



**EMPREGO DO MODELO
SUPERPARAMETRIZADO EM
EXPERIMENTO FATORIAL
DESBALANCEADO COM DOIS E TRÊS
FATORES**

ELIANA MARA MANSO

2004

ELIANA MARA MANSO

**EMPREGO DO MODELO SUPERPARAMETRIZADO EM
EXPERIMENTO FATORIAL DESBALANCEADO COM DOIS E TRÊS
FATORES**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

Orientador

Prof. Dr. Augusto Ramalho de Moraes

LAVRAS
MINAS GERAIS –
2004

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Manso, Eliana Mara

**Emprego do modelo superparametrizado em experimento fatorial
desbalanceado com dois e três fatores / Eliana Mara Manso. -- Lavras : UFLA,
2004.**

62 p. : il.

Orientador: Augusto Ramalho de Moraes.

Dissertação (Mestrado) – UFLA.

Bibliografia.

**I. Modelo superparametrizado. 2. Análise fatorial. I. Universidade Federal de
Lavras. II. Título.**

CDD-519.5354

ELIANA MARA MANSO

**EMPREGO DO MODELO SUPERPARAMETRIZADO EM
EXPERIMENTO FATORIAL DESBALANCEADO COM DOIS E TRÊS
FATORES**


Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de “Mestre”.

APROVADA em 18 de fevereiro de 2004.

Prof. Dr. Joel Augusto Muniz UFLA

Prof. Dr. Messias José Bastos de Andrade UFLA

Prof. Dr. Paulo César Lima UFLA


Prof. Dr. Augusto Ramalho de Moraes
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

Ao meu pai, *Luiz Carlos Manso Filho* (in memoriam).

À minha mãe, *Ana Thereza dos Santos Manso*.

Aos meus irmãos, cunhadas e sobrinhos,

DEDICO.

Aos meus filhos,

Fernanda e Armando Camargo Penteado Neto,

OFEREÇO.

A *DEUS*,
AGRADEÇO pela vida.

AGRADECIMENTOS

Ao Professor Dr. Augusto Ramalho de Moraes, pela sugestão do tema, orientação, amizade e confiança.

Ao Professor Dr. Joel Augusto Muniz, pela compreensão e incentivo sempre presentes.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de estudos.

Aos professores e do Departamento de Ciências Exatas da UFLA, Daniel Furtado Ferreira, Eduardo Bearzoti, Júlio Sílvio de Souza Bueno Filho, Lucas Monteiro Chaves, Luiz Henrique de Aquino, Marcelo Silva de Oliveira, Mário Xavier Ferrua Vivanco, Paulo César Lima, Ruben Delly Veiga e Thelma Sáfadi, pelos ensinamentos, incentivo e amizade.

As funcionárias do Departamento de Ciências Exatas, pela atenção e dedicação.

Aos meus colegas, Carla Regina Guimarães Brighenti, Denismar Alves Nogueira, Fabyano Fonseca e Silva, Francisco Luiz Vilas Boas, Janser Moura Pereira, José Antônio Leite, José Waldemar Silva, Marcelo Ângelo Cirillo, Maria Laucinéia Carari, Rafaela Cardoso Alves, Renata Pires Gonçalves e Vânia Fátima Lemes de Miranda, pela convivência, contribuição, companheirismo e, principalmente, pela verdadeira amizade.

A Luiz Cláudio Santos Pereira e Silvana Pereira Ramos Garcia, por me estenderem a mão nos momentos mais difíceis.

À Universidade Federal de Lavras e ao programa de pós-graduação em agronomia/estatística e experimentação agropecuária, pela oportunidade.

A todos que, de algum modo, contribuíram para que este trabalho fosse realizado.

SUMÁRIO

	Página
LISTA DE ABREVIATURAS E SIGLAS.....	i
LISTA DE TABELAS	ii
RESUMO	vi
ABSTRACT	vii
1 INTRODUÇÃO	01
2 REFERENCIAL TEÓRICO	03
2.1 Caracterização do modelo linear	09
2.2 Estimação.....	09
2.2.1 Estimação em modelos lineares.....	10
2.2.2 Funções lineares paramétricas estimáveis	11
2.2.3 O melhor estimador linear não viesado.....	12
2.2.4 Regras práticas de estimabilidade	13
2.3 Conectividade	14
2.4 Testabilidade.....	15
2.5 Modelo superparametrizado (Modelo-S).....	15
2.5.1 Obtenção de somas de quadrados.....	18
2.5.2 Notação $R(\)$	20
2.6 Funções estimáveis.....	23
2.7 Hipóteses estatísticas.....	25
2.8 Sistema estatístico: PROC GLM do SAS	33
3 MATERIAL E MÉTODOS	35
3.1 1º caso: fatorial 4 x 4 desbalanceado.....	35
3.1.1 Modelo superparametrizado: Modelo-S.....	36

3.2 2° caso: fatorial 4 x 4 desbalanceado com caselas vazias	38
3.2.1 Modelo superparametrizado: Modelo-S	39
3.3 3° caso: fatorial 4 x 5 x 4 desbalanceado	41
3.3.1 Modelo superparametrizado: Modelo-S	41
3.4 Procedimento GLM do sistema estatístico SAS	43
4 RESULTADOS E DISCUSSÃO	45
4.1 1° caso: fatorial 4 x 4 desbalanceado	45
4.1.1 Equações normais e aproximação de mínimos quadrados	50
4.1.2 Estimação	52
4.1.3 Hipóteses e somas de quadrados associados	56
4.2 2° caso: fatorial 4 x 4 desbalanceado com caselas vazias	60
4.2.1 Equações normais e aproximação de mínimos quadrados	70
4.2.2 Estimação	72
4.2.3 Hipóteses e somas de quadrados associadas	76
4.2.3.1 Hipóteses	76
4.2.3.2 Conseqüências da ordenação dos efeitos principais	79
4.2.3.3 Somas de quadrados	82
4.3 3° caso: fatorial 4 x 5 x 4 desbalanceado	86
4.3.1 Estimação	86
4.3.2 Hipóteses e somas de quadrados associadas	87
4.3.3 Comparações entre as médias	92
5 CONCLUSÕES	94
6 REFERÊNCIAS BIBLIOGRÁFICAS	95
ANEXOS	99

LISTA DE ABREVIATURAS E SIGLAS

BLUE	Best Linear Unbiased Estimator
BMDP	BioMeDical Statistic Package
EAD	Each ADjusted for the other
ENA	Sistema de equações normais associadas
Modelo-M	Modelo de médias de caselas
Modelo-S	Modelo superparametrizado
Notação R()	Procedimento para obtenção de somas de quadrados por meio da redução da soma de quadrados totais ao ajustar um modelo particular de interesse
PROC GLM	Procedimento General Linear Models
SAS	Statistical Analysis System
SEN	Sistema de equações normais
SPSS	Statistical Package for Social Science
SQ	Soma de quadrados
STATGRAPHICS	Statistical Graphics System
STP	Standard Parametric
WTM	WeighTed Means

LISTA DE TABELAS

Página

TABELA 1	Hipóteses associadas ao modelo de classificação dupla.....	26
TABELA 2	Somas de quadrados obtidas pelo método do ajuste de constantes.....	27
TABELA 3	Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com as diferentes fases da lua (1º caso).....	35
TABELA 4	Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com as diferentes fases da lua (2º caso).....	39
TABELA 5	Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com os meses e fases da lua na época de plantio (3º caso).	42
TABELA 6	Forma geral de funções estimáveis dos dados da Tabela 3, fornecida pelo PROC GLM do SAS.....	46
TABELA 7	Forma geral das funções estimáveis do Tipo I e II do SAS para os dados da Tabela 3, na ordem A-B, para os fatores A, B e interação AB	47
TABELA 8	Forma geral das funções estimáveis do Tipo I e II do SAS para os dados da Tabela 3, na ordem B-A, para os fatores A, B e interação BA	48
TABELA 9	Forma geral das funções estimáveis do Tipo III e IV do SAS para os dados da Tabela 3, na ordem A-B e na ordem B-A, para os fatores A, B e interação AB	49
TABELA 10	Funções estimáveis sobre o fator A (cultivares), dos Tipos I, II, III e IV	54

TABELA 11	Funções estimáveis sobre o fator B (fases da lua), dos Tipos I, II, III e IV.....	55
TABELA 12	Funções estimáveis sobre a interação	55
TABELA 13	Hipóteses testadas sobre o fator A (cultivares).....	58
TABELA 14	Hipóteses testadas sobre o fator B (fases da lua)	58
TABELA 15	Hipóteses testadas sobre a interação	58
TABELA 16	Análise de variância dos dados da Tabela 3, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo I e II, e respectivas hipóteses associadas	59
TABELA 17	Análise de variância dos dados da Tabela 3, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo III e VI, e respectivas hipóteses associadas	59
TABELA 18	Forma geral de funções estimáveis dos dados da Tabela 4, fornecida pelo PROC GLM do SAS, na ordem A-B	60
TABELA 19	Forma geral de funções estimáveis dos dados da Tabela 4, fornecida pelo PROC GLM do SAS, na ordem B-A	61
TABELA 20	Forma geral das funções estimáveis do Tipo I do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB	62
TABELA 21	Forma geral das funções estimáveis do Tipo I do SAS para os dados da Tabela 4, na ordem B-A, para os fatores A, B e interação BA.....	63
TABELA 22	Forma geral das funções estimáveis do Tipo II do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.....	64

TABELA 23	Forma geral das funções estimáveis do Tipo II do SAS para os dados da Tabela 4, na ordem B-A, para os fatores A, B e interação BA.....	65
TABELA 24	Forma geral das funções estimáveis do Tipo III do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.....	66
TABELA 25	Forma geral das funções estimáveis do Tipo III do SAS para os dados da Tabela 4, na ordem B-A, para os fatores A, B e interação BA.....	67
TABELA 26	Forma geral das funções estimáveis do Tipo IV do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.....	68
TABELA 27	Forma geral das funções estimáveis do Tipo IV do SAS para os dados da Tabela 4, na ordem B-A, para os fatores A, B e interação BA.....	69
TABELA 28	Funções estimáveis sobre o fator A (cultivares) dos Tipos I e II	74
TABELA 29	Funções estimáveis sobre o fator A (cultivares) dos Tipos III e IV	74
TABELA 30	Funções estimáveis sobre o fator B (fases da lua) dos Tipos I, II, III e IV	75
TABELA 31	Funções estimáveis sobre a interação	76
TABELA 32	Hipóteses testadas sobre o fator A (cultivares).....	80
TABELA 33	Hipóteses testadas sobre o fator B (fases da lua).....	81
TABELA 34	Hipóteses testadas sobre a interação.....	81
TABELA 35	Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo I, e respectivas hipóteses associadas	84

TABELA 36	Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo II, e respectivas hipóteses associadas	84
TABELA 37	Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo III, e respectivas hipóteses associadas	85
TABELA 38	Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo IV, e respectivas hipóteses associadas	85
TABELA 39	Análise de variância dos dados da Tabela 5, com base no modelo $y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varphi_{ij} + \eta_{ik} + \tau_{jk} + \rho_{ijk} + e_{ijkl}$	89
TABELA 40	Análise de variância dos dados da Tabela 5, com base no modelo $y_{jkil} = \mu + \beta_j + \gamma_k + \alpha_i + \tau_{jk} + \varphi_{ji} + \eta_{ki} + \rho_{jki} + e_{jkil}$	90
TABELA 41	Análise de variância dos dados da Tabela 5, com base no modelo $y_{kjil} = \mu + \gamma_k + \beta_j + \alpha_i + \tau_{kj} + \eta_{ik} + \varphi_{ji} + \rho_{kji} + e_{kjil}$	91
TABELA 42	Número de observações por casela, de acordo com as interações duplas, B-A, B-C e A-C	92
TABELA 43	Valores médios estimados da produção comercial de raízes de cenoura, em t/ha, de acordo com as cultivares e épocas (meses) de plantio	93
TABELA 44	Valores médios estimados da produção comercial de raízes de cenoura, em t/ha, de acordo com os plantios realizados nos diversos meses e nas diferentes fases da lua	93

RESUMO

MANSO, Eliana Mara. Emprego do modelo superparametrizado em experimento fatorial desbalanceado com dois e três fatores. 2004. 102 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária – Agronomia) – Universidade Federal de Lavras, Lavras, MG.*

Na pesquisa agropecuária é comum o estudo de vários fatores e frequentemente ocorrem perdas de observações, constituindo assim um experimento desbalanceado. Este é um problema conhecido há muitos anos e, apesar de terem sido desenvolvidas várias técnicas de estimação de parcelas perdidas, nenhuma é universalmente aceita como única. É necessário conhecer as hipóteses testadas por meio dos sistemas estatísticos e, ocorrendo caselas vazias, a interpretação é ainda mais complexa, pois, geralmente, as hipóteses sobre os efeitos principais de um dos fatores contêm os efeitos principais de outros fatores e os efeitos de interações. Um outro detalhe que influencia a formulação de hipóteses por envolver parâmetros sem interesse, quando ocorre perda de caselas é que, ao adotar-se um modelo superparametrizado, o número de parâmetros pode ser maior do que o número de caselas disponíveis para estimá-los. Diante disso, adotando o modelo superparametrizado, este trabalho teve como objetivo desenvolver esquemas de análises de variâncias de dados desbalanceados e/ou com caselas vazias, identificar e interpretar as hipóteses associadas às somas de quadrados por meio do procedimento General Linear Models (GLM) do Statistical Analysis System (SAS), que provêm quatro tipos de somas de quadrados. Foram analisados três casos distintos, utilizando dados referentes ao peso comercial de cenoura, provenientes de um experimento inteiramente ao acaso, tendo como fatores cultivares e fases da lua no plantio, no estudo do modelo com dois fatores e ainda o fator mês de plantio, no caso do modelo com três fatores. Em face dos resultados obtidos, verificou-se que, quando os dados são desbalanceados, as funções estimáveis de um fator envolvem os parâmetros relativos ao fator e os componentes das interações onde o fator está presente; as somas de quadrados dos Tipos I e II são equivalentes; as dos Tipos III equivalentes às do Tipo IV e a ordenação dos fatores principais não afeta as hipóteses do Tipo I. Entretanto, quando ocorrem caselas vazias no modelo com dois fatores, os quatro tipos de somas de quadrados para o fator principal de entrada são diferentes e a ordenação é fundamental para a obtenção das hipóteses do Tipo I. Quando ocorrem perdas de parcelas, a identificação das funções estimáveis é mais complexa e as hipóteses ficam de difícil interpretação. Nas funções estimáveis de interações ocorrem parâmetros da própria interação.

* Orientador: Augusto Ramalho de Moraes - UFLA

ABSTRACT

MANSO, Eliana Mara. Use of the overparametrized model in factorial experiments with two and three factors. 2004. 102 p. Dissertation (Master in Agronomy/Statistics and Agricultural Experimentation). - Federal University of Lavras, Lavras, MG.*

In agricultural research it is common to study various factors and this frequently leads to loss of observations, therefore forming an unbalanced experiment. This is a well-known problem for many years and in spite of there being developed various favorite lost work techniques, not one is universally accepted as the only one. It is necessary to know the tested hypothesis behind the statistical systems and when there are missing cells the interpretation is even more complex, but generally, the hypothesis on principal effects of one of these factors contains the principal effects of other factors and of effects of interactions. Another detail which influences the formulation of hypothesis by involving parameters without interest, when there is a loss of cells and, by adopting an overparametrized model, this work meets the objective, developing the schemes of variance analysis of unbalanced data and/or with missing cells, identify and interpret the hypothesis associated with square sums behind the procedure General Linear Models (GLM) of Statistical Analysis System (SAS), which prove four types of square sums. Three distinct cases were analysed, using data referring to commercial weight of carrots, arising from an experiment completely at random, using as factors cultivation and the moon's phases on planting, the model study as two factors and even the month of planting factor in the case of three factor models. With the obtained results, you can conclude when the data is unbalanced, the estimated functions of a factor involving the relative parameters of the factor and it's components of interaction where the factor is present; the square sum of Type I and II are equivalent; as are Type III's equivalent as Type IV's and the order of the principle factors do not effect the hypothesis of Type I. However when there are missing cells in the two-factor model, the four types of square sums for the principal factor input are different and the order is fundamental to obtain the hypothesis of Type I. When loss of parts happens, the identification of the estimate functions is more complex and the hypotheses are difficult to interpret. In the estimable functions of interactions present parameters of the interaction itself.

* Advisor: Augusto Ramalho de Morais – UFLA.

1 INTRODUÇÃO

Uma técnica estatística de suma importância na análise de experimentos é a análise de variância, amplamente utilizada nas mais diversas áreas da pesquisa científica.

Na pesquisa agropecuária é comum a necessidade de estudar alguns fatores que influenciam uma variável dependente. Um desses casos é quando se têm dois ou mais fatores simultaneamente, em um arranjo fatorial, de modo que se possam avaliar as interações entre fatores e os próprios fatores.

Nesse caso, se ocorrer que todas as combinações dos níveis dos fatores estejam presentes e suas freqüências sejam iguais, caracterizando um conjunto de dados balanceados, as técnicas de análise de variância são amplamente conhecidas e os resultados obtidos são facilmente analisados por meio de algum sistema computacional para análises estatísticas.

Porém, freqüentemente, ocorre perda de parcelas, por motivos alheios à vontade do pesquisador, como fatores naturais ou acidentais, ou os próprios planejamentos, constituindo assim um conjunto de dados desbalanceados.

Este é um problema conhecido há muitos anos e, apesar de terem sido desenvolvidas várias técnicas de estimação de parcelas perdidas, nenhuma é universalmente aceita como única. Quando o número de observações por caselas é distinto, há necessidade de se conhecer que hipóteses estão sendo testadas e se são de interesse do pesquisador. Além disso, com a ocorrência de caselas vazias, a interpretação das verdadeiras hipóteses é ainda mais difícil.

Outro problema que ocorre na presença de caselas vazias é que, ao adotar-se um modelo superparametrizado, o número de parâmetros pode ser maior do que o número de caselas disponíveis para estimá-los, influenciando a formulação de hipóteses que podem envolver parâmetros sem interesse, induzindo a hipóteses equivocadas. Ainda, há diferenças nas saídas das somas de

quadrados entre os sistemas computacionais, acarretando dúvidas nas interpretações.

O procedimento General Linear Models (GLM) do Statistical Analysis System (SAS) fornece quatro tipos de funções estimáveis, nas quais são embasados a construção das hipóteses e o cálculo da soma de quadrados associados, primordiais na análise de dados desbalanceados. Para uma perfeita interpretação da rotina do GLM é necessário domínio sobre os conceitos de estimabilidade e testabilidade.

Com isso, os objetivos deste trabalho são:

- i) utilizando o modelo superparametrizado, apresentar esquemas de análises de variâncias para experimentos com dois e três fatores, com dados desbalanceados e/ou caselas vazias;
- ii) desenvolver procedimentos para análise desses dados utilizando o sistema computacional SAS;
- iii) identificar e interpretar as hipóteses associadas às somas de quadrados.

2 REFERENCIAL TEÓRICO

Na análise de variância para dados balanceados há um consenso geral com relação aos procedimentos estatísticos de análises dos dados. Porém, quando se trata de dados desbalanceados e caselas vazias, os diferentes resultados apresentados por sistemas estatísticos geram confusão entre pesquisadores, não havendo nenhuma metodologia aceita universalmente como correta e única.

A história da análise de variância para dados desbalanceados é encontrada nos artigos originais de Yates (1933, 1934), que foram a base para o início da revolução computacional nos anos sessenta e para os métodos de análise usados atualmente.

Yates (1934) estabelece três métodos para cálculo das somas de quadrados:

- método das médias não ponderadas;
- método dos quadrados de médias ponderadas;
- método do ajuste de constantes, iniciando a parametrização sucessiva do modelo.

Snedecor & Cox (1935) fornecem sugestões para usar cinco métodos por eles discutidos, não mencionando um modelo nem uma declaração clara de uma hipótese:

- médias não ponderadas;
- números de subclasses esperadas;
- STP – *Standard parametric* (padrão paramétrico; quadrados de médias ponderadas de Yates);
- EAD – *Each adjusted for the other* (ajustamento de constantes de Yates);
- WTM – *Weighted means* (médias ponderadas).

Stevens (1948), citado por Herr (1986), descreve um algoritmo para executar a análise de variância em um experimento fatorial de três fatores com dados desbalanceados, apresentando alguns detalhes sobre a construção da análise, mas nenhum procedimento geral sobre o que fazer quando ocorre falta de observações.

Em análise de classificação dupla com dados desbalanceados, Kramer (1955) focaliza o problema da obtenção das somas de quadrados para os efeitos principais e interação e afirma que podem ocorrer situações em que o modelo, por ser aditivo, dê um resultado negativo para a soma de quadrados da interação. O autor afirma que o melhor método de análise, quando se supõe que não há interação, é o EAD (método de ajuste de constantes de Yates) e, não havendo suposição, é o STP (método dos quadrados de médias ponderadas de Yates), apresentando um STP modificado, mais poderoso.

Pela análise de mínimos quadrados de dados desbalanceados por meio do modelo superparametrizado, tanto para efeitos fixos como para aleatórios, Harvey (1960) fornece detalhes complexos para obter a soma de quadrados em condições explicitamente paramétricas, porém, não especificadas. Na ausência de interação, o autor obteve a soma de quadrados pelo método de ajuste das constantes e, caso contrário, pelo método das médias dos quadrados ponderados.

Francis (1973) compara alguns sistemas estatísticos para um mesmo conjunto de dados e observa que diferentes somas de quadrados são obtidas para os efeitos principais, alertando que, para o exemplo utilizado, um fatorial com classificação dupla e dados desbalanceados, o Statistical Analysis System (SAS) proporciona uma soma de quadrados negativa para a interação, quando é utilizado o proc anova.

Hocking & Speed (1975) comparam, para vários exemplos, a clássica análise de posto incompleto (superparametrizado) com a análise de posto completo (médias de caselas). Quando o conjunto de dados é desbalanceado,

ressaltam a importância das hipóteses testadas e sua associação com as somas de quadrados, principalmente quando se tem perda total de uma casela, pois o usuário raramente está atento à hipótese que é testada, importante fato que frequentemente é deixado ao capricho do método computacional empregado. No modelo superparametrizado nem todos os parâmetros são estimáveis, o que não ocorre no modelo de médias de caselas.

Os mesmos autores ressaltam, ainda, que o modelo superparametrizado é, com frequência, teoricamente mal interpretado pelo pesquisador, conduzindo a conceitos como “funções não estimáveis” e “hipóteses não testáveis”. Como resultado da superparametrização, pode-se impor certas condições nos parâmetros, mas, particularmente, quando se trata de dados desbalanceados, muita confusão existe na aplicação dessas condições. No desenvolvimento de testes de hipóteses, surgem expressões explícitas para testar hipóteses lineares da forma $\Lambda\beta = \eta$. Porém, na prática, a hipótese que está sendo testada raramente é declarada nesta forma e certas conveniências computacionais conduzem o usuário a testar hipóteses que não são, necessariamente, de interesse.

Searle et al. (1981) classificam os modelos em relação ao posto como completo ou incompleto e mostram uma relação entre eles. O modelo de posto completo é um suporte para a obtenção das somas de quadrados para o modelo de posto incompleto. Quando ocorre perda de observações, principalmente casela(s) vazia(s), as hipóteses associadas às somas de quadrados no modelo de posto completo não proporcionam correspondência perfeita com o modelo incompleto, dificultando a interpretação das hipóteses. Para facilitar a interpretação das hipóteses, o modelo de médias de caselas serviria como base.

Além do modelo com restrição Σ , aqueles autores comentam a respeito de outros dois modelos restritos, modelo com restrição zero e modelo de restrição ponderada, e mostram um exemplo com dados desbalanceados. Procuram estudar principalmente quatro sistemas estatísticos: BioMeDical

Statistic Package (BMDP2V), SAS/GLM, SAS/HARVEY e Statistical Package for Social Sciences (SPSS), ressaltando que as restrições do tipo Σ são usadas de forma explícita ou implícita pelo procedimento GLM - PROC GLM - do SAS, na soma de quadrados do Tipo III.

Boa parte das dúvidas originadas na análise de dados desbalanceados com casela(s) vazia(s), é resultado do uso do modelo superparametrizado de classificação dupla, segundo Jennings & Ward (1982). O uso do modelo de posto completo, principalmente o modelo de médias de caselas, reduz as dificuldades na interpretação das hipóteses. Os autores relacionam as hipóteses geradas no modelo superparametrizado com as proporcionadas pelo modelo de médias de caselas, concluindo que há necessidade de criar um vínculo com os usuários da estatística, para se ter uma única linguagem e, assim, uma boa comunicação.

Herr (1986) faz uma abordagem geral sobre a história dos primeiros trinta anos da análise de variância de dados desbalanceados em delineamentos fatoriais. Relata que, com a chegada dos pacotes estatísticos computacionais, a preocupação com os cálculos deveria ser um assunto morto, mas faltava ainda entender qual o método mais apropriado para análise de dados desbalanceados. Comenta que se Yates, em 1934, no caso do modelo aditivo ($\gamma_{ij} = 0$), ao declarar exatamente o modelo como o vemos, inclusive com as letras, também tivesse declarado explicitamente a hipótese que estava testando, em termos dos parâmetros do modelo, as complexidades computacionais do momento teriam sido o ponto de início da pesquisa nesses trinta anos.

O mesmo autor relata que quando o tamanho das caselas é proporcional, as somas de quadrados do Tipo I e II são iguais, na presença de interação; entretanto, há diferença entre as hipóteses.

Searle (1987) afirma que o desbalanceamento pode ser planejado, como os delineamentos experimentais quadrado latino e blocos incompletos

balanceados; ou não planejado, quando ocorrem perdas de observações de forma natural. Define dados desbalanceados como aqueles com números desiguais de observações nas caselas e os classifica em dois casos: o primeiro, com todas as caselas completas, ou seja, contendo observações, mesmo em número diferente de repetições em todos os tratamentos e o segundo, com alguma casela vazia ou com, pelo menos, uma combinação dos tratamentos ausente. Ocorrendo interação, e dependendo da forma como foram obtidas as somas de quadrados para os efeitos principais, podem ocorrer diferenças entre essas.

Camarinha Filho (1995) compara o SAS ao Statistical Graphics System (STATGRAPHICS), universalmente aceitos. O primeiro apresenta quatro tipos de somas de quadrados associadas às hipóteses e o segundo, duas delas. Utilizando dois exemplos práticos, objetivando a escolha de hipóteses mais adequadas para experimentos balanceados, desbalanceados com e sem caselas vazias, este autor mostrou que, para dados balanceados, as hipóteses não apresentaram diferenças; entretanto, para dados desbalanceados, as diferenças entre as hipóteses foram evidentes, agravando-se severamente para o caso de caselas vazias. Salienta também que é indispensável que os manuais de pacotes estatísticos sejam mais esclarecedores, a fim de possibilitar interpretações corretas das hipóteses.

Com o objetivo mostrar as diferenças existentes entre os modelos adotados pelos diferentes pacotes estatísticos utilizados na realização da análise de variância e relacioná-los com as hipóteses referentes às somas de quadrados resultantes, Nekatschalow (1997) aborda modelos de posto completo como alternativas ao modelo superparametrizado, de posto incompleto, em um conjunto de dados desbalanceados com caselas vazias. As hipóteses testadas são comparadas com as usualmente testadas pelo modelo superparametrizado, universalmente adotado devido à facilidade que proporciona ao usuário na descrição dos fatores de interesse em seus experimentos.

O autor verificou que os modelos são equivalentes, apresentando o mesmo vetor de aproximação de mínimos quadrados, portanto, com a mesma soma de quadrados dos parâmetros. A interação apresenta hipóteses semelhantes e a mesma soma de quadrados, resultado do ajuste em relação aos demais efeitos. As hipóteses determinadas pelo modelo de média de caselas são de mais simples compreensão do que do modelo superparametrizado. A ordenação dos efeitos principais é fundamental para a determinação das hipóteses do Tipo I e das somas de quadrados associadas, não interferindo na determinação das hipóteses dos demais tipos.

Santana & Iemma (2002) estudam as funções estimáveis e as hipóteses testáveis em delineamentos ortogonais e parcialmente ortogonais com três fatores, por meio do PROC GLM do SAS, objetivando possibilitar uma orientação aos pesquisadores, quanto a testabilidade e à forma das hipóteses de mais interesse, tanto para os efeitos principais, quanto para as interações.

Trabalhando com métodos alternativos para realização de testes de hipóteses em delineamentos experimentais, tendo como objetivo propor uma técnica, por meio de variáveis auxiliares, a fim de facilitar o desdobramento dos graus de liberdade de tratamentos, Nesi (2002) estudou um experimento fatorial 2×3 , inteiramente ao acaso, com dados desbalanceados e caselas vazias. Utilizando restrições paramétricas estimáveis, os resultados mostram que este método permite ao pesquisador testar hipóteses de seu interesse, com resultados equivalentes aos fornecidos pela função CONTRAST do PROC GLM do SAS.

Demarqui & Oikawa (2003) verificaram, por meio de um modelo de médias de caselas, com um fator de classificação e uma covariável, que interpretações equivocadas das somas de quadrados que realmente testam as hipóteses de interesse, podem levar a resultados equivocados e, também que a constituição de uma hipótese de interesse depende dos propósitos do estudo e dos dados disponíveis para a análise.

2.1 Caracterização do modelo linear

O modelo linear de Gauss-Markov, segundo Rao (1945) e Searle (1971), dentre outros, pode ser denotado por:

$$y = X\theta + e, \quad (1)$$

em que:

${}_n y_1$: é um vetor de realizações de variáveis aleatórias;

${}_n X_p$: é uma matriz conhecida (matriz de delineamento) de posto $k \leq \min(n,p)$;

${}_p \theta_1$: é um vetor de parâmetros desconhecidos;

${}_n e_1$: é um vetor de variáveis aleatórias não observáveis, tal que $e \sim N(\phi, I\sigma^2)$.

2.2 Estimação

As idéias apresentadas foram baseadas e adaptadas de conceitos que podem ser vistos em Rao (1945), Searle (1971) e Lemma (1988).

No modelo linear G.M., de acordo com (1) tem-se:

$$i) E[y] = E[X\theta + e] = X\theta + \phi = X\theta \quad (3)$$

$$\text{Var}[y] = \text{Var}[X\theta + e] = \text{Var}[e] = I\sigma^2 \quad (4)$$

ii) Como a matriz de covariâncias dos erros pode assumir caracterizações mais complexas do que apenas $I\sigma^2$, podem ser citadas duas situações:

- sendo $\text{Var}[e]$ ainda diagonal mas não uma matriz identidade com $I\sigma^2$ em evidências e as demais pressuposições inalteradas, $e \sim N(\phi, D\sigma^2)$, o modelo é definido como modelo linear ponderado de Gauss-Markov, ficando as equações normais como: $X'D^{-1}X\theta = X'D^{-1}y$, em que $\theta_p^0 = (X'D^{-1}X)^{-1}X'D^{-1}y$ é uma solução de mínimos quadrados ponderados.

- Tendo a matriz de variâncias e covariâncias, $\text{Var}[e]$, uma estrutura mais generalizada, e $\sim N(\phi, V\sigma^2)$, o modelo é definido como modelo linear generalizado de Gauss-Markov, ficando as equações normais como: $X'V^{-1}X\theta = X'V^{-1}y$, em que $\theta_p^0 = (X'V^{-1}X)^{-1}X'V^{-1}y$ é uma solução de mínimos quadrados generalizados, onde V é uma matriz positiva definida.

2.2.1 Estimação em modelos lineares

Considerando o modelo linear G.M., se X tem posto coluna completo, como em geral ocorre nos problemas de análise de regressão, todos os parâmetros são estimáveis individualmente, então $X'X$ é positiva definida e, portanto, não singular. Assim sendo, a solução única do Sistema de equações normais (SEN) é dada por $\hat{\theta} = (X'X)^{-1}X'y$, em que $\hat{\theta}$ é um estimador não viesado para θ , pois:

$$E[\theta^0] = E[(X'X)^{-1}X'y] = (X'X)^{-1}X'E[y] = (X'X)^{-1}X'X\theta = \theta \quad (5)$$

Pode-se dizer que se X tem posto coluna completo, então, a solução de quadrados mínimos θ coincide com o estimador de quadrados mínimos para θ (Lemma, 1988).

Quando X tem posto coluna incompleto ($k_{[n, X_p]} < p$), as equações normais não têm solução única, exceto em condições especiais como as restrições ou reparametrizações. Portanto $X'X$ não é positiva definida, $X'X$ é singular, implicando que não existe $(X'X)^{-1}$ e o sistema de equações normais é indeterminado e possui infinitas soluções. Dentre as possíveis soluções, para qualquer inversa generalizada, tem-se $\theta^0 = (X'X)^G X'y$. Então θ^0 não é um estimador não viesado para θ :

$$E[\theta^0] = E[(X'X)^G X'y] = (X'X)^G X'X\theta = H\theta \quad (6)$$

em que $H = (X'X)^G X'X$, mas sim para funções das componentes de θ que estão no vetor $H\theta$, já que existem outras. As funções determinadas por $H\theta$ são chamadas funções paramétricas estimáveis.

Se X tem posto coluna incompleto, os componentes do vetor θ não são individualmente estimáveis, então as soluções θ^0 não têm valor por si só, mas por meio de X , pois $X\theta^0 = Py$ é invariante para qualquer solução das equações normais. Não faz sentido obter $\text{Var}(\theta^0)$, mas sim de funções estimáveis $\lambda'\theta$, sendo $\lambda = X'a$ (em que 'a' é um vetor de possíveis estimadores não viesados).

2.2.2 Funções lineares paramétricas estimáveis

Definição (Rao, 1945): uma função linear paramétrica, $\lambda'\theta$, é estimável no modelo $y = X\theta + e$, G.M. se, e somente se, existe uma combinação linear das observações, $a'y$, tal que $E[a'y] = \lambda'\theta$, com $a \neq 0$. Se não existe $a'y$, tal que a igualdade se verifique, então $\lambda'\theta$ é dita não estimável nesse modelo.

Em demonstrações, esta definição é muito útil, mas, em aspectos práticos é bem pouco operacional, pois depende da dimensão do vetor y das observações.

Teorema (Rao, 1945): uma condição necessária e suficiente para que a função paramétrica $\lambda'\theta$ seja estimável no modelo $y = X\theta + e$, G.M. é $\lambda' \in C(X')$.

Teorema: cada condição dada a seguir é necessária e suficiente para que $\lambda'\theta$ seja estimável no modelo $y = X\theta + e$ G.M.:

i) $k[X] = k[X|\lambda]$

ii) $k[X'X] = k[X'X|\lambda]$

$\lambda'\theta$ é estimável $\Leftrightarrow k[X'] = k[X'|\lambda] = k[X'X|\lambda]$.

Corolário da definição: $\lambda'\theta$ é estimável se, e somente se, $X'a = \lambda$.

2.2.3 O melhor estimador linear não viesado

Sendo $\lambda'\theta$ estimável, pode apresentar vários estimadores não viesados, no entanto, tem-se interesse naquele que é classificado como melhor.

Definição: dado o modelo $y = X\theta + e$, G.M., sejam todas as possíveis combinações $a'y$, tais que $E[a'y] = \lambda'\theta$. Dentre elas, seja a^*y . Então a^*y é definido como o melhor estimador linear não viesado, *Best Linear Unbiased Estimator*, conhecido por BLUE, de $\lambda'\theta$, se e somente se:

$$\text{var}[a^*y] = \min\{\text{Var}[a'y], \forall a'y: E[a'y] = \lambda'\theta\},$$

O BLUE de $\lambda'\theta$ será representado por $\widehat{\lambda'\theta}$.

Com a finalidade de ressaltar sua importância e facilitar demonstrações, Lemma (1988, 1993) apresentou algumas regras para obtenção do BLUE.

Definição: dado o modelo $y = X\theta + e$, G.M., se a função paramétrica $\lambda'\theta$ é estimável, então, o sistema consistente $X'X\rho = \lambda$ é definido como sistema de equações normais associadas (ENA) a $\lambda'\theta$.

Se $\lambda'\theta$ é estimável, então $\lambda \in C(X'X)$ e existe ρ_1 tal que, $X'X\rho = \lambda$, e sempre consistente já que $\lambda'\theta$ é estimável. Então $\lambda' = \rho'X'X$.

Teorema: o sistema de equações normais associadas (ENA) é consistente.

Teorema: o vetor $X\rho^\circ$ é invariante para qualquer ρ° solução das equações normais associadas a $\lambda'\theta$, $X'X\rho = \lambda$.

Teorema: se $\lambda'\theta$ é uma função estimável no modelo linear $y = X\theta + e$, G.M., então seu BLUE pode ser obtido de modo único por: $\lambda'\theta = \rho^\circ X'y$, em que ρ° é qualquer solução do sistema de equações normais associadas a $\lambda'\theta$.

Teorema de Gauss Markov: se $\lambda'\theta$ é uma função estimável no modelo $y = X\theta + e$, G.M., então seu BLUE pode ser obtido de modo único por

$\lambda'\theta = \lambda'\theta^\circ$, em que θ° é qualquer solução das equações normais. Sua variância é dada por $var[\hat{\lambda'\theta}] = var[\lambda'\theta] = \lambda'(X'X)^G\lambda\sigma^2$.

2.2.4 Regras práticas de estimabilidade

Uma função paramétrica $\lambda'\theta$ é estimável se, e somente se $\lambda \in C(X')$, isto é, $\lambda' \in L(X)$. Estendendo o conceito de função linear paramétrica no modelo linear G.M. para um conjunto L de funções estimáveis, $L\theta$ dá um conjunto de funções paramétricas estimáveis, impondo que L tenha posto coluna completo. Para facilidade de identificação de funções serão apresentadas quatro regras:

- Funções básicas estimáveis: dado o modelo $y = X\theta + e$, G.M., em que $e \sim N(\phi, I\sigma^2)$, uma função linear paramétrica $\lambda'\theta$ é dita estimável se, e somente se, existir ao menos uma combinação linear das observações, dada por $a'y$, tal que $E[a'y] = a'E[y] = a'X\theta = \lambda'\theta$. Como $E[y] = X\theta$, então $X\theta$ é estimável e também as suas componentes. Assim, de acordo com a caracterização (parametrização) de $y = X\theta + e$, $X\theta$ indica um conjunto de funções que são estimáveis (lemma, 1985; lemma, 1988).
- Combinações lineares de funções estimáveis: Definição: dado o modelo $y = X\theta + e$, (G.M.) e seja θ um vetor de parâmetros, tal que, $\theta' = [\mu \ \alpha_1 \ \alpha_2 \dots \alpha_a \ \beta_1 \ \beta_2 \dots \beta_b \ \gamma_1 \ \gamma_2 \dots \gamma_c \dots]$, então funções lineares paramétricas do tipo $\sum_{i=1}^a c_i \alpha_i$, $\sum_{j=1}^b c_j \beta_j$, $\sum_{l=1}^c c_l \gamma_l$, ..., são ditas contrastes, se $\sum_{i=1}^l c_i = 0$. Os contrastes constituem um caso particular das combinações lineares das funções estimáveis ($\sum_{i=1}^l c_i = 0$), portanto, sendo estimáveis. Formam um subconjunto das funções estimáveis.

- Usando as equações normais: dado o modelo linear $y = X\theta + e$, uma função paramétrica $\lambda'\theta$ é estimável se, e somente se, puder ser obtida como combinação linear dos elementos do vetor $X'X\theta$ (lado esquerdo das equações normais). Quando isso ocorre, o BLUE de $\lambda'\theta$ é dado pela mesma combinação linear dos componentes de $X'y$ (lado direito das equações normais).
- Invariância de $\lambda'H$: dado o modelo linear $y = X\theta + e$, sendo $\lambda'\theta$ estimável, então existe o vetor $a : X'a = \lambda \Rightarrow \lambda' = a'X$, pós-multiplicando por $H = (X'X)^{-1}X'X$, tem-se $\lambda'H = a'X(X'X)^{-1}X'X = a'XX'X = a'X = \lambda' \therefore \lambda'H = \lambda'$.

2.3 Conectividade

De acordo com Santos (1994), conectividade é definida por Weeks & Willians (1964) como: “um modelo é conectado se todo contraste simples de diferenças de dois níveis de um fator é estimável” e por Raghavaro (1971): “um delineamento é dito conectado se todos os contrastes elementares são estimáveis.” Caso contrário, o delineamento é dito desconexo.

Searle (1987) faz uma discussão detalhada sobre dados conectados. Na ocorrência de caselas vazias, a regra prática de estimabilidade por meio das funções básicas estimáveis não detecta com facilidade casos em que haja desconexão.

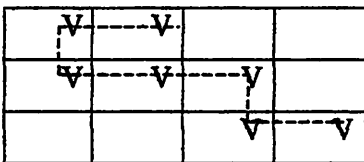


FIGURA 1 - Conjunto conexo

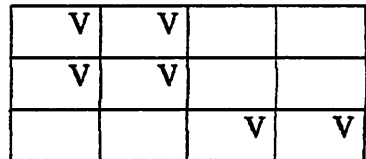


FIGURA 2 – Conjunto desconexo

Para detectar a conectividade, utiliza um exemplo no qual aplica a regra descrita como “geométrica”: os dados são ditos conexos, se for possível

caminhar nas caselas, entre linhas e colunas (Figura 1) e ditos desconexos, se ocorrer o contrário (Figura 2).

No modelo de classificação dupla sem interação, com os dados conexos, todas as médias de caselas são estimáveis, enquanto que, no modelo com interação, somente as médias de caselas não vazias são estimáveis.

2.4 Testabilidade

Uma hipótese é testável se for constituída de funções estimáveis no modelo. Os conceitos de testabilidade se relacionam aos de estimabilidade e conectividade.

2.5 Modelo superparametrizado (Modelo-S)

De acordo com Searle (1987), um modelo é dito superparametrizado quando explicita um parâmetro para cada um dos fatores envolvidos. Inclusive, Iemma (1995) e outros alegam que, apesar do Modelo-S ser de uso bastante difundido, possui alguns inconvenientes, dentre eles, o de possuir mais parâmetros do que o número de caselas para estimá-las, o que se agrava na presença de caselas vazias.

Doravante denominado Modelo-S e denotado matricialmente como em (1), por $y = X\theta + e$, pode assumir diversas caracterizações, de acordo com o interesse do pesquisador, por meio de parametrizações sucessivas. Segundo Speed & Hocking (1976), dentre outros, partindo de um modelo mais simples até o modelo com dois fatores fixos e com interação, tem-se:

$$y = X_1\theta_1 + e_1 \Rightarrow y_i = \mu + e_i \quad (S.1)$$

$$y = X_2\theta_2 + e_2 \Rightarrow y_{ik} = \mu + \alpha_i + e_{ik} \quad (S.2)$$

$$y = X_3\theta_3 + e_3 \Rightarrow y_{ij} = \mu + \beta_j + e_{jk} \quad (S.3)$$

$$y = X_4\theta_4 + e_4 \Rightarrow y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad (S.4)$$

$$y = X\theta + e \Rightarrow y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad (S.5)$$

em que:

$i = 1, 2, \dots, a$; níveis do fator A (α);

$j = 1, 2, \dots, b$; níveis do fator B (β);

$k = 1, 2, \dots, n_{ijt}$ repetições;

X_1 : primeira coluna da matriz X;

θ_1 : vetor do parâmetro μ ;

y_i : observação i -ésima;

μ : constante em todas as observações;

X_2 : matriz formada pelas colunas da matriz X correspondentes a μ e ao fator A;

θ_2 : vetor formado pelos parâmetros μ e α_i ;

y_{ik} : efeito entre o i -ésimo nível do fator A e a k -ésima repetição;

α_i : efeito do i -ésimo nível do fator A;

X_3 : matriz formada pelas colunas da matriz X correspondentes a μ e ao fator B;

θ_3 : vetor formado pelos parâmetros μ e β_j ;

β_j : efeito do j -ésimo nível do fator B;

X_4 : matriz formada pelas colunas da matriz X correspondentes a μ e aos fatores A e B;

θ_4 : vetor formado pelos parâmetros μ , α_i e β_j ;

γ_{ij} : efeito da interação entre o i -ésimo nível do fator A e o j -ésimo do fator B;

X: matriz do delineamento;

θ : vetor de parâmetros para o modelo completo;

y_{ijk} : valor observado na k -ésima repetição, avaliada no i -ésimo nível do fator A e no j -ésimo nível do fator B;

e_{ijk} : erro aleatório associado à observação y_{ijk} , considerado independente e normalmente distribuído com média zero e variância constante.

De acordo com Camarinha Filho (1995), observa-se que:

- de (S.1) a (S.5) há uma parametrização sucessiva e ordenada;
- todas as caracterizações do modelo são de posto coluna completo, exceto (S.1);
- (S.1) corresponde a um modelo aditivo, ou seja, todos os dados podem ser descritos como a adição de uma constante a um erro aleatório;
- os modelos (S.2) e (S.3) são denominados modelos de classificação simples e são providos da adição de um efeito fixo (α_i ou β_j) ao modelo anterior;
- os modelos (S.4) e (S.5) são denominados modelos de classificação dupla cruzadas, sem e com interação, respectivamente.

Wechsler (1998) relata que o modelo (S.5) não define claramente o que se entende por efeito de A, efeito de B e efeito da interação, em que os parâmetros α , β , e γ não são estimáveis, contanto que alguma restrição seja adotada (Searle, 1971), como as restrições usuais:
$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \gamma_{ij} = \sum_{j=1}^b \gamma_{ij} = 0,$$

recebendo assim o nome de modelo superparametrizado. O autor relata que num fatorial desbalanceado podem variar, além das somas de quadrados de A, B e interação, também as hipóteses de nulidade a elas associadas, dependendo das restrições e procedimentos computacionais empregados, causando os mal-entendidos sobre as hipóteses testadas pelos aplicativos.

A estimação de e , vetor de variáveis aleatórias não observáveis, de acordo com Searle (1987), depende da caracterização assumida e da solução do sistema de equações lineares, que é inconsistente. Uma solução pode ser obtida por meio do SEN, $X'X\theta = X'y$, sempre consistente (Searle, 1971; lemma, 1987). Qualquer solução exata, θ^o , do SEN é solução aproximada de mínimos quadrados para $y = X\theta$ inconsistente.

De acordo com os autores, se a matriz X tem posto coluna incompleto, $X'X$ é singular e o SEN é indeterminado. As soluções podem ser obtidas e são

soluções aproximadas de mínimos quadrados, dentre outras, por exemplo, $\theta^0 = (X'X)^G X'y$, em que $(X'X)^G$ é uma inversa generalizada qualquer de $X'X$. As soluções variam de acordo com a inversa generalizada escolhida. Quando X tem posto coluna incompleto, os componentes do vetor θ não são individualmente estimáveis.

Segundo Lemma (1995), o modelo superparametrizado é parte integrante da história dos modelos lineares e muito tem auxiliado os pesquisadores das ciências aplicadas, pois exhibe explicitamente os parâmetros sobre os quais estão geralmente concentradas as hipóteses de interesse.

O Modelo-S é utilizado pelo PROC GLM do SAS na apresentação de suas funções estimáveis e das estimativas de cada um dos parâmetros do modelo linear (Nesi, 2002).

2.5.1 Obtenção de somas de quadrados

Searle (1987) afirma que, para dados balanceados, obtêm-se as mesmas somas de quadrados para os diferentes sistemas estatísticos computacionais, o que não ocorre com dados desbalanceados. Para um mesmo conjunto de dados desbalanceados, os vários métodos não são facilmente interpretados e nem todos os sistemas estatísticos fazem a mesma análise, pois, além de utilizarem modelos distintos, a ordenação dos fatores é fundamental. Com isso, uma variedade de somas de quadrados associadas a uma mesma causa de variação pode ser encontrada nos diferentes sistemas estatísticos, gerando dúvidas ao usuário, em interpretar e escolher, dentre as saídas, qual seria a análise de variância mais adequada para seu conjunto de dados, o que pode levar a erros.

O autor estabelece a relação entre os tipos de somas dos quadrados do SAS, I, II, III e IV, caracterizadas, respectivamente, como seqüencial, cada uma

ajustada para os demais, restrições com Σ , a soma de quadrados depende da posição da casela vazia e conseqüentes hipóteses associadas:

- Tipo I = II = III = IV para os dados balanceados;
- Tipo II = III = IV para modelos sem interação;
- Tipo III = IV para os modelos com todas as caselas completas.

Para cada tipo de somas de quadrados, as saídas do GLM do SAS fornecem as funções estimáveis, proporcionando cada função uma base para o usuário associar as hipóteses a cada soma de quadrados.

Sendo $\hat{y} = X\theta^\circ$, a aproximação de mínimos quadrados para o vetor de observações y , admitindo-se o modelo (1), invariante para qualquer θ° solução das equações normais: $\hat{y} = X(X'X)^G X'y = Py$, em que $P = X(X'X)^G X' = XX^+$ = XX^1 , é o projetor ortogonal de y sobre o espaço gerado pelas colunas de X , $C(X)$, representado graficamente na Figura 3.

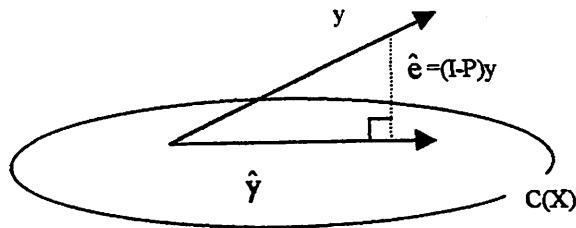


FIGURA 3. Projeção ortogonal de y sobre $C(X)$.

Com a ortogonalidade, por meio do Teorema de Pitágoras, pode-se obter a decomposição ortogonal:

$$\|y\|^2 = \|\hat{y}\|^2 + \|\hat{e}\|^2 \quad (7)$$

em que,

- $\|y\|^2 = y'y = \sum y^2$: soma de quadrados total, não corrigida;
- $\|\hat{y}\|^2 = \hat{y}\hat{y} = \sum \hat{y}^2 = y'Py = \theta^{\circ\circ} X'y$: soma de quadrados de parâmetros;

- $\|\hat{e}\|^2 = \hat{e}'\hat{e} = \sum \hat{e}^2 = y'y - \theta'X'y = y'(I-P)y$: soma de quadrados do resíduo.

Segundo Lemma (1993), com dados balanceados há ortogonalidade e obtém-se P de forma trivial, porém, com dados desbalanceados é mais difícil a obtenção dos projetores ortogonais, no que se refere às hipóteses sobre linhas, colunas e interação, principalmente na presença de caselas vazias.

Por intermédio de outras partições sobre X e θ , pode-se obter outras regras para determinar hipóteses sobre funções lineares paramétricas de interesse. Em geral, na partição da soma de quadrados dos parâmetros estão contidas as hipóteses de interesse do pesquisador.

Lemma (1995) faz um alerta sobre o valor da soma de quadrados da interação em todas as análises para um mesmo conjunto de dados, que é invariante para qualquer tipo de soma de quadrados. Ele firmou o postulado de que “modelo com interação é bom para testar interação”, sendo mais agravante na ocorrência de caselas vazias, pois o teste da interação não testa que todas as interações são nulas, apenas aquelas que são estimáveis e, portanto, testáveis.

2.5.2 Notação R()

Consagrada por Searle (1971), a utilização da notação R() está ligada ao método de ajuste de constantes, descrito por Yates (1934). É um procedimento para obtenção de somas de quadrados por meio da redução da soma de quadrados totais ao ajustar um modelo particular de interesse.

Speed & Hocking (1976) colocam as idéias de parametrização sucessiva por meio do modelo superparametrizado. As parametrizações sucessivas e ordenadas facilitam a interpretação da notação R() e de certas somas de quadrados a elas associadas. O termo R() pode ser interpretado como a medida de variação em y explicada pelo modelo ajustado.

Os autores discutem a redução nas somas de quadrados para dados desbalanceados. Alegam que apesar de haver apenas uma notação $R(\cdot)$, há dois procedimentos distintos quanto ao cálculo da redução. Considerando o modelo (1), representado por $Y = X\theta + e$, evidenciam e ilustram as diferenças entre os procedimentos, demonstrando que um não é um caso especial do outro.

Procedimento 1: assumindo o modelo (1), de posto incompleto e superparametrizado, Searle (1971) considera que este procedimento consiste em aplicar a notação $R(\cdot)$ definida por $R(\theta) = \hat{\theta}'X'Y$, sendo $\hat{\theta}$ uma solução para o sistema de equações normais $X'X\hat{\theta} = X'Y$. Pelo particionamento da matriz em $X = [X_1 : X_2]$ e $\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$, e tomando-se o modelo $Y = X_2\theta_2 + e$, portanto $R(\theta_2) = \hat{\theta}_2'X_2'Y$, sendo que $\hat{\theta}_2$ é uma solução para $X_2'X_2\hat{\theta}_2 = X_2'Y$, determina-se a redução: $R(\theta_1|\theta_2) = R(\theta_1, \theta_2) - R(\theta_2)$.

Procedimento 2: consiste em impor ao modelo acima descrito um conjunto de restrições não estimáveis, obtendo um modelo reparametrizado de posto completo, $Y = X^*\theta^* + e$. Faz-se o particionamento semelhante ao modelo anterior, obtendo-se $Y = X_2^*\theta_2^* + e$. Aplica-se notação $R(\cdot)$ definida por $R(\hat{\theta}_2^*) = (\hat{\theta}_2^*)'(X_2^*)'Y$, em que $\hat{\theta}_2^* = [(X_2^*)'X_2^*]^{-1}(X_2^*)'Y$ é solução única para o sistema de equações normais.

Speed & Hocking apresentam distinção entre os procedimentos: no procedimento 2, $R(\hat{\theta}_2^*)$ não é sempre definido; caso contrário, pode apresentar falta de unicidade; o procedimento 2 não é um caso especial do 1, no caso de dados desbalanceados, apresentando somas de quadrados diferentes.

Estes autores ressaltam que, em ambos os procedimentos as hipóteses testadas não são claras e a hipótese geral $H\theta = \gamma$ não pode ser testada pelos procedimentos, tornando a análise limitada. Conclui-se que, tratando-se de dados

desbalanceados, as somas de quadrados fornecidas pelos sistemas computacionais poderão ser interpretadas erroneamente se o usuário tomar a notação $R(\)$ no seu valor aparente.

O Modelo-S é excelente para a interpretação da notação $R(\)$. Searle (1987) sugere a redução da soma de quadrados e como exemplo do ajuste do modelo (S.5) com a notação $R(\)$, têm-se:

$$R(\theta) = y'X(X'X)^{-1}X'y = \theta'X'y = S.Q. \text{ Parâmetros} \quad (8)$$

em que θ^0 é uma solução para $X'X\theta = X'y$.

Particionando convenientemente a matriz do delineamento (X) e o vetor de parâmetros (θ), usando a notação $R(\)$, pode-se associar a cada modelo, do mais simples ao mais completo, uma soma de quadrados dos parâmetros. Por exemplo, segundo Lemma & Pierri (1997), dentre outros:

$$X = [X_1 : X_2 : X_3 : X_4]; \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

- $y_i = \mu + e_i \Rightarrow y = X_1\theta_1 + e_1; \theta_1' = [\mu]$

- $\rightarrow R(\mu) = \theta_1'X_1'y = \text{correção} = SQ_{\text{Parâmetros}} \text{ do modelo considerando } \mu.$

- $y_{ik} = \mu + \alpha_i + e_{ik} \Rightarrow y = X_2\theta_2 + e_2; \theta_2' = [\mu, \alpha]$

- $\rightarrow R(\mu, \alpha) = \theta_2'X_2'y = SQ_{\text{Parâmetros}} \text{ do modelo considerando } \mu \text{ e } \alpha.$

$$R(\alpha|\mu) = R(\mu, \alpha) - R(\mu) \quad (9)$$

- $y_{ijk} = \mu + \beta_j + e_{ijk} \Rightarrow y = X_3\theta_3 + e_3; \theta_3' = [\mu, \beta]$

- $\rightarrow R(\mu, \beta) = \theta_3'X_3'y = SQ_{\text{Parâmetros}} \text{ do modelo considerando } \mu \text{ e } \beta$

$$R(\beta|\mu) = R(\mu, \beta) - R(\mu) \quad (10)$$

- $y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \Rightarrow y = X_4\theta_4 + e_4; \theta_4' = [\mu, \alpha, \beta]$

- $\rightarrow R(\mu, \alpha, \beta) = \theta_4'X_4'y = SQ_{\text{Parâmetros}} \text{ do modelo considerando } \mu, \alpha \text{ e } \beta$

$$R(\alpha|\mu,\beta) = R(\mu,\alpha,\beta) - R(\mu,\beta) \quad (11)$$

$$R(\beta|\mu,\alpha) = R(\mu,\alpha,\beta) - R(\mu,\alpha) \quad (12)$$

$$\blacksquare \quad y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \Rightarrow y = X\theta + e; \quad \theta' = [\mu, \alpha, \beta, \gamma]$$

$$\rightarrow R(\mu,\alpha,\beta,\gamma) = \theta' X' y = SQ_{\text{Parâmetros}} \text{ do modelo completo}$$

$$R(\gamma|\mu,\alpha,\beta) = R(\mu,\alpha,\beta,\gamma) - R(\mu,\alpha,\beta) \quad (13)$$

2.6 Funções estimáveis

A ocorrência de caselas com dados perdidos é um problema conhecido há muitos anos e, apesar de terem sido desenvolvidos métodos especializados para certos casos, nenhum é universalmente aceito como único e correto, segundo Freund (1980), que estudou a perda de observações e de caselas inteiras, com suas implicações nas funções estimáveis e na formulação de hipóteses. O autor utiliza, como exemplo, um experimento fatorial 2 x 3 com uma casela perdida e compara os procedimentos e a análise dos dados por dois métodos: (a) pelo PROC GLM do SAS que usa a metodologia das funções estimáveis; e, (b) pelo método da reparametrização. Ele analisou apenas dois tipos de funções estimáveis do GLM, os Tipos III e IV, pois, na maioria das análises, são idênticos e normalmente correspondem às estimativas associadas à soma parcial de quadrados, porém na presença de caselas vazias, comentou as diferenças existentes. O procedimento do GLM, apesar de mais difícil, fornece meios para o usuário especificar uma função estimável dos parâmetros do modelo.

O autor relata que, quando se trata de casela perdida, os resultados apresentados pelas duas metodologias são confusos, não conduzindo a uma conclusão, pois os métodos incluem o uso de outros tipos de inversas generalizadas, estimando dados para as caselas perdidas e impondo outras restrições, dificultando a obtenção das funções úteis aos pesquisadores. As

somas de quadrados podem diferir entre si, dependendo de qual casela foi perdida. Não há sustentação que o procedimento GLM tem a solução correta para tais situações, mas fornece informação sobre a natureza de possíveis estimativas e, assim, permite ao pesquisador decidir se os resultados têm qualquer significado prático.

Searle et al. (1980) discorrem que os quatro tipos de funções estimáveis fornecidas pelo PROC GLM do SAS possuem coeficientes arbitrários e que são obtidas por expressões gerais, constituindo as bases para testar hipóteses por meio da tradicional estatística F para dados desbalanceados. As somas de quadrados são interpretadas pelas funções estimáveis, sendo mais informativas que os valores fornecidos por $R(\)$:

- Tipo I: valores $R(\)$ são seqüenciais;
- Tipo II: valores $R(\)$ para cada fator ajustado a todos os outros que não o contêm;
- Tipo III: valores $R(\)$ para cada fator ajustado a todos os outros;
- Tipo IV: fixa uma hipótese e testa-a pela estatística F; contrariamente aos tipos anteriores, não pretende proporcionar funções estimáveis para explicar a soma de quadrados predefinida.

Os autores relacionam a idéia de estimabilidade com o método dos mínimos quadrados.

Searle (1987) ressalva que, no teste de hipóteses decorrente do conceito de funções estimáveis, as hipóteses sobre diferenças de efeitos de filas (linhas ou colunas) não são testáveis, ou seja, hipóteses referentes a contraste de filas.

Tomando-se o modelo linear (1) em que $E[y] = X\theta$, o interesse principal da análise é estimar os elementos de θ , no modelo de posto coluna completo, ou alguma combinação linear de seus elementos. Isto é realizado pelo cálculo de combinações lineares dos componentes do vetor de observações y , a fim de estimar uma função linear específica de θ , expressa por $L\theta$. Por exemplo,

Ky terá, como valor esperado, $E[Ky] = KX\theta$, portanto, $L\theta$ é estimável se uma combinação linear das linhas de X , que é igual a $L = KX$, puder ser encontrada.

O manual do SAS (SAS, 1990) cita que o procedimento GLM, utilizando essas idéias, fornece uma base para o estudo das estimabilidade de funções. Opta por calcular uma inversa generalizada $G2$, que contém numerosos zeros e linhas não nulas linearmente independentes. O GLM usa as linhas não nulas para representar $L = (X'X)^{G2}X'X$, análoga à matriz H , pois G , não única, é uma inversa generalizada qualquer de $X'X$, gerando vários conjuntos equivalentes de funções estimáveis, enquanto que a inversa generalizada $G2$, de $(X'X)^{G2}X'X$, é única e H fornece um “ L ” único.

Com base em $L\theta$, o PROC GLM gera quadros com os coeficientes da forma geral e dos quatro tipos de funções estimáveis, denominadas Tipo I, II, III e IV, que podem ser utilizados para o reconhecimento das hipóteses testadas, principalmente em casos de desbalanceamentos. A forma geral das funções estimáveis é obtida por $L\theta$, envolvendo todos os parâmetros do modelo e a cada linha não nula linearmente independente da matriz $L = (X'X)^{G2}X'X$ é associado o símbolo L , que representa o posto máximo da hipótese associada.

Se a matriz X não tem posto coluna completo, de acordo com lemma (1993), então o vetor θ^0 não é estimador não viesado no conceito de Rao (1945) e $E[\theta^0] = (X'X)^G X'X\theta = H\theta$, portanto, a função $H\theta$ é estimável.

2.7 Hipóteses estatísticas

Segundo Speed et al. (1978), na análise de modelos lineares para delineamentos experimentais com dados balanceados há um consenso geral entre as hipóteses testadas para efeitos principais e interações e, infelizmente, quando se trata de uma análise desbalanceada, há vários métodos que não conduzem a um resultado único. Embasados nisso, o propósito dos autores é

rever os métodos de análise e relacionar as somas de quadrados associadas às hipóteses lineares que estão sendo testadas.

Os autores utilizam um modelo de médias de caselas (Modelo-M), $Y = W\mu + e$, sendo μ o p vetor da média de caselas, sujeito às restrições $G\mu = 0$, e G uma matriz, n x p, das conhecidas relações lineares entre as médias de caselas. Como exemplo, consideram um modelo com dois fatores sem interação, $y_{ijk} = \mu_{ij} + e_{ijk}$, sujeito a $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$, em que $i, i' = 1 \dots a$; $j, j' = 1 \dots b$; $e, k = 0, 1, \dots, n_{ij}$. O modelo sem as restrições, com interação, geralmente escrito na forma superparametrizada como $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$, e com as restrições, pode ser reparametrizado como $y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$, que é o modelo padrão de dois fatores sem interação.

TABELA 1. Hipóteses associadas ao modelo de classificação dupla.

		Efeitos	
Linhas Efeito principal A	H1	$H_0^1: \bar{\mu}_{i.} = \bar{\mu}_{i'}$	
	H2	$H_0^2: \sum_j \frac{n_{ij}\mu_{ij}}{n_{i.}} = \sum_j \frac{n_{i'j}\mu_{i'j}}{n_{i'}}$	
	H3	$H_0^3: \sum_j (n_{ij})\mu_{ij} = \sum_{i'} \sum_j \frac{n_{ij}n_{i'j}\mu_{i'j}}{n_{i'}}$	
	H4	$H_0^4: \bar{\mu}_{i1.} = \bar{\mu}_{i'1}$	
Colunas Efeito principal B	H5	$H_0^5: \bar{\mu}_{.j.} = \bar{\mu}_{.j'}$	
	H6	$H_0^6: \sum_i \frac{n_{ij}\mu_{ij}}{n_{i.}} = \sum_i \frac{n_{ij'}\mu_{ij'}}{n_{j'}}$	
	H7	$H_0^7: \sum_i (n_{ij})\mu_{ij} = \sum_{j'} \sum_i \frac{n_{ij}n_{ij'}\mu_{ij'}}{n_{i.}}$	
	H8	$H_0^8: \bar{\mu}_{1j.} = \bar{\mu}_{1j'}$	
Interação	H9	$H_0^9: \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$	

Notas: Sendo para todo o i, i', j, j' , com $i \neq i'$. $H_1^0 \dots H_1^9$: diferentes hipóteses possíveis; $\bar{\mu}_{i.}$: média ponderada das caselas da linha i; n_{ij} : freqüência da casela ij; μ_{ij} : média da casela ij; n_i : soma das freqüências das caselas da linha i; $n_{j'}$: soma das freqüências das caselas da coluna j; $\bar{\mu}_{.j.}$: média ponderada das caselas da coluna j.

Na Tabela 1, os autores relacionaram nove hipóteses aos diferentes métodos, ressaltando ainda que, na ocorrência de casela vazia, as hipóteses H1, H4, H5, H8 e H9 não podem ser testadas.

Para os efeitos principais, há quatro formas de calcular as somas de quadrados; ocorrendo interação, independente do método, o mesmo resultado é obtido. Em um modelo sem interação, existe uma equivalência das somas de quadrados obtidas para os efeitos principais.

Speed et al. (1978) fazem uma descrição detalhada dos métodos comuns de cada análise, mostrando as características essenciais e as relações, inclusive para um conjunto de dados desbalanceados, com casela vazia e interação. A Tabela 2 apresenta, de forma resumida, as somas de quadrados descritas usando a notação $R(\)$, pelo método do ajuste de constantes (modelo superparametrizado).

TABELA 2. Somas de quadrados obtidas pelo método do ajuste de constantes.

Efeitos	Hipóteses	Soma de quadrados
Fator A	H1	$^{(1)}R(\alpha \mu)$
	H2	$^{(1)}R(\alpha \mu,\beta)$
	H3	$^{(2)}R(\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma})$
Fator B	H5	$^{(1)}R(\beta \mu)$
	H6	$^{(1)}R(\beta \mu,\alpha)$
	H7	$^{(2)}R(\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma})$
Interação	H9	$R(\gamma \mu,\alpha,\beta)$

Notas: μ : constante inerente a todas as observações; α : efeito do fator A; β : efeito do fator B; γ : efeito da interação AxB; $\hat{\alpha},\hat{\beta},\hat{\gamma}$: o ponto indica que o efeito do fator está ajustado para todos os outros efeitos considerados no modelo.

⁽¹⁾: Podem ser obtidas pelo método do ajuste de constantes no modelo irrestrito.

⁽²⁾: Podem ser obtidas pelo método do ajuste de constantes no modelo restrito.

Segundo os autores, dentre os métodos citados, quatro são incorporados ao SAS-76, como parte do PROC GLM:

- Método Tipo I: o método de ordenação a priori de Overall & Spiegel (1969), similar ao método do ajuste de constantes, em que H2 e H7 são as hipóteses para efeitos principais.
- Método Tipo II: o método de delineamento experimental de Overall & Spiegel (1969) é o mais comum, baseado no próprio método de constantes, sendo as somas de quadrados ajustadas, $R(\alpha|\mu,\beta)$ e $R(\beta|\mu,\alpha)$, usadas para testar efeitos principais, em que H3 e H7 são as hipóteses testadas.
- Método Tipos III e IV: equivalentes ao método de médias ou mínimos quadrados completos de Overall & Spiegel (1969), em que H1 e H5 são as hipóteses dos efeitos principais.

Quando existem caselas vazias, as somas de quadrados Tipos III e IV não coincidem necessariamente. Os autores relatam ainda que alguns programas, inclusive o Tipo III do SAS, utilizam uma modificação no procedimento R(), $\gamma_{ij} = 0$ se $n_{ij} = 0$, generalizando as somas de quadrados, mas nenhuma indicação é dada sobre as hipóteses que estão sendo testadas. As hipóteses geradas por meio do Tipo IV podem não ser únicas.

Os modelos superparametrizados são de grande interesse, pois mostram explicitamente os parâmetros e, na presença de casela vazia, surge uma dificuldade maior na interpretação das hipóteses a serem testadas, havendo diferentes formas de obtenção das somas de quadrados, especificamente para os efeitos principais. Dentro de um modelo de médias de caselas (em geral, de posto completo), conhecido como Modelo-M, é uma alternativa ao excesso de parâmetros além de tornar as hipóteses de mais fácil interpretação, para um modelo com classificação dupla, principalmente na presença de interação.

Iemma (1993) fornece as hipóteses mais comuns sobre linhas, colunas e interação. Para tratar efeitos de linhas:

- hipóteses sobre médias ponderadas de linhas;
- hipóteses sobre médias ponderadas de linhas ajustadas para colunas;

- hipóteses sobre médias não ponderadas de linhas;
- hipóteses sobre médias não ponderadas de linhas, segundo as caselas vazias.

E de modo semelhante, para testar os efeitos de colunas.

Verificou que, com a ocorrência de caselas vazias, algumas médias populacionais correspondentes as caselas vazias não são estimáveis e podem conduzir a problemas com relação à estimação das funções paramétricas, refletindo sobre a interpretação das verdadeiras hipóteses testadas pelas somas de quadrados. Dentre os problemas, podem-se relacionar os seguintes:

- ocorrências de desconexões;
- em geral, as hipóteses sobre médias não ponderadas de linhas e colunas têm uma interpretação diferente das que possuem as caselas ocupadas;
- a hipótese sobre a interação não testa que todas as interações são nulas, como em H_9 ;
- a interpretação das hipóteses está relacionada à posição e ao número das caselas vazias;
- cautela na escolha dos graus de liberdade da interação.

Com dados desbalanceados e algumas caselas vazias, objetivando estruturação de hipóteses e somas de quadrados a elas associados, Santos (1994) apresenta uma comparação entre os resultados fornecidos por doze pacotes, ressaltando aqui apenas a saída do SAS:

- Modelos com interação: com todas as caselas ocupadas, as hipóteses de interesse são as de Tipo III e IV, objetivando testar efeitos puros de cada fator envolvido e a hipótese de interação para testar efeitos de interações. Com algumas caselas vazias, a hipótese de interesse é a hipótese sobre interação, pois como as demais contêm a presença de outros fatores, não há como testar um efeito livre dos outros, tornando difícil sua interpretação.

No Modelo-S ou Modelo-M, uma estratégia prática de análise seria verificar a significação da interação: significativa, estudar o comportamento

de um fator dentro de cada nível fixado do outro; não significativa, excluir a interação do modelo e executar uma nova análise de variância.

- Modelos sem interação: com todas as caselas ocupadas, as hipóteses de interesse são as de Tipo II = III = IV, objetivando testar efeitos puros de cada fator envolvido.

Mondardo (1994), por meio de exemplos com dados desbalanceados com caselas vazias, aplica a teoria disponível para obtenção das hipóteses testadas pelos quatro tipos de somas de quadrados fornecidas pelo SAS, a fim de traduzir esta teoria em uma linguagem mais simples e acessível, facilitando a construção, interpretação e a escolha mais adequada dentre essas hipóteses, possibilitando ao pesquisador realizar sua análise de forma correta e consciente.

Segundo a autora, em termos do modelo superparametrizado, todos os tipos de hipóteses envolvem os parâmetros referentes à interação, não se podendo testar os efeitos principais isoladamente; somente a hipótese sobre a interação é livre de parâmetros sem interesse. Como não se pode testar se todas as interações são nulas, só se pode abandonar o modelo com interação caso se baseie no conhecimento do pesquisador. A soma de quadrados Tipo I para efeitos principais não apresenta aparente interesse e mesmo em termos dos parâmetros do Modelo de médias de caselas, são difíceis de interpretar, porque podem envolver parâmetros do outro fator, ou mesmo em conjuntos completos não ortogonais de modelos sem interação, além de parâmetros da interação nos modelos com interação. A Tipo II, apesar de testar hipóteses de efeitos principais ajustados, sempre contém combinações lineares dos parâmetros relativos à interação e coeficientes pouco usuais. Na do Tipo III, aumentando-se o número de caselas vazias e o número de níveis do fator, também ela se torna complexa e foge ao interesse do pesquisador. A soma de quadrados Tipo IV testa hipóteses bem mais simples e fáceis de interpretar; entretanto, com a perda

de informação, só é considerada uma parte dos dados com os quais elas podem trabalhar, como se fossem dados balanceados.

Os tipos de hipóteses construídas pelo PROC GLM, para um modelo de dois fatores e interação, segundo Iemma (1995), incorporam quatro tipos de hipóteses para testar efeitos de linha (fator A), quatro tipos de hipóteses para testar efeitos de coluna (fator B) e uma para interação. Utilizando-se um modelo de média de caselas, a Tabela 1 exibe as hipóteses abaixo:

- hipóteses do Tipo I: a igualdade entre os efeitos de um fator é verificada pelas médias ponderadas dos efeitos desse fator. Não é feito qualquer ajuste e, quando as hipóteses são feitas no Modelo-S, ocorrem parâmetros relativos aos efeitos da interação e do outro fator;
- hipóteses do Tipo II: a igualdade entre os efeitos de um fator é verificada pelas médias ponderadas dos efeitos desse fator, ajustada de modo a eliminar de forma explícita os efeitos dos parâmetros do outro fator;
- hipóteses do Tipo III: sobre médias não ponderadas dos efeitos de um fator, ajustadas para os demais efeitos do outro fator. Não são levadas em consideração as frequências das caselas;
- hipóteses do Tipo IV: semelhante à anterior, dependendo da posição das caselas vazias;
- hipóteses sobre a interação: Searle (1987) relata que quando todas as caselas estão ocupadas, para 'a' níveis do fator A e 'b' níveis do fator B, têm-se $1/4ab(a-1)(b-1)$ interações, das quais, no máximo, somente $(a-1)(b-1)$ são linearmente independentes (L.I.). Com caselas vazias, menos do que $1/4ab(a-1)(b-1)$ interações podem ser estimadas e também podem ocorrer menos do que $(a-1)(b-1)$ que sejam L.I., modificando o número de graus de liberdade da interação.

Com todas as caselas ocupadas, a hipótese sobre interação testa se todas as interações são nulas; com caselas vazias, ocorrem interações que não fazem

parte da hipótese. Lemma (1995) alerta que o teste da interação (de H_0^2) não deve ser utilizado como padrão de escolha entre um modelo com ou sem interação, porque não corresponde à hipótese de que todas as interações são nulas.

Nekastschalow (1997) relata que, para qualquer das hipóteses anteriores, de modo geral, pode ser posta na forma $H_0: k'\theta = c$ vs. $H_1: k'\theta \neq c$, em que, $k'\theta$ seja uma função paramétrica estimável no modelo proposto, em geral, na análise de variância, $c=0$ para as hipóteses anteriores. A soma de quadrados associada a essas hipóteses fica:

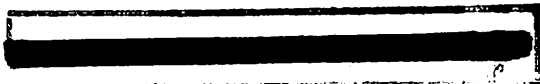
$$SQH_0 = (k'\theta^0 - c)' [k'(X'X)^G k]^{-1} (k'\theta^0 - c) \quad (14)$$

Considerando a hipótese de nulidade, obtém-se

$$SQH_0 = (k'\theta^0)' [k'(X'X)^G k]^{-1} (k'\theta^0) \quad (15)$$

Com o objetivo de estudar as funções estimáveis e as hipóteses testáveis, por meio do PROC GLM do SAS, tanto para os efeitos principais quanto para as interações, em delineamentos ortogonais e parcialmente ortogonais com três fatores, Santana & Lemma (2002) verificaram que, na presença de interações, as funções estimáveis apresentam, além de parâmetros do próprio fator, parâmetros das interações nas quais o fator está presente e, independente do modelo conter ou não conter interações, as hipóteses sobre médias ponderadas (Tipo I) são equivalentes às hipóteses sobre médias ponderadas ajustadas (Tipo II). Sendo conjunto completo, é condição obrigatória para a ortogonalidade parcial ou plena, ocorrendo também a equivalência entre as hipóteses sobre médias não ponderadas ajustadas (Tipos III e IV).

Discorrem ainda que, nos delineamentos ortogonais, a igualdade entre as hipóteses dos Tipos I e II ocorre para todas as interações; porém, nos delineamentos parcialmente ortogonais, ocorre apenas para as interações constituídas de fatores que não forem ortogonais entre si. As funções estimáveis



nessas interações apresentam parâmetros da própria interação e parâmetros da interação tripla.

2.8 Sistema estatístico: PROC GLM do SAS

Faz-se necessário o domínio sobre os conceitos de estimabilidade e testabilidade, para que ocorra uma interpretação da rotina do General Linear Models (GLM) do Statistical Analysis System (SAS, 1990). Com base nos modelos superparametrizados, o procedimento acima fornece quatro formas de execução das análises de variância, denominadas somas de quadrados (SQ), conforme mencionado anteriormente:

- SQ Tipo I: refere-se à soma de quadrados seqüencial; é obtida com a redução do modelo de um parâmetro por vez (efeito), na ordem inversa à de entrada do mesmo modelo;
- SQ Tipo II: refere-se à soma de quadrados de uma fonte de variação ajustada para as demais, que não contenham o efeito que se está testando;
- SQ Tipo III: refere-se ao ajuste de cada fator de variação para todos os efeitos remanescentes do modelo, utilizando o modelo com restrição do tipo $\Sigma = 0$;
- SQ Tipo IV: tem seu valor alterado em função das caselas vazias.

Segundo Santana & Iemma (2002), o PROC GLM apenas fornece hipóteses do Tipo I e conseqüente soma de quadrados associada, se o fator ou a interação de interesse forem declarados em primeiro lugar no modelo. Informações mais detalhadas no estudo das funções estimáveis e das hipóteses testáveis em delineamentos ortogonais e parcialmente ortogonais com mais de dois fatores, por meio do PROC GLM do SAS, são apresentados pelos autores.

As funções estimáveis são apresentadas pelo comando MODEL quando as opções:

- 'E', que provê a forma geral das funções estimáveis: E1, E2, E3 e E4 (funções estimáveis associadas a cada tipo de soma de quadrados) e
- 'SS', que provê as somas de quadrados de cada tipo respectivamente: SS1, SS2, SS3, SS4,

são requeridas após a especificação do modelo.

As opções XPX e I fornecem a matriz $X'X$ e a sua inversa generalizada $G2$, que é utilizada para obtenção das somas de quadrados Tipo I, II e IV.

Maiores informações podem ser obtidas no Manual do SAS (1990).

3 MATERIAL E MÉTODOS

3.1 1º caso: fatorial 4 x 4 desbalanceado

Os dados utilizados são provenientes de experimento realizado no Setor de Olericultura do Departamento de Agricultura da Universidade Federal de Lavras (UFLA), em Lavras, Minas Gerais e referem-se ao peso comercial de raízes de cenoura.

Os dados relativos à produção comercial das raízes, em t/ha, estão apresentados na Tabela 3.

TABELA 3. Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com as diferentes fases da lua (1º caso).

Fator A (cult)	Fator B (fases da lua)											
	Nova			Crescente			Cheia			Minguante		
Nan	48,40	10,70	25,70	30,40	20,40	47,15	38,45	41,05	30,95
	13,40	14,45	9,45	31,85	21,45	36,80
	17,90	23,85	22,90	14,80	20,50	3,80	22,90	34,55	22,40	5,40	18,10	17,05
	36,45	38,55	33,30	30,45	36,45	33,20	38,80	26,95	37,50	29,20	28,65	25,45
	37,55	41,45	18,15	33,85	10,90	15,30	37,75	28,85	30,40	27,15	19,80	15,20
Kur	68,25	70,90	66,30	68,85	72,00	68,90	64,15	56,15	64,75
	23,05	34,10	26,65	47,60	45,95	40,25
	30,70	36,85	40,75	33,40	29,65	28,75	39,05	39,20	39,50	30,65	16,35	37,45
	33,85	39,05	42,00	38,35	33,40	42,40	47,55	53,05	54,70	36,25	34,35	28,45
	50,00	44,85	52,45	51,80	33,55	21,40	42,90	37,55	35,65	31,70	38,95	23,65
Car	65,00	49,60	42,90	52,35	65,50	67,40	51,55	46,05	52,95
	19,00	25,05	20,20	35,40	38,60	40,10
	35,50	40,40	30,35	26,05	36,75	17,20	41,45	37,10	31,60	26,20	19,85	23,65
	44,25	33,85	38,90	38,25	41,50	39,90	55,25	54,65	53,80	24,30	37,95	30,30
	49,70	45,90	31,00	48,75	17,95	22,95	35,60	29,95	38,55	39,85	28,05	31,80
Bra	78,25	61,25	71,35	75,10	49,65	76,60	71,00	78,95	79,40
	28,95	19,30	41,90	44,25	38,90	33,70
	36,45	42,15	48,80	34,10	33,05	29,55	37,95	44,65	52,80	25,00	25,75	30,10
	57,65	52,85	54,80	41,25	41,15	44,80	56,95	63,45	55,80	45,35	58,60	55,45
	56,00	54,85	54,15	55,05	30,70	29,15	55,25	38,75	44,25	46,55	50,85	50,65

Fonte: Cristiane R. B. Aguirre Ramos. Departamento de Agricultura. UFLA, Lavras, MG.

Nota: ... : não se dispõe do dado.

Para a realização das análises considerou-se o delineamento experimental inteiramente casualizado, com quinze repetições. Os tratamentos foram dispostos em esquema fatorial 4 x 4, constituído por quatro cultivares, sendo três nacionais (Kuronan, Carandaí e Brasília) e uma importada (Nantes), e quatro épocas de plantio, caracterizadas pelas diferentes fases da lua.

A parcela experimental era de 2m de comprimento por 1m de largura. Os tratos culturais foram realizados de acordo com a necessidade da cultura.

O objetivo deste experimento foi testar se as diferentes cultivares e as diversas fases da lua influenciam a produção comercial de raízes de cenoura.

3.1.1 Modelo superparametrizado: Modelo-S

Para o modelo linear superparametrizado, o modelo dado em (1) assume a caracterização

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

em que:

y_{ijk} : observação referente a i -ésima cultivar na j -ésima lua e na k -ésima repetição;

μ : constante inerente à todas as observações;

α_i : efeito da i -ésima cultivar (fator A);

β_j : efeito da j -ésima lua (fator B);

γ_{ij} : efeito da interação entre o i -ésimo nível do fator A e o j -ésimo do fator B;

e_{ijk} : erro experimental associado à observação y_{ijk} considerado independente e normalmente distribuído com média zero e variância constante;

$i = 1, 2, 3$ e 4 níveis do fator A(α) que correspondem respectivamente às cultivares Nantes, Kuronan, Carandaí e Brasília;

$j = 1, 2, 3$ e 4 níveis do fator B (β) que correspondem, respectivamente, às fases da lua nova, crescente, cheia e minguante;
 $k = 0, \dots, n_{ij}$ é o número de observações por casela.

Com a finalidade de verificar se a ordem de entrada dos parâmetros no modelo afetava ou não as somas de quadrados e suas hipóteses, foram tomadas duas ordens para os efeitos principais:

- para a ordenação A-B:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk};$$

- para a ordenação B-A:

$$y_{jik} = \mu + \beta_j + \alpha_i + \gamma_{ji} + e_{jik}.$$

Quando os dados são desbalanceados, a ordem dos fatores no modelo é fundamental na formulação das hipóteses do Tipo I, que é seqüencial: o primeiro fator é ajustado em relação à constante, o segundo fator é ajustado tanto em relação à constante como ao outro fator.

Para o experimento apresentado na Tabela 3, para a ordem A-B, o modelo linear está representação matricial ($y = X\theta + e$) a seguir.

Para a ordem B-A, as colunas 2, 3, 4 e 5, representantes dos níveis do fator A, trocam de posição com as colunas 6, 7, 8 e 9, representantes do fator B, influenciando nas colunas de interação.

48,40	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	μ	
10,70	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_1	
25,70	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_2	
17,90	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_3	
23,85	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_4	
22,90	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_1	
36,45	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_2	
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	β_3		
31,70	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	β_4
38,95	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	γ_{11}
23,65	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	γ_{13}
65,00 =	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	$\gamma_{14} + e$
49,60	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	γ_{21}
42,90	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	γ_{22}
35,50	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	γ_{23}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	γ_{24}	
30,10	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{31}
45,35	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{32}
58,60	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{33}
55,45	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{34}
46,55	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{41}
50,85	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{42}
50,65	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	γ_{43}

3.2 2º caso: fatorial 4 x 4 desbalanceado com caselas vazias

A Tabela 4 apresenta alguns dados relativos ao experimento do primeiro caso, envolvendo a produção comercial de raízes de cenoura, em t/ha.

O delineamento experimental foi o inteiramente casualizado, com três repetições, com os tratamentos em esquema fatorial 4 x 4, constituído pelas combinações dos fatores cultivar e épocas de plantio caracterizadas pelas diferentes fases da lua.

Para efeito de estudo de dados provenientes de experimento desbalanceado com caselas vazias, foram suprimidas algumas observações aleatoriamente por sorteio.

TABELA 4. Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com as diferentes fases da lua (2º caso).

Fator A (cultivares)	Fator B (fases da lua)			
	nova	crescente	cheia	Minguante
Nantes	37,55	x	37,75	27,15
	41,45	x	28,85	19,80
	18,15	x	30,40	15,20
Kuronan	50,00	51,80	42,90	31,70
	44,85	33,55	x	38,95
	52,45	21,40	35,65	23,65
Carandai	49,70	48,75	35,60	39,85
	45,90	17,95	29,95	28,05
	31,00	22,95	38,55	31,80
Brasília	56,00	55,05	55,25	x
	54,85	30,70	38,75	x
	54,15	29,15	44,25	x

Fonte: Cristiane R. B. Aguirre Ramos. Departamento de Agricultura. UFLA, Lavras, MG.

Nota: x: dado omitido propositalmente para este estudo.

3.2.1 Modelo superparametrizado: Modelo-S

O Modelo-S assume a mesma caracterização do primeiro caso, inclusive para as duas ordens dos efeitos dos fatores principais do modelo dado em (1).

Para o experimento apresentado na Tabela 4, para a ordem A-B, o modelo linear está representado matricialmente na próxima página ($y = X\theta + e$).

Para a ordem B-A, as colunas 2, 3, 4 e 5, representantes dos níveis do fator A, trocam respectivamente de posição com as colunas 6, 7, 8 e 9, representantes do fator B, influenciando nas colunas de interação.

3.3 3º caso: fatorial 4 x 5 x 4 desbalanceado

Na Tabela 5 apresentam-se os dados de produção de raízes comerciais (t/ha) relativos ao experimento do primeiro caso, agora constituído por um fatorial 4 x 5 x 4, envolvendo as quatro cultivares (Kuronan, Carandai, Brasília e Nantes), cinco meses do ano (março, abril, maio, junho e julho) e as quatro fases da lua (cheia, crescente, minguante e nova). Utilizou-se o delineamento inteiramente casualizado, com três repetições.

3.3.1 Modelo superparametrizado: Modelo-S

Para o modelo linear superparametrizado, o modelo dado em (1) assume a caracterização

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varphi_{ij} + \eta_{ik} + \tau_{jk} + \rho_{ijk} + e_{ijkl}$$

em que:

y_{ijkl} : valor observado na l -ésima repetição da k -ésima fase da lua, no j -ésimo mês de plantio e na i -ésima cultivar, com $l = 1, 2, 3$ repetições;

μ : constante inerente a todas as observações;

α_i : efeito da i -ésima cultivar (fator A), com $i = 1, \dots, 4$ níveis do fator A (α);

β_j : efeito do j -ésimo mês (fator B), com $j = 1, \dots, 5$ níveis do fator B (β);

γ_k : efeito da k -ésima fase da lua (fator C), com $k = 1, \dots, 4$ níveis do fator C (γ);

φ_{ij} : efeito da interação entre o i -ésimo nível do fator A e o j -ésimo do fator B;

η_{ik} : efeito da interação entre o i -ésimo nível do fator A e o k -ésimo do fator C;

τ_{jk} : efeito da interação entre o j -ésimo nível do fator B e o k -ésimo do fator C;

ρ_{ijk} : efeito da interação entre i -ésimo nível do fator A, o j -ésimo nível do fator B e o k -ésimo do fator C;

e_{ijkl} : erro experimental associado à observação y_{ijkl} e considerado independente e normalmente distribuído com média zero e variância constante.

TABELA 5 – Produção comercial de raízes de cultivares de cenoura, em t/ha, de acordo com os meses e fases da lua na época de plantio (3º caso).

F a t o r A	Fator B (mês)																			
	Março			Abril			Maio			Junho			Julho							
	Fator C (fases da lua)			Fator C (fases da lua)			Fator C (fases da lua)			Fator C (fases da lua)			Fator C (fases da lua)							
	No	Cr	Ch	M	N	Cr	Ch	M	N	Cr	Ch	M	N	Cr	Ch	M	N	Cr	Ch	M
N a n	48,40	30,40	38,45	.	.	.	13,40	31,85	17,90	14,80	22,90	5,40	36,45	30,45	38,80	29,20	37,55	33,85	37,75	27,15
	10,70	20,40	41,05	.	.	.	14,45	21,45	23,85	20,50	34,55	18,10	38,55	36,45	26,95	28,65	41,45	10,90	28,85	19,80
	25,70	47,15	30,95	.	.	.	9,45	36,80	22,90	3,80	22,40	17,05	33,30	33,20	37,50	25,45	18,15	15,30	30,40	15,20
K u r	68,25	68,85	64,15	.	.	.	23,05	47,60	30,70	33,40	39,05	30,65	33,85	38,35	47,55	36,25	50,00	51,80	42,90	31,70
	70,90	72,00	56,15	.	.	.	34,10	45,95	36,85	29,65	39,20	16,35	39,05	33,40	53,05	34,35	44,85	33,55	37,55	38,95
	66,30	68,90	64,75	.	.	.	26,65	40,25	40,75	28,75	39,50	37,45	42,00	42,40	54,70	28,45	52,45	21,40	35,65	23,65
C a r	65,00	52,35	51,55	.	.	.	19,00	35,40	35,50	26,05	41,45	26,20	44,25	38,25	55,25	24,30	49,70	48,75	35,60	39,85
	49,60	65,50	46,05	.	.	.	25,05	38,60	40,40	26,75	37,10	19,85	33,85	41,50	54,65	37,95	45,90	17,95	29,95	28,05
	42,90	67,40	52,95	.	.	.	20,20	40,10	30,35	17,20	31,60	23,65	38,90	39,90	53,80	30,30	31,00	22,95	38,55	31,80
B r a	78,25	75,10	71,00	.	.	.	28,95	44,25	36,45	34,10	37,95	25,00	57,65	41,25	56,95	45,35	56,00	55,05	55,25	46,55
	61,25	49,65	78,95	.	.	.	19,30	38,90	42,15	33,05	44,65	25,75	52,85	41,15	63,45	58,60	54,85	30,70	38,75	50,85
	71,35	76,60	79,40	.	.	.	41,90	33,70	48,80	29,55	52,80	30,10	54,80	44,80	55,80	55,45	54,15	29,15	44,25	50,65

Fonte: Cristiane R. B. Aguirre Ramos. Departamento de Agricultura. UFLA, Lavras, MG.

Nota: . : não se dispõe do dado.

Para verificar se a ordem de entrada dos fatores afetava os resultados das somas de quadrados e de suas respectivas hipóteses associadas, foram consideradas as seguintes ordens de entrada dos fatores principais:

- para a ordenação A-B-C:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varphi_{ij} + \eta_{ik} + \tau_{jk} + \rho_{ijk} + e_{ijkl};$$

- para a ordenação B-C-A:

$$y_{jkil} = \mu + \beta_j + \gamma_k + \alpha_i + \tau_{jk} + \varphi_{ji} + \eta_{ki} + \rho_{jki} + e_{jkil};$$

- para a ordenação C-B-A:

$$y_{kjil} = \mu + \gamma_k + \beta_j + \alpha_i + \tau_{kj} + \eta_{ik} + \varphi_{ji} + \rho_{kji} + e_{kjil}.$$

A representação matricial para o modelo linear superparametrizado ($y=X\theta + e$), para a ordem A-B-C, para os dados apresentados na Tabela 5, se encontra na próxima página.

3.4 Procedimento GLM do sistema estatístico SAS

É necessário um programa informando os dados, o tipo do modelo adotado e também a solicitação das funções estimáveis e conseqüentes somas de quadrados.

As macros para realizar a análise dos dados da Tabela 3 (1º caso), da Tabela 4 (2º caso) e da Tabela 5 (3º caso), estão, respectivamente, nos Anexos A, B e C.

48,40	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	μ		
10,70	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_1		
25,70	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_2		
30,40	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_3		
20,40	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	α_4		
47,15	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_1		
38,45	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_2		
...	β_3		
64,75	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_4		
23,05	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	β_5		
34,10	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	γ_1		
26,65	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	γ_2		
47,60	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	γ_3		
45,95	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	γ_4		
40,25	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	φ_1		
...	φ_2		
29,15	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	φ_3		
55,25	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	φ_4		
38,75	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	φ_5		
44,25	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	\vdots		
46,55	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ρ_{32}		
50,85	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ρ_{33}		
50,65	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ρ_{351}		
+e																											

4 RESULTADOS E DISCUSSÃO

4.1 1º caso: fatorial 4 x 4 desbalanceado

A forma geral das funções estimáveis, fornecida pelo PROC GLM do SAS, obtidas através de $L\theta$, independente da ordenação dos dados está apresentada na Tabela 6.

Têm-se 16 símbolos (L1, L2, L3, L4, L6, L7, L8, L10, L11, L12, L14, L15, L16, L18, L19 e L20) envolvidos, indicando o posto da matriz L, sendo o posto máximo da hipótese que pode ser avaliada; neste caso, tem 16 graus de liberdade, ou seja, podem ser avaliadas até 16 funções lineares paramétricas linearmente independentes. Os símbolos faltantes, L5, L9 e L13 correspondem a linhas formadas apenas por zeros, da matriz $L = (X'X)^{G2}(X'X)$.

As Tabelas 7, 8 e 9 mostram as funções estimáveis fornecidas pelo SAS. Na Tabela 7 estão as do Tipo I e II para a ordem A-B; na Tabela 8 as do Tipo I e II para a ordem B-A e na Tabela 9 as funções estimáveis do Tipo III e IV para ambas as ordens, pois se apresentam equivalentes.

TABELA 6. Forma geral de funções estimáveis dos dados da Tabela 3, fornecida pelo PROC GLM do SAS.

Parâmetros	Coefficientes
μ	L1
α_1	L2
α_2	L3
α_3	L4
α_4	L1-L2-L3-L4
β_1	L6
β_2	L7
β_3	L8
β_4	L1-L6-L7-L8
γ_{11}	L10
γ_{12}	L11
γ_{13}	L12
γ_{14}	L2-L10-L11-L12
γ_{21}	L14
γ_{22}	L15
γ_{23}	L16
γ_{24}	L3-L14-L15-L16
γ_{31}	L18
γ_{32}	L19
γ_{33}	L20
γ_{34}	L4-L18-L19-L20
γ_{41}	L6-L10-L14-L18
γ_{42}	L7-L11-L15-L19
γ_{43}	L8-L12-L16-L20
γ_{44}	L1-L2-L3-L4-L6-L7-L8-L10-L11-L12-L14-L15-L16-L18-L19-L20

TABELA 7. Forma geral das funções estimáveis Tipo I e II do SAS para os dados da Tabela 3, na ordem A-B, para os fatores A, B e interação AB.

		A	B	AB
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
AB	1 1	0.2353*L2	0.25*L6	L10
	1 2	0.2353*L2	0.25*L7	L11
	1 3	0.2941*L2	0.25*L8	L12
	1 4	0.2353*L2	-0.25*L6-0.25*L7-0.25*L8	-L10-L11-L12
	2 1	0.2353*L3	0.25*L6	L14
	2 2	0.2353*L3	0.25*L7	L15
	2 3	0.2941*L3	0.25*L8	L16
	2 4	0.2353*L3	-0.25*L6-0.25*L7-0.25*L8	-L14-L15-L16
	3 1	0.2353*L4	0.25*L6	L18
	3 2	0.2353*L4	0.25*L7	L19
	3 3	0.2941*L4	0.25*L8	L20
	3 4	0.2353*L4	-0.25*L6-0.25*L7-0.25*L8	-L18-L19-L20
	4 1	-0.2353*L2-0.2353*L3-0.2353*L4	0.25*L6	-L10-L14-L18
	4 2	-0.2353*L2-0.2353*L3-0.2353*L4	0.25*L7	-L11-L15-L19
	4 3	-0.2941*L2-0.2941*L3-0.2941*L4	0.25*L8	-L12-L16-L20
	4 4	-0.2353*L2-0.2353*L3-0.2353*L4	-0.25*L6-0.25*L7-0.25*L8	L10+L11+L12+L14+L15+L16 +L18+L19+L20

Nota: As funções estimáveis para os Tipos I e II são equivalentes para a ordem A-B.

TABELA 8. Forma geral das funções estimáveis Tipo I e II do SAS para os dados da Tabela 3, na ordem B-A, para os fatores B, A e interação BA.

		B	A	BA
μ		0	0	0
B	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
A	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
BA	1 1	0.25*L2	0.2353*L6	L10
	1 2	0.25*L2	0.2353*L6	L11
	1 3	0.25*L2	0.2941*L6	L12
	1 4	0.25*L2	0.2353*L6	-L10-L11-L12
	2 1	0.25*L3	0.2353*L7	L14
	2 2	0.25*L3	0.2353*L7	L15
	2 3	0.25*L3	0.2941*L7	L16
	2 4	0.25*L3	0.2353*L7	-L14-L15-L16
	3 1	0.25*L4	0.2353*L8	L18
	3 2	0.25*L4	0.2353*L8	L19
	3 3	0.25*L4	0.2941*L8	L20
	3 4	0.25*L4	0.2353*L8	-L18-L19-L20
	4 1	-0.25*L2-0.25*L3-0.25*L4	-0.2353*L6-0.2353*L7-0.2353*L8	-L10-L14-L18
	4 2	-0.25*L2-0.25*L3-0.25*L4	-0.2353*L6-0.2353*L7-0.2353*L8	-L11-L15-L19
	4 3	-0.25*L2-0.25*L3-0.25*L4	-0.2941*L6-0.2941*L7-0.2941*L8	-L12-L16-L20
	4 4	-0.25*L2-0.25*L3-0.25*L4	-0.2353*L6-0.2353*L7-0.2353*L8	L10+L11+L12+L14+L15+L16 +L18+L19+L20

Nota: As funções estimáveis para os Tipos I e II são equivalentes para a ordem B-A.

TABELA 9. Forma geral das funções estimáveis Tipo III e IV do SAS para os dados da Tabela 3, na ordem A-B e na ordem B-A, para os fatores A, B e interação AB.

		A	B	AB
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
AB	1 1	0.25*L2	0.25*L6	L10
	1 2	0.25*L2	0.25*L7	L11
	1 3	0.25*L2	0.25*L8	L12
	1 4	0.25*L2	-0.25*L6-0.25*L7-0.25*L8	-L10-L11-L12
	2 1	0.25*L3	0.25*L6	L14
	2 2	0.25*L3	0.25*L7	L15
	2 3	0.25*L3	0.25*L8	L16
	2 4	0.25*L3	-0.25*L6-0.25*L7-0.25*L8	-L14-L15-L16
	3 1	0.25*L4	0.25*L6	L18
	3 2	0.25*L4	0.25*L7	L19
	3 3	0.25*L4	0.25*L8	L20
	3 4	0.25*L4	-0.25*L6-0.25*L7-0.25*L8	-L18-L19-L20
	4 1	-0.25*L2-0.25*L3-0.25*L4	0.25*L6	-L10-L14-L18
	4 2	-0.25*L2-0.25*L3-0.25*L4	0.25*L7	-L11-L15-L19
	4 3	-0.25*L2-0.25*L3-0.25*L4	0.25*L8	-L12-L16-L20
	4 4	-0.25*L2-0.25*L3-0.25*L4	-0.25*L6-0.25*L7-0.25*L8	L10+L11+L12+L14+L15+L16 +L18+L19+L20

Nota: As funções estimáveis para os Tipos III e IV são equivalentes para a ordem A-B, como também para os Tipos III e IV da ordem B-A

$(X'X)^{G2}$																				$X'y = \theta^0$		
5	-5	-5	-5	0	5	5	5	0	5	5	5	0	5	5	5	0	0	0	0	0	7957,95	42,096
-5	10	5	5	0	-10	-10	0	-5	-5	-5	0	-5	-5	-5	0	0	0	0	0	0	1356,05	-19,088
-5	5	10	5	0	-5	-5	0	-10	-10	0	-5	-5	-5	-5	0	0	0	0	0	0	2168,00	-7,796
-5	5	5	10	0	-5	-5	-5	0	-5	-5	0	-10	-10	-10	0	0	0	0	0	0	1950,70	-10,758
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2483,20	0
-5	5	5	5	0	-10	-5	-5	0	-10	-5	-5	0	-10	-5	-5	0	0	0	0	0	2106,75	13,617
-5	5	5	5	0	-5	-10	-5	0	-5	-10	-5	0	-5	-10	-5	0	0	0	0	0	1824,35	2,917
-5	5	5	5	0	-5	-5	-9	0	-5	-5	-9	0	-5	-5	-9	0	0	0	0	0	2447,95	9,194
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1568,90	0
5	-10	-5	-5	0	20	10	0	10	5	5	0	10	5	5	0	0	0	0	0	0	354,90	-7,050
5	-10	-5	-5	0	10	20	10	0	5	10	5	0	5	10	5	0	0	0	0	0	297,20	-1,158
5	-10	-5	-5	0	-5	5	-9	0	5	5	9	0	5	5	9	0	0	0	0	0	427,85	-3,679
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	276,10	0
5	-5	-10	-5	0	10	5	5	0	20	10	10	0	10	5	5	0	0	0	0	0	575,95	0,079
5	-5	-10	-5	0	-5	-10	-5	0	5	10	5	0	10	5	10	5	0	0	0	0	522,45	6,321
5	-5	-10	-5	0	-5	-9	0	-5	5	9	0	10	18	0	5	5	0	0	0	0	658,00	0,372
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	411,60	0
5	-5	-5	-10	0	-10	-5	-5	0	10	5	0	10	5	0	20	10	0	0	0	0	507,35	-2,675
5	-5	-5	-10	0	-5	-10	-5	0	5	10	5	0	10	5	10	20	10	0	0	0	474,55	5,292
5	-5	-5	-10	0	-5	-9	0	-5	5	9	0	10	18	0	10	18	0	0	0	0	592,75	-1,015
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	376,05	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	668,55	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	540,15	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	769,35	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	505,15	0

$\frac{1}{60}$

e a aproximação de quadrados mínimos, \hat{y} , para o vetor y das observações:

$$\hat{y} = X\theta^0 = \begin{bmatrix} 29,575 \\ 29,575 \\ 29,575 \\ 29,575 \\ 29,575 \\ 29,575 \\ 29,575 \\ \vdots \\ 34,300 \\ 34,300 \\ 34,300 \\ 42,279 \\ 42,279 \\ 42,279 \\ \vdots \\ 42,096 \\ 42,096 \\ 42,096 \\ 42,096 \\ 42,096 \\ 42,096 \\ 42,096 \end{bmatrix}$$

Verifica-se que o vetor da aproximação de quadrados mínimos é constituído pelas médias das caselas.

4.1.2 Estimação

Para este conjunto de dados e caracterização do modelo linear adotado, as Tabelas 10 e 11 mostram um conjunto de funções estimáveis sobre os efeitos principais e a Tabela 12, mostra as funções estimáveis sobre a interação. Elas foram obtidas utilizando-se as funções estimáveis fornecidas pelo SAS (Tabelas 7 e 9), para o modelo dado, na ordem A-B.

Atribuindo-se valores arbitrários aos coeficientes fornecidos nas funções estimáveis, obtêm-se as funções estimáveis que irão constituir conjuntos de hipóteses que poderão ser testadas. Neste estudo, atribuiu-se valor um para cada um dos coeficientes, zerando os demais, de acordo com o interesse na obtenção de uma dada função estimável.

Para as funções estimáveis do Tipo I para o fator A, a primeira linha é escrita, atribuindo-se valor um para o coeficiente L2 e zero para os coeficientes L3 e L4. Em seguida, atribui-se valor um para o coeficiente L3 e zero para os coeficientes L2 e L4, obtendo-se a segunda linha. Finalmente, para a terceira linha atribui-se valor um para o coeficiente L4 e zero para os coeficientes L2 e L3. Maiores detalhes podem ser encontrados em Mondardo (1994).

A análise dos dados mostra que as funções estimáveis (Tabelas 10 e 11) de um fator envolvem apenas parâmetros relativos ao próprio fator e os componentes das interações nas quais o referido fator está presente.

As funções estimáveis do Tipo I são construídas de modo dependente da ordem dos parâmetros do modelo; cada efeito é ajustado somente para os efeitos precedentes no modelo. A construção das funções estimáveis Tipo II é de forma independente para cada fator.

Para o fator A, as funções estimáveis do Tipo I não envolvem o parâmetro do outro fator, apenas a interação e são equivalentes às do Tipo II. Observa-se que, nas três funções $\alpha_1 - \alpha_4$, $\alpha_2 - \alpha_4$ e $\alpha_3 - \alpha_4$, os coeficientes das interações formadas pelos níveis do fator A para um mesmo nível do fator B são iguais, por exemplo, $4/17\gamma_{11} = 4/17\gamma_{21} = 4/17\gamma_{31}$. As funções estimáveis do Tipo III são equivalentes às do Tipo IV e envolvem também apenas a interação, sendo que todos os coeficientes das interações possuem o mesmo valor, ou seja, 1/4.

Quanto ao fator B, os quatro tipos de funções estimáveis são equivalentes e não envolvem o outro parâmetro, apenas a interação, cujos coeficientes são iguais a 1/4.

Provavelmente, isto se deve ao fato de que nas caselas que apresentam desbalanceamento ocorreram perdas iguais, ou seja, apresentam doze repetições, enquanto que nas caselas completas ocorrem em número de quinze repetições. As perdas para os níveis do fator A ocorreram em números equivalentes; já para o fator B, apenas no nível três (lua cheia) não ocorreram perdas.

No modelo superparametrizado, as funções estimáveis auxiliam na interpretação das hipóteses apresentadas e, conseqüentemente, das somas de quadrados associadas às hipóteses. Não são de fácil compreensão.

TABELA 10. Funções estimáveis sobre o fator A (cultivares), dos Tipos I, II, III e IV.

Funções estimáveis	
T	L2=1
I	L3=0
P	L4=0
.....	
O	L2=0
	L3=1
I	L4=0
.....	
e	L2=0
II	L3=0
(1)	L4=1
.....	
T	L2=1
I	L3=0
P	L4=0
.....	
O	L2=0
	L3=1
III	L4=0
.....	
e	L2=0
IV	L3=0
(2)	L4=1

⁽¹⁾ As funções estimáveis para os Tipos I e II são equivalentes.

⁽²⁾ As funções estimáveis para os Tipos III e IV são equivalentes.

TABELA 11. Funções estimáveis sobre o fator B (fases da lua), dos Tipos I, II, III e IV.

Funções estimáveis	
T	L6=1
I	L7=0
P	L8=0
O	L6=0
I, II	L7=1
III e IV	L8=0
	L6=0
	L7=0
	L8=1

$$\beta_1 - \beta_4 + 1/4 (\gamma_{11} - \gamma_{14} + \gamma_{21} - \gamma_{24} + \gamma_{31} - \gamma_{34} + \gamma_{41} - \gamma_{44})$$

$$\beta_2 - \beta_4 + 1/4 (\gamma_{12} - \gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} + \gamma_{42} - \gamma_{44})$$

$$\beta_3 - \beta_4 + 1/4 (\gamma_{13} - \gamma_{14} + \gamma_{23} - \gamma_{24} + \gamma_{33} - \gamma_{34} + \gamma_{43} - \gamma_{44})$$

As funções estimáveis para os tipos I, II, III e IV são equivalentes.

TABELA 12. Funções estimáveis sobre a interação.

Funções estimáveis	
L10=1; L11=0; L12=0; L14=0; L15=0; L16=0; L18=0; L19=0; L20=0	$\gamma_{11} - \gamma_{14} - \gamma_{41} + \gamma_{44}$
L10=0; L11=1; L12=0; L14=0; L15=0; L16=0; L18=0; L19=0; L20=0	$\gamma_{12} - \gamma_{14} - \gamma_{42} + \gamma_{44}$
L10=0; L11=0; L12=1; L14=0; L15=0; L16=0; L18=0; L19=0; L20=0	$\gamma_{13} - \gamma_{14} - \gamma_{43} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=1; L15=0; L16=0; L18=0; L19=0; L20=0	$\gamma_{21} - \gamma_{24} - \gamma_{41} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=0; L15=1; L16=0; L18=0; L19=0; L20=0	$\gamma_{22} - \gamma_{24} - \gamma_{42} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=0; L15=0; L16=1; L18=0; L19=0; L20=0	$\gamma_{23} - \gamma_{24} - \gamma_{43} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=0; L15=0; L16=0; L18=1; L19=0; L20=0	$\gamma_{31} - \gamma_{34} - \gamma_{41} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=0; L15=0; L16=0; L18=0; L19=1; L20=0	$\gamma_{32} - \gamma_{34} - \gamma_{42} + \gamma_{44}$
L10=0; L11=0; L12=0; L14=0; L15=0; L16=0; L18=0; L19=0; L20=1	$\gamma_{33} - \gamma_{34} - \gamma_{43} + \gamma_{44}$

As funções estimáveis da interação apresentaram somente parâmetros da interação.

4.1.3 Hipóteses e somas de quadrados associadas

Por meio das funções estimáveis, pode-se reconhecer a hipótese que está sendo testada.

Para o modelo superparametrizado, de classificação dupla cruzada, com dados desbalanceados, as hipóteses testadas para os quatro tipos de somas de quadrados envolvem os parâmetros referentes à interação, sendo apenas a hipótese sobre a interação livre de parâmetros sem interesse.

O Modelo-S é utilizado inicialmente pelo PROC GLM do SAS.

Percebe-se, na Tabela 13, para os quatro tipos de hipóteses no Modelo-S, sobre o fator A e na Tabela 14 sobre o fator B, que todas as hipóteses envolvem os parâmetros referentes à interação, não podendo assim testar os efeitos principais isoladamente.

Para o fator A, as hipóteses do Tipo I, apesar de não sofrerem ajuste, são equivalentes às do Tipo II (Tabela 13). Apresentam coeficientes confusos e um contraste entre os níveis do fator principal, seguidos dos efeitos da interação e, aparentemente, não têm interesse prático.

As hipóteses do Tipo III e IV (Tabela 14) são equivalentes, fato concordante com o comentário de Herr (1986), pois tratam-se de caselas de tamanho proporcionais. São de aparente interesse, por apresentarem contraste entre os níveis do fator A e coeficientes mais simples para a interação.

Neste caso, sendo os dados desbalanceados e com todas as caselas ocupadas, as hipóteses de interesse são: do Tipo III e IV quando se objetiva testar efeitos puros de cada fator envolvido, e as hipóteses de interação para testar efeitos de interações, confirmando a citação de Santos (1994).

A igualdade das somas de quadrados associadas às hipóteses Tipo I = Tipo II para os fatores A e B; às hipóteses do Tipo III = Tipo IV; conseqüentemente, $R(\alpha|\mu) = R(\alpha|\mu,\beta)$ e $R(\beta|\mu) = R(\beta|\mu,\alpha)$. A diferença da

hipótese do Tipo II da hipótese do Tipo III, na presença de interação com desbalanceamento, ou seja, $R(\alpha|\mu,\beta) \neq R(\hat{\alpha}|\hat{\mu},\hat{\beta},\hat{\gamma})$ e $R(\beta|\mu,\alpha) \neq R(\hat{\beta}|\hat{\mu},\hat{\alpha},\hat{\gamma})$, também foi obtida por Santana (2000).

Para o fator B, os quatro tipos de hipóteses são equivalentes, apesar do desbalanceamento (número diferente de repetição). Elas apresentam contraste entre os níveis do fator e coeficientes simples para a interação.

Para este conjunto de dados, sem caselas vazias, a hipótese sobre a interação tem $(a-1)(b-1) = (4-1)(4-1) = 9$ graus de liberdade e, conseqüentemente, nove funções linearmente independentes podem ser testadas. Nota-se que as hipóteses sobre a interação (Tabela 15) são livres de parâmetros sem interesse, ou seja, de parâmetros dos fatores principais.

A alteração da ordem de especificação do modelo dos efeitos principais, na ordem B, A, B-A, não alterou a determinação das hipóteses do Tipo I e das somas de quadrados associadas, como geralmente ocorre.

Em se tratando de um modelo com interação com dados desbalanceados, porém sem caselas vazias, com a apresentação da análise de variância é possível apresentar, efetivamente, independente da ordenação dos efeitos principais, duas somas de quadrados: a primeira é a soma de quadrados associada às hipóteses dos Tipos I e II, que são iguais (Tabela 16) e a segunda é a soma de quadrados associada às hipóteses dos Tipos III e IV, que são também iguais (Tabela 17). Portanto, confirma a afirmativa feita por Speed et al. (1978), Searle (1987) e Lemma (1995) de que, no caso de dados desbalanceados, sem caselas vazias, as hipóteses dos Tipos III e IV são iguais.

TABELA 13. Hipóteses testadas sobre o fator A (cultivares).

Hipóteses	Hipóteses testadas
Tipo I	$\alpha_1 - \alpha_4 + 1/17 (4\gamma_{11} + 4\gamma_{12} + 5\gamma_{13} + 4\gamma_{14} - 4\gamma_{41} - 4\gamma_{42} - 5\gamma_{43} - 4\gamma_{44}) = 0$
e II ⁽¹⁾	$\alpha_2 - \alpha_4 + 1/17 (4\gamma_{21} + 4\gamma_{22} + 5\gamma_{23} + 4\gamma_{24} - 4\gamma_{41} - 4\gamma_{42} - 5\gamma_{43} - 4\gamma_{44}) = 0$
H_0^1 e H_0^2	$\alpha_3 - \alpha_4 + 1/7 (4\gamma_{31} + 4\gamma_{32} + 5\gamma_{33} + 4\gamma_{34} - 4\gamma_{41} - 4\gamma_{42} - 5\gamma_{43} - 4\gamma_{44}) = 0$
Tipo III	$\alpha_1 - \alpha_4 + 1/4 (\gamma_{11} + \gamma_{12} + \gamma_{13} + \gamma_{14} - \gamma_{41} - \gamma_{42} - \gamma_{43} - \gamma_{44}) = 0$
e IV ⁽²⁾	$\alpha_2 - \alpha_4 + 1/4 (\gamma_{21} + \gamma_{22} + \gamma_{23} + \gamma_{24} - \gamma_{41} - \gamma_{42} - \gamma_{43} - \gamma_{44}) = 0$
H_0^3 e H_0^4	$\alpha_3 - \alpha_4 + 1/4 (\gamma_{31} + \gamma_{32} + \gamma_{33} + \gamma_{34} - \gamma_{41} - \gamma_{42} - \gamma_{43} - \gamma_{44}) = 0$

⁽¹⁾ As hipóteses para os Tipos I e II são equivalentes.

⁽²⁾ As hipóteses para os Tipos III e IV são equivalentes.

TABELA 14. Hipóteses testadas sobre o fator B (fases da lua).

Hipóteses	Hipóteses testadas
Tipo I, II, III e IV ⁽¹⁾	$\beta_1 - \beta_4 + 1/4 (\gamma_{11} - \gamma_{14} + \gamma_{21} - \gamma_{24} + \gamma_{31} - \gamma_{34} + \gamma_{41} - \gamma_{44}) = 0$
H_0^5, H_0^6	$\beta_2 - \beta_4 + 1/4 (\gamma_{12} - \gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} + \gamma_{42} - \gamma_{44}) = 0$
H_0^7 e H_0^8	$\beta_3 - \beta_4 + 1/4 (\gamma_{13} - \gamma_{14} + \gamma_{23} - \gamma_{24} + \gamma_{33} - \gamma_{34} + \gamma_{43} - \gamma_{44}) = 0$

⁽¹⁾ As hipóteses para os Tipos I, II, III e IV são equivalentes.

TABELA 15. Hipóteses testadas sobre a interação.

Hipóteses	Hipóteses testadas
H_0^9	$\gamma_{11} - \gamma_{14} - \gamma_{41} + \gamma_{44} = 0$
	$\gamma_{12} - \gamma_{14} - \gamma_{42} + \gamma_{44} = 0$
	$\gamma_{13} - \gamma_{14} - \gamma_{43} + \gamma_{44} = 0$
	$\gamma_{21} - \gamma_{24} - \gamma_{41} + \gamma_{44} = 0$
	$\gamma_{22} - \gamma_{24} - \gamma_{42} + \gamma_{44} = 0$
	$\gamma_{23} - \gamma_{24} - \gamma_{43} + \gamma_{44} = 0$
	$\gamma_{31} - \gamma_{34} - \gamma_{41} + \gamma_{44} = 0$
	$\gamma_{32} - \gamma_{34} - \gamma_{42} + \gamma_{44} = 0$
	$\gamma_{33} - \gamma_{34} - \gamma_{43} + \gamma_{44} = 0$

Nota: Para as ordens A-B e B-A as hipóteses sobre interação são equivalentes

TABELA 16. Análise de variância dos dados da Tabela 3, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo I e II, e respectivas hipóteses associadas.

Ordem A-B						Ordem B-A							
CV	GL	Tipo I		Tipo II		SS	CV	GL	Tipo I		Tipo II		SS
		H_0	R()	H_0	R()				H_0	R()	H_0	R()	
A	3	H_0^1	R($\alpha \mu$)	H_0^2	R($\alpha \mu,\beta$)	13 301,300	B	3	H_0^5	R($\beta \mu$)	H_0^6	R($\beta \mu,\alpha$)	3 285,757
B	3	H_0^6	R($\beta \mu,\alpha$)	H_0^6	R($\beta \mu,\alpha$)	3 285,757	A	3	H_0^2	R($\alpha \mu,\beta$)	H_0^2	R($\alpha \mu,\beta$)	13 301,300
A-B	9	H_0^9	R($\gamma \mu,\alpha,\beta$)	H_0^9	R($\gamma \mu,\alpha,\beta$)	471,015	B-A	9	H_0^9	R($\gamma \mu,\alpha,\beta$)	H_0^9	R($\gamma \mu,\alpha,\beta$)	471,015
Res	188					30 933,712	Res	188					30 933,712

59

TABELA 17. Análise de variância dos dados da Tabela 3, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo III e IV, e respectivas hipóteses associadas.

Ordem A-B						Ordem B-A							
CV	GL	Tipo III		Tipo IV		SS	CV	GL	Tipo III		Tipo IV		SS
		H_0	R()	H_0	R()				H_0	R()	H_0	R()	
A	3	H_0^3	R($\alpha \mu,\beta,\gamma$)	H_0^4	—	13 146,842	B	3	H_0^7	R($\beta \mu,\alpha,\gamma$)	H_0^8	—	3 285,757
B	3	H_0^7	R($\beta \mu,\alpha,\gamma$)	H_0^8	—	3 285,757	A	3	H_0^3	R($\alpha \mu,\beta,\gamma$)	H_0^4	—	13 146,842
A-B	9	H_0^9	R($\gamma \mu,\alpha,\beta$)	H_0^9	—	471,015	B-A	9	H_0^9	R($\gamma \mu,\alpha,\beta$)	H_0^9	—	471,015
Res	188					30 933,712	Res	188					30 933,712

4.2 2º caso: fatorial 4 x 4 desbalanceado com caselas vazias

A forma geral das funções estimáveis liberadas pelo SAS, obtidas por meio de $L\theta$ para os dados da Tabela 4, com duas caselas ausentes, está apresentada na Tabela 18 para a ordem A-B e na Tabela 19, para a ordem B-A.

TABELA 18. Forma geral de funções estimáveis dos dados da Tabela 4, fornecida pelo PROC GLM do SAS, na ordem A-B

Parâmetros	Coefficientes
μ	L1
α_1	L2
α_2	L3
α_3	L4
α_4	L1-L2-L3-L4
β_1	L6
β_2	L7
β_3	L8
β_4	L1-L6-L7-L8
γ_{11}	L10
γ_{13}	L11
γ_{14}	L2-L10-L11
γ_{21}	L13
γ_{22}	L14
γ_{23}	L15
γ_{24}	L3-L13-L14-L15
γ_{31}	L17
γ_{32}	L18
γ_{33}	-L1+L2+L3+L4+L6+L7+L8-L10-L11-L13-L14-L15-L17-L18
γ_{34}	L1-L2-L3-L6-L7-L8+L10+L11+L13+L14+L15
γ_{41}	L6-L10-L13-L17
γ_{42}	L7-L14-L18
γ_{43}	L1-L2-L3-L4-L6-L7+L10+L12+L13+L14+L17+L18

Têm-se 14 símbolos (L1, L2, L3, L4, L6, L7, L8, L10, L11, L13, L14, L15, L17 e L18) envolvidos, portanto, o posto máximo da hipótese tem 14 graus de liberdade.

TABELA 19. Forma geral de funções estimáveis dos dados da Tabela 4, fornecida pelo PROC GLM do SAS, na ordem B-A.

Parâmetros	Coefficientes
μ	L1
α_1	L2
α_2	L3
α_3	L4
α_4	L1-L2-L3-L4
β_1	L6
β_2	L7
β_3	L8
β_4	L1-L6-L7-L8
γ_{11}	L10
γ_{12}	L11
γ_{13}	L12
γ_{14}	L2-L10-L11-L12
γ_{22}	L14
γ_{23}	L15
γ_{24}	L3-L14-L15
γ_{31}	L17
γ_{32}	L18
γ_{33}	-L1+L2+L3+L4+L6+L7+L8-L10-L11-L12-L14-L15-L17-L18
γ_{34}	L1-L2-L3-L6-L7-L8+L10+L11+L12+L14+L15
γ_{41}	L6-L10-L17
γ_{42}	L7-L11-L14-L18
γ_{43}	L1-L2-L3-L4-L6-L7+L10+L11+L14+L17+L18

Têm-se 14 símbolos (L1, L2, L3, L4, L6, L7, L8, L10, L11, L12, L14, L15, L17 e L18) envolvidos, sendo o posto máximo da hipótese de 14 graus de liberdade.

As Tabelas 20 a 27 mostram as funções estimáveis fornecidas pelo SAS.

TABELA 20. Forma geral das funções estimáveis Tipo I do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.

		A	B	AB
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	-0.0606*L3-0.0833*L4	L6	0
	2	-0.3333*L2-0.0606*L3-0.0833*L4	L7	0
	3	-0.1515*L3-0.0833*L4	L8	0
	4	0.3333*L2+0.2727*L3+0.25*L4	-L6-L7-L8	0
AB	1 1	0.3333*L2	0.2868*L6+0.125*L7+0.0294*L8	L10
	1 3	0.3333*L2	0.0392*L6+0.125*L7+0.3064*L8	L11
	1 4	0.3333*L2	-0.326*L6-0.25*L7-0.3358*L8	-L10-L11
	2 1	0.2727*L3	0.277*L6+0.0466*L8	L13
	2 2	0.2727*L3	0.0392*L6+0.375*L7+0.0564*L8	L14
	2 3	0.1818*L3	0.0196*L6+0.2157*L8	L15
	2 4	0.2727*L3	-0.3358*L6-0.375*L7-0.3186*L8	-L13-L14-L15
	3 1	0.25*L4	0.2745*L6+0.0196*L8	L17
	3 2	0.25*L4	0.0368*L6+0.375*L7+0.0294*L8	L18
	3 3	0.25*L4	0.027*L6+0.2966*L8	-L10-L11-L13-L14-L15-L17-L18
	3 4	0.25*L4	-0.3382*L6-0.375*L7-0.3456*L8	L10+L11+L13+L14+L15
	4 1	-0.3333*L2-0.3333*L3-0.3333*L4	0.1618*L6-0.125*L7-0.0956*L8	-L10-L13-L17
	4 2	-0.3333*L2-0.3333*L3-0.3333*L4	-0.076*L6+0.25*L7-0.0858*L8	-L14-L18
4 3	-0.3333*L2-0.3333*L3-0.3333*L4	-0.0858*L6-0.125*L7+0.1814*L8	L10+L13+L14+L17+L18	

TABELA 21. Forma geral das funções estimáveis Tipo I do SAS para os dados da Tabela 4, na ordem B-A, para os fatores B, A e interação BA.

		B	A	BA
μ	0	0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	-0.0833*L2-0.3333*L3-0.0606*L4	L6	0
	2	-0.0833*L2-0.1515*L4	L7	0
	3	-0.0833*L2-0.0606*L4	L8	0
	4	0.25*L2+0.3333*L3+0.2727*L4	-L6-L7-L8	0
BA	1 1	0.25*L2	0.375*L6+0.0294*L7+0.0368*L8	L10
	1 2	0.25*L2	0.2966*L7+0.027*L8	L11
	1 3	0.25*L2	0.0196*L7+0.2745*L8	L12
	1 4	0.25*L2	-0.375*L6-0.3456*L7-0.3382*L8	-L10-L11-L12
	2 2	0.3333*L3	0.125*L6+0.3064*L7+0.0392*L8	L14
	2 3	0.3333*L3	0.125*L6+0.0294*L7+0.2868*L8	L15
	2 4	0.3333*L3	-0.25*L6-0.3358*L7-0.326*L8	-L14-L15
	3 1	0.2727*L4	0.375*L6+0.0564*L7+0.0392*L8	L17
	3 2	0.1818*L4	0.2157*L7+0.0196*L8	L18
	3 3	0.2727*L4	0.0466*L7+0.277*L8	-L10-L11-L12-L14-L15-L17-L18
	3 4	0.2727*L4	-0.375*L6-0.3186*L7-0.3358*L8	L10+L11+L12+L14+L15
	4 1	-0.3333*L2-0.3333*L3-0.3333*L4	0.25*L6-0.0858*L7-0.076*L8	-L10-L17
	4 2	-0.3333*L2-0.3333*L3-0.3333*L4	-0.125*L6+0.1814*L7-0.0858*L8	-L11-L14-L18
	4 3	-0.3333*L2-0.3333*L3-0.3333*L4	-0.125*L6-0.0956*L7+0.1618*L8	L10+L11+L14+L17+L18

TABELA 22. Forma geral das funções estimáveis Tipo II do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.

		A	B	AB
μ	0	0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
AB	1 1	$0.375*L2+0.0294*L3+0.0368*L4$	$0.2868*L6+0.125*L7+0.0294*L8$	L10
	1 3	$0.375*L2+0.0564*L3+0.0392*L4$	$0.0392*L6+0.125*L7+0.3064*L8$	L11
	1 4	$0.25*L2-0.0858*L3-0.076*L4$	$-0.326*L6-0.25*L7-0.3358*L8$	-L10-L11
	2 1	$0.2966*L3+0.027*L4$	$0.277*L6+0.0466*L8$	L13
	2 2	$0.125*L2+0.3064*L3+0.0392*L4$	$0.0392*L6+0.375*L7+0.0564*L8$	L14
	2 3	$0.2157*L3+0.0196*L4$	$0.0196*L6+0.2157*L8$	L15
	2 4	$-0.125*L2+0.1814*L3-0.0858*L4$	$-0.3358*L6-0.375*L7-0.3186*L8$	-L13-L14-L15
	3 1	$0.0196*L3+0.2745*L4$	$0.2745*L6+0.0196*L8$	L17
	3 2	$0.125*L2+0.0294*L3+0.2868*L4$	$0.0368*L6+0.375*L7+0.0294*L8$	L18
	3 3	$0.0466*L3+0.277*L4$	$0.027*L6+0.2966*L8$	-L10-L11-L13-L14-L15-L17-L18
	3 4	$-0.125*L2-0.0956*L3+0.1618*L4$	$-0.3382*L6-0.375*L7-0.3456*L8$	L10+L11+L13+L14+L15
	4 1	$-0.375*L2-0.3456*L3-0.3382*L4$	$0.1618*L6-0.125*L7-0.0956*L8$	-L10-L13-L17
4 2	$-0.25*L2-0.3358*L3-0.326*L4$	$-0.076*L6+0.25*L7-0.0858*L8$	-L14-L18	
4 3	$-0.375*L2-0.3186*L3-0.3358*L4$	$-0.0858*L6-0.125*L7+0.1814*L8$	L10+L13+L14+L17+L18	

TABELA 23. Forma geral das funções estimáveis Tipo II do SAS para os dados da Tabela 4, na ordem B-A, para os fatores B, A e interação BA.

		B	A	BA
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
BA	1 1	$0.2868*L2+0.125*L3+0.0294*L4$	$0.375*L6+0.0294*L7+0.0368*L8$	L10
	1 2	$0.277*L2+0.0466*L4$	$0.2966*L7+0.027*L8$	L11
	1 3	$0.2745*L2+0.0196*L4$	$0.0196*L7+0.2745*L8$	L12
	1 4	$0.1618*L2-0.125*L3-0.0956*L4$	$-0.375*L6-0.3456*L7-0.3382*L8$	-L10-L11-L12
	2 2	$0.0392*L2+0.375*L3+0.0564*L4$	$0.125*L6+0.3064*L7+0.0392*L8$	L14
	2 3	$0.0368*L2+0.375*L3+0.0294*L4$	$0.125*L6+0.0294*L7+0.2868*L8$	L15
	2 4	$-0.076*L2+0.25*L3-0.0858*L4$	$-0.25*L6-0.3358*L7-0.326*L8$	-L14-L15
	3 1	$0.0392*L2+0.125*L3+0.3064*L4$	$0.375*L6+0.0564*L7+0.0392*L8$	L17
	3 2	$0.0196*L2+0.2157*L4$	$0.2157*L7+0.0196*L8$	L18
	3 3	$0.027*L2+0.2966*L4$	$0.0466*L7+0.277*L8$	-L10-L11-L12-L14-L15-L17-L18
	3 4	$-0.0858*L2-0.125*L3+0.1814*L4$	$-0.375*L6-0.3186*L7-0.3358*L8$	L10+L11+L12+L14+L15
	4 1	$-0.326*L2-0.25*L3-0.3358*L4$	$0.25*L6-0.0858*L7-0.076*L8$	-L10-L17
4 2	$-0.3358*L2-0.375*L3-0.3186*L4$	$-0.125*L6+0.1814*L7-0.0858*L8$	-L11-L14-L18	
4 3	$-0.3382*L2-0.375*L3-0.3456*L4$	$-0.125*L6-0.0956*L7+0.1618*L8$	L10+L11+L14+L17+L18	

TABELA 24. Forma geral das funções estimáveis Tipo III do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.

		A	B	AB
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
AB	1 1	$0.375*L2+0.0375*L3+0.0375*L4$	$0.2875*L6+0.125*L7+0.0375*L8$	L10
	1 3	$0.375*L2+0.0375*L3+0.0375*L4$	$0.0375*L6+0.125*L7+0.2875*L8$	L11
	1 4	$0.25*L2-0.075*L3-0.075*L4$	$-0.325*L6-0.25*L7-0.325*L8$	-L10-L11
	2 1	$0.275*L3+0.025*L4$	$0.275*L6+0.025*L8$	L13
	2 2	$0.125*L2+0.2875*L3+0.0375*L4$	$0.0375*L6+0.375*L7+0.0375*L8$	L14
	2 3	$0.275*L3+0.025*L4$	$0.025*L6+0.275*L8$	L15
	2 4	$-0.125*L2+0.1625*L3-0.0875*L4$	$-0.3375*L6-0.375*L7-0.3375*L8$	-L13-L14-L15
	3 1	$0.025*L3+0.275*L4$	$0.275*L6+0.025*L8$	L17
	3 2	$0.125*L2+0.0375*L3+0.2875*L4$	$0.0375*L6+0.375*L7+0.0375*L8$	L18
	3 3	$0.025*L3+0.275*L4$	$0.025*L6+0.275*L8$	-L10-L11-L13-L14-L15-L17-L18
	3 4	$-0.125*L2-0.0875*L3+0.1625*L4$	$-0.3375*L6-0.375*L7-0.3375*L8$	L10+L11+L13+L14+L15
	4 1	$-0.375*L2-0.3375*L3-0.3375*L4$	$0.1625*L6-0.125*L7-0.0875*L8$	-L10-L13-L17
4 2	$-0.25*L2-0.325*L3-0.325*L4$	$-0.075*L6+0.25*L7-0.075*L8$	-L14-L18	
4 3	$-0.375*L2-0.3375*L3-0.3375*L4$	$-0.0875*L6-0.125*L7+0.1625*L8$	L10+L13+L14+L17+L18	

TABELA 25. Forma geral das funções estimáveis Tipo III do SAS para os dados da Tabela 4, na ordem B-A, para os fatores B, A e interação BA.

		B	A	BA
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
BA	1 1	$0.2875*L2+0.125*L3+0.0375*L4$	$0.375*L6+0.0375*L7+0.0375*L8$	L10
	1 2	$0.275*L2+0.025*L4$	$0.275*L7+0.025*L8$	L11
	1 3	$0.275*L2+0.025*L4$	$0.025*L7+0.275*L8$	L12
	1 4	$0.1625*L2-0.125*L3-0.0875*L4$	$-0.375*L6-0.3375*L7-0.3375*L8$	-L10-L11-L12
	2 2	$0.0375*L2+0.375*L3+0.0375*L4$	$0.125*L6+0.2875*L7+0.0375*L8$	L14
	2 3	$0.0375*L2+0.375*L3+0.0375*L4$	$0.125*L6+0.0375*L7+0.2875*L8$	L15
	2 4	$-0.075*L2+0.25*L3-0.075*L4$	$-0.25*L6-0.325*L7-0.325*L8$	-L14-L15
	3 1	$0.0375*L2+0.125*L3+0.2875*L4$	$0.375*L6+0.0375*L7+0.0375*L8$	L17
	3 2	$0.025*L2+0.275*L4$	$0.275*L7+0.025*L8$	L18
	3 3	$0.025*L2+0.275*L4$	$0.025*L7+0.275*L8$	-L10-L11-L12-L14-L15-L17-L18
	3 4	$-0.0875*L2-0.125*L3+0.1625*L4$	$-0.375*L6-0.3375*L7-0.3375*L8$	L10+L11+L12+L14+L15
	4 1	$-0.325*L2-0.25*L3-0.325*L4$	$0.25*L6-0.075*L7-0.075*L8$	-L10-L17
4 2	$-0.3375*L2-0.375*L3-0.3375*L4$	$-0.125*L6+0.1625*L7-0.0875*L8$	-L11-L14-L18	
4 3	$-0.3375*L2-0.375*L3-0.3375*L4$	$-0.125*L6-0.0875*L7+0.1625*L8$	L10+L11+L14+L17+L18	

TABELA 26. Forma geral das funções estimáveis Tipo IV do SAS para os dados da Tabela 4, na ordem A-B, para os fatores A, B e interação AB.

		A	B	AB
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
AB	1 1	0.5*L2	0.3333*L6	L10
	1 3	0.5*L2	0.3333*L8	L11
	1 4	0	-0.3333*L6-0.3333*L8	-L10-L11
	2 1	0.3333*L3	0.3333*L6	L13
	2 2	0.3333*L3	0.5*L7	L14
	2 3	0.3333*L3	0.3333*L8	L15
	2 4	0	-0.3333*L6-0.5*L7-0.3333*L8	-L13-L14-L15
	3 1	0.3333*L4	0.3333*L6	L17
	3 2	0.3333*L4	0.5*L7	L18
	3 3	0.3333*L4	0.3333*L8	-L10-L11-L13-L14-L15-L17-L18
	3 4	0	-0.3333*L6-0.5*L7-0.3333*L8	L10+L11+L13+L14+L15
	4 1	-0.5*L2-0.3333*L3-0.3333*L4	0	-L10-L13-L17
4 2	-0.3333*L3-0.3333*L4	0	-L14-L18	
4 3	-0.5*L2-0.3333*L3-0.3333*L4	0	L10+L13+L14+L17+L18	

Nota: Existem outras funções estimáveis Tipo IV.

TABELA 27. Forma geral das funções estimáveis Tipo IV do SAS para os dados da Tabela 4, na ordem B-A, para os fatores B, A e interação BA.

		B	A	BA
μ		0	0	0
A	1	L2	0	0
	2	L3	0	0
	3	L4	0	0
	4	-L2-L3-L4	0	0
B	1	0	L6	0
	2	0	L7	0
	3	0	L8	0
	4	0	-L6-L7-L8	0
BA	1 1	0.3333*L2	0.5*L6	L10
	1 2	0.3333*L2	0.3333*L7	L11
	1 3	0.3333*L2	0.3333*L8	L12
	1 4	0	-0.5*L6-0.3333*L7-0.3333*L8	-L10-L11-L12
	2 2	0.5*L3	0.3333*L7	L14
	2 3	0.5*L3	0.3333*L8	L15
	2 4	0	-0.3333*L7-0.3333*L8	-L14-L15
	3 1	0.3333*L4	0.5*L6	L17
	3 2	0.3333*L4	0.3333*L7	L18
	3 3	0.3333*L4	0.3333*L8	-L10-L11-L12-L14-L15-L17-L18
	3 4	0	-0.5*L6-0.3333*L7-0.3333*L8	L10+L11+L12+L14+L15
	4 1	-0.3333*L2-0.3333*L4	0	-L10-L17
4 2	-0.3333*L2-0.5*L3-0.3333*L4	0	-L11-L14-L18	
4 3	-0.3333*L2-0.5*L3-0.3333*L4	0	L10+L11+L14+L17+L18	

Nota: Existem outras funções estimáveis Tipo IV.

TABELA 28. Funções estimáveis sobre o fator A (cultivares) dos Tipos I e II.

Funções estimáveis	
T	L2=1; L3=0
I	L4=0
P	L2=0
O	L3=1
I	L2=0 L3=0 L4=1
T	L2=1 L3=0
I	L4=0
P	L2=0
O	L3=1
II	L2=0 L3=0 L4=1

$$\alpha_1 - \alpha_4 + 1/3 (-\beta_2 + \beta_4 + \gamma_{11} + \gamma_{13} + \gamma_{14} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/33 (-2\beta_1 - 2\beta_2 + 5\beta_3 + 9\beta_4 + 9\gamma_{21} + 9\gamma_{22} + 6\gamma_{23} - 9\gamma_{24} - 11\gamma_{41} + 11\gamma_{42} - 11\gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/12 (-\beta_1 - \beta_2 - \beta_3 + 3\beta_4 + 3\gamma_{31} + 3\gamma_{32} + 3\gamma_{33} - 3\gamma_{34} - 4\gamma_{41} + 4\gamma_{42} - 4\gamma_{43})$$

$$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 2\gamma_{43})$$

$$\alpha_2 - \alpha_4 + 0,0294\gamma_{11} + 0,0564\gamma_{13} - 0,0858\gamma_{14} + 0,2966\gamma_{21} + 0,3064\gamma_{22} + 0,2157\gamma_{23} + 0,1814\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,0466\gamma_{33} - 0,0956\gamma_{34} - 0,3456\gamma_{41} - 0,3358\gamma_{42} - 0,3186\gamma_{43}$$

$$\alpha_3 - \alpha_4 + 0,0368\gamma_{11} + 0,0392\gamma_{13} - 0,076\gamma_{14} + 0,027\gamma_{21} + 0,0392\gamma_{22} + 0,0196\gamma_{23} - 0,0858\gamma_{24} + 0,2745\gamma_{31} + 0,2868\gamma_{32} + 0,277\gamma_{33} + 0,1618\gamma_{34} - 0,3382\gamma_{41} - 0,326\gamma_{42} - 0,3358\gamma_{43}$$

TABELA 29. Funções estimáveis sobre o fator A (cultivares) dos Tipos III e IV.

Funções estimáveis	
T	L2=1; L3=0
I	L4=0
P	L2=0
O	L3=1
III	L2=0 L3=0 L4=1
T	L2=1 L3=0
I	L4=0
P	L2=0
O	L3=1
IV	L2=0 L3=0 L4=1

$$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 3\gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 22\gamma_{21} + 23\gamma_{22} + 22\gamma_{23} + 13\gamma_{24} + 2\gamma_{31} + 3\gamma_{32} + 2\gamma_{33} - 7\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 7\gamma_{24} + 22\gamma_{31} + 23\gamma_{32} + 22\gamma_{33} + 13\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$$

$$\alpha_1 - \alpha_4 + 1/2 (\gamma_{11} + \gamma_{13} - \gamma_{41} - \gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/3 (\gamma_{21} + \gamma_{22} + \gamma_{23} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/3 (\gamma_{31} + \gamma_{32} + \gamma_{33} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

TABELA 28. Funções estimáveis sobre o fator A (cultivares) dos Tipos I e II.

		Funções estimáveis
T	L2=1;	$\alpha_1 - \alpha_4 + 1/3 (-\beta_2 + \beta_4 + \gamma_{11} + \gamma_{13} + \gamma_{14} - \gamma_{41} - \gamma_{42} - \gamma_{43})$
	L3=0	
I	L4=0	
P	L2=0	$\alpha_2 - \alpha_4 + 1/33 (-2\beta_1 - 2\beta_2 + 5\beta_3 + 9\beta_4 + 9\gamma_{21} + 9\gamma_{22} + 6\gamma_{23} - 9\gamma_{24} - 11\gamma_{41} + 11\gamma_{42} - 11\gamma_{43})$
	L3=1	
O	L4=0	
I	L2=0	$\alpha_3 - \alpha_4 + 1/12 (-\beta_1 - \beta_2 - \beta_3 + 3\beta_4 + 3\gamma_{31} + 3\gamma_{32} + 3\gamma_{33} - 3\gamma_{34} - 4\gamma_{41} + 4\gamma_{42} - 4\gamma_{43})$
	L3=0	
	L4=1	
T	L2=1;	$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 2\gamma_{43})$
	L3=0	
I	L4=0	
P	L2=0	$\alpha_2 - \alpha_4 + 0,0294\gamma_{11} + 0,0564\gamma_{13} - 0,0858\gamma_{14} + 0,2966\gamma_{21} + 0,3064\gamma_{22} + 0,2157\gamma_{23} + 0,1814\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,0466\gamma_{33} - 0,0956\gamma_{34} - 0,3456\gamma_{41} - 0,3358\gamma_{42} - 0,3186\gamma_{43}$
	L3=1	
O	L4=0	
II	L2=0	$\alpha_3 - \alpha_4 + 0,0368\gamma_{11} + 0,0392\gamma_{13} - 0,076\gamma_{14} + 0,027\gamma_{21} + 0,0392\gamma_{22} + 0,0196\gamma_{23} - 0,0858\gamma_{24} + 0,2745\gamma_{31} + 0,2868\gamma_{32} + 0,277\gamma_{33} + 0,1618\gamma_{34} - 0,3382\gamma_{41} - 0,326\gamma_{42} - 0,3358\gamma_{43}$
	L3=0	
	L4=1	

TABELA 29. Funções estimáveis sobre o fator A (cultivares) dos Tipos III e IV.

		Funções estimáveis
T	L2=1;	$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 3\gamma_{43})$
	L3=0	
I	L4=0	
P	L2=0	$\alpha_2 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 22\gamma_{21} + 23\gamma_{22} + 22\gamma_{23} + 13\gamma_{24} + 2\gamma_{31} + 3\gamma_{32} + 2\gamma_{33} - 7\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$
	L3=1	
O	L4=0	
III	L2=0	$\alpha_3 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 7\gamma_{24} + 22\gamma_{31} + 23\gamma_{32} + 22\gamma_{33} + 13\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$
	L3=0	
	L4=1	
T	L2=1;	$\alpha_1 - \alpha_4 + 1/2 (\gamma_{11} + \gamma_{13} - \gamma_{41} - \gamma_{43})$
	L3=0	
I	L4=0	
P	L2=0	$\alpha_2 - \alpha_4 + 1/3 (\gamma_{21} + \gamma_{22} + \gamma_{23} - \gamma_{41} - \gamma_{42} - \gamma_{43})$
	L3=1	
O	L4=0	
IV	L2=0	$\alpha_3 - \alpha_4 + 1/3 (\gamma_{31} + \gamma_{32} + \gamma_{33} - \gamma_{41} - \gamma_{42} - \gamma_{43})$
	L3=0	
	L4=1	

e a aproximação de mínimos quadrados, \hat{y} , para o vetor y das observações:

$$\hat{y} = X\theta^0 = \begin{bmatrix} 32,383 \\ 32,383 \\ 32,383 \\ 32,333 \\ 32,333 \\ 32,333 \\ 20,717 \\ 20,717 \\ 20,717 \\ 49,100 \\ 49,100 \\ 49,100 \\ 35,583 \\ 35,583 \\ 35,583 \\ 39,275 \\ 39,275 \\ 31,433 \\ 31,433 \\ \vdots \\ 46,083 \\ 46,083 \\ 46,083 \end{bmatrix}$$

Nota-se que o vetor da aproximação de mínimos quadrados é constituído pelas médias das caselas.

4.2.2 Estimação

Nas Tabelas 28 e 29 têm-se os quatro tipos de funções estimáveis sobre os efeitos principais para o fator A; na Tabela 30, para o fator B e na Tabela 31, as funções estimáveis sobre a interação. Elas foram obtidas utilizando as funções estimáveis fornecidas pelo SAS (Tabelas 20, 22, 24 e 26) para o modelo (dado em 1), na ordem A-B.

As funções estimáveis que serão aplicadas nas hipóteses testadas, no Modelo-S, são obtidas atribuindo-se valor um para cada um dos coeficientes e zerando os demais.

A construção das funções estimáveis Tipo II é feita de forma independente para cada fator. Para o Tipo I isto não ocorre, pois seu ajuste é seqüencial e a sua obtenção para um determinado fator depende do fator que o antecede. Para o fator B, devido ao ajuste seqüencial das funções estimáveis do Tipo I, são iguais às funções estimáveis Tipo II.

As funções estimáveis do Tipo III, quando se assume o coeficiente $L6=1$ zerando os demais, em $\beta_1 - \beta_4$, é aproximadamente igual às do Tipo I e II; como exemplo, o coeficiente da interação $0,2868\gamma_{11}$ da função estimável do Tipo I e II é aproximado do coeficiente da interação $0,2875\gamma_{11}$ (ou $23/80\gamma_{11}$) da função estimável Tipo III. Entretanto, quando o coeficiente $L7=1$ e os demais zero, em $\beta_2 - \beta_4$, a função estimável Tipo I = Tipo II = Tipo III.

Quanto à alteração da ordem das variáveis no modelo, em B, A, B-A, apenas as funções estimáveis do Tipo I sofrem alteração.

Na presença de interação não se obtêm funções estimáveis exclusivamente sobre os efeitos principais, ocorrendo, normalmente, um contraste entre os níveis de um fator, seguido de outros parâmetros, ou seja, os efeitos do outro fator e/ou de interações, reforçando o alerta de Camarinha Filho (1995), de que o problema principal não está apenas na interpretação das hipóteses embasadas nessas funções estimáveis, mas sim que vários usuários nem imaginam o que estão testando.

TABELA 28. Funções estimáveis sobre o fator A (cultivares) dos Tipos I e II.

Funções estimáveis	
T	$L_2=1;$ $L_3=0$
I	$L_4=0$
P	$L_2=0$
O	$L_3=1$
	$L_4=0$
I	$L_2=0$
	$L_3=0$
	$L_4=1$
T	$L_2=1;$ $L_3=0$
I	$L_4=0$
P	$L_2=0$
O	$L_3=1$
	$L_4=0$
II	$L_2=0$
	$L_3=0$
	$L_4=1$

$$\alpha_1 - \alpha_4 + 1/3 (-\beta_2 + \beta_4 + \gamma_{11} + \gamma_{13} + \gamma_{14} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/33 (-2\beta_1 - 2\beta_2 + 5\beta_3 + 9\beta_4 + 9\gamma_{21} + 9\gamma_{22} + 6\gamma_{23} - 9\gamma_{24} - 11\gamma_{41} + 11\gamma_{42} - 11\gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/12 (-\beta_1 - \beta_2 - \beta_3 + 3\beta_4 + 3\gamma_{31} + 3\gamma_{32} + 3\gamma_{33} - 3\gamma_{34} - 4\gamma_{41} + 4\gamma_{42} - 4\gamma_{43})$$

$$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 2\gamma_{43})$$

$$\alpha_2 - \alpha_4 + 0,0294\gamma_{11} + 0,0564\gamma_{13} - 0,0858\gamma_{14} + 0,2966\gamma_{21} + 0,3064\gamma_{22} + 0,2157\gamma_{23} + 0,1814\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,0466\gamma_{33} - 0,0956\gamma_{34} - 0,3456\gamma_{41} - 0,3358\gamma_{42} - 0,3186\gamma_{43}$$

$$\alpha_3 - \alpha_4 + 0,0368\gamma_{11} + 0,0392\gamma_{13} - 0,076\gamma_{14} + 0,027\gamma_{21} + 0,0392\gamma_{22} + 0,0196\gamma_{23} - 0,0858\gamma_{24} + 0,2745\gamma_{31} + 0,2868\gamma_{32} + 0,277\gamma_{33} + 0,1618\gamma_{34} - 0,3382\gamma_{41} - 0,326\gamma_{42} - 0,3358\gamma_{43}$$

TABELA 29. Funções estimáveis sobre o fator A (cultivares) dos Tipos III e IV.

Funções estimáveis	
T	$L_2=1;$ $L_3=0$
I	$L_4=0$
P	$L_2=0$
O	$L_3=1$
	$L_4=0$
III	$L_2=0$
	$L_3=0$
	$L_4=1$
T	$L_2=1;$ $L_3=0$
I	$L_4=0$
P	$L_2=0$
O	$L_3=1$
	$L_4=0$
IV	$L_2=0$
	$L_3=0$
	$L_4=1$

$$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 3\gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 22\gamma_{21} + 23\gamma_{22} + 22\gamma_{23} + 13\gamma_{24} + 2\gamma_{31} + 3\gamma_{32} + 2\gamma_{33} - 7\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 7\gamma_{24} + 22\gamma_{31} + 23\gamma_{32} + 22\gamma_{33} + 13\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43})$$

$$\alpha_1 - \alpha_4 + 1/2 (\gamma_{11} + \gamma_{13} - \gamma_{41} - \gamma_{43})$$

$$\alpha_2 - \alpha_4 + 1/3 (\gamma_{21} + \gamma_{22} + \gamma_{23} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

$$\alpha_3 - \alpha_4 + 1/3 (\gamma_{31} + \gamma_{32} + \gamma_{33} - \gamma_{41} - \gamma_{42} - \gamma_{43})$$

TABELA 30. Funções estimáveis sobre o fator B (fases da lua) dos Tipos I, II, III e IV.

		Funções estimáveis	
T	L6=1	$\beta_1 - \beta_4 + 0,2868\gamma_{11} + 0,0392\gamma_{13} - 0,326\gamma_{14} + 0,277\gamma_{21} + 0,0392\gamma_{22}$	
I	L7=0	$+0,0196\gamma_{23} - 0,3358\gamma_{24} + 0,2745\gamma_{31} + 0,0368\gamma_{32} + 0,027\gamma_{33} - 0,3382\gamma_{34}$	
P	L8=0	$+0,1618\gamma_{41} - 0,076\gamma_{42} - 0,0858\gamma_{43}$	

O	L6=0		
I	L7=1	$\beta_2 - \beta_4 + 1/8 (\gamma_{11} + \gamma_{13} - 2\gamma_{14} + 3\gamma_{22} - 3\gamma_{24} + 3\gamma_{32} - 3\gamma_{34} - \gamma_{41} + 2\gamma_{42} - \gamma_{43})$	
e	L8=0		
II	L6=0	$\beta_3 - \beta_4 + 0,0294\gamma_{11} + 0,3064\gamma_{13} - 0,3358\gamma_{14} + 0,0466\gamma_{21} + 0,0564\gamma_{22}$	
(1)	L7=0	$+0,2157\gamma_{23} - 0,3186\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,2966\gamma_{33} - 0,3456\gamma_{34} -$	
	L8=1	$0,0956\gamma_{41} - 0,0858\gamma_{42} + 0,1814\gamma_{43}$	

T	L6=1	$\beta_1 - \beta_4 + 1/80 (23\gamma_{11} + 3\gamma_{13} - 26\gamma_{14} + 22\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 27\gamma_{24} + 22\gamma_{31} +$	
I	L7=0	$3\gamma_{32} + 2\gamma_{33} - 27\gamma_{34} + 13\gamma_{41} - 6\gamma_{42} - 7\gamma_{43})$ ⁽²⁾	
P	L8=0		
O	L6=0		
	L7=1	$\beta_2 - \beta_4 + 1/8 (\gamma_{11} + \gamma_{13} - 2\gamma_{14} + 3\gamma_{22} - 3\gamma_{24} + 3\gamma_{32} - 3\gamma_{34} - \gamma_{41} + 2\gamma_{42} - \gamma_{43})$ ⁽³⁾	
	L9=0		

III	L6=0	$\beta_3 - \beta_4 + 1/80 (3\gamma_{11} + 23\gamma_{13} - 26\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 22\gamma_{23} - 27\gamma_{24} + 2\gamma_{31} +$	
	L7=0	$3\gamma_{32} + 22\gamma_{33} - 27\gamma_{34} - 7\gamma_{41} - 6\gamma_{42} + 13\gamma_{43})$	
	L8=1		

T	L6=1	$\beta_1 - \beta_4 + 1/3 (\gamma_{11} - \gamma_{14} + \gamma_{21} - \gamma_{24} + \gamma_{31} - \gamma_{34})$	
I	L7=0		
P	L8=0		
O	L6=0		
	L7=1	$\beta_2 - \beta_4 + 1/2 (\gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34})$	
	L9=0		

IV	L6=0	$\beta_3 - \beta_4 + 1/3 (\gamma_{13} - \gamma_{14} + \gamma_{23} - \gamma_{24} + \gamma_{33} - \gamma_{34})$	
	L7=0		
	L8=1		

(1) As funções estimáveis do Tipo I e II são iguais.

(2) Assumindo o coeficiente L6=1 e zerando os demais, a função estimável do Tipo III se aproxima as do Tipo I e II

(3) Assumindo o coeficiente L7=1 e zerando os demais, a função estimável do Tipo III é equivalente as do Tipo I e II

TABELA 31. Funções estimáveis sobre a interação.

Funções estimáveis	
$L_{10}=1; L_{11}=0; L_{13}=0; L_{14}=0; L_{15}=0; L_{17}=0; L_{18}=0$	$\gamma_{11} - \gamma_{14} - \gamma_{33} + \gamma_{34} - \gamma_{41} + \gamma_{43}$
$L_{10}=0; L_{11}=1; L_{13}=0; L_{14}=0; L_{15}=0; L_{17}=0; L_{18}=0$	$\gamma_{13} - \gamma_{14} - \gamma_{33} + \gamma_{34}$
$L_{10}=0; L_{11}=0; L_{13}=1; L_{14}=0; L_{15}=0; L_{17}=0; L_{18}=0$	$\gamma_{21} - \gamma_{24} - \gamma_{33} + \gamma_{34} - \gamma_{41} + \gamma_{43}$
$L_{10}=0; L_{11}=0; L_{13}=0; L_{14}=1; L_{15}=0; L_{17}=0; L_{18}=0$	$\gamma_{22} - \gamma_{24} - \gamma_{33} + \gamma_{34} - \gamma_{42} + \gamma_{43}$
$L_{10}=0; L_{11}=0; L_{13}=0; L_{14}=0; L_{15}=1; L_{17}=0; L_{18}=0$	$\gamma_{23} - \gamma_{24} - \gamma_{33} + \gamma_{34}$
$L_{10}=0; L_{11}=0; L_{13}=0; L_{14}=0; L_{15}=0; L_{17}=1; L_{18}=0$	$\gamma_{31} - \gamma_{33} - \gamma_{41} + \gamma_{43}$
$L_{10}=0; L_{11}=0; L_{13}=0; L_{14}=0; L_{15}=0; L_{17}=0; L_{18}=1$	$\gamma_{32} - \gamma_{33} - \gamma_{42} + \gamma_{43}$

4.2.3 Hipóteses e somas de quadrados associadas

4.2.3.1 Hipóteses

As hipóteses no Modelo-S sobre os efeitos principais estão descritas na Tabela 32 para o fator A (cultivares) e na Tabela 33 para o fator B (fases da lua). A Tabela 34 descreve as hipóteses sobre o efeito da interação.

A interpretação das hipóteses apresenta dificuldade. Todos os tipos de hipóteses envolvem os parâmetros referentes à interação, portanto, não se pode testar isoladamente os efeitos principais. Com exceção da hipótese sobre a interação, as demais carregam componentes do outro fator e/ou interação.

A hipótese sobre a interação é sempre ajustada para os demais fatores, portanto, todos os tipos de somas de quadrados se equivalem e são iguais a $R(\gamma|\mu, \alpha, \beta)$, fato que vem confirmar a afirmação de Lemma (1995).

i) Tipo I

Para as hipóteses do Tipo I cada linha da hipótese apresenta um contraste entre níveis do efeito principal em questão.

A hipótese Tipo I para o fator A (cultivares), H_0^1 , é sobre médias ponderadas de linhas. Envolve, além dos parâmetros α , também os parâmetros β e γ , sendo a soma de quadrados Tipo I equivalente a $R(\alpha|\mu)$, não ajustada para o fator B e interação. Tem $a-1 = 4-1 = 3$ graus de liberdade e, conseqüentemente, com três símbolos (L2, L3 e L4) associados à hipótese.

É de difícil interpretação, tornando-se praticamente impossível visualizar o que está sendo testado. Contém coeficientes confusos, pois é uma hipótese sobre médias ponderadas de linhas não ajustadas para colunas, em diferentes freqüências.

A hipótese Tipo I para o fator B (fases da lua), H_0^5 , é sobre médias ponderadas de colunas ajustadas para linhas. Não envolve o parâmetro α , pois o ajuste é seqüencial para as somas de quadrados Tipo I, equivalente a $R(\beta|\mu, \alpha)$, para o modelo na ordem A, B e A-B. Tem $b-1 = 4-1 = 3$ graus de liberdade e, conseqüentemente, três símbolos (L6, L7 e L8) associados à hipótese.

Mesmo não contendo os parâmetros α , envolve funções pouco comuns dos parâmetros γ , tornando-se, assim, de difícil interpretação.

Para a interação entre cultivares e fases da lua, A-B, a hipótese Tipo I é livre dos outros parâmetros. Com sete símbolos L (L10, L11, L13, L14, L15, L17, L18) envolvidos, a hipótese associada tem sete graus de liberdade.

Se não houvesse caselas vazias, haveria $(a-1)(b-1) = 9$ graus de liberdade para a interação. Mas a interação tem $S - (a+b-1) = 14-7 = 7$ graus de liberdade e, conseqüentemente, sete funções linearmente independentes podem ser testadas.

ii) Tipo II

Nas hipóteses do Tipo II não há presença de contraste do outro fator, apenas da interação.

Para o fator A, a hipótese H_0^2 é sobre médias ponderadas de linhas ajustadas para colunas. Em termos de notação $R(\)$, é dada por $R(\alpha|\mu,\beta)$. Mesmo não envolvendo os parâmetros β , é difícil de interpretar, pois envolve combinações lineares complexas dos parâmetros γ .

Devido ao ajuste sequencial das funções estimáveis Tipo I, para o fator B, H_0^5 e H_0^6 , são iguais e ambas equivalem a $R(\beta|\mu,\alpha)$.

As hipóteses para a interação são iguais, pois aparecem sempre ajustadas para os demais efeitos.

iii) Tipo III

A hipótese Tipo III para o fator A, H_0^3 , é sobre médias não ponderadas de linhas. Em termos de notação $R(\)$, se equivalem a $R(\hat{\alpha}|\mu,\hat{\beta},\hat{\gamma})$.

Para o fator B, a hipótese H_0^7 é sobre médias não ponderadas de colunas. Em termos de notação $R(\)$, equivalem a $R(\hat{\beta}|\mu,\hat{\alpha},\hat{\gamma})$.

Ambas trazem coeficientes confusos, de difícil interpretação.

iv) Tipo IV

A hipóteses do Tipo IV envolvem um subconjunto dos dados em que não aparecem caselas vazias, portanto, sendo menos complexas. Não se baseiam na notação $R(\)$.

Para o fator A, a hipótese do Tipo IV, H_0^4 , é sobre médias não ponderadas de linhas, semelhante à do Tipo III, porém, depende da posição das caselas vazias.

Quanto ao fator B, a hipótese do Tipo IV, H_0^8 , é sobre médias não ponderadas de colunas, também semelhante à do Tipo III, dependendo da posição das caselas vazias.

Como ocorrem caselas vazias, há interações que não fazem parte da hipótese, portanto, não se deve usar a hipótese sobre a interação para verificar se o modelo com interação é adequado para o conjunto de dados.

4.2.3.2 Conseqüências da ordenação dos efeitos principais

A alteração da ordem de especificação do modelo, no PROC GLM, na ordem B, A, B-A, é fundamental para a determinação das hipóteses do Tipo I e das somas de quadrados associadas.

A hipótese do Tipo I, para o fator B, é uma hipótese sobre médias não ajustadas, com notação $R(\beta|\mu)$, que, na ordenação A, B, A-B, não aparece em nenhum tipo, enquanto que, para o fator A, é ajustada para B e representada por $R(\alpha|\mu, \beta)$, sendo a mesma hipótese Tipo II dada pela ordenação A, B, A-B. Para as demais hipóteses e para a interação, permanecem inalteradas.

Como também foi observado por Camarinha Filho (1995), Nesi (2002) dentre outros autores, quando se escrevem, em termos do modelo superparametrizado, todas as hipóteses testadas pelos quatro tipos de somas de quadrados, fornecidas pelo PROC GLM do SAS, nota-se que somente aquelas referentes à interação não apresentam algum parâmetro “perturbador” na sua forma. Isto sugere que modelos que envolvem interação são adequados para avaliar a interação, não sendo próprios para avaliar os efeitos principais pois, para esses, ocorrem muitos coeficientes “perturbadores”, levando a formar funções bastante complexas e de difícil compreensão.

TABELA 32. Hipóteses testadas sobre o fator A (cultivares).

Hipóteses	Hipótese testadas
Tipo I	$\alpha_1 - \alpha_4 + 1/3 (-\beta_2 + \beta_4 + \gamma_{11} + \gamma_{13} + \gamma_{14} - \gamma_{41} - \gamma_{42} - \gamma_{43}) = 0$
H_0^1	$\alpha_2 - \alpha_4 + 1/33 (-2\beta_1 - 2\beta_2 + 5\beta_3 + 9\beta_4 + 9\gamma_{21} + 9\gamma_{22} + 6\gamma_{23} - 9\gamma_{24} - 11\gamma_{41} + 11\gamma_{42} - 11\gamma_{43}) = 0$
	$\alpha_3 - \alpha_4 + 1/12 (-\beta_1 - \beta_2 - \beta_3 + 3\beta_4 + 3\gamma_{31} + 3\gamma_{32} + 3\gamma_{33} - 3\gamma_{34} - 4\gamma_{41} + 4\gamma_{42} - 4\gamma_{43}) = 0$
	$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 2\gamma_{43}) = 0$
Tipo II	$\alpha_2 - \alpha_4 + 0,0294\gamma_{11} + 0,0564\gamma_{13} - 0,0858\gamma_{14} + 0,2966\gamma_{21} + 0,3064\gamma_{22} + 0,2157\gamma_{23} + 0,1814\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,0466\gamma_{33} - 0,0956\gamma_{34} - 0,3456\gamma_{41} - 0,3358\gamma_{42} - 0,3186\gamma_{43} = 0$
H_0^2	$\alpha_3 - \alpha_4 + 0,0368\gamma_{11} + 0,0392\gamma_{13} - 0,076\gamma_{14} + 0,027\gamma_{21} + 0,0392\gamma_{22} + 0,0196\gamma_{23} - 0,0858\gamma_{24} + 0,2745\gamma_{31} + 0,2868\gamma_{32} + 0,277\gamma_{33} + 0,1618\gamma_{34} - 0,3382\gamma_{41} - 0,326\gamma_{42} - 0,3358\gamma_{43} = 0$
Tipo III	$\alpha_1 - \alpha_4 + 1/8 (3\gamma_{11} + 3\gamma_{13} + 2\gamma_{14} + \gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34} - 3\gamma_{41} - 2\gamma_{42} - 3\gamma_{43}) = 0$
H_0^3	$\alpha_2 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 22\gamma_{21} + 23\gamma_{22} + 22\gamma_{23} + 13\gamma_{24} + 2\gamma_{31} + 3\gamma_{32} + 2\gamma_{33} - 7\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43}) = 0$
	$\alpha_3 - \alpha_4 + 1/80 (3\gamma_{11} + 3\gamma_{13} - 6\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 7\gamma_{24} + 22\gamma_{31} + 23\gamma_{32} + 22\gamma_{33} + 13\gamma_{34} - 27\gamma_{41} - 26\gamma_{42} - 27\gamma_{43}) = 0$
Tipo IV	$\alpha_1 - \alpha_4 + 1/2 (\gamma_{11} + \gamma_{13} - \gamma_{41} - \gamma_{43}) = 0$
H_0^4	$\alpha_2 - \alpha_4 + 1/3 (\gamma_{21} + \gamma_{22} + \gamma_{23} - \gamma_{41} - \gamma_{42} - \gamma_{43}) = 0$
	$\alpha_3 - \alpha_4 + 1/3 (\gamma_{31} + \gamma_{32} + \gamma_{33} - \gamma_{41} - \gamma_{42} - \gamma_{43}) = 0$

TABELA 33. Hipóteses testadas sobre o fator B (fases da lua).

Hipóteses	Hipótese testadas
Tipo I e II ⁽¹⁾ H_0^5 e H_0^6	$\beta_1 - \beta_4 + 0,2868\gamma_{11} + 0,0392\gamma_{13} - 0,326\gamma_{14} + 0,277\gamma_{21} + 0,0392\gamma_{22} + 0,0196\gamma_{23} - 0,3358\gamma_{24} + 0,2745\gamma_{31} + 0,0368\gamma_{32} + 0,027\gamma_{33} - 0,3382\gamma_{34} + 0,1618\gamma_{41} - 0,076\gamma_{42} - 0,0858\gamma_{43} = 0$
	$\beta_2 - \beta_4 + 1/8 (\gamma_{11} + \gamma_{13} - 2\gamma_{14} + 3\gamma_{22} - 3\gamma_{24} + 3\gamma_{32} - 3\gamma_{34} - \gamma_{41} + 2\gamma_{42} - \gamma_{43}) = 0$
	$\beta_3 - \beta_4 + 0,0294\gamma_{11} + 0,3064\gamma_{13} - 0,3358\gamma_{14} + 0,0466\gamma_{21} + 0,0564\gamma_{22} + 0,2157\gamma_{23} - 0,3186\gamma_{24} + 0,0196\gamma_{31} + 0,0294\gamma_{32} + 0,2966\gamma_{33} - 0,3456\gamma_{34} - 0,0956\gamma_{41} - 0,0858\gamma_{42} + 0,1814\gamma_{43} = 0$
Tipo III H_0^7	$\beta_1 - \beta_4 + 1/80 (23\gamma_{11} + 3\gamma_{13} - 26\gamma_{14} + 22\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} - 27\gamma_{24} + 22\gamma_{31} + 3\gamma_{32} + 2\gamma_{33} - 27\gamma_{34} + 13\gamma_{41} - 6\gamma_{42} - 7\gamma_{43}) = 0$
	$\beta_2 - \beta_4 + 1/8 (\gamma_{11} + \gamma_{13} - 2\gamma_{14} + 3\gamma_{22} - 3\gamma_{24} + 3\gamma_{32} - 3\gamma_{34} - \gamma_{41} + 2\gamma_{42} - \gamma_{43}) = 0$
	$\beta_3 - \beta_4 + 1/80 (3\gamma_{11} + 23\gamma_{13} - 26\gamma_{14} + 2\gamma_{21} + 3\gamma_{22} + 22\gamma_{23} - 27\gamma_{24} + 2\gamma_{31} + 3\gamma_{32} + 22\gamma_{33} - 27\gamma_{34} - 7\gamma_{41} - 6\gamma_{42} + 13\gamma_{43}) = 0$
Tipo IV H_0^8	$\beta_1 - \beta_4 + 1/3 (\gamma_{11} - \gamma_{14} + \gamma_{21} - \gamma_{24} + \gamma_{31} - \gamma_{34}) = 0$
	$\beta_2 - \beta_4 + 1/2 (\gamma_{22} - \gamma_{24} + \gamma_{32} - \gamma_{34}) = 0$
	$\beta_3 - \beta_4 + 1/3 (\gamma_{13} - \gamma_{14} + \gamma_{23} - \gamma_{24} + \gamma_{33} - \gamma_{34}) = 0$

⁽¹⁾ As hipóteses estimáveis, H_0^5 e H_0^6 , correspondente aos Tipos I e II são equivalentes.

TABELA 34. Hipóteses testadas sobre a interação.

Hipóteses	Hipótese testadas
H_0^9	$\gamma_{11} - \gamma_{14} - \gamma_{33} + \gamma_{34} - \gamma_{41} + \gamma_{43} = 0$
	$\gamma_{13} - \gamma_{14} - \gamma_{33} + \gamma_{34} = 0$
	$\gamma_{21} - \gamma_{24} - \gamma_{33} + \gamma_{34} - \gamma_{41} + \gamma_{43} = 0$
	$\gamma_{22} - \gamma_{24} - \gamma_{33} + \gamma_{34} - \gamma_{42} + \gamma_{43} = 0$
	$\gamma_{23} - \gamma_{24} - \gamma_{33} + \gamma_{34} = 0$
	$\gamma_{31} - \gamma_{33} - \gamma_{41} + \gamma_{43} = 0$
	$\gamma_{32} - \gamma_{33} - \gamma_{42} + \gamma_{43} = 0$
	$\gamma_{11} - \gamma_{14} - \gamma_{33} + \gamma_{34} - \gamma_{41} + \gamma_{43} = 0$
	$\gamma_{13} - \gamma_{14} - \gamma_{33} + \gamma_{34} = 0$

4.2.3.3 Somas de quadrados

Ha cinco formas de apresentar as somas de quadrados para o conjunto de dados da Tabela 4. Observa-se, na Tabela 35, que tem-se um valor para a soma de quadrados do Tipo I para os fatores A e B na ordem A-B e outro na ordem B-A; e nas Tabelas 36, 37 e 38, as somas de quadrados Tipo II, III e IV se equivalem nas duas ordenações. Resultados concordantes foram obtidos por Nekatschalow (1997), em um fatorial 3 x 4 com duas caselas vazias.

A ordem de entrada dos fatores é fundamental para a análise. Portanto, as somas de quadrados das Tabelas 35 e 36, para o fator B, são as mesmas, quando a ordenação é A-B e, para o fator A, quando a ordenação é B-A, devido à hipótese associada ser do Tipo II, fato que sempre ocorre. Isso confirma a citação de Camarinha Filho (1995), de que a segunda linha das tabelas de análise de variância para somas de quadrados do Tipo I e II, testa a mesma hipótese para o segundo fator na ordenação do modelo.

Em se tratando de um modelo com interação com dados desbalanceados e caselas vazias, as quatro somas de quadrados fornecidas pelo PROC GLM do SAS diferem entre si e as somas de quadrados Tipo I dependem da ordem de entrada dos parâmetros no modelo, dado que são obtidas seqüencialmente. A Tabela 35 mostra que $R(\alpha|\mu) \neq R(\alpha|\mu,\beta)$ e $R(\beta|\mu) \neq R(\beta|\mu,\alpha)$. Portanto, se a ordem de entrada dos parâmetros é A, B, A-B, ou seja, o modelo é $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$, as somas de quadrados são obtidas primeiramente supondo-se o modelo $y_{ijk} = \mu + \alpha_i + e_{ijk}$, após o modelo $y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$, e, afinal, o modelo completo, gerando respectivamente $R(\alpha|\mu) = 1542,597$, $R(\beta|\mu,\alpha) = 1270,737$ e $R(\gamma|\mu,\alpha,\beta) = 175,559$. Semelhantemente, se a ordem de entrada é B, A, B-A, são obtidas as somas de quadrados por meio dos modelos $y_{ijk} = \mu + \beta_j + e_{ijk}$, $y_{ijk} = \mu + \beta_j + \alpha_i + e_{ijk}$ e, por fim, o modelo completo $y_{ijk} = \mu + \beta_j + \alpha_i + \gamma_{ij} + e_{ijk}$, obtendo-se, respectivamente, $R(\beta|\mu) = 1424,937$,

$R(\alpha|\mu,\beta) = 1388,370$ e $R(\gamma|\mu,\alpha,\beta) = 175,559$. Para as demais somas de quadrados, a ordem dos fatores no modelo não altera o resultado. Resultados semelhantes também foram obtidos por Lemma (1995), que trabalhou com um fatorial 2×3 , com dados desbalanceados e uma casela vazia.

Para a soma de quadrados associada à hipótese do Tipo II (Tabela 36), para o fator A, têm-se $R(\alpha|\mu,\beta) = 1388,370$, e para o fator B, $R(\beta|\mu,\alpha) = 1270,737$; do Tipo III (Tabela 37) para o fator A mostra que $R(\hat{\alpha}|\hat{\mu},\hat{\beta},\hat{\gamma}) = 1379,424$ e para o fator B, $R(\hat{\beta}|\hat{\mu},\hat{\alpha},\hat{\gamma}) = 1264,779$, e do Tipo IV (Tabela 38) para o fator A a soma de quadrados é 1162,994 e para o fator B a soma de quadrados é 880,698.

A soma de quadrados para a interação é invariante, $R(\gamma|\mu,\alpha,\beta) = R(\gamma|\hat{\mu},\hat{\alpha},\hat{\beta}) = 2488,430$, confirmando o postulado citado por Lemma (1995) de que “modelo com interação é bom para testar interação”, e que o teste da interação (H_0^9) é bem mais geral e que é testável se não há caselas vazias.

A soma de quadrados Tipo IV não é única, o que causa uma maior dificuldade de interpretação, pois, com uma reordenação dos dados, há possibilidade de obter outras funções estimáveis do Tipo IV, induzindo a outras somas de quadrados. Ocorre uma nota a respeito na saída do PROC GLM.

TABELA 35. Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo I e respectivas hipóteses associadas.

Ordem A-B					Ordem B-A				
C.V.	G.L	H_0	R()	SS	C.V.	G.L	H_0	R()	
A	3	H_0^1	$R(\alpha \mu)$	1 542,597	B	3	H_0^5	$R(\beta \mu)$	1 424,937
B (aj)	3	H_0^6	$R(\beta \mu,\alpha)$	1 270,737	A (aj)	3	H_0^2	$R(\alpha \mu,\beta)$	1 388,370
A-B	7	H_0^9	$R(\gamma \mu,\alpha,\beta)$	175,559	B-A	7	H_0^9	$R(\gamma \mu,\alpha,\beta)$	175,559
Resíduo	27			2 488,830	Resíduo	27			2 488,830

84

TABELA 36. Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo II e respectivas hipóteses associadas.

Ordem A-B					Ordem B-A				
C.V.	G.L	H_0	R()	SS	C.V.	G.L	H_0	R()	
A (aj)	3	H_0^2	$R(\alpha \mu,\beta)$	1 388,370	B (aj)	3	H_0^6	$R(\beta \mu,\alpha)$	1 270,737
B (aj)	3	H_0^6	$R(\beta \mu,\alpha)$	1 270,737	A (aj)	3	H_0^2	$R(\alpha \mu,\beta)$	1 388,370
A-B	7	H_0^9	$R(\gamma \mu,\alpha,\beta)$	175,559	B-A	7	H_0^9	$R(\gamma \mu,\alpha,\beta)$	175,559
Resíduo	27			2 488,830	Resíduo	27			2 488,830

TABELA 37. Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo III e respectivas hipóteses associadas.

Ordem A-B					Ordem B-A				
C.V.	G.L	H_0	R()	SS	C.V.	G.L	H_0	R()	SS
A	3	H_0^3	$R(\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma})$	1 379,424	B	3	H_0^7	$R(\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma})$	1 264,799
B	3	H_0^7	$R(\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma})$	1 264,799	A	3	H_0^3	$R(\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma})$	1 379,424
A-B	7	H_0^9	$R(\hat{\gamma} \hat{\mu},\hat{\alpha},\hat{\beta})$	175,559	B-A	7	H_0^9	$R(\hat{\gamma} \hat{\mu},\hat{\alpha},\hat{\beta})$	175,559
Resíduo	27			2 488,830	Resíduo	27			2 488,830

TABELA 38. Análise de variância dos dados da Tabela 4, por meio do modelo com dois fatores e interação, expondo somas de quadrados Tipo IV e respectivas hipóteses associadas.

Ordem A-B					Ordem B-A				
C.V.	G.L	H_0	R()	SS	C.V.	G.L	H_0	R()	SS
A	3 ⁽¹⁾	H_0^4	—	1 162,994	B	3 ⁽¹⁾	H_0^8	—	880,698
B	3 ⁽¹⁾	H_0^8	—	880,698	A	3 ⁽¹⁾	H_0^4	—	1 162,994
A-B	7	H_0^9	—	175,559	B-A	7	H_0^9	—	175,559
Resíduo	27		—	2 488,830	Resíduo	27		—	2 488,830

⁽¹⁾ Existem outras hipóteses testáveis Tipo IV que produzem diferentes somas de quadrados.

4.3 3º caso: fatorial 4 x 5 x 4 desbalanceado

Têm-se 68 símbolos (L1, L2, L3, L4, ...) envolvidos, indicando que o posto máximo da hipótese tem 68 graus de liberdade.

4.3.1. Estimação

As funções estimáveis para os efeitos principais e para as interações são formadas de acordo com a(s) combinação(ões) entre os fatores que apresentam frequências de caselas proporcionais. Para a ordenação A-B-C, segundo o modelo $y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varphi_{ij} + \eta_{ik} + \tau_{jk} + \rho_{ijk} + e_{ijkl}$, a forma das funções apresentou as características a seguir.

Dentre os efeitos principais, as funções estimáveis para o fator A, cultivares de cenoura, envolvem parâmetros do próprio fator (α) e parâmetros das interações nas quais o fator ocorre. Fato semelhante ocorre com o fator C (γ), fases da lua. Entretanto, para o fator B (β), meses do ano, têm-se funções do Tipo I com os seguintes parâmetros: β , γ , φ , η , τ e ρ , que são parâmetros correspondentes ao próprio fator, ao fator C e a todas as interações, A-B, A-C, B-C e A-B-C, respectivamente.

Entre as interações de segunda ordem, a interação A-B apresenta funções estimáveis do Tipo II iguais às do Tipo III, apresentando parâmetros da própria interação, (φ), parâmetros da interação A-C, (η) e parâmetros da interação de maior grau, (ρ). A interação A-C mostra funções estimáveis do Tipo I = Tipo II = Tipo III, com parâmetros da própria interação A-C, (η) e parâmetros da interação tripla, (ρ). A interação B-C apresenta igualdade entre os quatro tipos de funções estimáveis, com parâmetros da própria interação (τ) e parâmetros da interação de maior grau (ρ).

4.3.2 Hipóteses e somas de quadrados associadas

Nas Tabelas 39, 40 e 41 encontram-se os resultados das análises de variância sobre o modelo, para as ordenações A-B-C, B-C-A e C-B-A, respectivamente.

Constata-se que, entre os efeitos principais, o fator A apresenta igualdade entre somas de quadrados associadas às hipóteses dos Tipos I, II e IV, sendo iguais às seguintes decomposições: $R(\alpha|\mu) = R(\alpha|\mu, \beta, \gamma) = 13301300$; o fator B apresenta soma de quadrados iguais para as hipóteses dos Tipos II e III, com $R(\beta|\mu, \alpha, \gamma) = R(\hat{\beta}|\hat{\mu}, \hat{\alpha}, \hat{\gamma}) = 15056,509$.

Entretanto, o fator C mostra igualdade entre as somas de quadrados associadas às hipóteses dos Tipos I, II e III, sendo iguais as decomposições: $R(\gamma|\mu, \beta) = R(\gamma|\mu, \alpha, \beta) = R(\hat{\gamma}|\hat{\mu}, \hat{\alpha}, \hat{\beta}) = 1169,525$, quando não foi declarado em primeiro lugar (Tabelas 39 e 40). A Tabela 41 fornece a análise de variância para o modelo na ordenação C-B-A, mostrando a desigualdade entre as somas de quadrados associadas às hipóteses Tipos I e II, sendo $R(\gamma|\mu) \neq R(\gamma|\mu, \alpha, \beta)$. Isso confirma a citação de Santana & Iemma (2002), de que a ordem de entrada dos parâmetros no modelo é importante para que o SAS dê uma soma de quadrados associada ao Tipo I, em que se tem uma soma de quadrados ajustada apenas para a constante μ , ou seja, $R(\gamma|\mu)$.

Pode ser observado que a interação A-C apresenta somas de quadrados iguais para as hipóteses dos Tipos I, II e III, ou seja, $R(\eta|\mu, \alpha, \beta, \gamma, \varphi) = R(\eta|\mu, \alpha, \beta, \gamma, \varphi, \tau) = R(\hat{\eta}|\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\varphi}, \hat{\tau}) = 413,222$. Entretanto, tal igualdade não ocorre com o Tipo I, quando a ordenação foi C-B-A. A interação A-B apresenta somas de quadrados iguais para os Tipos II e III, mostrando que $R(\phi|\mu, \alpha, \beta, \gamma, \eta, \tau) = R(\hat{\phi}|\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\eta}, \hat{\tau}) = 2757,251$; entretanto, na ordenação

C-B-A, a igualdade também se verificou para o Tipo I, novamente confirmando a importância da entrada dos parâmetros no modelo. A interação B-C apresenta somas de quadrados iguais para os quatro tipos, sendo $R(\tau|\mu,\alpha,\beta,\gamma) = R(\tau|\mu,\alpha,\beta,\gamma,\varphi,\eta) = R(\dot{\tau}|\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma},\dot{\varphi},\dot{\eta}) = 4386,327$, independente da ordenação no modelo. De modo análogo, para a interação tripla A-B-C, que apresenta todas somas de quadrados equivalentes, $R(\rho|\mu,\alpha,\beta,\gamma,\varphi,\eta,\tau) = R(\dot{\rho}|\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma},\dot{\varphi},\dot{\eta},\dot{\tau}) = 1344,267$.

As somas de quadrados associadas às hipóteses do Tipo III e IV para os efeitos principais e para as interações foram diferentes, com exceção da interação B-C, que é a única que apresenta caselas vazias. Para dados desbalanceados, sem caselas vazias, de acordo com Speed et al. (1978), Searle (1987) e Iemma (1995), as hipóteses dos Tipos III e IV são iguais.

Os resultados do presente estudo, portanto, contestam em parte estes autores. Observando-se o número de observações em cada casela, de acordo com a Tabela 42, a interação B-C (Tabela 42.2) é a única que apresenta o mesmo número, fato que pode ter compensado a falta das três caselas. Para as interações A-B e A-C (Tabelas 42.1 e 42.3), mesmo não faltando nenhuma casela, elas possuem número diferente de observações entre as caselas e, como apresentaram resultados diferentes entre os tipos de somas de quadrados, este passa a ser um fator que deve ser olhado com muito cuidado pelo pesquisador.

TABELA 39. Análise de variância para os dados da Tabela 5, com base no modelo $y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \varphi_{ij} + \eta_{ik} + \tau_{jk} + \rho_{ijk} + e_{ijkl}$.

C.V.	G.L.	R()	SS Tipo I	F	P > F
A	3	$R(\alpha \mu)$	13 301,300	81,60	0,0001*
B	4	$R(\beta \mu,\alpha)$	17 172,742	79,02	0,0001*
C	3	$R(\gamma \mu,\alpha,\beta)$	1 169,525	7,17	0,0002*
A-B	12	$R(\varphi \mu,\alpha,\beta,\gamma)$	2 815,043	4,32	0,0001*
A-C	9	$R(\eta \mu,\alpha,\beta,\gamma,\varphi)$	413,222	0,85	0,5761
B-C	9	$R(\tau \mu,\alpha,\beta,\gamma,\varphi,\eta)$	4 386,327	8,97	0,0001*
A-B-C	27	$R(\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau)$	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo II	F	P > F
A	3	$R(\alpha \mu,\beta,\gamma)$	13 301,300	81,60	0,0001*
B	4	$R(\beta \mu,\alpha,\gamma)$	15 056,509	69,28	0,0001*
C	3	$R(\gamma \mu,\alpha,\beta)$	1 169,525	7,17	0,0002*
A-B	12	$R(\varphi \mu,\alpha,\beta,\gamma,\eta,\tau)$	2 757,251	4,23	0,0001*
A-C	9	$R(\eta \mu,\alpha,\beta,\gamma,\varphi,\tau)$	413,222	0,85	0,5761
B-C	9	$R(\tau \mu,\alpha,\beta,\gamma,\varphi,\eta)$	4 386,327	8,97	0,0001*
A-B-C	27	$R(\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau)$	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo III	F	P > F
A	3	$R(\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma})$	12 945,459	79,42	0,0001*
B	4	$R(\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma})$	15 056,509	69,28	0,0001*
C	3	$R(\hat{\gamma} \hat{\mu},\hat{\alpha},\hat{\beta})$	1 169,525	7,17	0,0002*
A-B	12	$R(\hat{\varphi} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\eta},\hat{\tau})$	2 757,251	4,23	0,0001*
A-C	9	$R(\hat{\eta} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\tau})$	413,222	0,85	0,5761
B-C	9	$R(\hat{\tau} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\eta})$	4 386,327	8,97	0,0001*
A-B-C	27	$R(\hat{\rho} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\tau},\hat{\eta},\hat{\varphi})$	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo IV	F	P > F
A	3	—	13 301,300	81,60	0,0001*
B	4 ⁽¹⁾	—	12 533,793	57,67	0,0001*
C	3 ⁽¹⁾	—	2 015,852	12,37	0,0001*
A-B	12 ⁽¹⁾	—	2 618,421	4,02	0,0001*
A-C	9 ⁽¹⁾	—	279,013	0,57	0,8193
B-C	9	—	4 386,327	8,97	0,0001*
A-B-C	27	—	1 344,267	0,92	0,5881
Resíduo	136		7 389,358		

⁽¹⁾ Existem outras hipóteses testáveis Tipo IV que produzem diferentes somas de quadrados.

*: significativo a 1% pelo teste F.

TABELA 40. Análise de variância dos dados da Tabela 5, com base no modelo

$$y_{ijkl} = \mu + \beta_j + \gamma_k + \alpha_i + \tau_{jk} + \varphi_{ji} + \eta_{ki} + \rho_{jki} + e_{ijkl}.$$

C.V.	G.L.	R()	SS Tipo I	F	P > F
B	4	R($\beta \mu$)	17 172,742	79,02	0,0001*
C	3	R($\gamma \mu,\beta$)	1 169,525	7,17	0,0002*
A	3	R($\alpha \mu,\beta,\gamma$)	13 301,300	81,60	0,0001*
B-C	9	R($\tau \mu,\alpha,\beta,\gamma$)	4 386,327	8,97	0,0001*
B-A	12	R($\varphi \mu,\alpha,\beta,\gamma,\tau$)	2 815,043	4,32	0,0001*
C-A	9	R($\eta \mu,\alpha,\beta,\gamma,\varphi,\tau$)	413,222	0,85	0,5761
B-C-A	27	R($\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo II	F	P > F
B	4	R($\beta \mu,\alpha,\gamma$)	15 056,509	69,28	0,0001*
C	3	R($\gamma \mu,\alpha,\beta$)	1 169,525	7,17	0,0002*
A	3	R($\alpha \mu,\beta,\gamma$)	13 301,300	81,60	0,0001*
B-C	9	R($\tau \mu,\alpha,\beta,\gamma,\varphi,\eta$)	4 386,327	8,97	0,0001*
B-A	12	R($\varphi \mu,\alpha,\beta,\gamma,\eta,\tau$)	2 757,251	4,23	0,0001*
C-A	9	R($\eta \mu,\alpha,\beta,\gamma,\varphi,\tau$)	413,222	0,85	0,5761
B-C-A	27	R($\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo III	F	P > F
B	4	R($\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma}$)	15 056,509	69,28	0,0001*
C	3	R($\hat{\gamma} \hat{\mu},\hat{\alpha},\hat{\beta}$)	1 169,525	7,17	0,0002*
A	3	R($\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma}$)	12 945,459	79,42	0,0001*
B-C	9	R($\hat{\tau} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\eta}$)	4 386,327	8,97	0,0001*
B-A	12	R($\hat{\varphi} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\eta},\hat{\tau}$)	2 757,251	4,23	0,0001*
C-A	9	R($\hat{\eta} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\tau}$)	413,222	0,85	0,5761
B-C-A	27	R($\hat{\rho} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\eta},\hat{\tau}$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo IV	F	P > F
B	4 ⁽¹⁾	—	12 533,793	57,67	0,0001*
C	3 ⁽¹⁾	—	2 015,852	12,37	0,0001*
A	3	—	13 301,300	81,60	0,0001*
B-C	9	—	4 386,327	8,97	0,0001*
B-A	12 ⁽¹⁾	—	2 618,421	4,02	0,0001*
C-A	9 ⁽¹⁾	—	279,013	0,57	0,8193
B-C-A	27	—	1 344,267	0,92	0,5881
Resíduo	136		7 389,358		

⁽¹⁾ Existem outras hipóteses testáveis Tipo IV que produzem diferentes somas de quadrados.

*: significativo a 1% pelo teste F

TABELA 41. Análise de variância dos dados da Tabela 5, com base no modelo

$$y_{kijl} = \mu + \gamma_k + \beta_j + \alpha_i + \tau_{kj} + \eta_{ik} + \varphi_{ji} + \rho_{kji} + e_{kijl}.$$

C.V.	G.L.	R()	SS Tipo I	F	P > F
C	3	R($\gamma \mu$)	3 285,758	20,16	0,0001*
B	4	R($\beta \mu,\gamma$)	15 056,509	69,28	0,0001*
A	3	R($\alpha \mu,\beta,\gamma$)	13 301,300	81,60	0,0001*
C-B	9	R($\tau \mu,\alpha,\beta,\gamma$)	4 386,327	8,97	0,0001*
C-A	9	R($\eta \mu,\alpha,\beta,\gamma,\tau$)	471,015	0,96	0,4733
B-A	12	R($\varphi \mu,\alpha,\beta,\gamma,\tau,\eta$)	2 757,251	4,23	0,0001*
C-B-A	27	R($\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo II	F	P > F
C	3	R($\beta \mu,\alpha,\gamma$)	1 169,525	7,17	0,0002*
B	4	R($\gamma \mu,\alpha,\beta$)	15 056,509	69,28	0,0001*
A	3	R($\alpha \mu,\beta,\gamma$)	13 301,300	81,60	0,0001*
C-B	9	R($\tau \mu,\alpha,\beta,\gamma,\varphi,\eta$)	4 386,327	8,97	0,0001*
C-A	9	R($\eta \mu,\alpha,\beta,\gamma,\varphi,\tau$)	413,222	0,85	0,5761
B-A	12	R($\varphi \mu,\alpha,\beta,\gamma,\eta,\tau$)	2 757,251	4,23	0,0001*
C-B-A	27	R($\rho \mu,\alpha,\beta,\gamma,\varphi,\eta,\tau$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo III	F	P > F
C	3	R($\hat{\gamma} \hat{\mu},\hat{\alpha},\hat{\beta}$)	1 169,525	7,17	0,0002*
B	4	R($\hat{\beta} \hat{\mu},\hat{\alpha},\hat{\gamma}$)	15 056,509	69,28	0,0001*
A	3	R($\hat{\alpha} \hat{\mu},\hat{\beta},\hat{\gamma}$)	12 945,459	79,42	0,0001*
C-B	9	R($\hat{\tau} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\eta}$)	4 386,327	8,97	0,0001*
C-A	9	R($\hat{\eta} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\tau}$)	413,222	0,85	0,5761
B-A	12	R($\hat{\varphi} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\eta},\hat{\tau}$)	2 757,251	4,23	0,0001*
C-B-A	27	R($\hat{\rho} \hat{\mu},\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\varphi},\hat{\eta},\hat{\tau}$)	1 344,267	0,92	0,5881

C.V.	G.L.	R()	SS Tipo IV	F	P > F
C	3 ⁽¹⁾	—	2 015,852	12,37	0,0001*
B	4 ⁽¹⁾	—	12 533,793	57,67	0,0001*
A	3	—	13 301,300	81,60	0,0001*
C-B	9	—	4 386,327	8,97	0,0001*
C-A	9 ⁽¹⁾	—	279,013	0,57	0,8193
B-A	12 ⁽¹⁾	—	2 618,421	4,02	0,0001*
C-B-A	27	—	1 344,267	0,92	0,5881
Resíduo	136		7 389,358		

⁽¹⁾ Existem outras hipóteses testáveis Tipo IV que produzem diferentes somas de quadrados.

*: significativo a 1% pelo teste F.

TABELA 42 - Número de observações por casela, de acordo com as interações duplas, B-A, B-C e A-C.

42.1. Interação B-A					42.2. Interação B-C				
B/A	Nan	Kur	Car	Bra	B/C	No	Cr	Ch	Mi
Mar	9	9	9	9	Mar	12	12	12	-
Abr	6	6	6	6	Abr	-	-	12	12
Mai	12	12	12	12	Mai	12	12	12	12
Jun	12	12	12	12	Jun	12	12	12	12
Jul	12	12	12	12	Jul	12	12	12	12

42.3. Interação A-C				
A/C	No	Cr	Ch	Mi
Nan	12	12	15	12
Kur	12	12	15	12
Car	12	12	15	12
Bra	12	12	15	12

4.3.3 Comparações entre as médias

As análises de variância dos dados originais (Tabelas 39, 40 e 41) do 3º caso (Tabela 5), sem supressão de observações, mostraram que os fatores principais, A (cultivares), (B) meses e (C) diferentes fases da lua, afetaram de modo significativo à produção comercial de raízes de cenoura. Portanto, as produções dos diferentes níveis dos fatores A, B e C não são estatisticamente iguais.

As interações de segunda ordem, A-B e B-C, apresentaram-se significativas ($P < 0,01$), sugerindo a existência de uma dependência entre os fatores. As médias de quadrados mínimos estimadas pelo procedimento *Lsmeans* do SAS, foram comparadas pelo teste de Tukey (5%).

Pelos dados da Tabela 43, observa-se que a cultivar Brasília apresentou maior produção comercial média de raízes de cenoura, em t/ha, quando os plantios foram realizados nos meses de maio, junho e julho.

TABELA 43. Valores médios estimados da produção comercial de raízes de cenoura, em t/ha, de acordo com as cultivares e épocas (meses) de plantio.

	'Nantes'	'Kuronan'	'Carandaí'	'Brasília'
Março	-	-	-	-
Abril	-	-	-	-
Maio	18,68 bB	33,52 aA	30,51 bA	36,70 bA
Junho	32,91 aB	40,28 aB	41,08 aB	52,34 aA
Julho	26,36 aC	38,70 aB	35,00 abB	47,18 aA

Nota: Médias seguidas pela mesma letra minúscula (coluna), ou pela mesma letra maiúscula (linha) não diferem entre si, pelo teste de Tukey, a 5% de probabilidade.
 - : médias não estimadas.

TABELA 44. Valores médios estimados da produção comercial de cenoura, em t/ha, de acordo com os plantios realizados nos diversos meses e nas diferentes fases da lua.

	Nova	Crescente	Cheia	Minguante
Março	54,88 aA	57,86 aA	56,28 aA	-
Abril	-	-	22,96 cB	37,90 aA
Maio	33,88 cA	25,63 cB	36,93 bA	22,96 bB
Junho	42,12 bAB	38,42 bB	49,87 aA	36,19 aB
Julho	44,67 bA	30,94 bcB	37,95 bAB	33,68 aB

Nota: Médias seguidas pela mesma letra minúscula (coluna), ou pela mesma letra maiúscula (linha) não diferem entre si, pelo teste de Tukey, a 5% de probabilidade.
 - : médias não estimadas

Quanto à interação meses versus fases da lua, verifica-se, pela Tabela 44, que o mês de plantio que apresentou maiores rendimentos foi março, para as fases da lua nova, crescente e cheia, que apresentaram produção semelhantes, variando de 54,9t/ha a 57,9t/ha. Já para os plantios realizados em maio, junho e julho, as melhores fases da lua para plantio das cultivares de cenoura foram a nova e a cheia.

5 CONCLUSÕES

i) Experimentos com dois e três fatores podem ser estudados utilizando-se o modelo superparametrizado.

ii) Conjuntos de funções estimáveis e conseqüentes hipóteses associadas podem ser obtidas usando-se o procedimento GLM do sistema SAS na análise de experimentos com dois e três fatores.

iii) Quando ocorrem perdas de parcelas, a identificação das funções estimáveis é mais complexa e as hipóteses ficam de difícil interpretação.

iv) Quando os dados são desbalanceados:

- as funções estimáveis de um fator envolvem os parâmetros relativos ao fator e os componentes das interações nas quais o fator está presente;
- as somas de quadrados dos Tipos I e II são iguais e as dos Tipos III e IV são iguais;
- a ordenação dos fatores principais não afeta as hipóteses do Tipo I.

v) Quando os dados são desbalanceados e com caselas vazias:

- os quatro tipos de somas de quadrados para o fator principal de entrada foram diferentes;
- a ordenação é fundamental para obtenção das hipóteses do Tipo I.

vi) Um pesquisador, ao analisar dados desbalanceados, em modelos com interação e na presença de caselas vazias, deve ter o cuidado de verificar as hipóteses a serem testadas, optando por um planejamento mais adequado aos seus propósitos.

REFERÊNCIAS BIBLIOGRÁFICAS

CAMARINHA FILHO, J. A. **Testes de hipóteses em modelos lineares com dados desbalanceados e caselas vazias**. 1995. 142p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

DEMARQUI, F. N.; OIKAWA, S. M. Somas de quadrados e hipóteses estatísticas associadas nos modelos com covariáveis para dados desbalanceados fornecidas pela sistema SAS. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 48.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 10., 2003, Lavras. Anais... Lavras: UFLA, 2003. p. 633-638.

FRANCIS, I. A comparison of several analysis of variance programs. *Journal of the American Statistical Association*, Alexandria, v. 68, n. 34, p. 860-865, 1973.

FREUND, R. J. The case of the missing cell. *The American Statistician*, Alexandria, v. 34, n. 2, p. 94-98, May 1980.

HARVEY, W. R. **Least-squares analysis of data with unequal subclass numbers**. Beltsville, MD: U. S. Department of Agriculture, 1960. (Agricultural Research Service Bulletin ARS-20-8).

HERR, D. G. On the history of ANOVA in unbalanced, factorial designs: the first 30 years. *The American Statistician*, Alexandria, v. 40, n. 4, p. 265-270, Nov. 1986.

HOCKING, R. R.; SPEED, F. M. A full rank analysis of some linear model problems. *Journal of the American Statistical Association*, Alexandria, v. 70, n. 351, p. 706-712, Sept. 1975.

IEMMA, A. F. **Análisis de varianza de experimentos con celdas vacías**. Córdoba, Argentina: Publicações, 1993. 102 p.

IEMMA, A. F. **Matrizes para estatística: um texto para profissionais de ciências aplicadas**. Piracicaba. ESALQ/USP, 1988. 339 p.

IEMMA, A. F. **Modelos lineares: uma introdução para profissionais da pesquisa agropecuária**. 2.ed.Londrina: Imprensa Oficial do Estado do Paraná, 1987. 263 p.

IEMMA, A. F. Que hipóteses estatísticas testamos através do SAS em presença de caselas vazias? *Scientia Agrícola*, Piracicaba, v. 52, n. 2, p. 210-220, maio/jun. 1995.

IEMMA, A. F.; PIERRI, S. H. V. **Ajuste de modelos fixos desbalanceados a través do sistema estatístico SAS**. Piracicaba: ESALQ. Departamento de Matemática e Estatística, 1997. 99 p.

JENNINGS, E.; WARD JR, J. H. Hypothesis identification in the case of the missing cell. *The American Statistician*, Alexandria, v. 36, n. 1, p. 25-27, Feb. 1982.

KRAMER, C. Y. On the analysis of variance of a two-way classification with unequal sub-class numbers. *Biometrics*, Washington, v. 11, n. 4, p. 441-452, Dec. 1955.

MONDARDO, M. **Estimabilidade de funções paramétricas com dados desbalanceados através do PROC GLM do SAS: aplicações à pesquisa agropecuária**. 1994. 166 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP.

NEKATSCHALOW, M. C. **Análise de variância: alternativas através de modelos de posto completo**. 1997. 123 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP.

NESI, C. N. **Métodos alternativos para realização de testes de hipóteses em delineamentos experimentais**. 2002. 123 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP.

OVERALL, J. E.; SPIEGEL, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin*, Washington, v. 72, n. 5, p. 311-322, 1969.

RAO, C. R. On the linear combination of observations and the general theory of least squares. *Sankhyā*, Calcutta, v. 7, p. 237-256, 1945.

SANTANA, D. G. **Delineamentos ortogonais e parcialmente ortogonais: teoria e aplicação.** 2000. 135 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP.

SANTANA, D. G.; IEMMA, A. F. Funções estimáveis e hipóteses testáveis nos delineamentos ortogonais e parcialmente ortogonais com três fatores. *Scientia Agrícola*, Piracicaba, v. 59, n. 2, p. 295-302, abr./jun. 2002.

SANTOS, E. S. **Testes de hipóteses com dados desbalanceados e interpretação de *softwares* mais utilizados.** 1994. 230 p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba.

SAS INSTITUTE. **User's guide: statistics version 6 edition**, Cary: Statistical Analysis System Institute, 1990. 846 p.

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

SEARLE, S. R. **Linear models for unbalanced data.** New York: John Wiley, 1987, 536 p.

SEARLE, S. R.; SPEED, F. M.; HENDERSON, H. V. Some computational and model equivalences in analyses of variance of unequal-subclass-numbers data. *The American Statistician*, Alexandria, v. 35, n. 1, p. 16-33, Feb. 1981.

SEARLE, S. R.; SPEED, F. M.; MILLIKEN, G. A. Population marginal means in the linear model: an alternative to least squares means. *The American Statistician*, Alexandria, v. 34, n. 4, p. 216-221, Nov. 1980.

SNEDECOR, G. W.; COX, G. **Disproportionate subclass numbers in tables of multiple classification.** Iowa: Iowa State Agriculture Experiment, 1935. (Iowa State Agriculture Experiment Station Research Bulletin, 180).

SPEED, F. M.; HOCKING, R. R. The use of the $R(\)$ -notation with unbalanced data. *The American Statistician*, Washington, v. 28, n. 1, p. 30-33, Feb. 1976.

SPEED, F. M.; HOCKING, R. R.; HACKNEY, O. P. Methods of analysis of linear models with unbalanced data. **Journal of the American Statistical Association**, Boston, v. 73, n. 361, p. 105-112, Mar. 1978.

WECHSLER, F. S. Fatoriais fixos desbalanceados: uma análise mal compreendida. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 3, p. 231-262, mar. 1998.

YATES, F. The analysis of multiple classifications with unequal numbers in the different classes. **Journal of the American Statistical Association**, Alexandria, v. 29, n. 1, p. 51-66, 1934.

YATES, F. The principles of orthogonality and confounding in replicated experiments. **Journal Agricultural Science**, Cambridge, v. 23 , pt. 1, p. 108-145, Jan. 1933.



ANEXOS

	Página
ANEXO A: Programa para análise de variância: 1º caso	100
ANEXO B: Programa para análise de variância: 2º caso	101
ANEXO C: Programa para análise de variância: 3º caso.....	102

ANEXO A

```

/*****
/*                               Programa para análise de variância                               */
/* 1º caso: fatorial 4x4 desbalanceado                                         */
/* Data: 11/05/2003                                                         */
*****/
data caso1;
  do A=1 to 4;                               /* níveis do fator A */
    do B=1 to 4;                               /* níveis do fator A */
      do rep=1 to 15;                          /* número de repetições */
input  A B peso;                             /* leitura das informações */
output;                                       /* saída dos dados */
end;
end;
end;
cards;                                         /* início de linhas de dados na ordem dada:
                                              nível de A, nível de B e variável resposta */
1      1      48.40
1      1      10.70
1      1      25.70
1      1      .          /* indica valor ausente */
1      1      .
1      1      .
1      1      17.90
.
.
4      4      58.60
4      4      55.45
4      4      46.55
4      4      50.85
4      4      50.65
;
proc print data=caso1;                       /* gera e imprime um relatório */
run;
proc glm;                                    /* análise de variância para dados
desbalanceados*/
  class A B;                                /* ordem de entrada dos dados: A, B, A-B */
  model peso=A B A*B/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;
proc glm;
  class B A;                                /* ordem de entrada dos dados: B, A, B-A */
  model peso=B A B*A/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;
quit;                                        /* fim de seção de programa */
                                           /* encerra o programa */

```

ANEXO B

```

/*****/
/*          Programa para análise de variância          */
/* 2º caso: fatorial 4x4 desbalanceado com caselas vazias */
/* Data: 11/05/2003 */
/*****/

data caso2;
  do A=1 to 4;
    do B=1 to 4;
      do rep=1 to 3;
input A B peso;
output;
end;
end;
end;
cards;
1 1 37.55
1 1 41.45
1 1 18.15
1 2 .
1 2 .
1 2 .
1 3 37.75
1 3 28.85
1 3 30.40
:
:
4 2 30.70
4 2 29.15
4 3 55.25
4 3 38.75
4 3 44.25
4 4 .
4 4 .
4 4 .
;
proc print data=caso2;
run;
proc glm;
  class A B;
  model peso=A B A*B/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;
proc glm;
  class B A;
  model peso=B A B*A/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;quit;

```

ANEXO C

```

/*****
/*          Programa para análise de variância          */
/* 3º caso: fatorial 4x5x4 desbalanceado                */
/* Data: 11/05/2003                                    */
*****/

data caso3;
  do A=1 to 4;
  do B=1 to 5;
  do C=1 to 4;
  do rep=1 to 3;
input A B C peso;
output;
end;
end;
end;
end;
cards;
1      1      1      48.40
1      1      1      10.70
1      1      1      25.70
1      1      2      30.40
1      1      2      20.40
1      1      2      47.15
1      1      3      38.45
.
.
4      5      3      44.25
4      5      4      46.55
4      5      4      50.85
4      5      4      50.65
;
proc print data=caso3;
run;
proc glm;
  class A B C;
  model peso=A B C A*B A*C B*C A*B*C/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;
proc glm;
  class B C A;
  model peso=B A C B*A B*C A*C B*A*C/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run;
proc glm;
  class C B A;
  model peso= C B A C*B C*A B*A C*B*A/E E1 E2 E3 E4 SS1 SS2 SS3 SS4 XPX I;
run; quit;

```