

**MODELOS LINEARES DE EFEITOS MISTOS: FORMULAÇÃO
GERAL E UTILIZAÇÃO DE ALGUNS SISTEMAS
COMPUTACIONAIS ESTATÍSTICOS**

MARIA IMACULADA DE LIMA MONTEBELO

Licenciada em Ciências

Orientador: Prof. Dr. ANTONIO FRANCISCO IEMMA

Tese apresentada à Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, para obtenção do título de Doutor em Agronomia. Área de concentração: Estatística e Experimentação Agronômica.

PIRACICABA

Estado de São Paulo - Brasil

Setembro - 1997

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

Montebelo, Maria Imaculada de Lima

Modelos lineares de efeitos mistos: formulação geral e utilização de alguns sistemas computacionais estatísticos / Maria Imaculada de Lima Montebelo. - - Piracicaba, 1997.

173 p.

Tese (doutorado) - Escola Superior de Agricultura Luiz de Queiroz, 1997.
Bibliografia.

1. Hipótese estatística 2. Inferência estatística 3. Informática aplicada 4.
Modelo matemático 5. Sistema de computador I. Título

CDD 519.56

**MODELOS LINEARES DE EFEITOS MISTOS: FORMULAÇÃO
GERAL E UTILIZAÇÃO DE ALGUNS SISTEMAS
COMPUTACIONAIS ESTATÍSTICOS**

MARIA IMACULADA DE LIMA MONTEBELO

Aprovada em: 14.11.1997

Comissão julgadora:

Prof. Dr. Antonio Francisco Iemma	ESALQ/USP
Prof. Dr. Carlos Tadeu dos Santos Dias	ESALQ/USP
Prof ^a . Dr ^a . Roseli Aparecida Leandro	ESALQ/USP
Prof. Dr. Eufrazio de Souza Santos	UFPE
Prof ^a . Dr ^a . Sheila Zambello de Pinho	FCA/UNESP

Prof. Dr. ANTONIO FRANCISCO IEMMA
Orientador



À Deus

OFEREÇO.

**Aos meus pais e
aos meus irmãos,**

AGRADEÇO.

**Ao meu esposo Francisco e
ao meu filho João Pedro, com
amor**

DEDICO.

AGRADECIMENTOS

Ao Prof. Dr. Antonio Francisco lemma, pela valiosa orientação e incentivos constantes que permitiram a realização deste trabalho.

Aos Professores do Departamento de Matemática e Estatística da ESALQ/USP, pelos ensinamentos recebidos.

Aos Funcionários do Departamento de Matemática e Estatística da ESALQ/USP, pela atenção e amizade.

Aos pesquisadores Antonio Cláudio Almeida, Cristian Andrés C. Acunã, Eliana Schammas, Maria Cristina Neves Oliveira, Prof. Dr. Cesar G. de Lima, Prof. Luiz M Salvarrey e Prof^a. Maria Izalina F. Alves, pelas assessorias recebidas durante o estudo dos sistemas estatísticos utilizados neste trabalho.

Aos colegas de pós-graduação, Cláudia, Dóris, João Gil, Liciano, Luiz, Pilar, Roseli e Silvio, pela agradável convivência durante o curso.

A amiga Prof^a. Dr^a. Dóris Albina Gomez Ticerán, pela amizade e incentivos recebidos durante o curso.

A Prof^a. Dr^a. Rosana Parente pelo apoio e estímulos durante a fase inicial do curso.

A Prof^a. Dr^a. Angela Maria C. Jorge Corrêa, pela amizade e sugestões apresentadas.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e tecnológico), pela bolsa de doutorado.

A todos que contribuíram para a realização deste trabalho.

SUMÁRIO

	Página
LISTA DE TABELAS	vi
RESUMO.....	xii
SUMMARY	xiv
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	3
2.1 Introdução ao estudo do modelo misto.	3
2.2 Estimação dos componentes de variância no modelo misto geral.....	12
3 METODOLOGIA.....	24
3.1 Aspectos do modelo misto.....	24
3.1.1 O modelo misto geral	24
3.1.2 Modelo Misto com classificação dupla.....	27
3.2 Estimação	29
3.2.1 Estimação da parte fixa do modelo misto.....	29
3.2.1.1 Estimação da parte fixa do modelo misto com classificação dupla..	33
3.2.1.2 Hipóteses e somas de quadrados associadas.....	36
3.2.2 Estimação dos componentes de variância.....	39
3.2.2.1 O método ANOVA	40
3.2.2.2 Os métodos de Henderson	41
3.2.2.2.1 Método 1	41
3.2.2.2.2 Método 2	44
3.2.2.2.3 Método 3.....	44
3.2.2.3 Estimadores de máxima verossimilhança.....	49
3.2.2.3.1 Estimadores de máxima verossimilhança - ML.....	50
3.2.2.3.2 Estimadores de máxima verossimilhança restrita - REML	52
3.2.2.4 Estimadores MINQUE e MIVQUE.....	53

3.2.2.4.1	Estimadores quadráticos não-viesados de norma mínima - MINQUE.....	53
3.2.2.4.2	MINQUE Iterativo: I - MINQUE.....	54
3.2.2.4.3	MINQUE(0)	54
3.2.2.4.4	Estimadores imparciais de variância quadrática mínima -MIVQUE.	55
3.2.3	Predição de variáveis aleatórias.....	55
4	SISTEMAS COMPUTACIONAIS ESTATÍSTICOS.....	59
4.1	Introdução.....	59
4.2	SAS - Statistical Analysis System (1992).....	60
4.2.1	Proc GLM - General Linear Models.....	60
4.2.2	Proc VARCOMP - Variance Components.....	62
4.2.3	Proc MIXED - Mixed Models	62
4.3	BMDP - Biomedical Computer Programs (1985).....	64
4.4	HARVEY - Mixed Model Least-Squares and Maximum Likelihood Computer Program (1990).....	65
4.5	MINITAB - Minitab Statistical Software (1986)	66
4.6	NTIA - Núcleo Tecnológico de Informática na Agricultura (1995).....	67
4.7	REML - Residual Maximum Likelihood (1989, versão 96K).....	68
4.8	SANEST - Sistema de Análise Estatística para Microcomputadores (1986, versão 2.1).....	68
4.9	SPSS - Statistical Package for Social Science (1994, versão 6.0).....	69
4.10	STATGRAPHICS - PLUS (1992, versão 6).....	69
4.11	STATISTICA (1994).....	70
5	EXEMPLO ILUSTRATIVO - Ajuste do modelo misto de classificação dupla.....	71
5.1	Ajuste do modelo misto de classificação dupla, através do SAS...	73
5.1.1	Procedimento GLM.....	73
5.1.1.1	Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia. (Dados da Tabela 8).....	77

5.1.1.1.1	Funções estimáveis e hipóteses testáveis.....	77
5.1.1.1.2	Hipóteses e somas de quadrados associadas.....	83
5.1.1.1.3	Testes de hipóteses para a análise de variância do modelo misto.	85
5.1.1.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia. (Dados da Tabela 9).....	92
5.1.1.2.1	Funções estimáveis e hipóteses testáveis.....	92
5.1.1.2.2	Hipóteses e somas de quadrados associadas.....	96
5.1.1.2.3	Testes de hipóteses para a análise de variância do modelo misto.	97
5.1.2	Procedimento VARCOMP.....	102
5.1.2.1	Estimativas dos componentes de variância, para os dados desbalanceados sem casela vazia. (Dados da Tabela 8).....	102
5.1.2.2	Estimativas dos componentes de variância, para os dados desbalanceado com casela vazia. (Dados da Tabela 9).....	107
5.1.3	Procedimento MIXED.....	111
5.1.3.1	Análise do modelo misto de classificação dupla com desbalan- -ceamento e sem casela vazias. (Dados da Tabela 8).....	114
5.1.3.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia. (Dados da Tabela 9).....	119
5.2	Ajuste do modelo linear misto de classificação dupla, através de outros sistemas estatísticos.....	123
5.2.1	Ajuste do modelo linear misto de classificação dupla, através do BMDP-3V	123
5.2.1.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	124
5.2.1.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	126
5.2.2	Ajuste do modelo linear misto de classificação dupla, através do HARVEY - LSMLMW.....	129

5.2.2.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	129
5.2.2.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	132
5.2.3	Ajuste do modelo linear misto de classificação dupla, através do MINITAB - GLM.....	133
5.2.3.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	133
5.2.3.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	134
5.2.4	Ajuste do modelo linear misto de classificação dupla, através do NTIA.-MODLIN.....	135
5.2.4.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	135
5.2.4.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	136
5.2.5	Ajuste do modelo linear misto de classificação dupla, através do REML.....	138
5.2.5.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	139
5.2.5.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	141
5.2.6	Ajuste do modelo linear misto de classificação dupla, através do SANEST.....	142
5.2.6.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	142
5.2.6.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	143

5.2.7	Ajuste do modelo linear misto de classificação dupla, através do SPSS.....	145
5.2.7.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	145
5.2.7.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	146
5.2.8	Ajuste do modelo linear misto de classificação dupla, através do STATGRAPHICS-PLUS.....	147
5.2.8.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	147
5.2.9	Ajuste do modelo linear misto de classificação dupla, através do STATISTICA.....	148
5.2.9.1	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e sem casela vazia.....	148
5.2.9.2	Análise do modelo misto de classificação dupla, com desbalan- -ceamento e casela vazia.....	149
6	UMA COMPARAÇÃO DOS SISTEMAS COMPUTACIONAIS ESTATÍSTICOS.....	150
6.1	Entrada de dados.....	151
6.2	O modelo estatístico.....	152
6.3	Testes de hipóteses.....	153
6.4	Somas de quadrados e hipóteses associadas.....	155
6.5	Estimativas dos componentes de variância.....	157
7	CONCLUSÕES.....	161
7.1	Sobre as hipóteses.....	161
7.2	Sobre as estimativas dos componentes de variância.....	162
7.3	Performance dos sistemas	163
	REFERÊNCIAS BIBLIOGRÁFICAS.....	164

LISTA DE TABELAS

		Página
1	Esperanças dos Quadrados médios para o modelo misto, com dados balanceados segundo Hocking (1973).....	7
2	Quadrados médios e as esperanças para o modelo misto de classificação dupla segundo Hocking (1985).....	9
3	Duas formulações do modelo misto de classificação dupla e balanceamento com A fixo e B aleatório segundo Schwarz (1993).....	11
4	Estatísticas para “testes de hipóteses” no modelo aleatório de classificação simples sob a suposição de normalidade segundo Searle (1987).....	22
5	Análise de Variância e esperanças dos quadrados médios.....	35
6	Partição da soma de quadrados total devido ao ajuste do modelo $y = Wb + e$	42
7	Somas de quadrados e esperanças dos quadrados médios para o modelo $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$	48
8	Experimento com dois fatores e dados desbalanceados com todas as caselas ocupadas.....	72
9	Experimento com dois fatores e dados desbalanceados com casela vazia	72
10	Configurações para o modelo-S com todas as caselas ocupadas.....	74
11	Configurações para o modelo-M com todas as caselas ocupadas.....	74
12	Configurações para o modelo-S com casela vazia.....	75
13	Configurações para o modelo-M com casela vazia.....	75
14	Somas de quadrados segundo o Proc GLM/SAS.....	76
15	Forma geral das funções estimáveis obtidas do procedimento GLM/SAS para o caso de todas as casela ocupadas.....	78
15.1	Forma geral das funções estimáveis obtidas do procedimento	

GLM/SAS para os dados sem casela vazia.....	78
16 Funções estimáveis do Tipo I segundo o Proc GLM/SAS para dados sem casela vazia.....	79
17 Funções estimáveis do Tipo II segundo o Proc GLM/SAS para dados sem casela vazia.....	81
18 Funções estimáveis do Tipo III segundo o Proc GLM/SAS para dados sem casela vazia.....	82
19 Funções estimáveis do Tipo IV segundo o Proc GLM/SAS para dados sem casela vazia.....	83
20 Análise de Variância dos dados sem casela vazia segundo o Proc GLM do SAS através do comando Model	84
21 Testes das hipóteses do Tipo I na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test	86
22 Testes das hipóteses do Tipo II na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test	89
23 Testes das hipóteses do Tipo III na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test	90
24 Testes das hipóteses do Tipo IV na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test	91
25 Forma Geral das funções estimáveis segundo o Proc GLM/SAS para dados com casela vazia.....	92
26 Funções estimáveis do Tipo I segundo o proc GLM/SAS com casela vazia.....	93
27 Funções estimáveis do Tipo II segundo o proc GLM/SAS com casela vazia.....	94
28 Funções estimáveis do Tipo III segundo o proc GLM/SAS com casela vazia.....	95

29	Funções estimáveis do Tipo IV segundo o proc GLM/SAS com casela vazia.....	96
30	Análise de Variância com casela vazia segundo o comando MODEL/GLM/SAS.....	97
31	Testes das hipóteses do Tipo I na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS.....	98
32	Testes das hipóteses do Tipo II na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS.....	99
33	Testes das hipóteses do Tipo III na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS.....	100
34	Testes das hipóteses do Tipo IV na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS.....	101
35	Estimativas do Tipo I para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS.....	103
36	Estimativas MIVQUE(0) para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS.....	104
37	Estimativas ML para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS.....	105
38	Estimativas REML para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS.....	106
39	Estimativas do Tipo I para os componentes de variância com casela vazia segundo Proc VARCOMP do SAS.....	107
40	Estimativas MIVQUE(0) para os componentes de variância com casela vazia segundo Proc VARCOMP do SAS.....	108
41	Estimativas ML para os componentes de variância com casela vazia segundo Proc VARCOMP do SAS.....	109
42	Estimativas REML para os componentes de variância com casela vazia segundo Proc VARCOMP do SAS.....	110

43	Configurações para o modelo-S com todas as caselas ocupadas.....	112
44	Configurações para o modelo-M com todas as caselas ocupadas.....	112
45	Configurações para o modelo-S com casela vazia.....	113
46	Configurações para o modelo-M com casela vazia.....	113
47	Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED e segundo o método REML	115
48	Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED e segundo o método ML	118
49	Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED e segundo o método MIVQUE(0).	119
50	Ajuste do modelo misto de classificação dupla com casela vazia através do Proc MIXED e segundo o método REML.....	120
51	Ajuste do modelo misto de classificação dupla com casela vazia através do Proc MIXED e segundo o método ML.....	121
52	Ajuste do modelo misto de classificação dupla, com casela vazia através do Proc MIXED e segundo o método MIVQUE(0).....	122
53	Ajuste do modelo misto sem casela vazia através do sistema BMDP-3V e segundo o método ML.....	125
54	Ajuste do modelo misto sem casela vazia através do sistema BMDP-3V e segundo o método REML.....	126
55	Ajuste do modelo misto com casela vazia através do sistema BMDP-3V e segundo o método ML.....	127
56	Ajuste do modelo misto com casela vazia através do sistema BMDP-3V e segundo o método REML.....	128
57	Ajuste do modelo misto sem casela vazia através do sistema HARVEY- -modelo 1.....	130
58	Ajuste do modelo misto sem casela vazia através do sistema HARVEY- -modelo 6.....	131

59	Ajuste do modelo misto com casela vazia através do sistema HARVEY- -modelo 1	132
60	Ajuste do modelo misto sem casela vazia através do sistema MINITAB- -GLM.....	133
61	Ajuste do modelo misto com casela vazia através do sistema MINITAB- -GLM.....	134
62	Ajuste do modelo misto sem casela vazia através do sistema NTIA- -MODLIN.....	135
63	Ajuste do modelo misto com casela vazia através do sistema NTIA- -MODLIN.....	136
64	Ajuste do modelo misto sem casela vazia através do sistema REML.....	140
65	Ajuste do modelo misto com casela vazia através do sistema REML.....	141
66	Ajuste do modelo misto sem casela vazia através do sistema SANEST- -modulo 3.....	142
67	Ajuste do modelo misto com casela vazia através do sistema SANEST- -modulo 3.....	143
68	Ajuste do modelo misto sem casela vazia através do sistema SPSS- -ANOVA.....	145
69	Ajuste do modelo misto com casela vazia através do sistema SPSS- -ANOVA.....	146
70	Ajuste do modelo misto sem casela vazia através do sistema STATGRAPHICS-PLUSANOVA.....	147
71	Ajuste do modelo misto sem casela vazia através do sistema STATISTICA-ANOVA.....	148
72	Telas fornecidas no ajuste do modelo misto com casela vazia através do sistema STATISTICA-ANOVA.....	149
73	Performance dos sistemas computacionais utilizados segundo a entrada de dados.....	151

74	Performance dos sistemas computacionais utilizados segundo a especificação dos termos no modelo.....	153
75	Performance dos sistemas computacionais utilizados segundo a determinação das E(QM) e testes F exatos ou aproximados.....	155
76	Performance dos sistemas computacionais para os dados com e sem casela vazia segundo as somas de quadrados e hipóteses associadas ao fator A fixo.....	157
77	Performance dos sistemas computacionais utilizados segundo as estimativas dos componentes de variância.....	158
78	Estimativas do Tipo I para os componentes de variância segundo os sistemas computacionais.....	159
79	Estimativas do Tipo II para os componentes de variância segundo os sistemas computacionais.....	159
80	Estimativas do Tipo III para os componentes de variância segundo os sistemas computacionais.....	159
81	Estimativas ML e REML para os componentes de variância para os dados sem casela vazia segundo os sistemas computacionais.....	160
82	Estimativas ML e REML para os componentes de variância para os dados com casela vazia segundo os sistemas computacionais.....	160

MODELOS LINEARES DE EFEITOS MISTOS: FORMULAÇÃO GERAL E UTILIZAÇÃO DE ALGUNS SISTEMAS COMPUTACIONAIS ESTATÍSTICOS

Autora: MARIA IMACULADA DE LIMA MONTEBELO

Orientador: PROF. DR. ANTONIO FRANCISCO IEMMA

RESUMO

Este estudo visa apresentar uma formulação atualizada dos modelos lineares de efeitos mistos e descrever as idéias básicas para a interpretação das hipóteses mais comumente testadas pelos pesquisadores das ciências aplicadas. Historicamente, o modelo misto tem sido analisado através de procedimentos computacionais apropriados para modelos de efeitos fixos modificados para a obtenção de estatísticas relevantes, através de comandos opcionais para especificar os termos dos erros. Tal estratégia, no entanto, pode não ser trivial, acarretando problemas de adequação e mesmo de interpretação das inferências realizadas.

Nesse contexto discutiu-se a utilização de sistemas computacionais estatísticos como o BMDP (1985), HARVEY (1990), MINITAB (1986), NTIA (1995), REML (1989), SANEST (1986), SAS (1992), SPSS (1994), STATGRAPHICS - PLUS (1992) e o STATISTICA (1994), na análise do modelo linear misto desbalanceado.

Para esses propósitos considera-se o modelo misto clássico com dois fatores, supondo-se o fator A fixo e o fator B aleatório, caracterizado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \text{ onde:}$$

$\mu + \alpha_i$ é a parte dos efeitos fixos do modelo, e

$\beta_j + (\alpha\beta)_{ij} + e_{ijk}$ é a parte dos efeitos aleatórios do modelo, com distribuições:

$\beta_j \overset{iid}{\cap} N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \overset{iid}{\cap} N(0, \sigma_{\alpha\beta}^2)$, $e_{ijk} \overset{iid}{\cap} N(0, \sigma_e^2)$, respectivamente e

β_j , $(\alpha\beta)_{ij}$ e e_{ijk} , são variáveis aleatórias independentes.

Para cada procedimento dos sistemas estatísticos mencionados, adotou-se um conjunto de sintaxes básicas para o ajuste do modelo misto de classificação dupla cruzada com interação.

Com base nas saídas emitidas discutiu-se a performance desses sistemas quanto às suas características metodológicas disponíveis, tais como: a entrada de dados, a flexibilidade quanto a especificação dos termos do modelo, a apresentação das expressões dos valores esperados dos quadrados médios com testes F exatos e aproximados. Visamos elucidar aos usuários das ciências aplicadas, estabeleceu-se um estudo descritivo sobre a análise da parte fixa do modelo, especificamente sobre as hipóteses que estão sendo testadas na análise de variância.

Foram considerados dois conjuntos de dados citados em lemma (1995a), para ilustrar esse desenvolvimento.

Os resultados observados no estudo comparativo indicaram que o sistema computacional SAS apresenta grande flexibilidade para o ajuste do modelo de efeitos mistos, através dos procedimentos GLM, VARCOMP e MIXED e fornece a análise de variância clássica mais completa que os demais, destacando-se a excelente performance do Proc MIXED. Por outro lado os sistemas computacionais MINITAB, NTIA, SANEST, SPSS, STATGRAPHICS e STATISTICA apresentam saídas limitadas para a análise do modelo misto e, não raro, requerem que o usuário complete a análise utilizando outros sistemas computacionais.

LINEAR MODELS OF MIXED EFFECTS: GENERAL WORDING AND THE USE OF SOME STATISTICAL COMPUTERIZED SYSTEMS

Author: MARIA IMACULADA DE LIMA MONTEBELO

Adviser: PROF. DR. ANTONIO FRANCISCO IEMMA

SUMMARY

This work aims to present an actual display of mixed effects linear models and to describe the basic ideas for the interpretation of the most common hypothesis tested by the applied sciences researchers. Historically, the mixed effects model has been analyzed through proper computerized procedures for fixed effect models, modified in order to get relevant statistics through optional commands to specify the error terms. Such strategy though, may not be so simple, bringing adequation problems, even interpretation of realized inferences.

In this context, the used of statistical computerized systems such as the BMDP (1985), HARVEY (1990), MINITAB (1986), NTIA (1995), REML (1989), SANEST (1986), SAS (1992), SPSS (1994), STATGRAPHICS-PLUS (1992) and the STATISTICA (1994), in discussed the mixed not balanced linear model analysis.

With this aim, a classic mixed model having two factors is considered supposing the fixed A and the random B, characterized by :

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} , \text{ where:}$$

$\mu + \alpha_i$ is the part of fixed effects of the model, and

$\beta_j + (\alpha\beta)_{ij} + e_{ijk}$ is the part random effects model, with

$\beta_j \overset{iid}{\sim} N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \overset{iid}{\sim} N(0, \sigma_{\alpha\beta}^2)$, $e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2)$, distributions respectively; and

β_j , $(\alpha\beta)_{ij}$ e e_{ijk} , are independent random variables.

For each mentioned statistical systems procedure, an asset of basic syntaxes was adopted for mixed model of adjustment of the two crossed with an interaction classification.

Based on the emitted outlets, the performance of these systems is discussed considering its available methodological characteristics such as data input, the flexibility considering the wording specification of the models, the expression of the expected value presentations of the average squares with exact and approximate F tests. Aiming to render clear to the applied sciences users, a descriptive study is set on the fixed part analysis, specifically on the variance analysis hypothesis tested.

Two data sets mentioned by lemma (1995a), should be considered to illustrate this development.

The results of reserved in the comparative study show that the SAS computational system presents a great flexibility for the adjusting model of mixed effects, using the GLM, VARCOMP and MIXED procedures, which provides a classic variance analysis more complete than the others, the excellent performance of the Proc MIXED is outstanding. On the other hand the MINITAB, NTIA, SANEST, SPSS, STATGRAPHICS and STATISTICA present limited outlets for the mixed model analysis and often require from the user to complete the analysis, the use of other computerized systems.

1 INTRODUÇÃO

Os modelos mistos são usados para descrever dados de experimentos cuja estrutura dos tratamentos envolve alguns fatores que são fixos e outros que são aleatórios.

A análise do modelo misto consiste, portanto, em dois tipos de análise, uma sobre a parte fixa e a outra sobre a parte aleatória do modelo. A literatura apresenta considerável discussão sobre o modelo, a qual se caracteriza pelo desenvolvimento de metodologias para a estimação por pontos e estimação pontual dos componentes de variância.

Em contraste com a análise de variância de modelos completamente aleatórios ou completamente fixos, a análise de modelos mistos pode apresentar dificuldades. As regras básicas para a análise de dados balanceados, estão, em geral, disponíveis na literatura, e há, de certo modo, um consenso por parte dos estatísticos. Este fato, no entanto, não ocorre com a análise de dados desbalanceados, pois os resultados podem ser de interpretação bastante complexa. Além disso, como pode ser visto em Schwarz (1993) alguns sistemas estatísticos apresentam resultados que podem gerar falácias, para os usuários não iniciados no trato dos modelos mistos.

Nesse contexto, o presente trabalho pretende abordar a análise da parte fixa do modelo misto, com a intenção de fornecer idéias básicas para a interpretação das hipóteses mais comumente testadas pelos pesquisadores das ciências aplicadas. Para esse objetivo segue-se a metodologia apresentada em lemma (1991, 1995-a) e Santos (1994).

Historicamente, o modelo misto tem sido analisado através de procedimentos computacionais apropriados para modelos de efeitos fixos modificados para a obtenção de estatísticas relevantes, através de comandos opcionais para especificar os termos dos erros. Tal estratégia, no entanto, pode não ser trivial, acarretando problemas de adequação e mesmo de interpretação das inferências realizadas. Visando-se amenizar o trabalho usuário não iniciado, apresenta-se um estudo comparativo de alguns sistemas computacionais estatísticos, quando utilizados para analisar o modelo misto de classificação cruzada com interação com dados desbalanceados.

Nesse contexto avalia-se as saídas dos sistemas computacionais estatísticos: BMDP (Biomedical Computer Programs, 1983), HARVEY (Mixed Model Least - Squares and Maximum Likelihood Computer Program, 1990), MINITAB (Minitab Statistical Software, 1986), NTIA (Núcleo Tecnológico, 1995), REML (Residual Maximum Likelihood, 1989), SANEST (Sistema de Análise Estatística para Microcomputadores, 1986), SAS (Statistical Analysis System, 1992), SPSS (Statistical Package for the Social Sciences, 1994), STATGRAPHICS - PLUS (1992) e o STATISTICA (1994).

Com isso, esse estudo tem por objetivos gerais:

- I) Fazer uma abordagem das principais hipóteses testadas através dos sistemas estatísticos citados, adotando-se o modelo linear misto de classificação dupla cruzada com interação.
- II) Estabelecer um estudo comparativo da performance dos sistemas estatísticos, para a análise do modelo misto, quanto às características metodológicas disponíveis, tais como: a entrada de dados, a flexibilidade quanto a especificação do modelo, a apresentação das expressões das esperanças matemáticas dos quadrados médios, e a apresentação das somas de quadrados e hipóteses associadas.

2 REVISÃO DE LITERATURA

2.1 Introdução ao estudo do modelo misto

A primeira ocorrência na literatura da descrição do modelo misto na forma de modelo linear, segundo citação feita em Searle et al. (1992) é atribuída a Jackson (1939), para estudar testes mentais. Nesse estudo, o autor descreve as observações através do modelo misto de classificação dupla sem interação, dado por:

$$y_{st} = A + B_s + C_t + Z_{st}, \text{ onde :}$$

A é uma medida de efeito comum a todos os indivíduos.

B_s é a medida do efeito de uma “prova”;

C_t é a medida do efeito de um indivíduo;

Z_{st} é o erro de mensuração.

y_{st} é o resultado do teste obtido pelo t -ésimo indivíduo, na s -ésima prova.

Os efeitos das “provas” são tratados como fixos e os efeitos dos “indivíduos” são considerados aleatórios.

Em livros textos o modelo misto foi introduzido por Mood (1950) quando o autor descreve os modelos alternativos contendo apenas efeitos fixos ou apenas efeitos aleatórios, ou contendo ambos, para análise de variância. Posteriormente, tais modelos foram desenvolvidos e discutidos em

detalhes por Crump (1951); Anderson & Bancroft (1952); Wilk & Kempthorne (1955); Scheffé (1956b) e Plackett (1960).

Com o objetivo de descrever uma grande variedade de situações experimentais nas quais um fator é assumido fixo e outro aleatório, Scheffé (1959) apresenta detalhadamente o modelo misto, onde a variável y_{ijk} , para todo i, j e k é variável aleatória com distribuição normal, descrita através do modelo:

$$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}, \text{ onde:}$$

y_{ijk} é a k -ésima observação referente a i -ésima linha com a j -ésima coluna

μ é a média geral;

α_i é o efeito fixo da i -ésima linha, tal que $\sum_{i=1}^I \alpha_i = 0$;

b_j é o efeito aleatório da j -ésima coluna;

c_{ij} é o efeito aleatório da interação da i -ésima linha com a j -ésima coluna, tal

que, $\sum_{i=1}^I c_{ij} = 0$, para todo j ;

e_{ijk} é o erro aleatório, com distribuição normal de média zero e variância σ_e^2 .

Scheffé(1959) propõe ainda uma estrutura de variâncias e covariâncias bem geral, definida em termos de uma matriz simétrica $I \times I$ com elementos $\sigma_{ii'}$, de modo que $\sigma_{ii'} = \sigma_{i'i}$, $\sigma_{i.} = \sigma_{.i}$ e $\sigma_b^2 = \sigma_{..}$, como a seguir:

$$\begin{aligned} \text{Cov}(b_j, b_{j'}) &= \delta_{jj'} \sigma_{..} \\ \text{Cov}(c_{ij}, c_{i'j'}) &= \delta_{jj'} (\sigma_{ii'} - \sigma_{i.} - \sigma_{i'.} + \sigma_{..}) \\ \text{Cov}(b_j, c_{ij'}) &= \delta_{jj'} (\sigma_{i.} - \sigma_{..}) \end{aligned}$$

desde que $E(c_{ij}c_{i'j'}) = \delta_{ij'} E(c_i c_{i'})$, $\delta_{ij'}$ é o parâmetro de não centralidade, e ainda, todas as demais covariâncias iguais a zero. Apresenta, também as seguintes expressões das esperanças matemáticas para os quadrados médios.

$$E(\text{QMA}) = \sigma_e^2 + K\sigma_{AB}^2 + JK\sigma_A^2,$$

$$\text{onde: } \sigma_A^2 = (I-1)^{-1} \sum_{i=1}^I \alpha_i^2 \text{ e } \sigma_{AB}^2 = (I-1)^{-1} \sum_{i=1}^I \text{Var}(c_{ij});$$

$$E(\text{QMB}) = \sigma_e^2 + IK\sigma_B^2, \text{ onde: } \sigma_B^2 = \text{Var}(b_j);$$

$$E(\text{QM AB}) = \sigma_e^2 + \left(\frac{K}{K} \right) \sigma_{AB}^2;$$

$$E(\text{QM ERRO}) = \sigma_e^2.$$

Argumenta ainda, o autor, que sob a suposição de normalidade, embora QMA e QMAB sejam independentes estatisticamente, sob a hipótese $H_0: \alpha_i = 0, \forall i$, o quociente $\frac{\text{QMA}}{\text{QMAB}}$ não tem em geral a distribuição exata de F, e então propõe o uso de um teste exato baseado na estatística T^2 de Hottelling.

Mood (1950), Searle (1971), entre outros, adotam para o modelo misto com dois fatores cruzados e interação, a seguinte formulação:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}, \text{ com: } i = 1, \dots, a, j = 1, \dots, b \text{ e } k = 1, \dots, n,$$

onde:

y_{ijk} é a k-ésima observação referente a i-ésima linha com a j-ésima coluna;

μ e τ_i , são os efeitos fixos do modelo;

β_j , $(\tau\beta)_{ij}$ e e_{ijk} , são variáveis aleatórias não correlacionadas de média zero

com:

$$\text{Var}(\beta_j) = \sigma_\beta^2$$

$$\text{Var}((\tau\beta)_{ij}) = \sigma_{\tau\beta}^2, \text{ onde } \sigma_\beta^2 \text{ e } \sigma_{\tau\beta}^2 \text{ são os componentes de variância.}$$

$$\text{Var}(e_{ijk}) = \sigma_e^2$$

Graybill (1961), entre outros, descreve o modelo misto,

$$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + e_{ijk}, \text{ com: } i = 1, \dots, a, j = 1, \dots, b \text{ e } k = 1, \dots, n,$$

onde:

y_{ijk} é a k -ésima observação referente a i -ésima linha com j -ésima coluna;

μ e α_i , para $i = 1, \dots, a$ são parâmetros fixos,

γ_j , $(\alpha\gamma)_{ij}$ e e_{ijk} , são variáveis aleatórias não correlacionadas de média zero e

com a seguinte estrutura de covariâncias,

$$\text{Var}(\gamma_j) = \sigma_\gamma^2$$

$$\text{Var}[(\alpha\gamma)_{ij}] = \frac{a-1}{a} \sigma_{\alpha\gamma}^2$$

$$\text{Cov}[(\alpha\gamma)_{ij}, (\alpha\gamma)_{i'j}] = -\frac{1}{a} \sigma_{\alpha\gamma}^2, \quad i \neq i'$$

$$\text{Var}(e_{ijk}) = \sigma_e^2.$$

Admite que $\sum_{i=1}^a (\alpha\gamma)_{ij} = 0$, para qualquer j e que as demais

covariâncias são nulas. Assim os componentes de variâncias são σ_γ^2 e $\sigma_{\alpha\gamma}^2$.

Parte da discussão do modelo misto de classificação dupla é centrada nas diferentes expressões do valor esperado do quadrado médio associado ao fator aleatório, quando a estrutura de dados é balanceada. Para o caso em que os dados y_{ijk} , com $i = 1, \dots, a$, $j = 1, \dots, b$ e $k = 1, \dots, n$, onde n é o número de observações por caselas e a o número de níveis do fator em linha pode-se encontrar a expressão $\sigma_e^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2$, em Mood (1950); Stell & Torrie (1960), entre outros.

A expressão $\sigma_e^2 + an\sigma_\beta^2$, para o componente de variância para o fator aleatório, onde a e n são como definidos anteriormente é encontrada em

discussões sobre o fato de incluir ou não o termo $\sigma_{\tau\beta}^2$ na expressão são encontradas em textos como Wilk & Kempthorne (1955); Hartley & Searle (1969).

Hartley (1967) desenvolve um procedimento geral para obter os valores numéricos dos coeficientes das expressões dos QM, para dados desbalanceados, denominado método de “synthesis”, que pode ser aplicado em modelos aleatórios e modelos mistos. Rao (1968) apresenta uma extensão do método de “synthesis” para o modelo misto. A vantagem do método é que qualquer programa computacional para analisar modelos de efeitos fixos pode ser adaptado para obter as esperanças dos quadrados médios. No entanto, como essas expressões dependem da natureza do algoritmo computacional utilizado, diferentes programas podem produzir diferentes resultados para os valores esperados.

Hartley & Searle (1969) tratam dos componentes de variância para o caso desbalanceado e argumentam que no estudo de Searle & Henderson (1961) o desenvolvimento das expressões dos valores esperados para o caso desbalanceado apresenta o termo da interação na $E(QM)$, associado ao fator aleatório, considerando-se o caso balanceado como um caso particular. Justificam a inclusão do termo, contrariando a opinião de muitos autores.

Hocking (1973) denomina MODELO I, o modelo geral atribuído a Scheffé (1959), MODELO II o modelo adotado por Mood (1950) e MODELO III o proposto por Graybill (1961). Ademais, descreve cada modelo através dos dois primeiros momentos das observações. Estabelece que os MODELOS II e III, são casos especiais do MODELO I. As aparentes diferenças entre esses modelos são traçadas com o objetivo de esclarecer o pesquisador na busca de justificativas para a escolha do modelo. Por sua vez, Hocking (1973) apresenta expressões para as $E(QM)$, conforme consta na Tabela 1.

Tabela 1. Esperanças dos Quadrados médios para o modelo misto com dados balanceados segundo Hocking (1973)

CAUSA	MODELO I	MODELO III *
	$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}$	$y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + e_{ijk}$
Fator fixo	$\sigma_e^2 + \frac{bn}{(a-1)} \sum_i \alpha_i^2 + n\sigma_{AB}^2$	$\sigma_e^2 + \frac{bn}{(a-1)} \sum_i \tau_i^2 + n\sigma_{\tau\beta}^2$
Fator aleatório	$\sigma_e^2 + an\sigma_B^2$	$\sigma_e^2 + an\sigma_\beta^2 + n\sigma_{\tau\beta}^2$
Interação	$\sigma_e^2 + n\sigma_{AB}^2$	$\sigma_e^2 + n\sigma_{\tau\beta}^2$
Erro	σ_e^2	σ_e^2

* modelo II (Mood, 1950) equivale a modelo III (Grabill, 1961) admitindo-se as seguintes relações: 1) $\gamma_j = \beta_j + (\tau\bar{\beta})_{.j}$, 2) $(\alpha\gamma)_{.j} = (\tau\beta)_{.j} + (\tau\bar{\beta})_{.j}$.

Hocking (1985) apresenta um resumo detalhado sobre a escolha entre os modelos I e III, sob o título "resolução de uma controvérsia no modelo misto de classificação dupla". Adota o modelo misto com dois fatores e interação, descrito por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk},$$

com $i = 1, \dots, a_1$; $j = 1, \dots, a_2$ e $k = 1, \dots, n$ e

$$E(y_{ijk}) = \mu_i = \mu + \alpha_i$$

e admite:

$$\beta_j \overset{iid}{\cap} N(0, \phi_2),$$

$$(\alpha\beta)_{ij} \overset{iid}{\cap} N(0, \phi_2) \text{ e}$$

$$e_{ijk} \overset{iid}{\cap} N(0, \phi_0).$$

Nesse contexto, propõe a análise apresentada na Tabela 2.

Tabela 2. Quadrados médios e as esperanças para o modelo misto de classificação dupla segundo Hocking (1985)

DESCRIÇÃO	QM	E(QM)
Linha	q_1/r_1	$\phi_0 + n\phi_{12} + na_2 \left(\sum_{i=1}^{a_1} \alpha_i^2 \right) / r_1$
Coluna	q_2/r_2	$\phi_0 + n\phi_{12} + na_1\phi_2$
Interação	q_{12}/r_{12}	$\phi_0 + n\phi_{12}$
Resíduo	q_0/r_0	ϕ_0

Com base na Tabela 2 onde, q_1 , q_2 , q_{12} e q_0 formam um conjunto de estatísticas suficientes, nesse caso somas de quadrados com r_1 , r_2 , r_{12} e r_0 graus de liberdade respectivamente, o autor comenta que um teste estatístico sugerido para a hipótese sobre as linhas, $H_0: \alpha_i = \mu_i - \bar{\mu} = 0$, $i = 1, \dots, a_i - 1$, para $i = 1, 2$ é:

$$F = (q_1 / r_1) / (q_{12} / r_{12}).$$

Salienta ainda que, com dados desbalanceados, a obtenção dos valores esperados para os quadrados médios não é um processo trivial.

Em trabalho recente, Schwarz (1993) discute o modelo misto:

$$y_{ijk} = \mu_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \text{ com, } i = 1, \dots, a; j = 1, \dots, b \text{ e } k = 1, \dots, n,$$

onde:

μ_i é a média marginal do nível i do fator A, a qual é frequentemente parametrizada como $\mu_i = \mu + \alpha_i$;

β_j : é o efeito do nível j do fator B;

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

ε_{ijk} : é o erro aleatório;

y_{ijk} : é o valor observado referente ao nível i do fator A e ao nível j do fator B, estabelece relações entre três parametrizações mais citadas na literatura, argumenta que duas são variantes de uma parametrização comum e conseqüentemente considera duas, e as apresenta como formulação 1 e formulação 2, conforme mostra a Tabela 3.

Nesse artigo, o autor avalia o tratamento que diferentes autores desenvolveram em livros textos e livros de referências, quanto a análise do modelo misto geral. Considera diversos aspectos da análise, tais como: formulação adotada, estimação dos efeitos fixos e aleatórios, e tipos de delineamentos utilizados. O autor salienta, ainda, que os livros textos abordam a segunda formulação, mas a discussão sobre a escolha é arbitrária e diz que alguns livros de referências tais como Searle (1971,1987) e Hocking (1985) apresentam uma discussão mais extensa.

O autor faz, também, um estudo de comparação dos sistemas computacionais SAS (1989), BMDP(1990) e SPSS (1990), quando utilizados para analisar o modelo misto, abordando os aspectos da formulação utilizada, e verifica que a formulação 1 é utilizada pelos sistemas, BMDP-3V e SAS (procedimentos GLM e VARCOMP). A formulação 2 é adotada pelo sistema BMDP-8V. O sistema SPSS apresenta maior flexibilidade, permite que a escolha da formulação seja feita pelo analista.

Tabela 3. Duas formulações do modelo misto de classificação dupla e balanceamento com A fixo e B aleatório segundo Schwarz (1993)

	formulação 1	formulação 2
SUPOSIÇÕES SOBRE OS PARÂMETROS		
	$\beta_j \overset{iid}{\cap} N(0, \sigma_B^2)$	$\beta_j \overset{iid}{\cap} N(0, \xi_B^2)$
	$\gamma_{ij} \overset{iid}{\cap} N(0, \sigma_{AB}^2)$	$\gamma_{ij} \overset{iid}{\cap} N(0, \frac{a-1}{a} \xi_{AB}^2)$
	$\varepsilon_{ijk} \overset{iid}{\cap} N(0, \sigma_e^2)$	$\varepsilon_{ijk} \overset{iid}{\cap} N(0, \xi^2)$
		$\sum_{i=1}^a \gamma_{ij} = 0, \text{ para todo } j$
		$Cov(\gamma_{ij}, \gamma_{i'j}) = -\frac{1}{a} \xi_{AB}^2, i \neq i'$
		β_j, γ_{ij} são independentes de ε_{ijk}
ESTRUTURAS DE COVARIÂNCIAS DAS OBSERVAÇÕES y_{ijk}		
$i = i', j = j', k = k'$	$\sigma_e^2 + \sigma_B^2 + \sigma_{AB}^2$	$\xi^2 + \xi_B^2 + \frac{a-1}{a} \xi_{AB}^2$
$i = i', j = j', k \neq k'$	$\sigma_B^2 + \sigma_{AB}^2$	$\xi_B^2 + \frac{a-1}{a} \xi_{AB}^2$
$i \neq i', j = j'$	σ_B^2	$\xi_B^2 - \frac{1}{a} \xi_{AB}^2$
ESPERANÇAS DOS QUADRADOS MÉDIOS		
A	$\sigma_e^2 + n\sigma_{AB}^2 + nbQ(\mu_1, \dots, \mu_a)$	$\xi^2 + n\xi_{AB}^2 + nbQ(\mu_1, \dots, \mu_a)$
B	$\sigma_e^2 + n\sigma_{AB}^2 + na\sigma_B^2$	$\xi^2 + na\xi_B^2$
A*B	$\sigma_e^2 + n\sigma_{AB}^2$	$\xi^2 + n\xi_{AB}^2$
ERRO	σ_e^2	ξ^2

2.2 Estimação dos componentes de variância no modelo misto geral

Um procedimento considerado pioneiro para a obtenção das estimativas dos componentes de variância, consiste em igualar as somas de quadrados obtidas na análise de variância com seus respectivos valores esperados.

Tradicionalmente conhecido como método ANOVA, tem dentre as primeiras publicações o artigo de Crump (1946) que, nesse estudo apresenta a metodologia com base no modelo aleatório de classificação dupla cruzada e dados balanceados. Ilustra o método através de um exemplo numérico e comenta que para classificações com números desiguais por caselas, há outros procedimentos como os propostos em Yates (1934), que podem ser mais adequados: os métodos conhecidos na literatura como o método de quadrado de médias ponderadas e o método do ajustamento de constantes.

Searle et al. (1992) citam que no período de 1950 a 1969 ocorreu o maior desenvolvimento dos métodos para estimativas de componentes de variâncias, tanto em termos de aperfeiçoamento dos disponíveis, quanto em termos de propostas de novos procedimentos baseados na máxima verossimilhança e no critério de norma mínima.

Um dos primeiros trabalhos clássicos na estimação dos componentes de variância é devido a Henderson (1953). Nele, o autor apresenta o desenvolvimento de três métodos de estimação, para dados desbalanceados do modelo aleatório ou misto, de classificação cruzada ou hierárquica, conhecidos como métodos 1, 2 e 3 de Henderson, respectivamente. Esses procedimentos estabelecem três conjuntos de formas quadráticas, as quais são utilizadas para a obtenção das estimativas dos componentes de variância.

O autor ilustra o desenvolvimento dessas metodologias com exemplos numéricos, e discute detalhadamente as análises em cada modelo sem explicitar, no entanto, as hipóteses de interesse. Pode-se verificar que o

método 1 é aplicado apenas para o modelo aleatório. O método 2 é uma adaptação do método 1 que pode ser aplicado ao modelo misto sem interação entre os efeitos fixos e aleatórios e, que o método 3 pode ser aplicado para o modelo misto em geral. Este, é usualmente conhecido como o método do ajuste de constantes.

Segundo Scheffé (1959) a falta de unicidade do método ANOVA para dados desbalanceados é devida à existência de mais estatísticas suficientes minimais do que são os componentes de variâncias no modelo, pois nesse caso o método não especifica um único conjunto de formas quadráticas.

Henderson et al. (1959) mostram que para o modelo misto geral,

$$y = X\beta + Zu + e,$$

onde: β é o vetor de efeitos fixos, enquanto u e e são vetores independentes de variáveis que normalmente distribuídas com média zero e matriz de variâncias e covariâncias $D\sigma^2$ e $R\sigma^2$ respectivamente. Então y tem distribuição normal multivariada com média $X\beta$ e matriz de variâncias e covariâncias $\Sigma = (R + ZDZ')\sigma^2$, os efeitos de β podem ser estimados maximizando-se a função de distribuição conjunta das observações e efeitos aleatórios. Provam ainda, nesse trabalho, a igualdade entre o estimador de β obtido pelo método 1 de Henderson através da solução para:

$$X'(R + ZDZ')^{-1}X\hat{\beta} = X'(R + ZDZ')^{-1}y,$$

e o estimador de β , obtido maximizando-se a função de distribuição conjunta:

$$f(y, u) = g(y / u)h(u),$$

ou seja:

$$X'WX\tilde{\beta} = X'Wy,$$

onde:

$$W = R^{-1} - R^{-1}Z(Z'R^{-1}Z + D^{-1})^{-1}Z^{-1}R^{-1}.$$

Searle & Henderson (1961), sugerem um procedimento simplificado utilizando a notação $R(\)$ para obter as estimativas dos componentes de variâncias para o modelo misto de classificação dupla com dados desbalanceados, com o fator fixo contendo um número moderado de níveis. Apresentam as somas de quadrados, seus respectivos valores esperados, mas não discutem explicitamente as hipóteses de interesse.

Thompson (1962) introduz a idéia de maximizar parte da função de verossimilhança, para modelos mistos, ou seja, sugere um procedimento que consiste em fator a função de máxima verossimilhança em duas partes, uma dependendo apenas dos componentes de variância. Esse fator é então maximizado com respeito aos componentes de variância, e o resultado é usado como o estimador dos componentes de variância. A maximização do segundo fator com respeito aos efeitos fixos usando essas estimativas, produzem uma estimativa para os efeitos fixos, os quais são considerados parâmetros de locação do modelo. Atualmente esse procedimento é conhecido como máxima verossimilhança restrita (REML) ou resíduo de máxima verossimilhança.

A eficiência relativa dos estimadores obtidos pelos três métodos de Henderson (1953) e pelos métodos propostos por Yates (1934) é analisada por Bush & Anderson (1963).

Hultquist & Graybill (1965) apresentam alguns teoremas que podem ser usados para a obtenção de estatísticas suficientes minimais para o modelo misto com classificação dupla, com números desiguais de observações por caselas. Para o modelo

$$y = X\tau + Z\beta + e,$$

onde: $X\tau$ representa a parte fixa, $Z\beta + e$ representa a parte aleatória do modelo, os autores demonstram que a dimensão de uma estatística suficiente

onde: $X\tau$ representa a parte fixa, $Z\beta + e$ representa a parte aleatória do modelo, os autores demonstram que a dimensão de uma estatística suficiente minimal é função dos postos de certas matrizes de $Z'X$. As estatísticas suficientes minimais são apresentadas em tabelas e algumas propriedades são discutidas.

Harville (1977) apresenta um trabalho de investigação de certos aspectos da estimabilidade de funções lineares de componentes de variância para classificação dupla de efeitos aleatórios, com número desigual de observações nas subclasses. Utiliza os métodos 1 e 3 de Henderson (1953), para a estimação dos componentes de variância, mas os aspectos relacionados às hipóteses testáveis não são investigados.

Hartley & Rao (1967) desenvolvem um procedimento de máxima verossimilhança para as estimativas das constantes desconhecidas e variâncias incluídas no modelo misto geral, descrito por :

$$y = X\alpha + U_1b_1 + \dots + U_cb_c + e,$$

onde :

X é uma matriz conhecida de ordem $n \times k$, $k \leq n$, $r(X) = k$;

U_i é uma matriz conhecida de ordem, $n \times m_i$, $m_i \leq n$, $r[U_i] = m_i$;

α é um vetor de constantes desconhecidas de ordem $k \times 1$;

b_i é um vetor $m_i \times 1$ de variáveis aleatórias independentes com $\cap N(0, \sigma_i^2)$;

e é um vetor $n \times 1$ de variáveis aleatórias independentes com $\cap N(0, \sigma_e^2 I)$.

Os vetores aleatórios b_1, b_2, \dots, b_c e e são mutuamente independentes.

Os autores aplicam o método em casos onde as matrizes de delineamento X e U_i , têm posto completo e assumem que U_i , tem em cada

consequentemente essa propriedade implica que todas as matrizes $U_i'U_i$ de ordem $m_i \times m_i$ são diagonais. Eles salientam a suposição adicional de que $m = \sum_{i=1}^c m_i$ é importante, e deve ser feita sobre as matrizes de delineamento, e que pode ser descrita como uma condição para a estimabilidade de α e σ_i^2 .

Nesse artigo, os autores definem a matriz $M = \left[\begin{array}{c|c|c} X & U_1 & \dots & U_c \end{array} \right]$ e a partir de M definem uma matriz W de ordem $n \times r$, da forma $W = \left[\begin{array}{c} X \\ U^* \end{array} \right]$, onde a matriz U^* de ordem $n \times (r - k)$, deve ter ao menos uma coluna para cada U_i , tal que, $k + c \leq r \leq k + m$. Apresentam, como uma consequência do modelo descrito, a matriz de variâncias e covariâncias dada por: $\sigma^2 H = \sigma^2 [I_n + \gamma_1 U_1 U_1' + \dots + \gamma_c U_c U_c']$, onde $\gamma_i = \sigma_i^2 / \sigma^2$. A eficiência assintótica das estimativas são discutidas. São obtidos testes de hipóteses e regiões de confiança.

Searle (1968) reformula toda a teoria dos métodos desenvolvidos por Henderson (1953) na forma de matrizes, considera o modelo linear:

$$y = X\beta + e,$$

a partição de β em $\beta' = (\beta_1', \beta_2', \dots, \beta_K')$, onde β_1 representa todos os efeitos fixos do modelo e os demais representam os efeitos aleatórios, com $E(y) = X_1 \beta_1$ e $Var(y) = X_2 Var(\beta_2) X_2' + \dots + X_K Var(\beta_K) X_K'$. Discute vários aspectos do modelo linear à luz dos resultados da redução:

$$R(\beta) = \hat{\beta}' X' y = y' X (X' X)^{-1} X' y,$$

argumenta que são considerações importantes, especialmente as propriedades que relacionam a análise de variância, às hipóteses testadas e aos componentes de variância. O interesse é centrado nos resultados dos valores esperados de formas quadráticas do tipo $y'Qy$, que para o modelo considerado resulta na expressão:

$$E(y'Qy) = (X_1\beta_1)'QX_1\beta_1 + \sum_{i=2}^k \sigma_i^2 \text{tr}(QX_iX_i') + \sigma_i^2 \text{tr}(Q).$$

Esse resultado geral é aplicado para desenvolver matricialmente os métodos 1, 2 e 3 de Henderson. Cada um dos procedimentos é ilustrado com exemplos numéricos. O autor, ainda, faz uma análise crítica e propõe algumas modificações para os métodos apresentando as seguintes conclusões:

i) O método 1 é de simples aplicação, utiliza as somas de quadrados da análise de variância e não é apropriado para modelos mistos.

ii) O método 2, não deve ser usado se existem interações entre os efeitos fixos e os efeitos aleatórios, utiliza as somas de quadrados da análise de variância.

iii) O método 3 é mais adequado para o modelo misto e para modelos envolvendo covariâncias entre o conjunto de efeitos aleatórios, utiliza a redução das somas de quadrados devida ao método do ajuste de constantes quando diferentes sub-modelos do modelo completo são ajustados.

Searle (1970), desenvolve uma expressão geral para os elementos da matriz de informação dos estimadores de máxima verossimilhança, para dados desbalanceados, no modelo linear misto geral. A expressão é utilizada para obter resultados explícitos para o modelo hierarquizado de classificação dupla e para o modelo aleatório de classificação simples.

Searle (1971), apresenta uma discussão dos procedimentos de estimação dos componentes de variâncias, a obtenção de intervalos exatos e aproximados e uma discussão superficial dos testes sobre os efeitos fixos do modelo.

Rao (1971a, 1971b, 1972), desenvolve um procedimento de estimação, baseado no critério de norma mínima, MINQUE, ou variância mínima MIVQUE. O procedimento MINQUE utiliza a restrição de que a norma euclidiana da matriz núcleo da forma quadrática seja mínima, enquanto que no procedimento MIVQUE a matriz núcleo é determinada de modo que os estimadores tenham variância mínima.

A partir da década de 70, os pesquisadores tem dado ênfase aos estudos de comparação dos diferentes métodos disponíveis na literatura, para obtenção das estimativas dos componentes de variância, utilizando dados simulados. Têm ocorrido, também, estudos sobre a performance dos diferentes sistemas computacionais, quanto a metodologia oferecida para a análise do modelo misto geral, principalmente quanto aos procedimentos disponíveis para a obtenção das estimativas dos componentes de variância.

Patterson & Thompson (1971) discutem as propriedades dos estimadores para dados balanceados e para dados desbalanceados. Afirmando que as soluções de REML são idênticas aos estimadores ANOVA, e que os estimadores ML e REML são mais adequados para os casos desbalanceados. Utilizam nesse estudo o modelo misto para o delineamento em blocos.

Hemmerle & Hartley (1973), propõem uma matriz conhecida como a transformação W , para facilitar os cálculos na utilização do método de obtenção dos estimadores de máxima verossimilhança para os efeitos fixos e os componentes de variância, para o modelo misto geral descrito em Hartley & Rao (1967). A implementação da matriz transformação W é ilustrada com exemplos numéricos, utilizando-se o método iterativo de Newton-Raphson.

Francis (1973) avalia a performance de quatro sistemas estatísticos. Proc ANOVA (Analysis of variance) do SAS, BMDX64 (Biomedical Computer Programs), CAROLINA (A Multivariate Analysis of Variance Program) do sistema OSIRIS (University of Michigan), usados para o procedimento da análise de variância. Discute, como ilustração, os resultados obtidos para

dados de classificação dupla com número desigual de observações por caselas.

Hocking & Kutner (1975), apresentam a forma das equações de verossimilhança como um procedimento de cálculo eficiente e estabelecem um critério de comparação com outros estimadores de componentes de variância propostos na literatura. Consideram as estimativas dos componentes de variância do modelo aleatório para o delineamento em blocos incompletos balanceados (BIB), para um conjunto de dados simulados com o objetivo de ilustrar as vantagens relativas dos procedimentos de estimação.

Brown (1976), discute as propriedades assintóticas e determina a distribuição limite dos estimadores MINQUE.

Speed & Hocking (1976) discutem o uso da notação $R(\)$, para o modelo de posto incompleto utilizada por Searle (1971) e estabelecem as diferenças entre dois procedimentos que denominaram "Procedimento 1" e "Procedimento 2". Enfatizam que o primeiro é aplicado aos modelos de posto incompleto e o segundo aos modelos reparametrizados de posto completo.

Miller (1977) discute as propriedades assintóticas dos estimadores de máxima verossimilhança (ML) para o modelo misto geral, descrito por Hartley & Rao (1967) e apresenta exemplos utilizando o teste da razão de verossimilhança.

Harville (1977), considerando os aspectos computacionais, relata e discute os problemas relacionados com os estimadores de máxima verossimilhança (ML e REML), o uso de algoritmos numéricos e os relacionamentos das aproximações de máxima verossimilhança com as estimações dos efeitos fixos e efeitos aleatórios.

Smith & Hocking (1978) desenvolvem analiticamente a inversa da matriz de covariâncias para o modelo linear misto geral da análise de variância com dados balanceados. Utilizam a matriz para a identificação de estatísticas suficientes e descrevem o desenvolvimento das equações de verossimilhança.

Estabelecem, ainda, relações com os estimadores ANOVA concluindo que: i) as propriedades para as grandes amostras são bem conhecidas, têm distribuição normal e as variâncias amostrais são conhecidas. ii) as estimativas dos efeitos fixos nos modelos mistos, através do método ANOVA não consideram as estimativas dos componentes de variância, como acontece com as estimativas do método ML.

Kackar & Harville (1984) investigam a performance das aproximações para erros padrões dos estimadores dos efeitos fixos e dos efeitos aleatórios do modelo misto geral no contexto da estimação dos efeitos do modelo aleatório de classificação simples balanceado e da estimação dos contrastes entre tratamentos para os delineamentos em blocos incompletos balanceados.

Swallow & Monahan (1984), através de simulação apresentam um estudo comparativo entre quatro metodologias de estimação dos componentes de variância, para o modelo aleatório de classificação simples desbalanceado. ANOVA, ML, REML e MIVQUE. Os resultados desse trabalho indicam que para os componentes de variância σ_a^2 e σ_e^2 : i) os estimadores ANOVA apresentam bom desempenho, exceto quando há sério desbalanceamento dos dados, ou seja, quando $\left(\frac{\sigma_a^2}{\sigma_e^2} > 1\right)$; ii) O ML é excelente quando $\frac{\sigma_a^2}{\sigma_e^2} < 0,5$; iii)

quando $\frac{\sigma_a^2}{\sigma_e^2} \geq 1$, os estimadores MIVQUE(O), os quais são parte do padrão do procedimento VARCOMP do SAS, não se apresentaram como adequados para estimar os componentes. Os autores salientam que esse método deve ser usado, apenas quando é confiável supor que a variância do fator aleatório A, isto é, σ_a^2 , é praticamente zero.

Seraphin (1984), através de simulação de dados, estabelece comparações dos procedimentos para estimativas de componentes de variância: MIVQUE, ML e do Método 3 de Henderson. Utiliza, nesse estudo, o

modelo aleatório de classificação dupla e o modelo misto de classificação dupla com interação. Na mesma linha de pesquisa Valério Filho (1991) apresenta um estudo comparativo, através de simulação de dados, entre os métodos para a estimação de componentes de variância disponíveis no Proc VARCOMP do SAS, ou sejam os estimadores ANOVA, ML, REML e MIVQUE(0). Esses estimadores foram utilizados em dois tipos de modelos mistos com dois fatores: cruzado sem interação e hierárquico.

Milliken & Johnson (1984) apresentam um dos estudos mais completos para analistas do modelo misto, considerando os casos de dados balanceados e desbalanceados. Os autores discutem o modelo misto sistematizando a análise em duas partes: a parte fixa e a parte aleatória, apresentam explicitamente a estimação dos efeitos fixos e aleatórios, testes de hipóteses de interesse, utilizando o sistema computacional estatístico SAS.

Searle (1987) apresenta uma discussão para o modelo misto, onde relata as vantagens e desvantagens dos diversos procedimentos de estimação dos componentes de variância, para os casos de dados balanceados e desbalanceados. Apresenta, também, para o modelo aleatório de classificação simples:

$$y_{ij} = \mu_i + e_{ij}, \text{ com } i = 1, \dots, a \text{ e } j = 1, \dots, n$$

onde: $\mu_i = \mu + \alpha_i$, o desenvolvimento da metodologia apresentada na Tabela 4.

Tabela 4. Estatísticas para “testes de hipóteses” no modelo aleatório de classificação simples sob a suposição de normalidade segundo Searle (1987).

Fórmula	Estatística	Hipótese
A	$N\bar{y}_{..}^2 / [R(\alpha \mu) / (a-1)] \approx F_{1,a-1}$	$H_0: \mu = 0$, para dados balanceados
B	$z = \frac{N\bar{y}_{..}^2}{\sqrt{\sigma_a^2 \left(\sum_{i=1}^a n_i^2 N \right) + \sigma_e^2}} \approx N(0,1)$	$H_0: \mu = 0$, para dados desbalanceados
C	$u = \frac{\sum_{i=1}^a w_i \bar{y}_i}{\sqrt{\sum_{i=1}^a w_i}} \approx N(0,1)$, para $w_i = n_i / (n_i \sigma_a^2 + \sigma_e^2)$	$H_0: \mu = 0$, para dados desbalanceados

As estatísticas z e u , requerem o conhecimento de σ_a^2 e σ_e^2 . Quando há balanceamento nos dados ($n_i = n$, para $\forall i$), tem-se: a fórmula B idêntica a fórmula C; e ainda, as fórmulas

B e C equivalentes a fórmula A, pois seus denominadores são estimados por $\frac{R(\alpha|\mu)}{a-1}$.

lemma (1991,1993a, 1995a) apresenta as principais idéias sobre a problemática de testar hipóteses através de modelos lineares de efeitos fixos em presença de amostras desbalanceadas. Discute os métodos mais utilizados na literatura e as hipóteses mais testadas em problemas práticos, bem como a sua relação com os quatro tipos de somas de quadrados fornecidas pelo Proc GLM do SAS.

Searle et al. (1992) apresentam uma discussão importante sobre os componentes de variâncias. Desenvolvem estudo contextualizando todos os tipos de modelos aleatórios e modelos mistos e discutem a metodologia implementada pelos sistemas estatísticos SAS, GENTAST 5 e BMDP, para a

obtenção das estimativas dos componentes de variância nos modelos mistos. Salientam que o SAS oferece os estimadores ANOVA (baseados no método 3 de Henderson), para as somas de quadrados do Tipo I, os estimadores MIVQUE(0) e os estimadores ML e REML que usam a transformação W de Hemmerle & Hartley (1973) através do proc VARCOMP. O GENTAST 5 contém a rotina REML desenvolvida por H. D. Patterson, em Edimburgo. O BMDP apresenta vários programas para estimativas dos componentes de variância, especificamente, o programa 3V foi um dos primeiros a fornecer as estimativas ML e REML em presença de dados desbalanceados, para qualquer modelo misto.

Schwarz (1993) apresenta um resumo detalhado para a análise do modelo misto geral, discutindo a literatura disponível, assim como os sistemas estatísticos que tratam do modelo misto. Ademais, comenta sobre a necessidade de um trabalho semelhante para subsidiar o usuário do modelo misto, no caso de dados desbalanceados.

Santos (1994) faz uma revisão detalhada dos métodos de análise de modelos lineares de efeitos fixos com classificação dupla com e sem interação. Nesse trabalho discute a performance de onze pacotes estatísticos e apresenta um estudo de estruturação das hipóteses e somas de quadrados a elas associadas

lemma & Perri (1997) apresentam, para o ajuste de modelos fixos desbalanceados através do sistema estatístico SAS, as idéias básicas sobre a parametrização, as equações normais, a estimação, a análise de variância. Discutem as funções estimáveis e as hipóteses fornecidas pelo SAS/GLM relacionando-as com as somas de quadrados associadas.

Perri & lemma (1997) apresentam uma revisão dos principais métodos utilizados para a análise do modelos mistos. Discutem as hipóteses fornecidas pelo SAS, sobre os efeitos fixos, que estão sendo testadas na análise de variância e relacionam-nas com as somas de quadrados associadas.

3 METODOLOGIA

3.1 Aspectos do modelo misto

3.1.1 O modelo misto geral

Segundo Searle (1968, 1971, 1987); Hocking (1985) e Searle et al. (1992) o modelo misto geral pode ser descrito por:

$$y = X\beta + Zv + e \quad (1)$$

onde:

y : é um vetor de variáveis aleatórias observáveis, de dimensão $n \times 1$;

$X\beta$: representa a ocorrência dos efeitos fixos, onde β é o vetor dos parâmetros desconhecidos dos efeitos fixos, de dimensão $p \times 1$ com a matriz modelo correspondente X , de dimensão $n \times p$;

Zv : representa a ocorrência dos efeitos aleatórios, v é o vetor dos efeitos aleatórios de dimensão $m \times 1$, com matriz modelo correspondente Z , de dimensão $n \times m$;

e : é um vetor dos erros aleatórios de variáveis independentes, definido por:

$$e = y - E(y / v),$$

onde :

$$E(y / v) = X\beta + Zv \quad (2)$$

$$E(y) = X\beta$$

assim, seja V a matriz de dispersão de y dada por:

$$V = \text{Var}(y), \quad (3)$$

então de (1), tem-se:

$$V = \text{Var}(Zv + e). \quad (4)$$

Considerando-se as seguintes pressuposições em (4):

$$\text{Cov}(v, e') = 0 \quad \text{e} \quad \text{Var}(e) = \sigma_e^2 I \quad (5)$$

tem-se:

$$V = Z\text{Var}(v)Z' + \sigma_e^2 I. \quad (6)$$

Agora, particionando-se v em r sub-vetores de dimensão q_i ,

$$v' = (v'_1, \dots, v'_r),$$

onde:

$$v'_i = v'_{1i}, \dots, v'_{q_i i},$$

q_i representa o número de elementos em v'_i , tal que $\sum_{i=1}^r q_i = q$, isto é o número de níveis do fator correspondente a v'_i

Particionando-se Z conforme (7), em r sub-matrizes de dimensões $n \times q_i$, tem-se, conforme a igualdade (7):

$$Z = (Z_1, \dots, Z_r). \quad (8)$$

De forma que o modelo descrito em (1), pode ser dado por:

$$y = X\beta + \sum_{i=1}^r Z_i v_i + e \quad (9)$$

com:

$$V = \text{Var}(y) = \text{Var} \left[\sum_{i=1}^r Z_i v_i + e \right]. \quad (10)$$

Portanto, considerando-se as pressuposições, para as esperanças, variâncias e covariâncias:

$$E(v_i) = 0 ; \text{ para todo } i = 1, 2, \dots, r \quad (11)$$

$$\text{Var}(v_i) = \sigma_i^2 I_{q_i} ; \text{ para todo } i \quad (12)$$

$$\text{Cov}(v_i, v_{i'}) = 0 ; \text{ para todo } i, i' \quad (13)$$

$$\text{Cov}(v_i, e') = 0 . \quad (14)$$

Usando as expressões (12) a (14) em (10), tem-se:

$$V = \sum_{i=1}^r Z_i Z_i' \sigma_i^2 + \sigma_e^2 I \quad (15)$$

e finalmente, definindo-se:

$$v_0 = e, \sigma_0^2 = \sigma_e^2 \text{ e } Z_0 = I \quad (16)$$

tem-se:

$$y = X\beta + \sum_{i=0}^r Z_i v_i \quad (17)$$

$$V = \sum_{i=0}^r Z_i Z_i' \sigma_i^2 \quad (18)$$

As expressões (17) e (18), constituem a descrição que é freqüentemente usada como forma geral do modelo misto. Naturalmente ela

pode ser mais complicada caso a estrutura de variâncias e covariâncias não tenha a forma diagonal, $I\sigma_i^2$.

Uma forma simples de (17) quando β é o vetor de médias de caselas é dada por:

$$y = W\mu + \sum_{i=0}^r Z_i v_i \quad (19)$$

$$\text{onde: } \varepsilon = \sum_{i=0}^r Z_i v_i, \quad (20)$$

tem-se:

$$y = W\mu + \varepsilon \quad (21)$$

Portanto, (19) e (21) são as caracterizações do modelo misto para o modelo de médias de caselas, onde a matriz W de (19), tem posto coluna completo.

3.1.2 Modelo Misto com classificação dupla

O modelo usual de classificação dupla com interação geralmente é escrito na forma superparametrizada. e segundo Searle (1971), um modelo é dito superparametrizado se ele exhibe explicitamente um parâmetro para cada fator do modelo. Dentre outros lemma (1991,1993a,1995a) e lemma & Perri (1997) apresentam esse modelo denotado como Modelo-S.

Considera-se aqui o modelo misto desbalanceado com dois fatores cruzados com interação, supondo-se o fator A fixo e o fator B aleatório, caracterizado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \quad (22)$$

com $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $k = 1, \dots, n_{ij}$,

onde :

y_{ijk} : é a k-ésima observação no i-ésimo nível do fator A e j-ésimo nível do fator B;

μ : é uma constante inerente a todas as observações;

α_i : é o efeito do i-ésimo nível do fator A

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

$(\alpha\beta)_{ij}$: é o efeito da interação entre o i-ésimo nível do fator A e o j-ésimo nível do fator B;

e_{ijk} : é o erro aleatório referente à k-ésima observação do i-ésimo nível do fator A e o j-ésimo nível do fator B; tem-se:

$$E(y_{ijk}) = \mu + \alpha_i = \mu_i, \quad (23)$$

onde:

$\mu + \alpha_i$ é a parte dos efeitos fixos do modelo, e:

$\beta_j + (\alpha\beta)_{ij} + e_{ijk}$ é a parte dos efeitos aleatórios do modelo, com distribuições:

$\beta_j \overset{iid}{\cap} N(0, \sigma_\beta^2)$, $(\alpha\beta)_{ij} \overset{iid}{\cap} N(0, \sigma_{\alpha\beta}^2)$, $e_{ijk} \overset{iid}{\cap} N(0, \sigma_e^2)$, respectivamente; e

β_j , $(\alpha\beta)_{ij}$ e e_{ijk} , são todos variáveis aleatórias independentes.

Na forma matricial, tem-se:

$$y = X_1\theta + X_2v_2 + X_3v_3 + \varepsilon, \quad \text{onde:} \quad (24)$$

$X_1\theta$ é parte fixa do modelo.

$X_2v_2 + X_3v_3 + \varepsilon$ é a parte aleatória do modelo com $v_i \overset{iid}{\cap} N(0, \sigma_i^2 I)$, $i = 2, 3$, e

$\varepsilon \overset{iid}{\cap} N(0, \sigma_e^2 I)$. Tem-se, assim, a seguinte população de parâmetros:

α , σ_β^2 , $\sigma_{\alpha\beta}^2$, σ_e^2 .

Segundo lemma (1991,1993a,1995a), conforme o interesse do pesquisador, o Modelo-S pode assumir diferentes caracterizações, desde o modelo mais simples, apenas com a constante μ , até o modelo mais completo. Assim tem-se

$$y = X_1\theta + \varepsilon^* \quad \leftrightarrow \quad y_{ijk} = \mu + \alpha_i + e_{ijk}^*, \quad (25)$$

onde:

$e_{ijk}^* = \beta_j + (\alpha\beta)_{ij} + e_{ijk}$, que são os efeitos aleatórios atribuídos a variável y_{ijk} .

Ainda segundo o autor citado, um modelo alternativo equivalente ao Modelo-S é o modelo de médias de caselas, Modelo-M, considerado excelente para simplificar a interpretação das hipóteses testadas, que é dado por:

$$y_{ijk} = \mu_i + \varepsilon_{ijk}^*, \quad (26)$$

onde : $\mu_i = \mu + \alpha_i$.

3.2 Estimação

Na análise do modelo misto tem-se, em geral, interesse na estimação e testes dos efeitos fixos e dos efeitos aleatórios e na estimação dos componentes de variância. Segundo Milliken & Johnson (1984), dentre outros, a análise consiste em duas partes, uma para a parte aleatória e outra para a parte fixa.

A análise da parte aleatória consiste de estimativas e testes de hipóteses sobre os componentes de variância. A análise da parte fixa consiste de estimativas e testes de hipóteses sobre funções estimáveis de θ .

3.2.1 Estimação da parte fixa do modelo misto

Segundo Rao (1945), conforme cita lemma (1991,1993a,1995a), “uma função linear paramétrica $\lambda'\theta$ é estimável no modelo linear de Gauss-Markov, se e somente se $\lambda \in C(X')$ ”.

Em termos práticos tem-se que, se $E(y) = X\theta$, então, a menos de desconexões, $X\theta$ é estimável. Um modelo é conectado se todo contraste simples de diferenças de dois níveis de um fator é estimável, uma discussão detalhada poderá ser encontrada em Searle (1987), lemma (1991, 1993a). Além disso, combinações lineares de funções estimáveis são estimáveis. Uma combinação linear $\lambda'\theta$ é estimável para esse modelo misto, se e somente se existe um vetor c , tal que:

$$E(c'y) = \lambda'\theta.$$

A estimativa de $\lambda'\theta$, uma função estimável de θ , é $\lambda'\hat{\theta}$, onde $\hat{\theta}$ é a solução para um conjunto de equações normais ou para equações de verossimilhança.

Segundo Milliken & Johnson (1984), Hocking (1985) e Searle et al. (1992) para o modelo linear de efeitos fixos definido por $y = X\beta + e$, com $e \sim (0, \sigma^2 I)$, entre outros métodos de estimação disponíveis na literatura, o mais utilizado é o dos mínimos quadrados. Esse método conduz ao sistema de equações normais $X'X\beta = X'y$. Segundo Searle (1971, 1987) lemma (1991,1995a), se a matriz X é de posto coluna completo, então $X'X$ é positiva definida e, portanto, não singular. Nesse contexto o sistema $X'X\beta = X'y$ tem solução única dada por $\hat{\beta} = (X'X)^{-1} X'y$ que é solução aproximada de mínimos quadrados para $y = X\beta$ e coincide com o estimador de mínimos quadrados para β . Isto nos garante que cada componente do vetor β é estimável e, portanto, testável. Por outro lado, se X não é de posto coluna

completo, então $X'X$ é singular e o sistema de equações normais sempre consistente, é indeterminado. Suas soluções exatas podem ser obtidas, entre outras, por $\beta^o = (X'X)^g X'y$, onde $(X'X)^g$ é uma matriz inversa generalizada qualquer de $X'X$ e são soluções aproximadas de mínimos quadrados para $y = X\beta$.

Essas equações podem ser usadas para o modelo misto dado em (1) ou (9), mas elas fornecem estimadores que são obtidos sem considerar os componentes de variância. Isto ocorre, porque, o método de estimação de mínimos quadrados minimiza $(y - X\beta)'(y - X\beta)$ com respeito à β , também chamado de sistema de mínimos quadrados ordinários: OLSE (Ordinary Least Squares Estimation) e fornece a estimativa de $X\beta$, dada por $OLSE(X\beta) = X(X'X)^g X'y$, não envolvendo a matriz de variâncias e covariâncias. Nesse caso, para os modelos mistos, eles nem sempre são estimadores do tipo BLUE (Best Linear Unbiased Estimator).

Conforme o Teorema de Gauss-Markov o BLUE de uma função estimável $k'\beta$ é dado por $k'\beta^o$, onde β^o é qualquer solução das equações normais. Assim para o modelo (1), considerando-se para o sistema de equações de mínimos quadrados generalizado: GLS (Generalized Least Squares) $(y - X\beta)' V^{-1}(y - X\beta)$, uma solução dada é $\beta_{GLSE}^o = (X'V^{-1}X)^g X'V^{-1}y$, (Generalized Least Squares Estimation).

Naturalmente, tem-se, por conseguinte:

$$BLUE(\lambda'X\beta) = \lambda'X(X'V^{-1}X)^g X'V^{-1}y \quad (27)$$

Desse modo, os componentes de variância são considerados nas estimativas das funções estimáveis $\lambda'X\beta$ dos efeitos fixos.

A variância de (27) é dada por:

$$\text{Var}(\text{BLUE}(\lambda'X\beta)) = \lambda'X(X'V^{-1}X)^s X'\lambda \quad (28)$$

Esse método de estimação deve ser o preferido sobre o uso do método de mínimos quadrados ordinários: $(X'X)\beta_{OLSE} = X'y$.

Uma dificuldade existente na utilização prática de (27) e (28) consiste no fato de que elas requerem $V = \text{Var}(y)$, ou seja, envolvem os componentes de variância, em muitas situações em que eles não são conhecidos. Uma alternativa interessante consiste na obtenção do conjunto de suas estimativas, que são utilizadas em lugar dos componentes em V . Substituí-se V por \hat{V} e assim tem-se:

$$\text{BLUE}_{\hat{V}}(\lambda'X\beta) = \lambda'X(X'\hat{V}^{-1}X)^s X'\hat{V}^{-1}y, \quad (29)$$

onde: $\hat{V} = \sum_{i=0}^r Z_i Z_i' \hat{\sigma}_i^2$

Por outro lado, para o modelo de médias de caselas descrito em (19) e (21), sendo W de posto coluna completo, μ é estimável, então a expressão (29), para o BLUE de μ é dada por:

$$\hat{\mu} = (W'V^{-1}W)^{-1} W'V^{-1}y, \text{ onde } \text{Var}(\hat{\mu}) = (W'V^{-1}W)^{-1} \quad (30)$$

Da mesma forma que em (29), pode-se substituir em (30) V por \hat{V} .

3.2.1.1 Estimação da parte fixa do modelo misto com classificação dupla

Particularmente, para o modelo de classificação dupla, como definido em (24), o estimador de mínimos quadrados de $\lambda'\theta$ é $\lambda'\theta_{LS}$, onde $\hat{\theta}_{LS} = (X_1' X_1)^s X_1' y$, ou qualquer outra solução para $\hat{\theta}$ em $X_1' X_1 \hat{\theta} = X_1' y$. Se os elementos de V são conhecidos, o melhor estimador linear não-viesado (BLUE) de $\lambda'\theta$ é $\lambda''\theta_{BLUE}$ onde:

$$\hat{\theta}_{BLUE} = (X_1' \hat{V}^{-1} X_1)^s X_1' \hat{V}^{-1} y. \quad (31)$$

Nesse caso, a matriz de covariância estimada é dada por:

$$\hat{V} = \hat{\sigma}_2^2 X_2 X_2' + \hat{\sigma}_3^2 X_3 X_3' + \hat{\sigma}_e^2 I_n \quad (32)$$

onde $\hat{\sigma}_2^2, \hat{\sigma}_3^2, \hat{\sigma}_e^2$ são os estimadores dos componentes de variância, os quais podem ser obtidos por qualquer um dos métodos: ANOVA, REML, ML, MINQUE ou MIVQUE.

De modo análogo, o estimador de mínimos quadrados ponderados para $\lambda'\theta$ é $\lambda''\theta_w$, onde:

$$\hat{\theta}_w = (X_1' \hat{V}^{-1} X_1)^- X_1' \hat{V}^{-1} y, \quad (33)$$

para qualquer solução para $\hat{\theta}$ em

$$X_1' \hat{V}^{-1} X_1 \hat{\theta} = X_1' \hat{V}^{-1} y.$$

Usualmente os pesquisadores desejam realizar testes de hipóteses sobre os efeitos principais e interações. Entretanto, a parte fixa do modelo

geralmente envolve efeitos principais e interações entre fatores fixos e aleatórios. Consequentemente, para o modelo misto, os valores esperados de muitos quadrados médios dos efeitos fixos apresentam a expressão a seguir

$$E(QM_{FIXO}) = \sigma_e^2 + C_1\sigma_1^2 + C_2\sigma_2^2 + \dots + C_k\sigma_k^2 + Q^2, \quad (34)$$

onde: C_k , para $k = 1, \dots, k$, são obtidos em função da estrutura dos dados conforme mostra a Tabela 2 e a forma quadrática dos efeitos fixos Q^2 tem o valor zero, se e somente se, a hipótese nula sobre os efeitos fixos é verdadeira.

Já, em delineamentos desbalanceados e mesmo em alguns casos de delineamentos balanceados, não existe um quadrado médio com esperança,

$$\sigma_e^2 + C_1\sigma_1^2 + C_2\sigma_2^2 + \dots + C_k\sigma_k^2.$$

Nesse caso o teste estatístico é $F = QM_{FIXO} / \tilde{Q}$, onde \tilde{Q} tem a construção:

$$\tilde{Q} = a_1QM_1 + a_2QM_2 + \dots + a_kQM_k + a_{k+1}QM_{k+1} \quad (35)$$

o qual sob condições da hipótese nula tem distribuição aproximada de F com r_{FIXO} e GL_{CORRG} graus de liberdade, onde GL_{CORRG} pode ser obtido através da aproximação de Satterthwaite, conforme apresenta a expressão:

$$GL_{CORRG} = \frac{\tilde{Q}^2}{(a_1QM_1)^2/r_1 + (a_2QM_2)^2/r_2 + \dots + (a_{k+1}QM_{k+1})^2/r_{k+1}}. \quad (36)$$

Apresenta-se na Tabela 5 a análise de variância, onde as somas de quadrados são obtidas através do método do ajuste de constantes.

Tabela 5. Análise de variância e esperanças dos quadrados médios.

CAUSA DE VARIAÇÃO	gl	SOMAS DE QUADRADOS	$E(QM)$
A	r_1	$R(\alpha \mu)$	$\sigma_e^2 + k_4\sigma_{\alpha\beta}^2 + k_5\sigma_\beta^2 + Q(\alpha)$
B	r_2	$R(\beta \mu, \alpha)$	$\sigma_e^2 + k_3\sigma_{\alpha\beta}^2 + k_2\sigma_\beta^2$
AxB	r_3	$R(\alpha\beta \mu, \alpha, \beta)$	$\sigma_e^2 + k_1\sigma_{\alpha\beta}^2$
RESÍDUO	r_4	S.Q.RESÍDUO	σ_e^2

Os coeficientes k_1, k_2, k_3, k_4 e k_5 , são constantes que dependem apenas da estrutura dos dados, isto é das matrizes de delineamentos, e podem ser obtidos das expressões das esperanças matemáticas das somas de quadrados e os valores a_i podem ser obtidos através das expressões a seguir:

$$a_1 = 1 - \frac{k_5}{k_2} - \frac{k_4}{k_1} + \frac{k_3 \times k_5}{k_1 \times k_2},$$

$$a_2 = \frac{k_4}{k_1} + \frac{k_3 \times k_5}{k_1 \times k_2} e$$

$$a_3 = \frac{k_5}{k_3}.$$

3.2.1.2 Hipóteses e somas de quadrados associadas

Para testar os efeitos de linhas, colunas e interações no modelo linear de classificação dupla, a literatura apresenta vários tipos de hipóteses. Dentre os vários tipos de hipóteses o sistema computacional SAS, através do procedimento GLM para o modelo de efeitos fixos, apresenta:

i) sobre os efeitos de linhas:

TIPO I: Hipóteses sobre médias ponderadas de linhas, \bar{L}_i .

$$H_o^{(1)}: \bar{L}_i = \bar{L}_{i'}, \forall (i, i'), i = 1, \dots, a: \text{ onde } a \text{ é o número de linhas}$$

com:

$$SQH_o^{(1)} = R(\alpha|\mu)$$

TIPO II: Hipóteses sobre médias ponderadas de linhas ajustadas para colunas, $(\bar{L}/C)_i, i = 1, \dots, a$

$$H_o^{(2)}: \bar{L}_i = (\bar{L}/C)_i, \forall i$$

com:

$$SQH_o^{(2)} = R(\alpha|\mu, \beta)$$

TIPO III: Hipóteses sobre médias não ponderadas de linhas, \bar{l}_i .

$$H_o^{(3)}: \bar{l}_i = \bar{l}_{i'}, \forall (i, i'), i = 1, \dots, a$$

com:

$$SQH_o^{(3)} = R(\alpha|\mu, \beta, (\alpha\beta))$$

TIPO IV: Hipóteses do Tipo IV. A igualdade entre os efeitos de linhas é verificada através de médias não ponderadas, assim elas podem ser semelhantes às do Tipo III. Se, no entanto, existem caselas vazias em geral, as hipóteses dos Tipo III e IV são diferentes. Ademais, em geral, na presença de caselas vazias, as hipóteses do Tipo IV não são únicas. Elas podem ser obtidas construindo-se contrastes entre médias de caselas que estão na mesma coluna, iniciando-se o processo sempre pela última linha. Maiores

detalhes e forma de obtenção são encontrados em lemma (1991 e 1995a); Santos (1994) e lemma & Perri (1997). As hipóteses do Tipo IV, bem como sua soma de quadrados, serão aqui denotadas por:

$$H_o^{(4)}:, \text{ e}$$

$$SQH_o^{(4)}, \text{ respectivamente.}$$

De forma análoga, tem-se:

ii) sobre os efeitos de colunas:

TIPO I: Hipóteses sobre médias ponderadas de colunas, \bar{C}_i .

$H_o^{(5)}: \bar{C}_j = \bar{C}_{j'}, \forall (j, j'), j = 1, \dots, b$; onde b representa o número de colunas.

$$SQH_o^{(5)} = R(\beta|\mu)$$

TIPO II: Hipóteses sobre médias ponderadas de colunas ajustadas para linhas, $(\bar{C}/L)_j$

$$H_o^{(6)}: \bar{C}_j = (\bar{C}/L)_j, \forall j, j = 1, \dots, b;$$

com:

$$SQH_o^{(6)} = R(\beta|\mu, \alpha)$$

TIPO III: Hipóteses sobre médias não ponderadas de colunas, \bar{c}_j .

$$H_o^{(7)}: \bar{c}_j = \bar{c}_{j'}, \forall (j, j')$$

com:

$$SQH_o^{(7)} = R(\beta|\mu, \alpha, (\alpha\beta))$$

TIPO IV: Hipóteses do Tipo IV. A igualdade entre os efeitos de colunas é verificada através de médias não ponderadas, assim elas podem ser semelhantes às do Tipo III. Se, no entanto, existem caselas vazias, as hipóteses dos Tipo III e IV podem ser diferentes. Ademais, em geral, na presença de caselas vazias, as hipóteses do Tipo IV não são únicas. Elas podem ser obtidas construindo-se contrastes entre médias de caselas que

estão na mesma linha, iniciando-se o processo sempre pela última coluna. Maiores detalhes e forma de obtenção são encontrados em lemma (1991 e 1995a); Santos (1994) e lemma & Perri (1997). Aqui ela será representada por:

$$H_o^{(8)}: , \text{ com } SQH_o^{(8)} = (B' \theta^0)' [B'(X'X)^G B]^{-1} (B' \theta^0)$$

iii) Hipóteses sobre a interação

Segundo lemma (1991, 1993a e 1995a), se os dados são desbalanceados com todas as caselas ocupadas, para o modelo com dois fatores de efeitos fixos, tem-se $\frac{1}{4}ab(a-1)(b-1)$ funções lineares estimáveis possíveis, para a linhas e b colunas, das quais $(a-1)(b-1)$ são linearmente independentes, ou seja $(a-1)(b-1)$ são graus de liberdade para a interação. A hipótese testável nesse contexto é:

$$H_o^{(9)}: \mu_{ij} + \mu_{i'j'} - \mu_{ij'} - \mu_{i'j} = 0, \quad \forall (i, j, i', j')$$

com:

$$SQH_o^{(9)} = R((\alpha\beta) | \mu, \alpha, \beta)$$

Em presença de caselas vazias essa regra não é válida . O teste da interação não deve ser utilizado como critério de escolha entre um modelo com e sem interação. Maiores detalhes são apresentados e discutidos em lemma (1991, 1993a e 1995a).

As formas de obtenção das somas de quadrados associadas às hipóteses fornecidas através do proc GLM do SAS, são detalhadamente discutidas em lemma (1991); Santos (1994) e lemma & Perri (1997).

3.2.2 Estimação dos componentes de variância

Para a estimativa do BLUE de uma função estimável dos parâmetros de efeitos fixos, é necessário o conhecimento das estimativas dos componentes de variância. Um problema com o qual o analista do modelo misto se depara, na estimação desses componentes especificamente para dados desbalanceados, reside no fato de existirem vários métodos disponíveis. Escolher um deles pode não ser uma questão tão simples. Em decorrência do avanço tecnológico e a facilidade de acesso ao equipamento da área de informática, a escolha tem estado entre um dos dois métodos baseados na máxima verossimilhança, até que ocorra maior aceitação de outras metodologias.

Nesse contexto, considerando-se que o conjunto dos tópicos que envolvem a estimação dos componentes de variância é muito extenso e complexo para um relato detalhado, optou-se por apresentar neste trabalho um resumo dos seguintes métodos disponíveis na literatura:

ANOVA: Análise de Variância.

Método 1 de Henderson

Método 2 de Henderson

Método 3 de Henderson

ML: Máxima Verossimilhança.

REML: Máxima Verossimilhança Restrita.

MINQUE: Estimador Quadrático Não-viesado de Norma Mínima.

MINQUE Iterativo: Estimador Quadrático Não-viesado de Norma Mínima Iterativo (I-MINQUE).

MINQUE(0): Estimador Quadrático Não-viesado de Norma Mínima (caso especial - matriz identidade como peso)

MIVQUE: Estimador Quadrático Não-viesado de Variância Mínima

3.2.2.1 O método ANOVA

Para dados balanceados usa-se, tradicionalmente, os estimadores obtidos pelo método da análise de variância (ANOVA), os quais são obtidos igualando-se os quadrados médios observados no quadro da ANOVA, como seus valores esperados (combinações lineares dos componentes de variância). Esse procedimento produz equações lineares, e as soluções são tomadas como os estimadores ANOVA.

Segundo Hocking (1985), Searle (1987) e Searle et al. (1992), o princípio do método ANOVA para dados balanceados pode ser generalizado para o caso dos dados desbalanceados. A generalização consiste em usar qualquer forma quadrática linearmente independente das observações no lugar das somas de quadrados.

Considerando-se o modelo definido em (1) e admitindo-se que σ^2 é o vetor dos componentes de variância, que devem ser estimados, e ainda que, q é um vetor de mesma ordem de σ^2 , o qual pode ser definido por:

$$q = \{ {}_c q_i \}_{i=0}^r = \{ y' A_i y \}_{i=0}^r, \text{ onde } q \text{ é um vetor de } r+1 \text{ formas}$$

quadráticas de y , tal que A_i é simétrica, onde:

$$E(q_i) = \text{tr}(A_i V) = \text{tr} \left[A_i \sum_{i=0}^r Z_i Z_i' \sigma_i^2 \right] = \sum_{j=0}^r \text{tr}(Z_i' A_i Z_i) \sigma_i^2, \text{ que pode ser}$$

escrito como:

$$E(q) = C \sigma^2, \tag{37}$$

onde, C é alguma matriz, tal que $C = \{ \text{tr}(Z_i' A_i Z_i) \}_i$ e $\sigma^2 = \{ {}_c \sigma_i^2 \}_{i=0}^r$.

De, (37), igualando-se a q , tem-se:

$$C \hat{\sigma}^2 = q \tag{38}$$

Se $C \hat{\sigma}^2 = q$ é não singular, tem-se:

$$\hat{\sigma}^2 = C^{-1}q \quad (39)$$

$Var(\hat{\sigma}^2) = C^{-1}Var(q)C^{-1}$, onde os elementos de $Var(q)$ são as variâncias e as covariâncias das respectivas formas quadráticas, elementos de q .

As equações (39), produzem as estimativas ANOVA, para os componentes de variância, com dados desbalanceados. Entretanto se a matriz C não é quadrada, então $C\sigma^2 = q$, pode apresentar mais equações do que componentes de variância, ou seja, o sistema de equações é inconsistente. Entretanto se C tem posto coluna completo, então $C'C$ é positiva definida e portanto não singular. Nesse caso o sistema $C'C\sigma^2 = C'q$ tem solução única dada por $\hat{\sigma}^2 = [C'C]^{-1}C'q$, e essa solução aproximada de mínimos quadrados para $C\sigma^2 = q$, coincide com o estimador de mínimos quadrados para σ^2 . Isto nos garante que cada elemento de σ^2 é estimável e portanto testável. Se C não tem posto coluna completo, σ^2 não é estimável. Suas vantagens e desvantagens são amplamente discutidas em Searle et al. (1992).

3.2.2.2 Os métodos de Henderson

Searle et al. (1992) enfatizam que os três procedimentos desenvolvidos por Henderson (1953) para as estimativas dos componentes de variância estabelecem três diferentes conjuntos de formas quadráticas que podem ser usados como elementos de q , de (39).

3.2.2.2.1 Método 1

As estimativas do método 1, para dados desbalanceados, são obtidas utilizando-se formas quadráticas análogas as somas de quadrados

usadas para dados balanceados. Assim, por exemplo, para o modelo de classificação dupla cruzada com interação:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad (40)$$

com : $i = 1, \dots, a$, $j = 1, \dots, b$ e $k = 1, \dots, n_{ij}$.

A soma de quadrado familiar, $bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$ para o caso balanceado ($n_{ij} = n$), apresenta a forma correspondente para o caso desbalanceado dada por:

$$SQA = \sum_{i=1}^a n_i (\bar{y}_{i..} - \bar{y}_{...})^2 \text{ que é usualmente representada por } R(\alpha|\mu).$$

As somas de quadrados para o caso desbalanceado podem ser vistas como uma generalização, nesse caso as quatro somas de quadrados obtidas por esse processo, são dadas por:

$$SQA = \sum_{i=1}^a n_i (\bar{y}_{i..} - \bar{y}_{...})^2, \text{ onde: } \bar{y}_{i..} = \frac{\sum_{j=1}^b \sum_{k=0}^{n_{ij}} y_{ijk}}{n_i}, \bar{y}_{...} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=0}^{n_{ij}} y_{ijk}}{n_{..}}, n_i = \sum_{j=1}^b n_{ij} \text{ e}$$

$$n_{..} = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$$

$$SQB = \sum_{j=1}^b n_j (\bar{y}_{.j.} - \bar{y}_{...})^2, \text{ onde: } \bar{y}_{.j.} = \frac{\sum_{i=1}^a \sum_{k=0}^{n_{ij}} y_{ijk}}{n_j}, n_j = \sum_{i=1}^a n_{ij}$$

$$SQAB^* = \sum_{i=1}^a \sum_{j=1}^b n_{ij} - \sum_{i=1}^a n_i \bar{y}_{i..}^2 - \sum_{j=1}^b n_j \bar{y}_{.j.}^2 + N \bar{y}_{...}^2, \text{ e} \quad (41)$$

$$SQR = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=0}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2, \text{ onde: } \bar{y}_{ij.} = \frac{\sum_{k=0}^{n_{ij}} y_{ijk}}{n_{ij}}$$

As estimativas são, então, obtidas igualando-se os valores esperados de cada uma das quatro formas quadráticas com seus valores esperados. Para o modelo aleatório resultam quatro equações lineares de componentes de variância.

Assim, para $N' = a - b + 1$, as equações são:

$$\begin{bmatrix} N - k'_1 & k_3 - k'_2 & k_3 - k'_{2,3} & a - 1 \\ k_4 - k'_1 & N - k'_2 & k_4 - k'_{2,3} & b - 1 \\ k'_1 - k_4 & k'_2 - k_3 & N - k_3 - k_4 + k'_{2,3} & N' \\ 0 & 0 & 0 & N - s \end{bmatrix} \begin{bmatrix} \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\beta^2 \\ \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} SQA \\ SQB \\ SQAB^* \\ SQRES \end{bmatrix}, \quad (42)$$

naturalmente, $E(SQRES) = (N - s)\sigma_e^2$, onde s é o número de caselas ocupadas, define-se:

$$k_1 = \sum_{i=1}^a n_i^2, \quad k_2 = \sum_{j=1}^b n_j^2, \quad k_3 = \sum_{i=1}^a \frac{\sum_{j=1}^b n_{ij}^2}{n_i}, \quad k_4 = \sum_{j=1}^b \frac{\sum_{i=1}^a n_{ij}^2}{n_j}, \quad k_{2,3} = \sum_i \sum_j n_{ij}^2,$$

e para qualquer k , define-se :

$$k'_r = k_r / N$$

Esse método fornece estimativas não viesadas com variância mínima, quando a estrutura de dados é balanceada, com a vantagem de ter relativa facilidade aritmética. Para os modelos mistos com dados desbalanceados, qualquer tentativa em usar o método 1 resulta em expressões para as esperanças dos quadrados médios que contêm funções de efeitos fixos, de modo que é aplicável apenas para os modelos completamente aleatórios. Pode-se adaptar essa metodologia para o modelo misto, considerando-se os efeitos fixos do modelo como aleatórios ou não existentes. Nesse caso as estimativas dos componentes de variância para os efeitos aleatórios do modelo são viesadas.

3.2.2.2 Método 2

O propósito do método 2 é fornecer uma metodologia que tenha a relativa facilidade aritmética do método 1, porém que seja aplicável ao modelo misto.

Primeiramente deve-se, para o modelo misto usual, $y = X\beta + Zv + e$, resolver as equações normais:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \beta^0 \\ v^0 \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

para β^0 . Assim considera-se o vetor ajustado para β^0 ou seja, $y_a = y - X\beta^0$, retira-se a média dos efeitos fixos e redefine-se como $X\beta$, excluindo-se a média geral, ou seja, o vetor $\mu 1$ e então a equação do modelo é dada por:

$$y = \mu 1 + X\beta + Zv + e \quad (43)$$

Determina-se $\beta^0 = Ly$, para L escolhido de forma que $y_a = y - X\beta^0$, assim, $\beta^0 = Ly$ é tal que:

$$y_a = \mu_0 1 + Zv + \varepsilon, \quad (44)$$

onde: $\mu_0 1 = \mu(I - XL)1 + (X - X'LX)\beta$ e $\varepsilon = (I - XL)e$, isto é, tal que a equação do modelo para y_a tem os efeitos aleatórios da forma como eles são em y , isto é, Zv . Então o procedimento seguinte é aplicar o método 1 em y_a . Searle (1968) apresenta uma descrição geral detalhada e estabelece as condições necessárias para tanto.

3.2.2.3 Método 3

O método 3, usualmente conhecido como o método de ajuste de constantes, baseia-se nas reduções nas somas de quadrados da análise do

modelo de efeitos fixos, ou seja nas reduções das somas de quadrados devida ao ajuste de um modelo completo e de vários sub-modelos dele. Fornece estimadores não viesados para os modelos mistos, mas eles podem envolver mais formas quadráticas do que são os componentes a serem estimados .

Para o modelo misto geral $y = X\beta + Zv + e$, supõe-se combinar β e v em um simples vetor b . Pode-se então escrever o modelo como:

$$y = Wb + e, \quad (45)$$

onde: b é o vetor de parâmetros, portanto contém os efeitos fixos e aleatórios. Assim no ajuste de um modelo de efeitos fixos, com a equação (45), é bem conhecido que o melhor estimador linear imparcial de Wb é :

$$BLUE(Wb) = Wb^0 = W(W'W)^{-1}W'y = WW^+y, \quad (46)$$

onde: W^+ é a inversa de Moore-Penrose e $(W'W)^{-1}$ é uma inversa generalizada de mínimos quadrados. Portanto a soma de quadrado do resíduo após o ajuste do modelo é:

$$SQRES = (y - Wb^0)'(y - Wb^0) = y'y - y'WW^+y \quad (47)$$

A partição de $y'y$ em duas somas de quadrados é resumida na Tabela 6 a seguir.

Tabela 6. Partição da soma de quadrados total devido ao ajuste do modelo $y = Wb + e$

Redução devido ao ajuste do modelo	$SQR = y'WW^+y$
Resíduo	$SQRES = y'y - y'WW^+y$
Total	$SQT = y'y$

Um teste para a hipótese $H_0: b = 0$, no modelo de efeitos fixos

$$\text{supondo-se que } y \approx N(Wb, \sigma_e^2 I_N), \text{ pode ser obtido por: } F = \frac{\frac{SQR}{r(W)}}{\frac{SQRES}{N - r(W)}}.$$

Para as estimativas dos componentes de variância do modelo misto a atenção está, em geral, voltada para as formas quadráticas baseadas na redução $R(b)$, a qual pode ser representada como:

$$R(b) = y'W(W'W)^{-1}W'y = y'WW^+y \quad (48)$$

Considerando-se a partição Wb , tal que:

$$E(y) = W_1b_1 + W_2b_2, \quad (49)$$

tem-se:

$$R(b_1|b_2) = y'[W_1 \ W_2][W_1 \ W_2]'y. \quad (50)$$

para $M_1 = I - W_1(W_1'W_1)^{-1}W_1' = M_1' = M_1^2$, com $M_1W_1 = 0$.

$$R(b_2|b_1) = y'M_1W_2(W_2'M_1W_2)^{-1}W_2'M_1y \quad (51)$$

As somas de quadrados definidas como em (51) são usadas pelo método 3. Embora nos modelos de efeitos completamente fixos tais somas de quadrados sejam utilizadas nos numeradores das estatísticas F , para as quais as suposições de normalidades são requeridas, o uso dessas somas de quadrados no método 3, para as estimativas dos componentes de variância, não implica em nenhuma suposição de normalidade.

Adaptando-se (51), para os modelos mistos, tem-se interesse em escrever Wb como $X\beta + Zy$, considerando-se para tanto a formulação geral de

$E[R(b_2|b_1)]$. Essa redução ilustra duas importantes propriedades do método 3: ou seja o valor esperado, $E[R(b_2|b_1)]$ é função somente das variâncias dos efeitos aleatórios, não contém termos devido aos efeitos fixos do modelo, como ocorre com o método 1; e não é afetada pelas interações dos efeitos fixos com os efeitos aleatórios, como ocorre no método 2.

Assim, define-se:

$$\text{Var}(b) = E(bb') - E(b)E(b'), \quad (52)$$

para cada um sub-vetor de b , tendo os elementos que são os efeitos devidos a um fator aleatório,

Com (52) tem-se:

$$V = \text{Var}(y) = \text{Var}(Wb) + \sigma_e^2 I_N = W\text{Var}(b)W' + \sigma_e^2 I$$

Para a forma quadrática de (51), onde $W = [W_1 \ W_2]$, tem-se:

$$E[R(b_2/b_1)] = \text{tr} \left\{ \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \left[M_1 W_2 (W_2' M_1 W_2)^{-1} W_2' M_1 \right] \begin{bmatrix} W_1 & W_2 \end{bmatrix} E(bb') \right\} + \sigma_e^2 \text{tr} \left[M_1 W_2 (W_2' M_1 W_2)^{-1} W_2' M_1 \right] \quad (53)$$

Usando-se $M_1 W_1 = 0$ de (51), considerando-se que $b' = [b_1' \ b_2']$ e a idempotência de $\left[M_1 W_2 (W_2' M_1 W_2)^{-1} W_2' M_1 \right]$, (53) reduz-se a seguinte expressão:

$$E[R(b_2|b_1)] = \text{tr} \left[W_2' M_1 W_2 E(b_2 b_2') \right] + \sigma_e^2 \left(r_{[W_1 \ W_2]} - r_{W_1} \right) \quad (54)$$

onde: r_{W_1} e $r_{[W_1 \ W_2]}$ são respectivamente os postos das matrizes W_1 e $[W_1 \ W_2]$.

Agora, voltando-se à equação do modelo misto, $y = X\beta + Zv + e$, e considerando-se:

$$M = I - XX^+ = I - X(X'X)^{-1}X' = M' = M^2, \text{ com } MX = 0, \quad (55)$$

tem-se de (54) que:

$$\begin{aligned}
E[R(v|\beta)] &= \text{tr}[Z'MZ]E(vv') + \sigma_e^2(r[X \ Z] - r[X]) = \\
&= \text{tr}\left[Z'MZ \left\{ \sum_{i=1}^r \sigma_i^2 I_{q_i} \right\}\right] + \sigma_e^2(r[X \ Z] - r[X]) = \\
&= \sum_{i=0}^r \text{tr}(Z_i'MZ_i) \sigma_i^2 (r[X \ Z] - r[X]).
\end{aligned} \tag{56}$$

Considerando-se o modelo misto de classificação dupla cruzada, com interação $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$, definido em (22), e admitindo-se os efeitos de α_i fixos, pode-se escrever:

$$y = \mu \mathbf{1} + X_1 \alpha + Z_1 \beta + Z_2 \gamma + e, \tag{57}$$

onde: $X = [\mathbf{1} \ X_1]$ e $Z = [Z_1 \ Z_2]$.

Nesse contexto, tem-se o conjunto de somas de quadrados, para as estimativas dos componentes de variância, que consta da Tabela 7.

Tabela 7. Somas de quadrados e esperanças dos quadrados médios para o modelo $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

SOMA DE QUADRADOS	E(SQ)
$R(\beta, \gamma \mu, \alpha) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha,)$	$r \left[M \left(Z_1 Z_1' \sigma_\beta^2 + Z_2 Z_2' \sigma_\gamma^2 \right) \right] + \sigma_e^2 [r(X \ Z) - r(X)]$
$R(\gamma \mu, \alpha, \beta) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta)$	$\lambda \sigma_\gamma^2 + \sigma_e^2 [r(X \ Z) - r(X)]$
$SQR = y'y - R(\mu, \alpha, \beta, \gamma)$	$(n - r[X \ Z]) \sigma_e^2$

* onde: $\lambda = \text{tr} \left\{ \left(I - [X \ Z_1] [X \ Z_1]^+ \right) Z_2 Z_2' \right\}$,

$[X \ Z_1]^+$ é a inversa de Moore-Penrose para $[X \ Z_1]$.

A partir desse conjunto, obtém-se as estimativas de $\hat{\sigma}_e^2, \hat{\sigma}_\beta^2, \hat{\sigma}_\gamma^2$, e os coeficientes são fornecidos em tabelas próprias a cada modelo, [maiores detalhes são dados em Searle et al. (1992).]

Assim, tem-se:

$$\hat{\sigma}_e^2 = \frac{SQR}{n-s} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=0}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2}{n-s},$$

onde: s representa o número de caselas ocupadas

$$\begin{aligned} \hat{\sigma}_\gamma^2 &= \frac{1}{\lambda} + [R(\gamma|\mu, \alpha, \beta) - (s-a-b+1)\hat{\sigma}_e^2], \\ \hat{\sigma}_\beta^2 &= \frac{1}{h} \left[\sum_{i=1}^a \sum_{j=1}^b n_{ij} \bar{y}_{ij}^2 - \sum_{i=1}^a n_i \bar{y}_{i.}^2 - (s-a)\hat{\sigma}_e^2 \right] - \hat{\sigma}_\gamma^2 \end{aligned} \quad (58)$$

Pode-se também verificar que:

$$E(R(\beta\gamma|\mu\alpha)) = \sigma_\beta^2 \left(n - \sum_i \frac{j}{n_i} \frac{\sum n_{ij}^2}{n_i} \right) + \sigma_\gamma^2 \left(n - \sum_i \frac{j}{n_i} \frac{\sum n_{ij}^2}{n_i} \right) + \sigma_e^2 (s-a), \quad (59)$$

3.2.2.3 Estimadores de máxima verossimilhança

Os estimadores de componentes de variância baseados nos métodos de máxima verossimilhança, ML e REML, são considerados segundo Searle et al. (1992) os mais adequados para dados desbalanceados, pois apresentam algumas propriedades tais como: consistência e normalidade assintótica, supondo-se normalidade dos dados. Ambos apresentam a desvantagem relacionada com as dificuldades de cálculo, que tem sido superada pelo rápido desenvolvimento dos sistemas estatísticos.

Os estimadores ML fornecem as estimativas dos efeitos fixos, entretanto, os estimadores REML não as fornecem. Com estruturas de dados balanceadas, as soluções REML são idênticas às soluções ANOVA, as quais têm propriedades ótimas, de mínima variância e para muitos analistas isso é uma característica suficiente para preferir as estimativas REML sobre as estimativas ML.

3.2.2.3.1 Estimadores de máxima verossimilhança - ML

Admitindo-se as propriedades essenciais do modelo linear $y = X\beta + Zv + e$ já descrito, tal como as definições apresentadas em (2), (17) e (18):

$$y = X\beta + \sum_{i=0}^r Z_i v_i, \quad E(y) = X\beta \quad \text{e} \quad V = \sum_{i=0}^r Z_i Z_i' \sigma_i^2$$

Para o vetor das observações $y \in N(X\beta, V)$, têm-se a função de verossimilhança, dada como:

$$L = L(\beta, V | y) = \frac{e^{-\frac{1}{2}(y-X\beta)' V^{-1}(y-X\beta)}}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}}}, \quad (60)$$

onde $|V|$ é o determinante de V .

As estimativas de ML utilizam como estimadores de β e V os valores de β e V , que maximizem L de (60). Frequentemente, maximiza-se L com respeito à β e à $\sigma^2 = [\sigma_0^2 \sigma_1^2 \dots \sigma_r^2]$, obtendo-se as seguintes equações de ML para os efeitos fixos e para os componentes de variâncias, respectivamente:

$$X'\tilde{V}^{-1}X\tilde{\beta} = X'\tilde{V}^{-1}y$$

$$\text{tr}(V^{-1}Z_iZ_i') = (y - X\tilde{\beta})' \tilde{V}^{-1}Z_iZ_i'\tilde{V}^{-1}(y - X\tilde{\beta}), \quad (61)$$

para $i = 0, 1, \dots, r$ e \tilde{V} , sendo a estimativa de máxima verossimilhança de V , isto é, $\tilde{V} = \sum_{i=0}^r Z_iZ_i'\tilde{\sigma}_i^2$.

As soluções de ML são obtidas por processos iterativos, para tanto, freqüentemente representa-se:

$$P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}, \quad (62)$$

e define-se: \tilde{P} de P , substituindo-se V por \tilde{V} , conseqüentemente, tem-se: $\tilde{V}^{-1}(y - X\tilde{\beta}) = \tilde{P}y$, onde \tilde{P} é projetor ortogonal. Assim, tem-se, as $r+1$ equações escritas por:

$$\left\{ \text{tr}(\tilde{V}^{-1}Z_iZ_i'\tilde{V}^{-1}Z_jZ_j') \right\} \{ \tilde{\sigma}_i^2 \} = \{ y'\tilde{P}Z_iZ_i'\tilde{P}y \} \quad (63)$$

Assim o estimador de ML para $X\beta$ é

$$MLE(X\beta) = X(X'\tilde{V}^{-1}X)^{-1}X'\tilde{V}^{-1}y, \quad (64)$$

com matriz de dispersão assintótica dada por:

$$\text{Var}[MLE(X\beta)] = X(X'\tilde{V}^{-1}X)^{-1}X' \quad (65)$$

As soluções ML são iterativas e por construção fornecem estimativas não-negativas para componentes de variância, ou seja quando um componente de variância é negativo atribuí-se um valor zero a esse componente. Isto já não ocorre com os métodos de momentos. Entretanto, são viesadas, pois não consideram a perda dos graus de liberdade resultante da estimação dos efeitos fixos. Searle (1987).

3.2.2.3.2 Estimadores de máxima verossimilhança restrita - REML

As estimativas de máxima verossimilhança restrita (REML) foram inicialmente desenvolvidas para certas situações de dados balanceados por Anderson & Bancroft (1952), e posteriormente estendidas para o modelo misto geral por Patterson & Thompson (1971).

Os estimadores de máxima verossimilhança restrita REML maximizam a função de verossimilhança de um vetor de combinações lineares das observações que são invariantes a $X\beta$.

Seja Ly o vetor de combinações lineares. Tem-se, então: $Ly = LX\beta + LZv + Le$, invariante a $X\beta$, se e somente se, $LX = 0$; mas $LX = 0$, somente se $L = TM$, para $M = I - X(X'X)^{-1}X'$, para algum T , onde L e T têm posto linha completo.

As equações REML podem ser obtidas das equações de ML em (64), substituindo-se \tilde{P} por \tilde{V} , ou seja:

$$\left\{ \text{tr}(\tilde{V}^{-1} Z_i Z_i') \right\} = \left\{ y' \tilde{P} Z_i Z_i' \tilde{P} y \right\}, \text{ para } i = 0, 1, \dots, r, \quad (66)$$

A maximização da função densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o

viés resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo.

3.2.2.4 Estimadores MINQUE e MIVQUE

Descreve-se aqui, resumidamente, os estimadores MINQUE, MIVQUE, MIVQUE(O) e I-MINQUE, os quais foram desenvolvidos com o objetivo de satisfazer alguns critérios como norma mínima (ou variância mínima) frequentemente são estimadores imparciais os quais são geralmente funções quadráticas das observações.

De modo geral, para o desenvolvimento dessas metodologias, admite-se que a estimativa de um componente de variância é obtida por uma combinação linear de componentes, digamos $p'\sigma^2$, onde p representa algum vetor conhecido e impõe-se como estimador de $p'\sigma^2$ uma forma quadrática $y'Ay$, tal que para o modelo misto definido em (2), (17) e (18), tem-se:

$$E(y'Ay) = tr(AV) + \beta'X'AX\beta$$

$$p'\sigma^2 = \sum_{i=0}^r tr(AZ_iZ_i')\sigma_i^2 + \beta'X'AX\beta \quad (67)$$

Se a igualdade (67) é verdadeira, então,

$$p_i = tr(AZ_iZ_i'), \quad X'AX = 0 \quad (68)$$

3.2.2.4.1 Estimadores quadráticos não-viesados de norma mínima - MINQUE

Em uma série de artigos Rao (1970,1971a,1971b,1972), sugere um procedimento para obtenção das estimativas de componentes de variâncias, que não requer a suposição de normalidade. Basta que se

minimize uma norma, a norma euclidiana (forma quadrática das observações) similar a uma variância generalizada.

Esse procedimento consiste em estimar $p' \sigma^2$ através de $y' Ay$, com $A = A'$, $AX = 0$ e $p_i = \text{tr}(AZ_i Z_i')$, quando $i = 0, \dots, r$ minimizando-se uma norma conhecida, a norma euclidiana. Efetivamente consiste em utilizar-se de alguns pesos pré-fixados w , da mesma ordem de σ^2 , e, substituir-se V por V_w a matriz com os pesos w em lugar de σ^2 , do mesmo modo substituir-se P de (62) por P_w . Assim, tem-se as seguintes equações:

$$\{\text{tr}[P_w Z_i Z_i' P_w Z_j Z_j']\} \hat{\sigma}^2 = \{y' P_w Z_i Z_i' P_w y\}, \quad (69)$$

tal que, para dados desbalanceados, diferentes valores de w são usados em P_w .

3.2.2.4.2 MINQUE iterativo: I-MINQUE

Esse procedimento consiste em utilizar-se das solução de (68), como um novo vetor de pesos w , a partir do qual um novo conjunto de equações pode ser estabelecido, produzindo um novo vetor de pesos. Esse processo contínuo é conhecido por MINQUE ITERATIVO.

3.2.2.4.3 MINQUE(0)

Trata-se de um caso particular da forma MINQUE, quando se utiliza a matriz identidade como peso. Nesse caso as equações reduzem-se a:

$$\{\text{tr}[MZ_i Z_i' MZ_j Z_j']\} \hat{\sigma}^2 = \{y' MZ_i Z_i' M y\}, \quad (70)$$

para $M = I - XX^+$.

3.2.2.4.4 Estimadores imparciais de variância quadrática mínima MIVQUE

Supondo-se que y tem distribuição normal, a solução MINQUE tem a propriedade de ser uma forma quadrática imparcial das observações com variância mínima, ou seja, é um estimador quadrático imparcial de variância mínima, conhecido por MIVQUE.

Em concordância com os pesos adotados, tem-se, MIVQUE(A), quando utiliza-se as estimativas ANOVA, e MIVQUE(0), quando se utiliza a matriz identidade, como matriz de variâncias e covariâncias.

3.2.3 Predição de variáveis aleatórias

Dado o modelo linear misto, definido em (1), pode-se ter interesse pelas questões:

- Para um vetor de valores observados y , como se pode “predizer” os valores de v ?

- Dentre todas as possíveis realizações de v , quais poderiam estar associadas com um particular valor de y que foi observado?

- Qual é o estimador da média ou qual é o valor esperado dos valores de v ?

Em outras palavras, “Qual é o estimador da média condicional, $E(v|y)$?

O interesse em $E(v|y)$ pode também ser motivado mais formalmente, pois, v e y são (realizações de) variáveis aleatórias, e qualquer que seja a

distribuição conjunta, pode-se mostrar que o preditor de v , que minimiza a soma de quadrados dos erros de “predição” é $E(v|y)$. Assim $E(v|y)$ é chamado o melhor preditor de v . Sob normalidade, tem-se:

$$\tilde{v} = E(v) + Cov(v, y')(Var(y))^{-1}(y - E(y)). \quad (71)$$

Assim \tilde{v} é considerado o melhor preditor linear de v sob normalidade.

E ainda, para o modelo linear misto dado em (1), admitindo-se:

$$\begin{aligned} Var(v) &= D(\sigma_i^2 I_{q_i}), \text{ ou simplesmente } Var(v) = D, \text{ onde: } D = (\sigma_i^2 I_{q_i}) \\ Cov(v, y') &= Cov(v, v'Z') = DZ' \\ Var(y) &= ZDZ' + \sigma_e^2 I = V, \end{aligned} \quad (72)$$

de (71) tem-se:

$$\tilde{v} = DZ'V^{-1}(y - X\beta^0), \text{ portanto uma função estimável de } X\beta, \text{ e}$$

assim tem-se:

$$BLUE(\tilde{v}) = DZ'V^{-1}(y - X\beta^0) \quad (73)$$

$$BLUE(X\beta) = X\beta^0 = X(X'V^{-1}X)^{-1}X'V^{-1}y, \quad (74)$$

onde: $X\beta^0$ é o BLUE de $X\beta$.

As propriedades e ilustrações do método de predição para modelos mistos, são apresentados em detalhes por Henderson(1984). Searle (1987) apresenta através do método dos mínimos quadrados generalizados a seguinte expressão

$$BLUE(X\beta) = X'(X'V^{-1}X)^{-1}X'V^{-1}y.$$

Combinando-se os conceitos de predição e do BLUE em (73) e (74), tem-se o que é freqüentemente referido como (especialmente na literatura de

genética animal) o melhor preditor linear imparcial: BLUP (Best Linear Unbiased Predictor). Assim, de (27), tem-se:

$$BLUP(v) = DZ'V^{-1}(y - X\beta^0) \quad (75)$$

Quando D e σ^2 são conhecidos, ou seja, quando $\sigma_1^2, \dots, \sigma_r^2$ e σ_e^2 , são conhecidos, o cálculo de (75) não apresenta dificuldades, a não ser problemas de inversão de matrizes de grande dimensão, como no caso da matriz V , pois sua ordem é igual ao número de observações. No entanto, é possível reduzir essa dificuldade usando-se “as equações do modelo misto”. Para a obtenção de tais equações assume-se temporariamente que os efeitos que v representa são fixos, assim, tem-se: $Var(y) = I\sigma_e^2$, e pode-se escrever,

$$E(Y) = \begin{bmatrix} X & Z \end{bmatrix} \begin{bmatrix} \beta \\ v \end{bmatrix}, \text{ onde as equações de mínimos quadrados ordinários para}$$

β e v , são dadas por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

e, adicionando-se $\sigma_e^2 D^{-1}$ à $Z'Z$, tem-se:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \sigma_e^2 D^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (76)$$

As equações de (76) são conhecidas como “as equações do modelo misto”, e apresentam a vantagem de requerer a inversão de uma matriz de ordem $p+m$, onde p e m são os números de parâmetros fixos e aleatórios, respectivamente, e $p+m$ freqüentemente é bem menor do que n , o número de observações, isto é se a matriz V , tem ordem n e a matriz D tem ordem $p+m$ bem menor do que n , Searle (1987).

Henderson et al. (1959) mostram que β e v de (76) são de fato o $BLUP(v)$ definido em (75). Para isso definem V matriz de variâncias e covariâncias de y , como:

$$V = ZDZ' + \sigma_e^2 I \quad \text{e} \quad V^{-1} = I - Z(Z'Z + D^{-1})^{-1}Z'.$$

Para o desenvolvimento precedente, assume-se que V é conhecida. Quando não é esse o caso, deve-se estimar as variâncias utilizando-se um dos métodos disponíveis na literatura. Assim em (74), substituindo-se V por \hat{V} , tem-se:

$$X\beta_\nu^0 = X(X'\hat{V}^{-1}X)^s X\hat{V}^{-1}y$$

e de (75), tem-se:

$$BLUP(v)_\nu = \hat{D}Z'\hat{V}^{-1}(y - X\beta_\nu^0) \quad (77)$$

onde, \hat{D} é D (75) com $\hat{\sigma}_i^2$ substituindo σ_i^2 , para $i = 0, \dots, r$.

Um aspecto interessante das “*equações do modelo misto*” é que elas podem ser usadas em procedimentos iterativos para os cálculos das estimativas de máxima verossimilhança (ML) e de máxima verossimilhança restrita (REML) dos componentes de variância. Maiores detalhes do relacionamento entre essas equações e as estimativas ML e REML são apresentados em Harville (1977) e Searle et al. (1992).

4 SISTEMAS COMPUTACIONAIS ESTATÍSTICOS

4.1 Introdução

A metodologia exposta no capítulo anterior desse estudo revela a necessidade da implementação da análise numérica pelos sistemas computacionais estatísticos, ou seja, dos métodos numéricos, dos métodos de inversão de matriz de grande ordem. Sem esses recursos, várias das teorias expostas teriam a aplicação prática limitada e até mesmo impossibilitada. A contribuição dos sistemas estatísticos computacionais deve ser, portanto, caracterizada não só pela agilidade de cálculos mas também por outros fatores, como resultados emitidos corretamente e adequação dos manuais.

Apresentam-se a seguir, resumidamente, as características básicas dos diferentes sistemas estatísticos utilizados nesse estudo. A razão da escolha desses sistemas é justificada pelos seguintes fatos: i) por serem os mais utilizados universalmente em análise de dados, ii) pela disponibilidade de acesso à documentação existente, iii) pelo conhecimento prévio do sistema, iv) por serem disponíveis.

Observa-se que, em termos de textos didáticos, apenas os mais recentes da área de modelos lineares e temas correlatos têm a preocupação de descrever e discutir as propriedades dos sistemas estatísticos. Dentre eles podem ser citados: Milliken & Johnson (1984); Searle (1987) e Searle et al. (1992).

4.2 SAS - Statistical Analysis System (1992)

O sistema estatístico SAS apresenta os conjuntos de procedimentos GLM, VARCOMP e MIXED para a análise de dados.

4.2.1 Proc GLM - General Linear Models

O Proc GLM é o procedimento básico para a análise de dados não balanceados, permite especificar modelos com fatores cruzados ou hierarquizados, de efeitos fixos ou mistos fornece o ajuste de mínimos quadrados para o modelo linear geral de efeitos fixos, dado por:

$$y = X\beta + \varepsilon,$$

onde:

y : é o vetor $n \times 1$ dos dados observados;

X : é uma matriz $n \times p$ conhecida de delineamento, de posto r , tal que

$$r \leq p \leq n;$$

β : é o vetor $p \times 1$ de parâmetros dos efeitos fixos desconhecidos;

ε : é o vetor $n \times 1$ de erros aleatórios.

O GLM realiza a análise padrão assumindo que os componentes de ε são independentes, onde $\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2 I)$ e fornece as estimativas de mínimos quadrados ordinários, $\hat{\beta} = (X'X)^{g_2} X'y$, onde $(X'X)^{g_2}$ representa uma inversa generalizada g_2 , Searle (1971, 1987).

Para estudos da estimabilidade de funções e da testabilidade de hipóteses, o Proc GLM fornece a forma geral das funções estimáveis $L'\beta$, as funções estimáveis dos Tipos I à IV, assim como as somas de quadrados para cada efeito especificado no modelo, emite as somas de quadrados dos Tipos I, II, III e IV, definidas em 3.2.1.2. Admitindo-se o modelo com dois fatores cruzados e interação esse procedimento fornece, para o fatores A, B e A*B, respectivamente, as somas de quadrados:

-do Tipo I (sequencial)

$$\begin{aligned} R(\alpha|\mu) \\ R(\beta|\mu, \alpha) \\ R(\gamma|\mu, \alpha, \beta) \end{aligned}$$

O método Tipo I do Proc GLM/SAS é similar aos método ajuste de constantes, método de Henderson 3, ou seja, utiliza as somas de quadrados sugeridas pelo método de ordenação a priori de Overall e Spiegel (Speed et al., 1978).

-do Tipo II (ajustada);

$$\begin{aligned} R(\alpha|\mu, \beta) \\ R(\beta|\mu, \alpha) \\ R(\gamma|\mu, \alpha, \beta) \end{aligned}$$

Esse método é um dos mais comuns procedimentos baseados no método ajuste de constantes no qual as somas de quadrados ajustadas, isto é, $R(\alpha|\mu, \beta)$ e $R(\beta|\mu, \alpha)$ são usadas no teste para os efeitos principais. As somas de quadrados ajustadas são sugeridas pelo método delineamento experimental de Overall e Spiegel (Speed et al., 1978).

-do Tipo III (restrição- Σ);

$$\begin{aligned} R(\dot{\alpha}|\dot{\mu}, \dot{\beta}, \dot{\gamma}) \\ R(\dot{\beta}|\dot{\mu}, \dot{\alpha}, \dot{\gamma}) \\ R(\dot{\gamma}|\dot{\mu}, \dot{\alpha}, \dot{\beta}) \end{aligned}$$

e do Tipo IV (Tipo contraste).

As somas de quadrados dos Tipos III e IV são equivalentes ao método de médias de quadrados ponderados ou o método mínimos quadrados completos de Overall e Spiegel (Speed et al., 1978).

No tocante ao ajuste do modelo linear misto geral, o Proc GLM permite especificar os efeitos aleatórios envolvidos no modelo e, através do comando “**Random**” fornece as expressões das E(QM) para cada efeito do modelo. Emite ainda, automaticamente, através da opção “**Test**”, os testes de hipóteses para a análise de variância do modelo misto, para cada efeito, através das expressões dos erros obtidas das E(QM), para os quatro tipos de somas de quadrados.

4.2.2 Proc VARCOMP - Variance Components

O Proc VARCOMP fornece as estimativas ANOVA, ML, REML e MIVQUE(0) para os componentes de variância dos modelos de efeitos aleatórios ou de efeitos fixos.

As estimativas ANOVA ou Tipo I obtidas através do Proc VARCOMP são baseadas no método 3 de Henderson, descrito em 3.2.2.2.3. As estimativas REML e ML, utilizam a transformação **W**, descrita em Hartley & Rao (1967) e apresentada em 2.2. O processo iterativo inicia-se com as estimativas MIVQUE(0).

4.2.3 Proc MIXED - Mixed Models

O Proc MIXED estende o ajuste do modelo linear geral, pois permite uma especificação mais geral da matriz de variâncias e covariâncias. Adota o modelo linear geral,

$$y = X\beta + Zv + \varepsilon ,$$

onde:

y, X, β , são, como definidos em 4.2.1;

Z : é uma matriz conhecida de delineamento, que representa a incidência dos efeitos aleatórios;

v : é um vetor de parâmetros desconhecidos dos efeitos aleatórios;

ε : é um vetor de erros,

onde:

Para a análise padrão o procedimento MIXED assume que:

$$E \begin{bmatrix} v \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ e } Var \begin{bmatrix} v \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}, \text{ de forma que a variância de } y \text{ é dada}$$

por, $V = ZGZ' + R$, tal que, $R = \sigma^2 I$. Nesse contexto permite, através da matriz R , especificar estruturas de variâncias e covariâncias mais complexas.

O Proc MIXED ajusta o modelo misto geral através do método de mínimos quadrados ponderados, o qual fornece o valor de β que minimiza $(y - X\beta)' V^{-1} (y - X\beta)$. Segundo Harville (1977) e Searle et al. (1992), em muitas situações nas quais a matriz V^{-1} é desconhecida, a melhor aproximação para as estimativas de mínimos quadrados ponderados é obtida quando se utiliza a matriz de estimativas dos componentes de variâncias, obtida através dos métodos baseados na máxima verossimilhança: ML ou REML.

O Proc MIXED apresenta três métodos para as estimativas dos componentes de variância: o método MIVQUE(0), o método ML, e o método REML, sendo os dois últimos baseados na verossimilhança. Tal procedimento consiste em construir a função objetivo, isto é a função completa de ML ou a função restrita de REML e maximizar a função sob todos os parâmetros desconhecidos. Esse procedimento utiliza uma modificação do método iterativo

de Newton-Raphson para obter numericamente \hat{G} e \hat{R} . Apresenta as soluções para as equações do modelo misto dadas em Henderson (1984).

Para o estudo da estimabilidade de funções e da testabilidade das hipóteses, o procedimento MIXED constroe as funções estimáveis da forma $L' \begin{pmatrix} \beta \\ v \end{pmatrix}$, e apresenta para os efeitos fixos especificados no modelo, apenas, os testes de hipóteses do Tipo III.

4.3 BMDP - Biomedical Computer Programs (1985)

O BMDP apresenta um conjunto de programas destinados para as análises que envolvem metodologias estatísticas simples e avançadas. Dentre os programas elaborados para a análise de variância e covariância encontra-se o programa 3V, o qual é destinado para a análise do modelo linear misto geral, como descrito em Hartley & Rao (1967).

O BMDP-3V é citado na literatura como o primeiro programa desenvolvido especificamente para a análise de variância do modelo misto geral desbalanceado. Esse programa ajusta o modelo linear misto através do método da máxima verossimilhança, fornece as estimativas ML e REML dos componentes de variância, as estimativas e os desvios padrões dos parâmetros dos efeitos fixos. Utiliza como procedimento iterativo uma combinação do método score de Fisher e o algoritmo de Newton Raphson. Apresenta para as estimativas ML o teste da razão de verossimilhança para todos os efeitos envolvidos no modelo linear.

4.4 HARVEY - Mixed Model Least-Squares and Maximum Likelihood Computer Program (1990)

O Harvey é um sistema que foi desenvolvido para efetuar análises estatísticas de dados da área de genética. Ele apresenta dois conjuntos de modelos que podem ser usados para ajustar o modelo linear misto geral. Os modelos definidos como modelos de 1 a 7, são executados através do procedimento LSMLMW (Mixed Model Least-Squares), o qual permite analisar modelos de efeitos fixos, aleatórios ou mistos.

Os modelos 8 e 9 são executados através do procedimento MIXMDL (Mixed Model Maximum Likelihood), que utiliza as equações do modelo misto, descritas em Henderson (1984) e fornece estimativas MINQUE para os componentes de variância.

O procedimento LSMLMW, através dos modelos de 1 a 7, fornece a análise do modelo misto através do método 3 de Henderson. Descreve-se a seguir os modelos 1 e 6 que são utilizados nesse estudo.

$$\text{Modelo 1: } y_{ij} = \mu + F_i + e_{ij},$$

onde :

F_i : representa todos os efeitos fixos do modelo, além da média.

e_{ij} : representa o erro aleatório

A variância do erro, σ^2 pode ser representada pela matriz $I\sigma_e^2$ (mínimos quadrados ordinários). Assim o ajuste para modelo em estudo é obtido através da especificação: MODEL 1 Y = A B A*B/ varcomp = (B, A*B). O comando

varcomp = (B, A*B) define os efeitos aleatórios e fornece as somas de quadrados equivalentes as do Tipo III do Proc GLM/SAS, para os dois casos de desbalanceamentos considerados.

$$\text{Modelo 6 : } y_{ijk} = \mu + a_i + B_j + (aB)_{ij} + F_k + e_{ijk} ,$$

onde:

y_{ijk} : é a observação referente ao i -ésimo nível de a , j -ésimo nível de B e

k -ésimo nível de K ;

a_i : é um conjunto de efeitos aleatórios de classificação cruzada;

B_j : é um conjunto de efeitos fixos;

$(aB)_{ij}$: é um conjunto de interações entre os efeitos aleatórios e fixos;

F_k : representa o conjunto de outros efeitos fixos incluídos no modelo.

Não são consideradas pelo modelo interações entre a_i e F_k . O ajuste dos dados através do sistema HARVEY-modelo 6 é efetuado através do comando MODEL 6 Y = B A B*A. A saída fornece as somas de quadrados correspondentes as do Tipo II do Proc GLM.

4.5 MINITAB - Minitab Statistical Software (1986)

Entre os módulos que compõem o sistema estatístico MINITAB, o módulo GLM é específico para ajustar o modelo linear geral, envolvendo delineamentos com estrutura balanceada ou desbalanceada. Esse módulo permite especificar modelos com fatores cruzados ou hierarquizados, de efeitos fixos ou mistos, e permite, também, através do comando Test especificar o

termo do erro, para testar os efeitos fixos envolvidos no modelo misto de dois fatores. Os testes são efetuados através de somas de quadrados ajustadas.

A execução do módulo GLM, através dos comandos:

MTB > GLM 'Y' = a b a*b;

SUBC> Test A / a*b,

fornece automaticamente as somas de quadrados seqüenciais, isto é Tipo I e as somas de quadrados parciais, ou seja, Tipo III as quais equivalem respectivamente as somas de quadrados do Tipo I e do Tipo III do Proc GLM/SAS.

4.6 NTIA - Núcleo Tecnológico de Informática na Agricultura (1995)

O sistema NTIA em desenvolvimento no Brasil, por técnicos da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) reúne um conjunto de módulos para a análise estatística de dados. Dentre os diversos módulos, o procedimento MODLIN utiliza o método de mínimos quadrados para ajustar modelos lineares do tipo:

$$y = X\beta + e,$$

onde:

y , X , β e e são como definidos em 4.2.1

O conjunto de comandos para o procedimento MODLIN permite especificar o modelo linear a ser ajustado aos dados e o termo do erro adequado aos testes dos efeitos fixos declarados no modelo. Permite, ainda, elaborar hipóteses sobre os parâmetros e fornece automaticamente as somas de quadrados seqüenciais e parciais, as quais correspondem as somas de quadrados do Tipo I e do Tipo III do Proc GLM/SAS.

4.7 REML - Residual Maximum Likelihood (1989, versão 96K)

O REML é um programa para determinar as estimativas dos componentes de variância, no caso de dados desbalanceados. Esse programa fornece as estimativas REML, também conhecidas como estimativas de máxima verossimilhança residual, ou ainda como máxima verossimilhança modificada ou restrita. A descrição detalhada dessa metodologia é dada por Patterson & Thompson (1971) e Harville (1977).

A utilização desse procedimento fornece, basicamente:

- estimativas eficientes dos parâmetros dos efeitos fixos do modelo.
- estimativas dos componentes de variância.
- predição de efeitos aleatórios e da variável resposta.

4.8 SANEST- Sistema de Análise Estatística para Microcomputadores (1986, versão 2.1)

Sistema computacional estatístico em desenvolvimento no Brasil, por pesquisadores da Universidade Federal de Pelotas, construído em linguagem BASIC e destinado à análise de dados. A sua versão 2.1 compõe-se de oito módulos. O módulo 2 especifica a análise de variância e permite executar análise de dados balanceados de classificações simples ou múltiplas, cruzadas ou hierárquicas, envolvendo até cinco fatores. O módulo 3 estabelece a análise de variância de dados de classificação dupla não-balanceada, para o modelo de efeitos fixos ou mistos, com ou sem interação. O método de estimação utilizado é o do ajustamento de constantes. Considerando-se um modelo misto com dois fatores cruzados e interação, através da saída desse sistema pode-se

verificar que as somas de quadrados para os fatores A sendo fixo e B aleatório apresentam-se segundo as opções:

Na ordem A, B (modelo com interação)

B não ajustado e B ajustado, o fator A é ajustado automaticamente, ao passo que o fator B, faz-se a opção pelo ajuste ou não e correspondem as somas de quadrados dos Tipos I e II do sistema Proc GLM/SAS.

4.9 SPSS- Statistical Package for Social Science (1994, versão 6.0)

O sistema computacional estatístico SPSS é um sistema integrado de programas computacionais construído inicialmente para analisar dados das ciências sociais.

O modelo linear univariado ou multivariado é ajustado através da caixa de diálogo MANOVA, que fornece a análise de variância para estrutura com n-classificações, através do método de mínimos quadrados e permite ajustar modelos que envolvam até cinco fatores e cinco covariáveis, para dados balanceados ou desbalanceados. Esse procedimento não fornece comandos para especificar se os efeitos são fixos ou aleatórios. No caso de dados desbalanceados com todas as caselas ocupadas fornece as somas de quadrados através de modelos do tipo Σ , que equivalem as somas de quadrados do Tipo III do Proc GLM/SAS.

4.10 STATGRAPHICS - PLUS (1992, versão 6)

O sistema computacional STATGRAPHICS-PLUS reúne um conjunto de procedimentos estatísticos em telas de comandos, que permitem executar a análise de dados. Através da tela de comando ANOVA-Multifator o sistema emite as estimativas de mínimos quadrados, para os modelos lineares

univariados, aos quais é permitido o uso de até dez fatores e três covariáveis para estruturas de dados balanceados e desbalanceados. Para o caso balanceado emite as somas de quadrados do Tipo I e do Tipo III, já para o caso desbalanceado apenas as somas de quadrados do Tipo III está disponível.

O procedimento ANOVA-Multifator apresenta opções nas telas de comando que permitem especificar o termo do erro adequado para os testes de hipóteses.

4.11 STATISTICA (1994)

O sistema estatístico computacional STATISTICA reúne um conjunto de telas de procedimentos para a análise de dados. Para o ajuste do modelo linear geral, apresenta o módulo geral ANOVA/MANOVA, para a análise de variância univariada e multivariada. O módulo ANOVA utiliza, por padrão, o modelo de médias, para ajustar o modelo linear univariado. Para as estruturas de dados com desbalanceamento completo, fornece as somas de quadrados obtidas através do método de Regressão as quais apresentam as mesmas descrições $R()$, das somas de quadrados do Tipo III do SAS, segundo descrevem Speed et al. (1978). Para desbalanceamento incompleto, ou seja com algumas caselas vazias, fornece hipóteses do tipo contraste, de difícil interpretação e que podem não ser de interesse do pesquisador.

5 EXEMPLO ILUSTRATIVO - Ajuste do modelo misto de classificação dupla

Apresenta-se, inicialmente, um estudo descritivo do ajuste do modelo linear misto, visando evidenciar as somas de quadrados e as hipóteses que estão sendo testadas na análise de variância, através dos procedimentos GLM e MIXED do sistema SAS e posteriormente compara-se as saídas emitidas pelos sistemas computacionais estatísticos: BMDP, HARVEY, MINITAB, NTIA, REML, SANEST, SAS, SPSS, STATGRAPHICS e STATISTICA. Para a análise do modelo misto, estabelece-se um estudo descritivo sobre as hipóteses que estão sendo testadas na análise de variância e as somas de quadrados associadas.

Para esses propósitos considera-se o modelo misto clássico, com dois fatores (A fixo e B aleatório) de classificação cruzada, com interação e dados desbalanceados, conforme (22). Para cada procedimento dos sistemas estatísticos citados, adotou-se um conjunto de sintaxe básica para o ajuste do modelo misto.

Com o objetivo de ilustrar esse desenvolvimento foram considerados os dois conjuntos de dados citados em Lemma (1995a), conforme apresentam as Tabelas 8 e 9.

Tabela 8. Experimento com dois fatores e dados desbalanceados com todas caselas ocupadas

Fator A	Fator B			
	j = 1	j = 2	j = 3	
i = 1	43	41	42	
	45	42	44	
	46	44	46	
	53	-	-	
	187(4)	127(3)	132(3)	446(10)
i = 2	40	35	37	
	40	37	39	
	40	40	40	
	43	40	40	
	163(4)	152(4)	156(4)	471(12)
	350(8)	279(7)	288(7)	917(22)

Fonte: IEMMA (1995a)

Tabela 9. Experimento com dois fatores e dados desbalanceados com casela vazia

Fator A	Fator B			
	j = 1	j = 2	j = 3	
i = 1	43	41	42	
	45	-	44	
	88(2)	41(1)	86(2)	215(5)
i = 2	40	35	-	
	40	37	-	
	-	33	-	
	80(2)	105(3)		185(5)
	168(4)	146(4)	86(2)	400(10)

Fonte: IEMMA (1995a)

A seguir apresentam-se os ajustes dos dados referentes as Tabelas 8 e 9, fornecidos pelos procedimentos GLM, VARCOMP e MIXED, do SAS.

5.1 Ajuste do modelo misto de classificação dupla, através do SAS

Apresenta-se a seguir o ajuste do modelo linear através dos procedimentos GLM, VARCOMP e MIXED do sistema computacional estatístico SAS.

5.1.1 Procedimento GLM

O modelo linear ajustado pelo Proc GLM é o modelo linear usual de efeitos fixos descrito em 4.2.1. Assim, para os dados da Tabela 8, tem-se:

1) O sistema de equações lineares, correspondente ao modelo-S:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{131} \\ y_{132} \\ y_{133} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \\ y_{231} \\ y_{232} \\ y_{233} \\ y_{234} \end{bmatrix} = \begin{bmatrix} 43 \\ 45 \\ 46 \\ 53 \\ 41 \\ 42 \\ 44 \\ 42 \\ 44 \\ 46 \\ 40 \\ 40 \\ 40 \\ 43 \\ 35 \\ 37 \\ 40 \\ 40 \\ 37 \\ 39 \\ 40 \\ 40 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{23} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ e_{114} \\ e_{121} \\ e_{122} \\ e_{123} \\ e_{131} \\ e_{132} \\ e_{133} \\ e_{211} \\ e_{212} \\ e_{213} \\ e_{214} \\ e_{221} \\ e_{222} \\ e_{223} \\ e_{224} \\ e_{231} \\ e_{232} \\ e_{233} \\ e_{234} \end{bmatrix}$$

II) As seguintes configurações apresentadas nas Tabelas 10 e 11.

Para o modelo-S:

Tabela 10. Configurações para o modelo-S com todas caselas ocupadas

	j = 1	j = 2	j = 3
i = 1	$\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}(4)$	$\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}(3)$	$\mu + \alpha_1 + \beta_3 + (\alpha\beta)_{13}(3)$
i = 2	$\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}(4)$	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}(4)$	$\mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23}(4)$

Para o modelo-M:

Tabela 11. Configurações para o modelo-M com todas caselas ocupadas

	j = 1	j = 2	j = 3
i = 1	$\mu_{11}(4)$	$\mu_{12}(3)$	$\mu_{13}(3)$
i = 2	$\mu_{21}(4)$	$\mu_{22}(4)$	$\mu_{23}(4)$

Equivalentemente, para os dados da Tabela 9, tem-se:

III) O sistema de equações lineares, correspondente ao modelo-S:

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{223} \end{bmatrix} = \begin{bmatrix} 43 \\ 45 \\ 41 \\ 42 \\ 44 \\ 40 \\ 40 \\ 35 \\ 37 \\ 33 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{131} \\ e_{132} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \\ e_{223} \end{bmatrix}$$

IV) As configurações descritas nas Tabelas 12 e 13.

Para o modelo-S:

Tabela 12. Configurações para o modelo-S com casela vazia

	j = 1	j = 2	j = 3
i = 1	$\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}(2)$	$\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}(1)$	$\mu + \alpha_1 + \beta_3 + (\alpha\beta)_{13}(2)$
i = 2	$\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}(2)$	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}(3)$	

Para o modelo-M:

Tabela 13. Configurações para o modelo-M com casela vazia

	j = 1	j = 2	j = 3
i = 1	$\mu_{11}(2)$	$\mu_{12}(1)$	$\mu_{13}(2)$
i = 2	$\mu_{21}(2)$	$\mu_{22}(3)$	-

O Proc GLM apresenta um conjunto de declarações e opções que permitem obter o ajuste do modelo linear. Neste estudo para ajustar o modelo misto de classificação dupla, foram consideradas apenas as declarações básicas a seguir:

```
proc glm;
class a b
model y = a b a*b / e e1 e2 e3 e4 ss1 ss2 ss3 ss4;
random b a*b / test ;
```

Com as declarações:

i) **class a b**; definem-se as variáveis classificatórias, as quais serão utilizadas na análise.

ii) **model y = a b a*b / e e1 e2 e3 e4 ss1 ss2 ss3 ss4**;

- **a b a*b**, define-se o modelo matemático pelo qual os dados foram obtidos, especificando-se as variáveis dependentes e independentes.

- **e e1 e2 e3 e4 ss1 ss2 ss3 ss4**, são as opções da declaração **model** mais usadas, onde:

- **ss1 ss2 ss3 ss4**: emitem respectivamente as somas de quadrados do Tipo I,II,III,IV, conforme a Tabela 14:

- **e e1 e2 e3 e4**: fornecem respectivamente a forma geral de funções estimáveis e as funções estimáveis do Tipo I, II, III, IV, para cada efeito do modelo e as somas de quadrados correspondentes, considerando-se todos os efeitos como sendo fixos, apesar da opção **Test**.

Tabela 14. Somas de quadrados segundo o Proc GLM/SAS

Causa de Variação	TIPO I (SEQÜENCIAL)	TIPO II (AJUSTADA)	TIPO III (RESTRIÇÃO Σ)	TIPO IV
A	$R(\alpha \mu)$	$R(\alpha \mu, \beta)$	$R(\dot{\alpha} \dot{\mu}, \dot{\beta}, (\alpha\dot{\beta}))$	$SQH_0^{(4)}$
B	$R(\beta \mu, \alpha)$	$R(\beta \mu, \alpha)$	$R(\dot{\beta} \dot{\mu}, \dot{\alpha}, (\alpha\dot{\beta}))$	$SQH_0^{(8)}$
AxB	$R(\alpha\beta \mu, \alpha, \beta)$	$R(\alpha\beta \mu, \alpha, \beta)$	$R(\alpha\dot{\beta} \dot{\mu}, \dot{\alpha}, \dot{\beta})$	$R(\alpha\beta \mu, \alpha, \beta)$

onde:

i) As somas de quadrados do Tipo I são seqüenciais, assim elas são obtidas em conformidade com a ordem dos fatores no modelo. Este tipo de soma de quadrados testa hipóteses sobre médias ponderadas não ajustadas.

ii) As somas de quadrados do Tipo II correspondem a notação $R(\)$ em que cada efeito é ajustado para os demais, exceto para interações. Nem sempre são provenientes de uma partição ortogonal da SQ dos parâmetros.

iii) As somas de quadrados do Tipo III dizem respeito às médias não ponderadas de linhas e / ou colunas.

iv) As somas de quadrados do Tipo IV testam hipóteses que são geradas automaticamente pelo Proc GLM do SAS, e que podem ser usualmente interpretadas através de contrastes entre médias não ponderadas.

Se não existem caselas vazias, são equivalentes as do Tipo III, caso contrário as somas de quadrados dos Tipos III e IV são, em geral, diferentes e referem-se à hipóteses diferentes. Maiores detalhes são encontrados em lemma (1991, 1993a e 1995a) e Santos (1994).

O comando **Random**, declara os fatores aleatórios e fornece os quadrados médios esperados para os efeitos considerados no modelo. A opção **Test** fornece os testes de hipóteses para a análise de variância do modelo misto considerando-se os quadrados médios esperados apropriados como termos do erro. Para o ajuste do modelo misto, são esses os testes que devem ser considerados. Apresenta-se os resultados obtidos para o ajuste do modelo misto considerado, utilizando-se como exemplos os dois conjuntos de dados descritos nas Tabelas 8 e 9.

5.1.1.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia. (Dados da Tabela 8)

Descreve-se a seguir o ajuste do modelo utilizando-se os dados da Tabela 8, ou seja, considera-se aqui o conjunto de dados desbalanceado com todas caselas ocupadas.

5.1.1.1.1 Funções estimáveis e hipóteses testáveis

Nas Tabelas 15 e 15.1 descreve-se a saída do proc GLM relacionada com as funções estimáveis e hipóteses testáveis, segundo a ordenação A, B.

Tabela 15. Forma geral das funções estimáveis obtidas do procedimento GLM/SAS para o caso de todas caselas ocupadas

Efeito	Coefficientes	A*B	
INTERCEPTO	L1		1 1 L7
A	1 L2		1 2 L8
	2 L1-L2		1 3 L2-L7-L8
B	1 L4		2 1 L4-L7
	2 L5		2 2 L5-L8
	3 L1-L4-L5		2 3 L1-L2-L4-L5+L7+L8

Tem-se, então a forma geral das funções estimáveis, para esse modelo,

$$L\beta = L_1\mu + L_2\alpha_1 + (L_1 - L_2)\alpha_2 + L_4\beta_1 + L_5\beta_2 + (L_1 - L_4 - L_5)\beta_3 + \\ + L_7(\alpha\beta)_{11} + L_8(\alpha\beta)_{12} + (L_2 - L_7 - L_8)(\alpha\beta)_{13} + (L_4 - L_7)(\alpha\beta)_{21} + \\ + (L_5 - L_8)(\alpha\beta)_{22} + (L_1 - L_2 - L_4 - L_5 + L_7 + L_8)(\alpha\beta)_{23}$$

Atribuindo-se valores aos coeficientes L 's, obtém-se as formas específicas de funções de interesse.

Tabela 15.1 Forma geral das funções estimáveis segundo o Proc GLM/SAS para os dados sem casela vazia

Efeito	Coefficientes	A*B	
INTERCEPTO	L1		1 1 L7
A	1 L2		1 2 L8
	2 L1-L2		1 3 L2-L7-L8
B	1 L4		2 1 L4-L7
	2 L5		2 2 L5-L8
	3 L1-L4-L5		2 3 L1-L2-L4-L5+L7+L8

Tem-se, então a forma geral das funções estimáveis, para esse modelo,

$$L\beta = L_1\mu + L_2\alpha_1 + (L_1 - L_2)\alpha_2 + L_4\beta_1 + L_5\beta_2 + (L_1 - L_4 - L_5)\beta_3 + \\ + L_7(\alpha\beta)_{11} + L_8(\alpha\beta)_{12} + (L_2 - L_7 - L_8)(\alpha\beta)_{13} + (L_4 - L_7)(\alpha\beta)_{21} + \\ + (L_5 - L_8)(\alpha\beta)_{22} + (L_1 - L_2 - L_4 - L_5 + L_7 + L_8)(\alpha\beta)_{23}$$

Atribuindo-se valores aos coeficientes L 's, obtêm-se as formas específicas de funções de interesse. Descreve-se nas Tabelas 16 a 19 as funções estimáveis e as hipóteses que são testadas, considerando-se o modelo-M.

Tabela 16. Funções estimáveis do Tipo I segundo o Proc GLM/SAS para os dados sem casela vazia

Hipóteses testáveis	
Tipo I Funções Estimáveis para: A	
Efeito	Coefficientes
INTERCEPTO	0
A	1 L2 2 -L2
B	1 0.0667*L2 2 -0.0333*L2 3 -0.0333*L2
A*B	11 0.4*L2 12 0.3*L2 13 0.3*L2 21 -0.3333*L2 22 -0.3333*L2 23 -0.3333*L2
$H_0^{(9)}: \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10} - \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12} = 0$ <p>ou</p> $H_0^{(9)}: \alpha_1 - \alpha_2 + \frac{1}{15}\beta_1 - \frac{1}{30}\beta_2 - \frac{1}{30}\beta_3 + \frac{4}{10}\alpha\beta_{11} + \frac{3}{10}\alpha\beta_{12} + \frac{3}{10}\alpha\beta_{13} - \frac{1}{3}\alpha\beta_{21} - \frac{1}{3}\alpha\beta_{22} - \frac{1}{3}\alpha\beta_{23} = 0$	
Tipo I Funções Estimáveis para: B	
INTERCEPTO	0
A	1 0 2 0
B	1 L4 2 L5 3 -L4-L5
A*B	11 0.4737*L4 12 -0.0226*L4+0.4286*L5 13 -0.4511*L4-0.4286*L5 21 0.5263*L4 22 0.0226*L4+0.5714*L5 23 -0.5489*L4-0.5714*L5
$H_0^{(6)}: \begin{cases} \frac{4\mu_{11} + 4\mu_{21}}{8} = \frac{4\bar{L}_1 + 4\bar{L}_2}{8} \\ \frac{3\mu_{12} + 4\mu_{22}}{7} = \frac{3\bar{L}_1 + 4\bar{L}_2}{7} \end{cases}$ <p>ONDE:</p> $\bar{L}_1 = \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10}, \bar{L}_2 = \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12}$ <p>ou:</p> $H_0^{(6)}: \begin{cases} \frac{3}{10}\mu_{11} - \frac{3}{20}\mu_{12} - \frac{3}{20}\mu_{13} + \frac{1}{3}\mu_{21} - \frac{1}{6}\mu_{22} - \frac{1}{6}\mu_{23} = 0 \\ -\frac{3}{35}\mu_{11} + \frac{3}{10}\mu_{12} - \frac{9}{70}\mu_{13} - \frac{4}{21}\mu_{21} + \frac{8}{21}\mu_{22} + \frac{8}{21}\mu_{23} = 0 \end{cases}$	
Tipo I Funções Estimáveis para: A*B	
INTERCEPTO	0
A	1 0 2 0
B	1 0 2 0 3 0
A*B	11 L7 12 L8 13 -L7-L8 21 -L7 22 -L8 23 L7+L8
$H_0^{(9)}: \mu_{ij} - \mu_{j'} - \mu_{ij'} + \mu_{ij''} = 0, i < i', j < j''$	

Fato:

$$\text{Sendo, no modelo-M, } H_0^{(1)}: \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10} - \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12} = 0 \text{ e}$$

lembrando que $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, por simples substituição em $H_0^{(1)}$ e um pouco de álgebra elementar, facilmente se obtém $H_0^{(1)}$ no modelo-S, como mostra o desenvolvimento algébrico a seguir:

Substituindo-se $\mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$ em $H_0^{(1)}$, tem-se:

$$\begin{array}{r} \frac{4}{10}\mu_{11} = \frac{4}{10}\mu + \frac{4}{10}\alpha_1 + \frac{4}{10}\beta_1 + \frac{4}{10}\alpha\beta_{11} \\ \frac{3}{10}\mu_{12} = \frac{3}{10}\mu + \frac{3}{10}\alpha_1 + \frac{3}{10}\beta_2 + \frac{3}{10}\alpha\beta_{12} \\ \frac{3}{10}\mu_{13} = \frac{3}{10}\mu + \frac{3}{10}\alpha_1 + \frac{3}{10}\beta_3 + \frac{3}{10}\alpha\beta_{13} \\ \frac{4}{12}\mu_{21} = \frac{4}{12}\mu + \frac{4}{12}\alpha_2 + \frac{4}{12}\beta_1 + \frac{4}{12}\alpha\beta_{21} \\ \frac{4}{12}\mu_{22} = \frac{4}{12}\mu + \frac{4}{12}\alpha_2 + \frac{4}{12}\beta_2 + \frac{4}{12}\alpha\beta_{22} \\ \frac{4}{12}\mu_{23} = \frac{4}{12}\mu + \frac{4}{12}\alpha_2 + \frac{4}{12}\beta_3 + \frac{4}{12}\alpha\beta_{23} \end{array}$$

tem-se:

$$\alpha_1 - \alpha_2 + \frac{1}{15}\beta_1 - \frac{1}{30}\beta_2 - \frac{1}{30}\beta_3 + \frac{4}{10}\alpha\beta_{11} + \frac{3}{10}\alpha\beta_{12} + \frac{3}{10}\alpha\beta_{13} - \frac{1}{3}\alpha\beta_{21} - \frac{1}{3}\alpha\beta_{22} - \frac{1}{3}\alpha\beta_{23}.$$

Portanto,

$$H_0^{(1)}: \alpha_1 - \alpha_2 + \frac{1}{15}\beta_1 - \frac{1}{30}\beta_2 - \frac{1}{30}\beta_3 + \frac{4}{10}\alpha\beta_{11} + \frac{3}{10}\alpha\beta_{12} + \frac{3}{10}\alpha\beta_{13} - \frac{1}{3}\alpha\beta_{21} - \frac{1}{3}\alpha\beta_{22} - \frac{1}{3}\alpha\beta_{23} = 0,$$

ou como pode ser visto na tabela 16, fazendo-se $L_2 = 1$, nos coeficientes das funções estimáveis do Tipo I, para o fator A:

$$H_0^{(1)}: \alpha_1 - \alpha_2 + 0,0667\beta_1 - 0,0333\beta_2 - 0,0333\beta_3 + 0,4\alpha\beta_{11} + 0,3\alpha\beta_{12} + 0,3\alpha\beta_{13} - 0,3333\alpha\beta_{21} - 0,3333\alpha\beta_{22} - 0,3333\alpha\beta_{23} = 0$$

Com raciocínio análogo, pode-se converter todas as demais hipóteses desse estudo, do modelo-M para o modelo-S.

Naturalmente, a apresentação das hipóteses no modelo-M pode simplificar sobremaneira a interpretação por parte dos usuários não iniciados.

Tabela 17. Funções estimáveis do Tipo II segundo o Proc GLM/SAS para os dados sem casela vazia

Funções estimáveis	Hipóteses testáveis
Tipo II Funções Estimáveis para: A	
Efeito	Coefficientes
INTERCEPTO	0
A	1 L2 2 -L2
B	1 0 2 0 3 0
A*B	11 0.3684*L2 12 0.3158*L2 13 0.3158*L2 21 -0.3684*L2 22 -0.3158*L2 23 -0.3158*L2
$H_0^{(2)}: \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10} = \frac{4\bar{C}_1 + 3\bar{C}_2 + 3\bar{C}_3}{10}$ $H_0^{(2)}: \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12} = \frac{4\bar{C}_1 + 4\bar{C}_2 + 4\bar{C}_3}{12}$ <p>onde:</p> $\bar{C}_1 = \frac{4\mu_{11} + 4\mu_{21}}{8}; \bar{C}_2 = \frac{3\mu_{12} + 4\mu_{22}}{7}; \bar{C}_3 = \frac{3\mu_{13} + 4\mu_{23}}{7}$	
Tipo II Funções Estimáveis para: B	
Efeito	Coefficientes
INTERCEPTO	0
A	1 0 2 0
B	1 L4 2 L5 3 -L4-L5
A*B	11 0.4737*L4 12 -0.0226*L4+0.4286*L5 13 -0.4511*L4-0.4286*L5 21 0.5263*L4 22 0.0226*L4+0.5714*L5 23 -0.5489*L4-0.5714*L5
$H_0^{(6)}: \begin{cases} \frac{4\mu_{11} + 4\mu_{21}}{8} = \frac{4\bar{L}_1 + 4\bar{L}_2}{8} \\ \frac{3\mu_{12} + 4\mu_{22}}{7} = \frac{3\bar{L}_1 + 4\bar{L}_2}{7} \end{cases}$ <p>ONDE:</p> $\bar{L}_1 = \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10}; \bar{L}_2 = \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12}$ <p>ou</p> $H_0^{(6)}: \begin{cases} \frac{3}{10}\mu_{11} - \frac{3}{20}\mu_{12} - \frac{3}{20}\mu_{13} + \frac{1}{3}\mu_{21} - \frac{1}{6}\mu_{22} - \frac{1}{6}\mu_{23} = 0 \\ \frac{3}{35}\mu_{11} + \frac{3}{10}\mu_{12} - \frac{9}{70}\mu_{13} - \frac{4}{21}\mu_{21} + \frac{8}{21}\mu_{22} + \frac{8}{21}\mu_{23} = 0 \end{cases}$	
Tipo II Funções Estimáveis para: A*B	
Efeito	Coefficientes
INTERCEPTO	0
A	1 0 2 0
B	1 0 2 0 3 A*B
	11 L7 12 L8 13 -L7-L8 21 -L7 22 -L8 23 L7+L8
$H_0^{(9)}: \mu_{ij} - \mu_{i'j'} - \mu_{rj} + \mu_{rj'} = 0, i < i', j < j'$	

Tabela 18. Funções estimáveis do Tipo III segundo o Proc GLM/SAS para os dados sem casela vazia

Funções estimáveis	Hipóteses testáveis
<p>Tipo III Funções Estimáveis para: A</p>	
<p>Efeito Coeficientes</p> <p>INTERCEPTO 0</p> <p>A 1 L2</p> <p> 2 -L2</p> <p>B 1 0</p> <p> 2 0</p> <p> 3 0</p> <p>A*B 11 0.3333*L2</p> <p> 12 0.3333*L2</p> <p> 13 0.3333*L2</p> <p> 21 -0.3333*L2</p> <p> 22 -0.3333*L2</p> <p> 23 -0.3333*L2</p>	$H_0^{(3)}: \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} = \frac{\mu_{21} + \mu_{22} + \mu_{23}}{3}$
<p>Tipo III Funções Estimáveis para: B</p>	
<p>Efeito Coeficientes</p> <p>INTERCEPTO 0</p> <p>A 1 0</p> <p> 2 0</p> <p>B 1 L4</p> <p> 2 L5</p> <p> 3 -L4-L5</p> <p>A*B 11 0.5*L4</p> <p> 12 0.5*L5</p> <p> 13 -0.5*L4-0.5*L5</p> <p> 21 0.5*L4</p> <p> 22 0.5*L5</p> <p> 23 -0.5*L4-0.5*L5</p>	$H_0^{(7)}: \begin{cases} \frac{\mu_{11} + \mu_{21}}{2} = \frac{\mu_{12} + \mu_{22}}{2} \\ \frac{\mu_{11} + \mu_{21} + \mu_{12} + \mu_{22}}{2} = \mu_{13} + \mu_{23} \end{cases}$
<p>Tipo III Funções Estimáveis para: A*B</p>	
<p>Efeito Coeficientes</p> <p>INTERCEPTO 0</p> <p>A 1 0</p> <p> 2 0</p> <p>B 1 0</p> <p> 2 0</p> <p> 3 0</p> <p>A*B 11 L7</p> <p> 12 L8</p> <p> 13 -L7-L8</p> <p> 21 -L7</p> <p> 22 -L8</p> <p> 23 L7+L8</p>	$H_0^{(9)}: \mu_{ij} - \mu_{i'j'} - \mu_{ij'} + \mu_{i'j} = 0, i < i', j < j'$

Tabela 19. Funções estimáveis do tipo IV segundo o Proc GLM/SAS para os dados sem casela vazia

Funções estimáveis	Hipóteses testáveis
<p>Tipo IV Funções Estimáveis para: A</p> <pre> INTERCEPTO 0 A 1 L2 2 -L2 B 1 0 2 0 3 0 A*B 11 0.3333*L2 12 0.3333*L2 13 0.3333*L2 21 -0.3333*L2 22 -0.3333*L2 23 -0.3333*L2 </pre>	$H_0^{(3)}: \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} = \frac{\mu_{21} + \mu_{22} + \mu_{23}}{3}$
<p>Tipo IV Funções Estimáveis para: B</p> <pre> INTERCEPTO 0 A 1 0 2 0 B 1 L4 2 L5 3 -L4-L5 A*B 11 0.5*L4 12 0.5*L5 13 -0.5*L4-0.5*L5 21 0.5*L4 22 0.5*L5 23 -0.5*L4-0.5*L5 </pre>	$H_0^{(7)}: \begin{cases} \frac{\mu_{11} + \mu_{21}}{2} = \frac{\mu_{12} + \mu_{22}}{2} \\ \frac{\mu_{11} + \mu_{21}}{2} + \frac{\mu_{12} + \mu_{22}}{2} = \mu_{13} + \mu_{23} \end{cases}$
<p>Tipo IV Funções Estimáveis para: A*B</p> <pre> Effect Coeficientes INTERCEPTO 0 A 1 0 2 0 B 1 0 2 0 3 0 A*B 11 L7 12 L8 13 -L7-L8 21 -L7 22 -L8 23 L7+L8 </pre>	$H_0^{(9)}: \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0, i < i', j < j'$

Obs. As hipóteses do tipo III e do tipo IV, neste caso são equivalentes.

5.1.1.1.2 Hipóteses e somas de quadrados associadas

A Tabela 20 apresenta a saída do proc GLM relacionada à análise de variância, emitida pela especificação **model**, a qual admite todos os efeitos como sendo fixo.

Tabela 20. Análise de Variância dos dados sem casela vazia segundo o Proc GLM do SAS através do comando **Model**

Variável Dependente: Y							
Causa de Variação	GL	Soma de Quadrados		Valor F	Pr > F		
		Quadrados	Médios				
Modelo	5	206.606061	41.321212	6.60	0.0016		
Erro	16	100.166667	6.260417				
Total	21	306.772727					
	R-Quadrado	C.V.	Raiz QME	MédiaY			
	0.673482	6.002815	2.50208	41.681			
Variável Dependente: Y							
Causa de	GL	H_0	$R(\)$	SQ Tipo I	Quadrado Médio	Valor F	Pr > F
A	1	$H_0^{(1)}$	$R(\alpha \mu)$	156.122727	156.122727	24.94	0.0001
B	2	$H_0^{(6)}$	$R(\beta \mu, \alpha)$	47.856767	23.928383	3.82	0.0440
A*B	2	$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$	2.626566	1.313283	0.21	0.8130
Causa de	GL	H_0	$R(\)$	SQ Tipo II	Quadrado Médio	Valor F	Pr > F
A	1	$H_0^{(2)}$	$R(\alpha \mu, \beta)$	144.421053	144.421053	23.07	0.0002
B	2	$H_0^{(6)}$	$R(\beta \mu, \alpha)$	47.856767	23.928383	3.82	0.0440
A*B	2	$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$	2.626566	1.313283	0.21	0.8130
Causa de	GL	H_0	$R(\)$	SQ Tipo III	Quadrado Médio	Valor F	Pr > F
A	1	$H_0^{(3)}$	$R(\dot{\alpha} \dot{\mu}, \dot{\beta}, \alpha\dot{\beta})$	141.066667	141.066667	22.53	0.0002
B	2	$H_0^{(7)}$	$R(\dot{\beta} \dot{\mu}, \dot{\alpha}, \alpha\dot{\beta})$	49.077694	24.538847	3.92	0.0412
A*B	2	$H_0^{(9)}$	$R(\alpha\dot{\beta} \dot{\mu}, \alpha, \dot{\beta})$	2.626566	1.313283	0.21	0.8130
Causa de	GL	H_0	$R(\)$	SQ Tipo IV	Quadrado Médio	Valor F	Pr > F
A	1	$H_0^{(4)}$	$R(\dot{\alpha} \dot{\mu}, \dot{\beta}, \alpha\dot{\beta})$	141.066667	141.066667	22.53	0.0002
B	2	$H_0^{(8)}$	$R(\dot{\beta} \dot{\mu}, \dot{\alpha}, \alpha\dot{\beta})$	49.077694	24.538847	3.92	0.0412
A*B	2	$H_0^{(9)}$	$R(\alpha\dot{\beta} \dot{\mu}, \alpha, \dot{\beta})$	2.626566	1.313283	0.21	0.8130

* Os testes apresentados consideram todos os efeitos fixos, de modo que não devem ser considerados apropriados para o modelo misto

Fatos :

1) Verifica-se conforme esperado pois não há caselas vazias, que as somas de quadrados dos tipos III e IV são as mesmas.

2) Ademais, dado que a interação está ajustada para todos os efeitos “anteriores”, as quatro somas de quadrados a ela relacionadas são iguais, isto pode ser verificado também no caso de ocorrência de caselas vazias.

5.1.1.1.3 Testes de hipóteses para a análise de variância no modelo misto

O Proc GLM apresenta, através do comando **Random** com a opção **Test**, os testes de hipóteses, utilizando-se dos quadrados médios esperados para cada efeito declarado no **Model**. Conforme mostram as Tabelas de 21 à 24, estes são os testes considerados apropriados para o modelo misto.

Tabela 21. Testes das hipóteses do Tipo I na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test

Causa de Variação	Esperança do Quadrado Médio Tipo I					
A	Var(Erro) + 3.6727 Var(A*B) + 0.0364 Var(B) + Q(A)					
B	Var(Erro) + 3.691 Var(A*B) + 7.3 Var(B)					
A*B	Var(Erro) + 3.609 Var(A*B)					
Testes de Hipóteses para a Análise de Variância do Modelo Misto						
Variável Dependente: Y						
Causa de Variação: A						
Erro: 0.005*QM(B) + 1.0126*QM(A*B) - 0.0175*QM(Erro)						
	Denominador		Denominador			
GL	QM Tipo I	$H_0^{(1)}$	GL	QM	Valor F	Pr > F
1	156.12272727		2.01	1.3391714664	116.5816	0.0083
Causa de Variação: B						
Erro: 1.0227*QM(A*B) - 0.0227*QM(Erro)						
	Denominador		Denominador			
GL	QM Tipo I	$H_0^{(I-B)}$	GL	QM	Valor F	Pr > F
2	23.928383459		1.60	1.2009420524	19.9247	0.0743
Causa de Variação: A*B						
Erro: QM(Erro)						
	Denominador		Denominador			
GL	QM Tipo I	$H_0^{(I-AB)}$	GL	QM	Valor F	Pr > F
2	1.313283208		16	6.2604166667	0.2098	0.8130

Fatos:

1) Verifica-se na Tabela 21, que a soma de quadrados do Tipo I para o fator A fixo, é a mesma já descrita na Tabela 20. No entanto, os quadrados médios, os graus de liberdade e o valor de F são diferentes. O mesmo fato pode ser verificado em relação as Tabelas 22, 23 e 24. Isso mostra, de certo modo, que a hipótese $H_0^{(1)}$ continua sendo testada no comando **Random**, embora de modo diferente:

2) Para os efeitos aleatórios, isto não se verifica, pois as hipóteses são conceitualmente diferentes. Conforme o modelo descrito em (22), tem-se hipóteses do tipo $H_0: \sigma_i^2 = 0$, de modo que a Tabela 21, apresenta :

$$H_0^{(I-B)}: \sigma_{\beta_1}^2 = \sigma_{\beta_2}^2 = \sigma_{\beta_3}^2, H_0^{(I-AB)}: \sigma_{\alpha\beta_{11}}^2 = \dots = \sigma_{\alpha\beta_{23}}^2.$$

O mesmo ocorre nas Tabelas 22, 23 e 24.

3) Verifica-se que na Tabela 20, tem-se: $F_{OBS} = \frac{QMA}{QMR} = 24,94$, e na

Tabela 21, tem-se: $F_{OBS} = \frac{QMA}{\tilde{Q}} = 116,5816$, onde \tilde{Q} é o denominador corrigido,

conforme a expressões dos valores esperados das somas de quadrados. Assim pode-se observar que a expressão do erro apropriado para o fator A fixo, que consta na Tabela 21 é a combinação linear, como definida em (35).

$\tilde{Q} = a_1 QMB + a_2 QMAB + a_3 QMR = 1,33917$, onde:

$$a_1 = 1 - \frac{0,0364}{7,3} - \frac{3,6727}{3,609} + \frac{3,691 \times 0,0364}{3,609 \times 7,3} = -0,0175,$$

$$a_2 = \frac{3,6727}{3,609} - \frac{3,691 \times 0,0364}{3,609 \times 7,3} = 1,0126 \quad e$$

$$a_3 = \frac{0,0364}{7,3} = 0,005, \text{ são obtidos segundo método de ajuste de constantes}$$

apresentado na Tabela 5.

4) Pode verificar também , que os graus de liberdade são corrigidos pela fórmula de Satterthwaite, definida em (36).

$$GL_{CORRG} = \frac{\tilde{Q}^2}{\frac{(a_1 QMR)^2}{gl_{RES}} + \frac{(a_2 QMAB)^2}{gl_{AB}} + \frac{(a_3 QMB)^2}{gl_B}}$$

$$GL_{CORRG} = \frac{1,33917^2}{\frac{(-0,0175 \times 6,2604)^2}{16} + \frac{(1,0126 \times 1,3133)^2}{2} + \frac{(0,005 \times 23,9284)^2}{2}} = 2,01$$

5) Pode-se facilmente obter as estimativas para os componentes de variância, utilizando-se os resultados da Tabela 21, segundo método dos momentos têm-se:

$$\hat{\sigma}_\beta^2 = \frac{23,9283 - 6,2604 + 5,0596}{7,3} = 3,1134 \quad e$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{1,3113283 - 6,2604}{3,609} = -1,3708$$

Verifica-se que a estimativa para o componente da interação é negativa e segundo Hocking (1985) há diversas situações em que a matriz das estimativas de covariâncias são positivas definidas mas isso não garante que as estimativas dos componentes de variância sejam positivas. Estimativas negativas na prática podem ocorrer, com exceção de σ_e^2 .

Tabela 22. Testes das hipóteses do Tipo II na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção **Test**

Causa de Variação		Esperança do Quadrado Médio Tipo II			
A		Var(Erro) + 3.6391 Var(A*B) + Q(A)			
B		Var(Erro) + 3.691 Var(A*B) + 7.3 Var(B)			
A*B		Var(Erro) + 3.609 Var(A*B)			

Testes de Hipóteses para a Análise de Variância do Modelo Misto

Variável Dependente: Y

Causa de Variação: A

Erro: 1.0083*QM(A*B) - 0.0083*QM(Erro)

		Denominador	Denominador		
GL	QM Tipo II	$H_0^{(2)}$	GL	QM	Valor F Pr > F
1	144.42105263		1.85	1.2720570959	113.5335 0.0114

Causa de Variação: B

Erro: 1.0227*QM(A*B) - 0.0227*QM(Erro)

		Denominador	Denominador		
GL	QM Tipo II	$H_0^{(I-B)}$	GL	QM	Valor F Pr > F
2	23.928383459		1.60	1.2009420524	19.9247 0.0743

Causa de Variação: A*B

Erro: QM(Erro)

		Denominador	Denominador		
GL	QM Tipo II	$H_0^{(I-AB)}$	GL	QM	Valor F Pr > F
2	1.313283208		16	6.2604166667	0.2098 0.8130

Fato:

1) Verifica-se na Tabela 22, que a soma de quadrados do Tipo II para o fator A fixo é a mesma já descrita na Tabela 20. No entanto os quadrados médios, os graus de liberdade e o valor de F, são diferentes e determinados segundo as metodologias definidas em (35) e (36) as quais são ilustradas nos fatos da Tabela 22 o que indica que a Hipótese $H_0^{(2)}$, continua sendo testada pelo comando **Random**, embora como já visto, de modo diferente.

2) As estimativas dos componentes de variância podem ser obtidos das expressões das esperanças quadrados médios, as quais são fornecidas segundo a formulação 1 conforme Tabela 3, ou modelo III conforme Tabela 1.

Tabela 23. Testes das hipóteses do Tipo III. na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test

Causa de Variação		Esperança do Quadrado Médio Tipo III			
A	Var(Erro) + 3.6 Var(A*B) + Q(A)				
B	Var(Erro) + 3.609 Var(A*B) + 7.218 Var(B)				
A*B	Var(Erro) + 3.609 Var(A*B)				
Testes de Hipóteses para a Análise de Variância do Modelo Misto					
Variável Dependente: Y					
Causa de Variação: A					
Erro: 0.9975*QM(A*B) + 0.0025*QM(Erro)					
		Denominador	Denominador		
GL	QM Tipo III	$H_0^{(3)}$	GL	QM	Valor F Pr > F
1	141.06666667		2.05	1.3256510417	106.4131 0.0085
Causa de Variação: B					
Erro: QM(A*B)					
		Denominador	Denominador		
GL	QM Tipo III	$H_0^{(III-B)}$	GL	QM	Valor F Pr > F
2	24.538847118		2	1.313283208	18.6851 0.0508
Causa de Variação: A*B					
Erro: QM(Erro)					
		Denominador	Denominador		
GL	QM Tipo III	$H_0^{(III-AB)}$	GL	QM	Valor F Pr > F
2	1.313283208		16	6.2604166667	0.2098 0.8130

Fato:

Verifica-se, também as mesmas ocorrências das Tabelas 21 e 22. ou seja, a hipótese $H_0^{(3)}$ continua sendo testada pelo comando **Random**, só que de modo diferente e as expressões $E(QM)$ são fornecidas segundo a formulação 1 conforme Tabela 3, ou modelo III conforme Tabela 1.

Tabela 24. Testes das hipóteses do Tipo IV na Análise de Variância para o modelo misto sem casela vazia segundo Proc GLM/SAS e opção Test

Causa de Variação		Esperança do Quadrado Médio Tipo IV			
A		Var(Erro) + 3.6 Var(A*B) + Q(A)			
B		Var(Erro) + 3.609 Var(A*B) + 7.218 Var(B)			
A*B		Var(Erro) + 3.609 Var(A*B)			
Causa de Variação: A					
Erro: 0.9975*QM(A*B) + 0.0025*QM(Erro)					
		Denominador	Denominador		
GL	QM Tipo IV	$H_0^{(4)}$	GL	QM	Valor F Pr > F
1	141.06666667		2.05	1.3256510417	106.4131 0.0085
Causa de Variação: B					
Erro: QM(A*B)					
		Denominador	Denominador		
GL	QM Tipo IV	$H_0^{(IV-B)}$	GL	QM	Valor F Pr > F
2	24.538847118		2	1.313283208	18.6851 0.0508
Causa de Variação: A*B					
Erro: QM(Erro)					
		Denominador	Denominador		
GL	QM Tipo IV	$H_0^{(IV-AB)}$	GL	QM	Valor F Pr > F
2	1.313283208		16	6.2604166667	0.2098 0.8130

Fato:

Verifica-se também na Tabela 24, os mesmos fatos das Tabelas 21, 22 e 23. Nesse caso, sem presença de caselas vazias, para o fator A fixo, $H_0^{(3)}$ e $H_0^{(4)}$ são equivalentes. Ademais verifica-se as somas de quadrados do Tipo III, nesse caso são equivalentes $SQ^{III} = R(\alpha|\mu, \beta, \gamma)$ é ajustada para todos os efeitos fixos e aleatórios.

5.1.1.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia. (Dados da Tabela 9)

5.1.1.2.1 Funções estimáveis e hipóteses testáveis

Descreve-se, a seguir nas Tabelas 25 a 29, a saída do Proc GLM relacionado com as funções estimáveis e hipóteses testáveis no modelo-M, segundo a ordenação A, B.

Tabela 25. Forma Geral das funções estimáveis segundo o Proc GLM/SAS para dados com casela vazia

INTERCEPTO	L1
A	1 L2
	2 L1-L2
B	1 L4
	2 L5
	3 L1-L4-L5
A*B	1 1 L7
	1 2 -L1+L2+L4+L5-L7
	1 3 L1-L4-L5
	2 1 L4-L7
	2 2 L1-L2-L4+L7

$$\begin{aligned}
 L'\beta = & L_1\mu + L_2\alpha_1 + (L_1 - L_2)\alpha_2 + L_4\beta_1 + L_5\beta_2 + (L_1 - L_4 - L_5)\beta_3 + \\
 & + L_7(\alpha\beta)_{11} + (-L_1 + L_2 + L_4 + L_5 - L_7)(\alpha\beta)_{12} + (L_1 - L_4 - L_5)(\alpha\beta)_{13} + (L_4 - L_7)(\alpha\beta)_{21} + \\
 & + (L_1 - L_2 - L_4 + L_7)(\alpha\beta)_{22}.
 \end{aligned}$$

Tabela 26. Funções estimáveis do Tipo I segundo o proc GLM/SAS com casela vazia

Funções estimáveis		Hipóteses testáveis
Funções Estimáveis Tipo I para: A		
Efeito	Coefficientes	
INTERCEPTO	0	$H_0^{(1)}: \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5} - \frac{2\mu_{21} + 3\mu_{22}}{5} = 0$
A	1 L2	
	2 -L2	
B	1 0	
	2 -0.4*L2	
	3 0.4*L2	
A*B	11 0.4*L2	
	12 0.2*L2	
	13 0.4*L2	
	21 -0.4*L2	
	22 -0.6*L2	
Funções Estimáveis Tipo I para: B		
Efeito	Coefficientes	
INTERCEPTO	0	$H_0^{(6)}: \begin{cases} \frac{2\mu_{11} + 2\mu_{21}}{4} = \frac{2\bar{L}_1 + 2\bar{L}_2}{4} \\ \frac{\mu_{12} + 3\mu_{22}}{4} = \frac{\bar{L}_1 + 3\bar{L}_2}{4} \end{cases}$ ONDE: $\bar{L}_1 = \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5}, \bar{L}_2 = \frac{2\mu_{21} + 3\mu_{22}}{5}$
A	1 0	
	2 0	
B	1 L4	
	2 L5	
	3 -L4-L5	
A*B	11 0.7857*L4+0.4286*L5	
	12 0.2143*L4+0.5714*L5	
	13 -L4-L5	
	21 0.2143*L4-0.4286*L5	
	22 -0.2143*L4+0.4286*L5	
Funções Estimáveis Tipo I para: A*B		
Efeito	Coefficientes	
INTERCEPTO	0	$H_0^{(9)}: \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$
A	1 0	
	2 0	
B	1 0	
	2 0	
	3 0	
A*B	11 L7	
	12 -L7	
	13 0	
	21 -L7	
	22 L7	

Fato:

Conforme já descrito em 5.1.1.1.1, pode-se converter a hipótese do

Modelo-M, $H_0^{(1)}: \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5} - \frac{2\mu_{21} + 3\mu_{22}}{5} = 0$, para o Modelo-S:

$$H_0^{(1)}: \alpha_1 - \alpha_2 - 0,4\beta_2 + 0,4\beta_3 + 0,4\alpha\beta_{11} + 0,2\alpha\beta_{12} - 0,4\alpha\beta_{21} - 0,6\alpha\beta_{22} = 0$$

Tabela 27. Funções estimáveis do Tipo II segundo o proc GLM/SAS com casela vazia.

Funções Estimáveis

Hipóteses Testáveis

Funções Estimáveis Tipo II para: A

Efeito Coeficientes

INTERCEPTO

A 1 L2

2 -L2

B 1 0

2 0

3 0

A*B 11 0.5714*L2

12 0.4286*L2

13 0

21 -0.5714*L2

22 -0.4286*L2

$$H_0^{(2)}: \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5} = \frac{2\bar{C}_1 + \bar{C}_2 + 2\bar{C}_3}{5},$$

onde:

$$\bar{C}_1 = \frac{2\mu_{11} + 2\mu_{21}}{4}; \bar{C}_2 = \frac{\mu_{12} + 3\mu_{22}}{4}; \bar{C}_3 = \mu_{13}$$

Tipo II Funções Estimáveis para: B

Efeito Coeficientes

INTERCEPTO 0

A 1 0

2 0

B 1 L4

2 L5

3 -L4-L5

A*B 11 0.7857*L4+0.4286*L5

12 0.2143*L4+0.5714*L5

13 -L4-L5

21 0.2143*L4-0.4286*L5

22 -0.2143*L4+0.4286*L5

$$H_0^{(6)}: \begin{cases} \frac{2\mu_{11} + 2\mu_{21}}{4} = \frac{2\bar{L}_1 + 2\bar{L}_2}{4} \\ \frac{\mu_{12} + 3\mu_{22}}{4} = \frac{\bar{L}_1 + 3\bar{L}_2}{4} \end{cases}$$

ONDE:

$$\bar{L}_1 = \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5}, \bar{L}_2 = \frac{2\mu_{21} + 3\mu_{22}}{5}$$

:

Tipo II Funções Estimáveis para: A*B

Efeito Coeficientes

INTERCEPTO 0

A 1 0

2 0

B 1 0

2 0

3 0

A*B 11 L7

12 -L7

13 0

21 -L7

22

$$H_0^{(9)}: \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = C$$

Tabela 28. Funções estimáveis do Tipo III segundo o proc GLM/SAS com casela vazia

Funções estimáveis	Hipóteses testáveis
Funções Estimáveis Tipo III para: A	
Efeito	Coefficientes
INTERCEPTO	0
A	1 L2 2 -L2
B	1 0 2 0 3 0
A*B	11 0.5*L2 12 0.5*L2 13 0 21 -0.5*L2 22 -0.5*L2
$H_0^{(3)}: \frac{\mu_{11} + \mu_{12}}{2} = \frac{\mu_{21} + \mu_{22}}{2}$	
Funções Estimáveis Tipo III para: B	
Efeito	Coefficientes
INTERCEPTO	0
A	1 0 2 0
B	1 L4 2 L5 3 -L4-L5
A*B	11 0.75*L4+0.25*L5 12 0.25*L4+0.75*L5 13 -L4-L5 21 0.25*L4-0.25*L5 22 -0.25*L4+0.25*L5
$H_0^{(7)}: \begin{cases} \frac{\mu_{11} + \mu_{21}}{2} = \frac{\mu_{12} + \mu_{22}}{2} \\ \frac{\mu_{11} + \mu_{21}}{2} = \mu_{13} \end{cases}$	
Funções Estimáveis Tipo III para: A*B	
Efeito	Coefficientes
INTERCEPTO	0
A	1 0 2 0
B	1 0 2 0 3 0
A*B	11 L7 12 -L7 13 0 21 -L7 22 L7
$H_0^{(9)}: \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$	

Tabela 29. Funções estimáveis do Tipo IV segundo o proc GLM/SAS com casela vazia

Funções estimáveis		Hipóteses testáveis	
Funções Estimáveis Tipo IV para: A			
Efeito	Coefficientes		
INTERCEPTO	0	$H_0^{(4)}: \mu_{21} + \mu_{22} = \mu_{11} + \mu_{12}$	
A	1 L2		
	2 -L2		
B	1 0		
	2 0		
	3 0		
A*B	11 0.5*L2		
	12 0.5*L2		
	13 0		
	21 -0.5*L2		
	22 -0.5*L2		
Funções Estimáveis para Tipo IV: B			
INTERCEPTO	0		$H_0^{(8)}: \begin{cases} \mu_{13} = \mu_{11} \\ \mu_{13} = \mu_{12} \end{cases}$
A	1 0		
	2 0		
B	1 L4		
	2 L5		
	3 -L4-L5		
A*B	11 L4		
	12 L5		
	13 -L4-L5		
	21 0		
	22 0		
Funções Estimáveis Tipo IV para: A*B			
INTERCEPTO	0	$H_0^{(9)}: \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$	
A	1 0		
	2 0		
B	1 0		
	2 0		
	3 0		
A*B	11 L7		
	12 -L7		
	13 0		
	21 -L7		
	22 L7		

5.1.1.2.2 Hipóteses e somas de quadrados associadas

A Tabela 30 descreve a saída do Proc GLM, relacionada à análise de variância, emitida pela especificação **Model**, que admite todos os efeitos como sendo fixos.

Tabela 30. Análise de variância com casela vazia segundo o comando MODEL/GLM/SAS

Variável Dependente: Y						
Causa de Variação	GL	Soma de Quadrados	Quadrados Médio	Valor F	Pr > F	
Modelo	4	126.000000	31.500000	13.13	0.0073	
Erro	5	12.000000	2.400000			
Total Corrigido	9	138.000000				
	R-Quadrados	C.V.	Raiz QME	Y Médio		
	0.913043	3.872983	1.54919	40.0000		

Variável Dependente: Y							
Causa de	GL	H_0	$R(\cdot)$	SQ Tipo I	QM	Valor F	Pr > F
A	1	$H_0^{(1)}$	$R(\alpha \mu)$	90.0000000	90.0000000	37.50	0.0017
B	2	$H_0^{(6)}$	$R(\beta \mu, \alpha)$	34.2857143	17.1428571	7.14	0.0342
A*B	1	$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$	1.7142857	1.7142857	0.71	0.4366
Causa	GL	$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$	SQ Tipo II	QM	Valor F	Pr > F
A	1	$H_0^{(2)}$	$R(\alpha \mu, \beta)$	41.2857143	41.2857143	17.20	0.0089
B	2	$H_0^{(6)}$	$R(\beta \mu, \alpha)$	34.2857143	17.1428571	7.14	0.0342
A*B	1	$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$	1.7142857	1.7142857	0.71	0.4366
Causa	GL	$H_0^{(3)}$	$R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha\hat{\beta})$	SQ Tipo III	QM	Valor F	Pr > F
A	1	$H_0^{(7)}$	$R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha\hat{\beta})$	42.8571429	42.8571429	17.86	0.0083
B	2	$H_0^{(6)}$	$R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha\hat{\beta})$	27.5789474	13.7894737	5.75	0.0506
A*B	1	$H_0^{(9)}$	$R(\alpha\hat{\beta} \hat{\mu}, \hat{\alpha}, \hat{\beta})$	1.7142857	1.7142857	0.71	0.4366
Causa	GL	$H_0^{(4)}$	$R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha\hat{\beta})$	SQ Tipo IV	QM	Valor F	Pr > F
A	1•	$H_0^{(8)}$	$R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha\hat{\beta})$	42.8571429	42.8571429	17.86	0.0083
B	2•	$H_0^{(9)}$	$R(\alpha\hat{\beta} \hat{\mu}, \hat{\alpha}, \hat{\beta})$	6.0000000	3.0000000	1.25	0.3629
A*B	1	$H_0^{(9)}$	$R(\alpha\hat{\beta} \hat{\mu}, \hat{\alpha}, \hat{\beta})$	1.7142857	1.7142857	0.71	0.4366

• NOTA: Outras Hipóteses Testáveis do Tipo IV existem as quais podem fornecer somas de quadrados diferentes.

5.1.1.2.3 Testes de hipóteses para a análise de variância do modelo misto

O Proc GLM apresenta através da comando Random com a opção Test, os testes de hipóteses, utilizando-se dos quadrados médios esperados para

cada efeito declarado no comando **Model**. Esses conforme mostram as Tabelas 31 - 34 são os testes considerados apropriados para o modelo misto.

Tabela 31. Testes das Hipóteses do Tipo I na Análise de Variância para o modelo misto com casela vazia segundo a opção **Test** do SAS

Causa de Variação	Esperança Quadrado Médio Tipo I
A	$\text{Var}(\text{Erro}) + 2.2 \text{Var}(A*B) + 0.8 \text{Var}(B) + Q(A)$
B	$\text{Var}(\text{Erro}) + 1.9429 \text{Var}(A*B) + 2.8 \text{Var}(B)$
A*B	$\text{Var}(\text{Erro}) + 1.7143 \text{Var}(A*B)$

Testes de Hipóteses para a Análise de Variância para o Modelo Misto

Variável Dependente: Y

Causa de Variação: A

Erro: $0.2857*QM(B) + 0.9595*QM(A*B) - 0.2452*QM(\text{Erro})$

GL	$H_0^{(1)}$	Tipo I QM	GL	QM	Valor F	Pr > F
1		90	2.40	5.9542857143	15.1152	0.0446

Causa de Variação: B

Erro: $1.1333*QM(A*B) - 0.1333*QM(\text{Erro})$

GL	$H_0^{(I-B)}$	Tipo I QM	GL	QM	Valor F	Pr > F
2		17.142857143	0.69	1.6228571429	10.5634	0.3023

Causa de Variação: A*B

Erro: $QM(\text{Erro})$

GL	$H_0^{(I-AB)}$	Tipo I QM	GL	QM	Valor F	Pr > F
1		1.7142857143	5	2.4	0.7143	0.4

Tabela 32. Testes das Hipóteses do Tipo II na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS

Causa de Variação	Esperança	Quadrado Médio do Tipo II
A	$\text{Var}(\text{Erro}) + 1.7857 \text{Var}(A*B) + Q(A)$	
B	$\text{Var}(\text{Erro}) + 1.9429 \text{Var}(A*B) + 2.8 \text{Var}(B)$	
A*B	$\text{Var}(\text{Erro}) + 1.7143 \text{Var}(A*B)$	

Testes de Hipóteses para a Análise de Variância para o Modelo Misto

Variável Dependente: Y

Causa de Variação: A

Erro: $1.0417*QM(A*B) - 0.0417*QM(\text{Erro})$

GL	Denominador	Tipo II QM	GL	QM	Valor F	Pr > F
1	$H_0^{(2)}$	41.285714286	0.89	1.6857142857	24.4915	0.1491

Causa de Variação: B

Erro: $1.1333*QM(A*B) - 0.1333*QM(\text{Erro})$

GL	Denominador	Tipo II QM	GL	QM	Valor F	Pr > F
2	$H_0^{(II-B)}$	17.142857143	0.69	1.6228571429	10.5634	0.3023

Causa de Variação: A*B

Erro: $QM(\text{Erro})$

GL	Denominador	Tipo II QM	GL	QM	Valor F	Pr > F
1	$H_0^{(II-AB)}$	1.7142857143	5	2.4	0.7143	0.4366

Tabela 33. Testes das Hipóteses do Tipo III na Análise de Variância para o modelo misto com casela vazia segundo a opção **Test** do SAS

Causa de Variação	Esperança	Quadrado Médio	Tipo III
A	$\text{Var}(\text{Erro}) + 1.7143 \text{ Var}(A*B) + Q(A)$		
B	$\text{Var}(\text{Erro}) + 1.7684 \text{ Var}(A*B) + 2.6526 \text{ Var}(B)$		
A*B	$\text{Var}(\text{Erro}) + 1.7143 \text{ Var}(A*B)$		

Testes de Hipóteses para a Análise de Variância para o Modelo Misto

Variável Dependente: Y

Causa de Variação: A

Erro: QM(A*B)

GL	$H_0^{(3)}$	Denominador	Denominador	GL	QM	Valor F	Pr > F
		Tipo III	QM				
1		42.857142857		1	1.7142857143	25.0000	0.1257

Causa de Variação: B

Erro: $1.0316*QM(A*B) - 0.0316*QM(\text{Erro})$

GL	$H_0^{(III-B)}$	Denominador	Denominador	GL	QM	Valor F	Pr > F
		Tipo III	QM				
2		13.789473684		0.92	1.6926315789	8.1468	0.2610

Causa de Variação: A*B

Erro: QM(Erro)

GL	$H_0^{(III-AB)}$	Denominador	Denominador	GL	QM	Valor F	Pr > F
		Tipo III	QM				
1		1.7142857143		5	2.4	0.7143	0.4366

Fatos:

1) Verifica-se, através das Tabelas 33 e 34 que para o fator A fixo, que:

SQ do Tipo III = $SQH_0^{(3)} = SQH_0^{(4)} = 42,8571$, é ajustada para todos os efeitos fixos e aleatórios.

2) Através das expressões das esperanças matemáticas, conforme as Tabelas 31 a 34 e as Tabelas 21 a 24, observa-se que a expressão de Tipo III para o fator A fixo é a que mais se aproxima da interação.

Tabela 34. Testes das Hipóteses do Tipo IV na Análise de Variância para o modelo misto com casela vazia segundo a opção Test do SAS

Causa de Variação	Esperança Quadrado Médio Tipo IV
A	$\text{Var}(\text{Erro}) + 1.7143 \text{Var}(A*B) + Q(A)$
B	$\text{Var}(\text{Erro}) + 1.6 \text{Var}(A*B) + 1.6 \text{Var}(B)$
A*B	$\text{Var}(\text{Erro}) + 1.7143 \text{Var}(A*B)$

Testes de Hipóteses para a Análise de Variância para o Modelo Misto

Variável Dependente: Y

Causa de Variação: A

Erro: QM(A*B)

GL	$H_0^{(4)}$	Denominador Tipo IV QM	Denominador GL	QM	Valor F	Pr > F
1		42.857142857	1	1.7142857143	25.0000	0.1257

Causa de Variação: B

Erro: $0.9333*QM(A*B) + 0.0667*QM(\text{Erro})$

GL	$H_0^{(IV-B)}$	Denominador Tipo IV QM	Denominador GL	QM	Valor F	Pr > F
2		3	1.21	1.76	1.7045	0.4450

Causa de Variação: A*B

Erro: QM(Erro)

GL	$H_0^{(IV-AB)}$	Denominador Tipo IV QM	Denominador GL	QM	Valor F	Pr > F
1		1.7142857143	5	2.4	0.7143	0.4366

5.1.2 Procedimento VARCOMP

O Proc VARCOMP fornece as estimativas dos componentes de variância para o modelo linear misto, através dos métodos ANOVA, ML, REML e MIVQUE(0). Para a análise dos dados da Tabela 8 e 9 utilizou-se o conjunto de comandos básicos a seguir:

```
proc varcomp method = type1\MIVQUEO\ML\REML;  
  class a b;  
  model y = a b a*b / fixed =1;  
run;
```

O Proc VARCOMP assume que todos efeitos do modelo são aleatórios, a opção **Fixed = 1** no comando **Model** é utilizada para especificar que o primeiro fator declarado no modelo é fixo. O MIVQUE(0) é o método padrão

Apresenta a seguir as saídas obtidas através desse conjunto de sintaxe básica.

5.1.2.1 Estimativas dos componentes de variância, para os dados desbalanceados sem casela vazia. (Dados da Tabela 8)

As Tabelas 35-38 descrevem saídas do Proc VARCOMP do SAS para os dados da Tabela 8.

Tabela 35. Estimativas do Tipo I para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS

Procedimento da Análise de Variância para Estimação dos Componentes de Variância			
Variável dependente:			
Causa de	GL	SQ Tipo I	QM Tipo I
Varição			
A	$R(\alpha \mu)$	156.12272727	156.12272727
B	$R(\beta \mu, \alpha)$	47.85676692	23.92838346
A*B	$R(\alpha\beta \mu, \alpha, \beta)$	2.62656642	1.31328321
Erro		100.16666667	6.26041667
Total		306.77272727	

Causa de Variação	Esperanças dos Quadrados Médios
A	$\text{Var}(\text{Erro}) + 3.6727 \text{Var}(A*B) + 0.0364 \text{Var}(B) + Q(A)$
B	$\text{Var}(\text{Erro}) + 3.691 \text{Var}(A*B) + 7.3 \text{Var}(B)$
A*B	$\text{Var}(\text{Erro}) + 3.609 \text{Var}(A*B)$
Erro	$\text{Var}(\text{Erro})$

Componente de Variância	Estimativa
$\text{Var}(B)$	3.11334814
$\text{Var}(A*B)$	-1.37076823
$\text{Var}(\text{Erro})$	6.26041667

Fato:

Verifica-se os mesmos resultados obtidos na Tabela 21, para a análise do fator A fixo, as mesmas expressões dos valores esperados, com a adição dos resultados das estimativas dos componentes de variância, as quais são as mesmas determinadas com os resultados da Tabela 21. Observa-se que a estimativa negativa para o componente $\sigma_{\alpha\beta}^2$.

Tabela 36. Estimativas MIVQUE(0) para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS

Procedimento MIVQUE(0) para Estimação dos Componentes de Variância			
Matriz SQ			
Causa de Variação		B	A*B
B	106.76000000	53.96000000	
A*B	53.96000000	53.96000000	
Erro	14.60000000	14.60000000	
Causa de Variação	Erro	Y	
B	14.60000000	360.24000000	
A*B	14.60000000	185.44000000	
Erro	20.00000000	150.65000000	
Componente de Variância		Estimativa	
Var(B)		3.31060606	
Var(A*B)		-1.56783436	
Var(Erro)		6.26027666	

Fato:

Verifica-se na Tabela 36, os resultados obtidos através do método MIVQUE(0), para os componentes de variâncias, ou seja, a matriz das estimativas das somas de quadrados e as estimativas dos componentes de variância. Observa-se nesse caso, estimativa negativa para o componente $\sigma_{\alpha\beta}^2$, isto é o método atribui valor negativo para as estimativas dos componentes.

Tabela 37. Estimativas ML para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS

Procedimento Máxima Verossimilhança para Estimação dos Componentes de Variância

Variável Dependente: Y

Iteração	Objetivo	Var(B)	Var(A*B)	Var(Erro)
0	40.63405960	2.70197111	0	5.10936258
1	40.33297386	1.45028047	0	5.39182307
2	40.33270998	1.42174643	0	5.40236827
3	40.33270989	1.42227057	0	5.40217192
4	40.33270989	1.42227057	0	5.40217192

Critério Convergência met.

Procedimento Máxima Verossimilhança para Estimação dos Componentes de Variância

Variável Dependente: Y

Matriz das estimativas de covariância assintótica

	Var(B)	Var(A*B)	Var(Erro)
Var(B)	3.0952806838	0	-0.391890798
Var(A*B)	0	0	0
Var(Erro)	-0.391890798	0	3.0614119967

Fato:

Verifica-se na Tabela 37 os resultados obtidos através do método ML para os componentes de variâncias, ou seja, a matriz das estimativas das covariâncias assintóticas e as estimativas dos componentes de variância, nesse caso obtidas na quarta iteração. Observa-se que o método em questão atribui valor zero, para o componente σ_{af}^2 . A matriz de covariâncias apresenta

valores negativos para estimativas fora da diagonal, trata-se de estimativas de covariâncias essas podem assumir valores negativos.

Tabela 38. Estimativas REML para os componentes de variância sem casela vazia segundo o Proc VARCOMP do SAS

Procedimento Máxima Verossimilhança Restrita para Estimação dos Componentes de Variância

Variável Dependente: Y

Iteração	Objetivo	Var(B)	Var(A*B)	Var(Erro)
0	37.68888879	2.97216823	0	5.62029883
1	37.66562985	2.44534177	0	5.70613149
2	37.66562984	2.44569650	0	5.70606365
3	37.66562984	2.44569650	0	5.70606365

Critério de convergência met.

Procedimento Máxima Verossimilhança Restrita para Estimação dos Componentes de Variância

Variável Dependente: Y

Matriz das estimativas de covariâncias assintóticas

	Var(B)	Var(A*B)	Var(Erro)
Var(B)	10.329977872	0	-0.465543654
Var(A*B)	0	0	0
Var(Erro)	-0.465543654	0	3.6118198422

Fato:

Verifica-se na Tabela 38 os resultados obtidos através do método REML para os componentes de variâncias, ou seja, a matriz das estimativas das covariâncias assintóticas e as estimativas dos componentes de variância,

obtidas por critério de convergência, nesse caso com três iterações. Observa-se, que o método em questão atribui valor zero, para o componente $\sigma_{\alpha\beta}^2$.

5.1.2.2 Estimativas dos componentes de variância, para os dados desbalanceados com casela vazia. (Dados da Tabela 9)

As Tabelas 39-42 descrevem saídas do Proc VARCOMP do SAS para os dados da Tabela 9.

Tabela 39. Estimativas do Tipo I para os componentes de variância com casela vazia segundo o Proc VARCOMP do SAS

Procedimento da Análise de Variância para Estimação dos Componentes de Variância			
Variável Dependente: Y			
Causa de	GL	SQ Tipo I	QM Tipo I
Varição			
A	1	90.00000000	90.00000000
B	2	34.28571429	17.14285714
A*B	1	1.71428571	1.71428571
Erro	5	12.00000000	2.40000000
Total	9	138.00000000	
Causa de Varição		Esperanças dos Quadrados Médios	
A		Var(Erro) + 2.2 Var(A*B) + 0.8 Var(B) + Q(A)	
B		Var(Erro) + 1.9429 Var(A*B) + 2.8 Var(B)	
A*B		Var(Erro) + 1.7143 Var(A*B)	
Erro		Var(Erro)	
Componente de Variância	Estimativa		
Var(B)	5.54285714		
Var(A*B)	-0.40000000		
Var(Erro)	2.40000000		

Fato:

Verifica-se os mesmos resultados obtidos na Tabela 31, para as esperanças dos quadrados médios do Tipo I, com a adição dos resultados das estimativas dos componentes de variância. Observa-se estimativa negativa para o componente $\sigma_{\alpha\beta}^2$.

Tabela 40. Estimativas MIVQUE(O) para os componentes de variância com casela vazia segundo o Proc VARCOMP do SAS

Procedimento MIVQUE(0) para Estimação dos Componentes de Variância		
Matriz das SQ		
Causa de Variação	B	A*B
B	17.92000000	11.20000000
A*B	11.20000000	11.20000000
Erro	5.60000000	5.60000000
Causa de Variação	Erro	Y
B	5.60000000	128.00000000
A*B	5.60000000	80.00000000
Erro	8.00000000	48.00000000
Estimativa		
Componente de Variância	Y	
Var(B)	7.14285714	
Var(A*B)	-0.76923077	
Var(Erro)	1.53846154	

Fato:

Verifica-se na Tabela 40, os resultados obtidos através do método MIVQUE(0), para os componentes de variância, ou seja, a matriz das estimativas das somas de quadrados e as estimativas dos componentes de variância. Observa-se nesse caso, estimativa negativa para o componente $\sigma_{\alpha\beta}^2$.

Tabela 41. Estimativas ML para os componentes de variância com casela vazia segundo o Proc VARCOMP do SAS

Procedimento Máxima Verossimilhança para Estimação dos Componentes de Variância				
Variável Dependente: Y				
Iteração	Objetivo	Var(B)	Var(A*B)	Var(Erro)
0	12.75315264	7.26272753	0	1.56427978
1	11.90533165	4.02903010	0	1.73558220
2	11.70671925	2.73251896	0	1.92872790
3	11.70625332	2.68150037	0	1.94033552
4	11.70625323	2.68084154	0	1.94048838
5	11.70625323	2.68080984	0	1.94049574

Critério Convergência met.

Procedimento Máxima Verossimilhança para Estimação dos Componentes de Variância

Variável Dependente: Y

Matriz das estimativas de covariâncias assintóticas

	Var(B)	Var(A*B)	Var(Erro)
Var(B)	6.7812326443	0	-0.199781025
Var(A*B)	0	0	0
Var(Erro)	-0.199781025	0	1.0409288253

Fato:

Verifica-se na Tabela 41, os resultados obtidos através do método ML para os componentes de variâncias, ou seja, a matriz das estimativas das covariâncias assintóticas e as estimativas dos componentes de variância, nesse caso obtidas na quinta iteração. Observa-se que o método em questão atribui valor zero, para o componente $\sigma_{\alpha\beta}^2$.

Tabela 42. Estimativas REML para os componentes de variância com casela vazia segundo o Proc VARCOMP do SAS

Procedimento Máxima Verossimilhança Restrita para Estimação dos Componentes de Variância
Variável Dependente: Y

Iteração	Objetivo	Var(B)	Var(A*B)	Var(Erro)
0	10.51121607	9.07840942	0	1.95534972
1	10.11219132	5.03628763	0	2.16947775
2	10.09320020	4.30132940	0	2.25529223
3	10.09307228	4.35097318	0	2.24852689
4	10.09307222	4.35211341	0	2.24837341
5	10.09307222	4.35211341	0	2.24837341

Critério de convergência met.

Procedimento Máxima Verossimilhança Restrita para Estimação dos Componentes de Variância
Variável Dependente: Y

Matriz das estimativas de covariâncias assintóticas

	Var(B)	Var(A*B)	Var(Erro)
Var(B)	25.174897372	0	-0.30411339
Var(A*B)	0	0	0
Var(Erro)	-0.30411339	0	1.6338965993

Fato:

Verifica-se na Tabela 42, os resultados obtidos através do método ML para os componentes de variâncias, ou seja, a matriz das estimativas das covariâncias assintóticas e as estimativas dos componentes de variância, nesse caso obtidas na quinta iteração. Observa-se que o método em questão atribui valor zero, para o componente $\sigma_{\alpha\beta}^2$.

5.1.3 Procedimento MIXED

O procedimento MIXED ajusta o modelo misto geral, de modo que para os dados da Tabela 8, tem-se:

I) O sistema de equações lineares, correspondente ao modelo-S.

$$\begin{array}{c}
 y_{111} \\
 y_{112} \\
 y_{113} \\
 y_{114} \\
 y_{121} \\
 y_{122} \\
 y_{123} \\
 y_{131} \\
 y_{132} \\
 y_{133} \\
 y_{211} \\
 y_{212} \\
 y_{213} \\
 y_{214} \\
 y_{221} \\
 y_{222} \\
 y_{223} \\
 y_{224} \\
 y_{231} \\
 y_{232} \\
 y_{233} \\
 y_{234}
 \end{array}
 =
 \begin{array}{c}
 43 \\
 45 \\
 46 \\
 53 \\
 41 \\
 42 \\
 44 \\
 42 \\
 44 \\
 46 \\
 40 \\
 40 \\
 40 \\
 43 \\
 35 \\
 37 \\
 40 \\
 40 \\
 37 \\
 39 \\
 40 \\
 40
 \end{array}
 =
 \begin{array}{ccc}
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 1 & 0 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1 \\
 1 & 0 & 1
 \end{array}
 \begin{array}{c}
 \mu \\
 \alpha_1 \\
 \alpha_2
 \end{array}
 +
 \begin{array}{cccccccc}
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
 \end{array}
 \begin{array}{c}
 \beta_1 \\
 \beta_2 \\
 \beta_3 \\
 (\alpha\beta)_{11} \\
 (\alpha\beta)_{12} \\
 (\alpha\beta)_{13} \\
 (\alpha\beta)_{21} \\
 (\alpha\beta)_{22} \\
 (\alpha\beta)_{23}
 \end{array}
 +
 \begin{array}{c}
 e_{111} \\
 e_{112} \\
 e_{113} \\
 e_{114} \\
 e_{121} \\
 e_{122} \\
 e_{123} \\
 e_{131} \\
 e_{132} \\
 e_{133} \\
 e_{211} \\
 e_{212} \\
 e_{213} \\
 e_{214} \\
 e_{221} \\
 e_{222} \\
 e_{223} \\
 e_{224} \\
 e_{231} \\
 e_{232} \\
 e_{233} \\
 e_{234}
 \end{array}$$

II) As configurações.

Para o modelo-S.

Tabela 43. Configurações para o modelo-S com todas as caselas ocupadas

	j = 1	j = 2	j = 3
i = 1	$\mu + \alpha_1(4)$	$\mu + \alpha_1(3)$	$\mu + \alpha_1(3)$
i = 2	$\mu + \alpha_2(4)$	$\mu + \alpha_2(4)$	$\mu + \alpha_2(4)$

Para o modelo-M.

Tabela 44. Configurações para o modelo-M com todas as caselas ocupadas

	j = 1	j = 2	j = 3
i = 1	$\mu_1(4)$	$\mu_1(3)$	$\mu_1(3)$
i = 2	$\mu_2(4)$	$\mu_2(4)$	$\mu_2(4)$

Equivalentemente, para os dados do Tabela 9, tem-se

III) O sistema de equações lineares, correspondente ao modelo-S.

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{131} \\ y_{132} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{223} \end{bmatrix} = \begin{bmatrix} 43 \\ 45 \\ 41 \\ 42 \\ 44 \\ 40 \\ 40 \\ 35 \\ 37 \\ 33 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \end{bmatrix} + \begin{bmatrix} e_{111} \\ e_{112} \\ e_{121} \\ e_{131} \\ e_{132} \\ e_{211} \\ e_{212} \\ e_{221} \\ e_{222} \\ e_{223} \end{bmatrix}$$

IV) As seguintes configurações:

Para o modelo-S.

Tabela 45. Configurações para o modelo-S com casela vazia

	j = 1	j = 2	j = 3
i = 1	$\mu + \alpha_1(2)$	$\mu + \alpha_1(1)$	$\mu + \alpha_1(2)$
i = 2	$\mu + \alpha_2(2)$	$\mu + \alpha_2(3)$	

Para o modelo-M.

Tabela 46. Configurações para o modelo-M com casela vazia

	j = 1	j = 2	j = 3
i = 1	$\mu_1(2)$	$\mu_1(1)$	$\mu_1(2)$
i = 2	$\mu_2(2)$	$\mu_2(3)$	

Proc MIXED reúne um conjunto de comandos com opções as quais permitem ajustar o modelo misto geral e para esse trabalho considerou-se desse conjunto apenas os comandos básicos a seguir:

```
proc mixed method = reml/ml/mivque0;
class a b;
model y = a / e e3 ;
random b a*b / s;
run;
```

Com a declaração:

method = define o método que deve ser utilizado para obter as estimativas dos componentes de variância. Quando nenhum método é especificado, o Proc MIXED utiliza como padrão o método REML.

class a b, define-se as variáveis classificatórias, as quais serão utilizadas na análise.

model y = a, declara que o efeito **a** é fixo;

/ e e3 apresenta a forma geral das funções estimáveis e as funções do Tipo III para todos os efeitos declarados no **Model**. Apresenta nas saídas para os efeitos fixos o teste F usando as somas de quadrados do Tipo III.

random b a*b, define a matriz **Z** do modelo misto, ou seja especifica os efeitos aleatórios do modelo que compõem o vetor ν e a estrutura da matriz **G**;

5.1.3.1 Análise do modelo misto de classificação dupla com desbalanceamento e sem casela vazia. (Dados daTabela 8)

Descreve-se, através das Tabelas 47 a 49, as saídas do Proc MIXED para o ajuste do modelo linear de efeitos mistos, considerando-se os dados apresentados na Tabela 8.

Tabela 47. Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED do SAS e segundo o método REML

Histórico das Iterações da Estimação REML					
Iteração	Avaliações	Objetivo	Critério		
0	1	65.17203158			
1	3	62.45838994	0.00015408		
2	1	62.45320304	0.00000258		
3	1	62.45312161	0.00000000		
Critério de Convergência Met.					
Estimativas dos Parâmetros de Covariância (REML)					
Parâm Cov	Razão	Estimativa	Erro Padrão	Z	Pr > Z
B	0.42851801	2.44519236	3.21303619	0.76	0.4466
A*B	0.00000000	0.00000000			
Resíduo	1.00000000	5.70616007	1.90051094	3.00	0.0027
Informação do modelo ajustado para Y					
Descrição	Valor				
Observações	22.0000				
Estimativa da variância	5.7062				
Estimativa do erro padrão	2.3888				
Log Verossimilhança REML	-49.6053				
Critério de Informação de Akaike	-52.6053				
Critério Bayesiano de Schwartz	-54.0989				
-2 Log Verossimilhança REML	99.2107				
Chi-Quadrado LRT Modelo Nulo	2.7189				
GL LRT Modelo Nulo	2.0000				
P-Valor LRT Modelo Nulo	0.2568				
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.46975562	1.09453948	16	1.34	0.1981
B 2	-1.21700564	1.10294147	16	-1.10	0.2862
B 3	-0.25274998	1.10294147	16	-0.23	0.8216
A*B 1 1	0.00000000
A*B 1 2	-0.00000000
A*B 1 3	-0.00000000
A*B 2 1	0.00000000
A*B 2 2	-0.00000000
A*B 2 3	0.00000000
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de variação	GLN	GLD	F Tipo III	Pr > F	
A	1	2	25.78	0.0367	

Fatos: Observando-se as Tabelas 38 e 47:

1) Verifica-se que o Proc MIXED, fornece mais informações sobre o ajuste do modelo do que o Proc VARCOMP, por exemplo apresenta teste do Tipo III, para o efeito A fixo, nesse caso com base nas estimativas REML para os componentes de variância. O Proc VARCOMP apresenta a soma de quadrados do Tipo I, para o efeito A fixo.

2) O Proc MIXED fornece também as seguintes informações adicionais sobre o ajuste do modelo.

Observations	número de observações utilizadas na análise
Variance Estimate	estimativa σ^2
Standard Deviation Estimate	estimativa σ
REML Log Likelihood	valor máximo do log da verossimilhança restrita
Akaike's Information Criterion	o critério AIC que pode ser usado para comparar modelos com os mesmos efeitos fixos, mas com diferentes estruturas de covariâncias. O modelo com valor AIC mais perto de zero pode ser considerado o melhor.
Schwartz's Bayesian Criterion	o modelo com valor SBC mais perto de zero é o melhor
-2 REML Log Likelihood	compara modelos com os mesmos efeitos fixos mas com diferentes estruturas de covariâncias, através do teste da razão de verossimilhança (LRT).
Null Model LRT Chi-Square	teste da razão de verossimilhança comparando

	este modelo com o modelo ajustado pelo Proc GLM, com $R = \sigma^2 I$
Null Model LRT DF	os graus de liberdade para o Null Model LRT Chi-Square que corresponde ao número de parâmetros de covariância estimados pelo Proc GLM.
Null Model LRT -Value	a probabilidade de observar um Null Model LRT Chi-Square.

3) O Proc MIXED apresenta as soluções e testes para os efeitos aleatórios.

4) Verifica-se para o teste da razão de verossimilhança emite as informações:

$$\chi^2 = -2\text{Log verossimilhança REML do modelo nulo} - (-2\text{Log verossimilhança REML do modelo completo}) = 2.7189 .$$

gl = número de parâmetros aleatórios no modelo completo - número de parâmetros aleatórios no modelo nulo = 3 - 1 = 2.

5) O critério AIC é determinado por:

$$\begin{aligned} \text{AIC} &= \log \text{ de verossimilhança REML} - \text{o número de parâmetros aleatórios} \\ &= p(\theta) - q \\ &= -49,6053 - 3 = - 52,6053. \end{aligned}$$

Esses fatos, também, podem ser observados comparando-se as demais Tabelas que apresentam resultados dos ajustes através dos Proc MIXED, considerando-se o modelo misto, sem casela vazia e com casela vazia.

Tabela 48. Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED e segundo o método ML

Histórico das Iterações da Estimação ML					
Iteração	Avaliações	Objetivo	Critério		
0	1	64.32616987			
1	3	62.35020689	0.00049230		
2	1	62.33323399	0.00001641		
3	1	62.33271057	0.00000002		
4	1	62.33270989	0.00000000		
Critério de Convergência Met.					
Estimativas dos Parâmetros de Covariancia(MLE)					
Parâm Cov	Razão	Estimativa	Erro Padrão	Z	Pr > Z
B	0.26327765	1.42227110	1.76273427	0.81	0.4197
A*B	0.00000000	0.00000000			
Residual	1.00000000	5.40217172	1.75016792	3.09	0.0020
Informação do modelo ajustado para Y					
Descrição		Valor			
Observações		22.0000			
Estimativa da variância		5.4022			
Estimativa do erro padrão		2.3243			
Log Verossimilhança		-51.3830			
Critério de Informação de Akaike		-54.3830			
Critério Bayesiano de Schwartz		-56.0196			
-2 Log Verossimilhança		102.7660			
Chi-Quadrado LRT Modelo Nulo		1.9935			
GL LRT Modelo Nulo		2.0000			
P-Valor LRT Modelo Nulo		0.3691			
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.28089597	0.88885024	16	1.44	0.1688
B 2	-1.05718135	0.89622656	16	-1.18	0.2554
B 3	-0.22371462	0.89622656	16	-0.25	0.8061
A*B 1 1	0.00000000
A*B 1 2	-0.00000000
A*B 1 3	-0.00000000
A*B 2 1	0.00000000
A*B 2 2	-0.00000000
A*B 2 3	-0.00000000
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de Variação	GLN	GLD	F Tipo III	Pr > F	
A	1	2	27.45	0.0346	

Tabela 49. Ajuste do modelo misto de classificação dupla sem casela vazia através do Proc MIXED e segundo o método MIVQUE(0)

Estimativas dos Parâmetros de covariâncias (MIVQUE0)					
Cov Parm	Ratio	Estimate			
B	0.52882744	3.31060606			
A*B	0.00000001	0.00000006			
Residual	1.00000000	6.26027666			
Informação do modelo ajustado para Y					
Descrição					Valor
Observações					22.0000
Estimativa da variância					6.2603
Estimativa do erro padrão					2.5021
REML Log Likelihood					-49.6731
Critério de Informação de Akaike					-52.6731
Critério Bayesiano de Schwartz					-54.1667
-2 Log Verossimilhança REML					99.3461
Chi-Quadrado LRT Modelo Nulo					2.5834
GL LRT Modelo Nulo					2.0000
P-Valor LRT Modelo Nulo					0.2748
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.53830332	1.24228448	16	1.24	0.2335
B 2	-1.27528282	1.25121026	16	-1.02	0.3233
B 3	-0.26302050	1.25121026	16	-0.21	0.8362
A*B 1 1	0.00000003	0.00025021	16	0.00	0.9999
A*B 1 2	-0.00000003	0.00025021	16	-0.00	0.9999
A*B 1 3	-0.00000001	0.00025021	16	-0.00	1.0000
A*B 2 1	-0.00000000	0.00025021	16	-0.00	1.0000
A*B 2 2	0.00000000	0.00025021	16	0.00	1.0000
A*B 2 3	0.00000000	0.00025021	16	0.00	1.0000
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de Variação	GLN	GLD	F Tipo III	Pr > F	
A	1	2	23.44	0.0401	

5.1.3.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia. (Dados da Tabela 9)

Descreve-se nas Tabelas 50, 51 e 52 a análise do modelo misto de classificação dupla, segundo as saídas fornecidas pelo Proc MIXED, considerando-se os dados apresentados na Tabela 9.

Tabela 50. Ajuste do modelo misto de classificação dupla com casela vazia através do Proc MIXED do SAS e segundo o método REML

Histórico das Iterações da Estimação REML					
Iteração	Avaliações	Objetivo	Critério		
0	1	25.55295158			
1	3	21.33106714	0.00222342		
2	2	21.31523929	0.00028694		
3	1	21.31198256	0.00000321		
4	1	21.31194804	0.00000000		
Critério de Convergência Met.					
Estimativas dos Parâmetros de Covariância(REML)					
Parâm Cov	Razão	Estimativa	Erro Padrão	Z	Pr > Z
B	1.93549765	4.35180244	5.01684935	0.87	0.3857
A*B	0.00000000	0.00000000			
Residual	1.00000000	2.24841526	1.27826318	1.76	0.0786
Informação do modelo ajustado para Y					
Descrição		Valor			
Observações		10.0000			
Estimativa da variância		2.2484			
Estimativa do erro padrão		1.4995			
Log Verossimilhança REML		-18.0075			
Critério de Informação de Akaike		-21.0075			
Critério Bayesiano de Schwartz		-21.1266			
-2 Log Verossimilhança REML		36.0150			
Chi-Quadrado LRT Modelo Nulo		4.2410			
GL LRT Modelo Nulo		2.0000			
P-Valor LRT Modelo Nulo		0.1200			
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.71385320	1.35366405	5	1.27	0.2613
B 2	-2.03773154	1.38687887	5	-1.47	0.2017
B 3	0.32387833	1.45102937	5	0.22	0.8322
A*B 1 1	-0.00000000
A*B 1 2	0.00000000
A*B 1 3	0.00000000
A*B 2 1	0.00000000
A*B 2 2	-0.00000000
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de Variação	GLN	GLD	F Tipo III	Pr > F	
A	1	1	21.31	0.1358	

Tabela 51. Ajuste do modelo misto de classificação dupla com casela vazia através do Proc MIXED do SAS e segundo o método ML

Histórico das Iterações da Estimação ML					
Iteração	Avaliações	Objetivo	Critério		
0	1	25.68615918			
1	3	21.70765885	0.00012437		
2	1	21.70625800	0.00000044		
3	1	21.70625323	0.00000000		
Critério de Convergência Met.					
Estimativas dos Parâmetros de Covariância(MLE)					
Parâm Cov	Razão	Estimativa	Erro Padrão	Z	Pr > Z
B	1.38149385	2.68078946	2.65594529	1.01	0.3128
A*B	0.00000000	0.00000000			
Residual	1.00000000	1.94050047	1.02744193	1.89	0.0589
Informação do modelo ajustado para Y					
Descrição		Valor			
Observações		10.0000			
Estimativa da variância		1.9405			
Estimativa do erro padrão		1.3930			
Log Verossimilhança		-20.0425			
Critério de Informação de Akaike		-23.0425			
Critério Bayesiano de Schwartz		-23.4964			
-2 Log Verossimilhança		40.0850			
Chi-Quadrado LRT Modelo Nulo		3.9799			
GL LRT Modelo Nulo		2.0000			
P-Valor LRT Modelo Nulo		0.1367			
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.64556355	1.09795326	5	1.50	0.1942
B 2	-1.92881086	1.12880758	5	-1.71	0.1482
B 3	0.28324731	1.18811486	5	0.24	0.8210
A*B 1 1	-0.00000000
A*B 1 2	0.00000000
A*B 1 3	0.00000000
A*B 2 1	0.00000000
A*B 2 2	-0.00000000
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de Variação	GLN	GLD	F Tipo III	Pr > F	
A	1	1	25.81	0.1237	

Tabela 52. Ajuste do modelo misto de classificação dupla com casela vazia através do Proc MIXED e segundo o método MIVQUE(0)

Estimativas Parâmetros de Covariâncias (MIVQUE0)					
Parâm	Razão	Estimativa			
B	4.64285714	7.14285714			
A*B	0.00000001	0.00000000			
Resíduo	1.00000000	1.53846154			
Informação do modelo ajustado para Y					
Descrição	Valor				
Observações	10.0000				
Estimativa da variância	1.5385				
Estimativa do erro padrão	1.2403				
Log Verossimilhança REML	-18.3413				
Critério de Informação de Akaike	-21.3413				
Critério Bayesiano de Schwartz	-21.4605				
-2 Log Verossimilhança REML	36.6826				
Chi-Quadrado LRT Modelo Nulo	3.5733				
GL LRT Modelo Nulo	2.0000				
P-Valor LRT Modelo Nulo	0.1675				
Soluções para os efeitos aleatórios					
Parâmetro	Estimativa	Erro padrão	GL	T	Pr > T
B 1	1.82163638	1.63277499	5	1.12	0.3153
B 2	-2.22301389	1.65624339	5	-1.34	0.2372
B 3	0.40137751	1.70220977	5	0.24	0.8229
A*B 1 1	-0.00000001	0.00012403	5	-0.00	1.0000
A*B 1 2	0.00000001	0.00012403	5	0.00	1.0000
A*B 1 3	0.00000000	0.00012403	5	0.00	1.0000
A*B 2 1	0.00000001	0.00012403	5	0.00	0.9999
A*B 2 2	-0.00000001	0.00012403	5	-0.00	0.9999
Coeficientes do Tipo I para A			Coeficientes do Tipo III para A		
Parâmetro	linha 1		Parâmetro	linha 1	
INTERCEPTO	0		INTERCEPTO	0	
A 1	1		A 1	1	
A 2	-1		A 2	-1	
Testes para os efeitos fixos					
Causa de Variação	GLN	GLD	F Tipo III	Pr > F	
A	1	1	28.80	0.1173	

Fato:

Verifica-se através das Tabelas 47 à 52 que as estimativas REML, MIVQUE0 e ML para os componentes de variâncias, são as mesmas obtidas pelo Proc Varcomp.

5.2 Ajuste do modelo linear misto de classificação dupla, através de outros sistemas estatísticos

Apresenta-se, através das Tabelas 53 a 72, o ajuste dos dados referentes as Tabelas 8 e 9, com base nas saídas fornecidas pelos procedimentos disponíveis nos pacotes estatísticos: BMDP, HARVEY, MINITAB, NTIA, REML, SANEST, SAS, SPSS, STATGRAPHICS e STATISTICA. Para cada procedimento adotou-se um conjunto de sintaxe básica para a obtenção do ajuste do modelo linear misto.

5.2.1 Ajuste do modelo linear misto de classificação dupla, através do BMDP-3V

Dentre o conjunto de programas disponíveis pelo sistema computacional BMDP utilizou-se o procedimento BMDP-3V. Trata-se de um programa específico para a análise de modelos mistos desbalanceados, o qual é baseado nos métodos ML (Máxima Verossimilhança) e REML (Máxima Verossimilhança Restrita). Neste estudo considerou-se apenas as declarações básicas a seguir:

i) conjunto básico de sintaxe para ajustar o modelo misto de classificação dupla através do método ML:

```

/problem title='Analysis of a mixed model
/input variables=3. format=free.
/variable name = A, B, Y.
/group codes(A) = 1, 2.
/design dependent = Y.
    fixed = A.
    fname = A.
    random = B.
    random = A, B.
    rname = B, 'A*B'.
    method = ml.

```

```

/ HIPÓTESES fixed = 1.
/ HIPÓTESES random =1.
/ HIPÓTESES random =2.
/ end.

```

i) conjunto básico de sintaxe para ajustar o modelo misto de classificação

dupla através do método REML:

```

/ problem title='Analysis of a mixed model
/ input variables=3. format=free.
/ variable name = A, B, Y.
/ group codes(A) = 1, 2.
/ design dependent = Y.
      fixed = A.
      fname = A.
      random = B.
      random = A, B.
      rname = B, 'A*B'.
      method = reml.
/ end

```

A seguir apresenta-se os resultados obtidos para o ajuste do modelo misto considerado, utilizando-se os dados descritos nas Tabelas 8 e 9.

5.2.1.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

O procedimento BMDP-3V apresenta, através dos comandos básicos considerados, o teste de hipóteses baseado na razão de verossimilhança para o ajuste através do método ML, e para o ajuste através do método REML, apresenta também o teste para os efeitos fixos com base na matriz de variâncias e covariâncias assintóticas (F assintótico), ou seja, apresenta o teste de hipóteses Tipo III do SAS através do Proc MIXED conforme mostram as Tabelas 53 e 54.

Tabela 53. Ajuste do modelo misto sem casela vazia através do sistema BMDP-3V e segundo do método ML.

VARIÁVEL DEPENDENTE Y					
PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	EST/D.PADRÃO	PROBABILIDADE	
(TEORIA ASSINTÓTICA).					
VAR. ERR.	5.402	1.753			
CONSTANTE	41.861	0.850	49.249	0.000	
A	2.611	0.498	5.239	0.000	
B	1.422	1.780			
A*B	0.000	0.000			
-2*LOG (MÁXIMA VEROSSIMILHANÇA)			102.7660000		
MODELO RESTRITO - HIPÓTESE NUMBER 1				$H_0: \alpha_i = 0$	
PARÂMETRO	ESTIMATIVA	PADRÃO ERRO	EST/D.PADRÃO	PROBABILIDADE	
(TEORIA ASSINTÓTICA).					
VAR. ERR.	6.229	2.202			
CONSTANT	41.780	1.230	33.977	0.000	
A	0.000	0.000			
B	0.000	0.000			
A*B	7.348	5.271			
-2*LOG (MÁXIMA VEROSSIMILHANÇA)			112.6777000		
TEST RAZÃO DE VEROSSIMILHANÇA					
CHI-QUADRADO = 9.912					
GRAUS DE LIBERDADE 1					
PROBABILIDADE 0.002					
<div style="border: 1px solid black; border-radius: 15px; padding: 5px; display: inline-block;"> $9,912 = 112,6777 - 102,7660$ $g.1 = 3-2=1$ </div>					
MATRIZ DOS PARÂMETROS DE VARIÂNCIAS E COVARIÂNCIAS ASSINTÓTICAS					
	VAR. ERR.	CONSTANTE	A	B	A*B
	1	2	3	4	5
VAR. ERR.	1	3.0717			
CONSTANTE	2	0.0000	0.7225		
A	3	0.0000	0.0233	0.2484	
B	4	-0.4194	0.0000	0.0000	3.1688
A*B	5	0.0000	0.0000	0.0000	0.0000
CASELA GRUPO DE VARIÁVEIS DUMMY					
A					
1	*1	1.	1.		
2	*2	1.	-1.		
CASELA MÉDIA MÉDIA DEV.PADRÃO.					
OBSERVADA PRED. PRED.					
1		44.6000	44.4719	1.0087	
2		39.2500	39.2500	0.9614	
MATRIZ VARIÂNCIA E COVARIANCIA DAS MÉDIAS DE PREDIÇÃO					
	1	2			
1	1	1.0175			
2	2	0.4741	0.9243		
TESTES PAREADOS PARA AS MÉDIAS DE PREDIÇÃO					
	1	2			
1	1	0.0000			
2	2	-5.2388	0.0000		

Fato:

Verifica-se que o sistema BMDP-3V (1985) fornece as estimativas dos efeitos fixos e aleatórios, os desvios padrões, a matriz de covariâncias assintóticas dos parâmetros e o teste da razão de verossimilhança para todos os efeitos considerados no modelo a Tabela 53 apresenta apenas o teste para a hipótese $H_0: \alpha_i = 0$. Comparando-se a Tabela 48 com a Tabela 53, observa-

se que as saídas apresentadas pelo Proc MIXED do SAS contêm mais informações para o ajuste do modelo misto, nesse caso.

Tabela 54. Ajuste do modelo misto sem casela vazia através do sistema BMDP-3V e segundo do método REML

VARIÁVEL DEPENDENTE Y					
PARÂMETRO	ESTIMATIVA	PADRÃO	EST/D.PADRÃO	PROBABILIDADE (TEORIA ASSINTÓTICA).	
	ERRO				
VAR. ERR.	5.706	1.900			
CONSTANTE	41.852	1.038	40.314	0.000	
A	2.602	0.512	5.078	0.000	
B	2.446	3.214			
A*B	0.000	0.000			
-2*LOG (MÁXIMA VEROSSIMILHANÇA)			94.42317000		
MATRIZ DOS PARÂMETROS DE VARIÂNCIAS E COVARIÂNCIAS ASSINTÓTICAS					
	VAR. ERR.	CONSTANTE	A	B	A*B
	1	2	3	4	5
VAR. ERR.	1	3.6118			
CONSTANTE	2	0.0000	1.0777		
A	3	0.0000	0.0247	0.2625	
B	4	-0.4655	0.0000	0.0000	10.3298
A*B	5	0.0000	0.0000	0.0000	0.0000
TESTES PARA OS EFEITOS FIXOS BASEADOS NA MATRIZ DE VARIÂNCIAS E COVARIÂNCIAS ASSINTÓTICAS					
CAUSAS DE VARIAÇÃO	ESTATÍSTICA-F	GRAUS DE LIBERDADE		PROBABILIDADE	
CONSTANTE	1625.24	1	20	0.00000	
A	25.78	1	20	0.00006	

Fato:

Comparando-se a Tabela 47 com a Tabela 54, verifica-se que esse sistema apresenta para o efeito A fixo o teste do Tipo III do Proc MIXED pode-se observar que o nível descritivo obtido pelo BMDP-3V, com base em 1 e 20 graus de liberdade é diferente do obtido através do Proc MIXED do SAS, com base em 1 e 2 graus de liberdade.

5.2.1.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Descreve-se, através das Tabelas 55 e 56, as saídas da análise do modelo misto de classificação dupla, segundo o procedimento BMDP-3V(1985)

considerando-se os dados apresentados na Tabela 9.

Tabela 55. Ajuste do modelo misto com casela vazia através do sistema BMDP-3V e segundo do método ML

VARIÁVEL DEPENDENTE: Y					
PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	EST/D.PADRÃO	PROBABILIDADE (TEORIA ASSINTÓTICA).	
VAR. ERR.	1.940	1.036			
CONSTANTE	40.057	1.053	38.048	0.000	
A	2.558	0.503	5.080	0.000	
B	2.681	2.719			
A*B	0.000	0.000			
-2*LOG (MÁXIMA VEROSSIMILHANÇA) = 40.08503000					
MATRIZ DOS PARÂMETROS DE VARIÂNCIA E COVÂRIÂNCIAS ASSINTÓTICAS					
	VAR. ERR.	CONSTANT	A	B	A*B
	1	2	3	4	5
VAR. ERR.	1	1.0739			
CONSTANTE	2	0.0000	1.1084		
A	3	0.0000	-0.0325	0.2535	
B	4	-0.3418	0.0000	0.0000	7.3933
A*B	5	0.0000	0.0000	0.0000	0.0000
MODELO RESTRITO - HIPÓTESE NUMBER 1 ▶ $H_0: \alpha_i = 0$					
PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	EST/D.PADRÃO	PROBABILIDADE (TEORIA ASSINTÓTICA).	
	ERR.VAR.	2.318	1.464		
CONSTANTE	40.548	1.454	27.881	0.000	
A	0.000	0.000			
B	0.000	0.000			
A*B	9.286	6.729			
-2*LOG (MÁXIMA VEROSSIMILHANÇA) = 47.55829000					
MATRIZ DOS PARÂMETROS DE VARIÂNCIA E COVÂRIÂNCIAS ASSINTÓTICAS					
	ERR.VAR.	CONSTANTE	A	B	A*B
	1	2	3	4	5
ERR.VAR.	1	2.1443			
CONSTANTE	2	0.0000	2.1151		
A	3	0.0000	0.0000	0.0000	
B	4	0.0000	0.0000	0.0000	0.0000
A*B	5	-1.1710	0.0000	0.0000	45.2765
TESTE RAZÃO DE VEROSSIMILHANÇA					
CHI-QUADRADO = 7.473					
GRAUS DE LIBERDADE = 1					
PROBABILIDADE = 0.006					
CASELA	MÉDIA OBSERVADA	MÉDIA PRED.	DEV.PADRÃO. PRED.		
1	43.0000	42.6142	1.1388		
2	37.0000	37.4991	1.1945		
MATRIZ VARIÂNCIA E COVARIÂNCIA DAS MÉDIAS DE PREDIÇÃO					
	1	2			
	1	2			
1	1	1.2969			
2	2	0.8549	1.4267		
TESTES PAREADOS PARA MÉDIAS DE CASELAS DE PREDIÇÃO					
	1	2			
	1	2			
1	1	0.0000			
2	2	-5.0801	0.0000		
CASELA	GRUPO DE VARIÁVEIS DUMMY				
	A				
1	*1	1. 1.			
2	*2	1. -1			

Tabela 56 . Ajuste do modelo misto sem casela vazia através do sistema BMDP-3V e segundo do método REML

VARIÁVEL DEPENDENTE Y					
PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	EST/D.PADRÃO	PROBABILIDADE (TEORIA ASSINTÓTICA).	
VAR. ERR.	2.248	1.278			
CONSTANTE	40.065	1.305	30.713		0.000
A	2.528	0.548	4.617		0.000
B	4.352	5.017			
A*B	0.000	0.000			

-2*LOG (MÁXIMA VEROSSIMILHANÇA) 32.79609000

MATRIZ DOS PARÂMETROS DE VARIÂNCIA E COVÂRIÂNCIAS ASSINTÓTICAS

	VAR. ERR.	CONSTANTE	A	B	A*B
	1	2	3	4	5
VAR. ERR.	1	1.6339			
CONSTANTE	2	0.0000	1.7017		
A	3	0.0000	-0.0411	0.2998	
B	4	-0.3041	0.0000	0.0000	25.1749
A*B	5	0.0000	0.0000	0.0000	0.0000

TESTES PARA OS EFEITOS FIXOS BASEADOS NA MATRIZ DE VARIÂNCIAS E COVARIÂNCIAS ASSINTÓTICAS
CAUSAS DE VARIAÇÃO ESTATÍSTICA-F GRAUS DE LIBERDADE PROBABILIDADE

CONSTANTE	943.26	1	8	0.00000
A	21.31	1	8	0.00172

Fato:

Comparando-se as Tabelas 50 e 56, observa-se que o teste para o efeito A fixo é o do Tipo III fornecido pelo Proc MIXED. Faz-se as mesmas observações quanto ao nível descritivo do teste.

5.2.2 Ajuste do modelo linear misto de classificação dupla, através do HARVEY - LSMLMW

HARVEY - LSMLMW é o procedimento que permite analisar modelos de efeitos fixos, aleatórios ou mistos e reúne um conjunto de sete modelos, dentre esses selecionou-se os modelos 1 e 6, os quais são indicados pela literatura como mais adequados para analisar o modelo misto clássico de dois fatores. Neste estudo considerou-se apenas as declarações básicas a seguir:

```
TITLE 'ANALISE MISTO 1';
INPUT A 1 B 4 Y 7-8;
CLASSES A B;
MODEL1 Y= A B A*B / VARCOMP =(B, A*B);
```

```
TITLE 'ANALISE MISTO 1';
INPUT A 1 B 4 Y 7-8;
CLASSES A B;
MODEL6 Y= B A B*A;
```

A seguir, apresenta-se os resultados obtidos para o ajuste do modelo misto considerado, utilizando-se os dados descritos nas Tabelas 8 e 9.

5.2.2.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

Conforme mostra a Tabela 57, esse sistema através do modelo 1, apresenta as somas de quadrados do Tipo III, correspondentes as do SAS/GLM. Emite as estimativas dos componentes de variância obtidos através da análise de variância, as quais equivalem às obtidas pela opção **Test** do

Proc GLM/SAS. Para o modelo 6 o sistema emite as somas de quadrados do Tipo II, e as estimativas dos componentes de variâncias obtidos direto do quadro da análise de variância, resultados que correspondem aos obtidos pelo SAS/GLM, conforme mostra a Tabela 58.

Tabela 57. Ajuste do modelo misto sem casela vazia através do sistema HARVEY-modelo 1

CAUSA DE VARIAÇÃO		G.L		SQ.	QM	F	PROB
TOTAL	22	H_0	$R()$	306.772727			
TOTAL REDUCTION	6			206.606061	34.434343	5.500	.0029
MU-YM	1			.330707	.330707	.053	.8211
A	1	$H_0^{(3)}$	$R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha \hat{\beta})$	141.066667	141.066667	22.533	.0002
B	2	$H_0^{(7)}$	$R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha \hat{\beta})$	49.077694	24.538847	3.920	.0412
A X B	2			2.626566	1.313283	.210	.8130
RESÍDUO	16	$H_0^{(9)}$	$R(\alpha \hat{\beta} \mu, \alpha, \beta)$	100.166667	6.260417		

MÉDIA = 41.68182 ERRO DESVIO PADRÃO = 2.50208 CV = 6.00 R SQUARED = .673 R = .821

ESTIMATIVAS DOS COMPONENTES DE VARIÂNCIA DIRETO DA ANÁLISE
 K COMPONENTES EFEITOS ALEATÓRIOS (B) = 7.2180 GRAUS DE LIBERDADE = 2.
 SQ, CP, QM, CPM, COMPONENTES DE VARIÂNCIAS E COVÂRIÂNCIAS

TRB	LIN	COL	VAR	DEP	VAR	DEP	SQ OU CP	QM OU COV	COMPONENTES
1	1	1	Y		Y		49.07769424	24.53884712	2.53232422

K COMPONENTES EFEITOS ALEATÓRIOS (A * B) = 7.2180 GRAUS DE LIBERDADE = 2.
 SS, CP, MS, MCP, COMPONENTES DE VARIÂNCIAS E COVÂRIÂNCIAS

TRB	LIN	COL	VAR	DEP	VAR	DEP	SQ OU PC	QM OU COV	COMPONENTES
1	1	1	Y		Y		2.62656642	1.31328321	-.68538411

Fatos :

1) apresenta as somas de quadrados do Tipo III conforme o Proc GLM, sem a opção **Test**, o teste F apresentado utiliza o resíduo como termo do erro para o efeito fixo, fornece a mesma análise da Tabela 20.

2) apresenta as estimativas dos componentes de variância através do método dos momentos segundo o modelo 1 de Hocking (1973) e formulação 2 de Schwarz (1993), conforme o proc GLM/SAS, segundo a opção **Test**. Assim fornece informações suficientes para o analista obter o teste do Tipo III, para o efeito A fixo.

A Tabela 58, descreve-se os resultados obtidos através do ajuste do modelo 6.

Tabela 58. Ajuste do modelo misto sem casela vazia através do sistema HARVEY-modelo 6

ANÁLISE DE VARIÂNCIA DE MÍNIMOS QUADRADOS						
Y						
CAUSA DE VARIAÇÃO	D.F.	SOMAS DE QUADRADOS	QUADRADOS MÉDIOS	F	PROB	ERRO
B	2	47.856767	23.928383	3.822	.0440	RESÍDUO
A	1	144.421053	144.421053	109.969	.0090	B X A
B X A	2	2.626566	1.313283	.210	.8130	REMNR
RESÍDUO	16	100.166667	6.260417			

*** TESTES DE SIGNIFICÂNCIA PARA TODOS OS CONJUNTOS DE EFEITOS ONDE O RESÍDUO NÃO É USADO COMO O TERMO DO ERRO SÃO APROXIMADOS COM DADOS DESBALANCEADOS OU NA ANÁLISE DE COVARIÂNCIA

ESTIMATIVAS DOS COMPONENTES DE VARIÂNCIA DIRETO DA ANÁLISE

K COMPONENTES EFEITOS ALEATÓRIOS (BXA)= 3.6090 GRAUS DE LIBERDADE = 2.

SQ, PC, QM, PCM, COMPONENTES DE VARIÂNCIAS E COVARIÂNCIAS

TR	LIN	COL	VAR DEP	VAR DEP	SQ OU PC	QM OU COV	COMPONENTES
4	1	1	Y	Y	2.62656642	1.31328321	-1.37076823

K VALORES (B) SÃO: K2= 3.6910 K3= 7.3000 GRAUS DE LIBERDADE = 2.

ESTIMATIVA NEGATIVA DE COMPONENTES DE VARIÂNCIA PARA B X A

SQ, PC, QM, PCM, COMPONENTES DE VARIÂNCIAS E COVARIÂNCIAS

TRAB	ROW	COL	VAR DEP	VAR DEP	SQ OU PC	QM OU COV	COMPONENTES
4	1	1	Y	Y	47.85676692	23.92838346	2.42026942

Fatos:

1) apresenta as somas de quadrados do Tipo II do Proc GLM/SAS, o teste F aproximado é determinado para o efeito utilizando-se o QM(A*B) como o termo do erro.

2) Apresenta as estimativas dos componentes de variância, através do método dos momentos e segundo o modelo de Scheffé (1959), modelo 1 de Hocking (1973) ou formulação 2 de Schwarz (1993).

5.2.2.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Descreve-se, a seguir, através da Tabela 59 as saídas da análise do modelo misto de classificação dupla, segundo o modelo 1 considerando-se os dados apresentados na Tabela 9. Em decorrência da estrutura de dados da Tabela 9, desbalanceado com casela vazia e por ter poucas observações não foi possível obter a entrada de dados para obter-se o ajuste através do modelo 6 que requer que a primeira linha e a última coluna sejam completas.

Tabela 59. Ajuste do modelo misto com casela vazia através do sistema HARVEY-modelo 1

ANÁLISE DE VARIÂNCIA DE MÍNIMOS QUADRADOS						
CAUSA DE VARIACÃO	G.L.	Y	SQ.	QM	F	PROB
TOTAL	10	$H_0 \quad R()$		138.000000		
TOTAL REDUÇÃO	5		126.000000	25.200000	10.500	.0110
MU-YM	1		.235294	.235294	.098	.7668
A	1	$H_0^{(2)} \quad R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha \hat{\beta})$	42.857143	42.857143	17.857	.0083
B	2	$H_0^{(2)} \quad R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha \hat{\beta})$	27.578947	13.789474	5.746	.0506
A X B	1		1.714286	1.714286	.714	.4366
RESÍDUO	5	$H_0^{(2)} \quad R(\alpha \beta \mu, \alpha, \beta)$	2.000000	2.400000		

MÉDIA = 40.00000 DESVIO PADRÃO ERRO = 1.54919 CV = 3.87 R QUADRADO = .913 R = .956

ESTIMATIVAS DOS COMPONENTES DE VARIÂNCIA DIRETO DA ANÁLISE

K COMPONENTES EFEITOS ALEATÓRIOS (B) = 2.6526 GRAUS DE LIBERDADE = 2.
SS, CP, MS, MCP, COMPONENTES DE VARIÂNCIAS E COVÂRIÂNCIAS

TRB	LIN	COL	VAR DEP	VAR DEP	SQ OU PC	QM OU COV	COMPONENTES
1	1	1	Y	Y	27.57894737	13.78947368	4.29365079

K COMPONENTES EFEITOS ALEATÓRIOS (A *B) = 3.4286 GRAUS DE LIBERDADE = 1.
SQ, PC, QM, PCM, COMPONENTES DE VARIÂNCIAS E COVÂRIÂNCIAS

JOB	ROW	COL	VAR DEP	VAR DEP	SQ OU PC	QM OU COV	COMPONENTES
1	1	1	Y	Y	1.71428571	1.71428571	-20000000

Fato:

Pode-se observar os mesmos fatos da Tabela 57.

5.2.3 Ajuste do modelo linear misto de classificação dupla, através do MINITAB - GLM

MINITAB - GLM é o procedimento destinado para analisar o modelo linear geral balanceado ou desbalanceado. Neste estudo considerou-se o conjunto básico de comandos:

```
MTB > Retrieve 'A:\MISTOI.MTP';
SUBC> Portable.
Retrieving worksheet from file: A:\MISTOI.MTP
Worksheet was saved on 4/ 1/1994
MTB > GLM 'Y' = a b a*b;
SUBC> Test A / a*b.
```

5.2.3.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

O procedimento GLM apresenta, através da opção **Test**, as somas de quadrados e hipóteses do Tipo I e do Tipo III do Proc GLM/SAS, conforme mostra a Tabela 60.

Tabela 60. Ajuste do modelo misto sem casela vazia através do sistema MINITAB-GLM

Análise de variância para: Y

Causa	V	GL	H ₀	R ()	SQ	SEQ.	SQ. AJ	QM. AJ.	F	P
A	1				156.123		141.067	141.067	22.53	0.000
B	2		H ₀ ⁽¹⁾	R (α μ, β)	47.857		49.078	24.539	3.92	0.041
A*B	2		H ₀ ^(ε)	R (β μ, α)	2.627		2.627	1.313	0.21	0.813
Error	16		H ₀ ^(ε)	R (α β μ, α, β)	100.167		100.167	6.260		
Total	21				306.773					

OBSERVAÇÃO NÃO USUAL PARA: Y

Obs	Y	ajus.	D.p.aju	Resíduo	D.p.res
4	53.0000	46.7500	1.2510	6.2500	2.88R

R representa uma observação com um alto valor residual padronizado

Teste-f com denominador: A*B

Denominador QM = 1.3133 com 2 graus de liberdade

Numerador	DF	QM Seq	F	P
A	1	156.1	118.88	0.008

Fatos:

1) verifica-se que esse sistema computacional fornece as hipóteses do Tipo I e III, segundo a Tabela 20, ou seja, conforme a análise através do Proc GLM do SAS, sem a opção **Test** no comando **Random**.

2) verifica-se que para o efeito A fixo, apresenta o teste para a hipótese do Tipo I, utilizando-se o QM(A*B) como denominador conforme o sistema computacional HARVEY- modelo 1 conforme Tabela 57.

5.2.3.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Descreve-se através da Tabela 61 as saídas da análise do modelo misto de classificação dupla, segundo o procedimento GLM e a opção **Test**, considerando-se os dados apresentados na Tabela 9. Nesse caso de desbalanceamento o procedimento fornece apenas as somas de quadrados do Tipo I do Proc GLM/SAS, e não realiza o teste para os efeitos fixos do modelo.

Tabela 61. Ajuste do modelo misto com casela vazia através do sistema MINITAB-GLM

Análise de Variância para : YY

Causa de V	Modelo	gl	gl redução	SQ Seq
AA	1	1	1	90.000
BB	2	2	2	34.286
AA*BB	2	1+	1+	1.714
Erro	4	5	5	12.000
Total	9	9	9	138.000

H_0	$R()$
$H_0^{(2)}$	$R(\alpha \mu, \beta, \alpha\beta)$
$H_0^{(7)}$	$R(\beta \mu, \alpha, \alpha\beta)$
$H_0^{(9)}$	$R(\alpha\beta \mu, \alpha, \beta)$

+ posto deficiente devido a caselas vazias, desbalanceamento aninhado colineariedade, ou uma covariável não declarada.

5.2.4 Ajuste do modelo linear misto de classificação dupla, através do NTIA-MODLIN

O procedimento MODLIN é o módulo do sistema computacional NTIA que permite a análise do modelo linear geral de efeitos fixos balanceado, ou desbalanceado. Considera-se aqui apenas o conjunto básico de comandos, a seguir:

```
GENESE MISTO
NUM A B Y;
ARQUIVO M=ABREF(MISTO) A B Y;
{
LEIAF(M);
}
MODLIN MISTO
MOD Y = A [A*B] B [A*B] A*B;
}
```

5.2.4.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

O procedimento MODLIN apresenta os testes de hipóteses do Tipo I (sequencial) e do Tipo III do SAS (parciais) conforme mostra a Tabela 62.

Tabela 62. Ajuste do modelo misto sem casela vazia através do sistema NTIA- MODLIN

F. de variação gl		Quadro de análise de variância Variável Dependente : Y				
		H_0 R()	SQ Seqüencial	QM	Valor F	PR > F
A	1	$H_0^{(1)} R(\alpha \mu)$	156.12272727	156.12272727	118.8797	0.008
B	2	$H_0^{(I-B)} R(\beta \mu, \alpha)$	47.85676692	23.92838346	18.2203	0.052
A*B	2	$H_0^{(I-AB)} R(\alpha\beta \mu, \alpha, \beta)$	2.62656642	1.31328321	0.2098	0.813
Resíduo	16		100.16666667	6.26041667		
Total	21		306.77272727			

F. de variação gl		H_0 R()	SQ Parcial	QM	Valor F	PR > F
A	1	$H_0^{(3)} R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha\hat{\beta})$	141.06666667	141.06666667	107.4153	0.009
B	2	$H_0^{(III-B)} R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha\hat{\beta})$	49.07769424	24.53884712	18.6851	0.051
A*B	2	$H_0^{(III-AB)} R(\alpha\hat{\beta} \hat{\mu}, \hat{\alpha}, \hat{\beta})$	2.62656642	1.31328321	0.2098	0.81

Fato:

O sistema computacional NTIA fornece o teste F exato considerando como denominador o $QM(A*B)$, para testar os efeitos de linhas (fixos) e de colunas (aleatórios) e testa a interação com o resíduo.

5.2.4.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Descreve-se através da Tabela 63 as saídas da análise do modelo misto de classificação dupla, segundo o procedimento MODLIN considerando-se os dados apresentados na Tabela 9. Nesse caso de desbalanceamento o procedimento fornece apenas as somas de quadrados do Tipo I do SAS (sequencial).

Tabela 63. Ajuste do modelo misto com casela vazia através do sistema NTIA- MODLIN

Quadro de análise de variância		Variável Dependente : Y			
F. de v	gl	SQ Seqüencial	QM	Valor F	PR > F
A	1	90.00000000	90.00000000	* 52.5000	0.087
B	2	34.28571429	17.14285714	10.0000	0.218
A*B	1	1.71428571	1.71428571	0.7143	0.437
Resíduo	5	12.00000000	2.40000000		
Total	9	138.00000000			

Atenção: A matriz $X'X$ e' singular. Coluna associada ao efeito $A*B$ e' uma combinação linear de colunas associadas a efeitos que o precedem no modelo. Nesta situação não existe solução única para as equações normais, mesmo com a imposição das restrições nos parâmetros, e as somas de quadrados parciais não são fornecidas

Fato:

Verifica-se que o sistema computacional NTIA (1995), apresenta para o modelo misto em estudo o ajuste de mínimos quadrados, apresenta o teste F exato, não considera as expressões dos valores esperados dos QM, conforme a Tabela 31, a qual apresenta para o efeito fixo de linha o valor correto para o teste F, ou seja, ($F = 15,12$).

5.2.5 Ajuste do modelo linear misto de classificação dupla, através do REML

REML é um sistema destinado a análise de dados desbalanceados para a obtenção das estimativas dos componentes de variância através do método da Máxima Verossimilhança Restrita. Neste estudo foram consideradas as seguintes declarações:

i) conjunto básico de sintaxe para ajustar o modelo misto de classificação dupla com todas as caselas ocupadas através da rotina REML.

```
'title'
modelo misto1
'unit' 22
'factor'
a 2 *
  FACTOR a    CREATED
b 3 *
  FACTOR b    CREATED
'variate' y
'random' b + a.b
'fixed' a
RANDOM MODEL READ
'dependent' y
FIXED MODEL READ
'readfree' 1 y
43 45 46 53 41 42 44 42 44 46 40 40 40 43 35 37 40 40 37 39 40 40
DATA SET READ
'readfree' 1 a
1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
DATA SET READ
'readfree' 1 b
1 1 1 1 2 2 2 3 3 3 1 1 1 1 2 2 2 2 3 3 3 3
DATA SET READ
'print' 10
'components'
'tables' a b a.b
'endprint'
'go'
```

ii) conjunto básico de sintaxe para ajustar o modelo misto de classificação dupla com a casela vazia através da rotina REML.

```
'title'
modelo misto2
'unit' 10
'factor'
a 2 *
FACTOR a      CREATED
b 3 *
FACTOR b      CREATED
'variate' y
'random' b + a.b
'fixed' a
RANDOM MODEL READ
'dependent' y
FIXED MODEL READ
'readfree' 1 y
43 45 41 42 44 40 40 35 33 37
DATA SET READ
'readfree' 1 a
1 1 1 1 1 2 2 2 2 2
DATA SET READ
'readfree' 1 b
1 1 2 3 3 1 1 2 2 2
DATA SET READ
'print' 8
'components'
'tables' a b a.b
'endprint'
'go'
```

5.2.5.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

O procedimento REML apresenta, através dos comandos considerados, as estimativas eficientes dos parâmetros de efeitos fixos, as estimativas dos componentes de variância e fornece o estudo de predição para os efeitos aleatórios e para os valores observados. Conforme mostra a Tabela 64.

Tabela 64. Ajuste do modelo misto sem casela vazia através do sistema REML

REML - COMPONENTES DE VARIÂNCIA RESÍDUO DE MÁXIMA VEROSSIMILHANÇA
Nº DE ITERAÇÕES 3

ESTIMATIVAS COMPONENTES DE VARIÂNCIA

	b	a b	SIGMA QUADRADO
	3.070	0.0006266	6.266
ES	4.025	1.835	2.214

TABELA (B.L.U.E.'S) PARA a

1	2	MARGINAL
0.000	-5.199	-2.599

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
1.074

TABELA (B.L.U.P.'S) PARA b

1	2	3	MARGINAL
1.514	-1.255	-0.2594	4.967E-7

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
MÉDIA 1.159 MÁXIMO 1.177 MÍNIMO 1.150

TABELA (B.L.U.P.'S) PARA a POR b

	1	2	3	MARGINAL
1	0.0003148	-0.0002581	-5.673E-5	2.547E-11
2	-5.657E-6	1.894E-6	3.763E-6	-8.489E-12
MARGINAL	0.0001546	-0.0001281	-2.648E-5	8.489E-12

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
0.03540

Fato:

Comparando-se as saídas do sistema REML conforme mostra a Tabela 64 com as saídas obtidas através dos sistemas BMDP e Proc MIXED/SAS conforme mostram as Tabelas 54 e 47 respectivamente, observa-se diferentes valores para as estimativas REML dos componentes de variâncias em decorrência do procedimento iterativo. O REML utiliza o método de escore de Fisher, já o BMDP utiliza uma combinação da função escore de Fisher com o algoritmo de Newton-Raphson e o proc. MIXED/SAS utiliza o algoritmo de Newton-Raphson. Faz-se as mesmas observações para a Tabela 65.

5.2.5.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Descreve-se na Tabela 65 a seguir os resultados obtidos através da rotina REML para o ajuste dos dados da Tabela 9.

Tabela 65. Ajuste do modelo misto com casela vazia através do sistema REML.

REML - COMPONENTES DE VARIÂNCIA RESÍDUO DE MÁXIMA VEROSSIMILHANÇA
N.º DE ITERAÇÕES 5

ESTIMATIVAS COMPONENTES DE VARIÂNCIA

	b	a b	SIGMA QUADRADO
	4.413	0.0002276	2.276
EP	5.501	2.009	1.435

TABELA (B.L.U.E.'S) PARA a

1	2	MARGINAL
0.000	-5.055	-2.528

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
1.102

TABELA (B.L.U.P.'S) PARA b

1	2	3	MARGINAL
1.714	-2.038	0.3240	1.987E-8

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
MÉDIA 1.228 MÁXIMO 1.377 MÍNIMO 1.033

TABELA (B.L.U.P.'S) PARA a POR b

1	2	3	MARGINAL
1	-6.128E-5	4.456E-5	1.672E-5
2	0.0001497	-0.0001497	0.000
MARGINAL	4.422E-5	-5.258E-5	8.359E-6

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS PARES
MÉDIA 0.01925 MÁXIMO 0.02134 MÍNIMO 0.01509

ERRO PADRÃO PARA AS DIFERENÇAS ENTRE OS MESMO NÍVEL DOS FATORES

	a	b
MÉD	0.01925	0.01925
MAX	0.02134	0.02133
MIN	0.01509	0.01509

Fato: As mesmas observações da Tabela 64.

5.2.6 Ajuste do modelo linear misto de classificação dupla, através do SANEST

Nesse estudo utilizou-se o SANEST-MÓDULO 3 que se destina a análise de variância de dados de classificação dupla não-balanceado, para modelo lineares de efeitos fixos ou mistos com e sem interação. Descreve-se, a seguir, os resultados obtidos no ajuste dos dados das Tabelas 8 e 9.

5.2.6.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

Através de declarações específicas nas telas de comandos o procedimento SANEST- Módulo 3, apresenta para dados desbalanceados e com todas as caselas ocupadas as hipóteses do Tipo II do SAS, conforme mostra a Tabela 66.

Tabela 66. Ajuste do modelo misto sem casela vazia através do sistema Sanest- modulo 3

QUADRO DA ANÁLISE DA VARIÂNCIA						
CAUSAS DA VAR.	G.L.	H_0	$R(-)$	S.Q.	Q.M.	VALOR F PROB.>F
LINHAS AJUST.	1	$H_0^{(2)}$	$R(\alpha \mu, \beta)$	144.4210526	144.4210526	109.9831 0.00652
COLONAS AJUST.	2	$H_0^{(\mu-\beta)}$	$R(\beta \mu, \alpha)$	47.8567669	23.9283835	18.2225 0.05099
LIN*COL AJUST.	2	$H_0^{(\mu-\beta)}$	$R(\beta \mu, \alpha)$	2.6262409	1.3131204	0.2097 0.81448
RESÍDUO	16	$H_0^{(I-AB)}$	$R(\alpha\beta \mu, \alpha, \beta)$	100.1669922	6.2604370	
TOTAL	21			306.7727273		
MÉDIA GERAL AJUSTADA = 41.828941						
COEFICIENTE DE VARIÂÇÃO = 5.982 %						
S.Q. LINHAS ANO AJUSTADA = 156.1227273						
S.Q. COLUNAS ANO AJUSTADA = 59.5584416						
ANÁLISE DA VARIÂNCIA FIXANDO NÍVEIS						
CAUSAS DA VARIAÇÃO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F	
LIN/COL 1	1	72.0000000	72.0000000	11.5008	0.00107	
LIN/COL 2	1	32.1901507	32.1901507	5.1418	0.01852	
LIN/COL 3	1	42.8571429	42.8571429	6.8457	0.00725	
COL/LIN 1	2	34.9830078	17.4915039	2.7940	0.08975	
COL/LIN 2	2	15.5000000	7.7500000	1.2379	0.31658	
RESÍDUO	16	100.1669922	6.2604370			

Fato :

Para o efeito A fixo, apresenta as somas de quadrados e hipóteses do Tipo II do Proc GLM/SAS com a opção **Test**, conforme a Tabela 22.

5.2.6.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Através de declarações específicas nas telas de comandos o procedimento SANEST-Módulo 3 apresenta para o caso desbalanceado com casela vazia, as hipóteses do Tipo II do SAS, conforme mostra a Tabela 67.

Tabela 67. Ajuste do modelo misto com casela vazia através do sistema Sanest- modulo 3

QUADRO DA ANALISE DA VARIÂNCIA							
CAUSAS DA VARIAÇÃO	G.L.	H_0	$R()$	S.Q.	Q.M.	VALOR F	PROB.>F
LINHAS AJUST.	1	$H_0^{(2)}$	$R(\alpha \mu, \beta)$	41.2857143	41.2857143	24.0833	0.13434
COLUNAS AJUST	2	$H_0^{(U-B)}$	$R(\beta \mu, \alpha)$	34.2857143	17.1428571	10.0000	0.22562
LIN*COL AJUST.	1			1.7142857	1.7142857	0.7143	0.55961
RESÍDUO	5	$H_0^{(U-AB)}$	$R(\alpha\beta \mu, \alpha, \beta)$	12.0000000	2.4000000		
TOTAL	9			138.0000000			

MEDIA GERAL AJUSTADA =40.095238

COEFICIENTE DE VARIAÇÃO = 3.864 %

S.Q. LINHAS NÃO AJUSTADA = 90.0000000 $R(\alpha|\mu)$

S.Q. COLUNAS NÃO AJUSTADA = 83.0000000 $R(\beta|\mu)$

ANALISE DA VARIÂNCIA FIXANDO NÍVEIS

CAUSAS DA VARIAÇÃO	G.L.	S.Q.	Q.M.	VALOR F	PROB.>F
LIN/COL	11	16.0000000	16.0000000	6.6667	0.04862
LIN/COL	21	27.0000000	27.0000000	11.2500	0.02037
COL/LIN	12	6.0000000	3.0000000	1.2500	0.31511
COL/LIN	21	30.0000000	30.0000000	12.5000	0.01691
RESÍDUO	5	12.0000000	2.4000000		

Fato:

Verifica-se que as somas de quadrados e hipóteses apresentadas na Tabela 67 correspondem às do Proc GLM com a opção **Test**, conforme a Tabela 32.

5.2.7 Ajuste do modelo linear misto de classificação dupla, através do SPSS

SPSS - ANOVA trata-se de um procedimento de mínimos quadrados para analisar modelos lineares univariados balanceados e desbalanceados. Neste estudo considerou-se as declarações básicas a seguir:

```
manova Y by a(1,2) b(1,3) /
  print=parameters(estim) /
  omeans(variable(Y),table
(a)) /
  pmeans(variable(Y),table
(a)) /
  design = a vs 1, b vs 1, a by b=1 vs W/
  print=parameters(estim) /
  omeans(variable(Y),table
(a)) /
  design = muplus, a vs 1, b vs 1, a by b=1 vs W/
```

5.2.7.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

O procedimento SPSS- MANOVA através das telas de comandos, para o ajuste do modelo linear misto de classificação dupla desbalanceado com todas as caselas ocupadas, fornece as hipóteses do Tipo III do SAS, conforme mostra a Tabela 68.

Tabela 68. Ajuste do modelo misto sem casela vazia através do sistema SPSS-ANOVA

***** Análise de Variância -- delineamento 1 *****

Testes de Significância para Y usando somas de quadrados ÚNICA						
Causa de Variação	H_0 R ()	SQ	GL	QM	F	Sig de F
DENTRO DE CASELAS	$H_0^{(III-AB)}$ R($\alpha\beta \mu, \alpha, \beta$)	100,17	16	6,26		
A POR B (ERRO 1)		2,63	2	1,31	,21	,813
Erro 1		2,63	2	1,31		
A	$H_0^{(I)}$ R($\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha\hat{\beta}$)	141,07	1	141,07	107,42	,009
B	$H_0^{(II-B)}$ R($\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha\hat{\beta}$)	49,08	2	24,54	18,69	,051

Fato:

Verifica-se que a Tabela 68 apresenta, somas de quadrados e hipóteses do Tipo III do Proc GLM/ SAS com a opção **Test**, conforme a Tabela 23.

Observa-se que o valor da estatística F, para o fator fixo A obtido pelo sistema SPSS difere do apresentado na Tabela 23, em decorrência dos gl do denominador o Proc GLM/SAS apresenta o graus de liberdade corrigido, $gl=2,05$, enquanto o SPSS não fornece o valor corrigido ($gl = 2$).

5.2.7.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

O procedimento SPSS- ANOVA, através das telas de comandos, para o ajuste do modelo linear misto de classificação dupla desbalanceado com casela vazia, fornece as hipóteses que são de difícil interpretação, ou mesmo sem interesse prático. Conforme mostra a Tabela 69.

Tabela 69. Ajuste do modelo misto com casela vazia através do sistema SPSS-ANOVA

* AVISO	* As somas de quadrados ÚNICA são obtidas assumindo	*				
*	* os efeitos redundantes (possivelmente causados por	*				
*	* caselas perdidas) são atualmente nulos.	*				
*	* As hipóteses obtidas podem não ser de interesse	*				
*	* Diferentes reordenações do modelo ou dados,	*				
*	* ou diferentes contrastes podem resultar em diferentes	*				
*	* somas de quadrados ÚNICA.	*				

***** Análise de Variância -- delineamento 1 *****						
Testes de Significância para Y usando somas de quadrados ÚNICA						
Causa de Variação		SQ	GL	QM	F	Sig de F
DENTRO DE CASELAS	$H_0^{(9)} R(\alpha\beta \mu, \alpha, \beta)$	12,00	5	2,40		
A B POR B (ERRO)		1,71	1	1,71	,71	,437
Error 1		1,71	1	1,71		
A		27,00	1	27,00	15,75	,157
B		32,92	2	16,46	9,60	,222

5.2.8 Ajuste do modelo linear misto de classificação dupla, através do STATGRAPHICS-PLUS

O sistema STATGRAPHICS-PLUS apresenta um conjunto de telas de comandos para a análise do modelo linear geral balanceado e desbalanceado, através do procedimento ANOVA/MULTIFATOR.

5.2.8.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

Através de telas de comandos do procedimento ANOVA permite realizar o ajuste de mínimos quadrados para modelos mistos desbalanceados com todas as caselas ocupadas, quando há caselas vazias ele não executa a análise. A seguir descreve-se na Tabela 70, os resultados obtidos para o ajuste dos dados da Tabela 8.

Tabela 70. Ajuste do modelo misto sem casela vazia através do sistema STATGRAPHICS-ANOVA

Análise de Variância para y - Somas de quadrados do Tipo III							
Causa de variação	H ₀	R ()	Somas de Quad.	g.l.	Q. médios	F-razão	nível Sig.
Efeitos principais							
A:MISTOI.a	H ₀ ⁽³⁾	R (α $\mu, \beta, \alpha\beta$)	141.06667	1	141.06667	107.415(1)	.0092
B:MISTOI.b	H ₀ ⁽⁷⁾	R (β $\mu, \alpha, \alpha\beta$)	49.07769	2	24.53885	18.685(1)	.0508
INTERAÇÕES							
AB	H ₀ ⁽⁹⁾	R ($\alpha\beta$ μ, α, β)	2.6265664	2	1.3132832	.210(0)	.8130
RESÍDUO			100.16667	16	6.2604167		
TOTAL)			306.77273	21			

O VALORES PERDIDOS TEM SIDO EXCLUÍDOS.

F-RAZÃO SÃO BASEADAS NOS SEGUINTE QUÁDRADOS MÉDIOS:

(0)RESIDUAL

(1)AB

Fato:

Verifica-se que para o efeito de A fixo, o sistema fornece as somas de quadrados e hipóteses do Tipo III, conforme a Tabela 23. Apresenta as mesmas hipóteses e somas de quadrados do Tipo III do Proc GLM / SAS e não utiliza testes F aproximados e graus de liberdade corrigidos.

5.2.9 Ajuste do modelo linear misto de classificação dupla, através do STATISTICA

STATISTICA -ANOVA é um procedimento que reúne um conjunto de telas de comandos para analisar o modelo linear geral balanceado ou desbalanceado.

5.2.9.1 Análise do modelo misto de classificação dupla, com desbalanceamento e sem casela vazia

Através de telas de comandos o procedimento ANOVA/MANOVA permite realizar o ajuste de modelos mistos desbalanceados com todas as caselas ocupadas. Fornece as hipóteses do Tipo III do SAS, conforme mostra a Tabela 71.

Tabela 71. Ajuste do modelo misto através do sistema STATISTICA- ANOVA

Resumo para todos os efeitos;

1-A, 2-B

Efeito	gl	QM	H_0	$R()$	gl	QM	F	p-nível
1	1	141.0667	$H_0^{(3)}$	$R(\hat{\alpha} \hat{\mu}, \hat{\beta}, \alpha \hat{\beta})$	2	1.313283	107.4153	.009182
2	2	24.5388	$H_0^{(7)}$	$R(\hat{\beta} \hat{\mu}, \hat{\alpha}, \alpha \hat{\beta})$	16	6.260417	3.9197	.041172
12	2	1.3133	$H_0^{(9)}$	$R(\alpha \hat{\beta} \hat{\mu}, \alpha, \hat{\beta})$	16	6.26041	.2098	.812961

Fato:

Apresenta-se as somas de quadrados e hipóteses do Tipo III, conforme a Tabela 23. Para o efeito de A fixo, apresenta o teste F exato com denominador sendo o QM(A*B), já para os efeitos aleatórios apresenta-se o teste F exato onde o termo do erro é o QM(erro), não fornece os testes F aproximados e graus de liberdade corrigidos.

5.2.9.2 Análise do modelo misto de classificação dupla, com desbalanceamento e casela vazia

Para o caso de dados desbalanceados com caselas vazias o procedimento em estudo fornece hipóteses as quais são de difícil interpretação, ou do tipo contraste sem interesse prático conforme mostra a Tabela 72.

Tabela 72. Telas fornecidas no ajuste do modelo misto com casela vazia através do sistema STATISTICA- ANOVA

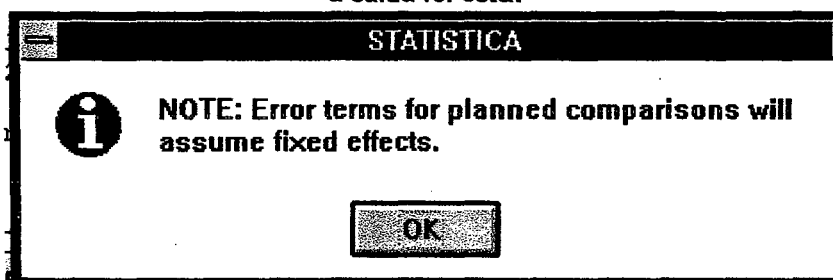
Continue...		following specific hypothesis will be tested (contrast):
A	B	C
G_1_1	G_0_0	.33
G_1_1	G_1_1	.33
G_1_1	G_2_2	.33
G_1_1	G_3_3	-.50
G_2_2	G_1_1	-.50
G_2_2	G_2_2	

GENERAL MANOVA Univar. Test	Error: Within				
	Sum of Squares	df	Mean Square	F	p-level
Effect	14544.35	1	14544.35	234.5863	.041506
Error	62.00	1	62.00		

Para esta opção:



a saída foi esta:



6 UMA COMPARAÇÃO DOS SISTEMAS COMPUTACIONAIS ESTATÍSTICOS

Apresenta-se, neste capítulo, um estudo comparativo entre os sistemas estatísticos: BMDP (1985), HARVEY (1990), MINITAB (1986), NTIA (1995), REML (1989) , SANEST (1986), SAS (1992), SPSS (1994), STATGRAPHICS -PLUS (1994), STATISTICA (1994) quando utilizados para analisar o modelo misto de classificação dupla cruzada com interação, para dados desbalanceados. Destaca-se que Santos (1994) apresenta uma discussão do uso de onze sistemas estatísticos para dados desbalanceados para o modelo de classificação dupla de efeitos fixos. Nesse trabalho o autor estabelece uma discussão sobre a performance dos sistemas, analisando-se a adequação de cada um para a formulação do modelo, para a obtenção de: somas de quadrados e hipóteses associadas, quadrados médios esperados e estimativas dos componentes de variância.

Os dez sistemas estatísticos aqui adotados foram utilizados para executar a análise do modelo linear misto de classificação dupla com interação, sendo o fator A (linha) fixo e o fator B (coluna) aleatório. Foram considerados os dados das Tabela 8 e 9 do exemplo ilustrativo.

Para cada procedimento adotou-se um conjunto de sintaxe básica já descritos em 5.1 e 5.2 A seguir apresenta-se nas Tabelas 73 à 82 um resumo da análise comparativa.

6.1 Entrada de dados

A entrada de dados para os sistemas BMDP-3V (1985), MINITAB (1986), NTIA (1995), SANEST (1986), SAS (1992), SPSS (1994), STATGRAPHICS -PLUS (1994) e STATISTICA (1994) segue uma estrutura livre por variável. O HARVEY (1990) exige que sejam especificadas as colunas de locação de cada variável dependente ou independente no arquivo de dados. O REML(1989) requer uma entrada padronizada de forma seqüencial. O SPSS (1994) exige que os níveis dos fatores envolvidos no modelo especificado, sejam inteiros consecutivos. Assim, os níveis devem ser decodificados antes de se executar a análise. A Tabela 73 apresenta um resumo das características dos sistemas.

Tabela 73. Performance dos sistemas computacionais utilizados segundo a entrada de dados

sistemas	Entrada de dados	
	padronizada	estrutura livre
BMDP-3V (1985)		x
HARVEY (1990)	x	
MINITAB (1986)		x
NTIA (1995)		x
REML(1989)	x	
SANEST (1986)		x
SAS (1992)		x
SPSS (1994)	x	
STATGRAPHICS-PLUS (1994)		x
STATISTICA (1994)		x

6.2 O modelo estatístico

Para o modelo estatístico em estudo efetua-se a comparação entre os sistemas, analisando-se a facilidade e disponibilidade de comandos para declarar os efeitos do modelo como fixo ou aleatório.

Os sistemas computacionais estatísticos utilizados neste trabalho requerem que todos os termos sejam explicitamente especificados no modelo ajustado.

Os sistemas BMDP-3V (1985), HARVEY (1990)-MODELO 1 , REML (1989) , SAS/GLM/VARCOMP/MIXED (1992), requerem que explicitamente sejam declarados todos os termos aleatórios do modelo, incluindo-se as interações. O HARVEY (1990)- modelo 6, segue as declarações designadas na sintaxe do modelo definido. Maiores detalhes podem ser obtidos no manual.

Os sistemas MINITAB (1986), NTIA (1995), SANEST (1986) e STATGRAPHICS -PLUS (1994) não apresentam mecanismo para declarar os efeitos aleatórios, mas apresentam opções para escolher o termo do erro para testar os efeitos de modo que permitem ao analista a escolha da formulação. O SPSS (1994) não apresenta opção para declarar se os efeitos são fixos ou aleatórios, e executa a análise para o modelo considerando-se todos os efeitos como sendo fixos. O STATITISTICA (1994) apresenta uma tela de comando de interpretação, a qual permite especificar, ou seja, definir o efeito principal aleatório. As propriedades de cada sistema são dadas a seguir, na Tabela 74.

Tabela 74. Performance dos sistemas computacionais utilizados segundo a especificação dos termos no modelo

sistemas	características: requer a especificação	
	de todos os termos do modelo	dos termos como fixo ou aleatório
BMDP-3V (1985)	x	x
HARVEY (1990)- modelo 1	x	x
- modelo 6	x	
MINITAB (1986)	x	
NTIA (1995)	x	
REML (1989)	x	x
SANEST (1986)	x	
SAS (1992)	x	x
SPSS (1994)	x	
STATGRAPHICS -PLUS (1994)	x	
STATISTICA (1994)	x	x*

* Apenas os efeitos principais

6.3 Testes de hipóteses

O procedimento SAS-GLM (1992) executa a análise de variância clássica do modelo misto através do método dos momentos, descrita em muitos textos. Assim apresenta: análise do Tipo I a qual se baseia no método de ajuste de constantes descrita na Tabela 5, análises do Tipo II e III, as quase baseiam no modelo III de Hocking (1973) apresentado na Tabela 1 ou na formulação 1 de Schwarz (1993) onde os quadrados médios esperados são utilizados para construir os testes F exatos ou aproximados, para os efeitos envolvidos no modelo.

O procedimento HARVEY (1990) através dos Modelos 1 e 6, fornece a análise clássica da variância com testes F exatos, as estimativas dos componentes de variância através do modelo 1 de Hocking (1973) apresentado na Tabela 1 ou a formulação 2 de Schwarz (1993) conforme mostra a Tabela 3.

Os procedimentos MINITAB (1986), NTIA (1995), SANEST (1986), SPSS (1994), STATGRAPHICS -PLUS (1994) e STATISTICA (1994), executam a análise de variância clássica para o modelo de efeitos fixos e com exceção do sistema computacional SPSS (1994) os demais sistemas computacionais permitem especificar o termo do erro que deve ser utilizado para testar os efeitos. Assim, fornecem o teste F exato para o efeito fixo A utilizando-se a interação como o termo do erro, isto é como denominador da razão F.

Conforme descritos anteriormente, os procedimentos BMDP-3V (1985), REML (1989) e SAS/MIXED (1992), são designados para a análise do modelo misto geral desbalanceado, onde os quadrados médios esperados não têm expressões usuais e os testes mais comuns não são adequados. Assim, o BMDP-3V (1985) utiliza o teste da razão de máxima verossimilhança ou teste F, assintóticos. O REML (1989) apresenta apenas as estimativas de máxima verossimilhança REML, ou seja, fornece os BLUP'S para os efeitos aleatórios e os BLUE'S para os efeitos fixos, de forma que o usuário deve completar a análise recorrendo-se a outros sistemas computacionais ou a cálculos manuais. O procedimento SAS/MIXED (1992) apresenta o teste do Tipo III, com a estatística F com base na matriz de variâncias e covariâncias assintóticas determinada utilizando-se as estimativas REML, ML ou MIVQUE(0).

Tabela 75. Performance dos sistemas utilizados segundo a determinação das E(QM) e testes F exatos ou aproximados

Sistemas	determina	Teste F		Outros
	E(QM)	exato	aproximado	
BMDP-3V (1985)	não é aplicável	-	-	Test F assintótico REML e teste razão de ML
HARVEY (1990)	apresenta apenas os coeficientes k dos componentes de variância.	sim	-	
MINITAB (1986)	não	sim	-	
NTIA (1995)	não	sim	-	
REML	não é aplicável	-	-	x
SANEST (1986)	não	-	-	
SAS	sim - Proc GLM	sim - Proc GLM	sim Proc GLM	Proc MIXED Test F assintotico REML , ML MIVQUE(0) e Teste da Razão de verossimilhança para os efeitos fixos.
SPSS (1994)	não	-	-	-
STATGRAPHICS -PLUS (1994)	não	sim	-	-
STATISTICA (1994)	não	sim	-	-

6.4 Somas de quadrados e hipóteses associadas

Para o efeito fixo do modelo misto clássico, com dois fatores cruzados e interação, os procedimentos fornecem as seguintes somas de quadrados e hipóteses associadas:

- SAS/GLM (1992) como já amplamente descrito, fornece os quatro tipos de somas de quadrados Tipo I, II, III e IV, conhecidas na literatura como somas de quadrados sequenciais, ajustadas e parciais obtidas segundo os métodos de ordenação a priori, de delineamento experimental e mínimos

quadrados completo respectivamente, os quais são sugeridos por Overall e Spiegel (Speed et al., 1978).

-HARVEY (1990)-Modelo 1 fornece as somas de quadrados parciais as quais equivalem as do Tipo III do SAS, para os dados desbalanceados com todas as caselas ocupadas e com casela vazia. Já o Modelo 6 emite as somas de quadrados ajustadas as quais equivalem as do Tipo II do SAS na ordenação B-A, para o caso de todas as caselas ocupadas.

- MINITAB (1986) através do procedimento GLM e NTIA (1995) através do procedimento MODLIN, emitem as somas de quadrados seqüencial e com restrição Σ (parciais), para o caso de todas as caselas ocupadas e apenas as do Tipo I para o caso de casela vazia.

- SANEST (1986) emite as somas de quadrados do tipo "ajustadas".

- SPSS (1994) para o caso desbalanceado com todas as caselas ocupadas fornece as somas de quadrados UNIQUE. as quais correspondem as somas de parciais. Para o caso de casela vazia, fornece uma hipótese do tipo contraste de difícil interpretação e emite um aviso que a hipótese não é única e pode não ser de interesse, processo semelhante ocorre com o sistema STATISTICA (1994).

- STATGRAPHICS-PLUS (1994) apresenta apenas a soma de quadrados do Tipo III para o caso de todas as caselas ocupadas, não emite saída para o caso de casela vazia.

Tabela 76. Performance dos sistemas computacionais para os dados com e sem casela vazia segundo as somas de quadrados e hipóteses associadas ao fator A fixo

Hipóteses	sem casela vazia		com casela vazia	
	somas de quadrados	sistemas	somas de quadrados	sistemas
Tipo I				
$H_0^{(1)} R(\alpha \mu)$	156.122727	SAS/GLM (1992) MINITAB (1986) NTIA (1995)	90.0000000	SAS/GLM (1992) MINITAB (1986) NTIA (1995)
Tipo II				
$H_0^{(2)} R(\alpha \mu, \beta)$	144.421053	SAS/GLM (1992) HARVEY (1990) - MODELO 6 (ordem B-A) SANEST (1986)	41.2857143 34.2857143 1.7142857	SAS/GLM (1992) SANEST (1986)
Tipo III				
$H_0^{(3)} R(\dot{\alpha} \dot{\mu}, \dot{\beta}, \alpha\dot{\beta})$	141.066667	SAS/GLM (1992) HARVEY (1990)- MODELO 1 MINITAB (1986) NTIA (1995) SPSS (1994) STATISTICA (1994) STATGRAPHICIS- PLUS (1994)	42.8571429	SAS/GLM (1992) HARVEY (1990) - MODELO 1
Tipo IV				
$H_0^{(4)} R(\dot{\alpha} \dot{\mu}, \dot{\beta}, \alpha\dot{\beta})$	141.066667	SAS/GLM (1992)	42.8571429	SAS/GLM (1992) SPSS (1994) * STATISTICA (1994)*

* hipóteses tipo contraste

6.5 Estimativas dos componentes de variância

Entre os sistemas computacionais estatísticos estudados apenas os sistemas BMDP-3V (1985), HARVEY (1990)-modelos 1 e 6, REML (1989) e

SAS/GLM/MIXED (1992), fornecem as estimativas dos componentes de variância, conforme mostra a Tabela 77.

Tabela 77. Performance dos sistemas computacionais utilizados segundo as estimativas dos componentes de variância

sistemas	método para as estimativas dos componentes de variância.
BMDP-3V (1985)	ML E REML
HARVEY (1990)- modelo 1 e modelo 6.	método dos momentos
MINITAB (1986)	nenhum
NTIA (1995)	nenhum
REML (1989)	REML
SANEST (1986)	nenhum
SAS (1992)	GLM oferece resultados suficientes para o analista completar os cálculos pelo método dos momentos. VARCOMP: ANOVA , REML , ML e MIVQUE(0) MIXED: REML, ML e MIVQUE(0)
SPSS (1994)	nenhum
STATGRAPHICS -	nenhum
PLUS (1994)	
STATISTICA (1994)	nenhum

Verifica-se através das Tabelas 78, 79 e 80 que apenas os sistemas computacionais HARVEY (1990)-modelos 1 e 6, e SAS/GLM (1992), apresentam as estimativas dos componentes de variância segundo o método dos momentos.

Tabela 78 Estimativas do Tipo I para os componentes de variância segundo os sistemas computacionais

PARÂMETRO	*estimativas	
	sem casela vazia	com casela vazia
	SAS/GLM/VARCOMP	SAS/GLM/VARCOMP
σ_{β}^2	3,1133	5,4286
$\sigma_{\alpha\beta}^2$	-1,3708	-0,4000
σ_e^2	6,2604	12,0000

*estimativas obtidas através do método ajuste de constantes (Henderson 3. Tabela.5)

Tabela 79 Estimativas do Tipo II para os dos componentes de variância segundo os sistemas computacionais

PARÂMETRO	ESTIMATIVAS		
	sem casela vazia		com casela vazia
	*SAS\GLM	**HARVEY (1990)-modelo 6	*SAS\GLM
σ_{β}^2	3,1134	2,4203	6,8300
$\sigma_{\alpha\beta}^2$	-1,3708	-1,3708	-6,0500
σ_e^2	6,2604	6,2604	12,0000

*estimativas segundo o modelo III (Hocking ,1973), conforme Tabela 1, (Hocking, 1985) conforme Tabela 2 , Formulação 1 Schwarz (1993) conforme Tabela 3.

** estimativas segundo o modelo I (Hocking, 1973) - Tabela 1; Formulação 2 (Schwarz, 1993) conforme Tabela 3.

Tabela 80 Estimativas do Tipo III para os componentes de variância segundo os sistemas computacionais

PARÂMETRO	ESTIMATIVAS			
	sem casela vazia		com casela vazia	
	*SAS\GLM	*HARVEY-modelo 1	*SAS\GLM	*HARVEY-modelo 1
σ_{β}^2	3,1134	2,5323	4,7300	4,2937
$\sigma_{\alpha\beta}^2$	-1,3708	-0,6854	-6,0000	-2,0000
σ_e^2	6,2604	6,26604	12,0000	12,0000

* estimativas segundo o modelo I (Hocking, 1973) - Tabela 1e Formulação 2 (Schwarz, 1993) conforme Tabela 3.

Verifica-se que apenas os sistemas computacionais BMDP-3V (1985), REML (1989) e SAS/MIXED (1992) fornecem as estimativas REML para os componentes de variância e as estimativas ML são fornecidas apenas pelos sistemas computacionais BMDP-3V (1985) e SAS/MIXED (1992) , para dados

desbalanceados e sem casela vazia (Tabela 8) e com casela vazia (Tabela 9) conforme mostram as Tabela 81 e 82 respectivamente.

Tabela 81 Estimativas ML e REML para os componentes de variância para os dados sem casela vazia segundo os sistemas computacionais

Parâmetro	SEM CASELA VAZIA				
	ESTIMATIVAS REML			ESTIMATIVAS ML	
	SAS/MIXED (1992)	BMDP-3V (1985)	REML (1989)	SAS/MIXED (1992)	BMDP-3V (1985)
σ_{β}^2	2,45	2,45	3,07	1,42	1,42
$\sigma_{\alpha\beta}^2$	0,00	0,00	0,00063	0,0	0,00
σ_e^2	5,71	5,71	6,27	5,40	5,40
iterações	3	-	3	4	-

Tabela 82 Estimativas ML e REML para os componentes de variância para os dados com casela vazia segundo os sistemas computacionais

Parâmetro	COM CASELA VAZIA				
	ESTIMATIVAS REML			ESTIMATIVAS ML	
	SAS/VARCOMP /MIXED	BMDP-3V (1985)	REML	SAS/VARCOMP /MIXED	BMDP-3V (1985)
σ_{β}^2	4,35	4,35	4,41	2,68	2,68
$\sigma_{\alpha\beta}^2$	0,00	0,00	0,0002276	0,00	0,00
σ_e^2	5,02	2,25	5,50	1,94	1,94
iteração1	4	0,0	5	3	-

As diferenças obtidas nas estimativas podem estar relacionadas com:

- 1) a metodologia de cálculo, já que os sistemas computacionais trabalham com diferentes métodos numéricos, por exemplo, o SAS/VARCOMP/MIXED (1992) utiliza o algoritmo de Newton-Raphson, o BMDP-3V (1985) utiliza uma combinação do método escore de Fisher e Newton-Raphson, já o REML utiliza o método escore de Fisher.
- 2) a capacidade dos sistemas computacionais relacionada com tecnologia, por exemplo segundo os manuais a precisão de cálculos na área de trabalho do Proc MIXED/SAS é de uma HP9000/720.

7 CONCLUSÕES

Os resultados do presente estudo permitem as seguintes conclusões:

7.1 Sobre as hipóteses

Para os modelos de efeitos mistos de classificação dupla cruzada, com interação e dados desbalanceados, os sistemas computacionais discutidos, apresentam para o efeito fixo do modelo, as seguintes saídas:

a) Quando não há casela vazia

Hipóteses do Tipo I: são emitidas pelos sistemas computacionais SAS/GLM (1992), MINITAB (1986) e NTIA (1995).

Hipóteses do Tipo II: são emitidas pelos sistemas computacionais SAS/GLM (1992), SANEST (1986) e HARVEY (1990) através do modelo 6

Hipóteses do Tipo III: são emitidas pelos sistemas computacionais SAS/GLM/ /MIXED (1992), HARVEY (1990) através do modelo 1, MINITAB (1986), NTIA (1995), SPSS (1994), STATISTICA (1994) e STATGRAPHICS-PLUS (1992).

Hipóteses do Tipo IV: é emitida pelo sistema computacional SAS/GLM/ (1992).

b) Quando há casela vazia

Hipóteses do Tipo I: são emitidas pelos sistemas computacionais SAS/GLM (1992), MINITAB (1986) e NTIA (1995).

Hipóteses do Tipo II: são emitidas pelos sistemas computacionais SAS/GLM (1992) e SANEST (1986).

Hipóteses do Tipo III: são emitidas pelos sistemas computacionais SAS/GLM/MIXED (1992) e HARVEY (1990) através do modelo 1 e STATGRAPHICS-PLUS (1992).

Hipóteses do Tipo IV: Nesse caso os sistemas computacionais SAS/GLM (1992), SPSS (1994) e STATISTICA (1994) apresentam hipóteses, que em geral não são únicas do tipo contraste e podem ser de difícil interpretação.

7.2 Sobre as estimativas dos componentes de variância

a) Entre os sistemas computacionais analisados o SAS (1992), o HARVEY (1990), o REML (1989) e o BMDP (1985) foram os únicos que forneceram as estimativas para os componentes de variância.

b) Apenas os sistemas BMDP (1986) e SAS/MIXED (1992) analisam o modelo através do método de máxima verossimilhança ML ou REML, no entanto, o SAS (1992) mostrou-se ser superior ao BMDP (1986) por apresentar o teste da razão de verossimilhança segundo as estimativas ML e REML e não apenas para as estimativas ML como apresenta o BMDP (1986).

c) O SAS/GLM (1992) através dos comandos **Random** e **Test** mostrou-se ser superior ao HARVEY (1990) para a análise de variância clássica do modelo misto, por fornecer através do método de momentos as expressões das esperanças do quadrados médios (Tipo I, II, III e IV) para as estimativas dos componentes de variância e por apresentar o teste F aproximado com graus de liberdade corrigido.

7.3 Performance dos sistemas

De um modo geral pode-se concluir para a performance desses sistemas que:

- O SAS (1992) apresenta grande flexibilidade para o ajuste do modelo de efeitos mistos. Através do procedimento GLM fornece a análise de variância clássica mais completa que os demais sistemas. Utilizando-se os procedimentos GLM, VARCOMP e MIXED do SAS (1992) obtém-se uma análise completa e adequada para o modelo de efeitos mistos, com dados desbalanceados.

- O BMDP-3V (1985) fornece saídas corretas e suficientes para a análise completa do modelo de efeitos mistos desbalanceado, mas muitos resultados requerem do usuário cálculos manuais.

- O HARVEY (1990) através dos modelos 1 e 6, também, fornece saídas suficientes, para o usuário do modelo misto.

- Os sistemas computacionais MINITAB (1986), NTIA (1995), SANEST (1986), SPSS (1994), STATGRAPHICS-PLUS (1992) e STATISTICA (1994) apresentam saídas limitadas para a análise do modelo misto e podem requer que o usuário complete a análise utilizando-se de outros sistemas. Ou seja, comportaram-se de modo semelhante por apresentarem a análise de variância através de metodologias apropriadas para modelos de efeitos fixos, adaptadas para o modelo misto, isto é, permitindo identificar o termo do erro. O SPSS (1994) só difere destes por permitir a especificação dos efeitos principais do modelo como aleatórios.

- O REML (1989) segundo o conjunto básico de sintaxe, adotado para o estudo, apresentou apenas as estimativas para os parâmetros, isto é os BLUE'S e BLUP'S.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, R. L. ; BANCROFT, T. A. **Statistical theory in research**. New York: McGraw-Hill, 1952. 399p.
- BARBIN, D. **Componentes de Variância**. Piracicaba: ESALQ/USP, 1993.108p.
- BMDP, STATISTICAL SOFTWARES. **Biomedical Computer Programs**. 3 edition. Los Angeles: University of California Press, 1985. 876p.
- BROWN, K.G. Asymptotic behavior of MINQUE-type estimators of variance components. **Annals of Mathematical of Statistics**, v.4, p.746-54, 1976.
- BUSH, N. ; ANDERSON, R.L. A comparison of three different procedures for estimating variance components. **Technometrics**, v.5, p.421-40, 1963.
- CRUMP, S. L. The estimation of variance components in analysis of variance. **Biometrics**, v.2, p. 7-11, 1946.
- CRUMP, S. L. The present status of variance components analysis. **Biometrics**, v.7, p. 1-16, 1951.

DIXON, W.J., **BMDP Statistical Software Manual**, Berkeley: University of California Press, 1990.

FISHER, R.A. Theory of statistical estimation. **Proc Cambridge Philos. Soc.**, v.22, p.700-25, 1925.

FRANCIS, I. A. Comparison of several analysis of variance programs. **Journal of the American Statistical Association**. v.68, p. 860-5, 1973.

GRAYBILL, F. A. **An introduction to linear statistical models**. New York: McGraw-Hill, 1961. 463p.

HARTLEY, H. O. Expectations, variances and covariances of ANOVA means squares by "Synthesis". **Biometrics**, v.23, p.105 - 14, 1969.

HARTLEY, H.O. ; RAO, J.N.K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**, v. 54, p.93-108, 1967.

HARTLEY, H. O. ; SEARLE, S. R. A discontinuity in mixed model analysis. **Biometrics**, v.25, p.573 - 96, 1969.

HARVEY, W.R. **User's guide for LSMLMW and MIXMDCL**. Versão PC-2. Copyright, 1990. 90p.

HARVILLE, D.A. Maximum likelihood approaches to variance components estimation and to related problems. **Journal of the American Statistical Association**, v. 73, p. 320-40, 1977.

HEMMERLE, W.J. ; HARTLEY, H.O. Computing maximum likelihood estimation for the mixed A.O.V. model using the w-transformation. **Technometric** v.15, p.819-31, 1973.

HENDERSON, C.R. Estimation of variance and covariance components. **Biometrics**,v.9, p.226-52, 1953.

HENDERSON, C.R. **Applications of linear models in animal breeding.** Guelph, Canada: University of Guelph, 1984. 462p.

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

HOCKING, R.R. A discussion of the two-way mixed model. **The American Statistician**, v.27, p.148-52, 1973.

HOCKING, R.R. **The analysis of linear models.** Monterey: Brooks/Cole Publishing Company, 1985. 385 p.

HOCKING, R.R. ; KUTNER, M.H. Some analytical and numerical comparisons of estimators for the mixed A.O.V. model. **Biometrics**, v.31, p.19-27, 1975.

HULTQUIST, R.A. ; GRAYBILL, F.A. Minimal sufficient statistics for the two-way classification mixed model design. **Journal of the American Statistical Association**, v.60, p.82-92, 1965.

- IEMMA, A.F. **Testes de hipóteses em modelos lineares com amostras desequilibradas**. Belgique. Gembloux. Faculdade de Ciências Agrônômicas, 1991. 105 p.
- IEMMA, A.F. **Análisis de varianza con datos desbalanceados**. Bogota, Universidad Nacional da Colombia, 1993a . 102 p.
- IEMMA, A. F. **Análise de variância de dados desbalanceados**. 4º Congresso Brasileiro de Usuários do "SAS". USP-ESALQ. Piracicaba, 1995a, 111p..
- IEMMA, A. F.. **Dados estatísticos desbalanceados: PROC SAS/GLM**. Semana Acadêmica - Departamento de Matemática, Instituto de Ciências Exatas, Universidade Federal do Rio de Janeiro, 1995-b. 130p.
- IEMMA, A. F. ; PALM, R.; ALVES, M.I.F. Statistical hypothesis and ortogonal projections for unbalanced data. **Biometrics Bulletin**. v.10, n.1, p.18, 1993a
- IEMMA, A. F.; PALM, R.; CLAUSTRIAUX, J.J. Sobre a construção de projetores ortogonais. **Revista de Matemática e Estatística**, São Paulo, **11**, 133-42, 1993b.
- IEMMA, A. F.; PERRI, S.H.V. **Ajuste de modelos fixos desbalanceados através do sistema estatístico SAS**. Piracicaba: ESALQ/USP, 1997. 99p.
- JACKSON, R.W.B. Reliability of mental tests. **Brit. J. Psychol**, v.29, p.267-87, 1939.

KACKAR, R.N. ; HARVILLE, D.A. Aproximations for standard error of estimators of fixed and random effect in mixed linear models. **Communications in Statistical - Part A. Theory and Methods**, v.10, p.1249-61, 1984.

KEMPTHORNE, O. **An introduction to genetic statistics**. New York : John Wiley, 1957.

LIMA, C.G. **Uso do MINITAB na estatística aplicada à experimentação agrônômica**. Piracicaba - ESALQ, 1993.

MILLER, J.J. Asymptotic properties of maximum likelihood estimation. The mixed model of the analysis of variance. **The Annals of Statistics**, v.5, p.746-62, 1977.

MILLIKEN, G. A. ; JOHNSON, D. E. **Analysis of messy data**. New York: Van Nostrand Reinhold Company, 1984. Vol.1. 473p.

MINITAB, Reference Manual, ed., **Minitab Statistical Computing System**. Boston: Duxbury Press. 1986.

MOOD, A. M. **Introduction to the theory of statistics**. New York : McGraw-Hill, 1950.

NTIA - MODULO MODLIN **Manual Sistema Operacional C**. Campinas: Empresa Brasileira de Pesquisa Agropecuária, 1995. 73p.

PATTERSON, H.D. ; THOMPSON, R.. Recovery of interblock information when block sizes are unequal. **Biometrika** ,v.58, p.545, 1971.

- PATTERSON, H.D. ; THOMPSON, R.; HUNTER, E.A. ; WILLIAMS, E.R.
Analysis of non-ortogonal data using REML, Edinburgh: R A Kempton,
1989. 39p.
- PERRI, S.H.V.; IEMMA, A.F. **Ajuste de modelos mistos através do sistema estatístico SAS**. Piracicaba: ESALQ/USP, 1997. 75p.
- PLACKETT, R. L.. Models in the analysis of variance. **Journal of the Royal Statistical Society B**, v.22, p.195 - 217, 1960.
- RAO, C.R. Estimation of heteroscedastic variances in linear models. **Journal of the American Statistical Association**. v.64, p.161-72, 1970.
- RAO, J. N. K. On Expectations, variances, and Covariances of ANOVA. Mean Squares by "Synththesis". **Biometrics**. v. 24, p. 963-978, 1968.
- REML, **Reference Manual** ed. , Edinburgh: R A Kemton, 1989. 39p.
- RAO, C.R. Estimation of variance and covariance components - MINQUE. Theory. **Journal Multivariate Analysis**, v.1, p.257-75, 1971 a.
- RAO, C.R. Minimum variance quadratic unbisead estimation of variance components. **Journal Multivariate Analysis**, v.1, p. 447-56, 1971 b.
- RAO, C.R. Estimation of variance and covariance components in linear models. **Journal of the American Statistical Association**. v.67, p.112-5, 1972.
- SANEST- Sistema de Análise Estatística para Microcomputadores. Pelotas. Universidade Federal de Pelotas, 1986.

SANTOS, E.S. Utilização de “softwares” estatísticos na interpretação de hipóteses com dados desbalanceados. Piracicaba, 1994. 175p. Tese (Doutorado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

SAS Institute Inc. **SAS User’s Guide: Statistics**, Version 6, 4 edition. Cary, North Carolina: Sas Institute Inc., 1992. 846p.

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

SCHEFFÉ, H. A “mixed model “ for the analysis of variance. **Annals of Mathematical of Statistics**, v.1, p.23 -6, 1956a.

SCHEFFÉ, H. Alternative models for the analysis of variance. **Annals of Mathematical of Statistics**, v.27, 251 - 71, 1956b.

SCHEFFÉ, H. **The analysis of variance**. New York: John Wiley, 1959. 477p.

SCHWARZ, C.J. The mixed-model ANOVA: The truth, the computer packages, the books, parte I: Balanced data. **The American Statistician**, v.47, p.48-59, 1993.

SEARLE, S.R. Another look at Henderson’s methods of estimating variance components. **Biometrics**, v.24, p.749-88, 1968.

- SEARLE, S.R. Large sample variances of maximum likelihood estimators of variance components using unbalanced data. **Biometrics**, v.26, p.505-24, 1970.
- SEARLE, S. R. **Linear models**. New York: John Wiley. 1971.
- SEARLE, S. R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536p.
- SEARLE, S.R. ; HENDERSON, C.R. Computing procedures for estimating components of variance in the two-way classification, mixed model. **Biometrics**, v.17, p. 607-16, 1961.
- SEARLE, S.R. ; CASELLA, G. ; McCULLOCH, C. **Variance Components**. New York: John Wiley, 1992. 501p.
- SERAPHIN, J.C. Comparação numérica de três estimadores de componentes de variância em alguns modelos genéticos - estatísticas de cruzamentos. Piracicaba, 1984. 74p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SMITH, D.W. ; HOCKING, R.R. Maximum likelihood analysis of the mixed model: The balanced case. **Communications in Statistical - Theory and Methods**, v. A7, n.10, p.1253-66, 1978.
- SNEDECOR, G.W. ; COCHRAN, W.G. **Statistical methods**, 6 ed.. Iowa, USA, The State University Press, 1967. 593p.
- SPEED, F.M. ; HOCKING, R.R. The use of the R () - notation with unbalanced data. **The American Statistial**, v. 30, p.30-3, 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**, v. 73, 105-112, 1978

SPSS, Statistical Package for the Social Sciences, Chicago: McGRAW-HILL, 1994.

STATGRAPHICS-Plus **Reference Manual**. Versão 6. Cambridge: Copyright, Manugistics, Inc., 1992.

STATISTICA for Windows. **General Conventions & Statistics**, v.1. Copyright, StatSoft, Inc., 1994. 1659p.

STEEL, R. G. D. ; TORRIE, J. H. **Principles and Procedures of Statistics**, New York, McGraw-Hill, 1960.

SWALLOW, W.H. ; MONAHAN, J.F. Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. **Technometrics**, v. 26, p.47-58, 1984.

THOMPSON, W.A. J. The problem of negative estimates of variances components. **Annals of Mathematical of Statistics**, v. 33, p.273-89,1962.

VALERIO FILHO, W.V. Comparação de métodos para estimação de componentes da variância através da simulação de dados. Piracicaba, 1991. 160p.. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz". Universidade de São Paulo.

WILK, M. B. ; KEMPTHORNE, O.. Fixed, mixed and random models. **Journal of the American Statistical Association**, v.50, p.1144 - 67, 1955.

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