 ** Valor-p bem menor que 0,01

Estruturas	Parâ- metros	AIC (Akaike)	BIC (Schwarz)	AICC (AIC c.)
Toeplitz Heterogênea: TOEPH	9	80,3	88,3	83,0
Auto-Reg. Het. de 1ª O.:ARH(1)	6	80,6	86,0	81,8
Antedependência de 1ª Ordem: ANTE(1)	9	86,0	94,0	88,8
Sem Estrutura: UN	15	89,0	102,4	97,2
Toeplitz: TOEP	5	117,4	121,8	118,2
Auto-Reg. de 1ª Ordem Médias Móveis: ARMA(1,1)	3	120,6	123,2	120,9
Auto-Reg. de 1ª Ordem:AR(1)	2	123,0	124,8	123,2
Huynh-Feldt: HF	6	123,3	128,7	124,6
Simetria Composta Heter.: CSH	6	124,6	129,9	125,8
Simetria Composta: CS	2	162,0	163,7	162,1
Componentes de Variância: VC	1	225,5	226,3	225,5

PROGRAMA M3: Toeplitz Heterogênea: TOEPH

```
proc mixed data=univ;  
class clone rep time;  
model y=clone time clone*time;  
repeated time / subject=rep(clone) type=toeph r rcorr;  
run;
```

Fit Statistics

-2 Res Log Likelihood	62.3
AIC (smaller is better)	80.3
AICC (smaller is better)	83.0
BIC (smaller is better)	88.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
8	161.21	<.0001

The Mixed Procedure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
clone	2	15	0.57	0.5797
time	4	60	335.49	<.0001
clone*time	8	60	3.40	0.0028

Modelando o TEMPO como uma variável de Regressão

PROGRAMA M4:

```
proc mixed data=univ;  
class clone rep;  
model y=clone time time*clone time*time  
time*time*clone / htype=1;  
repeated / type=arh(1) subject=rep(clone);  
run;
```

Fit Statistics

-2 Res Log Likelihood	92.5
AIC (smaller is better)	104.5
AICC (smaller is better)	105.6
BIC (smaller is better)	109.8

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
5	150.61	<.0001

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
clone	2	15	1.06	0.3703
time	1	66	1244.52	<.0001
time*clone	2	66	3.92	0.0245
time*time	1	66	42.41	<.0001
time*time*clone	2	66	0.72	0.4928

Ajustando as Equações de Regressão para cada CLONE

PROGRAMA M5:

```
proc mixed data=univ;  
class clone rep;  
model y=clone time*clone time*time*clone / noint s htype=1;  
repeated / type=arh(1) subject=rep(clone);  
run;
```

Ajustando as Equações de Regressão para cada CLONE

Solution for Fixed Effects

Effect	program	Estimate	Standard Error	DF	t Value	Pr > t
clone	A	-1.4047	0.1523	15	-9.22	<.0001
clone	B	-1.5150	0.1523	15	-9.95	<.0001
clone	C	-1.1629	0.1523	15	-7.63	<.0001
time*clone	A	2.5173	0.2300	66	10.94	<.0001
time*clone	B	2.8219	0.2300	66	12.27	<.0001
time*clone	C	2.6687	0.2300	66	11.60	<.0001
time*time*clone	A	-0.1063	0.03601	66	-2.95	0.0044
time*time*clone	B	-0.1328	0.03601	66	-3.69	0.0005
time*time*clone	C	-0.1671	0.03601	66	-4.64	<.0001

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
clone	3	15	340.51	<.0001
time*clone	3	66	417.46	<.0001
time*time*clone	3	66	14.61	<.0001

Equações Ajustadas

- CLONE A:

$$\hat{Y} = -1,4047 + 2,5173^{**} X - 0,1063^{**} X^2$$

- CLONE B:

$$\hat{Y} = -1,5150 + 2,8219^{**} X - 0,1328^{**} X^2$$

- CLONE C:

$$\hat{Y} = -1,1629 + 2,6687^{**} X - 0,1671^{**} X^2$$

X=Time ** p < 0,01

2.5. EXTENSÕES:

- ▶ Modelos Mistos Generalizados
- ▶ Modelos Mistos Não Lineares
- ▶ Modelos Mistos sob enfoque Bayesiano

2.6. REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control AIC**, v. 19, p. 716-723, 1974.

BARCIKOWSKI, R.S.; ROBEY, R.R. Decisions in single group repeated measures analysis: statistical tests and three computer packages. **The American Statistician**, v. 38, p. 148-150, 1984.

BOX, G.E.P. A general distribution theory for a class of likelihood ratio criteria. **Biometrika**, v. 36, p. 317-346, 1949.

BOX, G.E.P. Problems in the analysis of growth and wear curves. **Biometrics**, v. 6, p. 362-389, 1950.

BOX, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems. I. Effects of inequality of variance in the one-way classification. **Annals of Mathematical Statistics**, v. 25, p. 290-302, 1954a.

BOX, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. **Annals of Mathematical Statistics**, v. 25, p. 484-498, 1954b.

BURNHAM, K.P.; ANDERSON, D.R. **Model selection and multimodel inference: a practical information-theoretic approach**. New York: Springer-Verlag, 2002.

CROWDER, M.J.; HAND, D.J. **Analysis of repeated measures**. London: Chapman & Hall, 1990. 256 p.

CRUZ, C.D.; CARNEIRO, P.C.S.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. v.2, 3. ed. rev. e ampl. – Viçosa: Ed. UFV, 2014. 668 p.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. v.1, 4. ed. Viçosa :UFV, 2012. 514 p.

DAVIDIAN, M.; GILTINAN, D. **Nonlinear models for repeated measurement data**. New York: Chapman & Hall, 1995.

DIGGLE, P.J. An approach to the analysis of repeated measurements. **Biometrics**, v. 44, p. 959-971, 1988.

FEDERER, W.T.; KING, F. **Variations on split plot and split block experimental designs**. New York: John Wiley, 2007. 270 p.

FERNANDES, G.C.J. Repeated measure analysis of line-source sprinkler experiments. **Hort Science**, v. 26, n. 4, p. 339-342, 1991.

FERREIRA, D. F. **Estatística multivariada**. Lavras: Ed. UFLA, 2008. 661 p.

GEISSER, J.; GREENHOUSE, S.W. An extension of Box's results on the use of the F distribution in multivariate analysis. **Annals of the Mathematical Statistics**, v. 29, p. 855-891, 1958.

GREENHOUSE, S.W.; GEISSER, J. On methods in the analysis of profile data. **Psychometrika**, v. 32, p. 95-112, 1959.

GUERIN, L.; STROUP, W. A simulation study to evaluate PROC MIXED analysis of repeated measures data. In **Proceedings of the twelfth annual conference and applied statistics in agriculture**. Manhattan: Kansas State University, 2000.

HARRIS, R.J. **A primer of multivariate statistics**. New York: Academic Press, 1975. 332 p.

HENDERSON, C.R. The estimation of genetic parameters. **The annals of mathematical statistics**, v. 21, p. 309-310, 1950.

HENDERSON, C.R.; KEMPTHORNE, O.; SEARLE, S.R.; et al. Estimation of environmental and genetic trends from records subject to culling. **Biometrics**, v.15, p.192-218, 1959.

HENDERSON, C.R. **Applications of linear models in animal breeding**. University of Guelph, 1984.

HICKS, C.R. **Fundamental Concepts in the Design of Experiments**. 2^a ed., USA, Holt, Rinehart and Winston, Inc., 1973. 349p.

HUYNH, H.; FELDT, L.S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. **Journal of the American Statistical Association**, v. 65, p. 1582-1589, 1970.

HUYNH, H.; FELDT, L.S. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. **Journal of Educational Statistics**, v. 1, p. 69-82, 1976.

HUYNH, H.; FELDT, L.S. Some approximate tests for repeated measurement designs. **Psychometrika**, v. 43, n. 2, p. 161-175, 1978.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4. ed. USA, Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1998. 816 p.

KENWARD, M.G.; ROGER, J.H. Small sample inference for fixed effects from restricted maximum likelihood. **Biometrics**, v. 53, p. 983-997, 1997.

KIRK, R.E. **Experimental design: Procedures for the behavioral sciences**. 3. ed. Pacific Grove, CA: Brooks/Cole, 1995. 921 p.

KUEHL, R.D. **Statistical Principles of Research design and analysis**. Belmont California: Duxbury Press, 1994. 863 p.

LITTELL, R.C.; FREUND, R.J.; SPECTOR, P.C. **SAS system for linear models**. 3. ed. Cary, NC: SAS Institute Inc., 1991.

LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W. WOLFINGER, R.D.; SCHABENBERGER, O. **SAS for Mixed models**. 2. ed. Cary, NC, USA: SAS Institute Inc., 2006. 814 p.

MAUCHLY, J.W. Significance test for sphericity of a normal n-variate distribution. **Annals of Mathematical Statistics**, v. 11, p. 204-209, 1940.

McLEAN, R.A.; SANDERS, W.L. Approximating degrees of freedom for standard errors in mixed linear models. **Proceedings of the statistical computing section**. New Orleans: American Statistical Association, p.50-59, 1988.

McLEAN, R. A.; SANDERS, W.L.; STROUP, W.W. A unified approach to mixed linear models. **The American Statistician**, v. 45, p. 54-64, 1991.

MILLIKEN, G.A.; JOHNSON, D.E. **Analysis of messy data**: designed experiments. New York: Chapman & Hall, 1992. v. 1. 473 p.

MULLER, K.E.; BARTON, C.N. Approximate power for repeated-measures ANOVA lacking sphericity. **Journal of the American Statistical Association**, v. 84, n. 406, p. 549-555, 1989.

MORRISON, D. F. **Multivariate statistical methods**. 2. ed. New York: McGraw-Hill., 1976. 415 p.

MORRISON, D.F. **Multivariate statistical methods**. 2. ed. New Delhi: McGraw-Hill Book, 1981. 415 p.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 2009. 451 p.

POTTHOFF, R.F.; ROY, S.N. A generalized multivariate analysis of variance model useful especially for growth curve problems. **Biometrika**, v. 51, p. 313-326, 1964.

RAO, C.R. **Linear statistical inference and its applications**. 2. ed. New York: John Wiley, 1973. 624 p.

RENCHEER, A.C. **Methods of multivariate analysis**. 2. ed. USA: John Wiley, 2002. 738 p.

SAS Institute Inc. **SAS/STAT 9.1**. User's guide, Cary, NC, USA: SAS Institute Inc., v. 3, p. 1609-2227, 2004.

SCHWARZ, G. Estimating the dimension of a model. **Annals of statistics**, v. 6, p. 461-464, 1978.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. **Biometrics**, n. 2, p. 110-114, 1946.

SEARLE, S. R. **Linear models**. New York: John Wiley, 1971. 532p.

SEARLE, S. R. **Matrix algebra useful for statistics**. 2. ed. USA: John Wiley, 1992. 527p.

SINGER, J.M.; ANDRADE, D.F. Análise de dados longitudinais. SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7., Campinas, **Proceedings...** São Paulo, 1986. 106 p.

STROUP, W.W. Predictable functions and prediction space in the mixed model procedure. In Applications of mixed models in agriculture and related disciplines. Louisiana Agricultural Experiment Station, Baton Rouge. **Southern Cooperative Series Bulletin**, n. 343, p. 39-48, 1989.

VIVALDI, L.J. **Análise de experimentos com dados repetidos ao longo do tempo ou espaço**. Planaltina: Embrapa Cerrados, 1999. 52 p.

VON ENDE, C.N. Repeated-measures analysis: growth and other time-dependent measures. In: SCHEINER, S.M.; GUREVITCH, J. (Eds.). **Design and analysis of ecological experiments**. New York: Chapman & Hall, 1993. p. 113-137.

VONESH, F.E.; CHINCHILLI, V.M. **Linear and nonlinear models for the analysis of repeated measurements**. New York: Marcel Dekker, Inc., 1997. 560 p.

WOLFINGER, CHANG, M. Comparing the SAS GLM and MIXED procedures for repeated measures. Cary, NC, USA: **SAS Institute Inc.**, p. 1-11, 1999.

XAVIER, L. H. **Modelo univariado e multivariado para análise de medidas repetidas e verificação da acurácia do modelo univariado por meio de simulação**. 2000. 91 f. Dissertação (Mestrado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

XAVIER, L. H.; DIAS, C.T.S. Acurácia do modelo univariado para análise de medidas repetidas por simulação multidimensional. **Scientia Agricola**, v. 58, n. 2, p. 241-250, abr./jun. 2001.

MENSAGEM

“O estudante de hoje leva uma desvantagem competitiva se não tiver estudado os conceitos básicos de **Estatística**”.

“O mercado não contrata apenas em função da formação do candidato, mas aquele com potencial para aprender e resolver problemas”.

Prof. Adair José Regazzi



