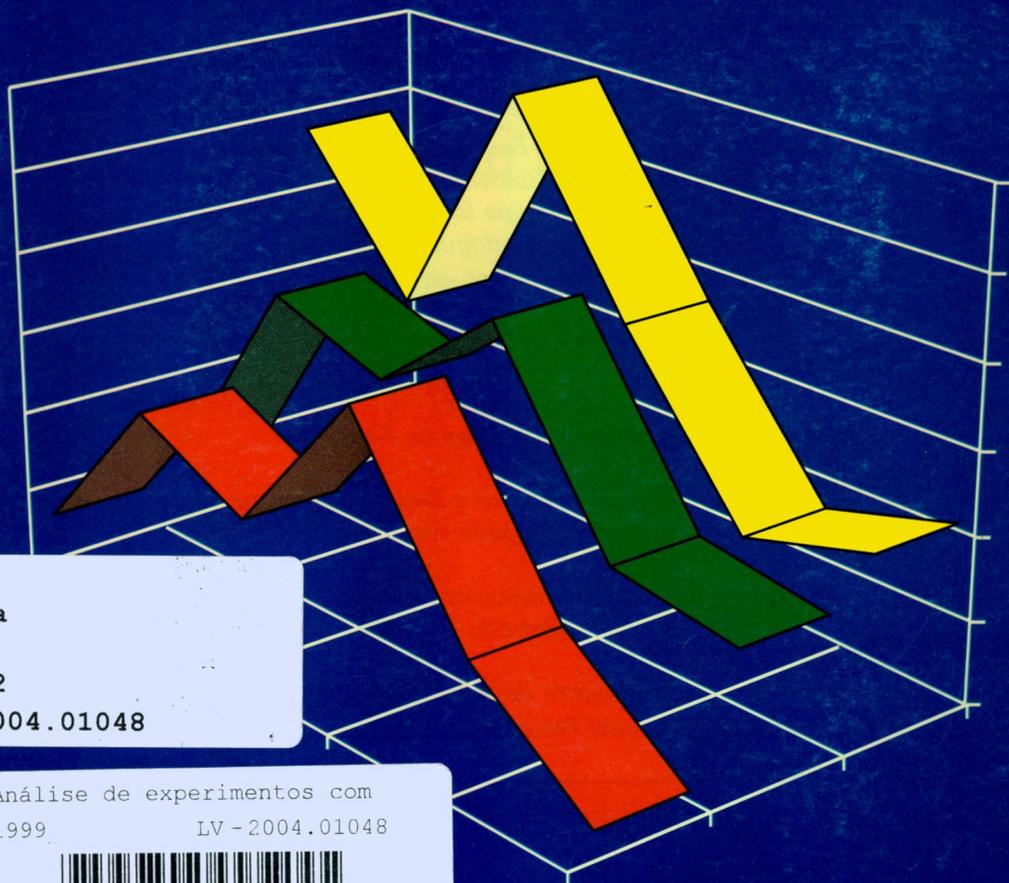


# ANÁLISE DE EXPERIMENTOS COM DADOS REPETIDOS AO LONGO DO TEMPO OU ESPAÇO

Lucio José Vivaldi



CPAC  
V855a  
1999  
ex. 2  
LV-2004.01048

Análise de experimentos com  
1999 LV-2004.01048



17538-2

***República Federativa do Brasil***

*Presidente*  
*Fernando Henrique Cardoso*

***Ministério da Agricultura e do Abastecimento***

*Ministro*  
*Marcus Vinícius Pratini de Moraes*

***Empresa Brasileira de Pesquisa Agropecuária***

*Diretor-Presidente*  
*Alberto Duque Portugal*

*Diretores-Executivos*  
*Dante Daniel Giacomelli Scolari*  
*Elza Ângela Battaglia Brito da Cunha*  
*José Roberto Rodrigues Peres*

***Embrapa Cerrados***

*Chefe-Geral*  
*Carlos Magno Campos da Rocha*

*Chefe Adjunto de Pesquisa e Desenvolvimento*  
*Eduardo Delgado Assad*

*Chefe Adjunto de Comunicação e Negócios*  
*Euzébio Medrado da Silva*

*Chefe Adjunto de Administração*  
*Ismael Ferreira Graciano*



---

*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Cerrados  
Ministério da Agricultura e do Abastecimento*

# **ANÁLISE DE EXPERIMENTOS COM DADOS REPETIDOS AO LONGO DO TEMPO OU ESPAÇO**

Lucio José Vivaldi

ISSN 1517-5111

Doc. - Embrapa Cerrados	Planaltina	n. 8	p.1-52	dez. 1999
-------------------------	------------	------	--------	-----------

Copyright © Embrapa – 1999  
Embrapa Cerrados. Documentos, 8

Exemplares desta publicação podem ser solicitados a:

Embrapa Cerrados  
BR 020, km 18, Rodovia Brasília/Fortaleza  
Caixa Postal 08223  
CEP 73301-970 – Planaltina, DF  
Telefone (61) 388-9898 – Fax (61) 388-9879

Tiragem: 200 exemplares

**Comitê de Publicações:** Eduardo Delgado Assad (Presidente), Maria Alice Bianchi, Daniel Pereira Guimarães, Leide Rovênia Miranda de Andrade, Marco Antonio de Souza, Carlos Roberto Spehar, José Luis Fernandes Zoby e Nilda Maria da Cunha Sette (Secretária-Executiva).

**Coordenação editorial:** Nilda Maria da Cunha Sette

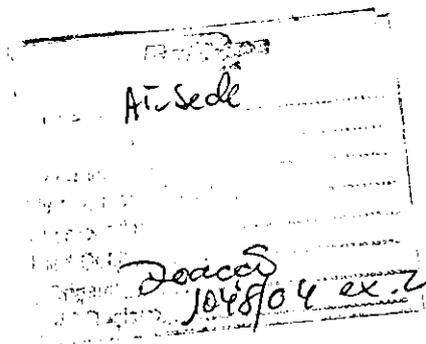
**Revisão gramatical:** Maria Helena Gonçalves Teixeira  
Nilda Maria da Cunha Sette

**Normalização bibliográfica:** Maria Alice Bianchi

**Diagramação e arte-final:** Wellington Cavalcanti

**Capa:** Wellington Cavalcanti

**Impressão e acabamento:** Jaime Arbués Carneiro, Divino Batista de Souza



V855a VIVALDI, Lucio José.

Análise de experimentos com dados repetidos ao longo do tempo ou espaço / Lucio José Vivaldi. - Planaltina: Embrapa Cerrados, 1999.

52p. - (Documentos / Embrapa Cerrados, ISSN 1517-5111; n.8).

1. Método estatístico. I. Título. II. Série.

001.422 - CDD 21

## SUMÁRIO

1 INTRODUÇÃO .....	6
2 INFERÊNCIA E CASUALIZAÇÃO .....	7
2.1 Definições .....	7
2.2 Repetição e casualização .....	9
2.3 Delineamento inteiramente casualizado .....	10
2.4 Modelo aleatório .....	12
2.5 Inferência experimental .....	15
3 DELINEAMENTOS EM PARCELAS DIVIDIDAS .....	21
4 EXPERIMENTOS COM MEDIDAS REPETIDAS NO TEMPO E NO ESPAÇO. ....	26
4.1 Exemplos .....	26
4.2 Análise .....	26
5 ANÁLISE MULTIVARIADA .....	30
6 ANÁLISE DE EXPERIMENTOS .....	34
7 REFERÊNCIAS BIBLIOGRÁFICAS .....	50

# ANÁLISE DE EXPERIMENTOS COM DADOS REPETIDOS AO LONGO DO TEMPO OU ESPAÇO

Lucio José Vivaldi<sup>1</sup>

**RESUMO** - Em uma grande parte dos experimentos uma variável resposta é observada ao longo do tempo, do espaço ou de ambos; a análise de cada resposta, isoladamente, é ineficiente pois não leva em consideração seu comportamento no tempo ou no espaço e as implicações daí advindas. Na análise multivariada, ignora-se que exista uma seqüência no tempo ou no espaço e os teste de significância são menos sensíveis do que os próprios da análise univariada. Dentro deste contexto, buscam-se procedimentos corretos para a análise de variância univariada de experimentos que considerem o tempo e o espaço como fatores. As dificuldades de se tomar o tempo e o espaço como fatores ocorrem, porque eles não são casualizados, ao contrário dos tratamentos aplicados às unidades experimentais, e por isso, a análise de experimentos com medidas repetidas no tempo tem sido objeto de muitos artigos. Para se compreender as técnicas e os procedimentos são necessários conhecimentos sobre casualização, análise multivariada e delineamentos em parcelas divididas. As três primeiras seções tratam destes tópicos enquanto as demais são específicas para a análise de experimentos com medidas repetidas no tempo ou no espaço.

## ANALYSIS OF EXPERIMENTS WITH REPEATED MEASUREMENTS ON TIME OR SPACE

**ABSTRACT** - In many experiments a response variable is observed at different time points, space points, or both; individual analysis at each point is not recommended because it does not consider the behavior of the response along the time or space and implications thereby. Multivariate analysis, on the other hand, does not regard the existence of a sequence on time or space, and its significance tests are less sensible than those provided by the univariate analysis. In this context we look for correct univariate analysis of variance procedures for situations where time and space are considered as factors. The difficulties in taking time and space as factors arise because they are not randomized, on the contrary of the treatments applied to the experimental plots; that is the reason we have many articles on this subject. To understand the techniques used in this kind of analysis, we need some knowledge about randomization, multivariate analysis and split-plot designs; they are treated on the first three sections while the others are specific for repeated measurements on time and space.

---

<sup>1</sup> Eng. Agr., Ph.D., Embrapa Cerrados e Prof. Adj. da UnB.

## 1 INTRODUÇÃO

Os experimentos casualizados têm sido os preferidos pelos pesquisadores para testar hipóteses ou estimar parâmetros de acordo com a natureza do estudo; o procedimento seguido depende do delineamento e das variáveis resposta que foram observadas. No caso mais freqüente as variáveis são medidas uma vez em cada unidade experimental e utilizadas na análise de variância; a alocação aleatória das parcelas aos tratamentos valida a análise da variância.

Outra situação é quando a variável resposta é observada ao longo do tempo (é medida semanalmente, mensalmente, etc.) ou ao longo do espaço (é medida em varias faixas de profundidade do solo). A análise de variância em cada tempo ignora a correlação entre diferentes tempos e é incompleta no estudo da interação tratamentos x tempo. A alternativa recomendada é considerar o tempo (ou o espaço) com um fator do experimento, colocar seus níveis em sub parcelas e analisá-lo como se ele tivesse sido planejado como um delineamento em parcelas divididas. Entretanto, como os tempos não foram casualizados (eles são sequenciais), os resultados da análise de variância ficam, em principio, comprometidos.

Nesse trabalho serão discutidas quais as conseqüências da casualização e como ela influencia nas inferências; estabelecer-se-á quando é válida a análise de variância de experimentos com dados repetidos no tempo na forma de parcelas divididas, abordar-se-á a alternativa da análise multivariada de variância e finalmente, será introduzida a análise de experimentos com dados repetidos no tempo via método da máxima verossimilhança com a utilização de várias estruturas de correlação entre os tempos.

A parte computacional da análise estatística será desenvolvida com detalhes de forma a permitir que o leitor use a metodologia descrita e para isso usar-se-á programas em linguagem SAS (1996). Acompanha o trabalho um disquete com os programas, dos quais as partes principais serão discutidas no texto.

## 2 INFERÊNCIA E CASUALIZAÇÃO

### 2.1 Definições

Para a descrição organizada de um experimento, precisa-se de definições e elas serão exibidas no decorrer do texto, sempre que necessário. A variável que é controlada pelo pesquisador e cujas características dependem do objetivo do experimento recebe o nome de **tratamento** ou de **fator**; ela vai causar mudança no sistema, que se refletirá no valor da variável resposta. Os tratamentos são estímulos aplicados ao sistema (experimento) que está sob controle parcial.

Os experimentos podem ter o objetivo de comparar tratamentos ou grupos de tratamentos; por exemplo, o interesse pode estar na comparação de cada tratamento com o controle ou entre os tratamentos propostos. Em ensaios de competição de variedades, o objetivo é comparar variedades ou grupos de variedades; esses experimentos são denominados experimentos comparativos, definição introduzida por Anscombe (1954). Outra classe de experimentos compreende aqueles cuja finalidade é a de estimar parâmetros, como no caso da estimação de uma superfície de resposta a níveis de nitrogênio e de fósforo aplicados ao milho. Em algumas situações ambos os propósitos coexistem.

A unidade básica de referência em um experimento é a **parcela** ou unidade experimental; é a entidade que vai dar origem às observações. Pode ser uma área contínua de terra, um animal ou grupo de animais, um vaso, uma bandeja de vermiculita com 100 sementes, depende do experimento. Em experimentos de campo, uma área de terra, em experimentos em casa de vegetação, um vaso com terra, em experimentos em laboratório, uma placa de "petri" ou uma lâmina.

Tratamento é um estímulo que é fornecido à parcela, produz um efeito e as variáveis observadas (respostas) refletem esse efeito;

entretanto, as variáveis ou fatores não controláveis também influenciam-nas e no final, o valor de uma variável resposta encerra a ação dos dois conjuntos de fatores: o controlado pelo pesquisador e o não controlado. Se as parcelas fossem iguais, a influência dos fatores não controlados seria desprezível. Utilizando o procedimento seguido por Kempthorne (1952), seja Y uma variável resposta, como a produção de uma parcela, então a resposta Y correspondente à parcela pode ser escrita da seguinte forma:

$$Y_{ik} = X_i + t_k \quad (2.1)$$

onde,

$Y_{ik}$  é o valor observado de Y na parcela i que recebeu o tratamento k;

$X_i$  é o valor da variável resposta Y na parcela i sem tratamento;

$t_k$  é o efeito do tratamento k, aplicado à parcela i.

No modelo apresentado em (2.1),  $X_i$  o valor de Y na parcela i, é uma propriedade da parcela i ( $X_i$  é chamado de efeito da parcela i); o valor de  $t_k$  para o tratamento k é o mesmo para qualquer parcela i onde ele for aplicado. A diferença entre duas parcelas i e j, por exemplo, que receberam o mesmo tratamento é:

$$Y_{ik} - Y_{jk} = (X_i + t_k) - (X_j + t_k) = (X_i - X_j) \quad (2.2)$$

a quantidade  $(X_i - X_j)$  advém da diferença entre as duas parcelas e o conjunto das diferenças desse tipo é que vai dar origem ao erro experimental. Em algumas áreas de pesquisa, como a física, a influência dos fatores não controláveis é desprezível, porque as parcelas são muito semelhantes e a diferença  $(X_i - X_j)$  é próxima de zero, tornando o experimento mais simples. Isto não ocorre nas ciências biológicas, o que leva à elaboração de uma estratégia com mais recursos no planejamento, na condução e na análise dos resultados, que possibilite estimar a influência dos tratamentos e dos fatores não controláveis ou erro experimental. Essa estratégia dá origem ao delineamento experimental.

Os fatores não controláveis que afetam o comportamento das variáveis resposta em um experimento, derivam da não igualdade das parcelas utilizadas; dois animais não são iguais, duas parcelas no campo não são iguais, duas folhas de uma planta não são iguais, duas árvores da mesma espécie não são iguais e assim, de acordo com a natureza do experimento sempre haverá parcelas diferentes. Em alguma fase do projeto de pesquisa, o pesquisador definirá quais as parcelas do experimento, ele procurará um conjunto homogêneo de parcelas; esse conceito necessita de reflexão. Um conjunto de parcelas é homogêneo quando a diferença entre elas é devido apenas a fatores não controláveis ou variáveis aleatórias. Em experimentos de campo, dificilmente conseguir-se-á homogeneidade em um conjunto de 30 parcelas de 40 m<sup>2</sup> e o dilema do pesquisador é dividir esse grupo em subgrupos homogêneos por meio de fatores controláveis como a declividade, por exemplo, utilizando o conceito de blocagem. Em experimentos com animais, também será difícil obter um conjunto de 30 carneiros que pode ser considerado homogêneo; idade, sexo, peso, por exemplo, são referenciais utilizados para se obter subgrupos homogêneos de animais. Na área biológica, a procura por conjuntos de parcelas persistirá, entretanto, o objetivo não é conseguir parcelas iguais e sim homogêneas, porque é o máximo que se pode atingir.

## **2.2 Repetição e casualização**

Diante das diferenças entre as parcelas devido aos fatores não controláveis - erro experimental - e que influenciam o valor das variáveis resposta, é necessário avaliar o grau dessa influência; para isso precisa-se de pelo menos duas parcelas como o mesmo tratamento, pois a diferença entre elas é uma medida do erro experimental. Observando-se duas parcelas  $i$  e  $j$  que receberam o mesmo tratamento  $k$  e considerando o modelo (2.1) observa-se que  $(X_i - X_j)$  em (2.2) é uma estimativa do erro experimental. Dessa forma, introduz-se o conceito de repetição em experimentos: a repetição do mesmo tratamento possibilita a estimação do erro experimental; como

ele vai ser estimado depende do tipo de delineamento a ser adotado.

Após a definição do conjunto de parcelas ou subconjuntos de parcelas homogêneas do experimento, em algum momento do processo experimental, os tratamentos vão ser aplicados às parcelas; a pergunta que se faz é: qual tratamento será aplicado ou designado para determinada parcela? Até meados da década de 1920, essa decisão era pessoal, baseada em argumentos subjetivos; Fisher (1926) preconizou que essa alocação fosse aleatória, no sentido de que todas as parcelas tivessem a mesma probabilidade de ser alocada a determinado tratamento (dentro de um conjunto homogêneo de parcelas). Essa prática é denominada de **casualização** e no decorrer do tempo passou a ser um dos princípios da experimentação; embora seja utilizada pela maioria dos pesquisadores, constata-se a falta de explanações sobre a real função da casualização: porque deve ser aplicada e quando, e as implicações no modelo e na análise estatística. As funções da casualização foram descritas por Cox (1958), no capítulo 5, e são as seguintes:

1. Permite estimar o erro experimental; a repetição é necessária, mas não é suficiente.
2. Possibilita o uso de testes de hipóteses.
3. Possibilita a estimação não viesada da diferença entre os efeitos de tratamentos.

### ***2.3 Delineamento inteiramente casualizado***

Uma situação experimental em que o pesquisador dispõe de um conjunto suficiente de parcelas homogêneas é ilustrada com o exemplo seguinte.

## Exemplo 2.1

Nematóide-do-cisto é uma praga que ataca a cultura da soja; na busca de soluções, um programa experimental está em andamento na Embrapa Cerrados (sob responsabilidade do pesquisador Ravi Sharma) e uma das linhas de pesquisa é o controle biológico, por meio de bactérias que atacam o nematóide e diminuem sua população. Para se testar a eficiência de cinco raças de bactérias no controle da população no solo, conduziu-se um experimento que incluiu os seguintes detalhes. Uma porção de terra foi homogeneizada, esterilizada, dividida em 30 vasos e cada vaso recebeu uma infestação de nematóide (esperando-se que diferentes parcelas recebessem infestação semelhante). Os 30 vasos são as parcelas e pela forma de obtenção o conjunto é homogêneo: os cinco tratamentos foram designados aleatoriamente aos vasos, cada um com seis repetições e foram colocados sobre uma mesa na casa de vegetação.

O delineamento – estratégia amostral – utilizado no experimento do exemplo 1.1, é denominado de **completamente ao acaso** ou **inteiramente casualizado**; o nome encerra o significado, isto é, não há restrições para a casualização e o conjunto de parcelas homogêneas é único. O modelo estatístico usualmente adotado para representar  $Y_{ik}$  o valor de uma variável resposta na parcela  $i$  que recebeu o tratamento  $k$  é:

$$Y_{ik} = \mu + t_k + e_{ik} \quad (2.3)$$

onde,

$\mu$ : média geral.

$t_k$ : efeito do tratamento  $k$ .

$e_{ik}$ : erro experimental, com distribuição normal de média zero, variância  $\sigma^2$  e independentes. Esse modelo é referido como o modelo normal clássico.

A análise de um experimento estatisticamente delineado constitui um exercício das várias técnicas existentes: análise de variância, análise de variância robusta, análise de variância espacial, análise de variância multivariada e outras. Por onde começar e quando parar depende de cada experimento; entretanto, se o planejamento foi adequado, todas, em princípios, podem ser aplicadas, condicionadas aos objetivos do experimento, natureza do experimento (competição de variedades de milho, ensaio com dietas em animais, testes para pesticidas) e existência de aplicativos de estatística (*software*). No momento, a análise de variância é a que vai ser utilizada, por ser a mais importante em análise de experimentos. Embora mais de uma variável resposta seja observada nos experimentos, somente uma será referenciada no modelo; por outro lado, a variável resposta considerada para análise é numérica, no sentido de que representa uma quantidade.

A análise de variância (e outras análises) de um experimento completamente casualizado está embasada no modelo 2.3; antes de examiná-la é necessário definir o modelo a ser adotado.

## **2.4 Modelo aleatório**

O desenvolvimento que se segue é específico para o delineamento inteiramente casualizado, e a extensão para outros delineamentos é encontrada em Hinkelmann e Kempthorne (1994); entretanto, um caso é suficiente para que o papel da casualização seja colocado da forma correta. Nesse delineamento, existe um conjunto de parcelas homogêneas e os tratamentos são alocados aleatoriamente a elas.

Considere então um experimento com  $v$  tratamentos,  $r$  repetições e que o pesquisador dispõe de  $N = rv$  parcelas homogêneas; o número de repetições pode variar de tratamento para tratamento sem causar modificações nos resultados que vão ser obtidos, entretanto, é mais simples considerar o caso onde todos os tratamentos têm o mesmo número de repetições. Seja  $Y$  a variável resposta que

será observada em cada parcela (matéria seca, número de sementes germinadas, população de nematóides). O primeiro passo é formular uma representação para a resposta em cada parcela e se nenhum tratamento for aplicado às parcelas, o valor da variável resposta dependerá somente das condições da parcela  $i$ . Seja  $X_i$  o valor da resposta  $Y$  na parcela  $i$  ( $i = 1, 2, \dots, N$ ), na ausência de tratamento (ou na presença de um tratamento padrão), isto é,  $X_i$  é o produto das próprias características da parcela  $i$ . Quando o tratamento é aplicado a parcela  $i$  espera-se que haja um efeito sobre as condições da parcela com reflexos no valor de  $Y$ ; seja então  $t_k$  o efeito do tratamento  $k$  na resposta  $Y$ . Com essas definições chega-se à representação mais simples de  $Y_{ik}$  o valor de  $Y$  na parcela  $i$  que recebeu o tratamento  $k$ , ou seja:

$$Y_{ik} = X_i + t_k \quad (2.4)$$

Quando o efeito de qualquer tratamento  $k$ ,  $t_k$  é o mesmo para toda parcela  $i$ , o modelo (2.4) é denominado **modelo aditivo**. Para melhor entendimento do modelo, seja:

$$\bar{X} = \frac{1}{N} \sum_i X_i \quad (2.5)$$

a média dos  $X_i$ 's e assim:

$$Y_{ik} = \bar{X} + t_k + (X_i - \bar{X}); \quad (2.6)$$

na nova representação de  $Y_{ik}$  (apesar de os modelos (2.4) e (2.6) serem iguais), observa-se que se todas as parcelas fossem iguais então  $(X_i - \bar{X})$  seria 0 (zero) para toda parcela  $i$ ; todavia, nos experimentos aqui considerados, isto não ocorre e  $(X_i - \bar{X})$  é o erro da parcela  $i$ .

O segundo passo é associar o modelo (2.6) ao processo de casualização; no modelo apresentado em (2.4) não há tal associação. **Certamente, as  $N$  parcelas homogêneas não constituem uma**

**amostra aleatória de um conjunto maior de parcelas (população);** o exemplo 2.1 evidencia esse fato: os trinta vasos não foram retirados aleatoriamente de uma população de vasos. Casualizar os tratamentos consiste em alocar r parcelas para cada tratamento, de tal forma que em cada parcela tenha a mesma probabilidade de receber o tratamento k. Como qualquer parcela pode receber qualquer tratamento, o que vai ocorrer numa parcela i devido à casualização, pode ser representada por variáveis aleatórias; assim na parcela i, o tratamento 1 pode ser aplicado ou não, ou o tratamento 2, e assim por diante. Considerando esse contexto probabilístico, o modelo que representa uma observação na parcela (ik) é o seguinte:

$$Y_{ik} = \mu + t_k + e_{ik} \quad (2.7)$$

onde os erros  $e_{ik}$  têm as seguintes propriedades:

$$1) \text{ Média: } E(e_{ik}) = 0 \quad (2.8)$$

$$2) \text{ Variância: } \text{Var}(e_{ik}) = \frac{1}{N} \sum (X_i - \bar{X})^2. \quad (2.9)$$

Fazendo-se

$$\sigma^2 = \frac{\sum_i (X_i - \bar{X})^2}{N-1}, \quad (2.10)$$

então:

$$3) \text{ Var } (e_{ik}) = \frac{N-1}{N} \sigma^2 \quad (2.11)$$

4) Covariâncias:

$$\text{Cov } (e_{ik}, e_{jl}) = -\frac{1}{N^2} \sigma^2 \quad (2.12)$$

As covariâncias não nulas, em (2.12), determinam que as observações em diferentes parcelas são correlacionadas; essa é uma consequência do modelo aditivo e da casualização. O modelo descrito de (2.7) até (2.12) é um modelo induzido pela casualização.

## 2.5 Inferência experimental

Nas seções anteriores, foram definidos os modelos: normal clássico, com variâncias homogêneas e erros independentes e o modelo aleatório induzido pela casualização; conduzido um experimento e obtidas as observações, o próximo passo é estabelecer julgamentos a partir das observações, isto é, inferir; inferência em estatística utiliza estimativas, testes de hipóteses e intervalos de confiança a partir de modelos probabilísticos aplicados às observações e, quando estas são originadas por meio de experimentos, o processo denomina-se inferência experimental. Nesta seção, será ilustrada uma metodologia para fazer inferências por meio de experimentos casualizados, e o modelo a ser seguido é o aleatório induzido pela casualização. Antes porém, é edificante uma síntese sobre os dois modelos.

O modelo normal clássico considera que as observações das parcelas são independentes e portanto não correlacionadas, isto é,  $e_{11}, e_{12}, \dots, e_{nv}$  em 1.3, é uma amostra aleatória de uma população infinita. Entretanto, pelo que já foi exposto, não existe tal amostra aleatória em experimentação: a população de parcelas é finita. Pode-se até argumentar que as  $N$  parcelas constituem uma amostra de uma possível população de parcelas, entretanto, não é uma amostra aleatória, pela forma como os pesquisadores costumam selecioná-las para determinado experimento. Por sua vez, a casualização implica a existência de correlação entre as parcelas, não havendo portanto, independência. A distribuição de  $X_i$  é desconhecida e no que se segue ela não é necessária. Nota-se, portanto, que os dois modelos são diferentes. Por que então o modelo normal é utilizado? Para responder a essa pergunta é necessário analisar um experi-

mento pelos dois modelos: modelo normal e modelo induzido pela casualização. Observa-se pelo que foi desenvolvido que o modelo normal é obtido por pressuposição enquanto o aleatório é por construção; no primeiro caso, não há influência da casualização e no segundo ela é utilizada para definir uma estrutura de covariância entre as parcelas (observações) e o modelo matemático para cada observação.

A comparação entre dois tratamentos,  $k_1$  e  $k_2$ , consiste em estimar a diferença  $(t_{k_1} - t_{k_2})$ ; em ambos os modelos o estimador é  $(\bar{Y}_{k_1} - \bar{Y}_{k_2})$  e tem variância  $2\frac{\sigma^2}{r}$ , onde  $\bar{Y}_{k_1}$  é a média do tratamento  $k_1$  no experimento. Para se estimar um contraste  $\sum_k \alpha_k t_k$ , em ambos os modelos, o estimador é  $\sum_k \alpha_k \bar{Y}_k$ , com variância  $\frac{\sum_k \alpha_k^2}{r} \sigma^2$  (se  $\sum_k \alpha_k = 0$ , então  $\sum_k \alpha_k t_k$  é um contraste entre tratamentos). Embora  $\sigma^2$  tenha definições diferente nos dois modelos, ele é um parâmetro desconhecido e tem de ser estimado pela análise de variância.

A análise de variância de experimentos é conhecida por todos os pesquisadores que conduzem experimentos e para não modificar essa base, deve-se examinar os resultados da análise, segundo o modelo probabilístico gerado pela casualização; como exemplo, considere um experimento para comparar três estirpes de *Rhizobium* (bactérias fixadoras de nitrogênio em leguminosas) em soja, conduzido em casa de vegetação e inteiramente casualizados. As parcelas, os tratamentos e os valores da variável resposta Y, número de nódulo por parcela, encontram-se na Tabela 2.1 a seguir:

**TABELA 2.1** Parcelas, tratamentos e valores de Y.

Parcela	1	2	3	4	5	6	7	8	9
Trat	A	A	A	B	B	B	C	C	C
Y	99	86	104	98	110	110	133	126	127

A análise de variância com o modelo normal é obtido por meio do método dos quadrados mínimos aplicado em (2.3) e apresenta os seguintes resultados :

F.V.	GL	SQ	QM	F
Tratamentos	2	1652,66	826,33	16,67
Erro	6	297,33	49,55	
<b>Total</b>	<b>8</b>	<b>1950,00</b>		

A estimativa de  $\sigma^2$  o erro experimental é dada pelo quadrado médio do erro e é 49,55 e o teste F leva à rejeição da hipótese nula ( $H_0$ ) de que os efeitos de tratamentos são nulos, uma vez que pela tabela de F,  $\text{Prob}(F \geq 16,67) = 0,0035 = 0,35\%$ . A estatística F reflete a evidência dos resultados contra  $H_0$  e sua distribuição sob  $H_0$  é fundamental para o teste de hipótese. No modelo normal, a distribuição dessa estatística é conhecida (distribuição F com  $m$  e  $n$  graus de liberdade); a pergunta principal é: Qual a distribuição da estatística F da análise de variância, sob  $H_0$ , quando o modelo é o aleatório? Para isto, aplica-se o teste de permutação, introduzido por Fisher (1935).

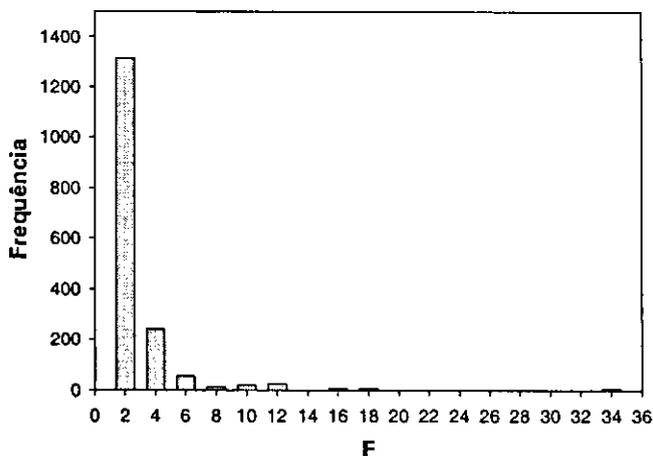
O experimento em análise tem três tratamentos, cada um com três repetições e são alocados aleatoriamente às nove parcelas; o número de diferentes casualizações é de 1680. Isto é, existem 1680 possíveis resultados da casualização e uma delas o pesquisador usou. O raciocínio do teste de permutação é o seguinte: Sob  $H_0$ , os resultados do experimento seriam os mesmos em qualquer casualização, ou seja, o valor de Y em cada parcela não dependeria do tratamento e sim das condições da própria parcela, isto é, seria  $X_i$  no modelo apresentado em (2.4); se isto é verdade, é indiferente o local onde o tratamento foi aplicado e assim pode-se considerar todas as 1680 casualizações, implicando 1680 análises de variância, mantendo fixo o valor observado em cada parcela, apenas tro-

cando de tratamento de acordo com as possíveis casualizações. A Tabela 2.2 abaixo contém os valores fixados às parcelas :

**TABELA 2.2 Efeito das parcelas sob hipótese nula.**

1	99	2	86	3	104
4	98	5	110	6	110
7	133	8	126	9	127

Cada casualização diferente gera um valor de F e assim com os 1680 valores constrói-se a distribuição exata da estatística sob  $H_0$  e seguindo o modelo aleatório, a Figura 2.1 representa a distribuição exata de F.



**FIG. 2.1 Distribuição de F segundo o modelo aleatório.**

Para calcular  $\text{Prob}(F \geq 16,07)$  (16,07 foi o valor de F obtido no experimento) conta-se o número de valores maiores ou iguais a 16,07 e divide-se por 1680. Como somente 12 valores apresentam essa característica,  $\text{Prob}(F \geq 16,07) = 12/1680 = 0,0071 = 0.71\%$ .

O valor de F observado no experimento está na região de rejeição (considerando  $\alpha=0,05$ ) e  $H_0$  é rejeitada; esse teste é exato e é independente de normalidade e de erros não correlacionados. Entretanto, mesmo com conclusões iguais para os dois modelos, o nível de significância do aleatório foi aproximadamente duas vezes maior do que o do normal. De fato, nos estudos de simulação de Hinkelmann e Kempthorne (1994), para  $r=4$  e  $v=8$ , os níveis de significância do modelo aleatório são ligeiramente maiores do que o do modelo normal; como no exemplo em questão, o experimento é pequeno ( $r=3$  e  $v=4$ ), essa diferença acentuou-se.

A comparação de dois tratamentos é feita pelo teste de Student no modelo normal e nesse experimento as médias foram:  $A=96,3$ ,  $B=106,0$  e  $C=128,8$ . A estatística “t” para testar  $H_0: C=B$ ,  $H_a: C \neq B$  é:

$$t = \frac{(128,8 - 106,0)}{\sqrt{(\frac{2}{3})49,55}} = 3,94 \quad (2.13)$$

que, sob  $H_0$ , tem distribuição de t de Student, com 6 graus de liberdade (gl), a  $\text{Prob}(t \geq 3,94) = 0,0076$  e rejeita-se  $H_0$ . Pelo modelo aleatório, calcula-se 1680 valores de t e procede-se como antes; feitos os cálculos, a  $\text{Prob}(t \geq 3,94) = 0,0077$ .

Em ambos os casos, o teste F e o teste t, os resultados do modelo normal são aproximações daqueles do modelo aleatório; por outro lado, como já foi visto anteriormente, as estimativas de contrastes de tratamentos e suas variâncias são iguais. Outro ponto importante é que  $\sigma^2$ , embora tenha definições diferentes nos dois modelos, é estimado pelo quadrado médio do erro na análise de variância, em ambos os modelos, o que leva também os contrastes de tratamento terem a mesma estimativa de variância.

No exemplo em análise, nenhuma distribuição conhecida foi assumida para os dados do experimento quando sob o modelo

aleatório, enquanto as pressuposições de independência e normalidade referem-se ao modelo normal; independência entre as parcelas não ocorre se houver casualização, e esta não leva à normalidade. Utilizando a linguagem SAS (1986) simulou-se um experimento com as observações tendo distribuição de Poisson, isto é, os valores de  $x_i$  em (2.4) constituem os elementos de uma amostra gerada pela distribuição de Poisson; foi estabelecida uma estrutura aditiva e os dados e a análise de variância do modelo normal estão na Tabela 2.3.

**TABELA 2.3** Valores gerados na simulação e análise de variância.

Tratamento	A	A	A	B	B	B	C	C	C
Y	26	23	20	29	23	23	30	32	37
<b>F. V.</b>	<b>GL</b>	<b>SQ</b>	<b>QM</b>	<b>F</b>	<b>P &gt; F</b>				
Tratamento	2	168	84	7,41	0,024				
Erro	6	68	11,33						
Total	8	236							

A Prob ( $F \geq 7,41$ ) obtida pelo modelo aleatório é 0,036. Os valores de Y não foram transformados para  $\sqrt{y}$ , procedimento recomendado para a distribuição de Poisson, e os resultados evidenciam que o modelo normal funciona. Os exemplos apresentados foram casos particulares, mas conclui-se o seguinte sobre o modelo normal:

- 1) É simples e demanda pouco cálculo;
- 2) É uma aproximação aceitável do modelo aleatório;
- 3) Sob aditividade, não há necessidade de transformações para se atingir normalidade.

Os estudos teóricos e de simulação vão continuar, mas os que existem até hoje levam à utilização do modelo normal na análise de experimentos casualizados e a inferência experimental continuará sendo feita através dele, embora os resultados sejam aproximações daqueles que seriam obtidos pelo modelo induzido pela casualização; deve-se compreender, entretanto, que a variável resposta não precisa ter distribuição normal (embora possa ter), que as observações são correlacionadas e que não existe uma amostra aleatória de parcelas. Por outro lado, a existência de observações discrepantes (*outliers*) e a heterogeneidade da variância, são sintomas de não aditividade, que comprometem a validade da análise de variância; nesses casos, deve-se usar transformações estabilizadoras da variância, como as do tipo Box e Cox (1964) ou no caso de proporções, arco-seno  $\sqrt{p}$ . Elas contribuem para melhorar a aditividade do modelo quando a resposta é a variável transformada.

Finalizando, sob aditividade, os dois modelos produzem resultados aproximados como os do teste F e *t* no exemplo desta seção e os cálculos no modelo normal são simples e poucos; estas são razões pelas quais o modelo normal é utilizado, entretanto, quando se trata de inferência experimental, o de referência é aquele gerado pela casualização.

### 3 DELINEAMENTOS EM PARCELAS DIVIDIDAS

Os experimentos fatoriais podem ser planejados e conduzidos em quaisquer dos delineamentos básicos: completamente casualizado, blocos incompletos, delineamentos em linhas e colunas. Cada combinação dos níveis dos fatores é um tratamento, e as parcelas são alocadas aleatoriamente aos tratamentos, respeitando-se a estrutura de parcelas de cada experimento. Entretanto, existem experimentos fatoriais em que a casualização é condicionada por objetivos específicos do experimento ou por condições experimentais; os experimentos para o estudo da interação entre variedade e espaçamento eram planejados em parcelas divididas, com as variedades nas parcelas e as densidades nas subparcelas; devido ao manejo,

um fator deve ser aplicado em parcelas maiores e o outro em menores; tomando-se como exemplo o delineamento em blocos ao acaso, os níveis do fator A são aplicados às parcelas e estas são divididas em subparcelas, onde são aleatoriamente alocados os níveis do fator B. Assim tem-se três fases da casualização: a dos blocos no campo experimental, a das parcelas dentro de blocos e a das subparcelas dentro de cada parcela. São exemplos clássicos de tratamentos de parcelas: variedades, níveis de irrigação, tipos de arado; e de tratamentos de subparcelas: densidade, espaçamento, nutrientes. As razões para esta estrutura variam, mas a consequência é a mesma: as estimativas envolvendo os tratamentos de subparcelas e a interação entre estes e os tratamentos das parcelas, têm maior precisão; analogamente, os respectivos testes de hipóteses também são mais sensíveis.

O modelo normal clássico deste delineamento é (considerando blocos casualizados):

$$Y_{ijk} = \mu + R_i + A_j + (RA)_{ij} + B_k + (AB)_{jk} + e_{ijk}, \quad (3.1)$$

onde,

$Y_{ijk}$ : variável resposta observada no nível k de B, nível j de A, no bloco i.

$\mu$ : média geral

$R_i$ : efeito do bloco i

$A_j$ : efeito do nível j do tratamento A

$(RA)_{ij}$ : erro aleatório, com média zero e variância  $\sigma_A^2$ , independentes, com distribuição normal

$B_k$ : efeito do nível k do tratamento B

$(AB)_{jk}$ : interação entre A e B

$e_{ijk}$ : erro aleatório com média zero e variância  $\sigma_B^2$ , independentes, com distribuição normal e independentes de  $(RA)_{ik}$

A ausência da interação (RA), blocos x tratamentos de parcelas, dá origem ao erro(a)

A variância da comparação entre dois tratamentos das parcelas é dada por:

$$\text{Var}(\bar{Y}_{.j} - \bar{Y}_{.l}) = 2 \frac{(\sigma_B^2 + b\sigma_A^2)}{rb}, \quad (3.2)$$

onde  $b$  é o número de tratamentos das subparcelas e  $r$  é o número de blocos; o numerador de (2.2) é estimado pelo erro (a), na análise da variância, e assim, a variância estimada em (3.2) é obtida pela substituição do numerador pelo erro (a). A variância da comparação, entre dois tratamentos das subparcelas é dada por:

$$\text{Var}(\bar{Y}_{.k} - \bar{Y}_{.l}) = 2 \frac{\sigma_B^2}{ra}, \quad (3.3)$$

onde  $a$  é o número de tratamentos das parcelas;  $\sigma_B^2$  é estimada pelo erro(b) da análise da variância. A variância da comparação entre dois tratamentos das subparcelas dentro de um tratamento de parcela é dada por:

$$\text{Var}(\bar{Y}_{.jk} - \bar{Y}_{.jl}) = 2 \frac{\sigma_B^2}{r}. \quad (3.4)$$

A variância da comparação entre dois tratamentos de parcelas dentro de um tratamento de subparcela é dada por:

$$\text{Var}(\bar{Y}_{.jk} - \bar{Y}_{.lk}) = 2 \frac{(\sigma_A^2 + \sigma_B^2)}{r} \quad (3.5)$$

e nesse caso não há uma estimativa direta da análise da variância para  $\sigma_A^2 + \sigma_B^2$  e temos que construir um erro sintético;  $\sigma_A^2 + \sigma_B^2$  em (3.5) é estimado por

$$\frac{\text{Erro}(a) + (b-1)\text{Erro}(b)}{b} \quad (3.6)$$

O grau de liberdade associado ao erro (3.6) é dado pela aproximação de Satterthwaite (1946),

$$gl = \frac{[Erro(a) + (b-1)Erro(b)]^2}{\frac{[Erro(a)]^2}{gl(a)} + \frac{[(b-1)Erro(b)]^2}{gl(b)}} \quad (3.7)$$

onde  $gl(a)$  e  $gl(b)$  são os graus de liberdade do Erro(a) e Erro(b) respectivamente; as variâncias e os graus de liberdade são necessários para a construção de intervalos de confiança para a diferença entre dois tratamentos ou para testar hipóteses.

O esquema de análise de variância com as fontes de variação e graus de liberdade estão na Tabela 3.1.

**TABELA 3.1** Esquema da análise de variância de um delineamento em parcelas divididas.

Fontes de variação	GL
Blocos	r-1
Tratamentos(A)	a-1
Erro(a)	(r-1)(a-1)
Tratamento(B)	b-1
Interação. A x B	(a-1)(b-1)
Erro(b)	a(r-1)(b-1)
Total	rab-1

Caso o delineamento seja outro, haverá modificação da primeira parte da tabela no que concerne a fontes de variação, mas a segunda será a mesma. O modelo acima, normal clássico, é uma aproximação aceitável do modelo induzido pela casualização. Ao falar em parcelas divididas é conveniente comentar as extensões desse delineamento; assim, as subparcelas podem também ser divididas, dando origem às subsubparcelas, e os c níveis de um terceiro fator C, são alocados aleatoriamente a elas e tem-se o delineamento em parcelas subdivididas (split-split plot); essa subdivisão pode continuar para os níveis de um fator D e assim por diante. As conseqüências são as mesmas da anterior: a cada nível de subdivi-

são, a precisão das comparações aumenta e aumentam também as complicações na comparação entre dois tratamentos das subparcelas dentro de um tratamento de subsubparcela (Cochran & Cox, 1957).

Um delineamento freqüentemente confundido com o de parcelas divididas é o delineamento em faixas ou em blocos divididos. Quando esses dois tipos de tratamento envolvem maquinária pesada ou são de difícil manejo, os níveis dos dois fatores precisam de parcelas grandes e assim divide-se o bloco em linhas e colunas, sorteiam-se os níveis do fator A nas colunas e os níveis do fator B nas linhas e isto independentemente em cada bloco. O modelo matemático desse delineamento é o seguinte:

$$Y_{ijk} = \mu + R_i + A_j + (RA)_{ij} + B_k + (RB)_{ik} + (AB)_{jk} + e_{ijk} \quad (3.8)$$

A ausência da interação blocos x tratamentos B (RB), origina o Erro(b) e a ausência da interação blocos x A x B (RAB) o Erro(c), representado aqui por  $e_{ijk}$ . O esquema da análise de variância está na Tabela 3.2. Vale ressaltar que a análise de variância de experimentos em blocos é válida quando não há interação dos tratamentos com blocos; essa interação não é exatamente zero, e os desvios são devidos à diferença entre as parcelas dentro do bloco, de natureza aleatória, e por isso origina um erro.

**TABELA 3.2** Esquema da análise de variância de um delineamento em blocos divididos.

Fontes de variação	GL
Blocos	r-1
Tratamentos(A)	a-1
Erro(a)	(r-1)(a-1)
Tratamentos(B)	b-1
Erro(b)	(r-1)(b-1)
Inter A x B	(a-1)(b-1)
Erro(c)	(r-1)(a-1)(b-1)
<b>Total</b>	<b>rab-1</b>

## **4 EXPERIMENTOS COM MEDIDAS REPETIDAS NO TEMPO E NO ESPAÇO.**

### **4.1 Exemplos**

**Exemplo 4.1** Um experimento foi realizado para estudar a produção de matéria seca de uma forrageira, sob quatro sistemas de manejo; foram feitos cinco cortes nos tempos  $t_1, t_2, t_3, t_4$  e  $t_5$  e as variáveis relevantes foram observadas em cada corte.

**Exemplo 4.2** Um experimento com laranjeiras foi conduzido para se estudar o efeito de vários métodos de poda na evolução da produção anual; em cada ano, foram registradas a produção e a incidência de pragas.

**Exemplo 4.3** Um experimento com cinco tratamentos (dietas), para se estudar o efeito da redução do nível de proteína em rações para leitões e da sua suplementação com aminoácidos; o peso e outras medidas foram registrados no início do experimento e a cada sete dias durante dois meses.

**Exemplo 4.4** Um experimento será realizado para se avaliar o efeito de diferentes sistemas de cultivo de cobertura do solo, na dinâmica da biomassa e da atividade microbiana ao longo de seis anos. As variáveis que as caracterizam serão observadas em cada estação e também em cinco profundidades.

Todos esses experimentos têm em comum o registro da mesma variável em tempos sucessivos (repetidos no tempo) e o último, em espaços consecutivos (repetidos no espaço - profundidades).

### **4.2 Análise**

A análise de experimentos, com medidas repetidas no tempo (ou no espaço), pode ser semelhante àquela de um experimento que tivesse sido planejado com parcelas divididas, considerando as

medidas no tempo como os tratamentos de subparcelas; o delineamento do experimento pode ser qualquer um. Ocorre que a sequência dos tempos  $t_1, t_2, \dots, t_v$  não é casualizada e, portanto, não se conhece a estrutura de covariância das observações no tempo. Para complementação das idéias, considere um delineamento em blocos ao acaso, com o seguinte modelo:

$$Y_{ij} = \mu + b_i + \tau_j + e_{ij} \quad (4.1)$$

onde:

$\mu$  é a média geral;

$b_i$  é o efeito de bloco;

$\tau_j$  é o efeito de tratamento;

$e_{ij}$  é o erro experimental;

a estrutura de variâncias e covariâncias dos  $Y_{ij}$ 's é dada pela casualização, mas se existem observações nos tempos  $t_1, t_2, \dots, t_v$ , para cada parcela (ij) do experimento têm-se as observações  $Y_{ij1}, Y_{ij2}, \dots, Y_{ijv}$ ; as variâncias e covariâncias dessas observações podem ser representadas pela matriz:

$$\Sigma = \begin{bmatrix} \text{var}(Y_{ij1}) & \text{cov}(Y_{ij1}, Y_{ij2}) & \dots & \text{cov}(Y_{ij1}, Y_{ijv}) \\ \text{cov}(Y_{ij2}, Y_{ij1}) & \text{var}(Y_{ij2}) & \dots & \text{cov}(Y_{ij2}, Y_{ijv}) \\ \dots & \dots & \dots & \dots \\ \text{cov}(Y_{ijv}, Y_{ij1}) & \text{cov}(Y_{ijv}, Y_{ij2}) & \dots & \text{var}(Y_{ijv}) \end{bmatrix} \quad (4.2)$$

ou,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1v} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2v} \\ \dots & \dots & \dots & \dots \\ \sigma_{v1} & \sigma_{v2} & \dots & \sigma_{vv} \end{bmatrix} \quad (4.3)$$

Estudando o assunto, Huynh & Feldt (1970) concluíram que se em (4.3)

$$\begin{aligned}\sigma_{ii} &= 2\gamma_i + \lambda, \\ \sigma_{ij} &= \gamma_i + \gamma_j \text{ se } i \neq j,\end{aligned}\tag{4.4}$$

para constantes  $\gamma_i$  e  $\lambda$ , então o teste F na análise de variância, como se o experimento fosse em parcelas divididas, é válido. Essas condições para a matriz  $\Sigma$  são conhecidas como condições (H-F); a dificuldade está em saber se no experimento em questão, as condições (H-F) existem. Embora estas sejam as condições mais referidas, o assunto começou a ser estudado por Box (1954); Box estabeleceu uma medida de afastamento das condições (H-F) (mesmo que na época elas ainda não tivessem sido formuladas) através dos elementos da matriz  $\Sigma$ , definida pelo, parâmetro  $\theta$ , onde:

$$\theta = \frac{v^2(\bar{\sigma}_{ii} - \bar{\sigma}_{..})^2}{(v-1)(\sum_i \sum_j \sigma_{ij}^2 - 2v \sum_i \bar{\sigma}_{i.}^2 + v^2 \bar{\sigma}_{..}^2)}\tag{4.5}$$

onde,

$$\bar{\sigma}_{..} = \frac{1}{v^2} \sum_i \sum_j \sigma_{ij}$$

$$\bar{\sigma}_{i.} = \frac{1}{v} \sum_j \sigma_{ij}$$

$$\bar{\sigma}_{ii} = \frac{1}{v} \sum_i \sigma_{ii}$$

Para  $v > 2$ ,  $\frac{1}{(v-1)} \leq \theta \leq 1$ . Quanto mais  $\theta$  se afasta de 1, mais se afasta das condições (H-F). Observa-se então, que se  $v=2$ , as

condições (H-F) sempre existem e o experimento pode ser analisado como parcelas divididas. Para ilustrar a recomendação de Box, considere um experimento em blocos ao acaso, com seis tratamentos e cinco repetições, onde a variável de interesse foi observada em cinco ocasiões. O esquema da análise de variância como se o experimento fosse em parcelas divididas está na Tabela 4.1.

**TABELA 4.1** Esquema da análise da variância do experimento.

Fontes de Variação	GL	GL(Box)
Blocos	5	5
Tratamentos	5	5
Erro(a)	25	25
Tempos	4	1
Tempos x Tratamentos	20	5
Erro(b)	120	30
<b>Total</b>	<b>179</b>	

A recomendação de Box é dividir os graus de liberdade para Tempos, Tempos x Tratamentos e Erro(b) por  $(v-1)$ , que no caso é 4 (como está na terceira coluna da tabela cima), e aplicar o teste F com esses novos valores. Para orientar essa recomendação, Box propôs um teste de hipótese da forma:  $H_0: \theta = 1$ ,  $H_a: \theta < 1$ ; entretanto, com o decorrer do tempo surgiram alternativas.

Um exame do procedimento de Box indica que os graus de liberdade para Tempos e Tempos x Tratamentos foram multiplicados por  $1/(v-1)$ , o menor valor que  $\theta$  pode ter; se  $\theta$  for maior então o teste passa a ser conservador e somente grandes diferenças vão ser significativas. Greenhouse & Geisser (1959) Hunch & Feldt (1970) conseguiram estimadores para  $\theta$ , conhecidos com estimadores (G-G) e (H-F); eles são usados para corrigir os graus de liberdade e aparecem em aplicativos de estatística como o SAS (1997).

Quando a resposta no tempo é semelhante à do Exemplo 4.1, podem-se utilizar modelos de regressão no tempo; se a matéria seca no tempo  $t_i$  é dada por  $y_i$ , então  $y_1$ ,  $y_1 + y_2$ ,  $y_1 + y_2 + y_3$  etc. são quantidades que crescem e portanto pode-se ajustar uma regressão no tempo, para se estimar, por exemplo, taxas de crescimento para cada sistema. Deve-se optar por equações simples, entretanto, isto depende da relação da resposta acumulada e o tempo.

Ocorre, ainda, o caso em que a matriz de covariâncias apresenta outro padrão ou não tem nenhum; por exemplo, as correlações podem diminuir ao longo do tempo ou do espaço. O procedimento MIXED do SAS (1997) permite analisar experimentos com medidas repetidas no tempo, admitindo vários tipos de matrizes de variâncias e covariâncias, inclusive uma matriz  $\Sigma$  qualquer; os resultados dessa análise vão subsidiar a decisão sobre qual matriz  $\Sigma$  utilizar.

Em alguns experimentos, entretanto, a resposta é observada no espaço e no tempo, como no exemplo 4.4; neste, os procedimentos são mais elaborados, mas as técnicas são as mesmas; em princípio, quer-se analisar o experimento como se ele tivesse sido planejado como parcelas subdivididas (*split-split-plot*), sendo os diferentes espaços os tratamentos das subparcelas e os tempos os tratamentos das subsubparcelas. Para consolidar conceitos e ilustrar operações serão utilizados os recursos do SAS na análise de um experimento com dados cedidos pela pesquisadora lêda Mendes de Carvalho, da Embrapa Cerrados (dados parcialmente modificados).

## 5 ANÁLISE MULTIVARIADA

Pode ocorrer que o afastamento das condições H-F para a análise univariada seja grande e nesse caso, tem-se a opção da análise de variância multivariada; aqui, as variáveis  $Y_1, Y_2, \dots, Y_v$ , observadas nos tempos 1, 2, ..., v, também têm uma matriz de variâncias e covariâncias  $\Sigma$ , desconhecida e sem restrições envolvendo

seus elementos (apenas sendo definida positiva). Na análise multivariada, as  $v$  variáveis são consideradas como  $v$  respostas, com matriz de variância  $\Sigma$ . A análise de variância multivariada está contida num capítulo maior de estatística, denominado Análise Multivariada, do qual faz parte grande elenco de outras análises. Para sua utilização em dados repetidos no tempo, vamos introduzir as idéias principais.

Considere um experimento com feijão, em blocos ao acaso, relatado em Seber (1984); os tratamentos consistiam em seis práticas de combate a *Leaf Miner* e as variáveis observadas foram:

$Y_1$  número de lesões por folha,

$Y_2$  peso de feijão por parcela,

$Y_3$  proporção de folhas infectadas (arco-seno ( $\sqrt{Y_3}$ )).

Numa análise de variância univariada, a hipótese nula ( $H_0$ ), de que os efeitos dos tratamentos são nulos, é testada independentemente em cada variável. Na multivariada, a hipótese nula passa a ser: os efeitos de tratamentos são nulos para as três variáveis e o teste é um só. Nessa análise, é aproveitada a correlação que existe entre as variáveis resposta; por exemplo,  $\text{Corr}(Y_1, Y_2) = -0,005$ ,  $\text{Corr}(Y_1, Y_3) = 0,264$  e  $\text{Corr}(Y_2, Y_3) = -0,185$ . A análise de variância para cada variável forneceu os seguintes valores de F:  $Y_1$ ,  $F = 9.25$  (rejeita-se  $H_0$ );  $Y_2$ ,  $F = 1.29$  (aceita-se  $H_0$ );  $Y_3$ ,  $F = 6.28$  (rejeita-se  $H_0$ ). A análise de variância multivariada utiliza matrizes de soma de quadrados e produtos para cada fonte de variação (em vez de soma de quadrados somente como na análise de variância univariada). No caso em questão, as matrizes relevantes são as de tratamentos e do erro,  $T$  e  $E$ , respectivamente, e são as seguintes:

$$T = \begin{bmatrix} 6,617 & -0,858 & 1,684 \\ -0,858 & 0,678 & -0,219 \\ 1,684 & -0,219 & 0,707 \end{bmatrix} \quad (5.1)$$

$$E = \begin{bmatrix} 1,031 & -0,007 & 0,156 \\ -0,007 & 1,578 & -0,135 \\ 0,156 & -0,135 & 0,338 \end{bmatrix} \quad (5.2)$$

Na diagonal, estão as somas de quadrados para cada variável e fora dela as soma de produtos; por exemplo, -0,858 é a soma de produtos para tratamentos entre  $Y_1$  e  $Y_2$ . Também está definida uma matriz para blocos, chamada de **B**, mas que não será usada aqui. A análise de variância multivariada é apresentada de forma semelhante à univariada e consta na Tabela 5.1.

**TABELA 5.1** Esquema da análise de variância multivariada.

Fontes de variação	GL	Matriz
Blocos	3	<b>B</b>
Tratamentos	5	<b>T</b>
Erro	15	<b>E</b>
<b>Total</b>	<b>23</b>	

Os quatro testes mais conhecidos para esse caso são: teste de Wilks, de Hotelling, de Pillai e de Roy; detalhes desses testes podem ser encontrados, por exemplo, em Chatfield & Collins (1975); aqui será feita uma síntese sobre suas características. O teste de Wilks é construído pela razão de verossimilhança e é definido pela estatística lambda de Wilks ( $\Lambda$ ), onde

$$\Lambda = \frac{\det(E)}{\det(E+T)} \quad (5.3)$$

e  $\det(E)$  significa determinante da matriz  $E$ . O teste de Hotelling é representado pela estatística  $T^2$ , onde

$$T^2 = Tr(TE^{-1})$$

e  $Tr(TE^{-1})$  significa traço da matriz  $TE^{-1}$ , ou seja, a soma dos elementos da diagonal da matriz  $TE^{-1}$ . O teste de Pillai usa a estatística  $U$ , onde

$$U = Tr[T(E+T)^{-1}].$$

O teste de Roy usa a estatística  $R$ , onde  $R =$  máximo autovalor da matriz  $TE^{-1}$ .

Os quatro testes não são equivalentes e não existe orientação em relação a qual escolher. Na análise do exemplo, será usado o teste de Wilks porque se assemelha ao teste F da análise da variância, embora os aplicativos, como o SAS, forneçam resultados dos quatro. Assim, tem-se que:

$$\Lambda = \frac{\det(E)}{\det(E+T)} = 0,05038.$$

Observe que se  $H_0$  é verdade então  $T$  tende para uma matriz nula e assim  $\Lambda$  converge para 1 e se  $H_0$  é rejeitada então,  $\Lambda$  aproxima-se de 0 (zero); portanto,  $\Lambda$  varia entre 0 e 1; logo  $H_0$  é rejeitada para valores próximos de 0 (zero), ao contrário do que acontece no teste F. A distribuição de  $\Lambda$  é complicada e usa-se o teste F (aproximado), obtido pela transformação de  $\Lambda$  (Rao (1977)) e usada pelo SAS (1990); para o exemplo,  $F=4,722$ , com 15 e 36,29 graus de liberdade e  $H_0$  é rejeitada.

Esta introdução sobre análise multivariada é necessária para que aqueles que usarem o SAS, entendam os resultados das diversas análises que esse aplicativo de estatística produz; de fato, prefere-se a análise univariada, pois a análise de variância multivariada ignora que existe uma seqüência no tempo e os testes de significância são menos sensíveis do que os testes univariados.

## 6 ANÁLISE DE EXPERIMENTOS

Os dados deste experimento fazem parte de um estudo relatado em Mendes (1998); um dos objetivos foi testar três tipos de cultivo de cobertura do solo: Pousio, centeio e trevo-vermelho. O delineamento experimental foi blocos casualizados. Uma das variáveis resposta, a biomassa, foi quantificada em cinco faixas de profundidade do solo:  $P_1, P_2, P_3, P_4$  e  $P_5$ ; para a análise, os dados foram modificados e parte deles estão no disquete que acompanha esse trabalho (o leitor deve exercitar essa seção com os programas do disquete e com dados próprios). O comportamento dos tratamentos, ao longo das cinco profundidades, está representado na Figura 6.1 e as médias estão na Tabela 6.1.

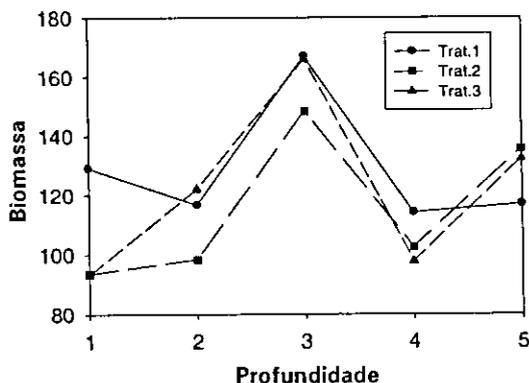


FIG. 6.1 Comportamento dos tratamentos.

TABELA 6.1 Médias dos tratamentos.

Tratam.	Profundidade					média
	1	2	3	4	5	
1	129,32	116,74	166,94	114,15	116,92	128,81
2	93,61	98,33	148,22	102,48	135,59	115,65
3	93,12	122,02	165,90	97,80	132,20	122,21
média	105,35	112,36	160,35	104,81	128,24	122,22

Pela Figura 6.1, há indicação que os perfis não são paralelos (existência de interação). A matriz de correlações no espaço é a seguinte:

	$P_1$	$P_2$	$P_3$	$P_4$
$P_2$	-0,07			
$P_3$	-0,08	0,57		
$P_4$	0,54	0,17	0,14	
$P_5$	-0,80	0,29	0,40	-0,67

As correlações entre as profundidades não evidenciam um padrão de comportamento em concordância com os perfis dos tratamentos. Se houver interesse, examinar-se-ão os resultados da análise de variância de cada profundidade.

Neste ponto será preciso elaborar a parte computacional do SAS utilizada na análise estatística dos dados; embora todos os programas estejam no disquete que acompanha o trabalho, serão descritas, aqui, as principais partes. A entrada dos dados é obtida pelo programa:

```
DATA TEMPO1;
INPUT PLOT REP TREAT PROF BIOMASS ;
DATALINES;
  2    1    1    1    153.205
  2    1    1    2    113.937
  .    .    .    .    .
 34    4    3    5    128.834
 34    4    3    5    105.398
;
PROC PRINT DATA=TEMPO1; RUN;
```

Para a análise de variância, como se o experimento tivesse sido em parcelas divididas e as condições (H-F) presentes, tem-se o programa :

```

TITLE2 ANÁLISE DA BIOMASS EM PARCELAS DIVIDIDAS(SEM CORREÇÃO DOS
GL);
PROC GLM DATA=TEMPO1;
CLASS REP TREAT PROF;
LABEL BIOMASS="BIOMASSA" PROF="PROFUNDIDADE";;
MODEL BIOMASS = REP TREAT REP*TREAT PROF PROF*TREAT
REP*PROF(TREAT);
TEST H=TREAT E=REP*TREAT;
TEST H=PROF PROF*TREAT E=REP*PROF(TREAT) ;
RUN;

```

Os resultados estão na Tabela 6.2.

**TABELA 6.2** Análise da variância do experimento.

Fontes de variação	GL	QM	F	Pr > F
Rep	3	174,45		
Treat	2	866,45	1,98	0,2200
Erro(a)	6	438,24		
Profundidade	4	6525,03	25,50	0,0001
Prof. X Treat	8	649,30	2,54	0,0300
Erro(b)	36	255,86		
Total	59			

Por essa análise, não houve evidência para a rejeição da hipótese nula para tratamentos ( $F = 1,98$ ); essa inferência ignora os perfis dos tratamentos. Como se esperava, há diferenças entre as profundidades ( $F = 25,50$ ); mais adiante essa interação será discutida.

Continuando, usa-se o procedimento GLM do SAS (1990), com o comando REPEATED para analisar os dados como medidas repetidas no espaço; para usar o comando REPEATED, os dados devem estar dispostos no arquivo (*data set*) da forma multivariada. No exemplo, a seqüência é: REP, TREAT,  $P_1, P_2, P_3, P_4, P_5$ , onde  $P_i$ , é a quantidade de biomassa na profundidade  $i$ . O programa de análise

divide-se em dois subprogramas: O primeiro para colocar os dados da forma multivariada, e o segundo para a análise estatística dos dados e segue abaixo:

Primeiro subprograma:

```
PROC SORT DATA = TEMPO1; BY TREAT REP ; RUN;
DATA TEMPO_M(KEEP = P1-P5 REP TREAT );
  ARRAY PP(5) P1-P5;
  DO PROF = 1 TO 5;
    SET TEMPO1;
    BY TREAT REP ;
    PP(PROF) = BIOMASS;
    IF LAST.REP THEN RETURN;
  END;
RUN;
TITLE2 DADOS PARA ANALISE MULTIVARIADA;
PROC PRINT DATA = TEMPO_M; RUN;
```

Segundo subprograma:

```
TITLE2 ANALISE MULTIVARIADA E UNIVARIADA COM CORREÇÃO DOS GL;
PROC GLM DATA = TEMPO_M;
CLASS REP TREAT ;
MODEL P1-P5 = REP TREAT ;
repeated PROFUNDIDADE 5 CONTRAST(1) / SUMMARY PRINTE ;
/*
repeated PROFUNDIDADE 5 HELMERT / SUMMARY PRINTE NOUNI;
repeated PROFUNDIDADE 5 POLYNOMIAL / SUMMARY PRINTE NOUNI;
repeated PROFUNDIDADE 5 PROFILE / SUMMARY PRINTE NOUNI;
repeated PROFUNDIDADE 5 MEAN / SUMMARY PRINTE NOUNI;
repeated PROFUNDIDADE 5 CONTRAST(5) / SUMMARY PRINTE NOUNI;
*/
RUN;
```

A primeira informação sobre as condições de H-F é dada pelo critério de Mauchly (1940) (ou teste para esfericidade) aplicada às transformações ortogonais; a estatística usada para o teste tem, sob normalidade, a distribuição de Qui-quadrado; neste caso, o critério transformado para Qui-quadrado vale 4,479, com 9 gl e Prob

$(\chi^2 > 4,479) = 0,88$ ; logo a hipótese nula de esfericidade (que corresponde à presença das condições H-F) é aceita e assim a análise univariada como se o experimento tivesse sido em parcelas divididas é válida. Entretanto, esse teste é criticado, porque é susceptível a desvios da normalidade. O teste está descrito em Milliken & Johnson (1992).

Continuando, o SAS fornece resultados dos testes da análise de variância multivariada para profundidade e a interação profundidade x tratamentos; no primeiro, a  $H_0$  é que não há diferenças entre as cinco profundidades e  $\Lambda$  de Wilks tem valor 0,040934 e pelo teste exato de F ( $F = 17,57$  com quatro e três gl) rejeita-se  $H_0$ , concordando com a análise univariada; no segundo caso, a  $H_0$  é que os perfis dos tratamentos ao longo do tempo são paralelos, isto é, não há interação entre tratamentos e profundidade;  $\Lambda$  de Wilks tem valor 0,16449 e pelo teste aproximado de F ( $F = 1,0992$  com 6 e 8 gl) não há interação profundidade x tratamento, discordando da análise univariada.

Por último, são apresentados os resultados da análise univariada com os graus de liberdade ajustados; nessa parte, o SAS executa a análise como se o experimento fosse inteiramente casualizado, não considerando outros delineamentos e portanto é necessária uma interferência externa. Os valores de  $\hat{\theta}$  (estimativa de  $\theta$ ), fornecidos pelo SAS devem ser utilizados para corrigir (quando for o caso) os graus de liberdade para profundidade, profundidade x tratamentos e Erro(b) e então observam-se os novos valores para  $(Pr > F)$ ; se o delineamento for inteiramente casualizado, isto é feito pelo próprio programa. No exemplo têm-se dois valores para  $\hat{\theta} = 0,6338$  (Greenhouse & Geisser) e  $\hat{\theta} = 1$  (Huynh & Feldt); quando  $\hat{\theta}$  é maior do que 1 usa-se o valor 1, como no caso em questão. Pelo valor de  $\hat{\theta}$  de H-F, pode-se analisar o experimento como se ele tivesse sido planejado em parcelas divididas, todavia, como exemplo vamos multiplicar os graus de liberdade pela  $\hat{\theta}$  de G-G; os resultados estão na Tabela 6.3.

**TABELA 6.3 Análise da variância com os graus de liberdade ajustados (G-G).**

Fontes de variação	GL	GL(Aj.)	QM	F	Pr > F	Pr > F G-G
Rep	3	3	174,45			
Treat	2	2	866,45	1,98	0,2200	
Erro(a)	6	6	438,24			
Prof	4	2,5	6525,00	25,50	0,0001	0.0001
Prof x Treat	8	5,1	649,30	2,54	0,0300	0.0560
Erro(b)	36	22,8	255,86			
<b>Total</b>	<b>59</b>					

O procedimento anterior está também descrito em Hinkelman & Kempthorne (1994).

Confrontando os resultados da análise multivariada (que admite  $\Sigma$  arbitrária e constante para cada tratamento), análise univariada (restrita às condições H-F) e a univariada com correções dos graus de liberdade), conclui-se que não há evidência suficiente para rejeitar a hipótese de nulidade dos tratamentos, há diferença entre as profundidades e não há evidência suficiente para admitir interação entre profundidade x tratamentos. A última afirmação corresponde a aceitar o paralelismo dos perfis dos tratamentos ao longo das profundidades.

O comando REPEATED do GLM possibilita cinco outros tipos de estudo:

- 1) Comparação entre uma profundidade e a média das demais, com o comando MEAN;
- 2) Comparação da profundidade k com a média das posteriores, com o comando HELMERT;

- 3) Estudo dos efeitos lineares, quadrático, etc., no tempo ou no espaço, com o comando POLYNOMIAL;
- 4) Estudo das diferenças entre dois tempos sucessivos, com o comando PROFILE;
- 5) Contraste entre determinada profundidade (por exemplo 1) e cada uma das outras, com o comando CONTRAST(1).

Qual deles vai ser aplicado, depende do tipo de estudo. Em cada um, o SAS produz a análise de variância simples de cada variável transformada; o comando PROFILE, por exemplo, cria quatro variáveis:  $PROFUN.1 = P_2 - P_1$ ,  $PROFUN.2 = P_3 - P_2$ ,  $PROFUN.3 = P_4 - P_3$ ,  $PROFUN.4 = P_5 - P_4$ . Como interpretar a análise de variância de  $P_2 - P_1$ , por exemplo? Se os efeitos dos tratamentos são os mesmos nas profundidades 1 e 2, então  $H_0$  será aceita. Pela análise de variância, não há diferença entre os tratamentos em nenhuma das quatro novas variáveis. Outra opção que cabe neste estudo é a comparação entre a profundidade 1 com cada outra, pelo comando CONTRAST(1); essa opção implica a criação das quatro variáveis:  $PROFUN.2 = P_1 - P_2$ ,  $PROFUN.3 = P_1 - P_3$ ,  $PROFUN.4 = P_1 - P_4$ ,  $PROFUN.5 = P_1 - P_5$ . Somente no último contraste há diferença entre os tratamentos.

Quando uma resposta é observada no tempo, como o peso de um animal ou a altura de uma planta, então o comando POLYNOMIAL tem utilidade, pois através dele pode-se estimar os efeitos lineares, quadráticos, cúbicos, etc.

O exame da matriz R de correlações, com valores negativos e positivos, leva-nos a conjecturar que a matriz de variâncias e covariâncias não tenha um padrão definido; o procedimento MIXED do SAS permite analisar o experimento admitindo vários tipos de matrizes, inclusive, uma matriz  $\Sigma$  qualquer. O programa encontra-se a seguir:

```

PROC MIXED DATA = TEMPO1;
CLASS TREAT REP PROF ;
MODEL BIOMASS = REP TREAT PROF PROF*TREAT ;
REPEATED PROF / TYPE = HF SUB = REP*TREAT RCORR;
RUN;

```

A opção chave neste programa é **TYPE=HF**, que significa que na análise do experimento vai ser usada uma matriz com as condições H-F; outros tipos de matrizes podem ser empregadas como: variâncias iguais e covariâncias iguais (TYPE=CS), matriz sem estrutura (TYPE=UN), etc; o manual do SAS (1997) contém uma listagem dessas matrizes e suas respectivas definições. O procedimento de análise funciona da seguinte forma: Obtêm-se os resultados para cada matriz e por meio dos critérios de informação de Akaike (AIC) e o Bayesiano de Schwarz (SBC), seleciona-se a matriz cujo critério deu o maior valor (Littel et al., 1996). Os resultados para os três tipos de matrizes citadas foram os seguintes:

CRITÉRIO	MATRIZ		
	HF	CS	UN
AKAIKE	-189,46	-193,41	-194,76
SCHWARZ	-194,68	-195,15	-207,79

Neste estudo, a matriz  $\Sigma$  tipo H-F apresenta os melhores resultados: AIC = -189,46 e SBC = -194,67; com esse modelo, pode-se aplicar a análise univariada sem ajustamento de graus de liberdade. A única discordância está na interação; considerando que o teste multivariado é mais conservador e que a matriz  $\Sigma$  H-F é a adequada, rejeita-se a hipótese de paralelismo e volta-se para a análise de variância quando as condições H-F ocorrem (a primeira que foi feita). Uma decisão diante desses resultados é mais difícil, entretanto, os computadores de alta velocidade de cálculo e os aplicativos de estatística, permitem o estudo de vários modelos, diminuindo o risco de se adotar um modelo inadequado.

No experimento analisado, a biomassa também foi avaliada em duas estações, outono e primavera e deparou-se com o problema: como analisar um experimento com medidas repetidas no espaço e no tempo? Para a análise que se segue foi simulada uma terceira estação, pois com duas, as condições H-F ocorrem. Será desenvolvido o procedimento presente em Milliken & Johnson (1992). Considerando as faixas de profundidade como subparcelas e as estações (tempo), como subsubparcelas, a análise do experimento como se ele tivesse sido planejado como parcelas subdivididas (split-split-plot), é obtida com o seguinte programa:

```
TITLE1 DADOS REPETIDOS NO ESPAÇO E TEMPO;
TITLE2 ANÁLISE DA BIOMASS COMO PARCELAS SUBDIVIDIDAS;
PROC GLM DATA=TUDO;
CLASS REP TREAT PROF SEASON;
LABEL BIOMASS="BIOMASSA" PROF="PROFUNDIDADE" SEASON="
ESTAÇÃO";
MODEL BIOMASS = REP TREAT REP*TREAT PROF PROF*TREAT
REP*PROF(TREAT) SEASON SEASON*TREAT SEASON*PROF SEASON
*PROF*TREAT REP*TREAT*PROF*SEASON;
TEST H=TREAT E=REP*TREAT;
TEST H=PROF PROF*TREAT E=REP*PROF(TREAT) ;
TEST H=SEASON SEASON*TREAT SEASON*PROF SEASON *PROF*TREAT
E=REP*TREAT*PROF*SEASON;
RUN;
```

Os resultados estão na Tabela 6.4, com duas colunas a mais para completar a análise.

**TABELA 6.4** Análise da variância do experimento.

FV	GL	QM	F	Pr > F	GL (G-G)	Pr > F(G-G)
rep	3	831,74				
treat	2	1166,89	2,42	0,170		
erro(a)	6	482,59				
prof	4	6843,92	20,46	0,001	3,81	0,001
treat*prof	8	570,89	1,71	0,130	5,62	0,160
erro(b)	36	334,49			25,28	
season	2	2736,46	18,58	0,001	1,58	0,001
prof*season	8	1910,57	12,97	0,001	6,30	0,001
treat*season	4	372,45	2,53	0,046	3,15	0,061
treat*prof*season	16	203,48	1,38	0,170	12,61	0,190
erro(c)	90	147,31			70,91	
<b>TOTAL</b>	<b>179</b>					

A análise nessas circunstâncias tem de considerar que nem as profundidades nem as estações foram casualizadas e portanto é necessário verificar as condições H-F em duas matrizes de variâncias e covariâncias: para as cinco profundidades e para as três estações. O primeiro passo será obter as estimativas de  $\theta$  de G-G e H-F, para as subsubparcelas (estações) e subparcelas (profundidades) respectivamente e isso se faz em duas fases. A primeira, considerando no SAS que o delineamento é de parcelas subdivididas (tratamentos nas parcelas e profundidades nas subparcelas) com medidas repetidas no tempo (três estações, que são as subsubparcelas); o programa é o seguinte:

```
PROC SORT DATA =TUDO; BY TREAT REP PROF SEASON ; RUN;
PROC PRINT DATA =TUDO; RUN;
```

```
DATA TUDO_M(KEEP = S1-S3 REP TREAT PROF );
  ARRAY SS(3) S1-S3;
  DO S = 1 TO 3;
    SET TUDO;
    BY TREAT REP PROF;
    SS(S) = BIOMASS;
    IF LAST.PROF THEN RETURN;
  END;
RUN;
TITLE2 DADOS PARA ANALISE MULTIVARIADA;
PROC PRINT DATA =TUDO_M; RUN;
```

```
TITLE2 ANALISE MULTIVARIADA E UNIVARIADA COM CORREÇÃO DOS GL;
TITLE3 ESTIMATIVAS DE G-G E H-F PARA AS ESTAÇÕES ;
PROC GLM DATA =TUDO_M;
CLASS REP TREAT PROF;
MODEL S1-S3 = REP TREAT REP*TREAT PROF PROF*TREAT ;
TEST H = REP TREAT E = REP*TREAT;
REPEATED ESTACAO 3 CONTRAST(3) /SUMMARY PRINTE ;
  /*
REPEATED ESTAÇÃO 3 HELMERT / SUMMARY PRINTE NOUNI;
REPEATED ESTAÇÃO 3 POLYNOMIAL / SUMMARY PRINTE NOUNI5
REPEATED ESTAÇÃO 3 PROFILE / SUMMARY PRINTE NOUNI;
REPEATED ESTAÇÃO 3 MEAN / SUMMARY PRINTE NOUNI;
REPEATED ESTAÇÃO 3 CONTRAST(5) / SUMMARY PRINTE NOUNI;
  */
RUN;
```

que produziu  $G-G = 0,7879$  ;  $H-F = 1$  para as estações. A segunda considera que o delineamento básico é o de blocos casualizados com medidas repetidas em cinco faixas de profundidade, tomando para a análise as médias das três estações de cada profundidade; o programa para essa análise é o seguinte:

```

PROC SORT DATA=TUDO; BY TREAT REP PROF SEASON; RUN;
PROC MEANS DATA=TUDO MEAN NOPRINT; BY TREAT REP PROF;
VAR BIOMASS;
OUTPUT OUT = TUDO2 MEAN = BIOMASS;
RUN;
PROC PRINT DATA=TUDO2; RUN;
PROC SORT DATA=TUDO2; BY TREAT REP PROF; RUN;
PROC PRINT DATA=TUDO2;
VAR REP TREAT PROF BIOMASS; RUN;

DATA TUDO3(KEEP=P1-P5 REP TREAT );
  ARRAY PP(5) P1-P5;
  DO PROF = 1 TO 5;
    SET TUDO2;
    BY TREAT REP;
    PP(PROF) = BIOMASS;
    IF LAST.REP THEN RETURN;
  END;
RUN;
TITLE2 DADOS PARA ANALISE MULTIVARIADA;
PROC PRINT DATA=TUDO3; RUN;

TITLE2 ANALISE MULTIVARIADA E UNIVARIADA COM CORREÇÃO DOS GL
PARA AS SUB PARCELAS;
TITLE3 CALCULO DO DA ESTIMATIVA DE TETA(BOX) PARA AS PROFUNDIDA-
DES;
PROC GLM DATA=TUDO3;
CLASS TREAT REP;
MODEL P1-P5 = TREAT REP;
REPEATED PROF 5 CONTRAST(1) /SUMMARY PRINTE;
/*
REPEATED PROFUNDIDADE 5 HELMERT / SUMMARY PRINTE NOUNI;
REPEATED PROFUNDIDADE 5 POLYNOMIAL / SUMMARY PRINTE NOUNI;
REPEATED PROFUNDIDADE 5 PROFILE / SUMMARY PRINTE NOUNI;
REPEATED PROFUNDIDADE 5 MEAN / SUMMARY PRINTE NOUNI;
REPEATED PROFUNDIDADE 5 CONTRAST(5) / SUMMARY PRINTE NOUNI;
*/
RUN;

```

com ele obteve-se  $G-G = 0,7023$ ;  $H-F = 1$ , para as profundidades. O valor de  $\theta$  para as subparcelas (profundidades) está entre 0,25 e 1, enquanto para as subsubparcelas (estações) está entre 0,5 e 1; quando a estimativa de  $\theta$  tem valor acima de 1, adota-se 1 para  $\hat{\theta}$ ; como uma ilustração, na Tabela 6.4, na sexta coluna, estão os graus de liberdade corrigidos pelo valor de  $G-G$  e na sétima, os respectivos níveis de significância. Embora os valores de  $\hat{\theta}$  de  $G-G$  foram utilizados na tabela anterior, os recomendados são os de  $H-F$ , e portanto, pelo valor igual a 1, existem as condições de  $H-F$  e volta-se à análise original, univariada, sem nenhum ajustamento de graus de liberdade. Os resultados do PROC MIXED do SAS, para cada caso, conduzem à escolha da análise univariada sem ajustamentos de graus de liberdade, dispostos na Tabela 6.4.

Cabe aqui um comentário sobre as possíveis matrizes que podem ser escolhidas para uso no PROC MIXED; foram utilizadas três, mas o manual do SAS (1997), p. 626, fornece uma coleção de delas; outro estudo pode modificar a conclusão atual e nesse caso prevalece a análise de variância fornecida pelo PROC MIXED.

Como referido antes, ocorrem situações experimentais em que tem significado estudar as medidas no tempo de forma acumulada; considere-se um experimento em blocos ao acaso, com seis repetições, com o objetivo de comparar três forrageiras (tratamentos 1, 2 e 3 respectivamente). A produção, em toneladas por hectare, acumulada em quatro épocas: tempo 1, tempo 2, tempo 3 e tempo 4 respectivamente - estão na Tabela 6.5.

**TABELA 6.5** Produção acumulada (Y) no tempo.

OBS	BLOCO	TRAT	TEMPO	Y
1	1	1	1	3.37
2	1	1	2	5.68
3	1	1	3	9.53
4	1	1	4	12.03

TABELA 6.5 (continuação)

OBS	BLOCO	TRAT	TEMPO	Y
5	1	2	1	2.22
6	1	2	2	3.15
7	1	2	3	3.15
8	1	2	4	5.07
9	1	3	1	2.30
10	1	3	2	4.85
11	1	3	3	6.06
12	1	3	4	8.79
13	2	1	1	1.53
14	2	1	2	4.24
15	2	1	3	6.69
16	2	1	4	9.12
17	2	2	1	3.62
18	2	2	2	5.50
19	2	2	3	7.24
20	2	2	4	8.46
21	2	3	1	2.71
22	2	3	2	3.90
23	2	3	3	4.53
24	2	3	4	5.49
25	3	1	1	5.09
26	3	1	2	9.27
27	3	1	3	10.50
28	3	1	4	14.11
29	3	2	1	3.01
30	3	2	2	3.83
31	3	2	3	5.53
32	3	2	4	9.21
33	3	3	1	2.49
34	3	3	2	3.75
35	3	3	3	4.38
36	3	3	4	6.34

TABELA 6.5 (continuação)

OBS	BLOCO	TRAT	TEMPO	Y
37	4	1	1	2.84
38	4	1	2	4.52
39	4	1	3	8.24
40	4	1	4	13.20
41	4	2	1	1.33
42	4	2	2	1.53
43	4	2	3	3.10
44	4	2	4	6.26
45	4	3	1	2.57
46	4	3	2	4.92
47	4	3	3	7.45
48	4	3	4	10.63
49	5	1	1	3.19
50	5	1	2	4.66
51	5	1	3	6.64
52	5	1	4	9.32
53	5	2	1	1.68
54	5	2	2	3.20
55	5	2	3	4.21
56	5	2	4	6.95
57	5	3	1	2.24
58	5	3	2	2.98
59	5	3	3	3.67
60	5	3	4	6.04
61	6	1	1	3.65
62	6	1	2	6.39
63	6	1	3	8.60
64	6	1	4	10.36
65	6	2	1	1.99
66	6	2	2	2.44
67	6	2	3	3.90
68	6	2	4	5.72
69	6	3	1	2.20
70	6	3	2	4.96
71	6	3	3	5.56
72	6	3	4	5.82

Pela Figura 6.2, conclui-se que, ao longo do tempo, a taxa de crescimento da produção para o tratamento 1, por período (no período estudado), é maior do que as dos tratamentos 2 e 3 e o comportamento linear da produção sugere um modelo de regressão simples para cada parcela; como os dados são correlacionados no tempo, utilizou-se o PROC AUTOREG do SAS (do módulo ETS) e estimou-se a taxa de crescimento ( $\beta_1$ ) para cada parcela, por meio do modelo:

$$Y_{ijt} = \beta_0 + \beta_1 t + e_{ijt} \quad (6.1)$$

onde  $t$  é o tempo. A Tabela 6.6 contém as referidas taxas.

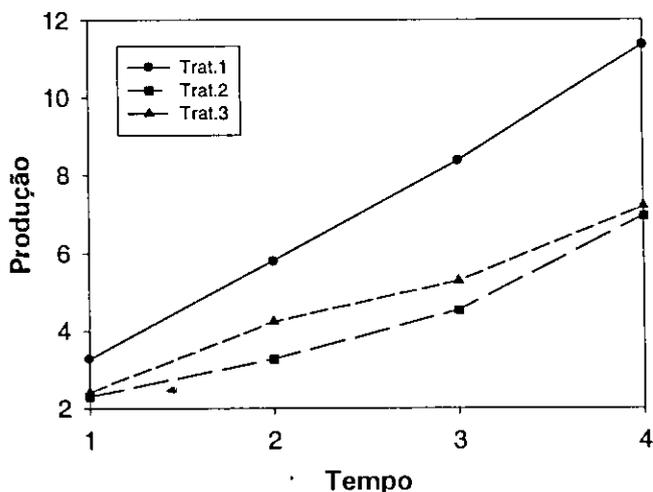


FIG. 6.2 Comportamento da produção no tempo.

A interpretação é a seguinte: para a parcela 18, bloco 6, tratamento 3, a taxa de crescimento foi de 1,11 t/ha por período (60 dias); o próximo passo é a análise de variância das taxas, cujo resultado está na Tabela 6.7.

**TABELA 6.6** Taxas de aumento da produção.

OBS	BLOCO	TRAT	TAXA
1	1	1	3.09
2	1	2	0.77
3	1	3	1.96
4	2	1	2.51
5	2	2	1.63
6	2	3	0.86
7	3	1	2.63
8	3	2	2.01
9	3	3	1.16
10	4	1	3.49
11	4	2	1.63
12	4	3	2.66
13	5	1	2.03
14	5	2	1.63
15	5	3	1.17
16	6	1	2.23
17	6	2	1.27
18	6	3	1.11

**TABELA 6.7** Análise da variância das taxas de aumento.

FV	GL	QM	F	Pr > F
Bloco	5	0,4484	1,83	
Tratamento	2	2,7576	11,24	0,0028
Erro	10	0,2452		
<b>Total</b>	<b>17</b>			

Dessa forma, pela regressão, transforma-se a variável resposta, concentrando a informação no tempo em um parâmetro, possibilitando uma conclusão quantitativa sobre os tratamentos. As taxas médias por tratamento são: trat 1 = 2,67; trat 2 = 1,49; trat 3 = 1,49; o tratamento 1 tem uma taxa de crescimento maior.

Cabe ressaltar que outros modelos de regressão podem ser aplicados, mas deve-se preferir os mais simples, para não complicar a análise. Por exemplo, o modelo

$$Y = A - \beta_0 e^{-\beta_1 t}, \quad \beta_0 > 0, \quad \beta_1 > 0, \quad (6.2)$$

pode ser usado quando o crescimento é não linear e cada parâmetro tem sua interpretação:  $A$  é o máximo que  $Y$  pode atingir e quanto maior  $\beta_1$ , mais forte é a curvatura.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

- ANSCOMBE, F.J. The validity of comparative experiments. **Journal of the Royal Statistical Society, Series A**, London, v.61, p.181-211, 1948.
- BOX, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems. **Annals of Mathematical Statistics**, Washington, DC, v.25, p.290-302, 1954.
- BOX, G.E.P.; COX, D.R. An analysis of transformations. **Journal of the Royal Statistical Society, Series B**, London, v.26, p.211-252, 1964.
- CHATFIELD, C.; COLLINS, A.J. **Introduction to multivariate analysis**. London: Chapman & Hall, 1976.
- COCHRAN, W.G.; COX, G.M. **Experimental designs**. 2.ed. New York: J. Wiley & Sons, 1957. 611p.
- COX, D.R. **Planning of experiments**. New York: J.Wiley, 1958.
- FISHER, R.A. The arrangement of field experiments. **Journal of Ministry of Agriculture, United Kingdom**, v.33, p.503-513, 1926.
- FISHER, R.A. **The design of experiments**. Edinburg: Oliver and Boyd, 1935.

- GREENHOUSE, S.W.; GEISSER, S. On methods in the analysis of profile data. **Psychometrika**, Baltimore, v.24, p.95-112, 1959.
- HINKELMAN, K.; KEMPTHORNE, O. **Design and analysis of experiments: introduction to experimental design**. New York: Wiley, 1994. v.1.
- HUYNH, H.; FELDT, L.S. Conditions under which mean square ratios in repeated measures designs have exact F-distributions. **Journal of the American Statistical Association**, Washington, DC, v.65, p.1582-89, 1970.
- KEMPTHORNE, O. **The design and analysis of experiments**. New York: J. Wiley, 1952.
- KEMPTHORNE, O. The randomization theory of experimental inference. **Journal of the American Statistical Association**, Washington, DC, v.50, p.946-67, 1955.
- KEMPTHORNE, O. Inference from experiments and randomization. In: SRIVASTAVA, N. ed. **A survey of statistical design and linear models**. Amsterdam: North-Holland, 1973. p.303-331.
- LITTELL, C.R.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D. **SAS system for mixed models**. Cary, NC: SAS Institute / SAS Campus Drive, 1996.
- MAUCHLY, J.W. Significance test for sphericity of a normal n-variate distribution. **Annals of Mathematical Statistics**, Washington, DC, v.11, p.608-610, 1940.
- MENDES, I. de C. **Microbiology of soil aggregates recovered from different crop management systems**. Corvallis: Oregon State University, 1997. 138p. Ph.D. Thesis.
- MILLIKEN G.A.E.; JOHNSON, D.E. **Analysis of messy data: designed experiments**. New York: Chapman & Hall, 1992. v.1.
- SAS INSTITUTE (Cary, NC). **SAS / STAT user's guide, version 6**. 4.ed. Cary, NC: SAS Institute / SAS Campus, 1990. v.2

**SAS INSTITUTE (Cary, NC). SAS software: changes and enhancements through release 6.12. Cary, NC: SAS Institute / SAS Campus, 1997.**

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

**SEBER, G.A.F. Multivariate observations. New York: J. Wiley, 1984.**