# ANÁLISE DE VARIÂNCIA: ALTERNATIVAS ATRAVÉS DE MODELOS DE POSTO COMPLETO

# MARCOS CUSTÓDIO NEKATSCHALOW Engenheiro Agrônomo

Orientador: Antonio Francisco Iemma

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Mestre em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA Estado de São Paulo - Brasil Janeiro -1997

## Dados Internacionais de Catalogação na Publicação (CIP) DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - Campus "Luiz de Queiroz"/USP

Nekatschalow, Marcos Custódio

Análise de variância: alternativas através de modelos de posto completo / Marcos Custódio Nekatschalow. - - Piracicaba: 1997.
123 p.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 1997.

Bibliografia.

1. Análise de variância 2. Modelo matematico I. Título.

CDD 519.535

## ANÁLISE DE VARIÂNCIA: ALTERNATIVAS ATRAVÉS DE MODELOS DE POSTO COMPLETO

### MARCOS CUSTÓDIO NEKATSCHALOW

Aprovada em: 09.04.97

Comissão julgadora:

Prof. Dr. Antonio Francisco lemma (ESALQ/USP) Profa Dra Sônia Maria de Stefano Piedade Prof. Dr. Lúcio Benedicto Kroll

(ESALQ/USP) (UNESP/Botucatu)

Prof.Dr. Antonio Franc

Ao meu pai Victor Nekatschalow(in memorian)

À minha mãe Thereza Custódio

Aos meus irmãos Sérgio, André e Paulo

DEDICO

A minha mulher, Lúcia

OFEREÇO

#### **AGRADECIMENTOS**

- Ao Professor Antonio Francisco Iemma, pela orientação, amizade e confiança, fundamentais para conclusão deste estudo.
- À CAPES, pela bolsa de estudo concedida.
- A todos professores do Departamento de Matemática e Estatística pela transmissão dos conhecimentos ministrados e apoio.
- Aos amigos e colegas da ESALQ, em especial do Departamento de Matemática e Estatística e do Departamento de Genética.
- Aos amigos César Gonçalves de Lima e Roseli Aparecida Leandro pelo auxílio com sistemas estatísticos, realizando análises de variância.
- Aos funcionários do Departamento de Matemática e Estatística pelo apoio recebido.

## SUMÁRIO

LIS	STA D	DE TABELAS	ix
RE	SUM	0	X
su	MMA	RY	xii
1.	INTF	RODUÇÃO	1
2.	REV	ISÃO BIBLIOGRÁFICA	4
3.	MET	ODOLOGIA	28
	3.1.	CARACTERIZAÇÃO	28
	3.2.	ESTIMAÇÃO	29
		3.2.1. As idéias de RAO (1945)	29
		3.2.2. Definições e Teoremas Importantes	30
		3.2.3. Regras Práticas	31
	3.3.	TESTABILIDADE	33
	3.4.	PARTIÇÃO DAS SOMAS DE QUADRADOS	33
	3.5.	NOTAÇÃO R( )	35
	3.6.	MODELOS UTILIZADOS	36
		3.6.1 Introdução	36
		3.6.2. Modelo Superparametrizado (Modelo-S)	37
		3.6.3. Modelo de Médias de Caselas	
		(Modelo-M)	39
		3.6.4. Modelo com Restrição $\Sigma$ (Modelo- $\Sigma$ )	41
		3.6.5. Modelo com Restrição Zero	
		(Modelo-Z)	42

	3.	6.6.	Modelo	com	Restriç	ao i	onder	ada	
			(Modelo-	W)					43
	3.	6.7	Modelo c	ie Regr	essão (	Mode	lo-R)		44
	3.7. H	IPÓT	ESES ES	TATÍST	TCAS				46
	3.8. S	ISTE	MAS EST	ATÍSTI	cos	· · · · · · · · ·			49
	3.	8.1.	SAS/PRO	C GLM	1				50
	3.	8.2.	BMDP						51
	3.	8.3.	GLIM			• • • • • • •			52
	3.	8.4.	SPSS/AN	IOVA					52
	3.	8.5.	MINITAB						53
4.	ASPEC	TOS	PRÁTICO	OS IMP	ORTAN	TES			54
			DUÇÃO						54
			OS UTIL						57
	4	2.1.	Modelo-	S					57
			4.2.1.1.	Caract	erizaçã	0			57
			4.2.1.2.	Proced	dimento	G	LM	do	
				Sistem	a Estat	ístico	SAS		58
			4.2.1.3.	Equaç	ões	Norn	nais	е	
				Aproxi	mação	de	Míni	mos	
				Quadr	ados				68
			4.2.1.4.	Estima	ıção				69
			4.2.1.5.	Hipóte	ses	Testa	das	na	
				Anális	e de Va	riânci	a		73
	4	.2.2.	Modelo-l	М					80
			4.2.2.1.	Carac	erizaçã	o	*******		80
			4.2.2.2.	Sistem	na Estat	ístico	BMDP		81
			4.2.2.3.	Equaç	ões	Norr	nais	е	
				Aproxi	mação	de	Míni	mos	
				Quadr	ados				82

	4.2.2.4.	Estimação	84
	4.2.2.5.	Hipóteses Testadas na	
		Análise de Variância	88
4.2.3.	Modelo-	Σ	95
	4.2.3.1.	Caracterização	95
	4.2.3.2.	Procedimento GLM do	
		Sistema Estatístico SAS	97
	4.2.3.3.	Equações Normais e	
		Aproximação de Mínimos	
		Quadrados	97
	4.2.3.4.	Hipóteses Testadas na	
		Análise de Variância	98
4.2.4.	Modelo-	Z	100
	4.2.4.1.	Caracterização	100
	4.2.4.2.	Sistema Estatístico GLIM	101
	4.2.4.3.	Equações Normais e	
		Aproximação de Mínimos	
		Quadrados	104
	4.2.4.4.	Hipóteses Testadas na	
		Análise de Variância	105
4.2.5.	Modelo-	W	105
	4.2.5.1.	Caracterização	105
	4.2.5.2.	Sistema Estatístico	
		SPSS/ANOVA	107
	4.2.5.3.	Equações Normais e	
		Aproximação de Mínimos	
		Quadrados	108
	4.2.5.4.	Hipóteses Testadas na	
		Análise de Variância	109
4.2.6.	Modelo-	R	109

	4.2.6.1.	Caracterização	109
	4.2.6.2.	Sistema Estatístico MINITAB.	112
	4.2.6.3.	Equações Normais e	
		Aproximação de Mínimos	
		Quadrados	113
	4.2.6.4.	Hipóteses Testadas na	
		Análise de Variância	114
5.	CONCLUSÕES		116
6.	REFERÊNCIAS BIBL	IOGRÁFICAS	119

## LISTA DE TABELAS

Tabela 01 - Hipóteses para linhas colunas e interação	. 13
Tabela 02 - Produtividade: Tonelada de Cana/ha	. 56
Tabela 03 - Funções estimáveis do Tipo I do SAS para a ordem A-B	.60
Tabela 04 - Funções estimáveis do Tipo I do SAS para a ordem B-A	.61
Tabela 05 - Funções estimáveis do Tipo II do SAS para a ordem A-B	. 62
Tabela 06 - Funções estimáveis do Tipo II do SAS para a ordem B-A	.63
Tabela 07 - Funções estimáveis do Tipo III do SAS para a ordem A-B	. 64
Tabela 08 - Funções estimáveis do Tipo III do SAS para a ordem B-A	. 65
Tabela 09 - Funções estimáveis do Tipo IV do SAS para a ordem A-B	. 66
Tabela 10 - Funções estimáveis do Tipo IV do SAS para a ordem B-A	. 67
Tabela 11 - Funções estimáveis sobre o fator A no Modelo-S	.71
Tabela 12 - Funções estimáveis sobre o fator B no Modelo-S	.72
Tabela 13 - Hipóteses testadas sobre o fator A no Modelo-S	.76
Tabela 14 - Hipóteses testadas sobre o fator B no Modelo-S	.77
Tabela 15 - SQ Tipo I do SAS¹	.78
Tabela 16 - SQ Tipo II do SAS²	.78
Tabela 17 - SQ Tipo III do SAS³	.79
Tabela 18 - SQ Tipo IV do SAS⁴	.79
Tabela 19 - Funções estimáveis sobre o fator A no Modelo-M	. 86
Tabela 20 - Funções estimáveis sobre o fator B no Modelo-M	. 87
Tabela 21 - Hipóteses testadas sobre o fator A no Modelo-M	.89
Tabela 22 - Hipóteses testadas sobre o fator B no Modelo-M	. 90
Tabela 23 - SQ com ajuste seqüencial no Modelo-M <sup>1</sup>	.93
Tabela 24 - SQ ajustadas para os efeitos principais no Modelo-M <sup>2</sup>	. 93
Tabela 25 - SQ Tipo III no Modelo-M <sup>3</sup>	. 94
Tabela 26 - SQ Tipo IV no Modelo-M⁴	. 94
Tabela 27 - Macros e Análises de Variância pelo GLIM	. 102

# ANÁLISE DE VARIÂNCIA: UMA ALTERNATIVA PARA MODELOS DE POSTO COMPLETO

Autor: Marcos Custódio Nekatschalow

Orientador: PROF. DR. Antonio Francisco lemma

#### RESUMO

É objeto deste trabalho mostrar, aos pesquisadores de ciências aplicadas, diferenças existentes entre os modelos adotados por diversos sistemas estatísticos utilizados para realização de análise de variância, relacionando-os com as hipóteses referentes às somas de quadrados resultantes, para conjunto de dados desbalanceados com caselas vazias.

Para tanto, são abordados, além do modelo superparametrizado de posto incompleto, os seguintes modelos lineares de posto completo: com Restrição  $\Sigma$ , de Médias de Caselas, com Restrição Zero, com Restrição Ponderada e de Regressão. Eles, são relacionados,

respectivamente, com os sistemas estatísticos: SAS/PROC GLM, BMDP, GLIM, SPSS/ANOVA e MINITAB.

Um conjunto de dados com duas caselas vazias, é utilizado para ilustrar que os diferentes sistemas proporcionam, muitas vezes, formas diversas de particionamento da soma de quadrados de parâmetros.

Nesse contexto, as hipóteses testadas pelos modelos aqui discutidos são comparadas com aquelas usualmente testadas através do modelo superparametrizado (posto incompleto), universalmente adotado pela facilidade que proporciona ao usuário na descrição dos fatores de interesse em seus experimentos.

Cada sistema possui sua forma peculiar de abordar conjunto de dados desbalanceados, marcadamente em presença de caselas vazias. E a sua posição pode ser, muitas vezes, fundamental para obtenção das somas de quadrados.

Os modelos apresentam semelhanças e diferenças.

Ressalta-se aqui, que mesmo em se tratando de modelos equivalentes, os diferentes métodos incorporados aos sistemas estatísticos podem levar a testar diferentes hipóteses. Proporcionando, em geral, no caso de caselas vazias diferentes somas de quadrados.

ANALYSIS OF VARIANCE: AN ALTERNATIVE TO

FULL-RANK-MODELS

Author: Marcos Custódio Nekatschalow

Advisor: PROF. DR. Antonio Francisco lemma

SUMMARY

The objective of this work is to demonstrate, to researchers of the applied sciences, differences between underlying models of various statistical systems used for Analysis of Variance, and to relate the hypotheses tested by these models and the resulting sums of squares for a set of unbalanced data with missing cells.

Other than the non-full-rank, overparametrized model, the following full-rank linear models are examined:  $\Sigma$  restriction, cells means, zero restriction, weighted restriction, and regression. These models are implicit in the following statistical systems, respectively: SAS/PROC GLM, BMDP, GLIM, SPSS/ANOVA e MINITAB.

A set of data with two missing cells is used to demonstrate that the sums of squares of the parameters are frequently partitioned in different manners by these systems.

In this context, the hypotheses tested by these models are compared with those usually tested by the overparametrized model (non-full-rank), which is universally adopted because of the ease in which the factors of interest are described by the user.

Each system has its own way of handling a set of unbalanced data, markedly when missing cells are present. The position of the missing cell(s) in relation to the rest of the data, can often be fundamental to the obtainment of the sums of squares.

The models have similarities and differences. It is emphasized that, although the models are equivalent, the different methods incorporated by the statistical systems may result in different hypotheses being tested. In the case of missing cells, the resulting sums of squares are usually different.

### 1. INTRODUÇÃO

Há cerca de trinta anos atrás, a realização de análises estatísticas exigia a conciliação entre dois aspectos. O primeiro deles, equipamentos apropriados: calculadoras e/ou computadores que usavam, em geral, linguagem Fortran. O segundo, material humano: técnicos especializados, estatísticos e programadores, com habilidades para executarem as análises. Esse binômio regia grande parte do universo da análise de dados.

Hoje estão a disposição microcomputadores, que são acessíveis ao grande público dos pesquisadores. Associado a esse aspecto, encontra-se disponível um grande número de sistemas estatísticos, que muitas vezes fazem com que o pesquisador sinta-se auto-suficiente para planejar, conduzir e analisar suas próprias pesquisas. Isso ocorre muitas vezes, mesmo que ele não tenha um sólido embasamento estatístico. As condições atuais impelem o pesquisador nessa direção.

Muitos dos sistemas estatísticos, são utilizados para realização de Análise de Variância. Entretanto, poucos manuais são esclarecedores quanto às dificuldades existentes, quando se trata da análise de

conjuntos de dados desbalanceados. Nesse contexto, por exemplo, em um experimento com classificação dupla cruzada, os usuários podem concluir, através de um dado sistema de análise, que há significância para este ou aquele efeito.

Dois aspectos básicos devem ser considerados.

O primeiro é que, em geral, os pesquisadores colocam a Análise de Variância sob o prisma de um modelo superparametrizado (posto coluna incompleto), no qual são explicitados todos os fatores e respectivos níveis, enquanto que muitos dos sistemas estatísticos utilizam modelos de posto coluna completo.

O segundo é que, as hipóteses sobre efeitos principais em um modelo com interação, carregam consigo, algumas vezes, além de componentes de interação. componentes do outro fator. Para agravar esse fato, os sistemas estatísticos, normalmente, não descrevem com detalhes as hipóteses associadas às somas de quadrados fornecidas. Além disso, muitos dos manuais são falhos nesse aspecto, não orientando sobre o método estatístico utilizado para obtenção da Análise de Variância. Dessa forma, torna-se difícil para o pesquisador concluir sobre o que está sendo testado, ou mesmo questionar qual seria o sistema estatístico mais apropriado para o conjunto de hipóteses que deseja testar.

Esse problema tem sido motivo de preocupação de pesquisadores consagrados internacionalmente. Fato

que pode ser verificado em muitas das publicações feitas na última década. É nessa direção que se desenvolve este estudo.

Nesse contexto, tem-se por objetivos:

- -Apresentar os modelos mais importantes de posto completo.
- -Contrastá-los com o modelo superparametrizado de posto incompleto.
- -Discutir as hipóteses estatísticas testadas por esses modelos.
- -Interpretar as hipóteses estatísticas testadas pelos referidos sistemas estatísticos.

### 2. REVISÃO BIBLIOGRÁFICA

Α análise de variância para dados desbalanceados é algo que preocupa grande parte dos pesquisadores. Εm particular, aqueles ligados à agropecuária, no tocante às hipóteses associadas somas de quadrados, obtidas em uma análise de variância pois, nesse ramo de atuação de pesquisa, é comum haver perda de observações, a utilização de delineamentos experimentais com desbalanceamento planejado, bem como a ocorrência de dados desbalanceados.

análise de variância para dados experimentais, que hoje se encontra largamente difundida mais diversas pesquisa científica, foi nas áreas da desenvolvida por R. A. Fisher. É interessante observar, como escreve YATES(1934), que as bases da estatística relacionadas experimental deram-se nas ciências agropecuária.

YATES(1933) e SNEDECOR & COCHRAN (1967), citados por SANTOS(1994), publicaram trabalhos, nos quais ponderam a respeito das conseqüências da desproporcionalidade de subclasses em modelos experimentais. Alertam a respeito das limitações nas interpretações dos resultados obtidos com base em tais

conjuntos de dados, pois, as soluções podem ser muito gerais.

A literatura cita também, BRANDT(1933). O autor trata de um caso de experimentos em delineamentos de classificação dupla, com dados desbalanceados. Apesar de fomentar a discussão a respeito deste assunto, o estudo trata somente o caso de fatorial 2<sup>n</sup>.

YATES(1934) realiza uma generalização do que BRANDT(1933) havia proposto. Estabelece três métodos para cálculo das somas de quadrados. São eles:

- Método de Análise Aproximado. É citado na bibliografia como Método das médias não ponderadas.
- Método das Médias dos Quadrados Ponderados.
- Método do Ajuste de Constantes. Este, foi o início das idéias a respeito do método, que hoje em dia, se conhece por parametrização sucessiva, muito útil quando as somas de quadrados são estudados através da notação R().

PATTERSON(1946), Segundo os métodos propostos anteriormente, podem ser vistos como uma correção ou um ajuste pela diferença entre as médias. método Nesse trabalho, sugere um que seria uma alternativa ao Método do ajuste de constantes, proposto por YATES(1934). Para tanto, considera a média marginal de cada fator para o ajuste da soma de quadrados ponderada para as subclasses.

A melhor forma de se realizar a análise de dados desbalanceados foi, e continuará sendo, motivo de controvérsia. VAJDA(1947) discorda de YATES(1934) a

respeito da melhor forma de se realizar este tipo de análise. Entretanto, YATES(1947), rebate as críticas, alegando que na presença de interação, a magnitude dos efeitos principais dependerá da forma como foram definidos.

Já FINNEY(1948), apesar de concordar com YATES(1934), mostra aspectos positivos em VAJDA(1947). Este procura dar um pouco mais de atenção ao que se refere à definição de efeitos principais. Ressalta, no entanto, para as "imprudências" contidas em VAJDA(1947). desbalanceados. Tratando-se de dados com vazias, porém com igual frequência, para as caselas com observações, a partição da soma de quadrados parâmetros pode ser feita de forma clara. Entretanto, quando há número desigual, há controvérsias. Além disso, problema da proporcionalidade de frequências, há tratado por VAJDA(1947). Ali, o autor argumenta que, as somas de quadrados devem ter tratamento diferenciado se as caselas mantiverem uma relação de proporcionalidade entre suas freqüências. Realça ainda que, na ausência de interação, ficar sensivelmente 0 problema pode simplificado.

HENDERSON(1953) descreve três métodos para obtenção das somas de quadrados. Estes métodos foram desenvolvidos visando aplicação no estudo de componentes de variância. Entretanto, os métodos 1 e 3 têm sido utilizados com sucesso na análise de modelos de efeitos fixos com dados desbalanceados:

- Método 1: utiliza o método do ajuste de constantes, ajustando os efeitos principais pela média. Apresenta a limitação de que a soma de quadrados correspondente à interação não está, em geral, relacionada às hipóteses de interesse do pesquisador.
- Método 3: utiliza o método de ajuste de constantes, aliando o conceito da ordenação "a priori". Esse método é utilizado atualmente por diversos sistemas estatísticos.

KRAMER(1955) levanta algumas dificuldades existentes na obtenção das somas de quadrados com dados desbalanceados. Em algumas situações pode ocorrer que a soma de quadrados da interação resulte negativa. O autor alega que a melhor forma de obtenção da soma de quadrados para dados desbalanceados é aquela que utiliza o método do ajuste de constantes.

Quando dados se tem um conjunto de desbalanceados e pretende-se realizar sua análise de variância, há diferentes propostas de como se obter sua soma de quadrados. Possivelmente, ter-se-á resultados diversos. ELSTON & BUSH(1964) discutem quanto às possíveis hipóteses que podem ser testadas, quando da presença de interação em experimentos com classificação dupla. Esses pesquisadores ressaltam a estreita relação entre testabilidade e estimabilidade. Dentro do modelo utilizado para análise de variância, propõem restrições do tipo  $\Sigma$ .

O método de médias não ponderadas, para as hipóteses de nulidade a respeito dos efeitos principais, é

revisto por GOSSLEE & LUCAS(1965). Os autores consideram o método razoável para se testar hipóteses sobre os efeitos principais.

Ao realizar uma análise de dados através de um sistema estatístico, muitas vezes o usuário não tem noção dos procedimentos adotados quanto às aproximações feitas. Sob esta ótica, LONGLEY(1967) procura resolver algumas dúvidas suscitadas por usuários. Faz também um levantamento a respeito da dimensão dos erros ocasionados pelas aproximações.

Recentemente, as análises de variância têm sido feitas, preferencialmente, por meio de sistemas estatísticos. FRANCIS(1973) compara alguns deles tocante à análise de variância, utilizando um exemplo de fatorial com classificação dupla, com um conjunto de dados desbalanceados. Um mesmo conjunto de dados é analisado através de diferentes sistemas. Observa que, diferentes resultados de somas de quadrados são obtidos para os efeitos principais, conforme o sistema adotado. feitas análises através dos seguintes sistemas: São SAS/PROC ANOVA (Statistical Analisys System), BMDX64 (Biomedical Package), CAROLINA (Multivariate Analisys of Variance Program) e OSIRIS (Inter-University Consortium for Political Research). Atenta que, para o conjunto de dados em questão, o SAS PROC ANOVA proporciona uma soma de quadrados negativa para a interação.

HOCKING & SPEED (1975) ressaltam a importância das hipóteses testadas e sua relação com as

somas de quadrados a elas associadas, quando o conjunto de dados é desbalanceado. Expõem o contraste entre um modelo de posto incompleto (superparametrizado) e um de posto completo (médias de caselas), evidenciando a necessidade de serem identificadas as hipóteses testadas. Mormente, quando se tem perda total de uma casela. No modelo de médias de caselas, em geral, todos os parâmetros são estimáveis; o que não ocorre no modelo superparametrizado. Conseqüentemente, todas as funções paramétricas no modelo de médias de caselas serão estimáveis, o que nem sempre acontecerá no modelo superparametrizado.

A notação R() para dados desbalanceados é discutida em SPEED & HOCKING(1976). Segundo os autores, há duas maneiras de se utilizar a notação R(). Evidenciam as diferenças, através dos modelos de posto incompleto(superparametrizado) e o modelo de posto completo(restrição do tipo  $\Sigma$ ), ilustrando as duas formas de se fazer uso deste tipo de notação. Para se fazer uso da notação R(), faz-se uma partição conveniente na matriz de delineamento e no vetor de parâmetros. Em termos práticos, as partições são feitas de acordo com a parametrização sucessiva do modelo.

No procedimento 1, através do modelo  $y=X\beta+e$ , onde  ${}_{n}y_{I}$  é um vetor de observações;  ${}_{n}X_{p}$ é uma matriz conhecida, de posto  $q\leq p$ ;  ${}_{p}\beta_{I}$  é um vetor de parâmetros;  ${}_{n}e_{I}\cap(\phi,I\sigma^{2})$  é um vetor de erros. Nesse

contexto, a notação R( ), é assim definida:  $R(\beta) = \hat{\beta}'X'y$ . Onde,  $\hat{\beta}$  é qualquer solução para o sistema de equações normais  $X'X\hat{\beta} = X'Y$ . Efetuando-se partição para X e  $\beta$  como  $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$  e  $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ , respectivamente, e tomando-se o modelo  $Y = X_2\beta_2 + e$ , obtém-se  $R(\beta_2) = \hat{\beta}_2'X_2'Y$ , onde  $\hat{\beta}_2'$  é qualquer solução para o sistema de equações normais  $X_2'X_2\hat{\beta}_2 = X_2'Y$ . Pode-se então definir :  $R(\beta_1|\beta_2) = R(\beta_1,\beta_2) - R(\beta_2)$ .

Procedimento 2 - Os autores anexam, então, ao modelo citado, um conjunto de restrições paramétricas não estimáveis. Com tal reparametrização, obtém-se o modelo:  $Y = X * \beta * + e$ . Este modelo possui posto completo, apresentando, portanto, solução única para seu sistema de equações normais:  $\hat{\beta}^* = (X *' X *)^{-l} X *' Y$ . Pode-se, então, fazer uma partição semelhante à feita para o primeiro modelo, sem restrição.

Os autores ressaltam diversos pontos que podem induzir os usuários de sistemas estatísticos a alguns erros, especificamente quando se trata de dados desbalanceados. O primeiro, é que no procedimento 2,  $R(\beta_2)$  nem sempre é definida, e quando é definida pode apresentar o problema da falta de unicidade. Em segundo, que os dois procedimentos, no caso de desbalanceamento, apresentam somas de quadrados distintas, sendo que o procedimento 2 não é um caso especial do procedimento 1.

O terceiro ponto consiste no fato que, em nenhum dos dois procedimentos, as hipóteses testadas são claramente, dificultando a interpretação dos quocientes entre os quadrados médios (Teste F). A respeito deste último aspecto, os autores ressaltam que as hipóteses resultantes de dados desbalanceados ficam mais difíceis de serem interpretadas através da notação R(). Ao final, os autores mostram que a hipótese geral,  $H'eta=\gamma$ , não pode ser testada através do uso da notação R( ). Os autores observam que, no caso de dados desbalanceados, as somas de quadrados obtidas para os efeitos principais, apresentar diferença modelos de posto podem nos completo e incompleto.

As idéias de estimabilidade e de testabilidade são discutidas por SEELY(1977). O autor discute a respeito de estimabilidade em um modelo superparametrizado, sem desbalanceamento. Comenta que para se testar  $\mathbf{H}'\beta = \delta$  vs.  $\mathbf{H}'\beta \neq \delta$ ,  $\mathbf{H}'\beta$  deve ser constituído por um conjunto de funções paramétricas estimáveis. Oferece um exemplo de um fator. Estes conceitos são aplicáveis no caso de haver balanceamento.

SPEED, HOCKING & HACKNEY(1978), apresentam uma discussão importante a respeito da análise de dados desbalanceados. A relação existente entre hipótese testada e notação R() é comentada, com ênfase nas diferenças entre os modelos superparametrizado e sujeito a restrições paramétricas

não estimáveis. Segundo os autores, quando há presença de casela(s) vazia(s), 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 médias de caselas(em modelo de geral, posto completo), tornam-se de mais fácil interpretação hipóteses que podem ser obtidas para um modelo com classificação dupla, principalmente na presença interação. A tabela 1 exibe as hipóteses sugeridas neste trabalho, levando-se em consideração o modelo de médias de caselas(Modelo-M). Mostra, também, sua relação com o modelo superparametrizado(ajuste de constantes) e o que é denominado de método de regressão(variáveis binárias). Sendo para este último, algo próximo do que ocorre para o modelo sob restrição Σ.

Ressaltam, entretanto, para que as hipóteses H1, H4, H5, H8 e H9 possam ser testadas, não deve haver casela vazia.

Dentre OS métodos. comentam OS de YATES(1934), OVERAL & SPIEGEL (1969) e HENDERSON (1953). Há uma relação entre estes diferentes métodos. Colocam 0 problema para um conjunto de dados desbalanceado, com casela(s) vazia(s) e interação.

Tabela 1: Hipóteses para linhas, colunas e interação

EFEITOS		Modelo-M	S.Q.
FATOR A	H1	$\overline{\mu}_{i.} = \overline{\mu}_{i'.}$	$^{1}$ SQH1= $R(\alpha \mu)$
(LINHAS)	H2	$\sum_{j} \frac{n_{ij} \mu_{ij}}{n_{i.}} = \sum_{i} \frac{n_{i'j} \mu_{i'j}}{n_{i'.}}$	<sup>1</sup> SQH2= $R(\alpha \mu,\beta)$
	Н3	$\sum_{j} (n_{ij}) \mu_{ij} = \sum_{i'} \sum_{j} \frac{n_{ij} n_{i'j} \mu_{i'j}}{n_{.j}}$	<sup>2</sup> SQH3= $R(\dot{\alpha} \dot{\mu},\dot{\beta},\dot{\gamma})$
	H4	$\mu_{il} = \mu_{i'l}$	SQH4
FATOR B	Н5	$\overline{\mu}_{.j} = \overline{\mu}_{.j'}$	$^{1}$ SQH5= $R(\beta \mu)$
(COLUNAS)	Н6	$\sum_{j} \frac{n_{ij} \mu_{ij}}{n_{.j}} = \sum_{i} \frac{n_{ij'} \mu_{ij'}}{n_{.j'}}$	$^{1}$ SQH6= $R(\beta \mu,\alpha)$
	<b>H</b> 7	$\sum_{j} (n_{ij}) \mu_{ij} = \sum_{j'} \sum_{i} \frac{n_{ij} n_{ij'} \mu_{ij'}}{n_{i.}}$	<sup>2</sup> SQH7= $R(\dot{\beta} \dot{\mu},\dot{\alpha},\dot{\gamma})$
	Н8	$\mu_{1j} = \mu_{1j'}$	SQH8
INTERAÇÃO	Н9	$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$	$R(\gamma \mu,\alpha,\beta)$

<sup>1/:</sup> SQH1, SQH2, SQH5 e SQH6 podem ser obtidas pelo método do ajuste de constantes no modelo irrestrito.

Assim, para os efeitos principais há quatro formas de calcular as somas de quadrados. No caso da interação. obtém-se 0 mesmo resultado, independentemente do método. Verificam que, no caso de um modelo sem interação, há por vezes uma equivalência efeitos das quadrados obtidas somas de para os principais.

 $<sup>\</sup>underline{2/}$ : SQH3 e SQH7 podem ser obtidas pelo método de ajuste de constantes no modelo restrito.

URQUHART & WEEKS(1978), expõem, para um modelo de classificação dupla, o problema de que havendo vazia. 0 modelo superparametrizado quando da suposição da inadequado presença interação. Para tanto, sugerem, como alternativa, o Modelo-M. Nesse caso, alegam que através desse modelo pesquisador pode definir funções de mais interpretação, através das médias populacionais das caselas. O modelo utilizado é:  $y = W\mu + e$ , com sistema de equações de solução única dada por  $\hat{\mu} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}$ . autores colocam a questão do que seria melhor: ou um estimador com variância mínima, porém com pouco ou nenhum interesse; ou um estimador com variância maior, mas com algo mais próximo do que o pesquisador procura?

HENDERSON & McALLISTER(1978) discutem a respeito, de um modelo de classificação dupla, desbalanceado. Descrevem três modelos: superparametrizado; de média de caselas e com restrição Σ. Discorrem a respeito dos estimadores em tais modelos.

Utilizando o mesmo modelo de médias de caselas, BURDICK(1979) explora alguns casos. Para um conjunto de dados desbalanceados, porém com todas as caselas ocupadas, obtém as somas de quadrados para linhas de duas maneiras. Verifica que pode ocorrer de as hipóteses de Linhas Ajustadas por Colunas e a de Média Simples de Linhas sejam equivalentes. Esta possibilidade ocorreria, por exemplo, em um experimento com duas

linhas, se e somente se:  $\binom{1}{n_{1J}} + \binom{1}{n_{2J}}$  for igual para cada coluna. No caso de um delineamento 2x2, as hipóteses seriam equivalentes se tivermos:  $\mathbf{n}_{11} = \mathbf{n}_{22}$  e  $\mathbf{n}_{12} = \mathbf{n}_{21}$ . Ressalta, ser didática a finalidade maior. Isto, em virtude da dificuldade de se de estender este procedimento para delineamentos mais complexos.

modelo de classificação Em um dupla. pesquisador ao realizar uma análise de variância, por meio de um sistema estatístico, por vezes não saberia distinguir na soma de quadrados para os fatores obtidas saidas das análises principais, nas proporcionadas por meio destes sistemas, qual hipótese estaria sendo testada. Ou, até mais conveniente, é se perguntar que hipóteses as somas testam. É o que SEARLE(1979) comenta, especialmente para dados desbalanceados. caselas vazias. alguns com para sistemas.

Através da ilustração de um exemplo de fatorial 2x3, FREUND(1980) discorre a respeito da perda de observações, bem como de caselas inteiras, com suas implicações nas funções estimáveis e hipóteses que se pode formular a respeito de um conjunto de dados com esta caracterização. O autor comenta que o PROC GLM do SAS procura lidar com este tipo de situação. Entretanto, dependendo de qual casela foi perdida, as somas de quadrados podem apresentar diferenças entre si.

SEARLE et alli (1980) exploram o conceito da média marginal da população (PMM) e sua relação com algumas saídas de sistemas estatísticos, como o SAS/GLM e o SAS HARVEY. Os autores procuram relacionar a idéia de estimabilidade com o método de mínimos quadrados.

Para um mesmo conjunto de dados, pode-se análise de variância por meio de diferentes modelos. Se os modelos apresentarem o mesmo espaço na matriz de delineamento, seriam modelos equivalentes. SEARLE et alli (1981) estudam para um modelo de classificação dupla, casos de dados balanceados; desbalanceados, com todas as caselas ocupadas; desbalanceados, com casela vazia. Entretanto, pode haver divergências quanto às hipóteses testadas ao levar-se em consideração a presença ou não de interação. Segundo os autores, a obtenção das somas de quadrados pode ser efetuada por diversos algoritmos. Pode-se utilizar o critério da inversa de parte da inversa. Outro procedimento, seria particionar a matriz de delineamento de forma conveniente. separando-se 0 conjunto parâmetros de cada fator em estudo. Nesse caso, seria utilizada a notação R().

Para uma melhor compreensão de como se realiza a análise de dados, classificam os modelos em relação ao posto: completo ou incompleto. O modelo de posto completo serve de ponte para obtenção das somas de quadrados para o modelo de posto incompleto. Entretanto, ressaltam que no caso de perda de

observações, mais especificamente quando em presença de casela(s) vazia(s), para as hipóteses associadas às somas de quadrados no modelo de posto completo não haveria uma correspondência perfeita com incompleto. Neste caso, salientam as hipóteses que tornam-se de mais difícil interpretação. O modelo de médias de caselas serviria como base para se obter mais facilmente as interpretações das hipóteses. Os autores mostram uma relação existente entre os modelos de posto completo e incompleto. Além do modelo com restrição  $\Sigma$ , os autores comentam a respeito de outros dois modelos restritos: modelo com restrição Zero e modelo de restrição ponderada. O primeiro, faz a restrição para que um nível de cada fator seja igual a zero; bem como as respectivas interações entre o s efeitos dos fatores principais envolvidos nas restrições, seja também igual a zero. Já no segundo modelo, a restrição feita é algo próximo do que se faz para o modelo com restrição Σ, mas levando-se em conta o peso dado pelo número de observações das marginais de cada fator. Os autores ainda expõem um exemplo com dados desbalanceados, mas com todas as caselas ocupadas. Os autores procuram estudar, principalmente, quatro sistemas estatísticos: BMDP2V. SAS/GLM, SAS/HARVEY e SPSS(Statistical Package for Social Sciences). Dentre outras conclusões, afirmam que: As restrições do tipo Σ são usadas de forma explícita

ou implícita pelo BMDP2, SAS/HARVEY (utilizando o método da inversa de parte da inversa), SPSS ANOVA opção 9 e pelo SAS GLM na soma de quadrados do Tipo III. Quando há presença de casela(s) vazia(s), o BMDP não oferece saída; o SPSS/ANOVA, que usa este modelo, calcula somas de quadrado sem interesse para o pesquisador.

STEINHORST(1982) trata de algumas dúvidas a respeito. Comenta que, por muitas vezes não fica claro, em alguns artigos e livros-textos o uso da notação R(), quando se faz a restrição do tipo  $\Sigma$ , passando-se de um modelo de posto incompleto para um de posto completo. O autor critica a forma como se faz uso do modelo de médias de caselas. Pondera que quando se tem um modelo misto ou aleatório, o modelo de médias de caselas não mostra bem as hipóteses. O autor então considera, dentre outros, os seguintes aspectos: estimadores para modelos mistos e aleatórios; componentes de variância para tais modelos, visando os testes adequados.

JENNINGS & WARD(1982) discorrem a respeito do problema de dados desbalanceados, no caso de modelo de classificação dupla, na presença de casela(s) vazia(s), e as dificuldades decorrentes. Argumentam que, boa parte das dúvidas originadas análise de dados na desbalanceados é resultado do do modelo uso superparametrizado. O uso do modelo de posto completo, mais especificamente o modelo de médias de caselas, faz reduzir as dificuldades na interpretação das hipóteses oferecidas por um conjunto de dados desbalanceados. Os autores relacionam as hipóteses geradas superparametrizado com as proporcionadas pelo modelo de médias de caselas. Ressaltam que FREUND(1980) oferece uma valiosa contribuição. Concluem que há necessidade de se criar um vínculo com os usuários da estatística, para comunicação. Para tanto, deve haver boa desenvolvimento dos conceitos, vocabulário e notação adequados. Acreditam que o modelo superparametrizado não cumpre com efeito esta finalidade. Entretanto, o modelo de posto completo não expõe as hipóteses com os parâmetros, com que se está normalmente acostumado. É necessário haver uma aproximação entre os algoritmos utilizados para realização das Análises de Variância e a linguagem utilizada pelos pesquisadores de forma geral. Esses, mais acostumados a explicitar todos os fatores envolvidos meio do modelo no delineamento. por superparametrizado.

Através da notação R() pode-se obter as somas de quadrados em uma análise de variância. Por exemplo, para um modelo de classificação dupla cruzada, dentro de um modelo superparametrizado, fazendo-se o ajuste seqüencial para um dos efeitos principais, tem-se por exemplo: A ajustado após a constante, o fator B e AB(interação). Na notação R() fica:  $\mathbf{R}(\alpha|\mu,\beta,\alpha\beta)$ .

SMITH e MURRAY(1983) propõem um modelo de médias de caselas modificado. Alegam ser um método

para evitar as confusões causadas pela estimação de mínimos quadrados de parâmetros em modelos de posto em um conjunto de incompleto, principalmente desbalanceados. Para tanto, trabalham em um exemplo de classificação dupla 2 x 2. Lembram que para "completar" o posto de um modelo superparametrizado, utiliza-se um conjunto de restrições paramétricas não estimáveis. Tal conjunto de restrições é colocado na forma de matriz, multiplicando o vetor de parâmetros. Particiona-se, então, matriz de tal forma a ter-se a segunda partição quadrada, com dimensões iguais ao posto correspondente ao conjunto de restrições. Então, substitui-se a segunda partição do vetor de parâmetros, em função da primeira partição, dentro do modelo. Observa-se a necessidade de haver uma correspondência entre o que é feito para o conjunto de restrições e а matriz de delineamento. Concluem que este procedimento tem cunho didático. Essa partição torna mais fácil a compreensão, mesmo em um modelo com restrição. Realçam a diferença entre estimar médias das caselas, quando há interação ou não. Também mostram a forma como se calculam as estimativas de mínimos quadrados para cada casela, conforme é feito através do SAS.

A idéia da necessidade de haver um conjunto de restrições não estimáveis é colocada por SEARLE(1984), para um modelo de classificação simples. Este conjunto de restrições é adicionado ao modelo, completando o posto. O autor refere-se também à forma

como o SAS utiliza esse procedimento para calcular as somas de quadrados.

A idéia de conectividade ao se propor uma análise de variância, para um conjunto de dados desbalanceados na presença de casela(s) vazia(s), é muito importante. MURRAY e SMITH(1985) exploram o assunto, utilizando principalmente o modelo de médias de caselas. em um modelo sem interação. Comentam rapidamente respeito а de análise de dados experimentos desconexos. Quando há caselas vazias, pode haver problemas para se estimar parâmetros de interesse do pesquisador. Há um teste simples para verificar se há ou não conectividade. Os autores discorrem a respeito do teste. A questão que se propõe é: há informação suficiente para se estimar todas as médias de caselas?

Uma resenha interessante revisão a respeito do assunto está presente em HERR(1986). O autor faz um resumo sobre os primeiros trinta anos da análise de variância de dados desbalanceados em experimentos de Foi classificação dupla cruzada. O início do desenvolvimento dos sistemas estatísticos que hoje conhecemos Α respeito da análise de dados desbalanceados, realiza uma síntese na qual se pergunta sobre qual(is) método(s) terá(ão) grande utilidade, quando se trata de sistemas estatísticos.

SEARLE(1987) faz uma excelente explanação sobre análise de variância para dados desbalanceados. Num primeiro instante, comenta que o desbalanceamento

pode ser planejado ou não. Como exemplo do primeiro caso, cita os delineamentos experimentais Quadrado Latino e Blocos Incompletos Balanceados. Já a segunda situação ocorre quando se tem um experimento planejado para ser balanceado, em que há perda de observações. Neste caso, há muitos pesquisadores que priorizam a estimação das parcelas perdidas. Não é o enfoque dado pelo autor.

O autor classifica os dados desbalanceados em dois casos: um, com todas as caselas ocupadas e outro com casela(s) vazia(s). Isto por que, dependendo da forma de como se obtém as somas de quadrados para os efeitos principais, havendo interação, poderá haver diferenças. Assim, quando há desbalanceamento, é de grande importância a presença, ou não, da interação.

O autor coloca os experimentos com dados balanceados como um caso particular, na realização de uma análise de variância. Esta, segundo ele, é a forma como os sistemas estatísticos deveriam atuar.

Os usuários, entretanto devem tomar algumas precauções. No caso de desbalanceamento, em diferentes sistemas, poder-se-á encontrar diferentes somas de quadrados associadas a uma mesma causa de variação. Com isto, o pesquisador poderá ter dúvidas quanto ao significado das somas de quadrados. Mesmo por que, ficará em dúvida qual análise de variância adotar para uma boa avaliação de seu conjunto de dados, confrontando as saídas de diferentes sistemas estatísticos. Como foi

comentado, estes podem utilizar modelos distintos, onde a ordenação dos fatores pode ser importante. Com isso, há possibilidade de se obter diferentes quadros de análise de variância.

Há um confronto entre sistemas dados desbalanceados. O autor deixa claro que, com dados balanceados, para diferentes sistemas teremos ainda as mesmas somas de quadrados associadas. Havendo diferentes somas de quadrados, mas com o mesmo número de graus de liberdade, tem-se uma razão de teste F diferente para os diferentes sistemas. A diferença entre as somas de quadrados pode ser de tal ordem, que altere a conclusão sobre se há significância, ou não para o teste. Desta forma, pode-se colocar que há diferenças entre o que está sendo testado em cada sistema. Ou seja, são diferentes hipóteses testadas. As saídas dos sistemas são insuficientes para se fazer uma avaliação mais profunda a respeito das hipóteses que estão sendo testadas.

Para dirimir muitas das dúvidas geradas pelo modelo superparametrizado, o autor propõe o modelo de médias de caselas, para conjunto de dados desbalanceados. Deve-se verificar se há conectividade, nos casos de caselas vazias. O autor mostra um método simples, descrito como um método de caminhamento.

São estudados os seguintes sistemas estatísticos: BMDP, GENSTAT, SAS GLM, SAS HARVEY e SPSS-X(2.1). O autor mostra os modelos, as restrições envolvidas, bem como as diferenças quanto a forma de

obtenção das somas de quadrados. Dependendo do sistema, no caso de modelos com interação, pode-se estar incorrendo em um erro grave de avaliação dos efeitos principais.

SINGH & SINGH(1989), em um modelo com interação, colocam a proposta de realizar a análise de variância utilizando um modelo de regressão, fazendo uso da notação R( ). Os autores descrevem primeiramente os modelos de médias de caselas е 0 modelo superparametrizado expondo, por fim, o modelo regressão. Colocam a correspondência existente entre os modelos de médias, de regressão e superparametrizado. Assim, realizam a análise de variância para um conjunto de dados desbalanceados, porém com todas as caselas ocupadas.

A conectividade de dados em modelos com dados desbalanceados com caselas vazias é tratado por MURRAY(1989). Procura estender o conceito a modelos multivariados.

IEMMA(1991) relata a respeito dos métodos de YATES(1934) e sua relação com a notação R(), alguns dos quais incorporados ao SAS. O autor comenta também a respeito da conectividade do modelo.

DALLAL(1992) apresenta um exemplo em que, para um mesmo conjunto de dados(balanceado), realiza a análise de variância em dois diferentes sistemas: SAS/GLM e SPSS. Admite duas formas de realizar a análise, tomando um experimento em classificação dupla

considera inicialmente 0 resíduo. cruzada. como usualmente se admite. Num segundo instante, define o aninhado. 0 resíduo efeito resultado como um contrastante entre as somas de quadrados obtidas, para os diferentes sistemas, nos diferentes enfoques.

IEMMA et alli (1993) para conjuntos de dados desbalanceados, descreve um método para obtenção de projetores ortogonais. A princípio mostra a relação entre análise de variância e o teorema de Pitágoras. Ressalta, ainda. caso de dados desbalanceados. aue no principalmente com caselas vazias, a igualdade não se verifica. Fornece um exemplo 2 x 3, com casela vazia. Constrói os projetores ortogonais, bem como as somas de associadas às hipóteses sobre médias ponderadas de linhas; médias ponderadas de colunas ajustadas para linhas; médias nãoponderadas de linhas; interação. Apresenta também as somas de quadrados fornecidas pelo SAS/GLM. Também foi tratado a respeito de equivalência de modelos, principalmente no que se refere a modelos de posto incompleto e completo.

Para esclarecer dúvidas surgidas em DALLAL(1992), SEARLE(1994) comenta o assunto. O resíduo do primeiro modelo, é colocado como um efeito aninhado dentro da interação. O autor descreve as hipóteses utilizadas para calcular somas de quadrados do Tipo III do SAS/GLM. Compara então, os dois modelos exemplifica através de um conjunto de dados desbalanceados. Relata a respeito do manual do SAS/GLM, e a forma como constrói as funções e hipóteses do tipo III. O autor solicita dos autores de programas estatísticos que se confeccione manuais que expliquem

com maiores detalhes os procedimentos envolvidos no cálculo das somas de quadrados.

Trabalhando com dados desbalanceados com caselas vazias, SANTOS(1994) estrutura as somas de quadrados e hipóteses a elas associadas, bem como sua relação com diversos sistemas estatísticos. Para dados desbalanceados, divide em casos de todas as caselas ocupadas e com algumas caselas vazias; cada caso, subdivide o estudo em modelos com e sem interação. Conclui que, dependendo do sistema utilizado, as somas de quadrados e as hipóteses associadas aos efeitos principais, e até mesmo da interação, serão diferentes. O trabalho do autor baseia-se, no modelo de médias de caselas e nas quatro somas de quadrados fornecidas pelo SAS/GLM. Reforça a idéia da necessidade da compreensão dos procedimentos envolvidos pelos diferentes sistemas.

MONDARDO(1994) estuda a respeito de funções paramétricas estimáveis e sua relação com o PROC GLM do SAS. Ressalta que, a localização das caselas vazias, é importante na formulação de hipóteses. Quando da formulação de hipóteses, o modelo superparametrizado pode envolver parâmetros interesse. Com relação às saídas do SAS/GLM: as do Tipo I e II, podem não ter real interesse para os pesquisadores; a do Tipo III, para o caso de haver somente uma casela vazia, é de utilidade do pesquisador, mas quando houver mais de uma torna-se complexa demais; já a do Tipo IV, oferece uma hipótese de fácil interpretação, porém com perda de informação. Esta última, depende da posição das caselas vazias. Para dados desbalanceados, hipóteses de nulidade não podem ser testadas. A autora, realiza um trabalho que auxilia pesquisadores a escolherem a melhor opção dentre as somas de quadrados da saída do SAS/GLM.

Este assunto também é estudado por CAMARINHA FILHO(1995), que compara o SAS ao STATGRAPHICS, utilizando três conjuntos de dados: balanceado, desbalanceado com todas as caselas ocupadas, desbalanceado com caselas vazias. Verifica que não há diferenças entre os sistemas para o primeiro conjunto de dados. Mas que para dados desbalanceados o tratamento deve ser cuidadoso, principalmente para modelos com interação e na presença de caselas vazias. Orienta pesquisadores a conhecer mais sobre os sistemas de que fazem uso, e ressalta a falta de maior detalhamento por parte dos manuais.

## 3. METODOLOGIA

# 3.1. CARACTERIZAÇÃO

Neste estudo, consideram-se modelos lineares, de Gauss-Markov (G.M) do tipo:

$$y = Mp + \varepsilon \tag{1}$$

onde:

 $_{\it n}\, y_{\it I}$  é um vetor de realizações de variáveis aleatórias;

 $_{n}M_{q}$  é uma matriz conhecida (matriz de delineamento) de posto  $k \leq \min(n,q);$ 

 $_{q}\,p_{I}\,$  é um vetor de parâmetros desconhecidos;

 $_n arepsilon_l$ é um vetor de variáveis aleatórias não observáveis, tal que  $arepsilon \cap (\phi, I\sigma^2)$ .

Nesse contexto:

$$E(y) = Mp (2)$$

# 3.2. ESTIMAÇÃO

## 3.2.1. As idéias de RAO (1945)

O objetivo inicial é estimar o vetor p, dado em (1), ou alguma combinação linear descrita por seus elementos. RAO (1945) apresenta a seguinte definição:

Definição 1: Uma função linear paramétrica,  $\lambda' p$  é dita estimável no modelo  $y = Mp + \varepsilon$ , G.M., se e somente se existir uma combinação linear das observações, a'y, tal que  $E[a'y] = \lambda' p$ . Se não existir a, tal que a igualdade se verifique,  $\lambda' p$  é dita não estimável nesse modelo.

Utilizando a definição 1 e (2), o autor propõe o seguinte teorema:

Teorema 1: Uma condição necessária e suficiente para que a função linear paramétrica  $\lambda'p$  seja estimável no modelo  $y=Mp+\varepsilon$ , G.M., é  $\lambda\in C(M')$ .

# 3.2.2. Definições e Teoremas Importantes

Definição 2: Dado o modelo de Gauss-Markov, sejam todas as possíveis combinações a'y, tais que  $E[a'y] = \lambda'p$ . Dentre elas, seja  $a^{*'}y$ . Então,  $a^{*'}y$  é definido como o melhor estimador linear não viesado Best Linear Univariate Estimator (BLUE) de  $\lambda'p$ , se e somente se:

$$var[a*'y] = min var[a'y], \ \forall \ a'y : E[a'y] = \lambda'p$$
  
Denota-se, geralmente: BLUE de  $\lambda'p = \lambda'p$ 

- Teorema 2: Se  $\lambda'p$  é uma função estimável no modelo linear  $y=Mp+\varepsilon$ , G.M., seu BLUE pode ser obtido de modo  $\wedge$  , único, por:  $\lambda'p=\rho^o$  M'y. Onde,  $\rho^o$  é qualquer solução do sistema de equações normais associadas a  $\lambda'p$ .
- Teorema 3: Teorema de Gauss-Markov.

Se  $\lambda'p$  é uma função estimável no modelo  $y=Mp+\varepsilon$ , G.M., então seu BLUE pode ser obtido de modo único por:  $\lambda'p=\lambda'p^0$ . Onde,  $p^0$  é qualquer solução das equações normais.

Definição 3: Dada a matriz <sub>p</sub>M<sub>q</sub> de posto k, então a matriz <sub>q</sub>M<sub>p</sub>, de posto k, que satisfaz as quatro condições seguintes é definida como inversa generalizada de Moore-Penrose:

a. 
$$MM^+M = M$$

b. 
$$M^{+}MM^{+} = M^{+}$$

c. 
$$MM^{+} = (MM^{+})' = M^{+}M'$$
 (Simétrica)

d. 
$$M^{+}M = (M^{+}M)' = M'M^{+'}$$
 (Simétrica)

Definição 4: Dada uma matriz  $_{\rm p}{
m M}_{\rm q}$ , toda matriz  $_{\rm q}{
m M}_p^l$  que satisfaz às duas condições seguintes, é inversa de mínimos quadrados de M:

a. 
$$MM^lM = M$$

b. 
$$MM^{l} = (MM^{l})' = M^{l'}M'$$
 (Simétrica)

# 3.2.3. Regras Práticas

A verificação da estimabilidade de uma função paramétrica no modelo em questão através da Definição 1, ou mesmo do Teorema 1 pode ser de difícil execução na prática. A aplicação dos conceitos de estimação pode exigir por parte do pesquisador de ciências aplicadas um

certo poder de abstração, além de um tempo razoável para sua execução.

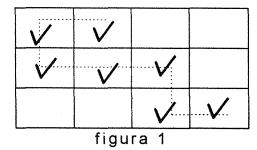
Uma regra prática é utilizar:

$$y = Mp + \varepsilon$$
, G.M.  $\Rightarrow E(y) = Mp$ 

Sendo Mp estimável, pode-se determinar as funções básicas estimáveis em cada caracterização de  $v=Mp+\varepsilon$ .

Toda combinação linear entre as funções básicas estimáveis também é estimável. Assim, através das funções básicas estimáveis, pode-se descrever outras funções estimáveis.

Nos casos em que ocorrem caselas vazias, esta regra não detecta com facilidade casos em que haja desconexão. Essa dificuldade pode ser mostrada através de um exemplo extraído de SEARLE (1987).



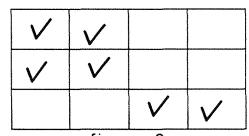


figura 2

Se for possível caminhar nas caselas, entre linhas e colunas, sem utilizar diagonais, os dados são ditos conexos, como na figura 1. Caso contrário, são ditos desconexos, como na figura 2. Essa regra é descrita como "geométrica". Maiores detalhes podem ser obtidos em MURRAY & SMITH (1985).

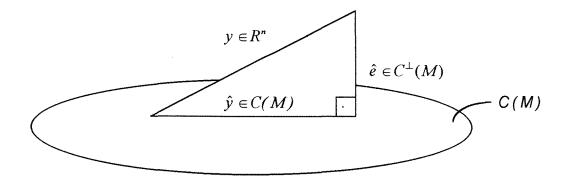
Pode-se estender o conceito da Definição 1, colocado para somente uma função estimável, para um conjunto L de funções estimáveis. Nesse contexto, tem-se para (1), que Lp oferece um conjunto de funções paramétricas estimáveis. Impõe-se que L tenha posto coluna completo.

#### 3.3. TESTABILIDADE

A noção de testabilidade está bastante próxima dos conceitos de estimabilidade e conectividade, vistos anteriormente. Para que uma hipótese seja testável, ela deve ser constituída de funções estimáveis no modelo.

# 3.4. PARTIÇÃO DAS SOMAS DE QUADRADOS

Admitindo-se o modelo (1), tem-se que  $\hat{y}=Mp^0$  é a aproximação de mínimos quadrados para o vetor de observações y, invariante para qualquer  $p^0$  solução das equações normais:  $\hat{y}=Mp^0=M(M'M)^GM'y=Py$ , onde  $P=M(M'M)^GM'=MM'=MM'$  é o projetor ortogonal de y sobre o espaço gerado pelas colunas de M, C(M). Graficamente, tem-se a seguir.



a ortogonalidade, pode-se utilizar o teorema de Pitágoras, obtendo-se a decomposição:  $\|y\|^2 = \|\hat{y}\|^2 + \|\hat{\varepsilon}\|^2$ , onde  $\|y\|^2 = y'y = \sum y^2$ , é a soma de quadrados total(não corrigida);  $\|\hat{y}\|^2 = \|Py\|^2 = y'Py = \sum \hat{y}^2 = p^{\sigma}M'y$ , soma quadrados de parâmetros;  $\|\hat{e}\|^2 = \hat{e}'\hat{e} = \sum \hat{e}^2 = y'y - p^{0'}M'y = y'(I-P)y$ , é a soma de quadrados do resíduo. Quando se tem um conjunto de dados balanceados, há ortogonalidade, obtendo-se P de forma trivial. Entretanto, em se tratando de dados desbalanceados, tal simplicidade não se verifica, tornando mais difícil a obtenção dos projetores ortogonais, no que se refere às hipóteses sobre linhas, colunas e interação, principalmente na presença de caselas vazias, IEMMA(1993).

Pode-se proceder ainda à partição da soma de quadrados dos parâmetros, pois nela, em geral, estão contidas as hipóteses que o pesquisador tem interesse de testar.

É importante a adequação do modelo aos dados, sem perder de vista a hipótese que se deseja testar

# 3.5. NOTAÇÃO R()

A idéia de utilização da notação R( ) está intimamente ligada ao método de ajuste de constantes, descrito por YATES(1934). A notação R() foi consagrada por SEARLE (1971). Tem merecido atenção de outros entre eles. pode-se citar SPEED HOCKING(1976) e IEMMA(1991) que colocam as idéias de parametrização sucessiva, através do modelo superparametrizado.

Aplicam-se estas idéias, no modelo (1). Particionando-se convenientemente o vetor de parâmetros e a matriz do delineamento, pode-se então, associar a cada modelo, do mais simples ao mais completo uma soma de quadrados de parâmetros pela notação R(). Fazendo-se, por exemplo,

$$M = \begin{bmatrix} M_1 & M_2 & M_3 & M_{23} \end{bmatrix}; \qquad p = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{23} \\ p_{23} \end{bmatrix}$$

por meio de parametrizações sucessivas, obtém-se as somas de quadrados de interesse.

Tem-se, por exemplo,

$$y = M_1 p_1 + \varepsilon^{(1)} \Rightarrow S.Q. p_1 = p_1^{0'} M_1' y = R(p_1)$$

$$y = M_2 p_2 + \varepsilon^{(2)} \Rightarrow S.Q. p_2 = p_2^{0'} M_2' y = R(p_2)$$

$$y = \begin{bmatrix} M_1 & M_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{bmatrix} \Rightarrow S.Q. [p_1, p_2] = \begin{bmatrix} p_1^{0'} & p_2^{0'} \end{bmatrix} \begin{bmatrix} M_1' \\ M_2' \end{bmatrix} y = R(p_1, p_2)$$

Pode-se obter a soma de quadrados de  $p_{1}$ , ajustado para  $p_{2}$ , denotado por  $R\!\left(p_{1}|p_{2}\right)$ , através de

$$R(p_1|p_2) = R(p_1, p_2) - R(p_2)$$

Pode-se obter a soma de quadrados de  $p_2$ , ajustado para  $p_4$ , denotado por  $R\!\left(p_2|p_I\right)$ , através de

$$R(p_2|p_1) = R(p_1, p_2) - R(p_1)$$

Repare que  $R(p_1, p_2, p_3, p_{23}) = S.Q.$  Parâmetros

#### 3.6. MODELOS UTILIZADOS

#### 3.6.1 Introdução

Os modelos utilizados aqui baseiam-se em (1). No caso de k<q, o modelo é de posto coluna incompleto, e os parâmetros não são estimáveis individualmente. Somente funções paramétricas serão estimáveis. Se k=q, o modelo é de posto coluna completo, e todos os parâmetros são estimáveis individualmente.

Dados os propósitos deste estudo, descrevemse aqui, modelos de posto completo como alternativas ao modelo superparametrizado, de posto incompleto. Esses, resultam de reparametrizações e/ou restrições impostas a um modelo de posto incompleto.

## 3.6.2. Modelo Superparametrizado (Modelo-S)

Segundo SEARLE(1987), um modelo é dito superparametrizado, caso explicite um parâmetro para cada um dos fatores envolvidos. Esse modelo, doravante denominado Modelo-S, é aqui denotado por

$$y = X\theta + e \tag{3.1}$$

onde:

 $_{n}\,y_{I}$  é um vetor de realizações de variáveis aleatórias;

 $_{n}X_{p}$  é uma matriz conhecida(matriz de delineamento) de posto k < min(n,p);

 $_{p} heta_{I}$  é um vetor de parâmetros desconhecidos;

 $_{n}e_{1}$  é um vetor de variáveis aleatórias não observáveis, tal que  $e \cap (\phi,I\sigma^{2})$ .

Utiliza-se aqui a idéia de parametrização sucessiva, vista anteriormente. Assim, por exemplo, no em um modelo com dois fatores fixos (A e B) com interação (AB) tem-se, parametrizando sucessivamente,

$$y = X_I \theta_I + e^{(I)} \Rightarrow y_i = \mu + e_i^{(I)}$$
 (S.1)

$$y = X_2 \theta_2 + e^{(2)} \Rightarrow y_{ij} = \mu + \alpha_i + e_{ij}^{(2)}$$
 (S.2)

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

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

O modelo completo, com interação, terá a seguinte caracterização:

$$y = X\theta + e \Rightarrow y_{iik} = \mu + \alpha_i + \beta_i + \gamma_{ii} + e_{iik}$$
 (S.5)

$$i=1, 2, ..., a; j=1, 2, ..., b; k=1, 2, ..., n_{ij}$$

Onde,  $\boldsymbol{a}$  e  $\boldsymbol{b}$  são os números de níveis dos fatores  $\boldsymbol{A}$  e  $\boldsymbol{B}$ , respectivamente, e  $\boldsymbol{n}_{ij}$  o número de observações na casela (i,j).

Conforme já visto, o sistema de equações lineares  $y = X\theta$  é inconsistente, entretanto, o sistema de equações normais(SEN), que é dado por  $X'X\theta = X'y$ , é sempre consistente. Qualquer solução exata do SEN é solução aproximada de mínimos quadrados para  $y = X\theta$ , inconsistente. Dessa forma, pode-se estimar **e**, que é um vetor de variáveis aleatórias não observáveis.

Como X é uma matriz com posto coluna incompleto, tem-se que XX não é não-singular, resultando em um SEN indeterminado. As soluções exatas podem ser obtidas , dentre outras, através de  $\theta^0 = (XX)^G XY$ , onde  $(XX)^G$  é uma inversa generalizada qualquer de XX.

Pode-se colocar as somas de quadrados em termos da notação R(). Alguns exemplos:

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

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

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

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

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

Apesar de ser de uso bastante difundido no meio dos pesquisadores de ciências aplicadas, o Modelo-S possui alguns inconvenientes. Dentre eles, o de possuir mais parâmetros do que o número de caselas para estimálos. Segundo HOCKING(1985), SEARLE(1987) e IEMMA(1993), esse problema se agrava nos casos em que há caselas vazias.

# 3.6.3. Modelo de Médias de Caselas (Modelo-M)

O modelo de médias de caselas, denominado de Modelo-M, tem a seguinte caracterização:

$$y = W\mu + e \tag{3.2}$$

onde,

<sub>n</sub> W<sub>k</sub> é uma matriz conhecida;

 $_{\it k}\mu_{\it 1}$  é um vetor de médias das populações das quais são retiradas as amostras aleatórias que compõem as caselas;

 $_{n}y_{1}$  e  $_{n}e_{1}$  são como foi descrito para (3.1).

Assim, em correspondência ao modelo (3.1), para o caso de dois fatores fixos com interação, tem-se respectivamente, através de parametrizações sucessivas,

$$y = W_1 \mu_1 + e^{(1)} \Rightarrow y_i = \mu + e_i^{(1)}$$
 (M.1)

$$y = W_2 \mu_2 + e^{(2)} \Rightarrow y_{ij} = \mu_i + e_{ij}^{(2)}$$
 (M.2)

$$y = W_3 \mu_3 + e^{(3)} \Rightarrow y_{ij} = \mu_j + e_{ij}^{(3)}$$
 (M.3)

$$y = W_4 \mu_4 + e^{(4)} \Rightarrow y_{ijk} = \mu_i + \beta_j + e^{(4)}_{ijk}$$
 (M.4)

O modelo completo, com interação, terá a seguinte caracterização:

$$y = W\mu + e \Rightarrow y_{ijk} = \mu_{ij} + e_{ijk}$$
 (M.5)

À exceção do (M.4), todos os modelos têm posto coluna completo

Os índices obedecem a mesma caracterização utilizada no Modelo-S.

Esse modelo possui vantagens em relação ao Modelo-S. Dentre elas, pode-se citar a maior simplicidade e clareza, ao descrever as hipóteses envolvidas. Sendo, em geral, de posto coluna completo, apresenta solução única para o SEN  $W'W\mu=W'y$ . Assim, a solução exata é dada por:  $\mu^0=(W'W)^{-1}W'y$ , que é a solução aproximada para o sistema inconsistente  $y=W\mu$ , coincide com o estimador de mínimos quadrados para  $\mu$ .

## 3.6.4. Modelo com Restrição $\Sigma$ (Modelo- $\Sigma$ )

O modelo com restrição  $\Sigma$ , aqui denominado Modelo- $\Sigma$  é superparametrizado. Tem a seguinte caracterização:

$$y = Z\dot{\theta} + e \tag{3.3}$$

onde,

 $_{n}Z_{k}$  é uma matriz conhecida(de posto coluna completo k);  $_{k}\dot{\theta}_{i}$ é um vetor de parâmetros desconhecidos  $_{n}y_{1}$  e  $_{n}e_{1}$  são como foi descrito para (3.1).

Este modelo que é semelhante ao Modelo-S, porém com algumas restrições paramétricas do tipo "soma igual a zero". Assume a seguinte caracterização:

$$y_{ijk} = \dot{\mu} + \dot{\alpha}_i + \dot{\beta}_j + \dot{\gamma}_{ij} + e_{ijk}$$

onde,

$$\dot{\mu} = \mu + \frac{\sum_{i} \alpha_{i}}{a} + \frac{\sum_{j} \beta_{j}}{b} + \frac{\sum_{i,j} \gamma_{ij}}{ab};$$

$$\dot{\alpha}_{i} = \alpha_{i} - \frac{\sum_{i} \alpha_{i}}{a};$$

$$\dot{\beta}_{j} = \beta_{j} - \frac{\sum_{j} \beta_{j}}{b};$$

$$\dot{\gamma}_{ij} = \gamma_{ij} - \frac{\sum_{i} \gamma_{ij}}{a} - \frac{\sum_{j} \gamma_{ij}}{b} + \frac{\sum_{i,j} \gamma_{ij}}{ab}$$

Resultando nas seguintes restrições.

$$\sum_{i} \dot{\alpha}_{i} = 0; \quad \sum_{j} \dot{\beta}_{j} = 0; \quad \sum_{i} \dot{\gamma}_{ij} = 0, \quad \forall \quad j; \quad \sum_{j} \dot{\gamma}_{ij} = 0, \quad \forall \quad i$$

Os índices obedecem a mesma caracterização utilizada no Modelo-S.

Esse modelo apresenta solução para o SEN,  $Z'Z\dot{\theta}=Z'y$ , dada por  $\dot{\theta}^o=(Z'Z)^{-1}Z'y$ , que é solução aproximada para o sistema inconsistente  $y=Z\dot{\theta}$ .

3.6.5. Modelo com Restrição Zero (Modelo-Z)

O modelo com restrição Zero, denominado aqui Modelo-Z é um modelo superparametrizado. Tem a seguinte caracterização:

$$y = A\pi + e \tag{3.4}$$

onde,

 $_{n}A_{k}$  é uma matriz conhecida (de posto coluna completo k);  $_{k}\pi_{l}$  é um vetor de parâmetros desconhecidos  $_{n}y_{1}$  e  $_{n}e_{1}$  são como foi descrito para (3.1).

Esse modelo é semelhante ao Modelo-S, porém com alguns parâmetros tomados como nulos. Assume a seguinte caracterização.

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

Sem perda de generalidade, pode-se fazer o primeiro nível igual a zero (poderia ser qualquer nível). Resulta nas seguintes restrições:

$$\alpha_{I} = 0$$
,  $\gamma_{Ij} = 0$ ,  $\forall$  j;

$$\beta_I = 0$$
,  $\gamma_{iI} = 0$ ,  $\forall$  i.

Os índices obedecem à mesma caracterização utilizada no Modelo-S.

Esse modelo apresenta solução para o SEN,  $A'A\pi=A'y$ , dada por  $\pi^o=(A'A)^{-l}\,A'y$ , que é solução aproximada de mínimos quadrados para o sistema inconsistente  $y=A\pi$ .

# 3.6.6. Modelo com Restrição Ponderada (Modelo-W)

Esse modelo é superparametrizado. É denominado aqui de Modelo-W. Tem a seguinte caracterização:

$$y = T\omega + e \tag{3.5}$$

onde,

 $_{\rm n}\mathcal{T}_{\it k}$  é uma matriz conhecida(de posto coluna completo k);  $_{\it k}\omega_{\it l}$  é um vetor de parâmetros desconhecidos

 $_{n}y_{1}$  e  $_{n}e_{1}$  são como foi descrito para (3.1).

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

Com as seguintes restrições:

$$\sum_{i} n_i \alpha_i = 0 ;$$

$$\sum_{j} n_{j} \beta_{j} = 0 \; ; \qquad$$

$$\sum_{i} n_{ij} \gamma_{ij} = 0, \ \forall \ j;$$

$$\sum_{i} n_{ij} \gamma_{ij} = 0, \ \forall \ i$$

Onde:  $n_i$ =número de observações para o índice i;  $n_j$ =número de observações para o índice j;  $n_{ij}$ =número de observações para os índices (i,j)

Os índices obedecem à mesma caracterização utilizada no Modelo-S.

Esse modelo apresenta solução para o SEN,  $T'T\omega=T'y$ , dada por  $\omega^\circ=(T'T)^{-1}T'y$ , que é solução aproximada de mínimos quadrados para o sistema inconsistente  $y=T\omega$ .

### 3.6.7 Modelo de Regressão (Modelo-R)

O modelo clássico de regressão, denominado Modelo-R, pode ser utilizado em delineamentos experimentais. Para tanto, faz-se uso de uma conveniente aplicação de variáveis binárias do tipo 0 e 1. Tem a seguinte caracterização:

$$y = U\delta + e \tag{3.6}$$

onde,

 $_{n}U_{k}$  é uma matriz conhecida(de posto coluna completo k);  $_{k}\delta_{I}$  é um vetor de parâmetros desconhecidos  $_{n}y_{I}$  e  $_{n}e_{I}$  são como foi descrito para (3.1).

Utiliza-se o mesmo exemplo utilizado para o Modelo-S. Fixando-se, por exemplo, o primeiro nível do fator A e o primeiro nível do fator B, pode-se tomar a seguinte variável indicadora para o Modelo-R, na ordem A-B:

Para o fator A,

$$u_{ij1} = \begin{cases} 1, \text{ se } i = 2(\text{nivel 2 do fator A}) \\ 0, \text{ caso contrario} \end{cases}$$

••

$$u_{ij(a-1)} = \begin{cases} 1, \text{ se } i = a(\text{nivel a do fator A}] \\ 0, \text{ caso contrario} \end{cases}$$

Para o fator B,

$$u_{ija} = \begin{cases} 1, se \ j = 2 \text{(nivel 2 do fator B)} \\ 0, caso \ contrario \end{cases}$$

••

$$u_{ij(a+b-2)} = \begin{cases} 1, & \text{se } j = b \text{(nivel } b \text{ do } fator B) \\ 0, & \text{caso } contrario \end{cases}$$

Tem-se então, a seguinte caracterização:

$$y_{ij} = \delta_0 + \underbrace{\sum_{r=1}^{a-1} u_{ijr} \delta_r}_{fator A} + \underbrace{\sum_{s=a}^{a+b-2} u_{ijs} \delta_s}_{fator B} + \underbrace{\sum_{r=1}^{a-1} \sum_{s=a}^{a+b-2} u_{ijr} u_{ijs} \delta_{rs-v}}_{A \times B} + e_{ijk}$$

Onde,  $v = n \acute{u}mero de casela(s) vazia(s)$ .

É evidente que se pode fixar qualquer nível do fator A. Esse modelo torna-se melhor utilizado quando é fixado um nível tido como padrão ou testemunha. Ao final, os demais níveis são descritos como contrastes com o padrão, descrito na forma  $\delta_i = \alpha_i - \alpha_a$ , sendo a o nível fixado.

Esse modelo apresenta solução para o SEN,  $U'U\delta=U'y$ , dada por  $\delta^0=(U'U)^{-1}U'y$ , que é solução aproximada de mínimos quadrados para o sistema inconsistente  $y=U\delta$ .

## 3.7. HIPÓTESES ESTATÍSTICAS

As hipóteses estatísticas desenvolvidas nesse estudo tem como base, aquelas apresentadas por SPEED, HOCKING & HACKNEY(1978). Considera-se aqui o modelo de médias de caselas, pois, conforme visto anteriormente, ele simplifica a interpretação das hipóteses estatísticas.

a) Hipóteses sobre linhas (fator A):

a.1)Hipótese sobre médias ponderadas de linhas (Hipótese do Tipo I).

Para esse tipo de hipótese não é feito qualquer ajuste, ocorrendo na forma explícita da hipótese, parâmetros relativos aos efeitos da interação e do fator B:

$$H_0^I: \overline{\mu}_i = \overline{\mu}_{i'}$$

a.2)Hipótese sobre médias ponderadas de linhas, ajustadas para colunas (Hipótese do Tipo II):

Para essa hipótese, é feito um ajuste, de modo a eliminar da forma explícita os efeitos dos parâmetros do fator B.

$$H_0^2: \sum_j \frac{n_{ij} \mu_{ij}}{n_{i.}} = \sum_j \frac{n_{i'j} \mu_{i'j}}{n_{i'.}}$$

a.3)Hipótese sobre médias não ponderadas de linhas(Hipótese do Tipo III):

Para essa hipótese não são levadas em consideração as frequências das caselas.

$$H_0^3: \sum_{j} (n_{ij}) \mu_{ij} = \sum_{i'} \sum_{j} \frac{n_{ij} n_{i'j} \mu_{i'j}}{n_{i'}}$$

a.4)Hipótese sobre médias não ponderadas de linhas(Hipótese do Tipo IV):

Semelhante a anterior, dependendo da posição da(s) casela(s) vazia(s).

$$H_0^4: \mu_{iI} = \mu_{i'I}$$

## b) Hipóteses sobre colunas:

b.1)Hipótese sobre médias ponderadas de colunas (Hipótese do Tipo I).

Para esse tipo de hipótese não é feito qualquer ajuste, ocorrendo na forma explícita da hipótese, parâmetros relativos aos efeitos da interação e do fator A:

$$H_0^5: \overline{\mu}_{.i} = \overline{\mu}_{.i'}$$

b.2)Hipótese sobre médias ponderadas de colunas, ajustadas para linhas (Hipótese do Tipo II):

Para essa hipótese, é feito um ajuste, de modo a eliminar da forma explícita os efeitos dos parâmetros do fator A.

$$H_0^6: \sum_{i} \frac{n_{ij} \mu_{ij}}{n_{.i}} = \sum_{i} \frac{n_{ij'} \mu_{ij'}}{n_{.i'}}$$

b.3)Hipótese sobre médias não ponderadas de colunas (Hipótese do Tipo III):

Para essa hipótese não são levadas em consideração as frequências das caselas.

$$H_o^7: \sum_{i} (n_{ij}) \mu_{ij} = \sum_{j'} \sum_{i} \frac{n_{ij} n_{ij'} \mu_{ij'}}{n_{i.}}$$

b.4)Hipótese sobre médias não ponderadas de colunas (Hipótese do Tipo IV):

Semelhante a anterior, dependendo da posição da(s) casela(s) vazia(s).

$$H_0^8: \mu_{Ii} = \mu_{Ii'}$$

c) Hipótese sobre a interação:

$$H_0^9: \mu_{ii} - \mu_{i'i} - \mu_{ii'} + \mu_{i'i'} = 0$$

Em um conjunto de dados com todas as caselas ocupadas, a hipótese sobre interação testa se todas as interações são nulas. Para um conjunto de dados com casela(s) vazia(s) há interações que não fazem parte da hipótese. Nesse contexto, não se pode propor a hipótese sobre interação como hipótese para verificar a adequação do modelo com interação ao conjunto de dados.

De modo geral, qualquer das hipóteses anteriores pode ser posta na forma  $H_o$ : k'p=c vs.  $H_i$ :  $k'p\neq c$ , sendo que, k'p deve ser uma função paramétrica estimável no modelo proposto. Geralmente na análise de variância, faz-se para as hipóteses anteriores, c=0.

Assim, a soma de quadrados associada a essas hipóteses fica:

$$SQH_0 = (k'p^0 - c)'[k'(M'M)^G k]^{-1}(k'p^0 - c)$$
 (3)

Considerando-se a hipótese de nulidade,  $H_{g}\colon k'p=\phi$ , obtém-se:

$$SQH_0 = (k'p^0)'[k'(M'M)^G k]^{-1}(k'p^0)$$
 (4)

## 3.8. SISTEMAS ESTATÍSTICOS

A seguir são expostos cinco sistemas estatísticos, que utilizam modelos de posto coluna completo para realizar Análise de Variância. Cada um deles é identificado com os modelos descritos anteriormente.

#### 3.8.1. SAS/PROC GLM

Dentro do SAS( $\mathbf{S}$ tatistical Analysis  $\mathbf{S}$ ystem), existe um procedimento para realizar Análise de Variância de dados desbalanceados denominado  $\mathrm{GLM}(\mathbf{G}$ eneral Linear Models). Com base nos modelos superparametrizados Modelo- $\mathbf{S}$  e Modelo- $\mathbf{\Sigma}$ , fornece quatro formas de execução

das Análise de Variância, denominadas SQ TIPO I, SQ TIPOII, SQ TIPO III e SQ TIPO IV. Se o conjunto de dados for balanceado não há diferença entre elas. SEARLE(1987) coloca que podem haver diferenças entre as somas de quadrados obtidas pelo SAS. No modelo com interação, em caselas vazias, presença de as quatro somas de quadrados diferem entre si. Nesse último, a SQ TIPO IV tem seu valor alterado em função da posição da(s) casela(s) vazia(s). Sendo que, a SQ TIPO I depende da impuser aos fatores, nos ordenação que se casos desbalanceados.

Segundo SEARLE(1987) para obtenção da SQ TIPO~III, é utilizado o modelo com restrição do tipo  $\Sigma$ =0. Para as outras é utilizada uma inversa generalizada denominada G2. Maiores detalhes, MONDARDO(1994).

#### 3.8.2. BMDP

O BMDP(BioMeDical Statistic Package), que teve seu início de desenvolvimento nos anos 60, possui diversas versões. SEARLE(1987) comenta que nos módulos 1V, 2V e 4V o "default" utiliza o Modelo- $\Sigma$ . Entretanto, o autor comenta que existe uma opção que permite realizar a ANOVA pelo Modelo-W. As hipóteses são formuladas a partir do modelo de média de caselas.

A versão mais recente do BMDP, 10V, faz uso do Modelo-M. Desta forma, um conjunto de dados desbalanceados possui um enfoque diferenciado, em relação aos modelos utilizados por outros sistemas. Isto por que as hipóteses são formuladas em relação à média populacional das caselas, facilitando sua interpretação, DIXON(1989).

Mesmo quando há freqüências diferentes para as caselas, as hipóteses testadas pelo BMDP são as mesmas que outros programas apresentam quando há um número constante de observações em cada casela, SANTOS(1994). Segundo o mesmo autor, no caso de haver caselas vazias, o programa não proporciona saída para a Análise de Variância.

#### 3.8.3. GLIM

O GLIM(Generalized Linear Interative Modeling System) possui uma saída semelhante àquela oferecida SQ TIPO I do SAS. Ou seja, depende da ordenação dada aos fatores principais, obtendo as somas de quadrados, ajustando em relação ao fator colocado anteriormente, SEARLE(1987). As somas de quadrados para os fatores principais independe da presença ou não da interação no modelo. O GLIM utiliza para realização da Análise de Variância o Modelo-Z, GLIM(1985).

#### 3.8.4. SPSS/ANOVA

0 SPSS(Statistical Package for Social Modelo-W Science) utiliza na Análise de Variância, quadrados SEARLE(1987). As somas de são feitas ajustando-se um fator em relação ao outro. Entretanto, as saídas ficam prejudicadas quando do desbalanceamento na presença de casela vazia. Nesse caso, dependendo da ordenação que se der aos efeitos principais, não apresentará saída para o segundo efeito principal.

Segundo SPSS/PC<sup>+</sup>(1986), o sistema estatístico oferece duas saídas, dependendo da opção escolhida: "UNIQUE" ou "SEQUENTIAL". SEARLE(1987), alerta que a primeira oferece uma saída semelhante à do Tipo III do

SAS; enquanto a segunda é seqüencial, dependendo da ordem de entrada dos fatores.

#### 3.8.5. MINITAB

0 MINITAB (MINITAB STATISTICAL SOFTWARE), proporciona de somas quadrados semelhantes às do TIPO I fornecidas pelo SAS. É feito um ajuste sequencial em relação ao fator precedente e portanto, dependerá da forma como são ordenados os fatores principais no modelo. Segundo SANTOS(1994), é possível ainda obter um quadro de Análise de Variância semelhante ao fornecido pelo SAS, SQ TIPO II. O MINITAB realiza a Análise de Variância utilizando um modelo de regressão, MINITAB(1991).

## 4. ASPECTOS PRÁTICOS IMPORTANTES

# 4.1. INTRODUÇÃO

A análise de um mesmo conjunto dados pode ser realizada através de diversos modelos. A verificação da adequação do modelo ao conjunto de dados deve ser feita com critério e rigor. Em particular, havendo caselas vazias. Nesse caso, comparando diferentes modelos utilizados para a análise, pode haver diferenças entre as somas de quadrados de hipóteses para os efeitos principais. Há uma evidência de que, entre os modelos, as hipóteses testadas podem ser diferentes.

Esse fato pode ser levado em consideração ao comparar sistemas estatísticos. Pesquisadores utilizando diferentes sistemas estatísticos podem querer comparar resultados obtidos em suas pesquisas. Para tanto, é necessário saber qual modelo utilizado e hipóteses testadas.

Nesse contexto, é importante evidenciar semelhanças e diferenças entre modelos, utilizados por alguns sistemas estatísticos, na formulação e interpretação das hipóteses testadas na análise de dados. Especificamente, no que se refere aos efeitos principais.

Com esses objetivos, o exemplo utilizado evidencia diferenças e semelhanças entre os modelos e sistemas. Colocam-se também as dificuldades existentes,

tendo os modelos de posto completo como alternativa ao modelo superparametrizado de posto incompleto.

O exemplo trata de uma simulação para um experimento em cana. São dois os fatores envolvidos: variedades de cana e fórmulas de adubação. Para o primeiro fator são três cultivares de cana, denominados C1, C2 e C3. Para o segundo fator, são quatro fórmulas, denominadas F1, F2, F3 e F4. São duas repetições em um delineamento com dois fatores cruzados com interação. A variável observada é toneladas de cana/ha.

É importante observar que o primeiro nível de cada fator é a ausência de adubo, nos quais a dose de cada um dos nutrientes é zero.

Tem-se que, i e j são os índices atribuídos aos níveis dos fatores A(dose de nitrogênio) e B(dose de fósforo), respectivamente, e  $n_{ij}$  o número de observações na casela (i,j).

Tabela 2: Produtividade de Tonelada de Cana/ha

		F	ATOR B	(FÓRMULAS	3)	
		j=1	j=2	j=3	j=4	
		45	50	60	70	
	i = 1	~ , <del>~</del>	55	65	75	
		45(1)	105(2)	125(2)	145(2)	420(7)
		45,00	52,50	62,50	72,50	60,00
FATOR A	***************************************	50	70	** , ***	80	
	i=2	60	65		90	
(VARIEDADES)		110(2)	135(2)	* ***	170(2)	415(6)
		55,00	67,50	The second secon	85,00	69,17
		75	75	85		
•	i=3	70	-,-	80		
		145(2)	75(1)	165(2)	<b>*</b>	385(5)
		72,50	75,00	82,50		77,00
		300(5)	315(5)	290(4)	315(4)	1220(18)
		60,00	63,00	72,50	78,75	67,77

Fonte: Dados fictícios

#### 4.2. MODELOS UTILIZADOS

#### 4.2.1. Modelo-S

# 4.2.1.1. Caracterização

No Modelo-S, foram tomadas duas ordens para os efeitos dos fatores principais.

Para a ordenação A-B

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

Para a ordenação B-A

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

A ordem dos fatores no modelo tem grande importância na formulação das hipóteses do Tipo I, que é seqüencial. O primeiro fator não é ajustado, enquanto o segundo fator é ajustado em relação ao primeiro. Essa informação será utilizada mais adiante.

Para a ordem A-B, tem-se matricialmente.

Para a ordem B-A, as colunas 2, 3, e 4 (fator A) trocam de posição com as colunas 5, 6, 7 e 8 (fator B). Consequentemente, esse resultado se faz sentir nas colunas da interação.

# 4.2.1.2. Procedimento GLM do Sistema Estatístico SAS

Conforme mencionado anteriormente, o SAS é construído para modelos superparametrizados.

A macro para realizar a análise é dada a seguir.

```
DATA DESBCVAZ;
  DO A=1 TO 3;
    DO B=1 TO 4;
      DO REP=1 TO 2;
       INPUT TON @;
       OUTPUT:
      END;
    END;
  END;
CARDS:
45 . 50 55 60 65 70 75
50 60 70 65 . . 80 90
75 70 75 . 85 80 . .
PROC PRINT;
RUN;
PROC GLM;
   CLASS A B:
   MODEL TON=A B A*B/
   E E1 E2 E3 E4
   SS1 SS2 SS3 SS4
   XPX I;
RUN;
PROC GLM:
   CLASS B A;
   MODEL TON=B A A*B/
   E E1 E2 E3 E4
   SS1 SS2 SS3 SS4
   XPX I:
RUN;
```

O primeiro PROC GLM é para a análise na ordem A-B. O segundo PROC GLM é para a análise na ordem B-A. As Tabelas de 3 a 10 mostram as funções estimáveis do SAS. As Tabelas 15, 16, 17 e 18 mostram as análises de variância que o SAS fornece através do PROC GLM.

Tabela 3: Funções estimáveis do Tipo I do SAS para ordem A-B

		Α	В	AB
μ		0	0	0
A	1	L 2	0	0
	2	L 3	0	0
	3	-L2-L3	0	0
В	1	-0.2571*L2-0.0667*L3	L5	0
	2	0.0857*L2+0.1333*L3	L6	0
	3	-0.1143*L2-0.4*L3	L 7	0
	4	0.2857*L2+0.3333*L3	-L5-L6-L7	0
ΑB	1 1	0.1429*L2	0.2648*L5+0.0228*L6+0.0091*L7	L 9
	1 2	0.2857*L2	0.0731*L5+0.4201*L6-0.032*L7	L10
	1 3	0.2857*L2	0.1416*L5+0.0639*L6+0.6256*L7	L11
	1 4	0.2857*L2	-0.4795*L5-0.5068*L6-0.6027*L7	-L9-L10-L11
	2 1	0.3333*L3	0.4886*L5+0.0594*L6+0.2237*L7	L13
	2 2	0.3333*L3	0.032*L5+0.4338*L6+0.1735*L7	-L9-L10-L11-L13
	2 4	0.3333*L3	-0.5205*L5-0.4932*L6-0.3973*L7	L9+L10+L11
	3 1	-0.4*L2-0.4*L3	0.2466*L5-0.0822*L6-0.2329*L7	-L9-L13
	3 2	-0.2*L2-0.2*L3	-0.105*L5+0.1461*L6-0.1416*L7	L9+L11+L13
	3 3	-0.4*L2-0.4*L3	-0.1416*L5-0.0639*L6+0.3744*L7	-L11

Tabela 4: Funções estimáveis do Tipo I do SAS para ordem B-A

		В	A	АВ
μ		0	0	0
В	1	L2	0	0
	2	L3	0	0
	3	L 4	0	0
	4	- L 2 - L 3 - L 4	0	0
Α	1	-0.3*L2-0.1*L3	L 6	0
	2	-0.1*L2-0.1*L3-0.5*L4	L 7	0
	3	0.4*L2+0.2*L3+0.5*L4	-L6-L7	0
AB	1 1	0.2*L2	0.21*L6+0.0183*L7	L9
	1 2	0.4*L3	0.2648*L6-0.0639*L7	L 1 0
	1 3	0.5*L4	0.3881*L6+0.2511*L7	L11
	1 4	-0.5*L2-0.5*L3-0.5*L4	0.137*L6-0.2055*L7	-L9-L10-L11
	2 1	0.4*L2	0.1461*L6+0.4475*L7	L13
	2 2	0.4*L3	-0.0091*L6+0.347*L7	-L9-L10-L11-L13
	2 4	-0.5*L2-0.5*L3-0.5*L4	-0.137*L6+0.2055*L7	L9+L10+L11
	3 1	0.4*L2	-0.3562*L6-0.4658*L7	-L9-L13
	3 2	0.2*L3	-0.2557*L6-0.2831*L7	L9+L11+L13
	3 3	0.5*L4	-0.3881*L6-0.2511*L7	-L11

Tabela 5: Funções estimáveis do Tipo II do SAS para ordem A-B

		Α	В	АВ
μ		0	0	0
Α	1	L 2	0	0
	2	L3	0	0
	3	-L2-L3	0	0
В	1	0	L5	0
	2	0	L 6	0
	3	0	L7	0
	4	0	-L5-L6-L7	0
AB	1 1	0.21*L2+0.0183*L3	0.2648*L5+0.0228*L6+0.0091*L7	L 9
	1 2	0.2648*L2-0.0639*L3	0.0731*L5+0.4201*L6-0.032*L7	L 1 0
	1 3	0.3881*L2+0.2511*L3	0.1416*L5+0.0639*L6+0.6256*L7	L11
	1 4	0:137*L2-0.2055*L3	-0.4795*L5-0.5068*L6-0.6027*L7	-L9-L10-L11
	2 1	0.1461*L2+0.4475*L3	0.4886*L5+0.0594*L6+0.2237*L7	L13
	2 2	-0.0091*L2+0.347*L3	0.032*L5+0.4338*L6+0.1735*L7	-L9-L10-L11-L13
	2 4	-0.137*L2+0.2055*L3	-0.5205*L5-0.4932*L6-0.3973*L7	L9+L10+L11
	3 1	-0.3562*L2-0.4658*L3	0.2466*L5-0.0822*L6-0.2329*L7	-L9-L13
	3 2	-0.2557*L2-0.2831*L3	-0.105*L5+0.1461*L6-0.1416*L7	L9+L11+L13
133344444	3 3	-0.3881*L2-0.2511*L3	-0.1416*L5-0.0639*L6+0.3744*L7	-L11

Tabela 6: Funções estimáveis do Tipo II do SAS para ordem B-A

m.nontaration.		В	А	АВ
μ		0	0	0
В	1	L 2	0	0
	2	L 3	0	0
	3	L 4	0	0
	4	-L2-L3-L4	0	0
Α	1	0	L6	0
	2	0	L7	0
	3	0	-L6-L7	0
AB	1 1	0.2648*L2+0.0228*L3+0.0091*L4	0.21*L6+0.0183*L7	L9
	1 2	0,0731*L2+0,4201*L3-0,032*L4	0.2648*L6-0.0639*L7	L 1 0
	1 3	0.1416*L2+0.0639*L3+0.6256*L4	0.3881*L6+0.2511*L7	L11
	1 4	-0.4795*L2-0.5068*L3-0.6027*L4	0.137*L6-0.2055*L7	-L9-L10-L11
	2 1	0.4886*L2+0.0594*L3+0.2237*L4	0.1461*L6+0.4475*L7	L13
	2 2	0.032*L2+0.4338*L3+0.1735*L4	-0.0091*L6+0.347*L7	-L9-L10-L11-L13
	2 4	-0.5205*L2-0.4932*L3-0.3973*L4	-0.137*L6+0.2055*L7	L9+L10+L11
	3 1	0.2466*L2-0.0822*L3-0.2329*L4	-0.3562*L6-0.4658*L7	-L9-L13
	3 2	-0,105*L2+0,1461*L3-0,1416*L4	-0.2557*L6-0.2831*L7	L9+L11+L13
	3 3	-0.1416*L2-0.0639*L3+0.3744*L4	-0.3881*L6-0.2511*L7	-L11

Tabela 7: Funções estimáveis do Tipo III do SAS para ordem A-B

***************************************		Α	В	АВ
μ		0	0	0
A	l	L 2	0	0
	2	L3	0	0
	3	-L2-L3	0	0
В	1	0	L 5	0
	2	0	L 6	0
	3	0	L 7	0
	4	0	-L5-L6-L7	o
AB	1 1	0.2857*L2	0.381*L5+0.0476*L6	L 9
	1 2	0.2857*L2	0.0476*L5+0.381*L6	L 1 0
	1 3	0.3143*L2+0.2*L3	0.0857*L5+0.0857*L6+0.6*L7	L11
	1 4	0:1143*L2-0.2*L3	-0.5143*L5-0.5143*L6-0.6*L7	-L9-L10-L11
	2 1	0.0571*L2+0.4*L3	0.4095*L5+0.0762*L6+0.2*L7	L13
	2 2	0.0571*L2+0.4*L3	0.0762*L5+0.4095*L6+0.2*L7	-L9-L10-L11-L13
	2 4	-0.1143*L2+0.2*L3	-0.4857*L5-0.4857*L6-0.4*L7	L9+L10+L11
	3 1	-0.3429*L2-0.4*L3	0.2095*L5-0.1238*L6-0.2*L7	-L9-L13
	3 2	-0.3429*L2-0.4*L3	-0.1238*L5+0.2095*L6-0.2*L7	L9+L11+L13
	3 3	-0.3143*L2-0.2*L3	-0.0857*L5-0.0857*L6+0.4*L7	-L11

Tabela 8: Funções estimáveis do Tipo III do SAS para ordem B-A

**************************************		В	Α	AB
μ	***************************************	0	0	0
В	i	L 2	0	0
	2	L3	O	0
	3	L 4	0	0
	4	-L2-L3-L4	0	0
Α	1	0	L 6	0
	2	0	L 7	0
	3	0	-L6-L7	0
AB	1 1	0.381*L2+0.0476*L3	0.2857*L6	L 9
	1 2	0.0476*L2+0.381*L3	0.2857*L6	L10
	1 3	0.0857*L2+0.0857*L3+0.6*L4	0.3143*L6+0.2*L7	L11
	1 4	-0.5143*L2-0.5143*L3-0.6*L4	0.1143*L6-0.2*L7	-L9-L10-L11
	2 1	0.4095*L2+0.0762*L3+0.2*L4	0.0571*L6+0.4*L7	L13
	2 2	0.0762*L2+0.4095*L3+0.2*L4	0.0571*L6+0.4*L7	-L9-L10-L11-L13
	2 4	-0.4857*L2-0.4857*L3-0.4*L4	-0.1143*L6+0.2*L7	L9+L10+L11
	3 1	0.2095*L2-0.1238*L3-0.2*L4	-0.3429*L6-0.4*L7	-L9-L13
	3 2	-0.1238*L2+0.2095*L3-0.2*L4	-0.3429*L6-0.4*L7	L9+L11+L13
	3 3	-0.0857*L2-0.0857*L3+0.4*L4	-0.3143*L6-0.2*L7	-L11

Tabela 9: Funções estimáveis do Tipo IV do SAS para ordem A-B

		A	В	АВ
μ		0	0	0
A	1	L 2	0	0
	2	L 3	0	0
	3	-L2-L3	0	0
В	1	0	L 5	0
	2	0	L 6	0
	3	0	L 7	0
	4	0	-L5-L6-L7	0
AB	1 1	0.3333*L2	0.5*L5	L 9
	1 2	0.3333*L2	0.5*L6	L 1 0
	1 3	0.3333*L2	L7	LII
	1 4	0	-0.5*L5-0.5*L6-L7	-L9-L10-L11
	2 1	0.5*L3	0.5*L5	L13
	2 2	0.5*L3	0.5*L6	-L9-L10-L11-L13
	2 4	0	-0.5*L5-0.5*L6	L9+L10+L11
	3 1	-0.3333*L2-0.5*L3	0	-L9-L13
	3 2	-0.3333*L2-0.5*L3	0	L9+L11+L13
	3 3	-0.3333*L2	0	-L11

Tabela 10: Funções estimáveis do Tipo IV do SAS para ordem B-A

		В	Α	AB
μ		0	0	0
В	1	L 2	0	0
	2	L 3	0	0
	3	L 4	0	0
	4	- L 2 - L 3 - L 4	0	0
A	1	0	L 6	0
	2	0	L 7	0
	3	0	-L6-L7	0
ΑB	1 1	0.5*L2	0.3333*L6	L 9
	1 2	0.5*L3	0.3333*L6	L10
	1 3	L 4	0.3333*L6	L 1 1
	1 4	-0.5*L2-0.5*L3-L4	0	-L9-L10-L11
	2 1	0.5*L2	0.5*L7	L13
	2 2	0.5*L3	0.5*L7	-L9-L10-L11-L13
	2 4	-0.5*L2-0.5*L3	0	L9+L10+L11
	3 1	0	-0.3333*L6-0.5*L7	-L9-L13
	3 2	0	-0.3333*L6-0.5*L7	L9+L11+L13
	3 3	0	-0.3333*L6	-L11

4.2.1.3. Equações Normais e Aproximação de Mínimos Quadrados

O Sistema de Equações Normais (SEN) para o Modelo-S, tem a seguinte apresentação.

1	r.,		_	_	: _	_			: .									_			_
	18		6	5	5	5	4	4	$I_{}$	.2	2	2	2	2	2	2	1	2	μ		1220
	7	7	0	0	1	2	2	2	1	2	2	2	0	0	0	0	0	0	$\alpha_{I}$		420
	6	0	6	0	2	2	0	2	0	0	0	0	2	2	2	0	0	o	$ \alpha_2 $		415
	5	0	0	5	2	1	2	0	0	0	0	0	0	0	0	2	I	2	$\alpha_3$		385
	5	1	2	2	5	0	0	0	1	0	0	0	2	0	0	2	0	0	$ \beta_I $		300
	5	2	2	1	0	5	0	0	0	2	0	0	0	2	0	0	1	0	$ \beta_2 $		315
	4	2	0	2	0	0	4	0	0	0	2	0	0	0	0	0	0	2	$ \beta_3 $		290
	4	2	2	0	0	0	0	4	0	0	0	2	0	0	2	0	0	0	$ \beta_4 $		315
$X'X\theta = X'y \Rightarrow$	I	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	γ 11		45
AAO = Ay = A	2	2	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	γ <sub>12</sub>	=	105
	2	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	γ 13		125
	2	2	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	o	Y 14		145
	2	0	2	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	$ \gamma_{21} $		110
	2	0	2	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	Y 22		135
	2	0	2	0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	γ 23		170
	2	0	0	2	2	0	0	0	0	0	0	0	0	0	0	2	0	0	Y 31		145
	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	γ <sub>32</sub>		75
	2	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	2	$\gamma_{33}$		165
	•				:				:												_

Uma das soluções,  $\theta^o$  e a aproximação de mínimos quadrados,  $\hat{y}$  são dadas a seguir.

## 4.2.1.4. Estimação

Agora, faz-se a extensão do que foi colocado nos itens (3.2.) e (3.3.). O interesse é apresentar, para esse modelo e conjunto de dados, conjunto de funções estimáveis que possibilitem a formulação de hipóteses.

Nas Tabelas 11 e 12 são mostradas as funções estimáveis sobre os efeitos principais. A seguir, as funções estimáveis sobre a interação. As funções estimáveis são obtidas através do PROC GLM do SAS, utilizando as tabelas de 3 a 10.

```
\begin{cases} \textit{para } L9 = 1; \ L10 = 0; \ L11 = 0; \ L13 = 0 \Rightarrow \gamma_{11} - \gamma_{14} - \gamma_{22} + \gamma_{24} - \gamma_{31} + \gamma_{32} = 0 \\ \textit{para } L9 = 0; \ L10 = 1; \ L11 = 0; \ L13 = 0 \Rightarrow \gamma_{12} - \gamma_{14} - \gamma_{22} + \gamma_{24} = 0 \\ \textit{para } L9 = 0; \ L10 = 0; \ L11 = 1; \ L13 = 0 \Rightarrow \gamma_{13} - \gamma_{14} - \gamma_{22} + \gamma_{24} + \gamma_{32} - \gamma_{33} = 0 \\ \textit{para } L9 = 0; \ L10 = 0; \ L11 = 0; \ L13 = 1 \Rightarrow \gamma_{21} - \gamma_{22} - \gamma_{31} + \gamma_{32} = 0 \end{cases}
```

Para obter as funções estimáveis apresentadas, atribui-se valores aos coeficientes fornecidos nas funções estimáveis utilizadas pelo SAS. Com o objetivo de aplicar as funções estimáveis nas hipóteses testadas atribui-se, neste estudo, valor 1 para um dos coeficientes, zerando os restantes. Maiores detalhes em MONDARDO (1994).

Observa-se que, para o Modelo-S com interação, não há possibilidade de se obter funções estimáveis exclusivamente sobre os efeitos principais. O que ocorre, em geral, é um contraste entre os níveis de um fator, acompanhado de outros parâmetros, que podem ser efeitos do outro fator e/ou de interações.

Tabela 11: Funções estimáveis sobre o fator A no Modelo-S

Função		Função estimável
	L2=1	
	L3=0	$\alpha_{1} - \alpha_{3} - 0,2571\beta_{1} + 0,0857\beta_{2} - 0,1143\beta_{3} + 0,2857\beta_{4} + 0,1429\gamma_{12} + 0,2857\gamma_{13} + 0,2857\gamma_{14} - 0,4\gamma_{31} - 0,2\gamma_{32} - 0,4\gamma_{33}$
Tinn I	L2=0	
	L3=1	$\alpha_2 - \alpha_3 - 0,0667\beta_1 + 0,1333\beta_2 - 0,4\beta_3 + 0,3333\beta_4 + 0,3333\gamma_{21} + 0,3333\gamma_{22} + 0,3333\gamma_{24} - 0,4\gamma_{31} - 0,2\gamma_{32} - 0,4\gamma_{33}$
enne annie mene atent titen beter at	L2=1	
	L3=0	$\alpha_{1} - \alpha_{3} + 0,21\gamma_{11} + 0,2648\gamma_{12} + 0,3881\gamma_{13} + 0,137\gamma_{14} + 0,1461\gamma_{21} - 0,0091\gamma_{22} - 0,137\gamma_{24} - 0,3562\gamma_{31} - 0,2557\gamma_{32} - 0,3881\gamma_{33} + 0,21\gamma_{11} + 0,21\gamma_{12} + 0,21\gamma_{11} + 0,21\gamma_{12} + 0,21\gamma_{11} + 0,21\gamma_{12} + 0,21\gamma_{12}$
Tinn II	L2=0	
	L3=1	$\alpha_2 - \alpha_3 + 0,0183\gamma_{11} - 0,0639\gamma_{12} + 0,2511\gamma_{13} - 0,2055\gamma_{14} + 0,4475\gamma_{21} + 0,347\gamma_{22} + 0,2055\gamma_{24} - 0,4658\gamma_{31} - 0,2831\gamma_{32} - 0,2511\gamma_{33} - 0,2831\gamma_{32} - 0,2831\gamma_$
anne anne anne anne este asse as	L2=1	
	L3=0	$\alpha_{1} - \alpha_{3} + 0,2857\gamma_{11} + 0,2857\gamma_{12} + 0,3143\gamma_{13} + 0,1143\gamma_{14} + 0,0571\gamma_{21} + 0,0571\gamma_{22} - 0,1143\gamma_{24} - 0,3429\gamma_{31} - 0,3429\gamma_{32} - 0,0000000000000000000000000000000000$
Tino III	L2=0	Q.3143 <sub>1722</sub>
	L3=1	$\alpha_2$ - $\alpha_3$ +0,2 $\gamma_{13}$ -0,2 $\gamma_{14}$ +0,4 $\gamma_{21}$ +0,4 $\gamma_{22}$ +0,2 $\gamma_{24}$ -0,4 $\gamma_{31}$ -0,4 $\gamma_{32}$ -0,2 $\gamma_{33}$
	L2=1	
	L3=0	$\alpha_1$ - $\alpha_3$ +0,3333 $\gamma_{11}$ +0,3333 $\gamma_{12}$ +0,3333 $\gamma_{13}$ -0,3333 $\gamma_{31}$ -0,3333 $\gamma_{32}$ -0,3333 $\gamma_{33}$
Tinn IV	L2=0	
	L3=1	$\alpha_2$ - $\alpha_3$ +0,5 $\gamma_{21}$ +0,5 $\gamma_{22}$ -0,5 $\gamma_{31}$ -0,5 $\gamma_{32}$

Tabela 12: Funções estimáveis sobre o fator B no Modelo-S

Função	***************************************	Função estimável
	L2=1	
	L3=0	$\beta_1$ - $\beta_4$ -0, $3\alpha_1$ -0, $1\alpha_2$ +0, $4\alpha_3$ +0, $2\gamma_{11}$ -0, $5\gamma_{14}$ +0, $4\gamma_{21}$ -0, $5\gamma_{24}$ +0, $4\gamma_{31}$
	L4=0	
	L2=0	
Tipo I	L3=1	$\beta_2$ - $\beta_4$ -0, 1 $\alpha_1$ -0, 1 $\alpha_2$ +0, 2 $\alpha_3$ +0, 4 $\gamma_{12}$ -0, 5 $\gamma_{14}$ +0, 4 $\gamma_{22}$ -0, 5 $\gamma_{24}$ +0, 2 $\gamma_{32}$
	L4=0	
	L2=0	
	L3=0	$\beta_3$ - $\beta_4$ -0, $5\alpha_2$ +0, $5\alpha_3$ +0, $5\gamma_{13}$ -0, $5\gamma_{14}$ -0, $5\gamma_{24}$ +0, $5\gamma_{33}$
	<u> </u>	
	L3=0	$\beta_1$ - $\beta_4$ +0,2648 $\gamma_{11}$ +0,0731 $\gamma_{12}$ +0,1416 $\gamma_{13}$ -0,4795 $\gamma_{14}$ +0,4886 $\gamma_{21}$ +0,032 $\gamma_{22}$ -0,5205 $\gamma_{24}$ +0,2466 $\gamma_{31}$ -0,105 $\gamma_{32}$ -0,1416 $\gamma_{33}$
	L4=0	$p_1$ - $p_4$ - $0$ , $20$ - $0$ , $11$ - $10$ , $11$ - $10$ , $13$ - $0$ , $11$ - $0$ , $10$ - $0$ , $11$ - $0$ , $10$ - $0$ , $11$ -
	L2=0	
Tipo II	L3=1	$\beta_2$ - $\beta_4$ +0,0228 $\gamma_{11}$ +0,4201 $\gamma_{12}$ +0,0639 $\gamma_{13}$ -0,5068 $\gamma_{14}$ +0,0594 $\gamma_{21}$ +0,4338 $\gamma_{22}$ -0,4932 $\gamma_{24}$ -0,0822 $\gamma_{31}$ +0,1416 $\gamma_{32}$ -0,0639 $\gamma_{33}$
	L4=0	
	L2=0	
	L3=0	$\beta_3$ - $\beta_4$ +0,0091 $\gamma_{11}$ -0,032 $\gamma_{12}$ +0,6256 $\gamma_{13}$ -0,6027 $\gamma_{14}$ +0,2237 $\gamma_{21}$ +0,1735 $\gamma_{22}$ -0,3973 $\gamma_{24}$ -0,2329 $\gamma_{31}$ -0,1416 $\gamma_{32}$ +0,3744 $\gamma_{33}$
	<u></u>	
	L2=1	
	L3=0	$\beta_1$ - $\beta_4$ +0,381 $\gamma_{11}$ +0,0476 $\gamma_{12}$ +0,0857 $\gamma_{13}$ -0,5143 $\gamma_{14}$ +0,4095 $\gamma_{21}$ +0,0762 $\gamma_{22}$ -0,4857 $\gamma_{24}$ +0,2095 $\gamma_{31}$ -0,1238 $\gamma_{32}$ -0,0857 $\gamma_{33}$
	L4=0	,
	L2=0	0.00.0070 +0.004 +0.0057 0.5440 +0.0700 +0.4005 0.4057 0.4000 +0.0005 0.0057
Tipo III	L3=1	$\beta_2$ - $\beta_4$ +0,0476 $\gamma_{11}$ +0,381 $\gamma_{12}$ +0,0857 $\gamma_{13}$ -0,5143 $\gamma_{14}$ +0,0762 $\gamma_{21}$ +0,4095 $\gamma_{22}$ -0,4857 $\gamma_{24}$ -0,1238 $\gamma_{31}$ +0,2095 $\gamma_{32}$ -0,0857 $\gamma_{33}$
	<u>L4=0</u> L2=0	
	L2=0 L3=0	$\beta_3$ - $\beta_4$ +0,6 $\gamma_{13}$ -0,6 $\gamma_{14}$ +0,2 $\gamma_{21}$ +0,2 $\gamma_{22}$ -0,4 $\gamma_{24}$ -0,2 $\gamma_{31}$ -0,2 $\gamma_{32}$ +0,4 $\gamma_{33}$
	L4=1	$p_3$ - $p_4$ - $v_1$ $v_1$ $v_2$ - $v_1$ $v_2$ $v_3$ - $v_1$ $v_2$ $v_3$ - $v_1$ $v_3$ $v_4$ - $v_4$ $v_3$ $v_4$ - $v_$
	L2=1	
	L3=0	$\beta_1$ - $\beta_4$ +0,5 $\gamma_{11}$ -0,5 $\gamma_{14}$ +0,5 $\gamma_{21}$ -0,5 $\gamma_{24}$
	L4=0	
	L2=0	
Tipo IV	L3=1	$\beta_2$ - $\beta_4$ +0,5 $\gamma_{12}$ -0,5 $\gamma_{14}$ +0,5 $\gamma_{22}$ -0,5 $\gamma_{24}$
	L4=0	
	L2=0	
	L3=0	$\beta_{3}$ - $\beta_{4}$ +0,5 $\gamma_{13}$ -0,5 $\gamma_{14}$
	L4=1	

# 4.2.1.5. Hipóteses Testadas na Análise de Variância

Em se tratando de um modelo com interação, e dado que apresenta caselas vazias, as quatro somas de quadrados fornecidas pelo SAS/GLM diferem entre si. Somente a soma de quadrados do Tipo I depende da ordenação dos efeitos dos fatores principais. Assim, para este conjunto de dados, é possível obter, efetivamente, cinco formas de apresentar a Análise de Variância: na Tabela 15 duas formas, nas ordens A-B ou B-A; nas Tabelas 16, 17 e 18 as ordens A-B e B-A se equivalem.

O pesquisador, ao realizar as análises através do SAS, pode optar pela Análise de Variância que melhor representar as hipóteses de interesse de sua pesquisa. Como o objetivo deste trabalho é o de apresentar didaticamente as possibilidades dos modelos e sistemas estatísticos em estudo, não houve a preocupação de comparar hipóteses mais adequadas para os dados da Tabela 2

O PROC GLM do SAS utiliza, a princípio, o Modelo-S. Enquanto as três primeiras somas de quadrados são únicas, a do Tipo IV não é. Na saída do PROC GLM aparece uma advertência a respeito. Isso, pela possibilidade de se obter outras funções estimáveis do Tipo IV para o mesmo conjunto de dados, gerando outras somas de quadrados.

Exceto a hipótese sobre a interação, todas as demais vêm carregadas de componentes do outro fator e/ou interações. Sendo assim, quando o pesquisador utilizar hipóteses sobre os efeitos principais, deve estar consciente disso.

É importante reparar nos diferentes valores de somas de quadrados de hipóteses para um mesmo efeito. Nas tabelas 13 e 14 estão descritas as hipóteses no Modelo-S, sobre os efeitos principais. A seguir, sobre o efeito da interação.

$$H_0^9: \begin{cases} \gamma_{11} - \gamma_{14} - \gamma_{22} + \gamma_{24} - \gamma_{31} + \gamma_{32} = 0 \\ \gamma_{12} - \gamma_{14} - \gamma_{22} + \gamma_{24} = 0 \\ \gamma_{13} - \gamma_{14} - \gamma_{22} + \gamma_{24} + \gamma_{32} - \gamma_{33} = 0 \\ \gamma_{21} - \gamma_{22} - \gamma_{31} + \gamma_{32} = 0 \end{cases}$$

Há dificuldade na interpretação das hipóteses. Por exemplo, nas do Tipo I cada linha da hipótese apresenta um contraste entre níveis do efeito principal em questão. A seguir vem um contraste entre níveis do outro fator, pois nenhum ajuste é feito, acompanhando efeitos de diversas interações. Para as hipóteses do tipo II, não há presença de contraste do outro fator, mas somente da interação. Os efeitos que não fazem parte do fator em questão, na hipótese são multiplicados por índices que, sendo um conjunto de dados desbalanceado com caselas vazias, dificulta ainda mais a interpretação das hipóteses.

Neste exemplo, a hipótese sobre a interação possui somente quatro linhas. Isso por que há duas Perde-se, portanto, dois caselas vazias. graus liberdade. A princípio, com todas as caselas ocupadas, seriam seis linhas. Apesar das caselas vazias, o número de liberdade para os efeitos graus principais permanece como em um conjunto de dados balanceados: dois graus de liberdade para o fator A e três para o Fator B. Para os dados do exemplo em questão, não há problema de desconexão.

As funções estimáveis auxiliam na interpretação das hipóteses apresentadas. É através delas que se originam as hipóteses testadas no Modelo-S e consequentes somas de quadrados de hipóteses.

Dentre os possíveis quadros de Análise de Variância, aqueles que fazem o ajuste seqüencial testam hipóteses ortogonais. A alteração na ordem do efeitos principais altera a decomposição feita para esses efeitos. Nesse caso, coincide os totais com o Total corrigido.

As somas de quadrados nas tabelas 16, 17 e 18 são não ortogonais. A alteração da ordem dos efeitos principais no modelo não altera seu valores.

Tabela 13: Hipóteses testadas sobre o fator A no Modelo-S

Hipótese testada
$\alpha_{1} - \alpha_{3} - 0,2571\beta_{1} + 0,0857\beta_{2} - 0,1143\beta_{3} + 0,2857\beta_{4} + 0,1429\gamma_{12} + 0,2857\gamma_{13} + 0,2857\gamma_{14} - 0,4\gamma_{31} - 0,2\gamma_{32} - 0,4\gamma_{33} = 0$
$\alpha_2 - \alpha_3 - 0,0667\beta_1 + 0,1333\beta_2 - 0,4\beta_3 + 0,3333\beta_4 + 0,3333\gamma_{21} + 0,3333\gamma_{22} + 0,3333\gamma_{24} - 0,4\gamma_{31} - 0,2\gamma_{32} - 0,4\gamma_{33} = 0$
$\alpha_{1}-\alpha_{3}+0,21\gamma_{11}+0,2648\gamma_{12}+0,3881\gamma_{13}+0,137\gamma_{14}+0,1461\gamma_{21}-0,0091\gamma_{22}-0,137\gamma_{24}-0,3562\gamma_{31}-0,2557\gamma_{32}-0,3881\gamma_{33}=0$
$\alpha_2 - \alpha_3 + 0,0183\gamma_{11} - 0,0639\gamma_{12} + 0,2511\gamma_{13} - 0,2055\gamma_{14} + 0,4475\gamma_{21} + 0,347\gamma_{22} + 0,2055\gamma_{24} - 0,4658\gamma_{31} - 0,2831\gamma_{32} - 0,2511\gamma_{33} = 0$
$\alpha_{1}-\alpha_{3}+0,2857\gamma_{11}+0,2857\gamma_{12}+0,3143\gamma_{13}+0,1143\gamma_{14}+0,0571\gamma_{21}+0,0571\gamma_{22}-0,1143\gamma_{24}-0,3429\gamma_{31}-0,3429\gamma_{32}-0,3143\gamma_{33}=0$
$\alpha_{2}-\alpha_{3}+0,2\gamma_{13}-0,2\gamma_{14}+0,4\gamma_{21}+0,4\gamma_{22}+0,2\gamma_{24}-0,4\gamma_{31}-0,4\gamma_{32}-0,2\gamma_{33}=0$
$\alpha_1$ - $\alpha_3$ +0,3333 $\gamma_{11}$ +0,3333 $\gamma_{12}$ +0,3333 $\gamma_{13}$ -0,3333 $\gamma_{31}$ -0,3333 $\gamma_{32}$ -0,3333 $\gamma_{33}$ =0
$\alpha_2$ - $\alpha_3$ +0,5 $\gamma_{21}$ +0,5 $\gamma_{22}$ -0,5 $\gamma_{31}$ -0,5 $\gamma_{32}$ =0

Tabela 14: Hipóteses testadas sobre o fator B no Modelo-S

Hipótese	Hipótese testada
	$\beta_{1}$ - $\beta_{4}$ -0,3 $\alpha_{1}$ -0,1 $\alpha_{2}$ +0,4 $\alpha_{3}$ +0,2 $\gamma_{11}$ -0,5 $\gamma_{14}$ +0,4 $\gamma_{21}$ -0,5 $\gamma_{24}$ +0,4 $\gamma_{31}$ =0
Tipo I $H_0^{\mathfrak s}$	$\beta_2$ - $\beta_4$ -0, 1 $\alpha_1$ -0, 1 $\alpha_2$ +0, 2 $\alpha_3$ +0, 4 $\gamma_{12}$ -0, 5 $\gamma_{14}$ +0, 4 $\gamma_{22}$ -0, 5 $\gamma_{24}$ +0, 2 $\gamma_{32}$ =0
	$\beta_3$ - $\beta_4$ -0,5 $\alpha_2$ +0,5 $\alpha_3$ +0,5 $\gamma_{13}$ -0,5 $\gamma_{14}$ -0,5 $\gamma_{24}$ +0,5 $\gamma_{33}$ =0
	$\beta_{1}-\beta_{4}+0,2648\gamma_{11}+0,0731\gamma_{12}+0,1416\gamma_{13}-0,4795\gamma_{14}+0,4886\gamma_{21}+0,032\gamma_{22}-0,5205\gamma_{24}+0,2466\gamma_{31}-0,105\gamma_{32}-0,1416\gamma_{33}=0$
Tipo II $H_{ heta}^{ extit{6}}$	$\beta_2 - \beta_4 + 0,0228\gamma_{11} + 0,4201\gamma_{12} + 0,0639\gamma_{13} - 0,5068\gamma_{14} + 0,0594\gamma_{21} + 0,4338\gamma_{22} - 0,4932\gamma_{24} - 0,0822\gamma_{31} + 0,1461\gamma_{32} - 0,0639\gamma_{33} = 0$
·	$\beta_{3}-\beta_{4}+0,0091\gamma_{11}-0,032\gamma_{12}+0,6256\gamma_{13}-0,6027\gamma_{14}+0,2237\gamma_{21}+0,1735\gamma_{22}-0,3973\gamma_{24}-0,2329\gamma_{31}-0,1416\gamma_{32}+0,3744\gamma_{33}=0$
	$\beta_{1}-\beta_{4}+0,381\gamma_{11}+0,0476\gamma_{12}+0,0857\gamma_{13}-0,5143\gamma_{14}+0,4095\gamma_{21}+0,0762\gamma_{22}-0,4857\gamma_{24}+0,2095\gamma_{31}-0,1238\gamma_{32}-0,0857\gamma_{33}=0$
Tipo III $H_{ heta}^{7}$	$\beta_2 - \beta_4 + 0,0476\gamma_{11} + 0,381\gamma_{12} + 0,0857\gamma_{13} - 0,5143\gamma_{14} + 0,0762\gamma_{21} + 0,4095\gamma_{22} - 0,4857\gamma_{24} - 0,1238\gamma_{31} + 0,2095\gamma_{32} - 0,0857\gamma_{33} = 0$
	$\beta_{3}-\beta_{4}+0,6\gamma_{13}-0,6\gamma_{14}+0,2\gamma_{21}+0,2\gamma_{22}-0,4\gamma_{24}-0,2\gamma_{31}-0,2\gamma_{32}+0,4\gamma_{33}=0$
	$\beta_{1}$ - $\beta_{4}$ +0,5 $\gamma_{11}$ -0,5 $\gamma_{14}$ +0,5 $\gamma_{21}$ -0,5 $\gamma_{24}$ =0
Tipo IV $H_{o}^{8}$	$\beta_2$ - $\beta_4$ +0,5 $\gamma_{12}$ -0,5 $\gamma_{14}$ +0,5 $\gamma_{22}$ -0,5 $\gamma_{24}$ =0
· ·	$\beta_3$ - $\beta_4$ +0,5 $\gamma_{13}$ -0,5 $\gamma_{14}$ =0

Tabela 15: SQ Tipo I do SAS<sup>1</sup>

Para a ordem A-B					Para a ordem B-A				
C.V.	G.L	$H_0$	S . Q .		C.V.	G.L	H <sub>0</sub>	S . Q .	(1) - (1) -
A	2	$H_0^I$		860,278	В	3	$H_0^5$		987,361
B(aj.)	3	$H_0^6$		1.612,991	A(aj.)	2	$H_0^2$		1.485,907
AxB	4	$H_0^9$		62,842	AxB	4	$H_0^{g}$		62,842
Resíduo	8	v		175,000	Resíduo	8	v		175,000

11: A ordem do fatores principais altera as somas de quadrados dos efeitos principais

Tabela 16: SQ Tipo II do SAS<sup>2</sup>

	Para a ordem A-B					Para a ordem B-A				
C.V.	G.L	$H_0$	S.Q.		C.V.	G.L	$H_0$	S . Q .		
A(aj.)	2	$H_0^2$		1.485,907	B(aj.)	3	$H_0^6$	***************************************	1.612,991	
B(aj.)	3	$H_0^6$		1.612,991	A(aj.)	2	$H_0^2$		1.485,907	
AxB	4	$H_0^9$		62,842	AxB	4	$H_0^9$		62,842	
Resíduo	8	v		175,000	Resíduo	8	v		175,000	

21: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

Tabela 17: SQ Tipo III do SAS<sup>3</sup>

Para a ordem A-B					Para a ordem B-A				
C.V.	G.L	$H_0$	S . Q .		C.V.	G.L	H <sub>0</sub>	S.Q.	
Α	2	$H_0^3$	***************************************	1.371,088	В	3	$H_0^7$	<del>, , , , , , , , , , , , , , , , , , , </del>	1.565,603
В	3	$H_{\theta}^{7}$		1.565,603	Α	2	$H_0^3$		1.371,088
AxB	4	$H_0^9$		62,842	AxB	4	$H_0^g$		62,842
Resíduo	8	v		175,000	Residuo	8	Ÿ		175,000

31: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

Tabela 18: SQ Tipo IV do SAS<sup>4</sup>

	Para	a ordem	A-B	Para a ordem B-A				
C.V.	G.L	$H_0$	S . Q .	C . V .	G.L	H <sub>0</sub>	S.Q.	Maka kun militar dan menjadi dan semini dan semini kenamat dan selah dan semini dan semini dan semini dan semi
A	2	$H_0^4$	1.225,806	В	3	$H_0^8$		1.475,000
В	3	$H_0^8$	1.475,000	A	2	$H_0^4$		1.225,806
AxB	4	$H_0^9$	62,842	AxB	4	$H_0^g$		62,842
Resíduo	8	v	175,000	Resíduo	8	v		175,000

4/: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

#### 4.2.2. Modelo-M

## 4.2.2.1. Caracterização

Nesse caso, o modelo na ordem A-B é

$$y_{ijk} = \mu_{ij} + e_{ijk}$$

Para a ordem B-A

$$y_{jik} = \mu_{ji} + e_{jik}$$

Matricialmente, pode-se escrever o modelo para a ordem A-B:

### 4.2.2.2. Sistema Estatístico BMDP

A macro para análise é dada a seguir.

```
TITLE IS 'Conjunto 1'.
/ INPUT
         VARIABLES = 3.
         FORMAT = FREE.
            NAMES = FATOR1, FATOR2, Y. FACTORS = FATOR1, FATOR2.
/ VARIABLE
/ BETWEEN
         CODES(FATOR1) = 1 TO 3.
         NAMES(FATOR1) = A1, A2, A3.
         CODES(FATOR2) = 1 TO 4.
         NAMES(FATOR2) = B1, B2, B3, B4.
/ WEIGHTS BETWEEN = EQUAL.
/ PRINT
           MARGINAL = ALL.
/ END
  1
      45
      50
  2
      55
1
  3
      60
  3
1
      65
  4
      70
  4
      75
1
2
  1
      50
2 2 2
  1
      60
  2
      70
      65
2
  4
      80
  4
      90
3
  1
      75
  1
      70
3
  2
      75
3
  3
      85
3 3
      80
/END
ANALYSIS
             PROCEDURE = FACTORIAL. /
END /
/ WEIGHT
             BETWEEN = SIZES.
            MARGINALS = NONE. NO CELLS.
/ PRINT
/ END
             PROCEDURE = FACTORIAL. /
ANALYSIS
END /
```

O BMDP 4V, que utiliza o Modelo-M para formulação das hipóteses, não realiza a análise de Variância para este conjunto de dados, pois como visto anteriormente, não realiza análise de dados com caselas vazias, que é o caso.

Quando se solicita a análise através desse sistema, fornece uma mensagem exigindo que não haja caselas vazias.

"\*\*\* WARNING \*\*\* BETWEEN DESIGN IS NOT COMPLETELY CROSSED.
IF THIS IS NOT INTENTIONAL, CHECK CELL
DESCRIPTIVE INFORMATION CAREFULLY.
COMPUTATIONS WILL CONTINUE UNDER THE
ASSUMPTION THAT CELLS NOT PRESENT ARE MISSING
FROM THE DESIGN, NOT MERELY FROM THE DATA.

Embora o BMDP não analise modelos com interação na presença de caselas vazias, ele é sabidamente eficiente e muito utilizado quando não há caselas vazias. Como um dos objetivos deste trabalho é o de divulgar modelos e sistemas, optou-se por apresentálo, mesmo sem utilidade prática imediata.

A partir deste momento, retoma-se a discussão através do SAS/PROC GLM, através de uma adaptação que se propõe a seguir. Espera-se com isso, facilitar ao pesquisador a construção das macros no BMDP e ampliar a utilização do SAS/GLM para o modelo de Médias de Caselas.

4.2.2.3. Equações Normais e Aproximação de Mínimos Quadrados

O modelo apresenta o seguinte SEN.

A solução,  $\mu^{o}$  e a aproximação de mímimos quadrados,  $\hat{y}$  são dadas a seguir.

$$\mu^{0} = (W'W)^{-1}W'y = \begin{bmatrix} 45.0 \\ 52.5 \\ 62.5 \\ 72.5 \\ 72.5 \\ 85.0 \\ 72.5 \\ 85.0 \\ 72.5 \\ 75.0 \\ 82.5 \end{bmatrix}; \qquad \hat{y} = W\mu^{0} = \begin{bmatrix} 45 \\ 52.5 \\ 62.5 \\ 72.5 \\ 72.5 \\ 67.5 \\ 85 \\ 85 \\ 85 \\ 72.5 \\ 75 \\ 82.5 \\ 82.5 \end{bmatrix}$$

Naturalmente, por se tratar de modelos equivalentes, a aproximação de mínimos quadrados através do Modelo-S e Modelo-M são iguais.

## 4.2.2.4. Estimação

O Modelo-M é de posto coluna completo. Isso significa que todos os parâmetros são estimáveis individualmente. Sendo assim, qualquer função paramétrica é estimável. Esse conceito se estende aos demais modelos de posto coluna completo. contexto, não se discute a estimação para os próximos modelos.

Embora, conforme dito anteriormente, o PROC GLM do SAS seja construído com base em modelos superparametrizados, ele pode ser utilizado na obtenção das funções estimáveis no Modelo-M. Para tanto, eliminase os efeitos dos fatores principais. Para uma melhor compreensão das hipóteses, pode-se esquematizar o conjunto de dados por:

$\mu_{II}^{(I)}$	$\mu_{_{I2}}^{(2)}$	$\mu_{I3}^{(2)}$	$\mu_{_{14}}^{(2)}$
$\mu_{2I}^{(2)}$	$\mu_{_{22}}^{(2)}$		$\mu_{\scriptscriptstyle 24}^{\scriptscriptstyle (2)}$
$\mu_{3I}^{(2)}$	$\mu_{\scriptscriptstyle 32}^{\scriptscriptstyle (I)}$	$\mu_{_{33}}^{(2)}$	<b>-</b>

Dessa forma, a título de exemplo, pode-se obter a função estimável do Tipo I para A, obtida da Tabela 19.

$$\begin{cases} \frac{\mu_{11} + 2\mu_{12} + 2\mu_{13} + 2\mu_{14}}{7} = \frac{2\mu_{31} + \mu_{32} + 2\mu_{33}}{5} \\ \frac{2\mu_{21} + 2\mu_{22} + 2\mu_{24}}{6} = \frac{2\mu_{31} + \mu_{32} + 2\mu_{33}}{5} \end{cases}$$

Outras formas de obtenção das funções estimáveis dos Tipos II e III podem ser obtidas em SEARLE(1987) e CAMARINHA FILHO(1995).

As Tabelas 19 e 20 mostram as funções estimáveis para os efeitos principais. Para a interação, são apresentadas a seguir.

Funções estimáveis sobre a interação no Modelo-M:

$$\begin{cases} \textit{para} \, L9 = 1; \, L10 = 0; \, L11 = 0; \, L13 = 0 \implies \mu_{11} - \mu_{14} - \mu_{22} + \mu_{24} - \mu_{31} + \mu_{32} \\ \textit{para} \, L9 = 0; \, L10 = 1; \, L11 = 0; \, L13 = 0 \implies \mu_{12} - \mu_{14} - \mu_{22} + \mu_{24} \\ \textit{para} \, L9 = 0; \, L10 = 0; \, L11 = 1; \, L13 = 0 \implies \mu_{13} - \mu_{14} - \mu_{22} + \mu_{24} + \mu_{32} - \mu_{33} \\ \textit{para} \, L9 = 0; \, L10 = 0; \, L11 = 0; \, L13 = 1 \implies \mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} \end{cases}$$

Nesse modelo as funções estimáveis são de mais simples interpretação. As funções estimáveis são utilizadas para descrever as hipóteses testadas através do Modelo-M. Facilitam sobremaneira sua compreensão.

Tabela 19: Funções estimáveis sobre o fator A no Modelo-M

Fun	ção	Função estimável
	L2=1	0.4400 1.0.0057 1.0.0057 0.4 0.0 0.4
Tipo I	L3=0	$0,1429\mu_{12}$ + $0,2857\mu_{13}$ + $0,2857\mu_{14}$ - $0,4\mu_{31}$ - $0,2\mu_{32}$ - $0,4\mu_{33}$
1150 1	L2=0	
	L3=1	$0,3333\mu_{21}$ + $0,3333\mu_{22}$ + $0,3333\mu_{24}$ - $0,4\mu_{31}$ - $0,2\mu_{32}$ - $0,4\mu_{33}$
	L2=1	
Time II	L3=0	$0,21\mu_{11}+0,2648\mu_{12}+0,3881\mu_{13}+0,137\mu_{14}+0,1461\mu_{21}-0,0091\mu_{22}-0,137\mu_{24}-0,3562\mu_{31}-0,2557\mu_{32}-0,3881\mu_{33}$
Tipo II	L2=0	
	L3=1	$0,0183\mu_{11}\text{-}0,0639\mu_{12}\text{+}0,2511\mu_{13}\text{-}0,2055\mu_{14}\text{+}0,4475\mu_{21}\text{+}0,347\mu_{22}\text{+}0,2055\mu_{24}\text{-}0,4658\mu_{31}\text{-}0,2831\mu_{32}\text{-}0,2511\mu_{33}$
	L2=1	
Time III	L3=0	$0,2857\mu_{11}+0,2857\mu_{12}+0,3143\mu_{13}+0,1143\mu_{14}+0,0571\mu_{21}+0,0571\mu_{22}-0,1143\mu_{24}-0,3429\mu_{31}-0,3429\mu_{32}-0,3143\mu_{33}$
Tipo III	L2=0	
	L3=1	$0,2\mu_{13}$ - $0,2\mu_{14}$ + $0,4\mu_{21}$ + $0,4\mu_{22}$ + $0,2\mu_{24}$ - $0,4\mu_{31}$ - $0,4\mu_{32}$ - $0,2\mu_{33}$
	L2=1	
	L3=0	$0,3333\mu_{11}$ + $0,3333\mu_{12}$ + $0,3333\mu_{13}$ - $0,3333\mu_{31}$ - $0,3333\mu_{32}$ - $0,3333\mu_{33}$
Tipo IV	L2=0	
	L3=1	$0,5\mu_{21}$ + $0,5\mu_{22}$ - $0,5\mu_{31}$ - $0,5\mu_{32}$

Tabela 20: Funções estimáveis sobre o fator B no Modelo-M

Função	0	Função estimável
	L2=1	
	L3=0	$0,2\mu_{11}$ - $0,5\mu_{14}$ + $0,4\mu_{21}$ - $0,5\mu_{24}$ + $0,4\mu_{31}$
	L4=0	***************************************
	L2=0	
Tipo I	L3=1	$0,4\mu_{12}$ - $0,5\mu_{14}$ + $0,4\mu_{22}$ - $0,5\mu_{24}$ + $0,2\mu_{32}$
	<u>L4=0</u> L2=0	***************************************
	L2=0 L3=0	$0,5\mu_{13}$ - $0,5\mu_{14}$ - $0,5\mu_{24}$ + $0,5\mu_{33}$
	L3=0 L4=1	$0,0\mu_{13}$ - $0,0\mu_{14}$ - $0,0\mu_{24}$ + $0,0\mu_{33}$
	L2=1	
	L3=0	$0,2648\mu_{11}+0,0731\mu_{12}+0,1416\mu_{13}-0,4795\mu_{14}+0,4886\mu_{21}+0,032\mu_{22}-0,5205\mu_{24}+0,2466\mu_{31}-0,105\mu_{32}-0,1416\mu_{33}$
	L4=0	0, = 0, = 0, = 0, = 0, = 0, = 0, = 0, =
	L2=0	
Tipo II	L3=1	$0,0228\mu_{11}+0,4201\mu_{12}+0,0639\mu_{13}-0,5068\mu_{14}+0,0594\mu_{21}+0,4338\mu_{22}-0,4932\mu_{24}-0,0822\mu_{31}+0,1416\mu_{32}-0,0639\mu_{33}$
•	L4=0	
	L2=0	
	L3=0	$0,0091\mu_{11}$ - $0,032\mu_{12}$ + $0,6256\mu_{13}$ - $0,6027\mu_{14}$ + $0,2237\mu_{21}$ + $0,1735\mu_{22}$ - $0,3973\mu_{24}$ - $0,2329\mu_{31}$ - $0,1416\mu_{32}$ + $0,3744\mu_{33}$
	_ <u></u>	
	L2=1	
	L3=0	$0,381\mu_{11}+0,0476\mu_{12}+0,0857\mu_{13}-0,5143\mu_{14}+0,4095\mu_{21}+0,0762\mu_{22}-0,4857\mu_{24}+0,2095\mu_{31}-0,1238\mu_{32}-0,0857\mu_{33}$
	L4=0	
	L2=0	0.0470
Tipo III	L3=1	$0,0476\mu_{11}$ + $0,381\mu_{12}$ + $0,0857\mu_{13}$ - $0,5143\mu_{14}$ + $0,0762\mu_{21}$ + $0,4095\mu_{22}$ - $0,4857\mu_{24}$ - $0,1238\mu_{31}$ + $0,2095\mu_{32}$ - $0,0857\mu_{33}$
	L4=0 L2=0	
	L3=0	$0,6\mu_{13}$ - $0,6\mu_{14}$ + $0,2\mu_{21}$ + $0,2\mu_{22}$ - $0,4\mu_{24}$ - $0,2\mu_{31}$ - $0,2\mu_{32}$ + $0,4\mu_{33}$
	L4=1	$0,0\mu13,0,0\mu14,0,2\mu21,0,2\mu22,0,4\mu24,0,2\mu31,0,2\mu32,0,4\mu33$
	L2=1	
	L3=0	$0,5\mu_{11}$ - $0,5\mu_{14}$ + $0,5\mu_{21}$ - $0,5\mu_{24}$
	L4=0	
	L2=0	
Tipo IV	L3=1	$0.5\mu_{12}$ - $0.5\mu_{14}$ + $0.5\mu_{22}$ - $0.5\mu_{24}$
	L4=0	
	L2=0	
	L3=0	$0,5\mu_{13}$ - $0,5\mu_{14}$
	L4=1	

# 4.2.2.5. Hipóteses Testadas na Análise de Variância

Com as hipóteses sobre os efeitos principais descritas nas Tabelas 21 e 22 e sobre a interação a seguir, com consequentes somas de quadrados, pode-se obter os quadros de Análise de Variância.

Hipótese sobre a interação:

$$H_0^9: \begin{cases} \mu_{11} - \mu_{14} - \mu_{22} + \mu_{24} - \mu_{31} + \mu_{32} = 0 \\ \mu_{12} - \mu_{14} - \mu_{22} + \mu_{24} = 0 \\ \mu_{13} - \mu_{14} - \mu_{22} + \mu_{24} + \mu_{32} - \mu_{33} = 0 \\ \mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0 \end{cases}$$

São apresentados os quadros de análise para as ordens A-B e B-A. É mostrado que, apesar de serem oito quadros no total, efetivamente são apenas cinco Análises de Variância possíveis. Somente para o quadro equivalente à soma de quadrados do Tipo I do PROC GLM do SAS é que há alteração quando muda-se a ordem dos efeitos principais.

As somas de quadrados de hipóteses no Modelo-M são obtidas conforme (4), apresentado no item (3.7.), de acordo com SEARLE(1987). Assim, cada hipótese tem uma soma de quadrados a ela associada, que possibilita testá-la.

Tabela 21: Hipóteses testadas sobre o fator A no Modelo-M

ipótese 	Hipótese testada
Tipo I	$0,1429\mu_{12}+0,2857\mu_{13}+0,2857\mu_{14}-0,4\mu_{31}-0,2\mu_{32}-0,4\mu_{33}=0$
$H_0^I$	$0,3333\mu_{21}+0,3333\mu_{22}+0,3333\mu_{24}-0,4\mu_{31}-0,2\mu_{32}-0,4\mu_{33}=0$
	$0,21\mu_{11}+0,2648\mu_{12}+0,3881\mu_{13}+0,137\mu_{14}+0,1461\mu_{21}-0,0091\mu_{22}-0,137\mu_{24}-0,3562\mu_{31}-0,2557\mu_{32}-0,3881\mu_{33}=0$
Tipo II $H_0^2$	$0,0183\mu_{11}-0,0639\mu_{12}+0,2511\mu_{13}-0,2055\mu_{14}+0,4475\mu_{21}+0,347\mu_{22}+0,2055\mu_{24}-0,4658\mu_{31}-0,2831\mu_{32}-0,2511\mu_{33}=0$
Tine 111	$0,2857\mu_{11}+0,2857\mu_{12}+0,3143\mu_{13}+0,1143\mu_{14}+0,0571\mu_{21}+0,0571\mu_{22}-0,1143\mu_{24}-0,3429\mu_{31}-0,3429\mu_{32}-0,3143\mu_{33}=0$
Tipo III $H_0^3$	$0, 2\mu_{13} - 0, 2\mu_{14} + 0, 4\mu_{21} + 0, 4\mu_{22} + 0, 2\mu_{24} - 0, 4\mu_{31} - 0, 4\mu_{32} - 0, 2\mu_{33} = 0$
Tipo IV	$0,3333\mu_{11}+0,3333\mu_{12}+0,3333\mu_{13}-0,3333\mu_{31}-0,3333\mu_{32}-0,3333\mu_{33}=0$
$H_0^4$	$0,5\mu_{21}+0,5\mu_{22}-0,5\mu_{31}-0,5\mu_{32}=0$

Tabela 22: Hipóteses testadas sobre o fator B no Modelo-M

Hipótese	Hipótese testada
	$0,2\mu_{11}$ - $0,5\mu_{14}$ + $0,4\mu_{21}$ - $0,5\mu_{24}$ + $0,4\mu_{31}$ = $0$
Tipo I $H_o^{\mathfrak s}$	$0,4\mu_{12}$ - $0,5\mu_{14}$ + $0,4\mu_{22}$ - $0,5\mu_{24}$ + $0,2\mu_{32}$ = $0$
	$0,5\mu_{13}$ - $0,5\mu_{14}$ - $0,5\mu_{24}$ + $0,5\mu_{33}$ = $0$
	$0,2648\mu_{11}+0,0731\mu_{12}+0,1416\mu_{13}-0,4795\mu_{14}+0,4886\mu_{21}+0,032\mu_{22}-0,5205\mu_{24}+0,2466\mu_{31}-0,105\mu_{32}-0,1416\mu_{33}=0$
Tipo II $H_{ heta}^{ heta}$	$0,0228\mu_{11}+0,4201\mu_{12}+0,0639\mu_{13}-0,5068\mu_{14}+0,0594\mu_{21}+0,4338\mu_{22}-0,4932\mu_{24}-0,0822\mu_{31}+0,1416\mu_{32}-0,0639\mu_{33}=0.0000000000000000000000000000000000$
	$0,0091\mu_{11}-0,032\mu_{12}+0,6256\mu_{13}-0,6027\mu_{14}+0,2237\mu_{21}+0,1735\mu_{22}-0,3973\mu_{24}-0,2329\mu_{31}-0,1416\mu_{32}+0,3744\mu_{33}=0$
	$0,381\mu_{11}+0,0476\mu_{12}+0,0857\mu_{13}-0,5143\mu_{14}+0,4095\mu_{21}+0,0762\mu_{22}-0,4857\mu_{24}+0,2095\mu_{31}-0,1238\mu_{32}-0,0857\mu_{33}=0$
Tipo III $H_{m{ heta}}^{7}$	$0,0476\mu_{11}+0,381\mu_{12}+0,0857\mu_{13}-0,5143\mu_{14}+0,0762\mu_{21}+0,4095\mu_{22}-0,4857\mu_{24}-0,1238\mu_{31}+0,2095\mu_{32}-0,0857\mu_{33}=0$
	$0,6\mu_{13}-0,6\mu_{14}+0,2\mu_{21}+0,2\mu_{22}-0,4\mu_{24}-0,2\mu_{31}-0,2\mu_{32}+0,4\mu_{33}=0$
	$0,5\mu_{11}$ - $0,5\mu_{14}$ + $0,5\mu_{21}$ - $0,5\mu_{24}$ = $0$
Tipo IV $H_0^8$	$0,5\mu_{12}$ - $0,5\mu_{14}$ + $0,5\mu_{22}$ - $0,5\mu_{24}$ = $0$
·	$0,5\mu_{13}$ - $0,5\mu_{14}=0$

Dessa forma, as somas de quadrados sobre as hipóteses são:

a)Sobre o fator A

Para testar  $H_0^I$ :

$$SQH_0^1 = (B_1'\mu^0)'[B_1'(W'W)^{-1}B_1]^{-1}(B_1'\mu^0) = 860,278$$

Para testar  $H_0^2$ 

$$SQH_0^2 = (B_2'\mu^0)'[B_2'(W'W)^{-1}B_2]^{-1}(B_2'\mu^0) = 1485,908$$

Para testar  $H_0^3$ 

$$SQH_0^3 = (B_3'\mu^0)'[B_3'(W'W)^{-1}B_3]^{-1}(B_3'\mu^0) = 1371,089$$

Para testar  $H_0^4$ 

$$SQH_0^4 = (B_4'\mu^0)'[B_4'(W'W)^{-1}B_4]^{-1}(B_4'\mu^0) = 1225,806$$

Onde,  $H_0^I$  a  $H_0^4$  representam, respectivamente, as hipóteses dos Tipos I a IV para o fator A.

b)Sobre o fator B

Para testar  $H_0^5$ 

$$SQH_0^5 = (B_5'\mu^0)'[B_5'(W'W)^{-1}B_5]^{-1}(B_5'\mu^0) = 987,361$$

Para testar 
$$H_0^6$$

$$SQH_0^6 = (B_6'\mu^0)'[B_6'(W'W)^{-1}B_6]^{-1}(B_6'\mu^0) = 1612,991$$

Para testar  $H_0^7$ 

$$SQH_0^7 = (B_7'\mu^0)'[B_7'(W'W)^{-1}B_7]^{-1}(B_7'\mu^0) = 1565,604$$

Para testar  $H_0^8$ 

$$SQH_0^8 = (B_8'\mu^0)'[B_8'(W'W)^{-1}B_8]^{-1}(B_8'\mu^0) = 1475,000$$

Onde,  $H_0^{\mathfrak H}$  a  $H_0^{\mathfrak H}$  representam, respectivamente, as hipóteses dos Tipos I a IV para o fator B.

c)Sobre a interação;

Para testar  $H_0^9$ 

$$SQH_0^9 = (B_9'\mu^0)'[B_9'(W'W)^{-1}B_9]^{-1}(B_9'\mu^0) = 62,842$$

 $H_{\theta}^{g}$  corresponde à hipótese sobre a interação.

Essas somas de quadrados proporcionam quadros de Análise de Variância, como os apresentados nas Tabelas de 23 a 26.

Tabela 23: SQ com ajuste seqüencial no Modelo-M<sup>1</sup>

	Para	a a ordem A	\-B	Para a ordem B-A					
C.V.	G.L.	hipótese	S.Q.	C.V.	G.L.	hipótese	S.Q.		
Α	2	$H_0^I$	$SQH_0^1 = 860,278$	В	3	$H_0^5$	$SQH_0^5 = 987,361$		
B(aj.)	3	$H_0^{\delta}$	$SQH_0^6 = 1612,991$	A(aj.)	2	$H_0^2$	$SQH_0^2 = 1485,908$		
AxB	4	$H_0^{g}$	$SQH_0^9 = 62,842$	AxB	4	$H_0^{g}$	$SQH_0^9 = 62,842$		
Resíduo	8	· ·	S.Q.Res.=175,00	Resíduo	8	v	S.Q.Res.=175,00		

1/: A ordem do fatores principais altera as somas de quadrados dos efeitos principais

Tabela 24: SQ ajustadas para os efeitos principais no Modelo-M<sup>2</sup>

	Para	a a ordem A	<b>\-В</b>	Para a ordem B-A				
C.V.	G.L.	hipótese	S.Q.	C.V.	G.L.	hipótese	S.Q.	
A(aj.)	2	$H_0^2$	$SQH_0^2 = 1485,908$	B(aj.)	3	$H_0^{\delta}$	$SQH_0^6 = 1612,991$	
B(aj.)	3	$H_0^6$	$SQH_0^6 = 1612,991$	A(aj.)	2	$H_0^2$	$SQH_0^2 = 1485,908$	
AxB	4	$H_0^9$	$SQH_0^9 = 62,842$	AxB	4	$H_0^{g}$	$SQH_0^9 = 62,842$	
Resíduo	8	v	S.Q.Res.=175,00	Resíduo	8	V	S.Q.Res.=175,00	

2/: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

Tabela 25: SQ Tipo III no Modelo-M<sup>3</sup>

Para a ordem A-B				Para a ordem B-A				
C.V.	G.L.	hipótese	S.Q.	C.V.	G.L.	hipótese	S.Q.	
A	2	$H_0^3$	$SQH_0^3 = 1371,089$	В	3	$H_0^7$	$SQH_0^7 = 1565,604$	
В	3	$H_0^7$	$SQH_0^7 = 1565,604$	Α	2	$H_0^3$	$SQH_0^3 = 1371,089$	
AxB	4	$H_0^{g}$	$SQH_0^9 = 62,842$	AxB	4	$H_{\theta}^{g}$	$SQH_0^9 = 62,842$	
Resíduo	8	v	S.Q.Res.=175,00	Resíduo	8	v	S.Q.Res.=175,00	

3/: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

Tabela 26: SQ Tipo IV no Modelo-M<sup>4</sup>

Para a ordem A-B				Para a ordem B-A				
C.V.	G.L.	hipótese	S.Q.	C.V.	G.L.	hipótese	S.Q.	
A	2	$H_0^4$	$SQH_0^4 = 1225,806$	В	3	$H_0^8$	$SQH_0^8 = 1475,000$	
3	3	$H_0^8$	$SQH_0^8 = 1475,000$	Α	2	$H_0^4$	$SQH_0^4 = 1225,806$	
АхВ	4	$H_o^g$	$SQH_0^9 = 62,842$	AxB	4	$H_0^{g}$	$SQH_0^9 = 62,842$	
Resíduo	8	v	S.Q.Res.=175,00	Resíduo	8	V	S.Q.Res.=175,00	

4/: A ordem do fatores principais não altera as somas de quadrados dos efeitos principais

#### 4.2.3. Modelo- $\Sigma$

## 4.2.3.1. Caracterização

O modelo tem a seguinte caracterização  $y_{ijk}=\dot{\mu}+\dot{\alpha}_i+\dot{\beta}_j+\dot{\gamma}_{ij}+e_{ijk}$ 

Apresentando as seguintes restrições.

$$\sum_{i} \dot{\alpha}_{i} = 0 \Rightarrow \dot{\alpha}_{1} + \dot{\alpha}_{2} + \dot{\alpha}_{3} = 0 \Leftrightarrow \dot{\alpha}_{3} = -\dot{\alpha}_{1} - \dot{\alpha}_{2}$$

$$\sum_{i} \dot{\beta}_{j} = 0 \Rightarrow \dot{\beta}_{1} + \dot{\beta}_{2} + \dot{\beta}_{3} + \dot{\beta}_{4} = 0 \Leftrightarrow \dot{\beta}_{4} = -\dot{\beta}_{1} - \dot{\beta}_{2} - \dot{\beta}_{3}$$

$$\sum_{i} \dot{\gamma}_{ij} = 0, \quad \forall \quad j \Rightarrow \begin{cases}
\dot{\gamma}_{11} + \dot{\gamma}_{21} + \dot{\gamma}_{31} = 0 \\
\dot{\gamma}_{12} + \dot{\gamma}_{22} + \dot{\gamma}_{32} = 0 \\
\dot{\gamma}_{12} + \dot{\gamma}_{23} = 0 \\
\dot{\gamma}_{14} + \dot{\gamma}_{24} = 0
\end{cases}$$

$$\sum_{i} \dot{\gamma}_{ij} = 0, \quad \forall \quad i \Rightarrow \begin{cases}
\dot{\gamma}_{11} + \dot{\gamma}_{12} + \dot{\gamma}_{13} + \dot{\gamma}_{14} = 0 \\
\dot{\gamma}_{21} + \dot{\gamma}_{12} + \dot{\gamma}_{13} + \dot{\gamma}_{14} = 0 \\
\dot{\gamma}_{21} + \dot{\gamma}_{22} + \dot{\gamma}_{24} = 0 \\
\dot{\gamma}_{31} + \dot{\gamma}_{32} + \dot{\gamma}_{33} = 0
\end{cases}$$

$$\Rightarrow \begin{cases}
\dot{\gamma}_{11} = -\dot{\gamma}_{12} - \dot{\gamma}_{13} - \dot{\gamma}_{13} \\
\dot{\gamma}_{12} = \dot{\gamma}_{12} \\
\dot{\gamma}_{13} = \dot{\gamma}_{13} \\
\dot{\gamma}_{21} = \dot{\gamma}_{12} + \dot{\gamma}_{14} + \dot{\gamma}_{32} \\
\dot{\gamma}_{22} = -\dot{\gamma}_{12} - \dot{\gamma}_{32} \\
\dot{\gamma}_{23} = -\dot{\gamma}_{13} - \dot{\gamma}_{32} \\
\dot{\gamma}_{32} = \dot{\gamma}_{32} \\
\dot{\gamma}_{33} = -\dot{\gamma}_{13}
\end{cases}$$

Há uma correspondência entre os parâmetros do Modelo- $\Sigma$  e do Modelo-S, para o conjunto de dados, expressa por:

$$\begin{split} \dot{\alpha}_1 &= \tfrac{2}{3}\alpha_1 - \tfrac{1}{3}\alpha_2 - \tfrac{1}{3}\alpha_3 \\ &+ \tfrac{7}{43}\gamma_{11} + \tfrac{26}{129}\gamma_{12} + \tfrac{7}{43}\gamma_{13} + \tfrac{6}{43}\gamma_{14} - \tfrac{5}{43}\gamma_{21} - \tfrac{10}{129}\gamma_{22} - \tfrac{6}{43}\gamma_{24} - \tfrac{2}{43}\gamma_{31} - \tfrac{16}{129}\gamma_{32} - \tfrac{7}{43}\gamma_{33} \\ \dot{\alpha}_2 &= -\tfrac{1}{3}\alpha_1 + \tfrac{2}{3}\alpha_2 - \tfrac{1}{3}\alpha_3 \\ &- \tfrac{1}{43}\gamma_{11} - \tfrac{16}{129}\gamma_{12} - \tfrac{1}{43}\gamma_{13} - \tfrac{7}{43}\gamma_{14} + \tfrac{13}{43}\gamma_{21} + \tfrac{26}{129}\gamma_{22} + \tfrac{7}{43}\gamma_{24} - \tfrac{12}{43}\gamma_{31} - \tfrac{10}{129}\gamma_{32} - \tfrac{1}{43}\gamma_{33} \\ \dot{\beta}_1 &= \tfrac{3}{4}\beta_1 - \tfrac{1}{4}\beta_2 - \tfrac{1}{4}\beta_3 - \tfrac{1}{4}\beta_4 \\ &+ \tfrac{12}{35}\gamma_{11} - \tfrac{2}{23}\gamma_{12} - \tfrac{8}{51}\gamma_{13} - \tfrac{8}{81}\gamma_{14} + \tfrac{9}{31}\gamma_{21} - \tfrac{6}{43}\gamma_{22} - \tfrac{8}{53}\gamma_{24} + \tfrac{5}{43}\gamma_{31} - \tfrac{1}{43}\gamma_{32} - \tfrac{12}{129}\gamma_{33} \\ \dot{\beta}_2 &= -\tfrac{1}{4}\beta_1 + \tfrac{3}{4}\beta_2 - \tfrac{1}{4}\beta_3 - \tfrac{1}{4}\beta_4 \\ &- \tfrac{2}{23}\gamma_{11} + \tfrac{2}{7}\gamma_{12} - \tfrac{2}{23}\gamma_{13} - \tfrac{20}{181}\gamma_{13} - \tfrac{5}{43}\gamma_{21} + \tfrac{10}{39}\gamma_{22} - \tfrac{6}{43}\gamma_{24} - \tfrac{2}{43}\gamma_{31} + \tfrac{9}{43}\gamma_{32} - \tfrac{7}{43}\gamma_{33} \\ \dot{\beta}_3 &= -\tfrac{1}{4}\beta_1 - \tfrac{1}{4}\beta_2 + \tfrac{3}{4}\beta_3 - \tfrac{1}{4}\beta_4 \\ &- \tfrac{8}{81}\gamma_{11} - \tfrac{20}{181}\gamma_{12} + \tfrac{41}{102}\gamma_{13} - \tfrac{9}{47}\gamma_{14} + \tfrac{3}{86}\gamma_{21} + \tfrac{1}{43}\gamma_{22} - \tfrac{5}{86}\gamma_{24} - \tfrac{8}{43}\gamma_{31} - \tfrac{9}{43}\gamma_{32} + \tfrac{15}{43}\gamma_{33} \\ \dot{\gamma}_{13} &= -\tfrac{5}{33}\gamma_{11} - \tfrac{6}{43}\gamma_{12} + \tfrac{15}{43}\gamma_{13} - \tfrac{5}{86}\gamma_{14} - \tfrac{3}{86}\gamma_{21} - \tfrac{1}{43}\gamma_{22} + \tfrac{5}{86}\gamma_{24} + \tfrac{8}{43}\gamma_{31} + \tfrac{7}{43}\gamma_{32} - \tfrac{15}{43}\gamma_{33} \\ \dot{\gamma}_{14} &= -\tfrac{4}{43}\gamma_{11} - \tfrac{6}{43}\gamma_{12} + \tfrac{15}{43}\gamma_{13} - \tfrac{5}{86}\gamma_{14} - \tfrac{3}{86}\gamma_{21} - \tfrac{1}{43}\gamma_{22} + \tfrac{5}{86}\gamma_{24} + \tfrac{8}{43}\gamma_{31} + \tfrac{7}{43}\gamma_{32} - \tfrac{15}{43}\gamma_{33} \\ \dot{\gamma}_{32} &= \tfrac{6}{63}\gamma_{11} - \tfrac{6}{43}\gamma_{12} + \tfrac{15}{43}\gamma_{13} - \tfrac{5}{86}\gamma_{14} - \tfrac{3}{86}\gamma_{21} - \tfrac{1}{43}\gamma_{22} + \tfrac{5}{86}\gamma_{24} + \tfrac{8}{43}\gamma_{31} + \tfrac{7}{43}\gamma_{32} - \tfrac{15}{43}\gamma_{33} \\ \dot{\gamma}_{32} &= \tfrac{6}{63}\gamma_{11} - \tfrac{6}{43}\gamma_{12} + \tfrac{15}{43}\gamma_{13} - \tfrac{5}{86}\gamma_{14} - \tfrac{3}{86}\gamma_{21} - \tfrac{1}{43}\gamma_{22} + \tfrac{5}{86}\gamma_{24} + \tfrac{8}{43}\gamma_{31} + \tfrac{7}{43}\gamma_{32} - \tfrac{15}{43}\gamma_{33} \\ \dot{\gamma}_{32} &= \tfrac{6}{63}\gamma_{11} - \tfrac{6}{63}\gamma_{12} + \tfrac{15}{43}\gamma_{13} - \tfrac{5}{86}\gamma_{14} - \tfrac{3}{86}\gamma_{21} - \tfrac{1}{43}\gamma_{22} + \tfrac{5}{86}$$

# Matricialmente, tem-se:

# 4.2.3.2. Procedimento GLM do Sistema Estatístico SAS

As Soma de Quadrados do Tipo III do SAS/PROC GLM utiliza o Modelo- $\Sigma$ , que pode ser solicitada através da especificação "SS3" (Soma de Quadrados do Tipo III) na macro. Sendo que, usando "E3" (Funções Estimáveis do Tipo III), obtém-se as funções estimáveis do Tipo III no Modelo-S.

4.2.3.3. Equações Normais e Aproximação de Mínimos Quadrados

O SEN tem a seguinte apresentação:

$$\begin{bmatrix} 18 & 2 & 1 & 1 & 1 & 0 & 1 & 3 & 1 & 3 \\ 2 & 12 & 5 & -3 & -1 & -2 & 1 & 1 & 1 & 1 \\ 1 & 5 & 11 & -2 & -1 & -4 & 0 & 2 & 0 & -1 \\ 1 & -3 & -2 & 9 & 4 & 4 & 1 & 3 & 1 & -2 \\ 1 & -1 & -1 & 4 & 9 & 4 & 0 & 0 & 0 & -1 \\ 0 & -2 & -4 & 4 & 4 & 8 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 7 & 3 & 3 & 2 \\ 3 & 1 & 2 & 3 & 0 & 0 & 3 & 9 & 3 & -2 \\ 1 & 1 & 0 & 1 & 0 & 0 & 3 & 3 & 7 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 3 & 3 & 7 & 0 \\ 3 & 1 & -1 & -2 & -1 & 0 & 2 & -2 & 0 & 5 \end{bmatrix} \dot{\gamma}_{12}$$

A solução,  $\dot{\theta}^0$  e a aproximação de mínimos quadrados,  $\hat{y}$  são dadas a seguir. Conforme comentado anteriormente, a aproximação de mínimos quadrados é invariante, pois os modelos aqui utilizados .são equivalentes

$$\dot{\theta}^{0} = (Z'Z)^{-1} Z'y = \begin{bmatrix} 69,520 \\ -11,395 \\ -0,872 \\ -13.386 \\ -4,520 \\ 2,543 \\ -1,104 \\ 1,831 \\ -0,988 \\ -2,267 \end{bmatrix}; \qquad \hat{y} = Z\dot{\theta}^{0} = \begin{bmatrix} 45 \\ 52,5 \\ 62,5 \\ 72,5 \\ 55 \\ 67,5 \\ 67,5 \\ 85 \\ 85 \\ 72,5 \\ 75 \\ 82,5 \\ 82,5 \end{bmatrix}$$

4.2.3.4. Hipóteses Testadas na Análise de Variância

Utilizando as idéias sobre parametrização sucessiva e a notação R(), obtém-se as seguintes somas de quadrados.

$$R(\dot{\mu}, \dot{\alpha}, \dot{\beta})_{\Sigma} = 85.162,158$$
  
 $R(\dot{\mu}, \dot{\alpha}, \dot{\gamma})_{\Sigma} = 83.659,396$   
 $R(\dot{\mu}, \dot{\beta}, \dot{\gamma})_{\Sigma} = 83.853,911$ 

 $R(\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma})_{\Sigma} = 85.225$ 

Para testar a hipótese 
$$H_0^3$$
:

$$R(\dot{\alpha}|\dot{\mu},\dot{\beta},\dot{\gamma})_{\Sigma} = R(\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma})_{\Sigma} - R(\dot{\mu},\dot{\beta},\dot{\gamma})_{\Sigma} = 85.225 - 83.853,25 = 1.371,089$$

Para testar a Hipótese  $H_0^7$ :

$$R(\dot{\beta}|\dot{\mu},\dot{\alpha},\dot{\gamma})_{\Sigma} = R(\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma})_{\Sigma} - R(\dot{\mu},\dot{\alpha},\dot{\gamma})_{\Sigma} = 1.565,604$$

Para testar a Hipótese  $H_0^9$ :

$$R(\dot{\gamma}|\dot{\mu},\dot{\alpha},\dot{\beta})_{\Sigma} = R(\dot{\mu},\dot{\alpha},\dot{\beta},\dot{\gamma})_{\Sigma} - R(\dot{\mu},\dot{\alpha},\dot{\beta})_{\Sigma} = 85.225 - 85.162,158 = 62,842$$

O Modelo- $\Sigma$  testa hipóteses do Tipo III. Como pode ser verificado comparando-se as somas de quadrados obtidas através do Modelo- $\Sigma$  com as do SAS, presentes na tabela 17.

#### 4.2.4. Modelo-Z

# 4.2.4.1. Caracterização

O modelo tem a seguinte caracterização Para a ordem A-B

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

Para a ordem B-A

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

Para o Modelo-Z, faz-se o primeiro nível de cada fator igual a zero, utiliza-se a seguinte notação: 0<sub>11</sub>, como descrito em SEARLE *et alii*(1981). Tem-se as seguintes restrições impostas.

$$\begin{cases} \alpha_{1} = 0 \\ \beta_{1} = 0 \end{cases} \Rightarrow \begin{cases} \gamma_{11} = 0 \\ \gamma_{12} = 0 \\ \gamma_{13} = 0 \\ \gamma_{14} = 0 \\ \gamma_{21} = 0 \\ \gamma_{31} = 0 \end{cases}$$

Como no Modelo-S, no Modelo-Z, alterando a ordem dos efeitos principais, a matriz de delineamento tem configuração diferente. Isso tem implicação semelhante àquele modelo.

Matricialmente para a ordem A-B, tem-se

4.2.4.2. Sistema Estatístico GLIM

Esse sistema utiliza o Modelo-Z. Para tanto, impõe uma restrição, fazendo o primeiro nível de cada fator igual a zero.

Na Tabela 27 são mostradas as macro para análise, nas ordens A-B e B-A, com as respectivas saídas. Pode-se montar um quadro de Análise de Variância utilizando-se as hipóteses do Tipo II para ambos os efeitos principais, como para o SAS (Tabela 16).

<u>Tabela 27</u>: Macros e Análises de Variância pelo GLIM

PARA A ORDEM AB	PARA A ORDEM BA					
[i] ? \$data 18 a b rep y\$ [i] ? \$dinput 'r:\glim1.txt' \$	[i] ? \$data 18 a b rep y\$ [i] ? \$dinput 'r:\glim1.txt' \$					
[i] ? \$look a b y\$	[i] ? \$look a b y\$					
[0] A B Y	[0] A B Y					
[o] 1 1.000 1.000 45.00	[o] 1 1.000 1.000 <b>45</b> .00					
[o] 2 1.000 2.000 50.00	[o] 2 1.000 2.000 50.00					
[o] 3 1.000 2.000 55.00	[o] 3 1.000 2.000 55.00					
[o] 4 1.000 3.000 60.00	[o] 4 1.000 3.000 60.00					
[o] 5 1.000 3.000 65.00	[o] 5 1.000 3.000 65.00					
[o] 6 1.000 4.000 70.00	[o] 6 1.000 4.000 70.00					
[o] 7 1.000 4.000 75.00	[o] 7 1.000 4.000 75.00					
[o] 8 2.000 1.000 50.00	[o] 8 2.000 1.000 50.00					
[o] 9 2.000 1.000 60.00	[o] 9 2.000 1.000 60.00					
[o] 10 2.000 2.000 70.00	[o] 10 2.000 2.000 70.00					
[0] 11 2.000 2.000 65.00	[o] 11 2.000 2.000 65.00					
[0] 12 2.000 4.000 80.00	[o] 12 2.000 4.000 80.00					
[o] 13 2.000 4.000 90.00	[o] 13 2.000 4.000 90.00					
[o] 14 3.000 1.000 75.00	[o] 14 3.000 1.000 75.00					
[o] 15 3.000 1.000 70.00	[o] 15 3.000 1.000 70.00					
[o] 16 3.000 2.000 75.00	[o] 16 3.000 2.000 75.00					
[o] 17 3.000 3.000 85.00	[o] 17 3.000 3.000 85.00					
[o] 18 3.000 3.000 80.00	[o] 18 3.000 3.000 80.00					
[i] ? \$factor a 3 b 4 \$yvar y \$fit:+a:+b:+a.b \$	[i] ? \$factor a 3 b 4 \$yvar y \$fit:+b:+a:+b.a \$					
[o] deviance = 2711.1	[o] deviance = 2711.1					
[o] residual df = 17	[o] residual df = 17					
[0]	[0]					
[o] deviance = 1850.8 (change = -860.3)	[o] deviance = 1723.7 (change = -987.4)					
[o] residual df = 15 (change = -2 )	[o] residual df = 14 (change = -3 )					
[0]	[0]					
[o] deviance = 237.84 (change = -1613.)	[o] deviance = 237.84 (change = -1486.)					
[o] residual df = $12$ (change = $-3$ )	[o] residual df = 12 (change = -2)					
[0]	[0]					
	[o] deviance = 175.00 (change = -62.84)					
[o] residual df = 8 (change = -4 )	[o] residual df = 8 (change = -4 )					
	[0]					
[i]?\$stop	[i] ? \$stop					

Ao comparar os sistemas estatísticos, podem haver diferenças de aproximação no cálculo das somas de quadrados. No caso do Glim, por exemplo, a soma de quadrados para A, na ordem A-B é 860,3. No caso do SAS é igual a 860,278. Testam, entretanto, hipóteses semelhantes.

Para tanto, toma-se a soma de quadrados do fator A, na ordem B-A e a soma de quadrados do fator B na ordem B-A.

Foi colocada a importância da ordem dos efeitos dos fatores principais no Modelo-Z. Assim, o GLIM depende da ordem de entrada dos efeitos dos fatores principais para a saída da Análise de Variância. Semelhante, no que se refere às hipóteses dos Tipos I e II,

testadas no Modelo-S, o GLIM faz um ajuste seqüencial dos efeitos dos fatores principais (SANTOS, 1994).

A primeira soma de quadrados é a do modelo (com respectivos graus de liberdade). Dessa, subtrai a soma de quadrados devido ao primeiro fator. Em seguida, subtrai a soma de quadrados devido ao segundo fator. A soma de quadrados restante é o resíduo.

4.2.4.3. Equações Normais e Aproximação de **M**ínimos Quadrados

O Modelo-Z apresenta o seguinte Sistema de Equações Normais.

A solução  $\pi^0$  e a aproximação de mínimos quadrados  $\hat{y}$  são dadas a seguir. Onde, se observa a invariância da aproximação de mínimos quadrados.

Essas soluções dizem respeito a uma comparação feita entre o nível 1 (neste exemplo), de cada efeito principal e os outros níveis. Bem como as respectivas interações. Assim, esse modelo é interessante quando se tem um tratamento dado como testemunha.

# 4.2.4.4. Hipóteses Testadas na Análise de Variância

As hipóteses testadas no Modelo-Z são semelhantes às formuladas no Modelo-S, onde a ordenação dos efeitos principais é importante. As somas de quadrados são as mesmas. As restrições impostas no Modelo-Z podem entretanto, levar a testar hipóteses diferentes daquelas testadas no Modelo-S.

#### 4.2.5. Modelo-W

## 4.2.5.1. Caracterização

Tem a caracterização para a ordem A-B:

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

Tem a caracterização para a ordem B-A:

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

Para o conjunto de dados tem-se as seguintes restrições.

$$\sum_{i} n_{i} \alpha_{i} = 0 \Rightarrow 7\alpha_{1} + 6\alpha_{2} + 5\alpha_{3} = 0 \Leftrightarrow \alpha_{3} = -\frac{7}{5}\alpha_{1} - \frac{6}{5}\alpha_{2};$$

$$\sum_{j} n_{j} \beta_{j} = 0 \Rightarrow 5\beta_{1} + 5\beta_{2} + 4\beta_{3} + 4\beta_{4} = 0 \Leftrightarrow \beta_{4} = -\frac{5}{4}\beta_{1} - \frac{5}{4}\beta_{2} - \beta_{3}.$$

$$\sum_{j} n_{ij} \gamma_{ij} = 0, \ \forall \ i \Rightarrow \begin{cases} \gamma_{11} + 2\gamma_{12} + 2\gamma_{13} + 2\gamma_{14} = 0 \\ 2\gamma_{21} + 2\gamma_{22} + 2\gamma_{24} = 0 \\ 2\gamma_{31} + \gamma_{32} + 2\gamma_{33} = 0 \end{cases}$$

$$\sum_{i} n_{ij} \gamma_{ij} = 0, \ \forall \ j \Rightarrow \begin{cases} \gamma_{11} + 2\gamma_{21} + 2\gamma_{31} = 0 \\ 2\gamma_{12} + 2\gamma_{22} + \gamma_{32} = 0 \\ 2\gamma_{12} + 2\gamma_{22} + \gamma_{32} = 0 \\ 2\gamma_{13} + 2\gamma_{33} = 0 \\ 2\gamma_{14} + 2\gamma_{24} = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \gamma_{11} = -2\gamma_{12} - 2\gamma_{13} - 2\gamma_{14} \\ \gamma_{12} = \gamma_{12} \\ \gamma_{13} = \gamma_{13} \\ \gamma_{21} = \gamma_{12} + \gamma_{14} + \frac{1}{2}\gamma_{32} \\ \gamma_{22} = -\gamma_{12} - \frac{1}{2}\gamma_{32} \\ \gamma_{24} = -\gamma_{14} \\ \gamma_{31} = \gamma_{13} - \frac{1}{2}\gamma_{32} \\ \gamma_{32} = \gamma_{32} \\ \gamma_{33} = -\gamma_{13} \end{cases}$$

## Matricialmente, tem-se para a ordem A-B

Para a ordem B-A, altera-se a ordem das colunas 2 e 3(fator A) com as colunas 4, 5 e 6(fator B), com implicação nas colunas da interação

#### 4.2.5.2. Sistema Estatístico SPSS/ANOVA

É utilizada uma versão "for windows". Nessa, existem duas opções de saídas: "SEQUENTIAL" e "UNIQUE". A primeira faz o ajuste seqüencial, utilizando o Modelo-W, com uma saída semelhante à do Tipo I do SAS; a segunda, faz o ajuste de um efeito, em relação aos efeitos restantes, utilizando o Modelo-Σ (SEARLE, 1987), com uma saída semelhante à do Tipo III do SAS. Faz-se a ressalva que, para a segunda, não deve haver casela vazia.

Quando há casela(s) vazia(s), as duas saídas são obtidas. Na saída da "UNIQUE" é colocada uma advertência de que as hipóteses relacionadas podem não ser de interesse do pesquisador. Para a "SEQUENTIAL", é importante a ordem dada aos efeitos principais. A seguir, as análises.

a)Para a ordem A-B:

Tests of Significance for VAR00003 using SEQUENTIAL Sums of Squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	175.00	8	21.88		
VAR00001	860.28	2	430.14	19.66	.001
VAR00002	1612.99	3	537.66	24.58	.000
VAR00001 BY VAR00002	62.84	4	15.71	.72	.603

b)Para a ordem B-A:
Tests of Significance for VAR00003 using SEQUENTIAL Sums of Squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	175.00	8	21.87		
VAR00002	987.36	3	329.12	15.05	.001
VAR00001	1485.91	2	742.95	33.96	.000
VAR00002 BY VAR00001	62.84	4	15.71	.72	.603

4.2.5.3. Equações Normais e Aproximação de **M**ínimos Quadrados

O Modelo-W apresenta o seguinte Sistema de Equações Normais.

A solução  $\omega^0$  e a aproximação de mínimos quadrados  $\hat{y}$  são dadas a seguir. Verificando-se a invariância comentada anteriormente.

$$\omega^{0} = (T'T)^{-1}T'y = \begin{bmatrix} 67,778 \\ -10,498 \\ 1,248 \\ -11,457 \\ -3,717 \\ 3,371 \\ -1,061 \\ 1,849 \\ -0.376 \\ -2,260 \end{bmatrix}; \qquad \hat{y} = T\omega^{0} = \begin{bmatrix} 45 \\ 52,5 \\ 62,5 \\ 62,5 \\ 72,5 \\ 75,5 \\ 85 \\ 85 \\ 72,5 \\ 75,82,5 \\ 75,82,5 \end{bmatrix}$$

4.2.5.4. Hipóteses Testadas na Análise de Variância

As somas de quadrados no Modelo-W são iguais às obtidas através do Modelo-S. As hipóteses formuladas a partir do Modelo-W podem entretanto, não ter total correspondência com aquelas formuladas a partir do Modelo-S.

#### 4.2.6. Modelo-R

## 4.2.6.1. Caracterização

Para o conjunto de dados apresentado utilizando-se o Modelo-R, fixa-se o primeiro nível dos fatores A e B. É interessante observar que a interação será resultado da combinação dos níveis correspondentes. Assim:, utilizando-se variáveis binárias

$$u_{ij1} = \begin{cases} 1, & \text{se } i = 2 \text{(nivel 2 do fator A)} \\ 0, & \text{caso contrario} \end{cases}$$

$$u_{ij2} = \begin{cases} 1, & \text{se } i = 3 \text{(nivel 3 do fator A)} \\ 0, & \text{caso contrario} \end{cases}$$

$$u_{ij3} = \begin{cases} 1, & \text{se } j = 2 \text{(nivel 2 do fator B)} \\ 0, & \text{caso contrario} \end{cases}$$

$$u_{ij4} = \begin{cases} 1, & \text{se } j = 3 \text{(nivel 3 do fator B)} \\ 0, & \text{caso contrario} \end{cases}$$

$$u_{ij5} = \begin{cases} 1, & \text{se } j = 4 \text{(nivel 4 do fator B)} \\ 0, & \text{caso contrario} \end{cases}$$

$$u_{ij5} = \begin{cases} 1, & \text{se } j = 4 \text{(nivel 4 do fator B)} \\ 0, & \text{caso contrario} \end{cases}$$

Fixando-se, primeiro nível do fator A e o primeiro nível do fator B, tem-se a seguinte caracterização para a ordem A-B

$$y_{ijk} = \delta_0 + \underbrace{u_{ij1}\delta_1 + u_{ij2}\delta_2}_{fator A} + \underbrace{u_{ij3}\delta_3 + u_{ij4}\delta_4 + u_{ij5}\delta_5}_{fator B} + \underbrace{u_{ij1}u_{ij3}\delta_6 + u_{ij1}u_{ij5}\delta_7 + u_{ij2}u_{ij3}\delta_8 + u_{ij2}u_{ij4}\delta_9 + e_{ijk}}_{i=2;j=2} + \underbrace{i=2;j=4}_{i=3;j=2} + \underbrace{i=3;j=3}_{interacao}$$

Tem a caracterização para a ordem B-A

$$y_{jik} = \delta_0 + \underbrace{u_{ij3}\delta_3 + u_{ij4}\delta_4 + u_{ij5}\delta_5}_{fator B} + \underbrace{+u_{ij1}\delta_1 + u_{ij2}\delta_2}_{fator A}$$

$$+\underbrace{u_{ij3}u_{ij1}\delta_6 + u_{ij3}u_{ij2}\delta_7 + u_{ij4}u_{ij2}\delta_8 + u_{ij5}u_{ij1}\delta_9 + e_{jik}}_{j=2;i=2}$$

$$\underbrace{j=2;i=2}_{interacao} \underbrace{j=3;i=3}_{j=3;i=3} \underbrace{j=4;i=2}_{j=4;i=2}$$

Como no Modelo-S, a alteração na ordem dos fatores é importante, modificando as somas de quadrados do Tipo I.

É importante observar que, para a interação, as variáveis binárias criadas são em função das caselas

em que ocorre observações. Pois, a princípio ter-se-ia seis combinações entre os dois fatores A e B. Mas, há duas caselas vazias, reduzindo para quatro o número de parâmetros para a interação.

Para a ordem A-B, tem-se matricialmente:

#### 4.2.6.2. Sistema Estatístico MINITAB

O MINITAB utiliza o Modelo-R. É um modelo, em termos de concepção, semelhante ao Modelo-Z. Pois, nos dois casos, promove-se uma comparação de um nível de cada fator com os demais níveis. Sendo que, no primeiro uma restrição; no segundo, uma reparametrização.

Para o MINITAB, a ordenação dos efeitos principais também é importante. Para saída é feito ajuste seqüencial. Como ocorre para o GLIM e SPSS/ANOVA-"SEQUENTIAL". A seguir as macros, com as respectivas saídas.

# a)Para a ordem A-B:

```
MTB > name c1='Fat_A' c2='Fat_B' c3='rep' c4='Prod'
MTB > set c1
DATA> 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3
DATA> end
MTB > set c2
DATA> 1 2 2 3 3 4 4 1 1 2 2 4 4 1 1 2 3 3
DATA> end
MTB > set c3
DATA> end
MTB > set c4
DATA> 45 50 55 60 65 70 75 50 60 70 65 80 90 75 70 75 85 80
DATA> end
MTB > glm c4=c1|c2
General Linear Model
Factor Levels Values
Fat_A
          3
                  2
Fat B
                       3
Analysis of Variance for Prod
Source
               Model DF Reduced DF Seq SS
                                      2
                                                    860.28
Fat_A
Fat_B
                   3
                                      3
                                                   1612.99
Fat_A*Fat_B
                               4+
                                                     62.84
Error
                   6
                                      8
                                                    175.00
Total
                   17
                                      17
                                                   2711.11
```

# b) Para a ordenação B-A:

MTB > name c1='Fat\_A' c2='Fat\_B' c3='rep' c4='Prod'
MTB > set c1
DATA> 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3
DATA> end
MTB > set c2
DATA> 1 2 2 3 3 4 4 1 1 2 2 4 4 1 1 2 3 3
DATA> end
MTB > set c3
DATA> 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 1 2
DATA> end
MTB > set c4
DATA> end
MTB > set c4
DATA> 45 50 55 60 65 70 75 50 60 70 65 80 90 75 70 75 85 80
DATA> end
MTB > glm c4=c2|c1

General Linear Model Factor Levels Values

Fat\_B 4 1 2 3 4 Fat\_A 3 1 2 3

Analysis of Variance for Prod

Model DF	Reduced DF	Seg SS	
3		3	987.36
2		2	1485.91
6	4+		62.84
6		8	175.00
17		17	2711.11
	3 2 6 6	3 2 6 4+	3 3 2 6 4+ 8

4.2.6.3. Equações Normais e Aproximação de Mínimos Quadrados

# O modelo apresenta o seguinte SEN.

[18	6	5	5	4	4	2	2	I	2	$\lceil \delta_{o} \rceil$		[1220]	
6	6	0	2	0	2	2	2	0	0	$\delta_{I}$		420	
5	0	5	1	2	0	0	0	I	2	$\delta_2$		395	
5	2	1	5	0	0	2	0	I	0	$\delta_3$		320	
4	0	2	0	4	0	0	0	0	2	$\delta_4$		315	
4	2	0	0	0	4	0	2	0	0	$\delta_s$	=	255	
2	2	0	2	0	0	2	0	0	0	$\delta_6$		135	
2	2	0	0	0	2	0	2	0	0	$\delta_7$		150	
1	0	1	1	0	0	0	0	1	0	$\delta_{s}$		75	
2	0	2	0	2	0	0	0	0	2	$\left[\delta_{g}\right]$		170	

A solução,  $\delta^0$  e a aproximação de mínimos quadrados,  $\hat{y}$  são dadas a seguir. Verificando-se a invariância comentada anteriormente.

$$\delta^{o} = (U U)^{-1} U y = \begin{bmatrix} 45 \\ 10 \\ 27.5 \\ 7.5 \\ 17.5 \\ 27.5 \\ 5 \\ 2.5 \\ -5 \\ -7.5 \end{bmatrix}; \qquad \hat{y} = U \delta^{o} = \begin{bmatrix} 45 \\ 62.5 \\ 62.5 \\ 72.5 \\ 72.5 \\ 55 \\ 67.5 \\ 67.5 \\ 85 \\ 85 \\ 72.5$$

4.2.6.4. Hipóteses Testadas na Análise de Variância

São utilizadas as idéias sobre parametrização sucessiva para obtenção das somas de quadrados. Utilizando a notação R( ).

$$R(\delta_{0}) = 82.688,889$$

$$R(\delta_{0}, \delta_{1}, \delta_{2}) = 83.549,167$$

$$R(\delta_{0}, \delta_{3}, \delta_{4}, \delta_{5}) = 83.676,25$$

$$R(\delta_{0}, \delta_{1}, \delta_{2}, \delta_{3}, \delta_{4}, \delta_{5}) = 85.162,158$$

$$R(\delta_{0}, \delta_{1}, \delta_{2}, \delta_{3}, \delta_{4}, \delta_{5}, \delta_{6}, \delta_{7}, \delta_{8}, \delta_{9}) = 85225$$

A soma de quadrados devido a  $H^1_{ heta}$  é dada por:

$$R(\delta_1, \delta_2 | \delta_0) = R(\delta_0, \delta_1, \delta_2) - R(\delta_0) = 83.549,167 - 82.688,889 = 860,278$$

A soma de quadrados devido a  $H_{ heta}^{ extstyle 5}$  é dada por:

$$R(\delta_3, \delta_4, \delta_5 | \delta_0) = R(\delta_0, \delta_3, \delta_4, \delta_5) - R(\delta_0) = 83.676, 25 - 82.688, 889 = 987, 361$$

A soma de quadrados devido a  $H_{ heta}^2$  é dada por:

$$R(\delta_1, \delta_2 | \delta_0, \delta_3, \delta_4, \delta_5) = R(\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5) - R(\delta_0, \delta_3, \delta_4, \delta_5) =$$
= 85.162.158 - 83.676.25 = 1.485.908

A soma de quadrados devido a  $H_{ heta}^{ extit{6}}$  é dada por:

$$R(\delta_3, \delta_4, \delta_5 | \delta_0, \delta_1, \delta_2) = R(\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5) - R(\delta_0, \delta_1, \delta_2) =$$

$$= 85.162,158 - 83.549,167 = 1.612,991$$

A soma de quadrados devido a  $H_{ heta}^{ heta}$  é dada por:

$$R(\delta_6, \delta_7, \delta_8, \delta_9 | \delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5) =$$

$$= R(\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8, \delta_9) - R(\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5) =$$

$$= 85.225 - 85.162, 158 = 62,842$$

## 5. CONCLUSÕES

# Face ao exposto, conclui-se que:

- -Os modelos discutidos são equivalentes, apresentando o mesmo vetor de aproximação de mínimos quadrados. Como conseqüência, apresentam a mesma soma de quadrados de parâmetros.
- -O SAS/PROC GLM é uma alternativa importante para determinar as funções estimáveis e consequentes hipóteses testáveis.
- -Para o exemplo discutido, as hipóteses dos Tipos I, II, III e IV diferem entre si, apresentando similaridade entre o Modelo-S e o Modelo-M. A hipótese do Tipo III apresenta semelhança com o Modelo- $\Sigma$ . Por sua vez, as hipóteses dos Tipos I e II no Modelo-S são semelhantes no Modelo-Z, Modelo-W e Modelo-R.

- -As hipóteses determinadas através do Modelo-M são de mais simples compreensão do que do Modelo-S.
- -A interação apresenta, em todos os modelos e possíveis quadros de análise de variância, hipóteses semelhantes e mesma soma de quadrados, resultado do ajuste em relação aos demais efeitos.
- -A ordenação dos efeitos principais é fundamental para a determinação das hipóteses do Tipo I, e das somas de quadrados associadas.
- -A ordenação dos efeitos principais não interfere na determinação das hipóteses dos Tipos II, III e IV e das somas de quadrados associadas.
- -O Modelo-Z e o Modelo-R apresentam, em suas soluções de mínimos quadrados, contrastes entre um tratamento fixado "como testemunha" e os demais.
- -Dos sistemas estatísticos estudados, o SAS/PROC GLM é o único que obtém somas de quadrados para os quatro tipos de hipóteses sobre os efeitos principais. O BMDP não realiza a análise, em modelos com interação na presença de caselas vazias, enquanto os demais têm como resultados somas de quadrados com ajuste

sequencial, comparável à Soma de Quadrados do Tipo I do SAS/PROC GLM.

- -O quadro de análise que testa hipóteses do Tipo II é fornecido naturalmente pelo SAS/PROC GLM. Para os demais sistemas utilizados, pode-se obter indiretamente tal quadro utilizando as somas de quadrados obtidas pelo ajuste sequencial dos efeitos principais. Para tanto, toma-se a soma de quadrados para o fator A, na ordem B-A, e a soma de quadrados para o fator B, na ordem A-B.
  -No que se refere às hipóteses testadas na realização da análise de variância através dos sistemas estatísticos, a documentação consultada mostrou-se, muitas vezes ineficaz para prestar esclarecimentos sobre o tema, mesmo nos casos do SAS e do BMDP, que se mostraram
- -Considera-se fato fortemente relevante, que o pesquisador ao analisar seus dados, através de sistemas estatísticos, em particular em modelos com interação na presença de caselas vazias, esteja consciente da importância de conhecer a verdadeira hipótese testada através do sistema adotado.

mais esclarecedores.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- BRANDT, A.E. The analysis of variance in a 2x's table with disproportionate frequencies. **Journal of the American Statistical Association**, **28**: 164, 1933.
- BURDICK, D.S.Rows adjusted for columns versus simple row means in two-way ANOVA: when are the two hypotheses equivalent? **Journal of the American Association**, **74**(366): 457-458, 1979.
- CAMARINHA FILHO, J.A. Testes de hipóteses em modelos lineares com dados desbalanceados e caselas vazias. Piracicaba, 1995. 142p. (Mestrado Escola Superior de Agricultura "Luiz de Queiroz"/USP)
- DALLAL, G.E. The computer analysis of factorial experiments with nested factors. **The American Statistician**, ,**46**(3): 240, 1992.
- DIXON, W.J. **BMDP Statistical Software Manual**, vol. 2. University of California Press, Berkeley, California, 1989. 542p.
- ELSTON, R.C. & BUSH, N. The hipotheses that can br tested when there are interactions in analysis of variance model. **Biometrics**, 681-699, 1964.
- FINNEY, D.J. Main effects and interactions. **Journal of the American**Statistical Association, 43: 566-571, 1948.
- FRANCIS, I. A comparison of several analysis of variance programs. **Journal of the American Statistical Association**, **68**(344): 860-865, 1973.
- FREUND, R.J. The case of the missing cell. **The American Statistician**, **34**(2): 94-98, 1980.
- GLIM. Manual Realease 3.77. Royal Statistical Society, London, United Kingdom, 1985, 263 p.
- GOSSLEE, D.G. & LUCAS, H.L. Analysis of variance of disproportionate data when interactions is present. **Biometrics**, 115-133, 1965.

- HENDERSON, C.R. Estimaton of variance and covariance components. **Biometrics**, 227-252, 1953.
- HENDERSON, C.R. & McALLISTER, A.J. The missing subclass problem in twoway fixed models. **Journal of Animal Science**, **46**(4): 1125-1137, 1978.
- HERR, D.G. On the history of ANOVA in unbalanced, factorial designs: the first 30 years. **The American Statistician**, **40**(4): 265-270, 1986.
- HOCKING, R.R. & SPEED, F.M. A full rank analysis of some linear model problems. **Journal of the American Statistical Association**, **70**(351): 706-712, 1975.
- HOCKING, R.R. **The Analysis of Linear Models**. Monterrey, Brooks/Cole, 1985, 385p.
- IEMMA, A.F. Modelos lineares:uma introdução para profissionais da pesquisa agropecuária.Imprensa Oficial do Estado do Paraná, Londrina, 1987, 263 p.
- IEMMA, A.F. Álgebra de matrizes com aplicações em estatística, ESALQ-USP, 1990, 411 p.
- IEMMA, A.F. **Testes de hipóteses com amostras equilibradas**. Gembloux, F.S.A., 1991, 101 p.
- IEMMA, A.F. Análisis de varianza de experimentos con celdas vácias. Córdoba, publicações, 1993, 102 p.
- IEMMA, A.F.; PALM, R.; CLAUSTRIAUX, J.J. Sobre a construção de projetores ortogonais. **Revista de Matemática e Estatística**, São Paulo, **11**: 133-142, 1993.
- JENNINGS, E. & WARD JR., J.H. Hypothesis identification in the case of the missing cell. **The American Statistician**, **36**(1): 25-27, 1982.
- KRAMER, C.Y. On the analysis of variance of two-way classification with unequal sub-class numbers. **Biometrics**, 441-452, 1955.

- LONGLEY, J.W. An appraisal of least squares programs for the eletronic computer from the point of view of the user. **Journal of the American Statistical Association**, **62**(319): 819-929, 1967.
- MINITAB Reference Manual: Realease 8. Minitab Inc., USA. 1991. 623p.
  MONDARDO, M. Estimabilidade de funções paramétricas com dados desbalanceados através do PROC GLM do SAS: aplicações à pesquisas agropecuária. Piracicaba, 1994. 166p. (Mestrado Escola Superior de Agricultura "Luiz de Queiroz"/USP)
- MURRAY, L.W. A note on connectedness in fixed effects MANOVA an GMANOVA models with missing cells. Commun. Statist.-Theory and Meth., 18(7):2527-2531, 1989.
- MURRAY, L.W. & SMITH, D.W. Estimability, testability, and conectedness in the cell means model. Commun. Statist.-Theory and Meth., 14(8): 1889-1917, 1985.
- NETER, J. & WASSERMANN, W. Applied Linear Statistical Models. Richard D. Irwin, Homewood, III, 1977.
- OKTABA, W. Tests of hypoteses for the general Gauss-Markov model. **Biom. Journal**, **26**(4): 415-424, 1984.
- OVERAL, J.E. & SPIEGEL, D.K. Concerning Least Squares Analysis of Experimental Data. **Psycological Bulletin, 72**:311-322, 1969
- PATTERSON, R.E. The use of adjusting factors in the analysis of data with disproportionate subclass numbers. **Journal of the American Statistical Association**, **41**: 335-346, 1946.
- PEIXOTO, J.L. Testable hypotheses in singular fixed linear models. **Commun. Statist.-Theory and Meth.**, **15**(6): 1957-1973, 1986.
- RAO, C.R. On the linear combination of observations and the general theory of least squares. **Sankhyã**, **73**: 105-112. 1945.

- RIO, M. del. A cordinate-free approach to testable hypotheses. **Commun. Statist.-Theory and Meth.**, **18**(9) 3493-3503, 1989.
- SANTOS, E.S. Utilização de "softwares" estatísticos na interpretação de hipóteses com dados desbalanceados. Piracicaba, 1994. 175p. (Mestrado Escola Superior de Agricultura "Luiz de Queiroz"/USP)
- SAS. **User's Guide: Statistics Version**, SAS Institute,5th edition. Cary, USA. 1987, 956 p.
- SEARLE, S.R. Linear Models, John Willey & Sons, New York, 1971.
- SEARLE, S.R. Annoted computer output for analysis of variance of unequalsubclass-numbers data. **The American Statistician**, **33**(4): 222-223, 1979.
- SEARLE, S.R. Restrictions and generalized inverses in linear models. **The**American Statistician, 38(1): 53-54, 1984.
- SEARLE, S.R. Linear models for unbalanced data. New York, John Wiley, 1987.
- SEARLE, S.R. Analysis of variance computing package output for unbalanced data from fixed-effects models with nested factors. **The American Statistician**, **48**(2): 148-153, 1994.
- 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 Stitistician**, **35**(1): 16-33, 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**, **34**(4): 216-221, 1980.
- SEELY, J. Estimability and linear hypotheses. **The American Statistician**, **31**(3): 121-123, 1977.
- SINGH, K.P. & SINGH, U. A technique for the analysis of unbalanced data. **Biom. Journal**, **31**(1): 3-17, 1989.

- SNEDECOR, G.W. & COCHRAN, W.G. Statistical Methods. Ames, Iowa State University Press, 6ª ed., 1967.
- SMITH, D. W. & MURRAY, L.W. A simplified treatment of the estimation of parameters and tests of hypotheses in constrained design models with unbalanced data. **The American Statistician**, **37**(2): 156-158, 1983.
- SPEED, F.M. & HOCKING, R.R. The use of the R( )-notation with unbalanced data. **The American Statistician**, **30**(1): 30-33, 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**, **73**(361): 105-112, 1978.
- SPSS/PC<sup>+</sup> User's Guide. SPSS Incorporation, USA, 1986. 340p. STEINHORST, R.K. Resolving current controversies in analysis of variance. **The American Statistician**, **36**(2): 138-139, 1982.
- VAJDA, S. Technique of the analysis of variance. Nature, 160: 27, 1947.
- URQUHART, N.S. & WEEKS, D.L. Linear models in messy data: some problems and alternatives. **Biometrics**, **34**: 696-705, 1978.
- YATES, F.The principles of orthogonality and confundig in replicated experiments. **Journal Agricultural Science**, **23**(parte 1): 108-145, 1933.
- YATES, F. The analysis of multiple classifications with unequal numbers in the different classes. **Journal of the American Statistical Association**, 1934.
- YATES, F. Technique of the analysis of variance. Nature, 160: 472-473, 1947.