### Modelos elípticos multiníveis

### Roberto Ferreira Manghi

Dissertação apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa: Estatística

Orientador: Prof. Dr. Gilberto Alvarenga Paula

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro do CNPq

São Paulo, dezembro de 2011

### Modelos elípticos multiníveis

Este exemplar corresponde à redação final da dissertação devidamente corrigida, defendida por Roberto Ferreira Manghi e aprovada pela Comissão Julgadora.

São Paulo, 08 de dezembro de 2011.

Comissão Julgadora:

Prof. Dr. Gilberto Alvarenga Paula (orientador) - IME/USP

Profa. Dra. Denise Aparecida Botter - IME/USP

Profa. Dra. Cibele Maria Russo Noveli - ICMC/USP

À minha Mãe, a mulher mais admirável que conheci, com muito amor.

## Agradecimentos

Gostaria de agradecer primeiramente aos meus familiares, em especial ao meu Pai **Roberto** e à minha Mãe **Raquel**, que sempre acreditaram nos meus objetivos e sempre me deram todo o suporte nesta jornada.

Ao meu irmão **Gustavo** e aos meus tios e tias, com muito carinho.

Dedico também este trabalho aos meus avós (in memorian).

À minha namorada **Paula**, a quem devo tanta compreensão e tanto amor durante todo o tempo desde a Graduação, nas felicidades e também nos momentos difíceis.

Ao meu orientador **Prof. Dr. Gilberto Alvarenga Paula**, pela honra e satisfação de trabalhar com um professor de tanto gabarito, tanto profissionalismo e atenção.

Aos professores do **Departamento de Estatística** do **Instituto de Matemática e Estatística** da **Universidade de São Paulo**, com os quais tive o privilégio de aprimorar meus conhecimentos.

Aos amigos e professores do **Departamento de Estatística** da **Universidade Federal de Pernambuco**, por sempre acreditarem em meu potencial e por sempre me incentivarem.

Ao **Prof. Dr. Paulo de Paula Mendes** e à **Profa. Dra. Emiko Shinozaki Mendes**, ambos professores da **Universidade Federal Rural de Pernambuco**, pelo apoio e pelos conselhos a respeito da vida acadêmica.

Aos colegas de pós-graduação do **Instituto de Matemática e Estatística** da **Universidade de São Paulo**, pelo convívio durante o tempo de Mestrado, uma experiência ímpar em meu desenvolvimento.

Ao apoio financeiro do **CNPq**.

### Resumo

Os modelos multiníveis representam uma classe de modelos utilizada para ajustes de dados que apresentam estrutura de hierarquia. O presente trabalho propõe uma generalização dos modelos normais multiníveis, denominada modelos elípticos multiníveis. Esta proposta sugere o uso de distribuições de probabilidade pertencentes à classe elíptica, envolvendo portanto todas as distribuições contínuas simétricas, incluindo a distribuição normal como caso particular. As distribuições elípticas podem apresentar caudas mais leves ou mais pesadas que as caudas da distribuição normal. No caso da presença de observações aberrantes, é sugerido o uso de distribuições com caudas pesadas no intuito de obter um melhor ajuste do modelo aos dados considerados discrepantes. Nesta dissertação, alguns aspectos dos modelos elípticos multiníveis são desenvolvidos, como o processo de estimação dos parâmetros via máxima verossimilhança, testes de hipóteses para os efeitos fixos e parâmetros de variância e covariância e análise de resíduos para verificação de características relacionadas aos ajustes e às suposições estabelecidas.

Palavras-chave: modelos multiníveis, modelos elípticos, estimação robusta.

### Abstract

Multilevel models represent a class of models used to adjust data which have hierarchical structure. The present work proposes a generalization of the multilevel normal models, named multilevel elliptical models. This proposal suggests the use of probability distributions belonging to the elliptical class, thus involving all symmetric continuous distributions, including the normal distribution as a particular case. Elliptical distributions may have lighter or heavier tails than the normal ones. In case of presence of outlying observations, it is suggested the use of heavy-tailed distributions in order to obtain a better fitted model to the discrepant observations. In this dissertation some aspects of the multilevel elliptical models are developed, such as the process of parameter estimation by maximum likelihood, hypothesis tests for fixed effects and variance-covariance parameters and residual analysis to check features related to the fitting and established assumptions. **Keywords:** multilevel models, elliptical models, robust estimation.

## Sumário

Li	sta d	le Figu	iras	vi
Li	Lista de Tabelas			x
1	Intr	Introdução		
	1.1	Introd	lução	1
<b>2</b>	Mo	delos I	Normais Multiníveis	6
	2.1	Motiv	ação	6
	2.2	Defini	ção do Modelo	12
	2.3	Métod	los de Estimação	15
		2.3.1	Máxima Verossimilhança (MV)	15
		2.3.2	Máxima Verossimilhança Restrita (MVR)	19
		2.3.3	Detalhes Computacionais no uso de MV e MVR	20
	2.4	Predic	ção dos Efeitos Aleatórios	21
	2.5	Coefic	eiente de Correlação Intraclasse	22
	2.6	Testes	de Hipóteses	23
		2.6.1	Testes de Hipóteses para os Efeitos Fixos	23
		2.6.2	Testes de Hipóteses para os Parâmetros de Variância e Covariância	26
	2.7	Estrat	égias na Seleção dos Modelos Multiníveis	28
		2.7.1	A Estratégia "Top-Down"	29
		2.7.2	A Estratégia "Step-Up"	30
		2.7.3	O Uso de Critérios de Informação na Seleção de Modelos	31
	2.8	Anális	se de Resíduos	32
	2.9	Aplica	nções	40

		2.9.1	Aplicação 1 - Experimento com Ratos	40
		2.9.2	Aplicação 2 - Estudo Educacional	55
3	Mo	delos I	Elípticos Multiníveis	75
	3.1	A Cla	sse de Distribuições Elípticas	75
	3.2	Defini	ção do Modelo	76
	3.3	Estim	ação dos Parâmetros por Máxima Verossimilhança	78
		3.3.1	Algoritmo ECME para a Estimação do Modelo $t$ -Student Multinível	80
	3.4	Prediq	ção dos Efeitos Aleatórios	83
	3.5	Coefic	eiente de Correlação Intraclasse	83
	3.6	Escolh	na do Valor do Parâmetro de Forma	84
	3.7	Testes	de Hipóteses	84
		3.7.1	Testes de Hipóteses para os Efeitos Fixos	84
		3.7.2	Testes de Hipóteses para os Parâmetros de Variância e Covariância	86
	3.8	Estrat	égias na Seleção dos Modelos Multiníveis	87
	3.9	Anális	se de Resíduos	87
	3.10	Aplica	ıções	92
		3.10.1	Aplicação 1 - Experimento com Ratos	92
		3.10.2	Aplicação 2 - Estudo Educacional	100
4	Con	clusõe	2S	107
$\mathbf{A}$	Not	ação N	Matricial dos Modelos Multiníveis	109
	A.1	Model	los Multiníveis com 2 Níveis	109
	A.2	Model	los Multiníveis com 3 Níveis	112
$\mathbf{R}$	eferê	ncias I	Bibliográficas	118

# Lista de Figuras

2.1	Gráficos de box-plot dos pesos dos ratos segundo o sexo para cada dosagem.	43
2.2	Gráficos de box-plot dos pesos dos ratos por tamanho da ninhada (machos	
	(a) e fêmeas (b))	51
2.3	Resíduos marginais e condicionais studentizados referentes ao modelo nor-	
	mal ajustado aos dados do experimento com ratos.	52
2.4	Estimativas das distâncias de Mahalanobis para as ninhadas e percentis	
	$0,\!95$ da distribuição qui-quadrado referentes ao modelo normal ajustado	
	aos dados do experimento com ratos.	52
2.5	Preditores lineares studentizados referentes ao modelo normal ajustado aos	
	dados do experimento com ratos.	53
2.6	Gráficos normais de probabilidades com envelopes gerados para os resí-	
	duos marginais e condicionais studentizados referentes ao modelo normal	
	ajustado aos dados do experimento com ratos	53
2.7	Gráfico normal de probabilidades com envelope gerado para os preditores	
	lineares studentizados referentes ao modelo normal ajustado aos dados do	
	experimento com ratos	54
2.8	Gráfico normal de probabilidades com envelope gerado para os resíduos con-	
	dicionais com confundimento mínimo referentes ao modelo normal ajustado	
	aos dados do experimento com ratos	54
2.9	Gráfico de dispersão bidimensional entre os escores de matemática no jar-	
	dim de infância e os ganhos nos escores dos alunos. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$	56
2.10	Gráfico de dispersão bidimensional entre os escores de matemática no jar-	
	dim de infância e os status socioeconômicos dos alunos.	57

2.11	Gráficos de box-plot dos ganhos nos escores pelo sexo dos alunos. $\ .\ .$ .	58
2.12	Gráficos de box-plot dos ganhos nos escores pela maioridade dos alunos	59
2.13	Gráficos de box-plot dos ganhos nos escores pela preparação em matemática	
	dos professores.	60
2.14	Gráficos de box-plot dos ganhos nos escores pelos anos de experiência em	
	ensino dos professores.	61
2.15	Gráficos de box-plot dos ganhos nos escores pelo conhecimento em mate-	
	mática dos professores	62
2.16	Gráficos de box-plot dos ganhos nos escores pelos percentuais de famílias	
	abaixo da linha de probreza nas proximidades das escolas	63
2.17	Gráficos de box-plot dos ganhos nos escores segundo as classes pertencentes	
	às quatro primeiras escolas amostradas	64
2.18	Resíduos marginais e condicionais studentizados referentes ao modelo nor-	
	mal ajustado aos dados do estudo educacional	68
2.19	Estimativas das distâncias de Mahalanobis para as escolas e percentis $0,\!95$	
	da distribuição qui-quadrado referentes ao modelo normal ajustado aos	
	dados do estudo educacional	69
2.20	Preditores lineares studentizados referentes ao modelo normal ajustado aos	
	dados do estudo educacional	70
2.21	Estimativas das distâncias de Mahalanobis para os efeitos aleatórios e per-	
	centis 0,95 da distribuição qui-quadrado referentes ao modelo normal ajus-	
	tado aos dados do estudo educacional	70
2.22	Gráficos normais de probabilidades com envelopes gerados para os resí-	
	duos marginais e condicionais studentizados referentes ao modelo normal	
	ajustado aos dados do estudo educacional	73
2.23	Gráfico normal de probabilidades com envelope gerado para os preditores	
	lineares studentizados referentes ao modelo normal ajustado aos dados do	
	estudo educacional.	73

2.24	Gráfico normal de probabilidades com envelope gerado para os resíduos con-	
	dicionais com confundimento mínimo referentes ao modelo normal ajustado	
	aos dados do estudo educacional.	74
3.1	Gráficos de dispersão bidimensional entre os pesos estimados e os resíduos	
	marginais e condicionais studentizados do modelo $t$ -Student ajustado aos	
	dados do experimento com ratos.	96
3.2	Estimativas das distâncias de Mahalanobis modificadas para as ninhadas	
	e percentis 0,95 da distribuição ${\cal F}$ referentes ao modelo $t\mbox{-Student}$ ajustado	
	aos dados do experimento com ratos	97
3.3	Preditores lineares studentizados do modelo $t$ -Student ajustado aos dados	
	do experimento com ratos	98
3.4	Gráficos normais de probabilidades com envelopes gerados para os resíduos	
	marginais e condicionais studentizados do modelo $t$ -Student ajustado aos	
	dados do experimento com ratos.	98
3.5	Gráfico normal de probabilidades com envelope gerado para os preditores	
	lineares studentizados do modelo $t$ -Student ajustado aos dados do experi-	
	mento com ratos.	99
3.6	Gráficos de dispersão bidimensional entre os pesos estimados e os resíduos	
	marginais e condicionais studentizados do modelo $t$ -Student ajustado aos	
	dados do estudo educacional	103
3.7	Estimativas das distâncias de Mahalanobis modificadas para as escolas e os	
	respectivos percentis 0,95 da distribuição ${\cal F}$ referentes ao modelo $t\mbox{-Student}$	
	ajustado aos dados do estudo educacional.	104
3.8	Preditores lineares studentizados do modelo $t$ -Student ajustado aos dados	
	do estudo educacional.	104
3.9	Gráficos normais de probabilidades com envelopes gerados para os resíduos	
	marginais e condicionais studentizados do modelo $t$ -Student ajustado aos	
	dados do estudo educacional	105

3.10	Gráfico normal de probabilidades com envelope gerado para os preditores	
	lineares studentizados do modelo $t$ -Student ajustado aos dados do estudo	
	educacional.	106

# Lista de Tabelas

2.1	Amostra do banco de dados referente ao experimento com ratos	41
2.2	Descrições das variáveis consideradas no experimento com ratos	42
2.3	Medidas descritivas dos pesos dos ratos por sexo e tratamento	43
2.4	Estimativas dos parâmetros do modelo (2.39)	47
2.5	Amostra do banco de dados referente ao estudo educacional	55
2.6	Descrições das variáveis consideradas no estudo educacional	71
2.7	Medidas descritivas de algumas variáveis consideradas no estudo educacional.	72
2.8	Estimativas dos parâmetros do modelo (2.41).	72
3.1	Estimativas dos parâmetros do modelo (3.20)	94
3.2	Estimativas dos parâmetros sob os modelos normal e $t$ -Student ajustados	
	aos dados do experimento com ratos	94
3.3	ICCs estimados dos modelos normal e $t$ -Student ajustados aos dados do	
	experimento com ratos	95
3.4	20 menores pesos estimados pelo modelo $t\mbox{-}Student$ ajustado aos dados do	
	experimento com ratos com as respectivas ninhadas	95
3.5	Estimativas dos parâmetros do modelo (3.22)	101
3.6	Estimativas dos parâmetros sob os modelos normal e $t$ -Student ajustados	
	aos dados do estudo educacional.	102
3.7	ICCs estimados dos modelo normal e $t$ -Student ajustados aos dados do	
	estudo educacional.	102
3.8	20 menores pesos estimados pelo modelo $t\mbox{-}Student$ ajustado para os dados	
	do estudo educacional com as respectivas escolas.	102

### Capítulo 1

## Introdução

### 1.1 Introdução

Os modelos multiníveis representam uma classe de modelos muito utilizada para estudos que apresentam dados com uma estrutura de hierarquia. Trabalhos que fornecem dados coletados nas ciências humanas e biológicas, por exemplo, possuem geralmente uma estrutura hierárquica. Pesquisas na área de hereditariedade envolvendo animais ou pessoas apresentam naturalmente dados dispostos em níveis, em que gerações são agrupadas dentro das famílias, uma vez que indivíduos de uma geração proveniente dos mesmos pais tendem a apresentar maiores similaridades em suas características físicas e mentais do que indivíduos escolhidos aleatoriamente na população como um todo. Para este caso, crianças de uma mesma família tendem a ter baixa estatura se seus pais são de baixa estatura, além da possibilidade da existência de problemas de formação provenientes de outros elementos, como por exemplo a influência de fatores ambientais na estatura das crianças. Em experimentos na forma de ensaios clínicos, como por exemplo no caso de indivíduos que são selecionados aleatoriamente dentre uma amostra também aleatória de hospitais, existe a característica de diferentes níveis nos dados coletados, uma vez que pacientes atendidos num mesmo hospital podem apresentar comportamentos similares considerando a evolução no tratamento de determinada doença em comparação com pacientes coletados de diferentes hospitais. Fatores relacionados aos hospitais podem ser determinantes na eficiência do tratamento considerado, tais como estrutura do prédio hospitalar, qualidade dos equipamentos, etc.

A literatura que trata de modelos multiníveis compreende, por exemplo, trabalhos realizados em Gelman e Hill (2007), Goldstein (2009), Raudenbush e Bryk (2002), Snijders e Bosker (1999) e West, Welch e Galecki (2007), porém a ideia de flexibilizar as suposições comumente assumidas nos modelos normais lineares para realização de estudos com dados apresentando estruturas mais complexas foi proposta em diversos outros trabalhos anteriores. Uma importante classe de modelos é a dos modelos lineares generalizados (MLGs), como vistos em McCullagh e Nelder (1999) e em Paula (2010). Eles permitem a modelagem de dados que seguem distribuições pertencentes à família exponencial (incluindo a distribuição normal como caso particular). Entretanto, esses modelos assumem independência entre as observações, suposição que nem sempre é satisfeita em determinados casos. Uma generalização desta classe de modelos compreende os modelos lineares generalizados hierárquicos (Lee e Nelder (1996)), em que o preditor linear definido nos MLGs agora pode apresentar componentes aleatórios.

Outros exemplos de grande importância são os modelos de componentes de variância, nos quais pode ser considerada a presença de efeitos aleatórios para determinados fatores considerados no estudo (ver, por exemplo, Crump (1946) e Searle, Casella e McCulloch (1992)). Nesses modelos, existe a possibilidade de ser imposta uma determinada estrutura de variância e covariância aos dados. Contudo, sob a presença de dados não balanceados o processo de estimação dos parâmetros não apresenta equações em forma fechada. Generalizações desta última classe de modelos foram propostas, como por exemplo os modelos mistos (Laird e Ware (1982) e Waternaux, Laird e Ware (1989)), aplicados eficientemente em dados de estrutura longitudinal, ou seja, dados que apresentam medidas coletadas para um mesmo indivíduo ao longo do tempo. Estes modelos também são aplicados em dados com medidas repetidas, que são caracterizados pela presença de variáveis observadas para um mesmo indivíduo sob, por exemplo, diferentes níveis de um determinado fator. McLean, Sanders e Stroup (1991) comparam resultados de aplicações envolvendo os modelos de análise de variância convencionais com o uso dos modelos mistos. Verbeke e Molenberghs (2003) discutem o uso de testes do tipo escore propostos por Silvapulle e Silvapulle (1995) em modelos de componentes de variância e Giampaoli e Singer (2009) discutem o uso do teste da razão de verossimilhanças para os parâmetros de variância e covariância em modelos mistos.

A existência de dados não balanceados motivou o uso de métodos iterativos para estimação dos parâmetros em modelos mistos. O algoritmo mais utilizado para o processo de estimação nestes modelos é o algoritmo EM (*Expectation-Maximization*), proposto por Dempster, Laird e Rubin (1977). Este algoritmo é utilizado inclusive para a obtenção das estimativas dos componentes de variância e covariância (Dempster, Rubin, e Tsutakawa (1981)).

Diversas técnicas desenvolvidas a partir dos modelos mistos podem ser utilizadas nos modelos multiníveis. Existem trabalhos específicos que detalham métodos de estimação dos parâmetros, testes de hipóteses e alguns procedimentos para a análise de resíduos e de diagnóstico em modelos multiníveis com distribuição normal. Raudenbush e Bryk (2002) consideram o uso destes modelos tratando cada nível individualmente, para uma posterior análise considerando todos os níveis. West, Welch e Galecki (2007) abordam o uso de modelos multiníveis de forma bastante didática, utilizando vários exemplos aplicados a dados com dois ou três níveis de hierarquia. Em Pires (2009) temos o desenvolvimento da análise de influência local através da curvatura normal em modelos multiníveis. Em Hilden-Minton (1995) temos algumas técnicas de diagnóstico para estes modelos. Em Raudenbush e Bryk (2002) e Natis (2000) temos ainda uma discussão acerca de importantes detalhes como a locação das variáveis preditoras e as várias interpretações dos parâmetros a partir do uso de determinadas locações.

Uma importante característica dos modelos normais multiníveis (assim como dos modelos mistos com distribuições normais dos erros e efeitos aleatórios) é a falta de robustez das estimativas de máxima verossimilhança na presença de observações aberrantes, ou seja, dados discrepantes em relação aos demais tendem a exercer maior influência nas estimativas dos parâmetros. Embora os modelos normais apresentem certa praticidade no seu uso, este problema pode implicar em erros na realização de inferência acerca dos parâmetros e, portanto, o modelo pode ser eventualmente mal especificado. Uma proposta para lidar com esse problema é o uso de distribuições de probabilidades com caudas mais pesadas que as caudas da distribuição normal, como por exemplo a distribuição *t*-Student, no intuito de melhor acomodação das observações discrepantes. Com relação a modelos que utilizam a distribuição citada, importantes trabalhos encontram-se disponíveis. Zellner (1976) discute as abordagens clássica e bayesiana em modelos de regressão com distribuição t-Student multivariada dos erros. Kowalsky, Mendoza-Blanco, Tu e Gleser (1999) também comparam técnicas inferenciais bayesianas e clássicas nestes modelos. Fernández e Steel (1999) ilustram algumas armadilhas encontradas em métodos bayesianos e de máxima verossimilhança nos ajustes de modelos de regressão t-Student com graus de liberdade desconhecidos. Lange, Little e Taylor (1989) tratam da robustez das estimativas de máxima verossimilhança em modelos de estrutura longitudinal com erros t-Student, para graus de liberdade fixos ou desconhecidos. Liu e Rubin (1995) apresentam extensões do algoritmo EM para as estimativas em modelos com distribuição t-Student multivariada. Pinheiro, Liu e Wu (2001) também utilizam extensões do algoritmo EM para estimação robusta dos parâmetros nos modelos mistos com a referida distribuição. Little (1988) aborda a estimação robusta em modelos com dados faltantes.

Considerando a classe de distribuições elípticas, importantes contribuições também foram desenvolvidas. Galea, Paula e Bolfarine (1997) tratam da análise de influência local em modelos elípticos e Galea, Paula e Uribe-Opazo (2003) abordam esta técnica de diagnóstico para o caso univariado. Osorio, Paula e Galea (2006) propõem o estudo de influência local em modelos elípticos com estrutura longitudinal. Savalli (2005) e Savalli, Paula e Cysneiros (2006) abordam o uso de testes do tipo escore para parâmetros de variância e covariância em modelos elípticos mistos e Cysneiros e Paula (2004) discutem o uso de testes de hipóteses unilaterais em modelos de regressão com distribuição t-Student multivariada dos erros.

O objetivo deste trabalho é propor uma generalização dos modelos normais multiníveis, denominada modelos elípticos multiníveis. Esta proposta sugere o uso de distribuições de probabilidade pertencentes à classe elíptica, envolvendo portanto todas as distribuições contínuas simétricas, incluíndo a distribuição normal como caso particular. As distribuições elípticas podem apresentar caudas mais leves ou mais pesadas do que as caudas da distribuição normal. No caso da presença de observações aberrantes, é sugerido o uso de distribuições com caudas pesadas no intuito de obter um melhor ajuste do modelo aos dados considerados discrepantes. No Capítulo 2 temos a motivação do uso de modelos multiníveis através de um exemplo hipotético, bem como a descrição dos métodos de estimação, testes de hipóteses e análise de resíduos nestes modelos considerando distribuição normal dos erros e efeitos aleatórios. Temos ainda neste capítulo a ilustração de duas aplicações, a primeira utilizando modelos para dados com dois níveis de hierarquia e a segunda abordando o uso dos modelos normais multiníveis para dados que apresentam três níveis de hierarquia. No Capítulo 3 temos o desenvolvimento da proposta dos modelos elípticos multiníveis. Primeiramente, a classe de distribuições elípticas é abordada, posteriormente é definido o modelo elíptico multinível propriamente dito. Em seguida são desenvolvidos os procedimentos para a estimação dos parâmetros via máxima verossimilhança, realização de testes de hipóteses e análise de resíduos. Finalmente, as aplicações trabalhadas no Capítulo 2 são desenvolvidas com o uso de modelos elípticos multiníveis com distribuição *t*-Student multivariada. Os resultados obtidos sob esta distribuição e sob a distribuição normal são comparados e discutidos. No Capítulo 4 temos as conclusões sobre o trabalho aqui desenvolvido e sugestões para trabalhos futuros.

### Capítulo 2

### Modelos Normais Multiníveis

Neste capítulo temos inicialmente um exemplo hipotético para a motivação do uso de modelos multiníveis. Posteriormente será discutida a teoria de estimação dos parâmetros, bem como as técnicas de inferência existentes para estes modelos. Em seguida temos dois exemplos de aplicações: o primeiro exemplo refere-se ao uso de modelos multiníveis em dados que apresentam dois níveis de hierarquia e o segundo exemplo faz uso destes modelos em um conjunto de dados que apresenta três níveis de hierarquia.

### 2.1 Motivação

Considere um exemplo de amostragem de alunos em uma determinada escola. Um dos possíveis interesses é a investigação de fatores que contribuem no rendimento dos alunos numa matéria específica. Supondo que exista uma variável Z que explique as notas dos alunos, um modelo inicial poderia ser proposto da forma

$$Y_i = \gamma_1 + \gamma_2 Z_i + \epsilon_i, \tag{2.1}$$

para i = 1, 2, ..., n, em que n é o número de alunos amostrados na determinada escola. Assim, em (2.1) temos que

- $Y_i$  representa o rendimento escolar do *i*-ésimo aluno da amostra;
- $Z_i$  é o valor da variável independente considerada no estudo para o *i*-ésimo aluno;
- $\gamma_1$  é o parâmetro que representa a média geral do rendimento dos alunos dado que o valor de Z é zero;

- $\gamma_2$  é o parâmetro associado à variável independente Z, ou seja,  $\gamma_2$  é o acréscimo obtido no valor médio do rendimento dos alunos quando o valor de Z é acrescido em uma unidade;
- $\epsilon_i$  é um erro aleatório associado a  $Y_i$ ,

com a suposição de que os erros são independentes entre si, com  $\epsilon_i \sim N(0, \sigma^2)$ , para i = 1, 2, ..., n.

Suponha agora que a amostra de n alunos tenha sido obtida a partir de uma amostra de J classes dentro da escola, e que o valor de  $\gamma_1$  pode apresentar diferentes valores para diferentes classes, ou seja, o valor médio do rendimento dos alunos dado que o valor de Z é zero depende da classe onde os alunos foram amostrados. Sendo assim, um modelo que leva em consideração essa suposição é dado por

$$Y_{ij} = \gamma_{1j} + \gamma_2 \mathbf{Z}_{ij} + \epsilon_{ij}, \qquad (2.2)$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, em que  $n_j$  é o número de alunos da amostra pertencentes à *j*-ésima classe, com  $\sum_{j=1}^{J} n_j = n$  e J o número total de classes consideradas. Desta forma, temos que

- $Y_{ij}$ é o rendimento do *i*-ésimo aluno pertencente à *j*-ésima classe amostrada;
- $\gamma_{1j}$ é o rendimento médio dos alunos da <br/> j-ésima classe dado o valor nulo de Z;
- $\gamma_2$  é o parâmetro associado à variável independente Z, ou seja, é o acrésimo no valor médio do rendimento dos alunos quando a variável Z é acrescida em uma unidade considerando todas as classes amostradas;

 $\epsilon_{ij}$  é um erro aleatório associado a  $Y_{ij}$ ,

com as suposições de que os erros são independentes entre si, sendo  $\epsilon_{ij} \sim N(0, \sigma^2)$ , para  $i = 1, 2, \ldots, n_j$  e  $j = 1, 2, \ldots, J$ . O parâmetro  $\gamma_{1j}$  pode, ainda, ser modelado por

$$\gamma_{1j} = \beta_1 + b_{1j}, \tag{2.3}$$

em que

- $\beta_1$  é o valor esperado de  $\gamma_{1j}$ ;
- $b_{1j}$  é um erro associado a  $\gamma_{1j}$ ,

com as suposições que os erros  $b_{1j}$ 's são independentes entre si e independentes dos erros  $\epsilon_{ij}$ 's, com  $b_{1j} \sim N(0, \sigma_b^2)$ , para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Se substituirmos (2.3) em (2.2) e fizermos  $\gamma_2 = \beta_2$ , obtemos

$$Y_{ij} = \beta_1 + \beta_2 \mathbf{Z}_{ij} + b_{1j} + \epsilon_{ij}, \qquad (2.4)$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Em (2.4) temos que  $\beta_1$  (o valor esperado dos  $\gamma_{1j}$ 's) é o valor médio dos rendimentos de todos os alunos considerados nas J classes da amostra, dado que o valor da variável Z é zero. Sendo assim, os erros  $b_{1j}$ 's associados aos  $\gamma_{1j}$ 's em (2.3) podem ser interpretados como "variações" na média geral dos rendimentos dos alunos entre as J classes consideradas no estudo.

Além das suposições consideradas anteriormente, poderia ser assumido que a magnitude da influência de Z no rendimento dos alunos depende da classe considerada. Assim, poderíamos ter

$$Y_{ij} = \gamma_{1j} + \gamma_{2j} Z_{ij} + \epsilon_{ij}, \qquad (2.5)$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, em que

 $Y_{ij}$ é o rendimento do *i*-ésimo aluno pertencente à *j*-ésima classe amostrada;

- $\gamma_{1j}$  é o rendimento médio dos alunos da *j*-ésima classe dado o valor nulo de Z;
- $\gamma_{2j}$  é o parâmetro associado à variável independente Z para os alunos da *j*-ésima classe, ou seja, é o acrésimo no valor médio do rendimento dos alunos pertencentes à *j*-ésima escola quando a variável Z é acrescida em uma unidade;
- $\epsilon_{ij}$  é um erro aleatório associado a  $Y_{ij}$ .

Os parâmetros  $\gamma_{1j} \in \gamma_{2j}$  podem ser dados por

$$\gamma_{1j} = \beta_1 + b_{1j};$$
  
 $\gamma_{2j} = \beta_2 + b_{2j},$ 
(2.6)

para  $j = 1, 2, \ldots, J$ , em que

- $\beta_1$  é o valor esperado de  $\gamma_{1j}$ ;
- $b_{1j}$  é um erro aleatório associado a  $\gamma_{1j}$ ;
- $\beta_2$  é o valor esperado de  $\gamma_{2j}$ ;
- $b_{2j}$  é um erro associado a  $\gamma_{2j}$ ,

com as suposições de que os  $b_{1j}$ 's são independentes entre si, os  $b_{2j}$ 's são também independentes entre si e independentes dos  $b_{1j}$ 's, e além disto tanto os  $b_{1j}$ 's quanto os  $b_{2j}$ 's são independentes dos  $\epsilon_{ij}$ 's, com  $b_{1j} \sim N(0, \sigma_b^2)$ ,  $b_{2j} \sim N(0, \sigma_b^2)$  e  $\epsilon_{ij} \sim N(0, \sigma^2)$  para  $i = 1, 2, \ldots, n_j$  e  $j = 1, 2, \ldots, J$ . A estrutura de variância e covariância entre os  $b_{1j}$ 's e os  $b_{2j}$ 's pode ser especificada segundo diferentes propostas, inclusive com  $\text{Cov}(b_{1j}, b_{2j}) \neq 0$ . Em West, Welch e Galecki (2007) temos alguns exemplos de matrizes de variância e covariância para os erros em modelos normais multiníveis tanto com dois níveis quanto com três níveis. Substituindo (2.6) em (2.5), obtemos

$$Y_{ij} = \beta_1 + b_{1j} + (\beta_2 + b_{2j})Z_{ij} + \epsilon_{ij},$$

o que leva a

$$Y_{ij} = \beta_1 + \beta_2 Z_{ij} + b_{1j} + b_{2j} Z_{ij} + \epsilon_{ij}.$$
(2.7)

Em (2.7), temos que

- $\beta_1$ é o rendimento médio para todos os alunos da amostra dado o valor nulo de Z;
- $\beta_2$  é o parâmetro associado à variável independente Z para todos os alunos das J classes consideradas na amostra, ou seja, é o acrésimo no valor médio do rendimento geral dos alunos quando a variável Z é acrescida em uma unidade.

Note que em (2.7) os valores dos  $b_{1j}$ 's podem ser interpretados como "variações" na média geral  $\beta_1$  dos rendimentos dos alunos (condicionais ao valor nulo da variável Z) entre as Jclasses. As quantidades representadas pelos  $b_{2j}$ 's podem ser interpretadas como "variações" na magnitude da influência da variável Z ( $\beta_2$ ) no rendimento médio dos alunos entre as classes consideradas.

A notação matricial para o modelo dado por (2.5) e (2.6) é

$$\mathbf{Y}_j = \mathbf{Z}_j \boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j; \tag{2.8}$$

$$\boldsymbol{\gamma}_j = \mathbf{T}_j \boldsymbol{\beta} + \boldsymbol{b}_j, \tag{2.9}$$

para j = 1, 2, ..., J, em que

$$\mathbf{Y}_{j} = (Y_{1j}, Y_{2j}, \dots Y_{n_{j}j})^{t}_{n_{j} \times 1}; \quad \mathbf{Z}_{j} = \begin{bmatrix} 1 & Z_{1j} \\ 1 & Z_{2j} \\ \vdots & \vdots \\ 1 & Z_{n_{j}j} \end{bmatrix}_{n_{j} \times 2}; \quad \boldsymbol{\gamma}_{j} = (\gamma_{1j}, \gamma_{2j})^{t}_{2 \times 1};$$

$$\boldsymbol{\epsilon}_{j} = (\epsilon_{1j}, \epsilon_{2j}, \dots, \epsilon_{n_{j}j})^{t}_{n_{j} \times 1}; \quad \mathbf{T}_{j} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2 \times 2}; \quad \boldsymbol{\beta} = (\beta_{1}, \beta_{2})^{t}_{2 \times 1}; \quad \boldsymbol{b}_{j} = (b_{1j}, b_{2j})^{t}_{2 \times 1}.$$

Substituindo (2.9) em (2.8), temos

$$\mathbf{Y}_{j} = \mathbf{Z}_{j} \left[ \mathbf{T}_{j} \boldsymbol{\beta} + \boldsymbol{b}_{j} \right] + \boldsymbol{\epsilon}_{j} = \mathbf{Z}_{j} \mathbf{T}_{j} \boldsymbol{\beta} + \mathbf{Z}_{j} \boldsymbol{b}_{j} + \boldsymbol{\epsilon}_{j}.$$
(2.10)

Fazendo  $\mathbf{X}_j = \mathbf{Z}_j \mathbf{T}_j$  em (2.10), obtemos

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j + \boldsymbol{\epsilon}_j, \ j = 1, 2, \dots, J.$$

Utilizando as suposições do modelo, obtemos que  $\boldsymbol{b}_j \sim N_2(\boldsymbol{0}, \boldsymbol{\Psi}_j)$ , em que

$$\Psi_j = \left[ \begin{array}{cc} \sigma_b^2 & 0 \\ 0 & \sigma_b^2 \end{array} \right]_{2 \times 2}$$

MANGHI, R. F.

Além, disso, temos  $\boldsymbol{\epsilon}_j \sim N_{n_j}(\mathbf{0}, \mathbf{R}_j)$ , em que  $\mathbf{R}_j = \sigma^2 \mathbf{I}_{n_j \times n_j}$ , sendo **I** a matriz identidade. Os  $\boldsymbol{\epsilon}_j$ 's são independentes entre si e independentes dos  $\boldsymbol{b}_j$ 's, para  $j = 1, 2, \ldots, J$ . De acordo com as distribuições de  $\boldsymbol{b}_j$  e  $\boldsymbol{\epsilon}_j$  e considerando a distribuição de  $\mathbf{Y}_j$  condicional ao valor de  $\boldsymbol{b}_j$  dada por

$$\mathbf{Y}_j | \boldsymbol{b}_j \sim N_{n_j} (\mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j, \mathbf{R}_j),$$

temos a distribuição conjunta de  $\mathbf{Y}_j$  <br/>e $\boldsymbol{b}_j$ dada por

$$\left[\begin{array}{c} \mathbf{Y}_{j} \\ \mathbf{b}_{j} \end{array}\right] \sim N_{n_{j}+2} \left( \left[\begin{array}{c} \mathbf{X}_{j}\boldsymbol{\beta} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{c} \mathbf{Z}_{j}\boldsymbol{\Psi}_{j}\mathbf{Z}_{j}^{t} + \mathbf{R}_{j} & \mathbf{Z}_{j}\boldsymbol{\Psi}_{j} \\ \boldsymbol{\Psi}_{j}\mathbf{Z}_{j}^{t} & \boldsymbol{\Psi}_{j} \end{array}\right] \right)$$

Portanto, a distribuição marginal de  $\mathbf{Y}_j$ é dada por

$$\mathbf{Y}_{j} \sim N_{n_{j}}(\mathbf{X}_{j}\boldsymbol{\beta}, \mathbf{Z}_{j}\boldsymbol{\Psi}_{j}\mathbf{Z}_{j}^{t} + \mathbf{R}_{j}),$$

ou ainda

$$\mathbf{Y}_j \sim N_{n_j}(\mathbf{X}_j \boldsymbol{\beta}, \mathbf{V}_j),$$

em que  $\mathbf{V}_j = \mathbf{Z}_j \Psi_j \mathbf{Z}_j^t + \mathbf{R}_j$ . Considerando uma notação para as J classes, sejam

$$\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_J^t)^t {}_{(\sum_{j=1}^J n_j) \times 1}; \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1^t | \mathbf{X}_2^t | \dots | \mathbf{X}_J^t \end{bmatrix}^t {}_{(\sum_{j=1}^J n_j) \times 2},$$
$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_J \end{bmatrix}_{(\sum_{j=1}^J n_j) \times 2J};$$
$$\mathbf{b} = (\mathbf{b}_1^t, \mathbf{b}_2^t, \dots, \mathbf{b}_J^t)^t_{2J \times 1}; \quad \mathbf{\epsilon} = (\mathbf{\epsilon}_1^t, \mathbf{\epsilon}_2^t, \dots, \mathbf{\epsilon}_J^t)^t_{(\sum_{j=1}^J n_j) \times 1}.$$

Assim, temos

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon}.$$

Desta forma, temos  $\boldsymbol{b} \sim N_{2J}(\boldsymbol{0}, \boldsymbol{\Psi}) \in \boldsymbol{\epsilon} \sim N_{\sum_{j=1}^{J} n_j}(\boldsymbol{0}, \mathbf{R})$ , em que

$$\Psi = \begin{bmatrix} \Psi_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Psi_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \Psi_J \end{bmatrix}_{2J \times 2J}; \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_J \end{bmatrix}_{(\sum_{j=1}^J n_j) \times (\sum_{j=1}^J n_j)}.$$

De acordo com as distribuições de b e  $\epsilon$  e considerando a distribuição de Y condicional ao valor de b dada por

$$\mathbf{Y}|\boldsymbol{b} \sim N_{\sum_{j=1}^{J} n_j}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b}, \mathbf{R}),$$

temos a distribuição conjunta de  ${\bf Y}$ e ${\pmb b}$ dada por

$$\left[\begin{array}{c} \mathbf{Y} \\ \mathbf{b} \end{array}\right] \sim N_{(\sum_{j=1}^{J} n_j)+2J} \left( \left[\begin{array}{c} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{c} \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \mathbf{R} & \mathbf{Z}\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\mathbf{Z}^t & \boldsymbol{\Psi} \end{array}\right] \right).$$

Portanto, a distribuição marginal de  $\mathbf{Y}$  é dada por

$$\mathbf{Y} \sim N_{\sum_{j=1}^{J} n_j}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \mathbf{R}),$$

ou ainda

$$\mathbf{Y} \sim N_{\sum_{j=1}^{J} n_j} \left( \mathbf{X} \boldsymbol{\beta}, \mathbf{V} \right),$$

sendo  $\mathbf{V} = \mathbf{Z} \mathbf{\Psi} \mathbf{Z}^t + \mathbf{R}.$ 

A partir do exemplo aqui considerado, pode ser constatado que as interpretações dos parâmetros em modelos multiníveis nem sempre são triviais, dependendo da estrutura de cada modelo e da alocação das variáveis preditoras. Maiores detalhes podem ser encontrados em Raudenbush e Bryk (2002) e em Natis (2000).

### 2.2 Definição do Modelo

Os modelos multiníveis com dois níveis podem ser definidos sob uma estrutura mais geral que aquela apresentada na Seção 2.1. Esta notação mais geral pode ser vista no Apêndice A.1. Quanto aos modelos multiníveis com três níveis, sua estrutura geral está ilustrada no Apêndice A.2. Nesta seção serão descritas as suposições sobre distribuições de probabilidade dos erros mais utilizadas no contexto destes modelos. Para o caso de modelos com três níveis, as suposições usuais são similares. Para o modelo com dois níveis, considere então

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j + \boldsymbol{\epsilon}_j,$$

para j = 1, 2, ..., J, com  $\mathbf{b}_j \sim N_P(\mathbf{0}, \mathbf{\Psi}_j)$  e  $\boldsymbol{\epsilon}_j \sim N_{n_j}(\mathbf{0}, \mathbf{R}_j)$ , sendo os  $\mathbf{b}_j$ 's independentes entre si e independentes dos  $\boldsymbol{\epsilon}_j$ 's, sendo estes últimos também independentes entre si. De acordo com as distribuições de  $\mathbf{b}_j$  e  $\boldsymbol{\epsilon}_j$  e considerando a distribuição de  $\mathbf{Y}_j$  condicional ao valor de  $\mathbf{b}_j$  dada por

$$\mathbf{Y}_{j}|\boldsymbol{b}_{j} \sim N_{n_{j}}\left(\mathbf{X}_{j}\boldsymbol{\beta} + \mathbf{Z}_{j}\boldsymbol{b}_{j}, \mathbf{R}_{j}\right), \qquad (2.11)$$

temos a distribuição conjunta de  $\mathbf{Y}_j$  e  $\boldsymbol{b}_j$  dada por

$$\left[ egin{array}{c} \mathbf{Y}_j \ m{b}_j \end{array} 
ight] \sim N_{n_j+P} \left( \left[ egin{array}{c} \mathbf{X}_j m{eta} \ m{0} \end{array} 
ight], \left[ egin{array}{c} \mathbf{Z}_j m{\Psi}_j \mathbf{Z}_j{}^t + \mathbf{R}_j & \mathbf{Z}_j m{\Psi}_j \ m{\Psi}_j \mathbf{Z}_j{}^t & \mathbf{W}_j \end{array} 
ight] 
ight),$$

para  $j = 1, 2, \ldots, J$ . Assim, temos a distribuição marginal de  $\mathbf{Y}_j$  dada por

$$\mathbf{Y}_j \sim N_{n_j}(\mathbf{X}_j \boldsymbol{\beta}, \mathbf{V}_j), \qquad (2.12)$$

em que  $\mathbf{V}_j = \mathbf{Z}_j \mathbf{\Psi}_j \mathbf{Z}_j^{\ t} + \mathbf{R}_j$ , para j = 1, 2, ..., J. A matriz  $\mathbf{\Psi}_j$  tem dimensão  $P \times P$ e a matriz  $\mathbf{R}_j$  tem dimensão  $n_j \times n_j$ , sendo ambas simétricas e positivas definidas para j = 1, 2, ..., J. Geralmente a matriz  $\mathbf{R}_j$  é diagonal, sendo em alguns casos  $\mathbf{R}_j = \sigma^2 \mathbf{I}$ , em que  $\mathbf{I}$  é a matriz identidade de dimensão  $n_j \times n_j$ . Utilizando a notação matricial geral para todas as J unidades ou "clusters" pertencentes ao nível 2 do modelo, ou seja, considerando a notação

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon},$$

com Y, X,  $\beta$ , Z,  $b \in \epsilon$  definidos como no apêndice A.1, obtemos que  $b \sim N_{JP}(\mathbf{0}, \Psi) \in \epsilon \sim N_{\sum_{j=1}^{J} n_j}(\mathbf{0}, \mathbf{R})$ , em que

Aqui, o vetor  $\boldsymbol{b}$  é independente do vetor  $\boldsymbol{\epsilon}$ . De acordo com as distribuições destes vetores e considerando a distribuição de  $\mathbf{Y}$  condicional ao valor de  $\boldsymbol{b}$  dada por

$$\mathbf{Y}|\boldsymbol{b} \sim N_{\sum_{j=1}^{J} n_j}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b}, \mathbf{R}),$$

temos a distribuição conjunta de  $\mathbf{Y} \in \boldsymbol{b}$  dada por

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{b} \end{bmatrix} \sim N_{(\sum_{j=1}^{J} n_j) + JP} \left( \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \mathbf{R} & \mathbf{Z}\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\mathbf{Z}^t & \boldsymbol{\Psi} \end{bmatrix} \right)$$

Neste caso as matrizes  $\Psi$  e **R** são ambas positivas definidas. Como cada uma das matrizes  $\Psi_j$ 's possui, no máximo, P(P+1)/2 elementos distintos, temos que a matriz  $\Psi$  possui, no máximo, JP(P+1)/2 elementos distintos. Analogamente, cada uma das matrizes  $\mathbf{R}_j$ 's possui, no máximo,  $n_j(n_j+1)/2$  elementos distintos. Assim, a matriz **R** possui, no máximo,  $\frac{1}{2} \sum_{j=1}^J n_j(n_j+1)$  elementos distintos. Assumindo um vetor  $\boldsymbol{\theta}_{\Psi}$  contendo os parâmetros distintos de  $\Psi$  e  $\boldsymbol{\theta}_{\mathbf{R}}$  um vetor que contém todos os parâmetros distintos de  $\mathbf{R}$ , podemos definir um vetor  $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\Psi}^t, \boldsymbol{\theta}_{\mathbf{R}}^t)^t$  que contém todos os parâmetros distintos distintos distintos de  $\Psi$  e  $\mathbf{R}$  conjuntamente.

Assim, temos a distribuição marginal de Y dada por

$$\mathbf{Y} \sim N_{\sum_{j=1}^{J} n_j} \left( \mathbf{X} \boldsymbol{\beta}, \mathbf{V} \right),$$

com

$$\mathbf{V} = \left[egin{array}{ccccccccc} \mathbf{V}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \dots & \mathbf{0} \\ dots & dots & \ddots & dots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{V}_J \end{array}
ight]_{(\sum_{j=1}^J n_j) imes (\sum_{j=1}^J n_j)},$$

ou seja,  $\mathbf{V} = \mathbf{Z} \mathbf{\Psi} \mathbf{Z}^t + \mathbf{R}$ . Neste caso, é suficiente apenas que a matriz  $\mathbf{V}$  seja positiva definida. Porém, se o intuito é preservar a estrutura original do modelo, os parâmetros em  $\mathbf{V}$  devem ser estimados sob a restrição de que as estimativas de  $\mathbf{\Psi}$  e  $\mathbf{R}$  sejam ambas positivas definidas. Na seção seguinte temos a abordagem de métodos de estimação nos modelos multiníveis com distribuição normal dos erros, a partir das suposições aqui ilustradas.

### 2.3 Métodos de Estimação

Para o contexto de modelos normais multiníveis, diversos trabalhos foram desenvolvidos no que se refere a técnicas de estimação dos parâmetros. Os dois métodos de estimação mais utilizados em modelos normais multiníveis são os métodos de **Máxima Verossimilhança** (MV) e **Máxima Verossimilhança Restrita** (MVR). Através destes, são obtidas as estimativas tanto dos parâmetros dos efeitos fixos quanto dos parâmetros referentes aos componentes de variância e covariância do modelo.

Existem muitos algoritmos que podem ser utilizados na obtenção das estimativas dos parâmetros, tanto por MV como por MVR. O algoritmo mais utilizado é o EM (*Expectation-Maximization*). Para maiores detalhes acerca deste algoritmo, ver Dempster, Laird e Rubin (1977) e Dempster, Rubin e Tsutakawa (1981).

Para o desenvolvimento dos métodos de estimação MV e MVR para modelos normais multiníveis, será considerado o modelo multinível com dois níveis dado em A.1, segundo as suposições de distribuições de probabilidades dos erros dadas na Seção 2.2. O procedimento para a estimação dos parâmetros em modelos com três níveis pode ser realizado de maneira similar.

#### 2.3.1 Máxima Verossimilhança (MV)

Em geral, o método de MV consiste em obter estimativas de parâmetros otimizando uma função de verossimilhança. Para a aplicação deste método, primeiro é obtida a verossimilhança como uma função dos parâmetros de um determinado modelo estatístico, baseada em suposições a respeito de distribuições de probabilidade. As estimativas de máxima verossimilhança dos parâmetros são os valores que maximizam a função de verossimilhança, isto é, os valores dos parâmetros que tornam os valores observados da variável dependente mais "prováveis", dadas as suposições de distribuições de probabilidade. Veja Casella e Berger (2002) para uma discussão mais detalhada acerca do referido método. De acordo com a distribuição marginal de  $\mathbf{Y}_j$  em (2.12), e considerando os valores observados de  $\mathbf{Y}_j$  denotados por  $\mathbf{y}_j$ , temos a função de densidade dada por

$$f(\mathbf{y}_j;\boldsymbol{\beta},\boldsymbol{\theta}) = (2\pi)^{-\frac{n_j}{2}} |\mathbf{V}_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^t \mathbf{V}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})\right],$$

para j = 1, 2, ..., J, em que  $|\mathbf{A}|$  refere-se ao determinante da matriz  $\mathbf{A}$ . Note que os elementos das matrizes  $\mathbf{V}_j$ , j = 1, 2, ..., J, são funções dos parâmetros pertencentes ao vetor  $\boldsymbol{\theta}$ . A contribuição do *j*-ésimo grupo ou "cluster" na função de verossimilhança é dada por

$$L_j(\boldsymbol{\beta}, \boldsymbol{\theta}) = f(\mathbf{y}_j; \boldsymbol{\beta}, \boldsymbol{\theta}), \qquad (2.13)$$

portanto, a função de verossimilhança é dada pelo produto das J contribuições definidas em (2.13), ou seja,

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{J} L_j(\boldsymbol{\beta}, \boldsymbol{\theta}) =$$
$$\prod_{j=1}^{J} (2\pi)^{-\frac{n_j}{2}} |\mathbf{V}_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})\right].$$

Assim, o logaritmo da função de verossimilhança é definido, neste caso, por

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = \log(L(\boldsymbol{\beta}, \boldsymbol{\theta})) = -\frac{1}{2}\log(2\pi)\sum_{j=1}^{J}n_j - \frac{1}{2}\sum_{j=1}^{J}\log|\mathbf{V}_j| - \frac{1}{2}\sum_{j=1}^{J}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^t \mathbf{V}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta}),$$
(2.14)

em que  $\log(x)$  refere-se ao logaritmo neperiano de x.

Para o vetor de parâmetros de variância e covariância  $\theta$  conhecido, os únicos parâmetros a serem estimados são aqueles pertencentes ao vetor de parâmetros dos efeitos fixos em  $\beta$ . Assim, (2.14) torna-se uma função apenas de  $\beta$ , e sua maximização é equivalente a minimizar a função

$$q(\boldsymbol{\beta}) = \frac{1}{2} \sum_{j=1}^{J} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}).$$
(2.15)

Esta função se assemelha à forma matricial para a soma de quadrado dos erros que é minimizada no modelo normal linear, porém com a inclusão das matrizes  $\mathbf{V}_{j}^{-1}$ 's. Neste caso, minimizar (2.15) é equivalente a utilizar o método de **Mínimos Quadrados Generalizados** (MQG), obtendo assim o valor ótimo de  $\boldsymbol{\beta}$  dado por

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \mathbf{X}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{X}_{j}\right)^{-1} \sum_{j=1}^{J} \mathbf{X}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{y}_{j}, \qquad (2.16)$$

ou ainda,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}, \qquad (2.17)$$

em que  $\mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_J^t)^t$ . Neste caso,  $\hat{\boldsymbol{\beta}}$  tem a menor variância dentre todos os estimadores lineares não-viesados de  $\boldsymbol{\beta}$ , ou seja,  $\hat{\boldsymbol{\beta}}$  é chamado de *Best Linear Unbiased Estimator* (BLUE).

Para o caso de  $\boldsymbol{\theta}$  desconhecido, é necessário maximizar (2.14) conjuntamente em  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$ . Contudo, pode ser utilizado o logaritmo da função de verossimilhança perfilada  $l_{pr}(\boldsymbol{\theta})$ . Esta função é obtida a partir de (2.14), substituindo  $\boldsymbol{\beta}$  pela expressão definida em (2.16). A função resultante é dada por

$$l_{pr}(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi)\sum_{j=1}^{J}n_j - \frac{1}{2}\sum_{j=1}^{J}\log|\mathbf{V}_j| - \frac{1}{2}\sum_{j=1}^{J}\mathbf{r}_j^{\ t}\mathbf{V}_j^{\ -1}\mathbf{r}_j,$$
(2.18)

em que

$$\mathbf{r}_{j} = \mathbf{y}_{j} - \mathbf{X}_{j} \hat{\boldsymbol{\beta}} = \mathbf{y}_{j} - \mathbf{X}_{j} \left( \sum_{j=1}^{J} \mathbf{X}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{X}_{j} \right)^{-1} \sum_{j=1}^{J} \mathbf{X}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{y}_{j} = \mathbf{y}_{j} - \mathbf{X}_{j} (\mathbf{X}^{t} \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^{t} \mathbf{V}^{-1} \mathbf{y}.$$

Em geral, a maximização de (2.18) em  $\boldsymbol{\theta}$  remete a um processo de otimização não linear, com restrições nos componentes de  $\boldsymbol{\theta}$  para garantir que as matrizes  $\Psi_j$  e  $\mathbf{R}_j$  sejam positivas definidas, para j = 1, 2, ..., J. Como não existe solução em forma fechada para este problema, a estimativa de  $\boldsymbol{\theta}$  pode ser obtida por métodos iterativos de otimização, como por exemplo o método escore de Fisher (ver, por exemplo, Longford(1987)). Após a estimativa de  $\boldsymbol{\theta}$  ser obtida, o valor da estimativa de  $\boldsymbol{\beta}$  pode ser obtido, inicialmente, obtendo a estimativa de  $\mathbf{V}$  dada por

$$\hat{\mathbf{V}} = \mathbf{Z}\hat{\Psi}\mathbf{Z}^t + \hat{\mathbf{R}}.$$
(2.19)

Finalmente, utilizamos  $\hat{\mathbf{V}}$  no lugar de  $\mathbf{V}$  em (2.17) para obter a estimativa de  $\boldsymbol{\beta}$  sendo dada por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{y}.$$
(2.20)

Devido ao fato de que em (2.20) estamos utilizando  $\hat{\mathbf{V}}$ , dizemos que  $\hat{\boldsymbol{\beta}}$  é o estimador empírico de menor variância entre todos os estimadores lineares não viesados de  $\boldsymbol{\beta}$ , ou seja,  $\hat{\boldsymbol{\beta}}$  é chamado de *Empirical Best Linear Unbiased Estimator* (EBLUE).

Para a obtenção da matriz de variância e covariância de  $\hat{\beta}$ , no caso de  $\theta$  conhecido podemos fazer

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1}.$$
(2.21)

No caso de  $\boldsymbol{\theta}$  desconhecido, é necessário substituir V por seu valor estimado  $\hat{\mathbf{V}}$ , e assim obter uma estimativa da matriz de variância e covariância de  $\hat{\boldsymbol{\beta}}$ , dada por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1}.$$
(2.22)

A expressão (2.22) pode ser utilizada no contexto de MV com  $\theta$  desconhecido ou através do método de máxima verossimilhança restrita (que será definido posteriormente).

Considerando D parâmetros em  $\boldsymbol{\theta}$ , temos a matriz de informação de Fisher para este vetor dada por

$$\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \left\{ \mathbf{K}_{\theta_l \theta_z} \right\},\,$$

em que

$$\mathbf{K}_{\theta_l \theta_z} = \frac{1}{2} \sum_{j=1}^{J} tr\left( \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_l} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_z} \right),$$

MANGHI, R. F.

para l, z = 1, 2, ..., D, sendo  $tr(\mathbf{A})$  o valor do traço da matriz  $\mathbf{A}$ .

#### 2.3.2 Máxima Verossimilhança Restrita (MVR)

O método de Máxima Verossimilhança Restrita (MVR) é uma proposta alternativa para a obtenção das estimativas dos parâmetros de variância e covariância nos modelos multiníveis. Patterson e Thompson (1971) propuseram o referido método, que leva em consideração a perda nos graus de liberdade resultante da estimação dos parâmetros dos efeitos fixos dos modelos. Derivações alternativas e mais gerais do método MVR são encontradas em Harville (1977), Cooper e Thompson (1977) e Verbyla (1990). Em algumas situações o método de MVR é preferido ao método de MV, pois este primeiro fornece estimativas não viesadas dos parâmetros de variância e covariância. Em Natis (2000) temos o desenvolvimento dos métodos de MV e de MVR no contexto de modelos normais multiníveis com dois níveis através do uso do modelo linear geral de Bayes (ver Lindley e Smith (1972)).

A estimativa via MVR do vetor  $\boldsymbol{\theta}$  é obtida a partir da otimização do logaritmo da função de verossimilhança restrita

$$l_{R} = -\frac{1}{2} \left[ \left( \sum_{j=1}^{J} n_{j} \right) - \left( \sum_{p=1}^{P} Q_{p} \right) \right] \log(2\pi) - \frac{1}{2} \sum_{j=1}^{J} \log|\mathbf{V}_{j}| -\frac{1}{2} \sum_{j=1}^{J} \mathbf{r}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{r}_{j} - \frac{1}{2} \sum_{j=1}^{J} \log|\mathbf{X}_{j}^{t} \mathbf{V}_{j}^{-1} \mathbf{X}_{j}|.$$

$$(2.23)$$

Aqui,  $\sum_{j=1}^{J} n_j$  é o número total de indivíduos pertencentes ao nível 1 do modelo e  $\sum_{p=1}^{P} Q_p$ é o número total de parâmetros que representam os efeitos fixos considerados no modelo. A ideia da utilização de (2.23) é obter uma estimativa não viesada de  $\boldsymbol{\theta}$ . Após obter a estimativa do referido vetor através da maximização de (2.23), obtém-se  $\hat{\mathbf{V}}$  através de (2.19),  $\hat{\boldsymbol{\beta}}$  por meio de (2.20) e  $\hat{Var}(\hat{\boldsymbol{\beta}})$  a partir de (2.22). É válido notar que, embora a equação que fornece a estimativa de  $\boldsymbol{\beta}$  no uso do método de Máxima Verossimilhança com  $\boldsymbol{\theta}$  desconhecido e no uso do método de Máxima Verossimilhança Restrita seja a mesma, em geral estas fornecem diferentes estimativas do vetor de parâmetros dos efeitos parâmetros de variância e covariância  $\boldsymbol{\theta}$ .

Além dos parâmetros em  $\beta$  referentes aos efeitos fixos e dos parâmetros em  $\theta$  que representam os parâmetros de variância e covariância nos modelos multiníveis, temos também a predição dos efeitos aleatórios no vetor **b**, através de um conhecido método que será descrito no tópico seguinte.

#### 2.3.3 Detalhes Computacionais no uso de MV e MVR

Alguns importantes detalhes devem ser considerados nos usos dos métodos de máxima verossimilhança e máxima verossimilhança restrita no contexto de modelos multiníveis. Uma importante etapa anterior à estimação dos modelos é a especificação das estruturas das matrizes  $\Psi_j$ 's e  $\mathbf{R}_j$ 's e, consequentemente, das matrizes  $\Psi$  e  $\mathbf{R}$ . Este procedimento pode implicar numa estimativa de  $\mathbf{V}$  mais próxima de seu valor verdadeiro. Contudo, uma má especificação das matrizes citadas pode remeter a valores de  $\boldsymbol{\theta}$  muito próximos da fronteira do espaço paramétrico, consequentemente o uso de alguns algoritmos nos métodos de MV ou de MVR pode apresentar problemas de convergência. Na prática, os programas já implementados em softwares estatísticos, como por exemplo o software gratuito  $\mathbf{R}$  disponível em http://www.r-project.org, possuem estruturas pré determinadas para as matrizes de variância e covariância. Para uma introdução a alguns programas já implementados no referido software ver, por exemplo, Crawley (2009).

Outro problema que requer atenção é a certificação de que as matrizes  $\Psi_j$ 's e  $\mathbf{R}_j$ 's são positivas definidas. Alguns procedimentos para garantir esta condição e, sendo assim, garantir que a estimação dos parâmetros será bem executada é a escolha de valores iniciais alternativos para a implementação de algoritmos que fazem uso de métodos iterativos, bem como a ocasional mudança de escala nas variáveis preditoras do modelo. Estes procedimentos, bem como uma discussão sobre os algoritmos mais utilizados no uso de MV e MVR para estimação dos parâmetros em  $\boldsymbol{\beta} \in \boldsymbol{\theta}$  podem ser vistos em West, Welch e Galecki (2007).

Na seção seguinte, temos uma discussão sobre algumas estratégias na construção dos modelos multiníveis. Estas estratégias possuem o objetivo de especificar estruturas de modelos o mais próximo possível da verdadeira estrutura dos dados, através de etapas que envolvem a especificação de diferentes aspectos dos modelos propriamente ditos.

#### 2.4 Predição dos Efeitos Aleatórios

Um dos aspectos dos modelos multiníveis, diferentemente dos modelos lineares clássicos, é a presença de diferentes níveis e, sendo assim, de variáveis aleatórias que descrevem variações nos parâmetros pertencentes a um determinado nível. Estas variáveis não podem ser estimadas pelos métodos acima descritos, porém assumem conjuntamente uma distribuição de probabilidade, usualmente a distribuição normal multivariada. Em alguns estudos, além da estrutura de variância e covariância associada a estas variáveis, sua predição fornece importante informação a respeito do comportamento das variáveis dependentes pertencentes a um determinado grupo ou "cluster".

A predição dos erros pertencentes à modelagem de parâmetros de um determinado nível pode ser obtida através de preditores lineares conhecidos como preditores de Bayes empíricos. Harville (1976) obteve este preditor através de uma extensão do Teorema de Gauss-Markov aplicado a modelos com presença de efeitos aleatórios. Laird (1982) e Laird e Ware (1982) consideram o uso do preditor linear de Bayes empírico no contexto de modelos mistos para dados com estrutura longitudinal.

Considerando (2.11) e (2.12), temos que a esperança condicional do vetor  $\mathbf{b}_j$  dado o valor observado de  $\mathbf{Y}_j = \mathbf{y}_j$  é dada por

$$E(\boldsymbol{b}_j | \mathbf{Y}_j = \mathbf{y}_j) = \boldsymbol{\Psi}_j \mathbf{Z}_j^{\ t} \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}),$$

para j = 1, 2, ..., J. Assim, quando o vetor  $\boldsymbol{\theta}$  é conhecido, temos a predição de  $\boldsymbol{b}_j$  dada por

$$\hat{\boldsymbol{b}}_j = \boldsymbol{\Psi}_j \mathbf{Z}_j{}^t \mathbf{V}_j{}^{-1} (\mathbf{Y}_j - \mathbf{X}_j \hat{\boldsymbol{\beta}}),$$

para j = 1, 2, ..., J. Para  $\boldsymbol{\theta}$  desconhecido, substituem-se as matrizes  $\mathbf{V}_j$ 's por suas respectivas estimativas. Considerando as J unidades do nível 2, temos

$$\hat{\boldsymbol{b}} = \boldsymbol{\Psi} \mathbf{Z}^t \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$
(2.24)

O preditor  $\hat{\boldsymbol{b}}$  possui a menor variância dentre todos os preditores lineares não viesados de  $E(\boldsymbol{b})$ . Sendo assim,  $\hat{\boldsymbol{b}}$  é chamado de *Best Linear Unbiased Predictor* (BLUP). Quando o vetor de parâmetros de variância e covariância  $\boldsymbol{\theta}$  é desconhecido, sua estimativa é utilizada nas matrizes  $\boldsymbol{\Psi} \in \mathbf{V}$ , bem como na obtenção do EBLUE de  $\boldsymbol{\beta}$ . Neste caso, dizemos que  $\hat{\boldsymbol{b}}$  possui a menor variância dentre todos os preditores lineares empíricos não viesados de  $E(\boldsymbol{b})$ , sendo denominado de *Empirical Best Linear Unbiased Predictor* (EBLUP).

### 2.5 Coeficiente de Correlação Intraclasse

Em geral, o Coeficiente de Correlação Intraclasse, ou Intraclass Correlation Coefficient (ICC), é uma quantidade que descreve a similaridade (ou homogeneidade) das medidas da variável resposta dentro de um determinado "cluster" ou em uma determinada unidade experimental (para o caso de dados com medidas repetidas ou estudos longitudinais). O ICC pode ser definido de diferentes formas, de acordo com a quantidade de níveis dos modelos considerados (ver West, Welch e Galecki (2007)). Para o modelo dado em (2.2) e (2.3), temos que

$$\operatorname{Cov}(Y_{ij}, Y_{i'j}) = \operatorname{Cov}(\beta_1 + \beta_2 Z_{ij} + b_{1j} + \epsilon_{ij}, \beta_1 + \beta_2 Z_{i'j} + b_{1j} + \epsilon_{i'j}) = \operatorname{Cov}(b_{1j} + \epsilon_{ij}, b_{1j} + \epsilon_{i'j}) = \operatorname{Var}(b_{1j}) = \sigma^2_b,$$

para todo  $i \neq i' = 1, 2, \dots, n_j$  e  $j = 1, 2, \dots, J$ . Por outro lado,

$$\operatorname{Var}(Y_{ij}) = \operatorname{Var}(Y_{i'j}) = \sigma^2{}_b + \sigma^2,$$

para  $i, i' = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Portanto, a correlação entre os valores dos rendimentos de dois alunos pertencentes à *j*-ésima classe é dada por

ICC = Cor(
$$Y_{ij}, Y_{i'j}$$
) =  $\frac{Cov(Y_{ij}, Y_{i'j})}{\sqrt{Var(Y_{ij})Var(Y_{i'j})}} = \frac{\sigma^2{}_b}{\sigma^2{}_b + \sigma^2},$  (2.25)

para todo  $i \neq i' = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Assim, o valor dado em (2.25) refere-se ao grau de similaridade entre os valores dos rendimentos de diferentes alunos numa mesma classe. Portanto, quanto maior o valor do coeficiente de correlação intraclasse, maior a similaridade entre os alunos em uma determinada classe.

MANGHI, R. F.

Uma vez descritos os métodos de estimação dos parâmetros e a técnica mais utilizada para predição dos erros aleatórios que explicam parâmetros de determinados níveis dos modelos normais multiníveis, temos a seguir o desenvolvimento de testes de hipóteses acerca dos parâmetros definidos nestes modelos.

#### 2.6 Testes de Hipóteses

Uma importante etapa na proposta de modelos multiníveis é a realização de testes de hipóteses acerca da significância de seus parâmetros. Neste contexto, os parâmetros referentes aos efeitos fixos do modelo, contidos no vetor  $\beta$ , podem ser testados com respeito às suas respectivas significâncias segundo vários tipos de testes, o mesmo sendo válido para os parâmetros de variância e covariância em  $\theta$  e, consequentemente, em V. Em Natis (2000) e Raudenbush e Bryk (2002) temos uma discussão sobre os procedimentos que serão detalhados nas próximas seções.

#### 2.6.1 Testes de Hipóteses para os Efeitos Fixos

#### Razão de Verossimilhanças (RV)

Os testes da Razão de Verossimilhanças (RV) compreendem uma classe de testes de hipóteses baseados na comparação das funções de verossimilhança de dois modelos (geralmente chamados de modelo completo e modelo reduzido, em que os parâmetros do modelo reduzido constituem um subconjunto dos parâmetros do modelo completo), uma vez definidas as hipóteses a serem testadas. Denotando por  $\hat{L}_{red}$  e  $\hat{L}_{comp}$  os valores das funções de verossimilhança avaliadas nas estimativas de máxima verossimilhança dos parâmetros para os modelos reduzido e completo respectivamente, temos que a estatística do teste é obtida fazendo

$$RV = -2\log\left(\frac{\hat{L}_{red}}{\hat{L}_{comp}}\right) = 2(\hat{l}_{comp} - \hat{l}_{red}), \qquad (2.26)$$

em que  $\hat{l}_{red}$  e  $\hat{l}_{comp}$  são os valores dos logaritmos das funções de verossimilhanças dos modelos reduzido e completo, respectivamente, avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Como exemplo do uso deste teste em modelos multiníveis, suponha que possamos particionar o vetor de parâmetros  $\boldsymbol{\beta}$  em  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\ t}, \boldsymbol{\beta}_2^{\ t})^t$ . Assim,
caso o interesse seja testar as hipóteses

$$H_0: \quad \boldsymbol{\beta}_1 = \mathbf{0},$$
$$H_1: \quad \boldsymbol{\beta}_1 \neq \mathbf{0},$$

Temos que, assintoticamente e sob a hipótese  $H_0$ , (2.26) segue aproximadamente uma distribuição qui-quadrado, com o valor dos graus de liberdade dado pela diferença entre o número de parâmetros contidos no vetor  $\beta$  e o número de parâmetros contidos em  $\beta_2$ . Para a realização deste teste é necessário assumir que a estrutura de variância e covariância dos modelos reduzido e completo são as mesmas. Quanto ao uso das estimativas de máxima verossimilhança em (2.26), temos que o uso das estimativas obtidas segundo o método de MVR não são apropriadas neste contexto (ver Morrell (1998), Pinheiro e Bates (2000) e Verbeke e Molenberghs (2000)).

#### Testes do tipo t

Em situações em que o interesse é testar a significância de um único parâmetro referente a um efeito fixo de uma covariável, o uso de testes do tipo t pode ser um procedimento alternativo ao uso do teste da razão de verossimilhanças. Para um determinado parâmetro  $\beta_{pq}$  contido no vetor de parâmetros  $\boldsymbol{\beta}$ , suponha que queiramos testar

$$H_0: \quad \beta_{pq} = 0,$$
$$H_1: \quad \beta_{pq} \neq 0.$$

Neste caso, a estatística do teste é dada por

$$t = \frac{\hat{\beta}_{pq}}{\widehat{\exp}(\hat{\beta}_{pq})},\tag{2.27}$$

onde  $\hat{\beta}_{pq}$  é o valor da estimativa de  $\beta_{pq}$  e  $\hat{ep}(\hat{\beta}_{pq})$  é o valor da estimativa do erro-padrão de  $\hat{\beta}_{pq}$ . Embora este teste seja utilizado no contexto de modelos multiníveis, é importante considerar o fato de que (2.27) em geral não segue uma distribuição *t*-Student exata sob a hipótese  $H_0$ . Além disto, o número de graus de liberdade associado a (2.27) sob  $H_0$  não é igual à diferença entre o número total de observações da variável resposta e o número de parâmetros em  $\beta$  a serem estimados. Contudo, existem alguns métodos para a obtenção de valores aproximados dos graus de liberdade a serem utilizados no teste (ver Kenward e Roger (1997) e Satterthwaite (1946)). Um método de obtenção dos graus de liberdades associados a estes tipos de testes será ilustrado posteriormente.

### Testes do tipo F

Testes do tipo F podem ser utilizados para testar hipóteses lineares acerca de múltiplos parâmetros no vetor  $\beta$ . Por exemplo, pode ser de interesse testar se alguns dos parâmetros associados com determinados níveis de um fator fixo são diferentes de zero. Em geral, quando o interesse é testar hipóteses do tipo

$$H_0: \quad \mathbf{A}\boldsymbol{\beta} = \mathbf{0},$$
$$H_1: \quad \mathbf{A}\boldsymbol{\beta} \neq \mathbf{0},$$

em que  $\mathbf{A}$  é uma matriz conhecida, a estatística F definida por

$$F = \frac{\hat{\boldsymbol{\beta}}^{t} \mathbf{A}^{t} [\mathbf{A} (\mathbf{X}^{t} \hat{\mathbf{V}} \mathbf{X})^{-1} \mathbf{A}^{t}]^{-1} \mathbf{A} \hat{\boldsymbol{\beta}}}{p(\mathbf{A})}$$
(2.28)

segue, sob a hipótese  $H_0$ , aproximadamente uma distribuição F, com graus de liberdade do numerador igual ao posto da matriz  $\mathbf{A}$   $(p(\mathbf{A}))$ , sendo  $p(\mathbf{A})$  dado pelo número de linhas (ou colunas) linearmente independentes da matriz  $\mathbf{A}$ . Os graus de liberdade do denominador da referida estatística podem ser aproximados por vários métodos (ver Verbeke e Molenberghs (2000), Kenward e Roger (1997) e Satterthwaite (1946)). Estes métodos são utilizados também para a obtenção dos graus de liberdade utilizados nos testes do tipo t, e levam em consideração a presença dos efeitos aleatórios e de erros correlacionados nos modelos considerados. Diferentes métodos podem remeter a diferentes valores dos graus de liberdade aproximados.

Nos exemplos de aplicações que serão apresentados a seguir, será adotado o método de obtenção dos graus de liberdade para os testes do tipo t e do tipo F condicionais sugerido em Pinheiro e Bates (2000), já implementado na função lme do software R. O referido método leva em consideração o número de grupos ou "clusters" pertencentes a cada nível dos dados, bem como o número de parâmetros dos efeitos fixos pertencentes a

cada nível. Um parâmetro fixo é dito pertencente ao *b*-ésimo nível do modelo se ele está associado a uma variável pertencente ao *b*-ésimo nível de hierarquia dos dados. No caso da existência de interação entre variáveis de diferentes níveis, o parâmetro associado a esta interação é dito pertencente ao menor nível encontrado dentre todas as variáveis presentes na determinada interação. Considerando um modelo multinível com B níveis, temos os graus de liberdade associados aos testes do tipo t e aos testes do tipo F condicionais (para o cálculo do denominador da estatística do teste) calculados por

$$GL_b = \#Gr_b - (\#Gr_{b+1} + \#Par_b),$$

para b = 1, 2, ..., B, em que:

 $#Gr_b$  é o número de observações (ou grupos) pertencentes a<br/>ob-ésimo nível dos dados;  $#Par_b$  é o número de parâmetros pertencentes a<br/>ob-ésimo nível do modelo.

Caso haja intercepto, este é considerado como pertencente a um nível de ordem B + 1, tendo este nível apenas um grupo, ou seja, adota-se  $\#Gr_{B+1} = 1$  e  $\#Par_{B+1} = 1$ . Contudo, para testar a significância deste parâmetro, os graus de liberdade associados ao teste do tipo t (e ao denominador da distribuição F para o teste condicional) são os mesmos utilizados para testar os parâmetros pertencentes ao primeiro nível do modelo, ou seja, é utilizado o valor de  $GL_1$ .

#### 2.6.2 Testes de Hipóteses para os Parâmetros de Variância e Covariância

### Razão de Verossimilhanças Restritas (RVR)

Quando o interesse é a realização de testes de hipóteses acerca dos parâmetros de variância e covariância nos modelos normais multiníveis, é recomendável que o teste da razão de verossimilhanças use as estimativas de máxima verossimilhança restrita dos referidos parâmetros, tanto no modelo de referência quanto no modelo reduzido. O método de MVR, em geral, reduz o viés das estimativas dos parâmetros em  $\theta$  em relação a estas mesmas estimativas obtidas por MV (ver Morrell (1998)). Para a realização deste teste é necessário que os modelos de referência e reduzido apresentem o mesmo conjunto de parâmetros em  $\beta$ . Denotando por  $\hat{L}_{R_{red}}$  e  $\hat{L}_{R_{comp}}$  os valores das funções de verossimilhança

restrita avaliadas nas estimativas de máxima verossimilhança restrita dos parâmetros para os modelos reduzido e completo, respectivamente, temos que a estatística do teste é dada por

$$RVR = -2\log\left(\frac{\hat{L}_{R_{red}}}{\hat{L}_{R_{comp}}}\right) = 2(\hat{l}_{R_{comp}} - \hat{l}_{R_{red}}), \qquad (2.29)$$

em que  $\hat{l}_{R_{red}}$  e  $\hat{l}_{R_{comp}}$  são os valores dos logaritmos das funções de verossimilhança restrita dos modelos reduzido e completo, respectivamente, avaliadas nas estimativas de máxima verossimilhança restrita dos parâmetros. Diferentemente do teste de razão de verossimilhanças para os parâmetros em  $\beta$ , a distribuição assintótica de (2.29) sob a hipótese nula depende dos valores dos parâmetros de variância e covariância sob a referida hipótese, mais especificamente se alguns destes valores encontram-se na fronteira do espaço paramétrico.

Quando os parâmetros de variância e covariância sob a hipótese nula não estão na fronteira do espaço paramétrico (por exemplo, se é de interesse testar se alguma covariância entre efeitos aleatórios é nula), temos que (2.29) segue assintoticamente uma distribuição qui-quadrado com o número de graus de liberdade dado pela diferença entre o número de parâmetros de variância e covariância do modelo de referência (modelo sob a hipótese alternativa) e o número de parâmetros do mesmo tipo a serem estimados no modelo reduzido (modelo sob a hipótese nula).

Para o caso onde pelo menos um dos parâmetros de variância e covariância encontramse na fronteira do espaço paramétrico sob a hipótese nula, o procedimento para a realização do teste de Razão de Verossimilhanças Restritas sofre algumas alterações. Por exemplo, no caso da existência de um único efeito aleatório no modelo e sendo de interesse testar se este efeito aleatório deve ser omitido, temos então as hipóteses

$$H_0: \quad \sigma^2{}_b = 0,$$
$$H_1: \quad \sigma^2{}_b \neq 0,$$

em que  $\sigma_b^2$  é a variância do referido efeito. Neste caso, Self e Liang (1987), Stram e Lee (1994) e Verbeke e Molenberghs (2000) mostraram que a estatística do teste RVR segue assintoticamente uma distribuição que é uma mistura de qui-quadrados  $\chi_0^2$  e  $\chi_1^2$ , ambas com peso 0.5. Aqui,  $\chi_p^2$  representa a distribuição qui-quadrado com p graus de liberdade e para p = 0 temos uma distribuição que assume valor 0 com probabilidade 1, ou seja, uma variável aleatória degenerada no valor 0. Considerando o caso em que o interesse é testar se um determinado efeito aleatório deve ser omitido num modelo em que existem dois efeitos aleatórios, temos que testar se a variância do primeiro efeito é nula, bem como a covariância entre os dois efeitos. Assim, para este exemplo, temos que a distribuição assintótica da estatística dada em (2.29) sob a hipótese nula é uma mistura de qui-quadrados  $\chi_1^2 e \chi_2^2$ , ambas com peso 0.5 (Verbeke e Molenberghs (2000)).

### Testes do tipo z

Uma alternativa ao teste de Razão de Verossimilhanças Restritas para testar componentes de variância e covariância em modelos normais multiníveis é o uso de testes do tipo z, conhecidos como testes Wald. Nestes testes, a estatística z é obtida dividindo-se a estimativa do componente de variância e covariância pela estimativa do seu correspondente erro-padrão. O valor p associado ao teste é obtido a partir da distribuição normal padrão. Assim, como o teste RVR, o teste de Wald é também um teste assintótico e apresenta propriedades indesejáveis quando o parâmetro a ser testado encontra-se na fronteira do espaço paramétrico sob a hipótese nula. Testes deste tipo estão detalhados em Raudenbush e Bryk (2002).

A seção seguinte trata de estratégias na construção dos modelos multiníveis. Estas estratégias devem atender a certos critérios que tornem as propostas de modelagem satisfatórias sob determinados aspectos, principalmente no sentido de adequabilidade do modelo adotado para explicar o comportamento de variáveis inerentes a determinado estudo. Estas estratégias são ilustradas na próxima seção.

# 2.7 Estratégias na Seleção dos Modelos Multiníveis

Para a obtenção de um modelo que explique da forma mais adequada possível os dados inerentes a um determinado estudo, são necessários alguns procedimentos como, por exemplo, a escolha da primeira proposta de modelagem. Neste contexto, um dos objetivos iniciais é a escolha do modelo mais simples possível que possa se adequar aos dados. Este procedimento é conhecido na literatura como **princípio da parcimônia**. Existem diversas alternativas para a escolha das estruturas das variáveis preditoras referentes aos

efeitos fixos, bem como para a estrutura de variância e covariância dos erros.

O processo de seleção de modelos multiníveis, assim como de outras classes de modelos, requer uma série de passos e de métodos de investigação a respeito dos aspectos dos ajustes. Sendo assim, não existe uma estratégia que seja aplicável a toda e qualquer modelagem, sendo necessária, assim, a utilização de uma estratégia que seja mais razoável para cada modelagem considerada. Em West, Welch e Galecki (2007) temos a descrição de dois métodos de seleção de modelos, que serão ilustrados a seguir.

### 2.7.1 A Estratégia "Top-Down"

Os passos aqui descritos para a construção de modelos são sugeridos em Verbeke e Molenberghs (2000), no contexto de modelos mistos para dados longitudinais. Contudo, esta estratégia pode ser utilizada em modelos multiníveis. O nome "*Top-Down*" é dado pelo fato de que este procedimento consiste em iniciar com a proposta de um modelo com o máximo de parâmetros dos efeitos fixos a serem considerados nos ajustes. Este método de construção de modelos consiste nos seguintes passos:

- 1. Inicie com uma estrutura bem especificada para os parâmetros de efeitos fixos: este passo envolve tipicamente a adição de tantas variáveis explicativas quanto forem possíveis em todos os níveis do modelo. Este passo tem como objetivo a certificação de que a variação sistemática na variável resposta seja bem explicada pelas variáveis explicativas, antes da proposta de estrutura de variância e covariância dos erros.
- 2. Selecione uma estrutura de erros aleatórios para os parâmetros do modelo: aqui temos a inclusão dos erros aleatórios que explicam os parâmetros de determinados níveis. Para os modelos multiníveis com dois níveis, neste passo são incluídos os erros aleatórios no nível 2, ou seja, os erros associados aos parâmetros do nível 1.
- 3. Selecione uma estrutura de variância e covariância para os erros do nível
  1: uma vez selecionada a estrutura de variáveis explicativas e também a estrutura de erros aleatórios que descrevem os parâmetros de todos os níveis do modelo, a

variação restante da variável dependente é dada pelos erros do nível 1 do modelo. Assim, é necessário propor uma estrutura de variância e covariância para estes erros.

 Reduza o modelo: este passo implica no uso de testes de hipóteses acerca dos parâmetros dos efeitos fixos contidos em β na determinação do conjunto de variáveis significativas do modelo.

# 2.7.2 A Estratégia "Step-Up"

Uma estratégia alternativa na seleção de modelos multiníveis é conhecida como "Step-Up", discutida em Raudenbush e Bryk (2002) e Snijders e Bosker (1999). Os passos desta estratégia são descritos a seguir.

- 1. Inicie com um modelo de médias para o nível 1: este passo envolve o ajuste de um modelo apenas com os interceptos no nível 1. Este modelo também deve incluir os erros referentes aos referidos interceptos no nível 2 e, caso o modelo seja proposto com mais níveis, devem ser especificados os erros aleatórios referentes aos demais níveis. Esta proposta permite acessar a variabilidade da variável dependente por meio apenas da variabilidade existente entre grupos ou "clusters", sem considerar o efeito de variáveis explicativas.
- 2. Adicione covariáveis ao nível 1 do modelo: aqui temos a inclusão de variáveis explicativas no primeiro nível do modelo. A inclusão destas variáveis pode ajudar a explicar a variabilidade da variável dependente, que no passo anterior era explicada apenas pela variabilidade entre grupos. Deve-se adicionar também os respectivos erros associados aos novos parâmetros agora existentes para cada variável incluída neste passo.
- 3. Adicione covariáveis aos níveis superiores:neste passo, variáveis explicativas são adicionadas no nível 2 para modelos com dois níveis, e no caso de modelos com três níveis, variáveis preditoras são adicionadas nos níveis 2 e 3. O intuito deste passo é explicar a variabilidade dos parâmetros dos níveis menores através de variáveis incluídas nos níveis maiores.

### 2.7.3 O Uso de Critérios de Informação na Seleção de Modelos

Uma prática comum nos procedimentos de seleção de modelos é o uso de critérios de informação, uma vez que o procedimento de seleção através de testes de hipóteses acerca dos parâmetros que venham a ser significativos em cada passo do processo não leva em consideração a quantidade de parâmetros especificada em cada modelo. Sendo assim, os critérios de informação são utilizados para a obtenção do modelo a ser aplicado a um determinado estudo. Os critérios mais utilizados neste contexto são o **Critério Informação de Akaike** (AIC) e o **Critério de Informação Bayesiano** (BIC). No contexto de modelos mistos temos maiores detalhes sobre o uso destes critérios em Verbeke e Molenberghs (2000). Estes critérios podem ser definidos para modelos multiníveis com dois níveis como sendo

$$AIC = -2\left\{\hat{l} - \sum_{p=1}^{P} Q_p\right\},\$$
$$BIC = -2\left\{\hat{l} - \left(\sum_{p=1}^{P} Q_p\right)\log\left(\sum_{j=1}^{J} n_j\right)\right\}$$

em que  $\hat{l}$  é o logaritmo da função de verossimilhança avaliado nas estimativas de máxima verossimilhança dos parâmetros,  $\sum_{p=1}^{P} Q_p$  é o número total de parâmetros de efeitos fixos do modelo e  $\sum_{j=1}^{J} n_j$  é o total de observações do nível 1 considerando todos os grupos ou "clusters" do nível 2 do modelo. A definição dos critérios AIC e BIC para modelos multiníveis com três níveis é obtida de forma similar. Contudo, para modelos estimados através do método MVR, temos

$$AIC = -2\left\{\hat{l}_R - \sum_{p=1}^P Q_p\right\},\$$
$$BIC = -2\left\{\hat{l}_R - \left(\sum_{p=1}^P Q_p\right)\log\left[\left(\sum_{j=1}^J n_j\right) - \left(\sum_{p=1}^P Q_p\right)\right]\right\},\$$

em que  $\hat{l}_R$  é o logaritmo da função de verossimilhança restrita avaliado nas estimativas de máxima verossimilhança restrita dos parâmetros. Na seleção de modelos utilizando critérios de informação, dada a escolha de um determinado critério, o modelo selecionado

é aquele que minimiza aquele critério. As estratégias de seleção *"Top-Down"* e *"Step-Up"* descritas anteriormente podem ser executadas incluindo a escolha dos modelos segundo a minimização do AIC, do BIC ou de outro critério de informação.

E válido citar também que, além da utilização de critérios de informação na seleção dos modelos normais multiníveis, um procedimento alternativo ao uso destes critérios é a realização de testes de hipóteses referentes aos parâmetros de efeitos fixos e/ou referentes aos parâmetros de variância e covariância, em cada passo do processo de seleção.

# 2.8 Análise de Resíduos

Após a realização de um ajuste de modelo multinível, é necessário realizar uma análise de diagnóstico, ou seja, executar uma verificação acerca das suposições assumidas no modelo, bem como uma investigação a respeito de algumas características do próprio ajuste. Um exemplo de uma importante característica a ser verificada é a eventual presença de sensibilidade com relação a observações discrepantes. O processo de análise de resíduos, bem como da análise de diagnóstico em geral, compreende vários procedimentos.

Os métodos de diagnóstico já estão bem definidos para modelos normais lineares, bem como para algumas outras classes de modelos, como os modelos lineares generalizados (ver, por exemplo, Paula (2010)). Para modelos mistos com distribuição normal dos erros, alguns trabalhos referentes a métodos de diagnóstico já foram realizados, como por exemplo em Nobre (2003) e Nobre e Singer (2007). Em Hilden-Minton (1995) temos propostas de técnicas de diagnóstico para modelos mistos e multiníveis. Em Raudenbush E Bryk (2002) temos exemplos de alguns procedimentos empíricos para a análise de resíduos nos modelos normais multiníveis. Em Pires (2009) temos o desenvolvimento da análise de influência local através da curvatura normal em modelos multiníveis com distribuição normal dos erros.

Como no caso de modelos multiníveis existe mais de uma fonte de variação, mais de um tipo de resíduo pode ser definido. Tais resíduos podem fornecer evidências necessárias para a verificação das suposições assumidas nesta classe de modelos. Sendo assim, teremos algumas definições de resíduos utilizadas em geral nas análises. Algumas das definições de resíduos aqui ilustradas apresentam estruturas obtidas de forma similar àquelas definidas para a classe de modelos mistos. Estas definições de resíduos, bem como algumas importantes propriedades, podem ser encontradas em Nobre (2003), Nobre e Singer (2007) e Hilden-Minton (1995). Assim, considerando o modelo

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon},$$

podemos definir o resíduo marginal como sendo

$$\hat{\boldsymbol{r}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}},\tag{2.30}$$

em que  $\hat{\beta}$  é o BLUE (ou EBLUE) do vetor de parâmetros  $\beta$ . É válido notar que  $\hat{r}$  é a estimativa do erro marginal

$$r = Y - E(Y) = Y - X\beta = Zb + \epsilon$$
.

Sendo assim, quando a matriz  $\mathbf{X}$  é a própria matriz de variáveis preditoras do modelo multinível, o resíduo aqui considerado pode ser útil, por exemplo, para verificar a suposição de linearidade entre  $\mathbf{E}(\mathbf{Y})$  e as covariáveis presentes em  $\mathbf{X}$ . Como exemplo do uso deste resíduo neste caso, Hilden-Minton (1995) sugere a construção de um gráfico de  $\hat{\mathbf{r}}$  contra os valores das covariáveis. É esperado que os elementos do resíduo marginal variem aleatoriamente em torno de zero, sob a suposição de linearidade entre  $\mathbf{Y}$  e as covariáveis pertencentes à matriz  $\mathbf{X}$ . Um gráfico destes resíduos contra seus respectivos índices pode revelar indícios da presença de observações discrepantes. Além disso, como  $\operatorname{Var}(r) = \operatorname{Var}(\mathbf{Y}) = \mathbf{V}$ , o resíduo dado em (2.30) pode ser útil também na verificação da estrutura de variância e covariância de  $\mathbf{Y}$ . Contudo, como estes resíduos podem apresentar diferentes variâncias, é conveninente utilizar os **resíduos marginais studentizados** 

$$\hat{r}_{ij}^* = \frac{\hat{r}_{ij}}{\widehat{\operatorname{ep}}(\hat{r}_{ij})},\tag{2.31}$$

com  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Aqui,  $\widehat{\operatorname{ep}}(\widehat{r}_{ij})$  é a estimativa do erro padrão do resíduo marginal  $\widehat{r}_{ij}$ . Definindo  $\mathbf{M} = \mathbf{V}^{-1}$ , e considerando que estamos utilizando o EBLUP de  $\boldsymbol{\beta}$  em (2.30), temos que

$$\hat{\boldsymbol{r}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^t\hat{\mathbf{M}}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = \hat{\mathbf{M}}^{-1}\hat{\mathbf{Q}}\mathbf{Y},$$

em que  $\mathbf{Q} = \mathbf{M} - \mathbf{M}\mathbf{X}(\mathbf{X}^{t}\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^{t}\mathbf{M}$ . A matriz  $\mathbf{Q}$  é simétrica, com  $\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q} = \mathbf{Q}$  e  $\mathbf{Q}\mathbf{X} = \mathbf{0}$ . Assim, temos a matriz de variância e covariância estimada do vetor de erros marginais dada por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{r}}) = \widehat{\operatorname{Var}}(\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{Q}}\boldsymbol{Y}) = \hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{Q}}\widehat{\operatorname{Var}}(\boldsymbol{Y})\hat{\boldsymbol{Q}}\hat{\boldsymbol{M}}^{-1} = \\ \hat{\boldsymbol{M}}^{-1}(\hat{\boldsymbol{Q}}\hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{Q}})\hat{\boldsymbol{M}}^{-1} = \hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{Q}}\hat{\boldsymbol{M}}^{-1}.$$
(2.32)

Logo, os elementos da diagonal principal de (2.32) podem ser utilizados na obtenção dos valores dados em (2.31).

Outro importante resíduo a ser considerado é o chamado **resíduo condicional**, definido por

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{b}},\tag{2.33}$$

em que  $\hat{\boldsymbol{\beta}}$  é o BLUE (ou EBLUE) do vetor  $\boldsymbol{\beta}$  e  $\hat{\boldsymbol{b}}$  é o BLUP (ou EBLUP) de  $\boldsymbol{b}$ . Com relação ao uso deste tipo de resíduo temos algumas propostas já sugeridas para modelos mistos. Em Pinheiro e Bates (2000) temos a sugestão do uso dos resíduos condicionais através da construção do gráfico de  $\hat{\boldsymbol{\epsilon}}$  contra os valores ajustados  $\hat{\mathbf{Y}}$  e gráficos dos percentis deste mesmo resíduo para verificar a suposição de normalidade do erro condicional

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{E}(\mathbf{Y}|\boldsymbol{b}) = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b}.$$

O resíduo  $\hat{\epsilon}$  também pode ser utilizado para identificar observações discrepantes presentes nos dados, através do gráfico destes resíduos contra seus respectivos índices. Similarmente ao caso dos resíduos marginais, os resíduos condicionais podem apresentar diferentes variâncias. Sendo assim, convém utilizar os **resíduos condicionais studentizados** 

$$\hat{\epsilon}_{ij}^* = \frac{\hat{\epsilon}_{ij}}{\widehat{\operatorname{ep}}(\hat{\epsilon}_{ij})},\tag{2.34}$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, em que  $\widehat{ep}(\hat{\epsilon}_{ij})$  é a estimativa do erro padrão do resíduo condicional  $\hat{\epsilon}_{ij}$ . Definindo

$$\mathbf{C} = \mathbf{\Psi}^{-1} + \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{Z},$$

temos que

$$\mathbf{M} = (\mathbf{R} + \mathbf{Z} \boldsymbol{\Psi} \mathbf{Z}^t)^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1} = \mathbf{R}^{-1} (\mathbf{I} - \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1}),$$

onde **I** é a matriz identidade de dimensão  $(\sum_{j=1}^{J} n_j) \times (\sum_{j=1}^{J} n_j)$ . Por outro lado,

$$\begin{split} \Psi \mathbf{Z}^t \mathbf{M} &= \Psi \mathbf{Z}^t (\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1}) = \Psi (\mathbf{C} - \mathbf{Z}^t \mathbf{R}^{-1} \mathbf{Z}) \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1} = \\ \Psi \Psi^{-1} \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1} &= \mathbf{C}^{-1} \mathbf{Z}^t \mathbf{R}^{-1}. \end{split}$$

Portanto, assumindo que em (2.33) estamos utilizando o EBLUE de  $\boldsymbol{\beta}$  e o EBLUP de  $\boldsymbol{b}$ , temos que

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{b}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\Psi}\mathbf{Z}^{t}\hat{\mathbf{M}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) =$$
  
$$\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{C}}^{-1}\mathbf{Z}^{t}\hat{\mathbf{R}}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{I} - \mathbf{Z}\hat{\mathbf{C}}^{-1}\mathbf{Z}^{t}\hat{\mathbf{R}}^{-1})(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) =$$
  
$$\hat{\mathbf{R}}\hat{\mathbf{M}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Como  $\boldsymbol{r} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{Y}$ , temos que

$$\hat{\boldsymbol{\epsilon}} = \hat{\mathbf{R}}\hat{\mathbf{M}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \hat{\mathbf{R}}\hat{\mathbf{M}}\hat{\mathbf{M}}^{-1}\hat{\mathbf{Q}}\mathbf{Y} = \hat{\mathbf{R}}\hat{\mathbf{Q}}\mathbf{Y}.$$

Assim, a matriz de variância e covariância estimada de  $\hat{\boldsymbol{\epsilon}}$  é dada por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\epsilon}}) = \widehat{\operatorname{Var}}(\hat{\mathbf{R}}\hat{\mathbf{Q}}\mathbf{Y}) = \hat{\mathbf{R}}\hat{\mathbf{Q}}\widehat{\operatorname{Var}}(\mathbf{Y})\hat{\mathbf{Q}}\hat{\mathbf{R}} = \hat{\mathbf{R}}(\hat{\mathbf{Q}}\hat{\mathbf{M}}^{-1}\hat{\mathbf{Q}})\hat{\mathbf{R}} = \hat{\mathbf{R}}\hat{\mathbf{Q}}\hat{\mathbf{R}}.$$
(2.35)

Desta forma, os termos da diagonal principal de (2.35) são utilizados na obtenção dos resíduos condicionais studentizados dados em (2.34).

Para a utilização dos resíduos condicionais na verificação da suposição de normalidade do vetor de erros  $\epsilon$ , Hilden-Minton (1995), Nobre (2003) e Nobre e Singer (2007) sugerem o uso dos **resíduos com confundimento mínimo**. A ideia é reduzir o confundimento do resíduo condicional com o vetor de efeitos aleatórios **b**. A partir dos referidos autores temos que

$$\hat{oldsymbol{\epsilon}} = \mathbf{R}\mathbf{Q}oldsymbol{\epsilon} + \mathbf{R}\mathbf{Q}\mathbf{Z}oldsymbol{b};$$
  
 $\hat{oldsymbol{r}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{M}]oldsymbol{r} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^t\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{M}](\mathbf{Z}oldsymbol{b} + oldsymbol{\epsilon}).$ 

De acordo com Hilden-Minton (1995), a capacidade de avaliação da normalidade de  $\boldsymbol{\epsilon}$  a partir de  $\hat{\boldsymbol{\epsilon}}$  diminui quando Var(**RQZb**) cresce com relação a Var(**RQ** $\boldsymbol{\epsilon}$ ). Considerando os vetores  $\mathbf{U}_l$ ,  $l = 1, 2, \ldots, \sum_{j=1}^J n_j$ , em que  $\mathbf{U}_l$  representa a *l*-ésima coluna da matriz identidade **I** de dimensão  $(\sum_{j=1}^J n_j) \times (\sum_{j=1}^J n_j)$ , o autor define a razão de confundimento do resíduo  $\hat{\epsilon}_l = \mathbf{U}_l^t \hat{\boldsymbol{\epsilon}}$  como sendo

$$RC_{\hat{\epsilon}_l} = \frac{\mathbf{U}_l^t \mathbf{R} \mathbf{Q} \mathbf{Z} \mathbf{\Psi} \mathbf{Z}^t \mathbf{Q} \mathbf{R} \mathbf{U}_l}{\mathbf{U}_l^t \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{U}_l} = 1 - \frac{\mathbf{U}_l^t \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{U}_l}{\mathbf{U}_l^t \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{U}_l}.$$

Temos que  $RC_{\hat{\epsilon}_l}$  é um valor entre 0 e 1. Quando  $RC_{\hat{\epsilon}_l} = 0$ , o resíduo  $\hat{\epsilon}_l$  não apresenta confudimento. Neste caso, é dito que o resíduo é **puro**. A ideia é minimizar a fração de confundimento, ou seja, para um vetor  $\boldsymbol{a}_l$  representando a *l*-ésima coluna de uma matriz  $\mathbf{A}$ , o objetivo é encontrar  $\mathbf{A}$  tal que  $\mathbf{A}\hat{\boldsymbol{\epsilon}}$  apresente confundimento mínimo. Isto é equivalente a maximizar cada uma das quantidades

$$\lambda_l = \frac{\boldsymbol{a}_l^t \mathbf{R} \mathbf{Q} \mathbf{R} \mathbf{Q} \mathbf{R} \boldsymbol{a}_l}{\boldsymbol{a}_l^t \mathbf{R} \mathbf{Q} \mathbf{R} \boldsymbol{a}_l},\tag{2.36}$$

para  $l = 1, 2, ..., \sum_{j=1}^{J} n_j$ . Este é um problema de autovalores e autovetores envolvendo matrizes semipositivas definidas. Nosso interesse restringe-se às condições de que  $\operatorname{Var}(\boldsymbol{a}_l^t \hat{\boldsymbol{\epsilon}}) > 0$ , para  $l = 1, 2, ..., \sum_{j=1}^{J} n_j$ . Então, a maximização da quantidade (2.36) deve ser restrita às condições

$$\boldsymbol{a}_l^{t} \mathbf{R} \mathbf{Q} \mathbf{R} \boldsymbol{a}_l > 0,$$

para  $l = 1, 2, ..., \sum_{j=1}^{J} n_j$ . Desde que a matriz **R** é geralmente de posto completo (sendo também diagonal), o interesse é o espaço não nulo gerado pelas colunas da matriz **Q**. Assumindo que o posto da matriz **Q** é dado por  $(\sum_{j=1}^{J} n_j) - \sum_{p=1}^{P} Q_p$ , em que  $\sum_{p=1}^{P} Q_p$  é o número de parâmetros dos efeitos fixos do modelo, são obtidas uma matriz **B** de dimensão  $(\sum_{j=1}^{J} n_j) \times [(\sum_{j=1}^{J} n_j) - \sum_{p=1}^{P} Q_p]$  e uma matriz diagonal  $\Lambda$  de dimensão  $[(\sum_{j=1}^{J} n_j) - \sum_{p=1}^{P} Q_p] \times [(\sum_{j=1}^{J} n_j) - \sum_{p=1}^{P} Q_p]$  tais que

$$\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{R}^{\frac{1}{2}}=\mathbf{B}\mathbf{\Lambda}\mathbf{B}^{t},$$

em que  $\mathbf{B}^t \mathbf{B} = \mathbf{I}$ , sendo  $\mathbf{I}$  a matriz identidade de dimensão  $(\sum_{p=1}^P Q_p) \times (\sum_{p=1}^P Q_p)$ . Fazendo

$$oldsymbol{a}_l = \mathbf{R}^{-rac{1}{2}} \mathbf{B} oldsymbol{\Lambda}^{-rac{1}{2}} oldsymbol{v}_l$$

para vetores quaisquer  $\boldsymbol{v}_l, l = 1, 2, \dots, (\sum_{j=1}^J n_j) - \sum_{p=1}^P Q_p$ , obtemos

$$\lambda_l = rac{oldsymbol{v}_l^t oldsymbol{\Lambda} oldsymbol{a}_l}{oldsymbol{a}_l^t oldsymbol{a}_l}.$$

Este problema é resolvido, então, tomando  $\boldsymbol{v}_l = \mathbf{U}_l, \ l = 1, 2, \dots, (\sum_{j=1}^J n_j) - \sum_{p=1}^P Q_p.$ Desta forma, ordenando os termos em  $\boldsymbol{\Lambda}$  da forma  $1 \ge \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{(\sum_{j=1}^J n_j) - \sum_{p=1}^P Q_p} \ge 0$ , obtemos que

$$oldsymbol{a}_l = \mathbf{R}^{-rac{1}{2}} \mathbf{B} oldsymbol{\Lambda}^{-rac{1}{2}} \mathbf{U}_l = rac{1}{\sqrt{\lambda_l}} \mathbf{R}^{-rac{1}{2}} \mathbf{B} \mathbf{U}_l = rac{1}{\sqrt{\lambda_l}} \mathbf{R}^{-rac{1}{2}} \mathbf{B}_l,$$

em que  $\mathbf{B}_l$  é a *l*-ésima coluna da matriz  $\mathbf{B}$ , para  $l = 1, 2, \ldots, (\sum_{j=1}^J n_j) - \sum_{p=1}^P Q_p$ . Hilden-Minton (1995) mostra que o resíduo  $\mathbf{a}_l^t \hat{\boldsymbol{\epsilon}}$  apresenta fração de confundimento  $1 - \lambda_l$ , para  $l = 1, 2, \ldots, (\sum_{j=1}^J n_j) - \sum_{p=1}^P Q_p$ , sendo estes resíduos não correlacionados entre si. Estes resíduos são então os resíduos com confundimento mínimo, utilizados para verificação da suposição de normalidade do vetor de erros  $\boldsymbol{\epsilon}$ .

Os BLUP's ou EBLUP's também podem ser utilizados nas análise de diagnóstico. As matrizes de variância e covariância estimadas dos BLUP's são dadas por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{b}}_j) = \hat{\boldsymbol{\Psi}}_j \mathbf{Z}_j^{\ t} \left[ \hat{\mathbf{V}}_j^{-1} - \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}_j^{\ t} \hat{\mathbf{V}}_j^{-1} \right] \mathbf{Z}_j \hat{\boldsymbol{\Psi}}_j,$$
(2.37)

para j = 1, 2, ..., J. Desta forma, os elementos da diagonal principal de (2.37) podem ser

utilizados na obtenção dos preditores lineares studentizados

$$\hat{b}_{ij}^* = \frac{\hat{b}_{ij}}{\widehat{\exp}(\hat{b}_{ij})},$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, onde  $\widehat{\operatorname{ep}}(\widehat{b}_{ij})$  é a estimativa do erro padrão de  $\widehat{b}_{ij}$ . Contudo, como citado em Laird e Ware (1982), a variância dada em (2.37) não considera a variação existente em  $b_j$ . Assim, é mais recomendado utilizar

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j) = \hat{\boldsymbol{\Psi}}_j - \hat{\boldsymbol{\Psi}}_j \mathbf{Z}_j^{\ t} \left[ \hat{\mathbf{V}}_j^{-1} - \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}_j^{\ t} \hat{\mathbf{V}}_j^{-1} \right] \mathbf{Z}_j \hat{\boldsymbol{\Psi}}_j,$$
(2.38)

para j = 1, 2, ..., J. As equações (2.37) e (2.38) são casos particulares das fórmulas mais gerais dadas em Harville (1976). Um gráfico dos preditores lineares studentizados contra seus respectivos índices pode dar indícios da existência de efeitos aleatórios desproporcionais em relação aos demais num determinado grupo.

No intuito de detectar grupos ou "clusters" atípicos em relação aos demais, existe a proposta do uso das **distâncias de Mahalanobis** dadas por

$$u_j = (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta}),$$

para j = 1, 2, ..., J. Sob o modelo dado em (2.12), temos que  $u_j$  segue uma distribuição qui-quadrado com  $n_j$  graus de liberdade (ver Little (1988), Lange, Little e Taylor (1989) e Copt e Victoria-Feser (2006)). Sendo assim, a construção de um gráfico dos valores estimados dos  $u_j$ 's contra seus respectivos índices pode servir para detectar grupos discrepantes.

Waternaux, Laird e Ware (1989) propõem utilizar as distâncias de Mahalanobis dos BLUP's

$$\zeta_j = \hat{\boldsymbol{b}}_j^t \widehat{\operatorname{Var}}(\hat{\boldsymbol{b}}_j - \boldsymbol{b}_j) \hat{\boldsymbol{b}}_j,$$

para j = 1, 2, ..., J na identificação de grupos ou "clusters" discrepantes, no que diz respeito aos efeitos aleatórios. Sob a hipótese de que o modelo está bem especificado, temos que  $\zeta_j$  segue aproximadamente uma distribuição qui-quadrado com P graus de liberdade, para  $\sum_{j=1}^{J} n_j$  suficientemente grande e  $n_j$  fixado, para j = 1, 2, ..., J. Para a verificação da suposição de normalidade dos efeitos aleatórios em **b**, pode ser construído um gráfico de percentis empíricos a partir dos BLUP's ou EBLUP's studentizados, utilizando os percentis teóricos da distribuição normal, no intuito de observar possíveis afastamentos de **b** com relação à distribuição citada.

Os resíduos studentizados marginais e condicionais, bem como os resíduos com confundimento mínimo e os BLUP's ou EBLUP's studentizados podem, ainda, ser utilizados para a construção de um tipo de banda de confiança através de simulações, denominada de envelope (ver Atkinson (1985)), no intuito de verificar se a distribuição dos erros (bem como dos efeitos aleatórios) está bem especificada, mesmo que para os casos dos resíduos marginais e condicionais possa existir confundimento. Uma proposta é a análise dos resíduos levando em consideração também o comportamento dos preditores lineares dados em (2.24). A construção da banda de confiança simulada implica na simulação de variáveis aleatórias com distribuição obtida a partir da distribuição da variável resposta utilizada no ajuste do modelo, tomando como valores dos parâmetros as estimativas obtidas no ajuste original. Assim são obtidos novos resíduos a partir do ajuste destas variáveis simuladas, considerando o mesmo grupo de variáveis explicativas do modelo original. Os passos para a obtenção da banda de confiança simulada com aproximadamente 95% de confiança para os resíduos marginais studentizados em um modelo normal multinível com dois níveis são os seguintes:

Passo 1: é gerado um vetor **y** a partir da distribuição  $\mathbf{Y} \sim N_{\sum_{j=1}^{J} n_j}(\mathbf{X}\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}});$ 

- Passo 2: o modelo é ajustado com a variável resposta gerada **y** e as mesmas variáveis explicativas do modelo original;
- Passo 3: são obtidos os resíduos marginais padronizados  $\hat{r}_{ij}^*$ ,  $i = 1, 2, ..., n_j$ , j = 1, 2, ..., J;
- Passo 4: os passos 1, 2 e 3 são repetidos L vezes, obtendo assim os resíduos marginais studentizados  $\hat{r}^*_{ijl}$ , para  $i = 1, 2, ..., n_j$ , j = 1, 2, ..., J e l = 1, 2, ..., L;
- Passo 5: cada um dos L vetores de resíduos é ordenado em ordem crescente, obtendo os resíduos ordenados  $\hat{r}^*_{(k)l}$ , para  $k = 1, 2, \ldots, \sum_{j=1}^J n_j$  e  $l = 1, 2, \ldots, L$ ;

Passo 6: são obtidos os percentis empíricos de ordem 0.025 e 0.975 a partir de cada vetor de resíduos  $\hat{\boldsymbol{r}}_{(k)}^* = (\hat{r}_{(k)1}^*, \hat{r}_{(k)2}^*, \dots, \hat{r}_{(k)L}^*)^t$ , para  $k = 1, 2, \dots, \sum_{j=1}^J n_j$ , denotados por  $\hat{r}_{(k)0.025}^*$  e  $\hat{r}_{(k)0.975}^*$ .

Desta forma, os limites correspondentes ao resíduo de ordem (k) são dados por  $\hat{r}^*_{(k)0.025}$  e  $\hat{r}^*_{(k)0.975}$ . Para a construção de bandas de confiança para os resíduos condicionais, resíduos com confundimento mínimo ou preditores lineares, os passos de execução necessários são similares.

# 2.9 Aplicações

Neste capítulo são desenvolvidos dois exemplos de aplicação dos modelos multiníveis. O primeiro exemplo considera uma análise envolvendo a aplicação de modelos multiníveis com dois níveis e o segundo exemplo trata de modelos multiníveis com três níveis. No primeiro caso, as unidades em análise (nível 1) estão agrupadas em unidades (ou "clusters") selecionadas aleatoriamente (nível 2). No segundo caso, temos que as unidades em análise (nível 1) estão agrupadas em unidades selecionadas aleatoriamente (nível 2) que, por sua vez, estão agrupadas em unidades também selecionadas aleatoriamente (unidades que formam o nível 3).

### 2.9.1 Aplicação 1 - Experimento com Ratos

## Descrição do Estudo

Os dados utilizados neste exemplo de aplicação de modelos multiníveis podem ser encontrados em Pinheiro e Bates (2000). O banco de dados remete a um estudo em que 30 ratos fêmeas foram aleatoriamente designados a receber uma das três dosagens (baixa, alta e controle) de um composto experimental. O objetivo do estudo foi comparar os pesos dos ratos recém nascidos nas respectivas ninhadas que receberam as dosagens alta e baixa com os pesos dos ratos recém nascidos cujas ninhadas receberam o tratamento controle. Enquanto 10 ratos fêmeas foram aleatorizados a receber um dos três tratamentos, três fêmeas que receberam a dosagem alta morreram, assim para estes casos as ninhadas não apresentaram dados. Além disso, o número de ratos nascidos em cada ninhada apresentaram grande variação, sendo o mínimo de 2 e o máximo de 18 ratos. Este é um estudo *desbalanceado*, pois tanto o número de ninhadas por tratamento quanto o número de ratos recém nascidos por ninhada não é constante para os três tratamentos considerados.

Estes dados apresentam dois níveis de hierarquia, pois cada rato fêmea foi aleatorizado a receber um dos três tratamentos, sendo os ratos nascidos dessas fêmeas pertecendo a somente uma ninhada remetente a determinada fêmea ("cluster"). Dizemos então que neste caso, os ratos filhotes (que constituem as unidades do nível 1) estão "aninhados" nas unidades do nível 2 (ninhadas ou fêmeas consideradas no estudo). Pesos de ratos recém nascidos provenientes de uma mesma ninhada podem apresentar correlação devido ao parentesco destes com relação à fêmea que gerou a ninhada. Sendo assim, pode ser razoável considerar um efeito aleatório para cada ninhada, assim como efeitos fixos para analisar as variações entre tratamentos. Considerando as 27 ninhadas temos um total de 322 ratos filhotes considerados neste exemplo.

## Análises Iniciais

Na Tabela 2.1 temos uma amostra do banco de dados utilizado na modelagem multinível dos dados referente ao experimento com ratos.

	Nível 1		Nível 2				
Ident.	Var. Dependente	Covariável	Ident.	Covariável	Covariável		
PUP_ID	WEIGHT	SEX	LITTER	TREATMENT	LITSIZE		
1	6,60	Male	1	Control	12		
2	7,40	Male	1	Control	12		
3	7, 15	Male	1	Control	12		
4	7,24	Male	1	Control	12		
9	6,95	Female	1	Control	12		
10	6,29	Female	1	Control	12		
132	5,65	Male	11	Low	16		
133	5,78	Male	11	Low	16		
258	5,09	Male	21	High	14		
259	5,57	Male	21	High	14		
260	5,69	Male	21	High	14		

Tabela 2.1: Amostra do banco de dados referente ao experimento com ratos.

Em seguida, através da Tabela 2.2 podemos obter as descrições das variáveis presentes

no banco de dados utilizado neste exemplo de aplicação.

Tabela	2.2:	Descrições	das	variáveis	consideradas	no	experimento	com	ratos.
		Ţ	Vari	áveis do	nível 1 (Ra	tos	3)		

PUP ID:	número de identificação do rato recém nascido;				
WEIGHT:	<b>WEIGHT</b> : peso (em gramas) do rato recém nascido (variável dependente);				
SEX:	sexo do rato recém nascido.				
Va	riáveis do nível 2 (Ninhadas)				
LITTER: número de identificação da ninhada;					
TREATMENT:	dosagem do composto experimental aleatorizado para cada ninhada (controle, baixa e alta);				
LITSIZE:	tamanho da ninhada (isto é, número de ratos recém nascidos observados na ninhada);				

A Tabela 2.3 contém as medidas descritivas do peso dos ratos filhotes por tratamento e sexo. Tanto para indivíduos do sexo masculino quanto para indivíduos do sexo feminino temos o peso médio maior para o grupo controle em relação às dosagens alta e baixa. O maior peso registrado foi de 8,33g, para um filhote do sexo masculino pertencente a uma ninhada que recebeu a dosagem controle do composto. O menor valor dos pesos registrados foi de 3,68g, para um filhote do sexo feminino pertencente a uma ninhada que também recebeu a dosagem controle do tratamento.

Através da Figura 2.1 pode-se perceber que os valores medianos dos pesos dos ratos recém nascidos são maiores para os machos que para as fêmeas nas ninhadas que receberam a dosagem controle e baixa, porém para ninhadas que receberam dosagem alta do complexo experimental, o valor mediano dos pesos dos machos é ligeiramente menor que esta mesma medida para os ratos filhotes do sexo feminino. Existem pontos discrepantes para todas as dosagens consideradas, assim como para os sexos masculino e feminino considerando as três dosagens do complexo.

Considerando o tamanho das ninhadas (ou seja, o número de ratos nascidos a partir de

Sexo Masculino								
Medidas								
Dosagem	Mínimo	1° quartil	Mediana	Média	$3^{\circ}$ quartil	Máximo	Desvio	
Controle	4,5700	6,0600	6,4100	6,4710	6,9400	8,3300	0,7540	
Baixa	5,2500	5,7500	6,0000	6,0250	6,2400	7,1300	0,3800	
Alta	5,0100	5,3600	5,6900	5,9180	6,2900	7,7000	0,6910	
Sexo Feminino								
Medidas								
Dosagem	Mínimo	1° quartil	Mediana	Média	3° quartil	Máximo	Desvio	
Controle	3,6800	5,8220	6,1750	6,1160	6,4480	7,5700	0,6860	
Baixa	4,7500	5,6000	5,8400	5,8380	6,1100	7,7300	0,4500	
Alta	4,4800	5,4150	5,7600	5,8520	6,2400	7,6800	0,6000	

Tabela 2.3: Medidas descritivas dos p	pesos dos ratos por sexo e tratamento.
---------------------------------------	--



Figura 2.1: Gráficos de box-plot dos pesos dos ratos segundo o sexo para cada dosagem. cada uma das 27 fêmeas), temos a partir da Figura 2.2 que o valor mediano dos pesos dos filhotes decresce à medida em que cresce o número de filhotes na ninhada, com exceção dos filhotes do sexo masculino provenientes de ninhadas pertencentes ao grupo que receberam baixa dosagem do composto experimental.

## Modelagem

Para o uso de modelos normais multiníveis para esta aplicação, foi utilizada a estratégia "*Top-Down*" de seleção de modelos, ou seja, aquela que considera inicialmente o número máximo de efeitos fixos a serem considerados nos ajustes, descrita na Seção 2.7.1. A sequência de passos seguidos nesta seleção estão descritos em West, Welch e Galecki (2007). Sendo assim, temos o modelo selecionado dado por

Equação do Nível 1:  
WEIGHT<sub>ij</sub> = 
$$\gamma_{1j} + \gamma_{2j}$$
SEX1<sub>ij</sub> +  $\epsilon_{ij}$ ;  
Equações do Nível 2:  
 $\gamma_{1j} = \beta_{11} + \beta_{12}$ TREAT1<sub>j</sub> +  $\beta_{13}$ TREAT2<sub>j</sub> +  $\beta_{14}$ LITSIZE<sub>j</sub> +  $b_{1j}$ ;  
 $\gamma_{2j} = \beta_{21} + b_{2j}$ ,  
(2.39)

com  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., 27, em que  $n_j$  é o número de filhotes pertencentes à *j*-ésima ninhada, num total de J = 27 ninhadas. As variáveis SEX1, TREAT1 e TREAT2 são obtidas a partir das variáveis SEX e TREATMENT, fazendo

$$SEX1_{ij} = \begin{cases} 1, & \text{se o } i\text{-}\text{\acute{e}simo filhote da } j\text{-}\text{\acute{e}sima ninhada \acute{e} fêmea}, \\ 0, & \text{se o } i\text{-}\text{\acute{e}simo filhote da } j\text{-}\text{\acute{e}sima ninhada \acute{e} macho}, \end{cases}$$

 $\text{TREAT1}_{j} = \begin{cases} 1, & \text{se a } j\text{-}\text{\'esima ninhada recebeu dosagem$ **alta** $do tratamento, \\ 0, & \text{caso contrário,} \end{cases}$ 

$$\text{TREAT2}_{j} = \begin{cases} 1, & \text{se a } j\text{-}\acute{\text{esima ninhada recebeu dosagem baixa do tratamento,} \\ 0, & \text{caso contrário.} \end{cases}$$

A quantidade LITSIZE<sub>j</sub> é o tamanho (número de filhotes) da *j*-ésima ninhada. Na primeira equação, referente ao nível 1 do modelo, temos WEIGHT<sub>ij</sub> representando o peso do *i*ésimo filhote pertencente à *j*-ésima ninhada,  $\gamma_{1j}$  o intercepto para a *j*-ésima ninhada,  $\gamma_{2j}$  o parâmetro associado ao sexo dos filhotes pertencentes à *j*-ésima ninhada e um erro aleatório  $\epsilon_{ij}$  associdado ao referido peso. É assumido que

$$\epsilon_{ij} \sim \begin{cases} N(0, \sigma_C^2), & \text{se a } j\text{-}\acute{esima ninhada recebeu a dosagem controle do tratamento;}\\ N(0, \sigma_{HL}^2), & \text{se a } j\text{-}\acute{esima ninhada recebeu a dosagem alta ou baixa do tratamento,} \end{cases}$$

com os  $\epsilon_{ij}$ 's independentes entre si. Esta suposição foi adotada de acordo com o modelo selecionado em West, Welch e Galecki (2007). Em um dos passos da seleção do modelo, estes autores testaram a hipótese de que as variâncias dos pesos dos filhotes cujas ninhadas receberam dosagens alta ou baixa do tratamento são iguais, sendo ainda o valor desta variância estatisticamente diferente daquela considerada para ninhadas que receberam o tratamento controle. Na segunda equação, referente ao nível 2 do modelo, temos os parâmetros  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13} \in \beta_{14}$ , sendo o primeiro a média geral dos interceptos  $\gamma_{1j}$  para ninhadas que não contém filhotes (LITSIZE $_j = 0$ ) e que receberam o tratamento controle, o segundo representando o efeito no intercepto  $\gamma_{1j}$  se a j-ésima ninhada recebeu a dosagem alta do composto, o terceiro represe<br/>tando o efeito em  $\gamma_{1j}$  se a j-ésima ninhada foi submetida à dosagem baixa do tratamento e o quarto sendo o parâmetro associado ao tamanho da ninhada. Temos ainda, nesta equação, um efeito aleatório relacionado a  $\gamma_{1j}$ , denotado por  $b_{1j}$ , para  $j = 1, 2, \ldots, 27$ . Assume-se que  $b_{1j} \sim N(0, \sigma_N^2)$ , sendo os  $b_{1j}$ 's independence ente si e independentes dos erros  $\epsilon_{ij}$ , para todo  $i = 1, 2, \ldots, n_j$  e  $j = 1, 2, \ldots, 27$ . Na terceira equação, também referente ao nível 2, assume-se  $b_{2j} = 0$  (ou seja, não há efeito aleatório para o parâmetro  $\gamma_{2j}$ ). Sendo assim, o parâmetro  $\beta_{21}$  indica que o efeito da variável SEX1 no peso médio dos ratos é o mesmo para todas as ninhadas.

O modelo (2.39) pode, ainda, ser expresso na forma

WEIGHT<sub>*ij*</sub> = 
$$\beta_{11} + \beta_{12}$$
TREAT1<sub>*j*</sub> +  $\beta_{13}$ TREAT2<sub>*j*</sub> +  $\beta_{14}$ LITSIZE<sub>*j*</sub> +  
 $\beta_{21}$ SEX1<sub>*ij*</sub> +  $b_{1j} + \epsilon_{ij}$ ; (2.40)

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., 27. Segundo o modelo como dado em (2.40) podemos dizer que  $\beta_{11}$  é o intercepto do modelo,  $\beta_{12}$  é o efeito fixo no valor esperado dos pesos dos ratos para ninhadas que receberam a dosagem alta do tratamento,  $\beta_{13}$  é o efeito fixo no peso esperado dos ratos para ninhadas que receberam a dosagem baixa do tratamento,  $\beta_{14}$ é o efeito fixo do tamanhos das ninhadas nos pesos esperados dos ratos e  $\beta_{21}$  é o efeito fixo nos pesos esperados dos ratos fêmeas. Temos, ainda, que os  $b_{1j}$ 's são os efeitos aleatórios das ninhadas nos pesos dos ratos e os  $\epsilon_{ij}$ 's são os erros aleatórios associados aos valores dos respectivos pesos dos filhotes considerados no experimento.

A notação matricial do modelo (2.39) para a *j*-ésima ninhada é, portanto, dada por

$$\mathbf{Y}_{j} = (\text{WEIGHT}_{1j}, \text{WEIGHT}_{2j}, \dots, \text{WEIGHT}_{n_{j}j})^{t}_{n_{j} \times 1};$$

$$\mathbf{Z}_{j} = \begin{bmatrix} 1 & \mathrm{SEX1}_{1j} \\ 1 & \mathrm{SEX1}_{2j} \\ \vdots & \vdots \\ 1 & \mathrm{SEX1}_{njj} \end{bmatrix}_{n_{j} \times 2}, \mathbf{T}_{j} = \begin{bmatrix} \mathrm{T}_{1j} & 0 \\ \mathbf{0} & \mathrm{T}_{2j} \end{bmatrix}_{2 \times 5}, \mathbf{X}_{j} = \mathbf{Z}_{j}\mathbf{T}_{j} = \begin{bmatrix} \mathrm{T}_{1j} & \mathrm{SEX1}_{1j} \\ \mathrm{T}_{1j} & \mathrm{SEX1}_{2j} \\ \vdots & \vdots \\ \mathrm{T}_{1j} & \mathrm{SEX1}_{njj} \end{bmatrix}_{n_{j} \times 5},$$

em que  $T_{1j} = (1, \text{TREAT1}_j, \text{TREAT2}_j, \text{LITSIZE}_j)_{1 \times 4}, \ T_{2j} = 1, \ \boldsymbol{\gamma}_j = (\gamma_{1j}, \gamma_{2j})^t_{2 \times 1}, \ \boldsymbol{\epsilon}_j = (\epsilon_{1j}, \epsilon_{2j}, \dots, \epsilon_{n_j j})^t_{n_j \times 1}, \ \boldsymbol{\beta} = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{21})^t_{5 \times 1} \in \boldsymbol{b}_j = (b_{1j}, 0)^t_{2 \times 1}.$ 

As estimativas dos parâmetros obtidas pelos métodos de máxima verossimilhança e máxima verossimilhança restrita são apresentados na Tabela 2.4. Para um nível de significância de, por exemplo, 5%, todos os efeitos fixos são marginalmente significativos pelos testes do tipo t, através dos dois métodos de estimação. As estimativas dos parâmetros dos efeitos fixos obtidas por MV e MVR apresentaram valores próximos para cada parâmetro. Com relação às estimativas dos erros padrão das estimativas, os valores obtidos via MVR são maiores em relação àqueles obtidos via MV, com exceção dos valores obtidos para a estimativa do parâmetro  $\beta_{21}$ . Neste caso, o valor estimado do erro padrão da estimativa deste parâmetro via MV é maior (0,0422) que esta mesma estimativa obtida via MVR (0,0420). Considerando cada parâmetro de variância e covariância, temos que as estimativas de máxima verossimilhança são menores do que estas mesmas estimativas obtidas por MVR.

Interpretando as estimativas obtidas por máxima verossimilhança, temos que a estimativa do intercepto é de 8,3280g. Fixados o tamanho da ninhada e o sexo dos ratos filhotes temos que, com relação aos ratos que receberam a dosagem controle do tratamento, o peso dos ratos que receberam a dosagem alta é, em média, 0,8648g menor, enquanto que o peso dos ratos que receberam a dosagem baixa do tratamento é, em média, 0,4344g menor em relação aos ratos submetidos a dosagem controle. Considerando o tamanho da ninhada, temos que o peso dos ratos diminui, em média, 0,1307g com o aumento de uma unidade no número de ratos recém nascidos na ninhada, fixados o sexo dos ratos filhotes e o tipo de tratamento aplicado à ninhada. Com relação ao peso dos machos, o peso de ratos recém nascidos fêmeas é, em média, 0,3419g menor, fixados o tipo de tratamento e o tamanho da ninhada. A interpretação das estimativas dos parâmetros obtidas via MVR é realizada de forma similar a esta aqui detalhada.

		MV			
Parâmetro	Estimativa	Erro padrão	G.l.	Valor t	Valor p
$\beta_{11}$	8,3280	0,2593	294	32,1197	< 0,001
$\beta_{12}$	-0,8648	0,1715	23	-5,0429	< 0,001
$\beta_{13}$	-0,4344	0,1424	23	-3,0500	0,0057
$\beta_{14}$	-0,1307	0,0175	23	-7,4646	< 0,001
$\beta_{21}$	-0,3419	0,0422	294	-8,0949	< 0,001
$\sigma_N^2$	0,0832	0,0265	_	_	_
$\sigma_C^2$	0,2639	0,0339	_	_	_
$\sigma^2_{HL}$	0,0913	0,0098	—	—	_
		MVR			
Parâmetro	Estimativa	Erro padrão	G.l.	Valor t	Valor p
$\beta_{11}$	8,3276	0,2741	294	30,3851	< 0,001
$\beta_{12}$	-0,8623	0,1829	23	-4,7136	< 0,001
0					
$\beta_{13}$	-0,4337	0,1523	23	-2,8481	0,0091
$egin{array}{c} eta_{13} \ eta_{14} \end{array}$	-0,4337 -0,1307	$0,1523 \\ 0,0186$	$\frac{23}{23}$	$-2,8481 \\ -7,0440$	0,0091 < 0,001
$\begin{array}{c} \beta_{13} \\ \beta_{14} \\ \beta_{21} \end{array}$	-0,4337 -0,1307 -0,3434	$0,1523 \\ 0,0186 \\ 0,0420$	$23 \\ 23 \\ 294$	-2,8481 -7,0440 -8,1685	$\begin{array}{c} 0,0091 \\ < 0,001 \\ < 0,001 \end{array}$
$\frac{\begin{array}{c}\beta_{13}\\\beta_{14}\\ \\ \hline \\ \hline \\ \sigma_N^2\end{array}}$	$-0,4337 \\ -0,1307 \\ -0,3434 \\ \hline 0,0990$	$0,1523 \\ 0,0186 \\ 0,0420 \\ -$	$23 \\ 23 \\ 294 \\ -$	$ \begin{array}{r} -2,8481 \\ -7,0440 \\ -8,1685 \\ \end{array} $	0,0091 < 0,001 < 0,001 -
$\frac{\begin{array}{c}\beta_{13}\\\beta_{14}\\ \\ \hline \\ \hline \\ \sigma_{21}^2\\ \\ \sigma_{C}^2\\ \\ \sigma_{C}^2\end{array}$	$-0,4337 \\ -0,1307 \\ -0,3434 \\ \hline 0,0990 \\ 0,2646$	$ \begin{array}{r} 0,1523\\ 0,0186\\ 0,0420\\ -\\ -\\ -\\ \end{array} $	$23 \\ 23 \\ 294 \\ - \\ -$	$ \begin{array}{r} -2,8481 \\ -7,0440 \\ -8,1685 \\ - \\ - \\ - \\ - \\ - \\ - \\ - \\ - \\ - \\ -$	$\begin{array}{c} 0,0091 \\ < 0,001 \\ < 0,001 \\ \hline - \\ - \end{array}$

Tabela 2.4: Estimativas dos parâmetros do modelo (2.39).

Utilizando as estimativas de  $\sigma_N^2$ ,  $\sigma_C^2 \in \sigma_{HL}^2$  podemos definir dois coeficientes de correlação intraclasse, que dependem do tratamento aplicado em cada ninhada. Considerando as estimativas de máxima verossimilhança dos componentes de variância e covariância citados, temos que os coeficientes de correlação intraclasse são estimados por

$$\begin{split} & \mathrm{I}\hat{\mathrm{C}}\mathrm{C}_{C} = \frac{\hat{\sigma}_{N}^{2}}{\hat{\sigma}_{N}^{2} + \hat{\sigma}_{C}^{2}} = \frac{0,0832}{0,0832 + 0,2639} = 0,2397;\\ & \mathrm{I}\hat{\mathrm{C}}\mathrm{C}_{HL} = \frac{\hat{\sigma}_{N}^{2}}{\hat{\sigma}_{N}^{2} + \hat{\sigma}_{HL}^{2}} = \frac{0,0832}{0,0832 + 0,0913} = 0,4768 \end{split}$$

em que ICC<sub>C</sub> é o coeficiente de correlação intraclasse para as ninhadas que receberam tratamento controle e ICC<sub>HL</sub> é o ICC para ninhadas que receberam as dosagens alta ou baixa do composto experimental. Desta forma, temos que a correlação intraclasse entre indivíduos de uma ninhada que recebeu dosagem controle do composto experimental é de 0,2397, menor que este mesmo coeficiente para ratos recém nascidos presentes em ninhadas que receberam as dosagens alta ou baixa do tratamento (assumindo valor 0,4768). Os mesmos coeficientes podem ser calculados a partir das estimativas de máxima verossimilhança restrita. Assim, temos

$$I\hat{C}C_{C} = \frac{\hat{\sigma}_{N}^{2}}{\hat{\sigma}_{N}^{2} + \hat{\sigma}_{C}^{2}} = \frac{0,0990}{0,0990 + 0,2646} = 0,2723;$$
  
$$I\hat{C}C_{HL} = \frac{\hat{\sigma}_{N}^{2}}{\hat{\sigma}_{N}^{2} + \hat{\sigma}_{HL}^{2}} = \frac{0,0990}{0,0990 + 0,0918} = 0,5189.$$

Utilizando as estimativas dos parâmetros de variância e covariância via MVR temos que a correlação entre indivíduos de uma ninhada que recebeu dosagem controle do composto experimental é menor para ratos recém nascidos presentes em ninhadas que receberam as dosagens alta ou baixa do tratamento. O ICC para indivíduos provenientes de ninhadas que receberam a dosagem controle é de 0,2723, enquanto que este mesmo coeficiente para ratos filhotes provenientes de ninhadas que receberam as dosagens alta ou baixa é de 0,5189. Note que, considerando os tratamentos, os ICCs obtidos via estimativas de máxima verossimilhança são menores que aquelas mesmas quantidades obtidas a partir das estimativas de máxima verossimilhança restrita. Isto pode ser ocasionado pelo fato de que as estimativas dos parâmetros de variância e covariância obtidas via MV são possivelmente mais viesadas do que aquelas mesmas estimativas obtidas por MVR.

## Análise de Resíduos

A partir do ajuste via MV do modelo (2.39), alguns gráficos podem ser construídos para verificação das suposições assumidas, bem como para algumas análises de importantes características do ajuste propriamente dito. Na Figura 2.3 temos os gráficos dos resíduos marginais e condicionais studentizados. Note que o indivíduo 66 pertencente à ninhada 6 apresenta valor discrepante destes dois resíduos em relação aos demais valores. Este filhote apresentou peso de 3,68g (o menor peso dentre todos os filhotes, como citado anteriormente), é do sexo feminino e a ninhada à qual ele pertencia recebeu o tratamento controle. A existência deste ponto discrepante para os gráficos dos resíduos condicional e marginal pode ser um indício de que o modelo está mal ajustado para esta observação ou, ainda, que a ninhada 6 apresenta valores dos pesos dos ratos discrepantes em relação a estas mesmas medidas para as outras ninhadas.

No gráfico da Figura 2.4, os pontos representam as estimativas das distâncias de Mahalanobis para os vetores de pesos em cada ninhada e as extremidades superiores dos segmentos de retas representam os valores dos percentis 0,95 de distribuições qui-quadrado com  $n_j$  graus de liberdade, para j = 1, 2, ..., 27. Sob o modelo (2.39) as distâncias de Mahalanobis seguem uma distribuição qui-quadrado com os mesmos graus de liberdade citados. Assim, pontos acima das extremidades superiores dos respectivos segmentos de retas podem indicar ninhadas discrepantes em relação às demais. Pode ser notado que as ninhadas 5, 6 e 18 apresentam valores das distâncias de Mahalanobis muito altas em relação aos percentis das distribuições qui-quadrado respectivas, sendo assim os conjuntos de valores dos pesos dos ratos para estas ninhadas podem ser considerados aberrantes em relação aos pesos dos ratos presentes nas demais ninhadas consideradas no estudo. Os pesos médios dos ratos para estas ninhadas são 6,89g, 6,97g e 6,14g, respectivamente. As ninhadas 5 e 6 receberam a dosagem controle do composto experimental, enquanto que a ninhada 18 foi submetia à dosagem baixa do tratamento.

Como no modelo (2.39) há apenas um efeito aleatório por ninhada, a raiz quadrada da distância de Mahalanobis para este efeito segue uma distribuição normal padrão sob a validade do respectivo modelo. A estimativa desta medida é equivalente ao preditor linear studentizado. Sendo assim, na Figura 2.5 temos os valores dos preditores lineares studentizados para cada ninhada. Estes valores podem ser comparados com os percentis da distribuição normal padrão. Valores acima de 2 ou abaixo de -2 são considerados aberrantes em relação aos demais. Assim, temos que as ninhadas 7, 9,..., 22 são aberrantes considerando seus respectivos efeitos aleatórios.

A partir da Figura 2.6 podemos observar que, considerando o gráfico normal de probabilidades com envelope gerado para os resíduos marginais studentizados, há indícios de que a suposição de normalidade dos erros é satisfeita, uma vez que a maioria dos pontos encontra-se dentro da banda de confiança simulada. Existe um ponto discrepante em relação aos demais, possivelmente referente ao rato 66 da ninhada 6. Considerando o gráfico normal de probabilidades com envelope gerado para os resíduos condicionais studentizados, temos que a maioria dos pontos não se encontra dentro da banda de confiança, sendo assim há indícios de que a suposição de normalidade dos erros do modelo considerado não é satisfeita.

Considerando o gráfico da Figura 2.7, temos que a suposição de normalidade dos efeitos aleatórios parece ser atendida, uma vez que todos os pontos encontram-se dispostos dentro da banda de confiança simulada.

Conforme a Figura 2.8 temos que os erros em  $\epsilon$  no modelo (2.39) não se afastam de uma distribuição normal, pois a maioria dos pontos encontra-se dentro da banda de confiança simulada.

Segundo as análises dos resíduos aqui desenvolvidas, a suposição de normalidade dos erros e efeitos aleatórios parece que em geral foi atendida. Contudo, foram identificadas ninhadas discrepantes que podem estar influenciando de forma desproporcional as estimativas dos parâmetros. Uma verificação mais detalhada destes possíveis problemas pode ser realizada através de outras técnicas de análise.



Figura 2.2: Gráficos de box-plot dos pesos dos ratos por tamanho da ninhada (machos  $(\mathbf{a})$  e fêmeas  $(\mathbf{b})$ ).



Figura 2.3: Resíduos marginais e condicionais studentizados referentes ao modelo normal ajustado aos dados do experimento com ratos.



Figura 2.4: Estimativas das distâncias de Mahalanobis para as ninhadas e percentis 0,95 da distribuição qui-quadrado referentes ao modelo normal ajustado aos dados do experimento com ratos.



Figura 2.5: Preditores lineares studentizados referentes ao modelo normal ajustado aos dados do experimento com ratos.



Figura 2.6: Gráficos normais de probabilidades com envelopes gerados para os resíduos marginais e condicionais studentizados referentes ao modelo normal ajustado aos dados do experimento com ratos.



Figura 2.7: Gráfico normal de probabilidades com envelope gerado para os preditores lineares studentizados referentes ao modelo normal ajustado aos dados do experimento com ratos.



Figura 2.8: Gráfico normal de probabilidades com envelope gerado para os resíduos condicionais com confundimento mínimo referentes ao modelo normal ajustado aos dados do experimento com ratos.

## 2.9.2 Aplicação 2 - Estudo Educacional

### Descrição do Estudo

O Estudo de Melhoria Instrucional (*Study of Instructional Improvement*, Hill, Rowan e Ball (2005)) foi realizado por pesquisadores da Universidade de Michigan, nos Estados Unidos, no intuito de investigar os escores de matemática de estudantes da primeira e terceira séries selecionados aleatoriamente em uma amostra aleatória de classes, pertencentes a uma amostra aleatória de escolas. Neste exemplo, temos 1190 estudantes do primeiro grau amostrados a partir de 312 classes, também amostradas a partir de uma amostra de 107 escolas. O estudo resulta em um conjunto de dados em três níveis, em que os estudantes (nível 1) estão distribúídos em classes (nível 2) que, por sua vez, estão distribuídas em escolas (nível 3). O objetivo neste exemplo é investigar a contribuição de variáveis dos níveis 1 (alunos), 2 (classes) e 3 (escolas) na diferença entre escores em matemática registrados no início do jardim de infância e no início do primeiro grau para cada um dos alunos selecionados.

### Análises Iniciais

Na Tabela 2.5 temos uma amostra do banco de dados utilizado na modelagem referente ao estudo educacional, com apenas algumas das variáveis presentes nesta ilustração.

Nível 1		Ni	ível 2	Nível 3		
Ident.	Var. Depend.	Ident.	Covariável	Ident.	Covariável	
CHILDID	MATHGAIN	CLASSID	YEARSTEA	SCHOOLID	HOUSEPOV	
1	32	160	1	1	0,082	
2	109	160	1	1	0,082	
3	56	160	1	1	0,082	
4	83	217	2	1	0,082	
5	53	217	2	1	0,082	
6	65	217	2	1	0,082	
7	51	217	2	1	0,082	
8	66	217	2	1	0,082	
9	88	217	2	1	0,082	
10	-7	217	2	1	0,082	

Tabela 2.5: Amostra do banco de dados referente ao estudo educacional.

Em seguida, através da Tabela 2.6 podemos obter as descrições das variáveis presentes

no banco de dados utilizado nesta aplicação.

Considerando a Tabela 2.7 temos que o ganho médio no escore dos alunos é de 57,57, próximo do valor mediano para esta mesma variável, que é de 56,00. Já o valor médio do escore no jardim de infância é alto (assumindo valor de 466,70) com relação à esta mesma medida considerada para o ganho no escore. Existe pelo menos um professor com menos de um ano de experiência em ensino, bem como pelo menos um professor com 40 anos de experiência.

De acordo com a Figura 2.9, nota-se que há uma relação aproximadamente linear entre os escores em matemática dos alunos registrados no jardim de infância e o ganho no escore da respectiva disciplina em relação ao primeiro grau. À medida que os valores dos escores no jardim de infância crescem, os ganhos nos escores decrescem.



Figura 2.9: Gráfico de dispersão bidimensional entre os escores de matemática no jardim de infância e os ganhos nos escores dos alunos.

Considerando o gráfico da Figura 2.10 não há evidências de relação linear entre os status socioeconômicos dos alunos e seus respectivos ganhos nos escores de matemática.

Os valores medianos dos ganhos nos escores são praticamente iguais para meninos e meninas, com alguns valores discrepantes nestes dois grupos, como pode ser visto nos gráficos da Figura 2.11. O mesmo pode ser dito em relação aos valores medianos dos



Figura 2.10: Gráfico de dispersão bidimensional entre os escores de matemática no jardim de infância e os status socioeconômicos dos alunos.

escores segundo a maioridade dos alunos, como pode ser verificado a partir dos gráficos de box-plot da Figura 2.12. Nota-se também a partir destes gráficos que existem indícios de uma pequena assimetria à direita nos ganhos nos escores segundo as variáveis explicativas consideradas.

Segundo a preparação em matemática dos professores, ou seja, o número de conteúdos dominados nesta área ou em áreas afins, temos que os valores medianos dos ganhos nos escores dos alunos segundo diferentes valores da variável MATHPREP não apresentam grandes diferenças, havendo valores discrepantes de ganhos nos escores para a maioria dos valores da preparação em matemática dos docentes (consequentemente das classes) considerados, como visto a partir da Figura 2.13. Nota-se também indícios de uma leve assimetria à direita nos ganhos nos escores de acordo com a preparação em matemática dos professores.

A partir da Figura 2.14, temos que não há grandes variações nos valores medianos dos ganhos nos escores segundo os diferentes valores da variável YEARSTEA (anos de experiência em ensino dos professores). Contudo, há alguns valores atípicos para a maioria dos anos de experiência registrados. Assim como observado a partir do gráfico anterior,



Figura 2.11: Gráficos de box-plot dos ganhos nos escores pelo sexo dos alunos.

há indícios de uma leve assimetria à direita nos ganhos nos escores, agora considerando os anos de experiência em ensino dos professores considerados no estudo educacional.

Em relação ao conhecimento em matemática dos professores, temos a partir da Figura 2.15 que os valores medianos dos ganhos nos escores em matemática dos alunos não apresentam nenhuma tendência ou grande diferença nas variações dos ganhos entre os diferentes valores da variável MATHKNOW.

Na Figura 2.16 pode ser notado que não há tendência dos valores medianos dos ganhos nos escores segundo os diferentes valores percentuais de famílias abaixo do índice de pobreza nas proximidades das escolas consideradas no estudo. As variabilidades registradas para cada conjunto de ganhos segundo cada um dos valores da variável HOUSEPOV são de magnitudes similares entre si.

Pode ser notado a partir da Figura 2.17 que, na escola 1 os valores medianos dos ganhos nos escores em matemática dos alunos segundo as duas classes pertencentes a esta escola são aproximadamente iguais, enquanto que nas escolas 2, 3 e 4 estes valores medianos apresentam valores mais afastados entre si, segundo suas respectivas classes. Existe então variabilidade entre os valores medianos dos ganhos nos escores para diferentes classes



Figura 2.12: Gráficos de box-plot dos ganhos nos escores pela maioridade dos alunos. numa determinada escola, assim como entre as diferentes escolas.

# Modelagem

Para a modelagem utilizada a partir desta aplicação, foi utilizada a estratégia "Step-Up" de seleção de modelos, descrita na Seção 2.7.2. Assim como no caso do exemplo de aplicação anterior, a sequência de procedimentos realizados para a seleção do modelo final estão descritos em West, Welch e Galecki (2007). Sendo assim, temos o modelo selecionado


Figura 2.13: Gráficos de box-plot dos ganhos nos escores pela preparação em matemática dos professores.

dado por

## Equação do Nível 1:

 $MATHGAIN_{ijk} = \gamma_{1j|k} + \gamma_{2j|k}MATHKIND_{ijk} + \gamma_{3j|k}MINORITY_{ijk} + \gamma_{4j|k}SES_{ijk} + \epsilon_{ijk};$ 

## Equações do Nível 2:

$$\begin{split} \gamma_{1j|k} &= \delta_{11k} + \xi_{1j|k}; \\ \gamma_{2j|k} &= \delta_{21k}; \\ \gamma_{3j|k} &= \delta_{31k}; \\ \gamma_{4j|k} &= \delta_{41k}; \\ \mathbf{Equações do Nível 3:} \\ \delta_{11k} &= \beta_{111} + \vartheta_{11k}; \\ \delta_{21k} &= \beta_{211}; \\ \delta_{31k} &= \beta_{311}; \\ \delta_{41k} &= \beta_{411}, \end{split}$$

(2.41)

em que  $i = 1, 2, ..., n_{jk}, j = 1, 2, ..., m_k$  e k = 1, 2, ..., 107. Na equação referente ao nível 1 do modelo, temos MATHGAIN<sub>ijk</sub> representando o ganho no escore em matemática do *i*-ésimo aluno pertencente à *j*-ésima classe da *k*-ésima escola,  $\gamma_{1j|k}$  o intercepto para a *j*-ésima classe da *k*-ésima escola,  $\gamma_{2j|k}$  o parâmetro associado à variável MATHKIND para



Figura 2.14: Gráficos de box-plot dos ganhos nos escores pelos anos de experiência em ensino dos professores.

a *j*-ésima classe da *k*-ésima escola, MATHKIND<sub>*ijk*</sub> o valor do escore em matemática no jardim de infância para o *i*-ésimo aluno pertencente à *j*-ésima classe da *k*-ésima escola,  $\gamma_{3j|k}$  o parâmetro associado à variável MINORITY para a a *j*-ésima classe da *k*-ésima escola, MINORITY<sub>*ijk*</sub> a variável indicadora de maioridade para o *i*-ésimo aluno pertencente à *j*-ésima classe da *k*-ésima escola,  $\gamma_{4j|k}$  o parâmetro associado à variável SES para a *j*-ésima classe da *k*-ésima escola, SES<sub>*ijk*</sub> o índice socioeconômico para o *i*-ésimo aluno pertencente à *j*-ésima classe da *k*-ésima escola e um erro aleatório  $\epsilon_{ijk}$  associdado ao valor MATHGAIN<sub>*ijk*</sub>. É assumido que  $\epsilon_{ijk} \sim N(0, \sigma^2)$ , com os  $\epsilon_{ijk}$ 's independentes entre si, para  $i = 1, 2, ..., n_{jk}, j = 1, 2, ..., m_k$  e k = 1, 2, ..., 107.

Na primeira equação referente ao nível 2 do modelo, temos  $\delta_{11k}$  representando a média geral dos interceptos  $\gamma_{1j|k}$ 's na k-ésima escola e  $\xi_{1j|k}$  sendo um erro aleatório associado ao valor de  $\gamma_{1j|k}$ . É assumido que  $\xi_{1j|k} \sim N(0, \sigma_C^2)$ , com os  $\xi_{1j|k}$ 's independentes entre si e independentes dos  $\epsilon_{ijk}$ 's, para  $i = 1, 2, \ldots, n_{jk}, j = 1, 2, \ldots, m_k$  e  $k = 1, 2, \ldots, 107$ . Na segunda equação referente ao nível 2 temos  $\delta_{21k}$  denotando que o valor de  $\gamma_{2j|k}$  é o mesmo para todas as  $m_k$  classes da escola k; na terceira equação referente ao nível 2 temos  $\delta_{31k}$  denotando que o valor de  $\gamma_{3j|k}$  é o mesmo para todas as  $m_k$  classes da escola ke analogamente, na quarta equação referente ao nível 2 temos  $\delta_{41k}$  denotando que o valor de  $\gamma_{4j|k}$  é o mesmo para todas as  $m_k$  classes da escola k.



Figura 2.15: Gráficos de box-plot dos ganhos nos escores pelo conhecimento em matemática dos professores.

Na primeira equação referente ao nível 3 do modelo, temos  $\beta_{111}$  representando a média geral dos  $\delta_{11k}$  para as 107 escolas e  $\vartheta_{11k}$  sendo um erro aleatório associado ao valor de  $\delta_{11k}$ . É assumido que  $\vartheta_{11k} \sim N(0, \sigma_E^2)$ , com os  $\vartheta_{11k}$ 's independentes entre si e independentes dos  $\epsilon_{ijk}$ 's e dos  $\xi_{j|k}$ 's, para  $i = 1, 2, \ldots, n_{jk}, j = 1, 2, \ldots, m_k$  e  $k = 1, 2, \ldots, 107$ . Na segunda equação referente ao nível 3 temos  $\beta_{211}$  denotando que o valor de  $\delta_{21k}$  é o mesmo para todas as 107 escolas; na terceira equação referente ao nível 3 temos  $\beta_{311}$  denotando que o valor de  $\delta_{31k}$  é o mesmo para todas as 107 escolas e, finalmente, na quarta equação referente ao nível 3 temos  $\beta_{411}$  denotando que o valor de  $\delta_{41k}$  é o mesmo para todas as 107 escolas.

O modelo (2.41) pode, ainda, ser expresso na forma

$$MATHGAIN_{ijk} = \beta_{111} + \beta_{211}MATHKIND_{ijk} + \beta_{311}MINORITY_{ijk} + \beta_{411}SES_{ijk} + \xi_{1j|k} + \vartheta_{11k} + \epsilon_{ijk},$$
(2.42)

para  $i = 1, 2, ..., n_{jk}$   $j = 1, 2, ..., m_k$  e k = 1, 2, ..., 107. Segundo o modelo expresso em (2.42) podemos dizer que  $\beta_{111}$  é o intercepto do modelo,  $\beta_{211}$  é o parâmetro fixo associado à variável MATHKIND,  $\beta_{311}$  é o efeito fixo da menoridade dos alunos no valor esperado do ganho no escore em matemática,  $\beta_{411}$  é o parâmetro fixo associado à variável SES,  $\xi_{1j|k}$ é o efeito aleatório da *j*-ésima classe pertencente à *k*-ésima escola nos ganhos nos escores



Figura 2.16: Gráficos de box-plot dos ganhos nos escores pelos percentuais de famílias abaixo da linha de probreza nas proximidades das escolas.

em matemática dos alunos desta mesma classe pertencente a esta mesma escola,  $\vartheta_{11k}$  é o efeito aleatório da k-ésima escola nos valores da variável MATHGAIN para a referida classe e  $\epsilon_{ijk}$  é o erro aleatório associado ao valor de MATHGAIN<sub>ijk</sub>.

A notação matricial do modelo (2.41) para a j-ésima classe pertencente à k-ésima escola é, portanto, dada por

$$\mathbf{Y}_{jk} = (\text{MATHGAIN}_{1jk}, \text{MATHGAIN}_{2jk}, \dots, \text{MATHGAIN}_{n_{jk}jk})^{t}_{n_{jk} \times 1};$$

$$\mathbf{G}_{jk} = \begin{bmatrix} 1 & \text{MATHKIND}_{1jk} & \text{MINORITY}_{1jk} & \text{SES}_{1jk} \\ 1 & \text{MATHKIND}_{2jk} & \text{MINORITY}_{2jk} & \text{SES}_{2jk} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{MATHKIND}_{n_{jk}jk} & \text{MINORITY}_{n_{jk}jk} & \text{SES}_{n_{jk}jk} \end{bmatrix}_{n_{jk} \times 4}^{,}$$
$$\mathbf{T}_{jk} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4 \times 4}^{,} \mathbf{L}_{k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4 \times 4}^{,}$$
$$\boldsymbol{\epsilon}_{jk} = (\epsilon_{1jk}, \epsilon_{2jk}, \dots, \epsilon_{n_{jk}jk})^{t} = \boldsymbol{\gamma}_{1j}; \boldsymbol{\gamma}_{jk} = (\gamma_{1|jk}, \gamma_{2|jk}, \gamma_{3|jk}, \gamma_{4|jk})^{t} + \boldsymbol{\gamma}_{1j};$$



Figura 2.17: Gráficos de box-plot dos ganhos nos escores segundo as classes pertencentes às quatro primeiras escolas amostradas.

$$\boldsymbol{\delta}_{k} = (\delta_{11k}, \delta_{21k}, \delta_{31k}, \delta_{41k})^{t}_{4\times 1}; \; \boldsymbol{\beta} = (\beta_{111}, \beta_{211}, \beta_{311}, \beta_{411})^{t}_{4\times 1};$$
$$\boldsymbol{\xi}_{jk} = (\xi_{1j|k}, 0, 0, 0)^{t}_{4\times 1}; \; \boldsymbol{\vartheta}_{k} = (\vartheta_{11k}, 0, 0, 0)^{t}_{4\times 1},$$

para  $j = 1, 2, \ldots, m_k$  e  $k = 1, 2, \ldots, 107$ . Considerando as  $m_k$  classes da k-ésima escola, temos

$$\mathbf{Y}_k = (\mathbf{Y}_{1k}^t, \mathbf{Y}_{2k}^t, \dots, \mathbf{Y}_{m_k k}^t)^t (\sum_{j=1}^{m_k} n_{jk}) \times 1;$$

$$\mathbf{G}_{k} = \begin{bmatrix} \mathbf{G}_{1k} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{2k} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_{mkk} \end{bmatrix}_{(\sum_{j=1}^{m_{k}} n_{jk}) \times (4m_{k})};$$
  
$$\boldsymbol{\gamma}_{k} = (\boldsymbol{\gamma}_{1k}^{t}, \boldsymbol{\gamma}_{2k}^{t}, \dots, \boldsymbol{\gamma}_{mkk}^{t})^{t}_{(4m_{k}) \times 1}; \quad \boldsymbol{\epsilon}_{k} = (\boldsymbol{\epsilon}_{1k}^{t}, \boldsymbol{\epsilon}_{2k}^{t}, \dots, \boldsymbol{\epsilon}_{mkk}^{t})^{t}_{(\sum_{j=1}^{m_{k}} n_{jk}) \times 1};$$
  
$$\mathbf{T}_{k} = [\mathbf{T}_{1k}^{t} | \mathbf{T}_{2k}^{t} | \dots | \mathbf{T}_{mkk}^{t} ]^{t}_{(4m_{k}) \times 4}; \quad \boldsymbol{\xi}_{k} = (\boldsymbol{\xi}_{1k}^{t}, \boldsymbol{\xi}_{2k}^{t}, \dots, \boldsymbol{\xi}_{mkk}^{t})^{t}_{(4m_{k}) \times 1};$$
  
$$\boldsymbol{b}_{k} = (\boldsymbol{\vartheta}_{k}^{t}, \boldsymbol{\xi}_{k}^{t})_{(4+4m_{k}) \times 1}; \quad \mathbf{X}_{k} = \mathbf{G}_{k} \mathbf{T}_{k} \mathbf{L}_{k}; \quad \mathbf{Z}_{k} = [\mathbf{G}_{k} \mathbf{T}_{k} | \mathbf{G}_{k}].$$

As estimativas dos parâmetros obtidas pelos métodos de máxima verossimilhança e máxima verossimilhança restrita são apresentados na Tabela 2.8. Para um nível de significância de, por exemplo, 5%, todos os efeitos fixos são marginalmente significativos pelos testes do tipo t, através dos dois métodos de estimação. As estimativas dos parâmetros dos efeitos fixos obtidas por MV e MVR apresentaram valores próximos para cada parâmetro. Considerando os erros padrão estimados para as estimativas dos parâmetros dos efeitos fixos, os valores obtidos via MVR são maiores em relação àqueles obtidos via MV. O mesmo acontece para as estimativas dos parâmetros de variância e covariância segundo os dois métodos de estimação.

Interpretando as estimativas obtidas por máxima verossimilhança, temos que o ganho médio no escore para alunos que obtiveram escore zero no jardim de infância, já atingiram a maioridade e que apresentam índice socioeconômico assumindo valor zero é de 282,3436. Considerados fixos o índice socioeconômico e a categoria de maioridade ou menoridade dos alunos, temos que à medida que o escore obtido no jardim de infância é aumentado em uma unidade, o ganho médio no escore em matemática diminue em 0,4701, ou seja, alunos que obtiveram escores baixos no jardim de infância tendem a obter maior recuperação no escore que alunos que já obtiveram escores mais altos no jardim de infância. Fixadas as variáveis MATHKIND e SES, temos que alunos na menoridade apresentam ganho médio no escore menor em 8,2850 em relação a alunos na maioridade, ou seja, alunos na maioridade apresentam ganho no escore maior em relação a alunos na menoridade. Considerados fixos o escore no jardim de infância e a categoria de maioridade ou menoridade dos alunos, temos que o aumento em uma unidade no índice socioeconômico implica no aumento médio do ganho no escore dos alunos em 5,3606, ou seja, quando maior o índice socioeconômico, maior o ganho no escore em matemática. A interpretação das estimativas dos parâmetros obtidas via MVR é realizada de forma similar a esta aqui detalhada.

Utilizando as estimativas de  $\sigma_E^2$ ,  $\sigma_C^2 e \sigma^2$  podemos definir duas estimativas de coeficientes de correlação intraclasse. A primeira leva em consideração a variabilidade dos ganhos nos escores entre diferentes escolas. Assim temos o primeiro coeficiente de correlação intraclasse dado por

$$ICC_E = \frac{\sigma_E^2}{\sigma_E^2 + \sigma_C^2 + \sigma^2}.$$
(2.43)

O valor de (2.43) é alto se os ganhos nos escores para indivíduos numa mesma escola apresentam pouca variabilidade em relação à variabilidade total dos ganhos nos escores considerando diferentes escolas.

O segundo coeficiente de correlação intraclasse considera a variabilidade dos ganhos nos escores entre diferentes classes de uma determinada escola e entre diferentes escolas. Desta forma, este coeficiente pode ser definido por

$$ICC_C = \frac{\sigma_E^2 + \sigma_C^2}{\sigma_E^2 + \sigma_C^2 + \sigma^2}.$$
(2.44)

Temos, assim, que o valor de (2.44) é alto se os ganhos nos escores para indivíduos numa mesma classe (e consequentemente pertencentes à mesma escola) apresentam pouca variabilidade em relação à variabilidade total dos ganhos nos escores considerando diferentes escolas. Considerando as estimativas de máxima verossimilhança dos parâmetros de variância e covariância do modelo (2.41), temos

$$\begin{split} & \text{I}\hat{\text{C}}\text{C}_{E} = \frac{\hat{\sigma}_{E}^{2}}{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2} + \hat{\sigma}^{2}} = \frac{72,5109}{72,5109 + 82,7348 + 732,9298} = 0,0816;\\ & \text{I}\hat{\text{C}}\text{C}_{C} = \frac{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2}}{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2} + \hat{\sigma}^{2}} = \frac{72,5109 + 82,7348 + 732,9298}{72,5109 + 82,7348 + 732,9298} = 0,1748. \end{split}$$

Para as estimativas de máxima verossimilhança restrita, temos

$$\begin{split} & \text{I}\hat{\text{C}}\text{C}_{E} = \frac{\hat{\sigma}_{E}^{2}}{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2} + \hat{\sigma}^{2}} = \frac{75,0358}{75,0358 + 82,8395 + 734,6077} = 0,0841;\\ & \text{I}\hat{\text{C}}\text{C}_{C} = \frac{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2}}{\hat{\sigma}_{E}^{2} + \hat{\sigma}_{C}^{2} + \hat{\sigma}^{2}} = \frac{75,0358 + 82,7348}{75,0358 + 82,8395 + 734,6077} = 0,1768. \end{split}$$

As estimativas dos ICCs utilizando o método de estimação de máxima verossimilhança restrita são maiores em relação às estimativas destes mesmos coeficientes utilizando o método de MV. Para os dois métodos de estimação temos que a variabilidade dos ganhos nos escores para alunos de diferentes escolas é menor que a variabilidade dos ganhos nos escores para alunos numa mesma classe (considerando a variabilidade total dos ganhos nos escores). Porém, ambos os valores dos coeficientes de correlação intraclasse podem ser considerados baixos, sendo eles obtidos por MV ou via MVR.

#### Análise de Resíduos

Nesta seção será realizada a análise dos resíduos obtidos a partir do ajuste via máxima verossimilhança do modelo (2.41), através da construção de alguns gráficos no intuito de verificar as suposições assumidas no modelo, bem como para algumas análises de importantes características do ajuste.

Na Figura 2.18 temos os gráficos dos resíduos marginais e condicionais studentizados. Note que o aluno 41 pertencente à classe 179 da escola 4 apresenta valor discrepante destes dois resíduos em relação aos demais valores. Este aluno apresentou ganho no escore de -110, ou seja, seu escore no ensino médio (324) foi menor que o seu escore registrado no jardim de infância (434). Este aluno está na menoridade e apresentou valor de índice socioecomônico de -0,37. A existência deste ponto discrepante para os gráficos dos resíduos condicional e marginal pode ser um indício de que o modelo está mal ajustado para esta observação.

Considerando as estimativas das distâncias de Mahalanobis para as escolas, temos a partir do gráfico da Figura 2.19 que as escolas 4, 47, 62 e 85 são discrepantes em relação às demais, pois os pontos econtram-se afastados das extremidades superiores dos segmentos de retas que representam os percentis 0,95 da distribuição qui-quadrado. Os ganhos médios nos escores dos alunos nestas escolas são de 35,17, 73,80, 54,00 e 52,80 respectivamente.



Figura 2.18: Resíduos marginais e condicionais studentizados referentes ao modelo normal ajustado aos dados do estudo educacional.

Através do gráfico da Figura 2.20, as escolas 3, 68, 70 e 76 são consideradas aberrantes no que diz respeito aos seus efeitos aleatórios. Com relação aos efeitos das classes nas escolas, nenhuma classe foi considerada atípica em relação aos seus respectivos efeitos aleatórios.

Considerando para cada escola o seu efeito aleatório e os efeitos de suas classes, temos a partir do gráfico da Figura 2.21 que a escola 76 pode ser considerada atípica em relação às demais, no que diz respeito aos efeitos aleatórios citados, pois o valor estimado da distância de Mahalanobis dos efeitos aleatórios está bem acima da extremidade superior do segmento de reta que representa o percentil da distribuição qui-quadrado associado à referida distância.

Podemos observar na Figura 2.22 que, tanto para o gráfico de percentis dos resíduos condicionais como para o gráfico referente aos resíduos marginais, os pontos mais extremos não se encontram dispostos dentro dos respectivos envelopes, principalmente aqueles dispostos nas partes superiores dos respectivos gráficos. Este resultado reforça o indício da presença de uma leve assimetria nos dados referentes ao estudo educacional Contudo, a proposta de uma distribuição com caudas mais pesadas do que as caudas da distribuição normal no ajuste de um modelo multinível pode ser eficiente para uma melhor acomodação destas observações.



Figura 2.19: Estimativas das distâncias de Mahalanobis para as escolas e percentis 0,95 da distribuição qui-quadrado referentes ao modelo normal ajustado aos dados do estudo educacional.

A partir do gráfico da Figura 2.23 podemos concluir que não há afastamento da suposição de normalidade dos efeitos aleatórios, pois os pontos encontram-se dispostos dentro do envelope simulado.

Considerando o gráfico da Figura 2.24, temos que os erros condicionais do modelo (2.41) seguem aproximadamente uma distribuição Normal, pois os pontos encontram-se dentro da banda de confiança simulada.

A partir dos ajustes dos modelos nas duas aplicações consideradas neste capítulo, podemos destacar a presença de observações e/ou grupos identificados como possivelmente aberrantes em ambos os casos. Isto pode implicar em falta de robustez das estimativas de máxima verossimilhança dos parâmetros. Uma alternativa é o uso de modelos multiníveis com distribuições de probabilidade que apresentam caudas mais pesadas em relação à distribuição normal. A t-Student é um exemplo de distribuição com esta característica. Sendo assim, no capítulo seguinte temos uma nova proposta referente aos modelos multiníveis, agora admitindo distribuições pertencentes à classe elíptica.



Figura 2.20: Preditores lineares studentizados referentes ao modelo normal ajustado aos dados do estudo educacional.



Figura 2.21: Estimativas das distâncias de Mahalanobis para os efeitos aleatórios e percentis 0,95 da distribuição qui-quadrado referentes ao modelo normal ajustado aos dados do estudo educacional.

## Tabela 2.6: Descrições das variáveis consideradas no estudo educacional. Variáveis do nível 1 (Alunos)

CHILDID:	número de identificação do aluno;				
MATHGAIN:	ganho no escore em matemática do aluno do início do jardim de infância ao início do primeiro grau (variável resposta);				
MATHKIND:	escore em matemática do aluno no início de seu jardim de infância;				
SEX:	sexo do aluno;				
MINORITY:	variável indicadora (0=estudante na maioridade, 1=estudante na menoridade);				
SES:	estatus socioeconômico do estudante.				
Variáveis do nível 2 (Classes)					
CLASSID:	número de identificação da classe;				
YEARSTEA:	anos de experiência em ensino do professor do primeiro grau;				
MATHPREP:	preparação em matemática do professor do primeiro grau: número de conteúdos dominados em matemática e cursos na área;				
MATHKNOW: conhecimento em matemática do professo do primeiro grau: uma escala baseada em um conjunto de 30 itens (maiores valores indicam maior conhecimento).					
Va	ariáveis do nível 3 (Escolas)				
SCHOOLID:	número de identificação da escola;				
HOUSEPOV:	percentual de famílias abaixo do				

índice de pobreza nas proximidades da escola.

Nível 1 (Alunos)									
Variável	Mínimo	1° quartil	Mediana	Média	3° quartil	Máximo	Desvio		
MATHKIND	290,00	439, 20	466,00	466,70	495,00	629,00	41,46		
MATHGAIN	-110,00	35,00	56,00	57, 57	77,00	253,00	34, 61		
SES	-1,61	-0,49	-0,03	-0,01	0, 40	3, 21	0,74		
Nível 2 (Classes)									
Variável	Mínimo	1º quartil	Mediana	Média	3° quartil	Máximo	Desvio		
YEARSTEA	0,00	4,00	10,00	12, 28	20,00	40,00	9,65		
MATHPREP	1,00	2,00	2, 30	2,58	3,00	6,00	0,96		
MATHKNOW	-2,50	-0,76	-0, 19	-0,08	0,62	2,61	0,99		
Nível 3 (Escolas)									
Variável	Mínimo	1° quartil	Mediana	Média	3° quartil	Máximo	Desvio		
HOUSEPOV	0,01	0,09	0, 15	0, 19	0, 26	0, 56	0, 14		

Tabela 2.7: Medidas descritivas de algumas variáveis consideradas no estudo educacional.

Tabela 2.8: Estimativas o	los parâmetros d	o modelo (	(2.41).
---------------------------	------------------	------------	---------

		MV			
Parâmetro	Estimativa	Erro padrão	G.l.	Valor t	Valor p
$\beta_{111}$	282,3436	10,8385	875	26,0500	< 0,001
$\beta_{211}$	-0,4701	0,0222	875	-21,1320	< 0,001
$\beta_{311}$	-8,2850	2,3343	875	-3,5492	< 0,001
$eta_{411}$	5,3606	1,2406	875	4,3210	< 0,001
$\sigma_E^2$	72,5109	26,4229	_	_	_
$\sigma_C^2$	82,7348	35,5971	—	—	—
$\sigma^2$	732,9298	42,4142	—	_	—
		MVR			
Parâmetro	Estimativa	Erro padrão	G.l.	Valor t	Valor p
$\beta_{111}$	282,4193	10,8406	875	26,0520	< 0,001
$\beta_{211}$	-0,4703	0,0223	875	-21,1367	< 0,001
$\beta_{311}$	-8,2909	2,3389	875	-3,5448	< 0,001
$eta_{411}$	5,3646	1,2407	875	4,3240	< 0,001
$\sigma_E^2$	75,0358	—	_	—	_
$\sigma_C^2$	82,8395	—	_	—	—
$\sigma^2$	734,6077	—	—	—	—



Figura 2.22: Gráficos normais de probabilidades com envelopes gerados para os resíduos marginais e condicionais studentizados referentes ao modelo normal ajustado aos dados do estudo educacional.



Figura 2.23: Gráfico normal de probabilidades com envelope gerado para os preditores lineares studentizados referentes ao modelo normal ajustado aos dados do estudo educacional.



Figura 2.24: Gráfico normal de probabilidades com envelope gerado para os resíduos condicionais com confundimento mínimo referentes ao modelo normal ajustado aos dados do estudo educacional.

## Capítulo 3

# Modelos Elípticos Multiníveis

No presente capítulo temos uma introdução à classe de distribuições elípticas. Posteriormente, temos a definição de uma nova proposta para modelagem de dados com estrutura de hierarquia, denominada modelos elípticos multiníveis. São apresentados o processo de estimação, a descrição de alguns tipos de testes de hipóteses para os parâmetros dos efeitos fixos e para os parâmetros de variância e covariância dos modelos multiníveis com distribuições elípticas e posteriormente são sugeridas técnicas gráficas para a análise de resíduos. As aplicações trabalhadas no Capítulo 2 são novamente abordadas sob esta nova proposta. Alguns resultados são exibidos e discutidos em paralelo com aqueles obtidos nos ajustes dos modelos normais multiníveis para estas mesmas aplicações citadas.

### 3.1 A Classe de Distribuições Elípticas

A classe de distribuições elípticas abrange todas as distribuições contínuas simétricas. Exemplos de distribuições que pertencem a esta classe são dados pela Normal, t-Student, Exponencial Potência, Logística I, Logística II, entre outras. Suas propriedades podem ser encontradas em Fang, Kotz e Ng (1990). Dizemos que um vetor aleatório  $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{n_jj})^t$  segue uma distribuição elíptica  $n_j$ -variada com vetor de médias  $\boldsymbol{\mu}_j$  e matriz de variância e covariância proporcional à matriz  $\mathbf{V}_j$  se sua função de densidade é dada por

$$f(\mathbf{y}_j) = |\mathbf{V}_j|^{-\frac{1}{2}}g(u_j),$$

em que  $u_j = (\mathbf{y}_j - \boldsymbol{\mu}_j)^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j)$  é a distância de Mahalanobis, a função  $g(\cdot)$  :  $\mathbb{R} \to [0,\infty)$  é tal que  $\int_0^\infty u^{n_j/2-1} g(u) du < \infty$ , sendo denominada função geradora de densidade (ver Fang, Kotz e Ng (1990)). Aqui,  $E(\mathbf{Y}_j) = \boldsymbol{\mu}_j$  e  $Var(\mathbf{Y}_j) = \alpha_j \mathbf{V}_j$ , em que  $\alpha_j$  é obtido a partir da função característica de  $\mathbf{Y}_j$ , que assume a forma  $\varphi_{\mathbf{Y}_j}(t) =$  $E(e^{it^t \mathbf{Y}_j}) = e^{it^t \boldsymbol{\mu}_j} \phi(t^t \mathbf{V}_j t)$ , para alguma função  $\phi$ . Temos então que  $\alpha_j = -2\phi'(\mathbf{0})$ , em que  $\phi'(t) = \partial \phi / \partial t|_{t=0}$ . Para o caso da distribuição t-Student (ver Lange, Little e Taylor (1989)) temos  $\alpha_j = \frac{\nu_j}{\nu_j - 2}$  para  $\nu_j > 2$ , em que  $\nu_j$  denota os graus de liberdade. Para o caso da Exponencial Potência (ver Gómez, Gómez-Villegas e Marín (1998)) temos que  $\alpha_j = 2^{(1+\eta_j)} \frac{\Gamma[(n_j+2)(1+n_j)/2]}{n_j \Gamma[n_j(1+\eta_j)/2]}$ , em que  $\Gamma(\cdot)$  é a função gama e  $\eta_j$  é um parâmetro de forma. Este parâmetro é definido de maneira diferente para cada distribuição elíptica. Denotase então que  $\mathbf{Y}_j \sim El_{n_j}(\boldsymbol{\mu}_j, \mathbf{V}_j, \eta_j)$ . A partir desta parte será considerada a notação simplificada  $\mathbf{Y}_j \sim El_{n_j}(\boldsymbol{\mu}_j, \mathbf{V}_j)$ . Temos ainda que, no caso de  $\mathbf{Y}_j \sim El_{n_j}(\mathbf{0}, \mathbf{I})$ , em que  $\mathbf{I}$ é a matriz identidade de dimensão  $n_j \times n_j$ ,  $\alpha_j$  é a variância das distribuições marginais univariadas obtidas a partir do vetor  $\mathbf{Y}_j$ . Para o caso da distribuição normal temos  $\alpha_j = 1$ .

## 3.2 Definição do Modelo

Os modelos elípticos multiníveis podem ser definidos a partir da estrutura apresentada para modelos normais multiníveis. Sendo assim, considerando o modelo multinível com dois níveis

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j + \boldsymbol{\epsilon}_j,$$

temos duas abordagens que podem ser consideradas. A primeira seria utilizar a formulação hierárquica

$$egin{array}{rcl} \mathbf{Y}_j | oldsymbol{b}_j &\sim & El_{n_j} (\mathbf{X}_j oldsymbol{eta} + \mathbf{Z}_j oldsymbol{b}_j, \mathbf{R}_j); \ oldsymbol{b}_j &\sim & El_P(oldsymbol{0}, oldsymbol{\Psi}_j); \ oldsymbol{\epsilon}_j &\sim & El_{n_j}(oldsymbol{0}, \mathbf{R}_j). \end{array}$$

A dificuldade da utilização desta abordagem reside no fato de que a distribuição conjunta de  $(\mathbf{Y}_j^t, \mathbf{b}_j^t, \mathbf{\epsilon}_j^t)^t$  não é necessariamente elíptica, portanto a distribuição marginal de  $\mathbf{Y}_j$ é analiticamente difícil de ser obtida. Contudo, para o caso particular da distribuição t-Student multivariada (ver Pinheiro, Liu e Wu (2001)) ao adotarmos para uma variável aleatória  $\tau_j$ 

$$\begin{aligned} \mathbf{Y}_{j} | (\boldsymbol{b}_{j}^{t}, \tau_{j})^{t} &\sim N_{n_{j}} (\mathbf{X}_{j} \boldsymbol{\beta} + \mathbf{Z}_{j} \boldsymbol{b}_{j}, \frac{1}{\tau_{j}} \mathbf{R}_{j}); \\ \boldsymbol{b}_{j} | \tau_{j} &\sim N_{P} (\mathbf{0}, \frac{1}{\tau_{j}} \boldsymbol{\Psi}_{j}); \\ \tau_{j} &\sim Gama(\frac{\nu_{j}}{2}, \frac{\nu_{j}}{2}), \end{aligned}$$
(3.1)

temos que

$$\mathbf{Y}_j \sim t_{n_j} (\mathbf{X}_j \boldsymbol{\beta}, \mathbf{Z}_j \boldsymbol{\Psi} \mathbf{Z}_j^t + \mathbf{R}_j, \nu_j),$$

em que  $\mathbf{Y}_j \sim t_{n_j}(\boldsymbol{\mu}_j, \mathbf{V}_j, \nu_j)$  denota que um vetor aleatório  $\mathbf{Y}_j$  segue uma distribuição t-Student  $n_j$ -variada com média  $\boldsymbol{\mu}_j$ , matriz de variância e covariância proporcional a  $\mathbf{V}_j$ e graus de liberdade  $\nu_j$ .

Para o caso da distribuição t-Student multivariada, a abordagem hierárquica citada acima não implica em grandes dificuldades de utilização. No entando, para demais distribuições pertencentes à classe elíptica, a obtenção da distribuição marginal de  $\mathbf{Y}_j$  em geral não é simples, sendo necessária a utilização de métodos de integração numérica. Desta forma, uma outra abordagem pode ser utilizada, considerando

$$\begin{bmatrix} \mathbf{Y}_j \\ \mathbf{b}_j \\ \mathbf{\epsilon}_j \end{bmatrix} \sim El_{n_j+P+n_j} \left( \begin{bmatrix} \mathbf{X}_j \boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_j \boldsymbol{\Psi}_j \mathbf{Z}_j^t + \mathbf{R}_j & \mathbf{Z}_j \boldsymbol{\Psi}_j & \mathbf{R}_j \\ \boldsymbol{\Psi}_j \mathbf{Z}_j^t & \boldsymbol{\Psi}_j & \mathbf{0} \\ \mathbf{R}_j & \mathbf{0} & \mathbf{R}_j \end{bmatrix} \right),$$

para j = 1, 2, ..., J. Neste caso, a partir das propriedades das distribuições elípticas (ver, por exemplo, Valle (1994)), temos que

$$\mathbf{Y}_{j} \sim El_{n_{j}}(\mathbf{X}_{j}\boldsymbol{\beta}, \mathbf{Z}_{j}\boldsymbol{\Psi}_{j}\mathbf{Z}_{j}^{t} + \mathbf{R}_{j}).$$
(3.2)

Sob esta proposta, temos que  $b_j$  e  $\epsilon_j$  são não correlacionados mas não necessariamente independentes, a não ser no caso da distribuição normal. Considerando as J unidades pertencentes ao nível 2 do modelo, podemos supor

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{b} \\ \mathbf{\epsilon} \end{bmatrix} \sim El_{\sum_{j=1}^{J} n_j + JP + \sum_{j=1}^{J} n_j} \left( \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \mathbf{R} & \mathbf{Z}\boldsymbol{\Psi} & \mathbf{R} \\ \boldsymbol{\Psi}\mathbf{Z} & \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

MANGHI, R. F.

Sendo assim, teríamos que

$$\mathbf{Y} \sim El_{\sum_{j=1}^{J} n_j}(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^t + \mathbf{R}).$$

## 3.3 Estimação dos Parâmetros por Máxima Verossimilhança

Para o vetor de parâmetros referentes aos efeitos fixos em  $\beta$  e um vetor  $\theta$  que contém todos os parâmetros de variância e covariância do modelo (3.2), temos a função de densidade de probabilidade de  $\mathbf{Y}_j$  dada por

$$f(\mathbf{y}_j; \boldsymbol{\beta}, \boldsymbol{\theta}) = |\mathbf{V}_j|^{-\frac{1}{2}} g(u_j),$$

em que  $u_j = (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})$  e  $\mathbf{V}_j = \mathbf{Z}_j \boldsymbol{\Psi} \mathbf{Z}_j^{-t} + \mathbf{R}_j$ , para  $j = 1, 2, \dots, J$ . A contribuição do *j*-ésimo grupo ou "cluster" na função de verossimilhança é dada por

$$L_j(\boldsymbol{\beta}, \boldsymbol{\theta}) = f(\mathbf{y}_j; \boldsymbol{\beta}, \boldsymbol{\theta}), \qquad (3.3)$$

portanto, a função de verossimilhança é dada pelo produto das J contribuições definidas em (3.3), ou seja,

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{j} L_j(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{j} |\mathbf{V}_j|^{-\frac{1}{2}} g(u_j).$$

Assim, o logaritmo da função de verossimilhança é definido, neste caso, por

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = \log(L(\boldsymbol{\beta}, \boldsymbol{\theta})) = -\frac{1}{2} \sum_{j=1}^{J} \log|\mathbf{V}_j| + \sum_{j=1}^{J} \log g(u_j).$$
(3.4)

Assumindo o vetor  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)^t$  e derivando (3.4) em relação a  $\boldsymbol{\beta}$  e a  $\theta_d$  obtemos (ver, por exemplo, Savalli, Paula e Cysneiros (2006))

$$\mathbf{S}_{\boldsymbol{\beta}} = \sum_{j=1}^{J} v(u_j) \mathbf{X}_j^{t} \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta});$$

$$\mathbf{S}_{\theta_d} = -\frac{1}{2} \sum_{j=1}^{J} \left[ tr\left( \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_d} \right) - v(u_j) (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_d} \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right],$$

para d = 1, 2, ..., D, em que  $v(u_j) = -2\omega(u_j)$ , sendo  $\omega(u_j) = g'(u_j)/g(u_j)$  e  $g'(u_j) = dg(u_j)/du_j$ . Fazendo  $\mathbf{S}_{\boldsymbol{\beta}} = \mathbf{0}$  obtemos o processo iterativo para estimação dos parâmetros dado por

$$\boldsymbol{\beta}^{(r+1)} = \left[\sum_{j=1}^{J} v(u_j)^{(r)} \mathbf{X}_j^{\ t} \mathbf{V}_j^{-1(r)} \mathbf{X}_j\right]^{-1} \left[\sum_{j=1}^{J} v(u_j)^{(r)} \mathbf{X}_j^{\ t} \mathbf{V}_j^{-1(r)} \mathbf{y}_j\right]; \quad (3.5)$$

$$\boldsymbol{\theta}^{(r+1)} = \arg \max_{\boldsymbol{\theta}} [l(\boldsymbol{\beta}^{(r+1)}, \boldsymbol{\theta})], \qquad (3.6)$$

para  $r = 1, 2, \ldots$  Valores iniciais devem ser utilizados para a inicialização do processo de estimação. Note que este processo é equivalente àquele utilizado para modelos normais multiníveis com distribuição marginal dos  $\mathbf{Y}_j$ 's dada por  $\mathbf{Y}_j \sim N_{n_j}(\sqrt{v(u_j)}\mathbf{X}_j\boldsymbol{\beta}, \mathbf{V}_j)$ . Sendo assim, dados valores de  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$ , rotinas computacionais já implementadas para modelos normais multiníveis podem ser utilizadas na obtenção das atualizações de  $\boldsymbol{\beta}$  e métodos de otimização não linear na obtenção das atualizações de  $\boldsymbol{\theta}$ , através de (3.6). A quantidade  $v(u_j)$  que aparece em (3.5) pode ser vista como um "peso" atribuído a cada grupo ou "cluster" no processo de estimação dos parâmetros do modelo. Já que a função  $g(u_j)$  é na maioria dos casos uma função positiva decrescente, em geral temos  $v(u_j) > 0$ . Para a distribuição t-Student  $v(u_j) = \frac{\nu_j + n_j}{\nu_j + u_j}$ , e para a distribuição Exponencial Potência temos  $v(u_j) = (1 + \eta_j)^{-1} u_j^{[(1+\eta_j)^{-1}]-1}$  (ver Savalli (2005)). No caso da distribuição t-Student,  $v(u_j)$  é inversamente proporcional à distância de Mahalanobis. Portanto, quanto maior for o valor de  $u_j$ , menor será o valor de  $v(u_j)$  e consequentemente menores pesos serão dados a observações aberrantes na estimação dos parâmetros.

As matrizes de informação de Fisher para  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$  são dadas, respectivamente, por

$$\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \sum_{j=1}^{J} \frac{4d_{g_j}}{n_j} \mathbf{X}_j^{t} \mathbf{V}_j^{-1} \mathbf{X}_j \quad e \quad \mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \{\mathbf{K}_{\theta_l \theta_z}\},\$$

em que

$$\mathbf{K}_{\theta_l\theta_z} = \sum_{j=1}^J \left\{ -\frac{c_{jlz}}{4} + \frac{f_{g_j}}{n_j(n_j+2)} \left[ c_{jlz} + 2tr \left( \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_l} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_z} \right) \right] \right\}$$

é o elemento pertencente à *l*-ésima linha e *z*-ésima coluna da matriz  $\mathbf{K}_{\theta\theta}$ , com  $d_{g_j} = E(\omega(u_j)^2 u_j)$ ,  $f_{g_j} = E(\omega(u_j)^2 u_j^2)$  e  $c_{jlz} = tr(\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_l}) tr(\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_z})$ , para l, z = 1, 2, ..., D. Para o caso da distribuição *t*-Student com  $\nu_j$  graus de liberdade temos  $d_{g_j} = \frac{n_j}{4} \left( \frac{\nu_j + n_j}{\nu_j + n_j + 2} \right)$ e  $f_{g_j} = \frac{n_j(n_j+2)}{4} \left( \frac{\nu_j + n_j}{\nu_j + n_j + 2} \right)$ . Para a distribuição Exponencial Potência com parâmetro de forma  $\eta_j$ , temos  $d_{g_j} = \frac{\eta_j^2}{2^{1/\eta_j}} \Gamma(\frac{n_j-2}{2\eta_j}+2) / \Gamma(\frac{n_j}{2\eta_j})$  e  $f_{g_j} = \frac{n_j(n_j+2\eta_j)}{4}$  (ver, por exemplo, Osorio, Paula e Galea (2007)). Os parâmetros  $\boldsymbol{\beta} \in \boldsymbol{\theta}$  são ortogonais e, sendo assim, temos a matriz de informação de Fisher para o vetor de parâmetros  $(\boldsymbol{\beta}^t, \boldsymbol{\theta}^t)^t$  dada por

$$\mathbf{K} = \left[egin{array}{cc} \mathbf{K}_{oldsymbol{eta}oldsymbol{eta}} & \mathbf{0} \ & \ & \mathbf{0} & \mathbf{K}_{oldsymbol{ heta}oldsymbol{ heta}} \end{array}
ight].$$

#### 3.3.1 Algoritmo ECME para a Estimação do Modelo t-Student Multinível

O algoritmo EM (ver Dempster, Laird e Rubin (1977)) é um processo iterativo para a obtenção de estimativas de máxima verossimilhança em modelos com dados faltantes. Assumindo o vetor de dados completos  $\mathbf{y}_{comp}$ , o vetor de dados observados  $\mathbf{y}_{obs}$  e o vetor de dados faltantes  $\mathbf{y}_{falt}$ , podemos dizer que  $\mathbf{y}_{comp} = (\mathbf{y}_{obs}{}^t, \mathbf{y}_{falt}{}^t)^t$ , ou seja, o vetor de dados completos é obtido "agrupando" os dados observados com os dados faltantes. Denotando o conjunto de parâmetros do modelo por  $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ ,  $f(\mathbf{y}_{comp}; \boldsymbol{\omega})$  a função de verossimilhança dos dados completos,  $L(\boldsymbol{\omega})$  a função de verossimilhança dos dados observados e  $Q(\boldsymbol{\omega}; \boldsymbol{\omega}')$ a esperança da função de verossimilhança dos dados completos condicional aos dados observados

$$Q(\boldsymbol{\omega}; \boldsymbol{\omega}') = E[f(\mathbf{y}_{\text{comp}}; \boldsymbol{\omega}) | \mathbf{y}_{\text{obs}}, \boldsymbol{\omega}'],$$

temos que cada iteração do algoritmo EM consiste em dois passos:

Passo E: faça  $Q(\boldsymbol{\omega}; \boldsymbol{\omega}^{(r)})$  como sendo uma função de  $\boldsymbol{\omega}$ ;

Passo M: encontre  $\boldsymbol{\omega}^{(r+1)}$  tal que  $Q(\boldsymbol{\omega}^{(r+1)}; \boldsymbol{\omega}^{(r)}) = \max_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} Q(\boldsymbol{\omega}; \boldsymbol{\omega}^{(r)}).$ 

Cada iteração aumenta o valor de  $L(\omega)$  e, sendo assim, o algoritmo EM geralmente converge para um valor máximo local ou global de  $L(\omega)$  (ver Dempster, Laird e Rubin (1977)). Para assegurar que o algoritmo converge para um máximo global, é recomendável utilizar o processo a partir de vários conjuntos distintos de valores iniciais. Quando o passo M do algortimo é difícil de ser implementado, geralmente é viável substituir este passo por uma sequência de passos restritos, ou seja, cada um maximizando  $Q(\omega; \omega^{(r)})$  sob um conjunto de restrições em  $\omega$ . Esta modificação no passo M define o chamado algoritmo ECM, descrito em Meng e Rubin (1993). Uma outra extensão do algoritmo EM é o algoritmo ECME (ver Liu e Rubin (1994)). Este último substitui o passo CM do algoritmo ECM por um passo CME, que maximiza restritamente a função  $Q(\omega; \omega^{(r)})$  ou a própria função de verossimilhança  $L(\omega)$ . Meng e van Dyk (1997) mostram que um passo "E" é necessário antes dos passos CM que maximizam a função Q. Liu and Rubin (1994) mostram que o ECME geralmente compartilha das mesmas propriedades de simplicidade e estabilidade do algoritmo EM, contudo o primeiro apresenta um tempo até a convergência menor, principalmente no caso da distribuição t-Student com graus de liberdade desconhecidos.

O desenvolvimento do algoritmo ECME utilizado para a estimação em modelos elípticos multiníveis com distribuição marginal t-Student é baseado na proposta sugerida por Pinheiro, Liu e Wu (2001) no contexto de modelos mistos sob esta mesma distribuição. Sendo assim, podemos considerar a formulação hierárquica (3.1) e tratar o vetor  $(\boldsymbol{b}_j^t, \tau_j)^t$  como sendo os dados faltantes. Outra alternativa é integrar (3.1) em  $\boldsymbol{b}_j$ , para  $j = 1, 2, \ldots, J$ . Sendo assim, temos

$$\mathbf{Y}_{j}|\tau_{j} \sim N_{n_{j}}(\mathbf{X}_{j}\boldsymbol{\beta}, \frac{1}{\tau_{j}}\mathbf{V}_{j}); \\
\tau_{j} \sim Gama(\frac{\nu_{j}}{2}, \frac{\nu_{j}}{2}).$$
(3.7)

Uma consequência dessa abordagem é a de que

$$\tau_j | \mathbf{Y}_j = \mathbf{y}_j \sim \operatorname{Gama}(\frac{\nu_j + n_j}{2}, \frac{\nu_j + u_j}{2}),$$

em que  $u_j = (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})$ . Desta forma temos que

$$E(\tau_j | \mathbf{Y}_j = \mathbf{y}_j) = \frac{\nu_j + n_j}{\nu_j + u_j}.$$
(3.8)

Aqui, apenas  $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_J)^t$  é considerado como vetor de dados faltantes. Considere um caso em que os graus de liberdade  $\nu_j$  são desconhecidos e assumidos iguais a  $\nu$  para todo  $j = 1, 2, \dots, J$ . Assim, temos o vetor de parâmetros  $\boldsymbol{\omega} = (\boldsymbol{\beta}^t, \boldsymbol{\theta}^t, \nu)^t$  e, portanto, temos a função de verossimilhança dos dados completos

$$f(\mathbf{y}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}, \nu) = f_1(\mathbf{y}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}) + f_2(\boldsymbol{\tau}; \nu) + c,$$

em que c é um valor que não depende dos parâmetros em  $\boldsymbol{\omega}$  e

$$f_1(\mathbf{y}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \sum_{j=1}^J \left\{ n_j \log |\mathbf{V}_j| + \tau_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right\};$$
  
$$f_2(\boldsymbol{\tau}; \boldsymbol{\nu}) = \sum_{j=1}^J \left\{ \frac{\nu}{2} [\log(\frac{\nu}{2}) + \log(\tau_j) - \tau_j] - \log(\tau_j) - \log\Gamma(\frac{\nu}{2}) \right\}.$$

Fazendo  $\hat{\tau}_j = E(\tau_j | \mathbf{y}_j, \hat{\boldsymbol{\omega}})$ , temos no passo E que

$$Q(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}}) = Q_1(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}}) + Q_2(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}}) = E[f_1(\mathbf{y}, \boldsymbol{\tau}; \boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{y}, \hat{\boldsymbol{\omega}})] + E[f_2(\boldsymbol{\tau}; \boldsymbol{\nu} | \mathbf{y}, \hat{\boldsymbol{\omega}})]$$
  
$$= -\frac{1}{2} \sum_{j=1}^J \left\{ n_j \log |\mathbf{V}_j| + \hat{\tau}_j (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right\} + \sum_{j=1}^J \left\{ \frac{\nu}{2} [\log(\frac{\nu}{2}) + E[\log(\tau_j) | \mathbf{y}, \hat{\boldsymbol{\omega}}] - \hat{\tau}_j] - E[\log(\tau_j) | \mathbf{y}, \hat{\boldsymbol{\omega}}] - \log\Gamma(\frac{\nu}{2}) \right\}.$$

Assim, o algoritmo ECM é dado por:

Passo E: obtenha  $Q(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}})$  como uma função de  $\boldsymbol{\omega}$ ;

Passo CM1: obtenha  $\hat{\boldsymbol{\beta}} \in \hat{\boldsymbol{\theta}}$  que maximize  $Q_1(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}});$ 

Passo CM2: dados  $\hat{\boldsymbol{\beta}} \in \hat{\boldsymbol{\theta}}$  obtidos no passo CM1, obtenha  $\nu$  que maximize  $Q_2(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}})$ .

Como  $E[\log(\tau_j)|\mathbf{y}, \hat{\boldsymbol{\omega}}]$  não possui forma fechada, o passo CM2 é modificado de modo a maximizar a verossimilhança condicional de  $\mathbf{Y}|(\hat{\boldsymbol{\beta}}^t, \hat{\boldsymbol{\theta}}^t)^t$ , ou seja, assumindo

$$\mathbf{Y}_{j} \sim t_{n_{j}}(\mathbf{X}_{j}\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}_{j}, \nu), \qquad (3.9)$$

para j = 1, 2, ..., J. Dessa forma, no passo CM2 a verossimilhança de **Y** a partir das distribuições dos **Y**<sub>j</sub>'s em (3.9) é maximizada em  $\nu$ . Assim, no lugar do passo CM2 temos um passo CME, que consiste em fazer

$$\hat{\nu} = \arg\max_{\nu} \sum_{j=1}^{J} \left\{ \log\Gamma(\frac{\nu+n_j}{2}) - \log\Gamma(\frac{\nu}{2}) + \frac{\nu}{2}\log\nu - \frac{(\nu+n_j)}{2}\log(\nu+\hat{u}_j) \right\},\$$

em que  $\hat{u}_j$  é a distância de Mahalanobis obtida através das estimativas de  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$ , para j = 1, 2, ..., J. Os passos E, CM1 e CME são realizados até a convergência da estimativa de  $\boldsymbol{\omega}$ . Dessa forma, temos o algoritmo ECME para a obtenção das estimativas de máxima verossimilhança de  $\boldsymbol{\beta}, \boldsymbol{\theta} \in \boldsymbol{\nu}$ .

## 3.4 Predição dos Efeitos Aleatórios

Similarmente ao caso normal, podemos considerar para a predição dos efeitos aleatórios a distribuição de  $b_j$  dados os valores observados de  $\mathbf{Y}_j$  dados por  $\mathbf{y}_j$ . Segundo Fang, Kotz e Ng (1990), temos que  $\mathbf{b}_j | \mathbf{Y}_j = \mathbf{y}_j$  segue uma distribuição elíptica. Quando o vetor  $\boldsymbol{\theta}$  é conhecido, podemos utilizar o preditor linear de Bayes empírico

$$\hat{\boldsymbol{b}}_j = \boldsymbol{\Psi}_j \mathbf{Z}_j{}^t \mathbf{V}_j{}^{-1} (\mathbf{y}_j - \mathbf{X}_j \hat{\boldsymbol{eta}})$$

para j = 1, 2, ..., J. No caso de  $\boldsymbol{\theta}$  desconhecido, sua estimativa de máxima verossimilhança é utilizada nas matrizes  $\Psi_j$  e  $\mathbf{V}_j$ . Considerando as J unidades do nível 2, temos

$$\hat{\boldsymbol{b}} = \boldsymbol{\Psi} \mathbf{Z}^t \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}), \qquad (3.10)$$

onde  $\mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_J^t)^t$ .

#### 3.5 Coeficiente de Correlação Intraclasse

Para os modelos elípticos multiníveis, o coeficiente de correlação intraclasse pode ser definido da mesma forma definida para modelos normais. Contudo, aqui a covariância entre indivíduos de mesmo grupo ou "cluster" é apenas proporcional aos parâmetros de variância e covariância considerados no modelo. Apesar deste fato, a interpretação do ICC pode ser a mesma utilizada nos modelos multiníveis com distribuição normal dos erros. Ao tratarmos das aplicações utilizando os modelos elípticos multiníveis, iremos utilizar as mesmas definições dadas na Seção 3.5, para o caso dos ICCs definidos a partir das estimativas de máxima verossimilhança dos parâmetros.

#### 3.6 Escolha do Valor do Parâmetro de Forma

Para o uso de distribuições elípticas em modelos multiníveis, assim como em modelos mistos, é necessário obter o valor do parâmetro de forma  $\eta$ , que define a curtose da distribuição. Dada a escolha da distribuição com a qual se pretende trabalhar, Savalli (2005) sugere a utilização de critérios de informação como por exemplo o AIC para a escolha do valor deste parâmetro. Analogamente a esta sugestão, o critério BIC também pode ser utilizado como critério na escolha do parâmetro de forma  $\eta$ .

#### 3.7 Testes de Hipóteses

Nesta seção serão discutidas algumas técnicas referentes a testes de hipóteses acerca dos parâmetros considerados nos modelos elípticos multiníveis.

#### 3.7.1 Testes de Hipóteses para os Efeitos Fixos

#### Razão de Verossimilhanças (RV)

Similarmente ao caso de modelos normais multiníveis, podemos utilizar o teste da razão de verossimilhanças

$$RV = -2\log\left(\frac{\hat{L}_{red}}{\hat{L}_{comp}}\right) = 2(\hat{l}_{comp} - \hat{l}_{red}), \qquad (3.11)$$

em que  $\hat{l}_{red}$  e  $\hat{l}_{comp}$  são os valores dos logaritmos das funções de verossimilhanças dos modelos reduzido e completo, respectivamente, avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Considerando o mesmo exemplo dado no uso deste teste em modelos normais multiníveis, suponha que possamos particionar o vetor de parâmetros  $\boldsymbol{\beta}$  em  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\ t}, \boldsymbol{\beta}_2^{\ t})^t$ . Assim, caso o interesse seja testar as hipóteses

$$H_0: \quad \boldsymbol{\beta}_1 = \mathbf{0},$$
$$H_1: \quad \boldsymbol{\beta}_1 \neq \mathbf{0},$$

temos que, assintoticamente e sob a hipótese  $H_0$ , (3.11) segue aproximadamente uma distribuição qui-quadrado, com o valor dos graus de liberdade dado pela diferença entre o número de parâmetros contidos no vetor  $\boldsymbol{\beta}$  e o número de parâmetros contidos em  $\boldsymbol{\beta}_2$ . Para a realização deste teste é necessário também assumir que as estruturas de variância e covariância dos modelos reduzido e completo são as mesmas.

#### Testes do tipo z

Dada a similaridade entre os modelos normais e elípticos multiníveis, é razoável considerar que, para um tamanho da amostra  $\sum_{j=1}^{J} n_j$  suficientemente grande,  $\hat{\boldsymbol{\beta}}$  segue uma distribuição aproximadamente normal com média  $\boldsymbol{\beta}$  e matriz de variância e covariância dada por  $\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}$  em analogia aos modelos elípticos mistos (Savalli, Paula e Cysneiros (2006)). Assim, no intuito de testar a significância de um único parâmetro referente a um efeito fixo de uma covariável, o uso de testes do tipo z pode ser um procedimento alternativo ao uso do teste da razão de verossimilhanças. Dado um parâmetro  $\beta_{pq}$  pertencente a  $\boldsymbol{\beta}$ , para testar

$$H_0: \quad \beta_{pq} = 0,$$
$$H_1: \quad \beta_{pq} \neq 0,$$

temos a estatística do teste dada por

$$z = \frac{\hat{\beta}_{pq}}{\widehat{\exp}(\hat{\beta}_{pq})},\tag{3.12}$$

em que  $\hat{\beta}_{pq}$  é a o valor da estimativa de máxima verossimilhança de  $\beta_{pq}$  e  $\hat{ep}(\hat{\beta}_{pq})$  é o valor da estimativa do erro-padrão de  $\hat{\beta}_{pq}$ , obtida a partir da matriz  $\mathbf{K}_{\beta\beta}^{-1}$ . O uso deste teste para pequenas amostras requer maiores estudos quanto à aproximação da distribuição da estatística do teste pela normal. Contudo, nos exemplos aqui desenvolvidos iremos utilizar este teste para comparação dos resultados obtidos nos modelos normais com a nova proposta de modelos aqui sugerida.

## 3.7.2 Testes de Hipóteses para os Parâmetros de Variância e Covariância Razão de Verossimilhanças

Para testar hipóteses a respeito dos parâmetros de variância e covariância em  $\theta$  nos modelos elípticos multiníveis, alguns importantes aspectos devem ser considerados. Assim como no caso do teste de RV para efeitos fixos, temos a estatística do teste dada por

$$RV = -2\log\left(\frac{\hat{L}_{red}}{\hat{L}_{comp}}\right) = 2(\hat{l}_{comp} - \hat{l}_{red}), \qquad (3.13)$$

em que  $l_{red}$  e  $l_{comp}$  são os valores dos logaritmos das funções de verossimilhanças dos modelos reduzido e completo, respectivamente, avaliadas nas estimativas de máxima verossimilhança dos parâmetros. Neste caso, é necessário que o vetor de parâmetros  $\beta$  seja o mesmo para os modelos completo e reduzido. Assim, como no caso normal, quando os parâmetros de variância e covariância sob a hipótese nula não se encontram na fronteira do espaço paramétrico, temos que (3.13) segue assintoticamente uma distribuição qui-quadrado com o número de graus de liberdade dado pela diferença entre o número de parâmetros de variância e covariância do modelo de referência (modelo sob a hipótese alternativa) e o número de parâmetros do mesmo tipo a serem estimados no modelo reduzido (modelo sob a hipótese nula).

#### Testes do Tipo Escore

Para  $D \ge 2$ , a estatística do teste da razão de veros similhanças para testar as hipóteses

$$H_0: \quad \boldsymbol{\theta} = 0,$$
$$H_1: \quad \boldsymbol{\theta} > 0,$$

com pelos menos uma desigualdade estrita em  $H_1$ , não segue uma distribuição quiquadrado, nem mesmo uma mistura de distribuições qui-quadrado (ver, por exemplo, Savalli, Paula e Cysneiros (2006)), quando  $\theta$  encontra-se na fronteira do espaço paramétrico sob a hipótese  $H_0$ . Neste caso, Savalli, Paula e Cysneiros (2006) propuseram um teste do tipo escore baseado em resultados obtidos em Silvapulle e Silvapulle (1995). Estes primeiros autores mostram que a estatística do teste, em geral, segue uma distribuição que é uma mistura de qui-quadrados. Estes testes possuem a vantagem de serem executados a partir das estimativas dos parâmetros obtidas apenas no modelo sob a hipótese nula. Exemplos de testes para um ou mais parâmetros de variância e covariância em modelos elípticos mistos são apresentados em Savalli (2005) e em Savalli, Paula e Cysneiros (2006).

## 3.8 Estratégias na Seleção dos Modelos Multiníveis

As estratégias na seleção de modelos elípticos multiníveis adequados no ajuste de dados referentes a um determinado estudo podem ser as mesmas utilizadas nos modelos normais multiníveis, como definidas na Seção 2.7, ou seja, as estratégias "*Top-Down*" e "*Step-Up*". O uso de critérios de informação também pode ser atodado nas etapas de seleção, como por exemplo os critérios AIC e BIC citados anteriormente. Também similarmente ao caso Normal, as etapas no processo de seleção podem ser realizadas a partir da realização dos testes de hipóteses para efeitos fixos e componentes de variância e covariância, como ilustrados nas seções anteriores.

## 3.9 Análise de Resíduos

No contexto de modelos elípticos multiníveis, podemos também definir alguns tipos de resíduos, de maneira análoga ao caso dos modelos multiníveis com distribuição normal dos erros. Sendo assim, podemos definir o **resíduo marginal** como sendo

$$\hat{\boldsymbol{r}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}},\tag{3.14}$$

em que  $\hat{\boldsymbol{\beta}}$  é a estimativa de máxima verossimilhança de  $\boldsymbol{\beta}$ . Assim como no caso normal, quando a matriz  $\mathbf{X}$  é a própria matriz de variáveis preditoras do modelo elíptico multinível, o resíduo aqui considerado pode ser útil para verificar a suposição de linearidade entre  $\mathbf{E}(\mathbf{Y})$  e as variáveis explicativas presentes em  $\mathbf{X}$ . Podemos também utilizar um gráfico de  $\hat{\boldsymbol{r}}$  contra os valores destas variáveis. É esperado que os elementos do resíduo marginal variem aleatoriamente em torno de zero, sob a suposição de linearidade entre  $\mathbf{Y}$  e as variáveis pertencentes à matriz  $\mathbf{X}$ . Um gráfico destes resíduos contra seus respectivos índices pode revelar indícios da presença de observações aberrantes. Contudo, esta análise deve ser realizada observando também os pesos  $v(u_j)$ 's estimados para cada grupo ou "cluster", pois observações identificadas como aberrantes podem receber pesos pequenos em relação às demais observações na estimação dos parâmetros em  $\beta$  e, desta forma, podem não apresentar grande influência nas estimativas dos parâmetros. Como estes resíduos podem possuir diferentes variâncias, é conveninente utilizar os **resíduos marginais studentizados** 

$$\hat{r}_{ij}^* = \frac{\hat{r}_{ij}}{\widehat{\exp}(\hat{r}_{ij})},\tag{3.15}$$

com  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Aqui,  $\widehat{\operatorname{ep}}(\widehat{r}_{ij})$  é a estimativa do erro padrão do resíduo marginal  $\widehat{r}_{ij}$ . Para a obtenção da matriz de variância e covariância estimada de (3.14), iremos utilizar a notação matricial para o estimador de máxima verossimilhança obtida em Savalli (2005) no contexto de modelos elípticos mistos. Sendo assim, definindo  $v(u_j) = v_j$ , podemos expandir a matriz  $\sum_{j=1}^{J} v_j \mathbf{X}_j^{\ t} \mathbf{V}_j^{-1} \mathbf{X}_j$  da forma

$$\sum_{j=1}^{J} v_j \mathbf{X}_j^{t} \mathbf{V}_j^{-1} \mathbf{X}_j = v_1 \mathbf{X}_1^{t} \mathbf{V}_1^{-1} \mathbf{X}_1 + v_2 \mathbf{X}_2^{t} \mathbf{V}_2^{-1} \mathbf{X}_2 + \ldots + v_j \mathbf{X}_J^{t} \mathbf{V}_J^{-1} \mathbf{X}_J =$$

$$\begin{bmatrix} \mathbf{X}_{1}{}^{t} \, \mathbf{X}_{2}{}^{t} \dots \mathbf{X}_{J}{}^{t} \end{bmatrix} \begin{bmatrix} v_{1} \mathbf{V}_{1}{}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & v_{2} \mathbf{V}_{2}{}^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & v_{J} \mathbf{V}_{J}{}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{1} \\ \mathbf{X}_{2} \\ \vdots \\ \mathbf{X}_{J} \end{bmatrix},$$

e fazendo

$$\mathbf{M}^* = \operatorname{diag}[v_1 \mathbf{V}_1^{-1}, v_2 \mathbf{V}_2^{-1}, \dots, v_J \mathbf{V}_J^{-1}],$$

podemos escrever a estimativa de máxima verossimilhança de  $\beta$ , na convergência do processo iterativo, como sendo

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{M}^* \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M}^* \mathbf{Y}.$$

Note que esta expressão permite obter uma estimativa da matriz de variância e covariância de  $\hat{\beta}$ , dada por

$$\begin{split} \widehat{\operatorname{Var}}(\hat{\boldsymbol{\beta}}) &= \widehat{\operatorname{Var}}[(\mathbf{X}^t \hat{\mathbf{M}}^* \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{M}}^* \mathbf{Y}] \\ &= (\mathbf{X}^t \hat{\mathbf{M}}^* \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{M}}^* \widehat{\operatorname{Var}}(\mathbf{Y}) \hat{\mathbf{M}}^* \mathbf{X} (\mathbf{X}^t \hat{\mathbf{M}}^* \mathbf{X})^{-1}. \end{split}$$

Desta forma, temos que os resíduos marginais são dados por

$$\begin{split} \hat{\boldsymbol{r}} &= \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} &= \mathbf{Y} - \mathbf{X} (\mathbf{X}^t \mathbf{M}^* \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M}^* \mathbf{Y} \\ &= [\mathbf{I} - \mathbf{X} (\mathbf{X}^t \mathbf{M}^* \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M}^*] \mathbf{Y} \\ &= \mathbf{M}^{*-1} [\mathbf{M}^* - \mathbf{M}^* \mathbf{X} (\mathbf{X}^t \mathbf{M}^* \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M}^*] \mathbf{Y} \\ &= \mathbf{M}^{*-1} \mathbf{Q}^* \mathbf{Y}, \end{split}$$

em que  $\mathbf{Q}^* = \mathbf{M}^* - \mathbf{M}^* \mathbf{X} (\mathbf{X}^t \mathbf{M}^* \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M}^*$  e I é a matriz identidade de dimensão  $(\sum_{j=1}^J n_j) \times (\sum_{j=1}^J n_j)$ . Portanto, a matriz de variância e covariância estimada dos resíduos marginais quando  $\boldsymbol{\theta}$  é desconhecido é dada por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{r}}) = \widehat{\operatorname{Var}}(\widehat{\boldsymbol{M}}^{*-1}\widehat{\boldsymbol{Q}}^*\boldsymbol{Y}) = \widehat{\boldsymbol{M}}^{*-1}\widehat{\boldsymbol{Q}}^*\widehat{\operatorname{Var}}(\boldsymbol{Y})\widehat{\boldsymbol{Q}}^*\widehat{\boldsymbol{M}}^{*-1} = \\ \widehat{\boldsymbol{M}}^{*-1}\widehat{\boldsymbol{Q}}^*\widehat{\boldsymbol{V}}^*\widehat{\boldsymbol{Q}}^*\widehat{\boldsymbol{M}}^{*-1}, \qquad (3.16)$$

onde

$$\mathbf{V}^* = \begin{bmatrix} \alpha_1 \mathbf{V}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \alpha_2 \mathbf{V}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \alpha_J \mathbf{V}_J \end{bmatrix}$$

Logo, os elementos da diagonal principal de (3.16) podem ser utilizados na obtenção dos valores dados em (3.15).

Na análise de resíduos dos modelos elípticos multiníveis também pode ser utilizado o **resíduo condicional** 

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{b}},$$

em que  $\hat{\boldsymbol{\beta}}$  é a estimativa de máxima verossimilhança de  $\boldsymbol{\beta}$  e  $\hat{\boldsymbol{b}}$  é preditor linear de  $\boldsymbol{b}$ . O

resíduo  $\hat{\boldsymbol{\epsilon}}$  pode ser utilizado para identificar observações aberrantes presentes nos dados, através do gráfico destes resíduos contra seus respectivos índices. De forma análoga ao caso do resíduo marginal, a identificação de observações aberrantes utilizando o resíduo  $\hat{\boldsymbol{\epsilon}}$  deve ser realizada observando-se também os pesos  $v(u_j)$ 's estimados. Também similarmente ao caso dos resíduos marginais, os resíduos condicionais podem apresentar diferentes variâncias. Sendo assim, convém utilizar os **resíduos condicionais studentizados** 

$$\hat{\epsilon}_{ij}^* = \frac{\hat{\epsilon}_{ij}}{\widehat{\exp}(\hat{\epsilon}_{ij})},\tag{3.17}$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, em que  $\widehat{ep}(\hat{\epsilon}_{ij})$  é a estimativa do erro padrão do resíduo condicional  $\hat{\epsilon}_{ij}$ . Podemos reescrever o resíduo condicional como sendo

$$\begin{aligned} \hat{\boldsymbol{\epsilon}} &= \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{b}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^{t}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= [\mathbf{I} - \mathbf{Z}\boldsymbol{\Psi}\mathbf{Z}^{t}\mathbf{V}^{-1}](\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{A}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \end{aligned}$$

em que  $\mathbf{A} = \mathbf{I} - \mathbf{Z} \Psi \mathbf{Z}^t \mathbf{V}^{-1}$ . Desta forma, podemos obter a matriz de variância e covariância estimada de  $\hat{\boldsymbol{\epsilon}}$ , quando  $\boldsymbol{\theta}$  é desconhecido, fazendo

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{\epsilon}}) = \widehat{\operatorname{Var}}[\hat{\mathbf{A}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})] = \hat{\mathbf{A}}\widehat{\operatorname{Var}}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}]\hat{\mathbf{A}}^{t} = \\ \hat{\mathbf{A}}\widehat{\mathbf{M}}^{*-1}\hat{\mathbf{Q}}^{*}\hat{\mathbf{V}}^{*}\hat{\mathbf{Q}}^{*}\hat{\mathbf{M}}^{*-1}\hat{\mathbf{A}}^{t}.$$
(3.18)

Portanto, os termos da diagonal principal de (3.18) são utilizados na obtenção dos resíduos condicionais studentizados dados em (3.17).

Assim como no caso dos modelos normais multiníveis, os **preditores lineares** também podem ser utilizados na análise de resíduos. Considerando

$$\hat{\boldsymbol{b}} = \boldsymbol{\Psi} \mathbf{Z}^t \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}),$$

temos que sua matriz de variância e covariância estimada é dada por

$$\widehat{\operatorname{Var}}(\hat{\boldsymbol{b}}) = \widehat{\operatorname{Var}}[\hat{\boldsymbol{\Psi}}\mathbf{Z}^{t}\widehat{\mathbf{V}}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})] 
= \hat{\boldsymbol{\Psi}}\mathbf{Z}^{t}\widehat{\mathbf{V}}^{-1}\widehat{\operatorname{Var}}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})\widehat{\mathbf{V}}^{-1}\mathbf{Z}\hat{\boldsymbol{\Psi}}$$

$$= \hat{\boldsymbol{\Psi}}\mathbf{Z}^{t}\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{M}}^{*-1}\widehat{\mathbf{Q}}^{*}\widehat{\mathbf{V}}^{*}\widehat{\mathbf{Q}}^{*}\widehat{\mathbf{M}}^{*-1}\widehat{\mathbf{V}}^{-1}\mathbf{Z}\hat{\boldsymbol{\Psi}}.$$
(3.19)

Portanto, os elementos da diagonal principal de (3.19) podem ser utilizados na obtenção dos **preditores lineares studentizados** 

$$\hat{b}_{ij}^* = \frac{\hat{b}_{ij}}{\widehat{\exp}(\hat{b}_{ij})},$$

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J, onde  $\widehat{\operatorname{ep}}(\widehat{b}_{ij})$  é a estimativa do erro padrão de  $\widehat{b}_{ij}$ . O gráfico dos preditores lineares studentizados contra seus respectivos índices pode detectar observações aberrantes com relação aos efeitos aleatórios em determinados grupos ou "clusters".

De acordo com Lange, Little e Taylor (1989), a **distância de Mahalanobis modi**ficada

$$\frac{u_j}{n_j} = \frac{1}{n_j} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})^t \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta}),$$

para o caso da distribuição t-Student multivariada segue uma distribuição  $F \operatorname{com} n_j$  graus de liberdade no numerador e  $\nu_j$  graus de liberdade no denominador. Esta quantidade pode ser utilizada, portanto, na identificação de grupos ou "clusters" aberrantes. Um gráfico dos valores estimados das distâncias de Mahalanobis modificadas contra seus respectivos índices pode ser utilizado para tal identificação. Também neste caso devem ser consideradas as estimativas dos pesos  $v(u_j)$ 's. Contudo, vale reforçar o fato de que, no caso da distribuição t-Student, o valor de  $v(u_j)$  é inversamente proporcional ao valor de  $u_j$  e, sendo assim, altos valores de  $u_j$  naturalmente implicam em valores mais baixos de  $v(u_j)$ .

Também analogamente ao caso normal podem ser utilizadas, ainda, as **distâncias de** Mahalanobis dos preditores lineares

$$\zeta_j = \hat{\boldsymbol{b}}_j^t \widehat{\operatorname{Var}}(\hat{\boldsymbol{b}}_j) \hat{\boldsymbol{b}}_j,$$

para j = 1, 2, ..., J na identificação de grupos ou "clusters" aberrantes, no que diz respeito aos efeitos aleatórios. Uma proposta é a utilização de um gráfico dos  $\zeta_j$ 's contra seus respectivos índices para a detecção de grupos aberrantes com relação aos seus respectivos efeitos aleatórios.

Assim como no caso de modelos normais multiníveis, os resíduos marginais, condicionais e preditores lineares studentizados podem ser utilizados na construção de bandas de confiança simuladas. Os passos necessários para a construção destas bandas de confiança são análogos àqueles descritos na Seção 2.8.

#### 3.10 Aplicações

Nesta seção, as aplicações discutidas no Capítulo 2 serão reavaliadas utilizando os modelos elípticos multiníveis. Mais especificamente, será utilizada a distribuição t-Student multivariada. Esta distribuição tem a propriedade de apresentar, em geral, caudas mais pesadas do que as caudas da distribuição normal. O objetivo de sua utilização nas aplicações é uma melhor acomodação de observações destacadas como possivelmente aberrantes. Além disso, a utilização da distribuição t-Student multivariada nos ajustes permite a estimação dos parâmetros de forma robusta, ou seja, atribuindo menores pesos para grupos ou "clusters" considerados aberrantes em relação aos demais contidos nos dados.

#### 3.10.1 Aplicação 1 - Experimento com Ratos

#### Modelagem

Para a modelagem deste aplicação, será utilizada a mesma estrutura do modelo discutido na Seção 2.9.1, ou seja

> Equação do Nível 1: WEIGHT<sub>ij</sub> =  $\gamma_{1j} + \gamma_{2j}$ SEX1<sub>ij</sub> +  $\epsilon_{ij}$ ; Equações do Nível 2:  $\gamma_{1j} = \beta_{11} + \beta_{12}$ TREAT1<sub>j</sub> +  $\beta_{13}$ TREAT2<sub>j</sub> +  $\beta_{14}$ LITSIZE<sub>j</sub> +  $b_{1j}$ ;  $\gamma_{2j} = \beta_{21}$ , (3.20)

com  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., 27, em que  $n_j$  é o número de filhotes pertencentes à *j*-ésima ninhada, num total de J = 27 ninhadas. Neste caso a suposição de distribuição conjunta dos erros e dos efeitos aleatórios seria a de que

$$\begin{bmatrix} \mathbf{Y}_j \\ \mathbf{b}_j \\ \mathbf{\epsilon}_j \end{bmatrix} \sim t_{n_j+1+n_j} \left( \begin{bmatrix} \mathbf{X}_j \boldsymbol{\beta} \\ 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_j \boldsymbol{\Psi}_j \mathbf{Z}_j^t + \mathbf{R}_j & \mathbf{Z}_j \boldsymbol{\Psi}_j & \mathbf{R}_j \\ \boldsymbol{\Psi}_j \mathbf{Z}_j^t & \boldsymbol{\Psi}_j & \mathbf{0} \\ \mathbf{R}_j & \mathbf{0} & \mathbf{R}_j \end{bmatrix}, \boldsymbol{\nu} \right),$$

em que  $\boldsymbol{\nu} = (\nu, \nu, \nu)^t$ ,  $R_j = \sigma_C^2 \mathbf{I}$  se a *j*-ésima ninhada foi submetida ao tratamento controle ou  $R_j = \sigma_{HL}^2 \mathbf{I}$  se a *j*-ésima ninhada foi submetida à dosagem alta ou baixa do composto experimental. Em ambos os casos temos  $\mathbf{I}$  a matriz identidade de dimensão  $n_j \times n_j$ . Contudo, para a obtenção da distribuição *t*-Student no modelo marginal foi considerada a formulação hierárquica

$$\begin{aligned} \mathbf{Y}_{j} | (\boldsymbol{b}_{j}, \tau_{j})^{t} &\sim N_{n_{j}} (\mathbf{X}_{j} \boldsymbol{\beta} + \mathbf{Z}_{j} \boldsymbol{b}_{j}, \frac{1}{\tau_{j}} \mathbf{R}_{j}); \\ \boldsymbol{b}_{j} | \tau_{j} &\sim N(0, \frac{1}{\tau_{j}} \boldsymbol{\Psi}_{j}); \\ \tau_{j} &\sim Gama(\frac{\nu}{2}, \frac{\nu}{2}), \end{aligned}$$
(3.21)

pois esta implica que

$$\mathbf{Y}_{j} \sim t_{n_{j}}(\mathbf{X}_{j}\boldsymbol{\beta}, \mathbf{Z}_{j}\boldsymbol{\Psi}_{j}\mathbf{Z}_{j}^{t} + \mathbf{R}_{j}, \boldsymbol{\nu}).$$

Através do algoritmo ECME descrito na Seção 3.3.1 foram obtidas as estimativas de máxima verossimilhança do parâmetros a partir da estrutura hierárquica (3.21). Estas estimativas são ilustradas na Tabela 3.1.

Considerando um nível de significância de 5% todos os parâmetros em  $\beta$  são considerados marginalmente significativos. A interpretação destes parâmetros é realizada de forma similar àquela obtida a partir dos ajustes efetuados na Seção 2.9.1. Contudo, em relação aos parâmetros de variância e covariância, vale notar que no caso dos modelos elípticos e, portanto, no modelo aqui considerado, as variâncias dos erros são apenas proporcionais a estas quantidades.

A Tabela 3.2 ilustra as estimativas de máxima verossimilhança dos parâmetros para os modelos normal e t-Student. As estimativas dos parâmetros em  $\beta$  sob os dois modelos

		ł		```	. /
Parâmetro	Estimativa	Erro padrão	G.l.	Valor z	Valor p
$\beta_{11}$	8,1630	0,2507	294	32,5554	< 0,001
$\beta_{12}$	-0,9044	0,1652	23	-5,4736	< 0,001
$\beta_{13}$	-0,4443	0,1360	23	-3,2668	0,0011
$\beta_{14}$	-0,1201	0,0171	23	-7,0111	< 0,001
$\beta_{21}$	-0,3005	0,0375	294	-8,0095	< 0,001
$\sigma_N^2$	0,0734	0,0256	_	_	_
$\sigma_C^2$	0,1170	0,0287	—	_	—
$\sigma^2_{HL}$	0,0782	0,0155	—	_	—
u	5,1424	—	_	_	_

Tabela 3.1: Estimativas dos parâmetros do modelo (3.20).

considerados apresentaram pouca diferença. Os valores estimados dos erros padrão das estimativas dos parâmetros no vetor  $\beta$  no modelo t-Student apresentam valores menores que estas mesmas quantidades obtidas a partir do modelo normal, ou seja, as estimativas obtidas a partir do modelo t-Student são aparentemente mais precisas em relação àquelas obtidas no modelo com erros normais. As estimativas dos parâmetros de variância e covariância são menores no caso do modelo t-Student.

	No	rmal	t-St	udent
Parâmetro	Estimativa	Erro padrão	Estimativa	Erro padrão
$\beta_{11}$	8,3280	0,2593	8,1630	0,2507
$\beta_{12}$	-0,8648	0,1715	-0,9044	0,1652
$\beta_{13}$	-0,4344	0,1424	-0,4443	0,1360
$\beta_{14}$	-0,1307	0,0175	-0,1201	0,0171
$\beta_{21}$	-0,3419	0,0422	-0,3005	0,0375
$\sigma_N^2$	0,0832	0,0265	0,0734	0,0256
$\sigma_C^2$	0,2639	0,0339	0, 1170	0,0287
$\sigma^2_{HL}$	0,0913	0,0098	0,0782	0,0155

Tabela 3.2: Estimativas dos parâmetros sob os modelos normal e t-Student ajustados aos dados do experimento com ratos.

Na Tabela 3.3 temos as estimativas dos coeficientes de correlação intraclasse obtidas a partir dos modelo t-Student aqui ajustado e a partir do modelo com distribuição Normal ajustado pelo método de Máxima Verossimilhança na Seção 2.9.1. Verifica-se que as estimativas dos ICC's são maiores sob o modelo t-Student, ou seja, este modelo capta uma maior similaridade entre os ratos provenientes de uma mesma ninhada que o modelo Normal, para os três tratamentos considerados no estudo. Contudo, assim como no modelo Normal, os valores dos coeficientes de correlação intraclasse estimados a partir do modelo t-Student também podem ser considerados baixos.

Tabela 3.3: ICCs estimados dos modelos normal e t-Student ajustados aos dados do experimento com ratos.

ICC	Normal	<i>t</i> -Student
$\hat{ICC}_C$	0,2397	0,3855
$\hat{ICC}_{HL}$	0,4768	0,4842

Os 20 menores pesos estimados para as ninhadas encontram-se na Tabela 3.4. Temos os 5 menores pesos atribuídos às ninhadas 6, 23, 12, 3 e 25 respectivamente, ou seja, as observações das variáveis referentes a estas ninhadas foram aquelas que exerceram as menores influências nas estimativas dos parâmetros.

Tabela 3.4: 20 menores pesos estimados pelo modelo t-Student ajustado aos dados do experimento com ratos com as respectivas ninhadas.

Peso	Ninhada	Peso	Ninhada	Peso	Ninhada	Peso	Ninhada
0,7440	6	1,8144	5	2,5039	18	3,1610	17
0,7799	23	2,2607	26	2,6354	13	3,1653	10
0,9555	12	2,2906	27	2,6406	24	3, 1919	8
1,5496	3	2,4109	19	2,7224	1	3,2033	21
1,7884	25	2,4976	22	3,0512	16	3,2529	4

#### Análise de Resíduos

De maneira análoga ao caso normal, alguns gráficos podem ser construídos para a realização de uma análise de resíduos obtidos a partir do ajuste do modelo aqui considerado. Na Figura 3.1 temos os gráficos dos resíduos marginais e condicionais studentizados contra os pesos atribuídos a cada ninhada no processo de estimação dos parâmetros. Note que o indivíduo 66 pertencente à ninhada 6 apresenta maiores valores (em valores absolutos) destes dois resíduos em relação aos demais filhotes. Este mesmo ponto foi detectado como aberrante na Seção 2.9.1. Pode ser observado, contudo, que o peso referente à ninhada 6 foi o menor dentre todas as ninhadas. Sendo assim, o rato 66, por pertencer à ninhada 6, não é considerado como aberrante, pois esta ninhada exerceu pouca influência nos valores da estimativas dos parâmetros do modelo.


Figura 3.1: Gráficos de dispersão bidimensional entre os pesos estimados e os resíduos marginais e condicionais studentizados do modelo *t*-Student ajustado aos dados do experimento com ratos.

Pelo gráfico da Figura 3.2 poderíamos identificar a ninhada 6 como aberrante por apresentar alto valor da estimativa da distância de Mahalanobis modificada em relação ao respectivo percentil da distribuição F representado pela extremidade superior do segmento de reta associado à referida distância. Contudo, a conclusão obtida a partir da Tabela 3.4 foi a de que observações pertencentes a esta ninhada exerceram baixa influência nas estimativas dos parâmetros.

Considerando a Figura 3.3 temos a ninhada 9 detectada como atípica em relação às demais, no que diz respeito aos efeitos aleatórios, apesar do valor do preditor linear studentizado para esta ninhada não ser muito discrepante em relação aos valores para as demais ninhadas.

Através do gráfico de percentis com envelope gerado para os resíduos marginais studentizados da Figura 3.4 podemos concluir que a distribuição t-Student está aparentemente bem especificada para os erros do modelo (3.20), pois a grande maioria dos pontos encontra-se disposta dentro do envelope simulado. Ainda considerando a Figura 3.4 temos que, de acordo com o gráfico de percentis com envelope gerado para os resíduos condicionais studentizados, não há afastamento da suposição de que os erros seguem uma distribuição t-Student.

Visualizando a Figura 3.5 nota-se que alguns pontos estão dispostos fora do envelope,



Figura 3.2: Estimativas das distâncias de Mahalanobis modificadas para as ninhadas e percentis 0,95 da distribuição F referentes ao modelo t-Student ajustado aos dados do experimento com ratos.

contudo não há sério afastamento da suposição de que os efeitos aleatórios seguem uma distribuição t-Student.



Figura 3.3: Preditores lineares studentizados do modelo t-Student ajustado aos dados do experimento com ratos.



Figura 3.4: Gráficos normais de probabilidades com envelopes gerados para os resíduos marginais e condicionais studentizados do modelo t-Student ajustado aos dados do experimento com ratos.



Figura 3.5: Gráfico normal de probabilidades com envelope gerado para os preditores lineares studentizados do modelo t-Student ajustado aos dados do experimento com ratos.

#### 3.10.2 Aplicação 2 - Estudo Educacional

#### Modelagem

Assim como no caso do experimento com ratos, para a modelagem desta aplicação será utilizada a mesma estrutura do modelo normal multinível discutido na seção 2.9.1, ou seja

#### Equação do Nível 1:

MATHGAIN<sub>*ijk*</sub> =  $\gamma_{1j|k} + \gamma_{2j|k}$ MATHKIND<sub>*ijk*</sub> +  $\gamma_{3j|k}$ MINORITY<sub>*ijk*</sub> +  $\gamma_{4j|k}$ SES<sub>*ijk*</sub> +  $\epsilon_{ijk}$ ; Equações do Nível 2:

$$\gamma_{1j|k} = \delta_{11k} + \xi_{1j|k};$$
  

$$\gamma_{2j|k} = \delta_{21k};$$
  

$$\gamma_{3j|k} = \delta_{31k};$$
  

$$\gamma_{4j|k} = \delta_{41k};$$

Equações do Nível 3:

$$\delta_{11k} = \beta_{111} + \vartheta_{11k}; \delta_{21k} = \beta_{211}; \delta_{31k} = \beta_{311}; \delta_{41k} = \beta_{411},$$
(3.22)

em que  $i = 1, 2, ..., n_{jk}$ ,  $j = 1, 2, ..., m_k$  e k = 1, 2, ..., 107. Para a estimação deste modelo poderíamos admitir a distribuição conjunta

$$\begin{bmatrix} \mathbf{Y}_k \\ \mathbf{b}_k \\ \mathbf{\epsilon}_k \end{bmatrix} \sim t_{\sum_{j=1}^{m_k} n_{jk} + (1+m_k) + \sum_{j=1}^{m_k} n_{jk}} \left( \begin{bmatrix} \mathbf{X}_k \boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}_k \boldsymbol{\Psi}_k \mathbf{Z}_k^t + \mathbf{R}_k & \mathbf{Z}_k \boldsymbol{\Psi}_k & \mathbf{R}_k \\ \boldsymbol{\Psi}_k \mathbf{Z}_k^t & \boldsymbol{\Psi}_k & \mathbf{0} \\ \mathbf{R}_k & \mathbf{0} & \mathbf{R}_k \end{bmatrix}, \boldsymbol{\nu} \right).$$

Contudo, o processo de estimação dos parâmetros foi realizado a partir da formulação hierárquica

$$\begin{aligned} \mathbf{Y}_k | (\boldsymbol{b}_k^{\ t}, \tau_k)^t &\sim & N_{\sum_{j=1}^{m_k} n_{jk}} (\mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{b}_k, \frac{1}{\tau_j} \mathbf{R}_k); \\ \boldsymbol{b}_k | \tau_k &\sim & N_{1+m_k} (\mathbf{0}, \frac{1}{\tau_k} \boldsymbol{\Psi}_k); \\ & \tau_k &\sim & Gama(\frac{\nu}{2}, \frac{\nu}{2}), \end{aligned}$$

MANGHI, R. F.

	1					
Parâmetro	Estimativa	Erro padrão	G.l.	Valor z	Valor p	
$\beta_{111}$	280,6996	10,6517	875	26,3525	< 0,001	
$\beta_{211}$	-0,4679	0,0219	875	-21,4082	< 0,001	
$\beta_{311}$	-7,7948	2,3092	875	-3,3755	0,0007	
$\beta_{411}$	5,0886	1,2183	875	4,1768	< 0,001	
$\sigma_E^2$	71,95597	25,8411	_	_	_	
$\sigma_C^2$	84,15703	28,0110	_	—	—	
$\sigma^2$	661,4466	38,2271	_	—	_	
u	19,7084	—	—	—	—	

Tabela 3.5: Estimativas dos parâmetros do modelo (3.22).

implicando em

$$\mathbf{Y}_k \sim t_{\sum_{j=1}^{m_k} n_{jk}} (\mathbf{X}_k \boldsymbol{eta}, \mathbf{Z}_k {\mathbf{\Psi}_k} {\mathbf{Z}_k}^t + \mathbf{R}_k, 
u)$$

O algoritmo ECME também foi utilizado para a estimação dos parâmetros neste modelo. Estas estimativas podem ser observadas na Tabela 3.5. Dado um nível de significância de 5% todos os parâmetros em  $\beta$  são considerados marginalmente significativos. A interpretação destes parâmetros é similar àquela realizada na Seção 2.9.2. Em relação aos parâmetros de variância e covariância, novamente é necessário lembrar que as variâncias dos erros neste modelo são apenas proporcionais a estas quantidades.

A Tabela 3.6 ilustra as estimativas de máxima verossimilhança dos parâmetros para os modelos normal e t-Student. Os valores estimados dos erros padrão das estimativas dos parâmetros no modelo t-Student apresentam valores menores do que estas mesmas quantidades obtidas a partir do modelo normal. As estimativas dos parâmetros de variância e covariância também são menores no caso do modelo t-Student, com exceção das estimativas do parâmetro  $\sigma_C^2$ , em que a estimativa obtida sob o modelo normal apresentou valor menor que esta mesma estimativa sob o modelo t-Student.

Na Tabela 3.7 podemos observar as estimativas dos coeficientes de correlação intraclasse obtidas a partir dos modelos t-Student e normal ajustados pelo método de máxima verossimilhança na Seção 2.9.2. Considerando alunos de uma mesma escola, temos que o coeficiente de correlação intraclasse é menor sob o modelo normal. Para o ICC definido considerando alunos de uma mesma classe, temos o valor estimado sob o modelo normal também menor do que o valor estimado sob o modelo t-Student, ou seja, a homogeneidade

obrado badronoman						
	No	rmal	t-Student			
Parâmetro	Estimativa	Erro padrão	Estimativa	Erro padrão		
$\beta_{111}$	282,3436	10,8385	280,6996	10,6517		
$\beta_{211}$	-0,4701	0,0222	-0,4679	0,0219		
$\beta_{311}$	-8,2850	2,3343	-7,7948	2,3092		
$\beta_{411}$	5,3606	1,2406	5,0886	1,2183		
$\sigma_E^2$	72,5109	26,4229	71,9560	25,8411		
$\sigma_C^2$	82,7348	35,5971	84,1570	28,0110		
$\sigma^2$	732,9298	42,4142	661,4466	38,2271		

Tabela 3.6: Estimativas dos parâmetros sob os modelos normal e t-Student ajustados aos dados do estudo educacional.

dos ganhos nos escores dos alunos considerando a mesma escola ou a mesma classe numa determinada escola é maior segundo o modelo t-Student. Contudo, assim como no modelo normal, os valores dos coeficientes de correlação intraclasse estimados a partir do modelo t-Student também podem ser considerados baixos.

 Tabela 3.7: ICCs estimados dos modelo normal e t-Student ajustados aos dados do estudo

 educacional.

ICC	Normal	t-Student
$\hat{\mathrm{ICC}}_E$	0,0816	0,0880
$\hat{\mathrm{ICC}}_C$	0,1748	0, 1909

A Tabela 3.8 ilustra os 20 menores pesos estimados para as escolas. Por exemplo, temos os 5 menores pesos atribuídos às escolas 4, 62, 47, 40 e 85, respectivamente, ou seja, as observações das variáveis referentes a estas escolas foram aquelas que exerceram a menor influência nas estimativas dos parâmetros.

Tabela 3.8: 20 menores pesos estimados pelo modelo *t*-Student ajustado para os dados do estudo educacional com as respectivas escolas.

Peso	Escola	Peso	Escola	Peso	Escola	Peso	Escola
0,3988	4	0,6944	70	0,7677	27	0,8250	90
0,5557	62	0,7179	102	0,7859	60	0,8517	31
0,5711	47	0,7227	8	0,8012	58	0,8553	24
0,6384	40	0,7292	49	0,8014	86	0,8581	2
0,6653	85	0,7583	75	0,8133	53	0,8601	97

#### Análise de Resíduos

Nesta seção, alguns gráficos são considerados na análise dos resíduos obtidos a partir do ajuste do modelo (3.22). Na Figura 3.6 temos os gráficos dos resíduos marginais e condicionais studentizados contra os pesos atribuídos a cada escola no processo de estimação dos parâmetros. Note que o aluno 41 pertencente à classe 179 da escola 4 apresenta maior valor (em valor absoluto) destes dois resíduos em relação aos demais. Este mesmo ponto foi detectado como aberrante na Seção 2.9.2. Pode ser observado, contudo, que o peso referente à escola 4 foi o menor dentre todas as escolas. Desta forma o aluno citado não é considerado aberrante pois a escola à qual ele pertence exerceu pouca influência nos valores da estimativas dos parâmetros do modelo.



Figura 3.6: Gráficos de dispersão bidimensional entre os pesos estimados e os resíduos marginais e condicionais studentizados do modelo *t*-Student ajustado aos dados do estudo educacional.

Pelo gráfico da Figura 3.7 poderíamos identificar as escolas 4, 40, 47 e 62 como aberrantes por apresentarem altos valores das estimativas da distância de Mahalanobis modificadas em relação aos respectivos percentis da distribuição F. Através da Tabela 3.8 podemos observar que observações pertencentes a estas escolas são aquelas que menos influenciam nas estimativas dos parâmetros.

A partir da Figura 3.8 temos as classes 179 pertencente à escola 4, 9 pertencente à escola 40 e 258 pertencente à escola 31 detectadas como atípicas em relação às demais



Figura 3.7: Estimativas das distâncias de Mahalanobis modificadas para as escolas e os respectivos percentis 0,95 da distribuição F referentes ao modelo t-Student ajustado aos dados do estudo educacional.

classes, no que diz respeito aos efeitos aleatórios das classes para todas as escolas.



Figura 3.8: Preditores lineares studentizados do modelo t-Student ajustado aos dados do estudo educacional.

Considerando os gráficos normais de probabilidades com envelopes gerados para os resíduos marginais e condicionais studentizados apresentados na Figura 3.9, podemos concluir que não há importantes afastamentos da distribuição *t*-Student para os erros do modelo (3.20), pois a maioria dos pontos encontra-se disposta dentro do envelope simulado. O fato de alguns pontos encontrarem-se fora do envelopes gerados nos extremos superiores

dos respectivos gráficos pode ser um indício de uma pequena assimetria à direita para a distribuição dos erros. Este indício já foi observado a partir do ajuste do modelo sob erros normais e na análise descritiva.



Figura 3.9: Gráficos normais de probabilidades com envelopes gerados para os resíduos marginais e condicionais studentizados do modelo *t*-Student ajustado aos dados do estudo educacional.

Segundo a Figura 3.10 os efeitos aleatórios seguem uma distribuição t-Student multivariada, pois os pontos do gráfico encontram-se dispostos dentro do envelope simulado.

De acordo com os resultados obtidos para as duas aplicações sob o modelo t-Student multinível considerando dados com dois e três níveis de hierarquia, podemos concluir que as estimativas dos parâmetros foram obtidas de forma robusta, ou seja, observações e grupos considerados atípicos exerceram pouca influência nas estimativas. Em geral, as estimativas dos erros padrão de  $\hat{\beta}$  apresentaram valores menores em relação a aqueles obtidos a partir dos ajustes dos modelos normais multiníveis, implicando em estimativas aparentemente mais precisas dos parâmetros em  $\beta$ . Estes resultados comprovam a eficiência no uso de modelos multiníveis com distribuições apresentando caudas mais pesadas em relação à distribuição normal. Contudo, outras distribuições pertencentes à classe elíptica podem ser utilizadas.



Figura 3.10: Gráfico normal de probabilidades com envelope gerado para os preditores lineares studentizados do modelo t-Student ajustado aos dados do estudo educacional.

### Capítulo 4

## Conclusões

O presente trabalho consiste na revisão da literatura já existente acerca dos modelos multiníveis com distribuição normal dos erros e efeitos aleatórios e na proposta de uma classe de modelos alternativa aos modelos normais multiníveis, denominada modelos elípticos multiníveis. Esta nova proposta permite o uso de distribuições pertencentes à classe elíptica para os erros e efeitos aleatórios do modelo. A classe de distribuições elípticas engloba totas as distribuições simétricas e contínuas, sendo a normal um caso particular. Desta forma, foi mostrado que o modelo elíptico multinível pode ser utilizado com eficiência na modelagem de dados com estrutura de hierarquia. Através de resultados já obtidos nos modelos elípticos mistos, foram desenvolvidos o processo de estimação dos parâmetros via máxima verossimilhança, a realização de testes de hipóteses e a análise de resíduos na verificação das suposições assumidas para os erros.

Também foi realizada a comparação dos resultados dos ajustes nos modelos normais multiníveis e nos modelos elípticos multiníveis com distribuição t-Student através de duas aplicações, sendo a primeira referente a um experimento cujo conjunto de dados apresentou dois níveis de hierarquia e a segunda inerente a um estudo cujos dados apresentaram três níveis de hierarquia. Pôde ser comprovado que o modelo elíptico multinível com distribuição t-Student tem a capacidade de melhor acomodação de observações aberrantes, pois a estimação dos parâmetros neste modelo é realizada de maneira robusta, ou seja, grupos ou "clusters" atípicos exercem menor influência nas estimativas dos parâmetros na classe de modelos proposta neste trabalho.

Desta forma, conclui-se que os modelos elípticos multiníveis compõem uma impor-

tante e abrangente classe de modelos, cujas propriedades e características podem ser amplamente estudadas a partir dos resultados aqui obtidos. Os bancos de dados utilizados nas aplicações aqui desenvolvidas podem ser acessados a partir do endereço eletrônico http://www-personal.umich.edu/~bwest/almmussp.html. As rotinas computacionais utilizadas neste trabalho foram desenvolvidas no software livre R em sua versão 2.11.1 e podem ser solicidadas através do email robertof23@gmail.com.

Como propostas para possíveis estudos podemos sugerir, entre outras:

- O desenvolvimento de técnicas de diagnóstico como a análise de influência local via curvatura normal na classe de modelos proposta;
- A obtenção do método de estimação de máxima verossimilhança restrita para os parâmetros de variância e covariância nos modelos elípticos multiníveis;
- Estudos de simulação para verificar o comportamento dos estimadores dos parâmetros em pequenas amostras;
- Aplicações dos modelos elípticos em outras áreas do conhecimento, inclusive com o uso de outras distribuições de probabilidade;
- Proposta do estudo de modelos multiníveis com distribuições assimétricas para os erros ou ainda o desenvolvimento de modelos multiníveis com estrutura semiparamétrica.

### Apêndice A

# Notação Matricial dos Modelos Multiníveis

### A.1 Modelos Multiníveis com 2 Níveis

Considerando P variáveis explicativas associadas ao nível 1 e  $\sum_{p=1}^{P} Q_p$  variáveis preditoras associadas ao nível 2, em que  $Q_p$  é o número de variáveis preditoras associadas ao p-ésimo parâmetro do nível 1, temos as equações gerais para o modelo multinível com dois níveis dadas a seguir.

Nível 1: 
$$Y_{ij} = \sum_{p=1}^{P} \gamma_{pj} Z_{pij} + \epsilon_{ij};$$
  
Nível 2:  $\gamma_{pj} = \sum_{q=1}^{Q_p} \beta_{pq} T_{pqj} + b_{pj},$ 
(A.1)

para  $i = 1, 2, ..., n_j$  e j = 1, 2, ..., J. Aqui,  $n_j$  é o número de indivíduos pertencentes ao *j*-ésimo "cluster" e J é o número total de "clusters" considerado para o modelo. Em (A.1), temos que:

 $Y_{ij}$ é o valor da variável resposta para <br/>o $i\mathchar`$ ésimo indivíduo pertencente a<br/>o $j\mathchar`$ ésimo "cluster";

- $Z_{pij}$  é o valor da *p*-ésima variável preditora associada ao *i*-ésimo indivíduo do *j*-ésimo "cluster";
- $\gamma_{pj}$ é o parâmetro associado à variável  $\mathbf{Z}_p;$
- $\epsilon_{ij}$ é o erro associado ao valor da variável respo<br/>sta do nível 1 $Y_{ij};$

 $T_{pqj}$  é o valor da q-ésima variável associada ao parâmetro  $\gamma_{pj}$ ;

 $\beta_{pq}$ é o parâmetro associado à referida variável do nível 2;

 $b_{pj}$ é o erro associado ao parâmetro  $\gamma_{pj}.$ 

Fazendo

$$\begin{split} \mathbf{Y}_{j} &= (Y_{1j}, Y_{2j}, \dots Y_{n_{j}j})^{t}_{n_{j} \times 1}; \quad \mathbf{Z}_{j} = \begin{bmatrix} Z_{11j} & Z_{21j} & \dots & Z_{P1j} \\ Z_{12j} & Z_{22j} & \dots & Z_{P2j} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1n_{j}j} & Z_{2n_{j}j} & \dots & Z_{Pn_{j}j} \end{bmatrix}_{n_{j} \times P} \\ \boldsymbol{\gamma}_{j} &= (\gamma_{1j}, \gamma_{2j}, \dots, \gamma_{Pj})^{t}_{P \times 1}, \quad \boldsymbol{\epsilon}_{j} = (\epsilon_{1j}, \epsilon_{2j}, \dots, \epsilon_{n_{j}j})^{t}_{n_{j} \times 1}, \\ \mathbf{T}_{j} &= \begin{bmatrix} \mathbf{T}_{1j} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2j} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_{Pj} \end{bmatrix}_{P \times \sum_{p=1}^{P} Q_{p}} \end{split}$$

em que  $\mathbf{T}_{pj} = (\mathbf{T}_{p1j}, \mathbf{T}_{p2j}, \dots, \mathbf{T}_{pQ_pj})_{1 \times Q_p}$  e **0** é um vetor (linha ou coluna) ou matriz de zeros, convenientemente adotada em cada caso;

$$\boldsymbol{\beta} = (\beta_{11}, \beta_{12}, \dots, \beta_{1Q_1}, \beta_{21}, \dots, \beta_{2Q_2}, \dots, \beta_{PQ_P})^t (\sum_{p=1}^P Q_p) \times 1$$

e  $\boldsymbol{b}_j = (b_{1j}, b_{2j}, \dots, b_{Pj})^t_{P \times 1}$ , obtemos as equações referentes aos níveis 1 e 2 dadas por

$$\mathbf{Y}_j = \mathbf{Z}_j \boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j; \tag{A.2}$$

$$\boldsymbol{\gamma}_{j} = \mathbf{T}_{j}\boldsymbol{\beta} + \boldsymbol{b}_{j}. \tag{A.3}$$

Para j = 1, 2, ..., J. Substituindo (A.3) em (A.2), temos

$$\mathbf{Y}_j = \mathbf{Z}_j \left[ \mathbf{T}_j \boldsymbol{\beta} + \boldsymbol{b}_j \right] + \boldsymbol{\epsilon}_j = \mathbf{Z}_j \mathbf{T}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j + \boldsymbol{\epsilon}_j.$$

Assim ,temos que

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \boldsymbol{b}_j + \boldsymbol{\epsilon}_j,$$

MANGHI, R. F.

para j = 1, 2, ..., J, em que  $\mathbf{X}_j = \mathbf{Z}_j \mathbf{T}_j$ . Caso algum parâmetro  $\gamma_{pj}$  tenha sido considerado fixo (ou seja,  $b_{pj} = 0$ ), é mais conveniente representar o modelo por

$$\mathbf{Y}_{j} = \mathbf{X}_{j} \boldsymbol{eta} + \mathbf{Z}_{j}^{*} \boldsymbol{b}_{j}^{*} + \boldsymbol{\epsilon}_{j},$$

em que  $\mathbf{Z}_{j}^{*}$  é a matriz  $\mathbf{Z}_{j}$  sem a sua *p*-ésima coluna e  $\boldsymbol{b}_{j}^{*}$  é o vetor  $\boldsymbol{b}_{j}$  rearranjado sem o componente  $b_{pj}$ , ou seja,  $\boldsymbol{b}_{j}^{*} = (b_{1j}, b_{2j}, \dots, b_{(p-1)j}, b_{(p+1)j}, \dots, b_{Pj})^{t}$ . A notação matricial do modelo considerando as J unidades pertencentes ao nível 2 pode ser obtida considerando

$$\begin{split} \mathbf{Y} &= (\mathbf{Y}_{1}^{t}, \mathbf{Y}_{2}^{t}, \dots, \mathbf{Y}_{J}^{t})^{t} (\sum_{j=1}^{J} n_{j}) \times 1, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_{1}^{t} | \mathbf{X}_{2}^{t} | \dots | \mathbf{X}_{J}^{t} \end{bmatrix}^{t} (\sum_{j=1}^{J} n_{j}) \times (\sum_{p=1}^{P} Q_{p}), \\ \mathbf{Z} &= \begin{bmatrix} \mathbf{Z}_{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_{J} \end{bmatrix}_{(\sum_{j=1}^{J} n_{j}) \times JP}, \quad \mathbf{b} = (\mathbf{b}_{1}^{t}, \mathbf{b}_{2}^{t}, \dots, \mathbf{b}_{J}^{t})^{t}_{JP \times 1}, \end{split}$$

e  $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^t, \boldsymbol{\epsilon}_2^t, \dots, \boldsymbol{\epsilon}_J^t)^t (\sum_{j=1}^J n_j) \times 1$ . Assim, temos o modelo na forma matricial dado por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{eta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon}.$$

No caso de algum parâmetro  $\gamma_{pj}$  fixo, temos

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\boldsymbol{b}^* + \boldsymbol{\epsilon},$$

em que

$$\mathbf{Z}^{*} = \begin{bmatrix} \mathbf{Z}_{1}^{*} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{2}^{*} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_{J}^{*} \end{bmatrix} e \ \boldsymbol{b}^{*} = (\boldsymbol{b}_{1}^{*t}, \boldsymbol{b}_{2}^{*t}, \dots, \boldsymbol{b}_{J}^{*t})^{t}.$$

### A.2 Modelos Multiníveis com 3 Níveis

Considerando P variáveis explicativas associadas ao nível 1,  $\sum_{p=1}^{P} Q_p$  variáveis preditoras associadas aos nível 2, onde  $Q_p$  é o número de variáveis explicativas associadas ao p-ésimo parâmetro do nível 1 e  $\sum_{p=1}^{P} \sum_{q=1}^{Q_p} S_{pq}$  variáveis preditoras associadas ao nível 3, onde  $S_{pq}$  é o número de variáveis preditoras associadas ao q-ésimo parâmetro do nível 2 associado ao p-ésimo parâmetro do nível 1, temos as equações gerais para o modelo multinível com três níveis dadas a seguir.

Nível 1: 
$$Y_{ijk} = \sum_{p=1}^{P} \gamma_{pj|k} G_{pijk} + \epsilon_{ijk};$$
  
Nível 2:  $\gamma_{pj|k} = \sum_{q=1}^{Q_p} \delta_{pqk} T_{pqjk} + \xi_{pj|k};$  (A.4)  
Nível 3:  $\delta_{pqk} = \sum_{s=1}^{S_{pq}} \beta_{pqs} L_{pqsk} + \vartheta_{pqk}.$ 

em que  $i = 1, 2, ..., n_{jk}$ ,  $j = 1, 2, ..., m_k$  e k = 1, 2, ..., K. Aqui,  $n_{jk}$  é o número de indivíduos pertencentes ao *j*-ésimo "cluster" do nível 2 pertencente ao *k*-ésimo "cluster" do nível 3;  $m_k$  é o número de "clusters" do nível 2 pertencentes ao *k*-ésimo "cluster" do nível 3 e *K* é o total de "clusters" do nível 3.

Em (A.4), temos que:

- $Y_{ijk}$  é o valor da variável resposta para o *i*-ésimo indivíduo pertencente ao *j*-ésimo "cluster" do nível 2 que por sua vez pertence ao *k*-ésimo "cluster" do nível 3;
- $G_{pijk}$  é o valor da *p*-ésima variável preditora associada ao *i*-ésimo indivíduo do *j*-ésimo "cluster" do nível 2 pertencente *k*-ésimo "cluster" do nível 3;
- $\gamma_{pj|k}$ é o parâmetro associado à variável  $G_p$  no *j*-ésimo "cluster" do nível 2 contido no k-ésimo "cluster" do nível 3;

 $\epsilon_{ijk}$  é o erro associado ao valor da variável resposta do nível 1  $Y_{ijk}$ ;

 $T_{pqjk}$ é o valor da q-ésima variável preditora associada ao parâmetro  $\gamma_{pj|k}$ ;

 $\delta_{pqk}$ é o parâmetro associado à variável  $T_{pq}$  no k-ésimo "cluster";

 $\xi_{pj|k}$ é o erro associado ao parâmetro  $\gamma_{pj|k};$ 

 $\mathcal{L}_{pqsk}$ é o valor da s-ésima variável explicativa associada ao parâmetro $\delta_{pqk};$ 

 $\beta_{pqs}$ é o parâmetro associado à variável  $\mathcal{L}_{pqs};$ 

 $\vartheta_{pqk}$ é o erro associado ao parâmetro  $\delta_{pqk}.$ 

Para uma primeira notação matricial deste modelo, sejam

$$\mathbf{Y}_{jk} = (Y_{1jk}, Y_{2jk}, \dots Y_{n_{jk}jk})^{t}_{n_{jk} \times 1}; \quad \mathbf{G}_{jk} = \begin{bmatrix} G_{11jk} & G_{21jk} & \dots & G_{P1jk} \\ G_{12jk} & G_{22j} & \dots & G_{P2jk} \\ \vdots & \vdots & \ddots & \vdots \\ G_{1n_{jk}jk} & G_{2n_{jk}jk} & \dots & G_{Pn_{jk}jk} \end{bmatrix};$$

$$\mathbf{Y}_{jk} = (\gamma_{1j|k}, \gamma_{2j|k}, \dots, \gamma_{Pj|k})^{t}{}_{P \times 1}; \quad \mathbf{\epsilon}_{jk} = (\epsilon_{1jk}, \epsilon_{2jk}, \dots, \epsilon_{n_{jk}jk})^{t}{}_{n_{jk} \times 1};$$
$$\mathbf{T}_{jk} = \begin{bmatrix} \mathbf{T}_{1jk} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2jk} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_{Pjk} \end{bmatrix};$$

em que

$$\mathbf{T}_{pjk} = \left( \mathbf{T}_{p1jk}, \mathbf{T}_{p2jk}, \dots, \mathbf{T}_{pQ_pjk} \right)_{1 \times Q_p};$$

 $\boldsymbol{\delta}_{k} = (\delta_{11k}, \delta_{12k}, \dots, \delta_{1Q_{1}k}, \delta_{21k}, \delta_{22k}, \dots, \delta_{2Q_{2}k}, \dots, \delta_{P1k}, \delta_{P2k}, \dots, \delta_{PQ_{P}k})^{t} \sum_{(\sum_{p=1}^{P} Q_{p}) \times 1};$ 

$$\boldsymbol{\xi}_{jk} = (\xi_{1j|k}, \xi_{2j|k}, \dots, \xi_{Pj|k})^t_{P \times 1};$$

$$\boldsymbol{\beta} = (\beta_{111}, \dots, \beta_{11S_{11}}, \beta_{121}, \dots, \beta_{12S_{12}}, \dots, \beta_{PQ_P1}, \dots, \beta_{PQ_PS_{PQ_P}})^t (\sum_{p=1}^{P} \sum_{q=1}^{Q_p} S_{pq}) \times 1;$$

$$\mathbf{L}_{k} = \begin{bmatrix} \mathbf{L}_{11k} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{12k} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \mathbf{L}_{1Q_{1}k} & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{L}_{PQ_{p}k} \end{bmatrix};$$

em que

$$\mathbf{L}_{pqk} = \left( \mathbf{L}_{pq1k}, \mathbf{L}_{pq2k}, \dots, \mathbf{L}_{pqS_{pq}k} \right)_{1 \times S_{pq}}$$

е

$$\boldsymbol{\vartheta}_{k} = \left(\vartheta_{11k}, \ldots, \vartheta_{1Q_{1}k}, \vartheta_{21k}, \ldots, \vartheta_{2Q_{2}k}, \ldots, \vartheta_{P1k}, \ldots, \vartheta_{PQ_{P}k}\right)^{t} (\sum_{p=1}^{P} Q_{p}) \times 1^{\cdot}$$

Então, temos

$$\mathbf{Y}_{jk} = \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk} + \boldsymbol{\epsilon}_{jk}; \tag{A.5}$$

$$\boldsymbol{\gamma}_{jk} = \mathbf{T}_{jk} \boldsymbol{\delta}_k + \boldsymbol{\xi}_{jk}; \tag{A.6}$$

$$\boldsymbol{\delta}_k = \mathbf{L}_k \boldsymbol{\beta} + \boldsymbol{\vartheta}_k. \tag{A.7}$$

Substituindo (A.7) em (A.6) e posteriormente (A.6) em (A.5), obtemos

$$\mathbf{Y}_{jk} = \mathbf{G}_{jk} [\mathbf{T}_{jk} (\mathbf{L}_k \boldsymbol{eta} + \boldsymbol{\vartheta}_k) + \boldsymbol{\xi}_{jk}] + \boldsymbol{\epsilon}_{jk} = \mathbf{G}_{jk} \mathbf{T}_{jk} \mathbf{L}_k \boldsymbol{eta} + \mathbf{G}_{jk} \mathbf{T}_{jk} \boldsymbol{\vartheta}_k + \mathbf{G}_{jk} \boldsymbol{\xi}_{jk} + \boldsymbol{\epsilon}_k.$$

Assim, temos o modelo dado por

$$\mathbf{Y}_{jk} = \mathbf{X}_{jk}\boldsymbol{\beta} + \mathbf{Z}_{jk}\boldsymbol{b}_{jk} + \boldsymbol{\epsilon}_{jk},$$

em que  $\mathbf{X}_{jk} = \mathbf{G}_{jk}\mathbf{T}_{jk}\mathbf{L}_k$ ,  $\mathbf{Z}_{jk} = [\mathbf{G}_{jk}\mathbf{T}_{jk}|\mathbf{G}_{jk}]$  e  $\boldsymbol{b}_{jk} = (\boldsymbol{\vartheta}_k^t, \boldsymbol{\xi}_{jk}^t)^t$ . Caso alguns dos parâmetros  $\gamma_{pj|k}$ 's sejam considerados fixos para quaisquer valores de  $p = 1, 2, \ldots, P$  (ou seja,  $\boldsymbol{\xi}_{pjk} = 0$  para estes parâmetros e consequentemente  $\boldsymbol{\vartheta}_{pqk} = 0$  para todo p cujos parâmetros  $\gamma_{pj|k}$ 's são fixos e para todo  $q = 1, 2, \ldots, Q_p$ ), temos o modelo dado por

$$\mathbf{Y}_{jk} = \mathbf{X}_{jk} \boldsymbol{\beta} + \mathbf{Z}_{jk}^{**} \boldsymbol{b}_{jk}^{**} + \boldsymbol{\epsilon}_{jk},$$

em que  $\mathbf{Z}_{jk}^{**} = [\mathbf{G}_{jk}\mathbf{T}_{jk}^{**}|\mathbf{G}_{jk}^{*}]$ , sendo  $\mathbf{G}_{jk}^{*}$  a matriz  $\mathbf{G}_{jk}$  sem a *p*-ésima coluna e  $\mathbf{T}_{jk}^{**}$  é a matriz  $\mathbf{T}_{jk}$  sem as colunas que contém os valores das variáveis  $\mathbf{T}_{p1jk}, \mathbf{T}_{p2jk}, \ldots, \mathbf{T}_{pQ_pjk}$ , para todos os valores de *p* cujos parâmetros  $\gamma_{pj|k}$ 's são fixos. O vetor  $\mathbf{b}_{jk}^{**}$  é obtido rearranjando o vetor  $\mathbf{b}_{jk}$  de modo que os efeitos referentes aos parâmetros considerados fixos sejam omitidos, de forma análoga ao caso do modelo multinível com dois níveis. No caso de apenas alguns dos parâmetros  $\delta_{pqk}$ 's serem considerados fixos para quaisquer valores de  $p = 1, 2, \ldots, P$  e  $q = 1, 2, \ldots, Q_p$  (ou seja,  $\vartheta_{pqk} = 0$  para os valores de p e q tais que os  $\delta_{pqk}$ 's sejam considerados fixos), temos o modelo dado por

$$\mathbf{Y}_{jk} = \mathbf{X}_{jk} \boldsymbol{eta} + \mathbf{Z}_{jk}^* \boldsymbol{b}_{jk}^* + \boldsymbol{\epsilon}_{jk},$$

em que  $\mathbf{Z}_{jk}^* = [\mathbf{G}_{jk}\mathbf{T}_{jk}^*|\mathbf{G}_{jk}]$ , sendo  $\mathbf{T}_{jk}^*$  a matriz  $\mathbf{T}_{jk}$  sem as colunas que contém os valores das variáveis  $\mathbf{T}_{pqjk}$ , para todos os valores de  $p \in q$  tais que os parâmetros  $\delta_{pqk}$ 's sejam fixos. O vetor  $\boldsymbol{b}_{jk}^*$  é obtido de forma análoga ao vetor  $\boldsymbol{b}_{jk}^{**}$ .

Fazendo

$$\mathbf{Y}_{k} = (\mathbf{Y}_{1k}^{t}, \mathbf{Y}_{2k}^{t}, \dots, \mathbf{Y}_{m_{k}k}^{t})^{t}{}_{(\sum_{j=1}^{m_{k}} n_{jk}) \times 1}; \quad \mathbf{G}_{k} = \begin{bmatrix} \mathbf{G}_{1k} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{2k} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_{m_{k}k} \end{bmatrix}_{(\sum_{j=1}^{m_{k}} n_{jk}) \times (Pm_{k})};$$

$$\boldsymbol{\gamma}_{k} = (\boldsymbol{\gamma}_{1k}^{t}, \boldsymbol{\gamma}_{2k}^{t}, \dots, \boldsymbol{\gamma}_{m_{k}k}^{t})^{t}{}_{(Pm_{k})\times 1}; \quad \boldsymbol{\epsilon}_{k} = (\boldsymbol{\epsilon}_{1k}^{t}, \boldsymbol{\epsilon}_{2k}^{t}, \dots, \boldsymbol{\epsilon}_{m_{k}k}^{t})^{t}{}_{(\sum_{j=1}^{m_{k}} n_{jk})\times 1};$$
$$\mathbf{T}_{k} = [\mathbf{T}_{1k}^{t} | \mathbf{T}_{2k}^{t} | \dots | \mathbf{T}_{m_{k}k}^{t} ]^{t}{}_{(Pm_{k})\times (\sum_{p=1}^{P} Q_{p})}; \quad \boldsymbol{\xi}_{k} = (\boldsymbol{\xi}_{1k}^{t}, \boldsymbol{\xi}_{2k}^{t}, \dots, \boldsymbol{\xi}_{m_{k}k}^{t})^{t}{}_{(Pm_{k})\times 1},$$

Temos o modelo dado por

$$\mathbf{Y}_k = \mathbf{G}_k \boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k; \tag{A.8}$$

$$\boldsymbol{\gamma}_k = \mathbf{T}_k \boldsymbol{\delta}_k + \boldsymbol{\xi}_k; \tag{A.9}$$

$$\boldsymbol{\delta}_k = \mathbf{L}_k \boldsymbol{\beta} + \boldsymbol{\vartheta}_k. \tag{A.10}$$

Substituindo (A.10) em (A.9) e posteriormente (A.9) em (A.8), obtemos

$$\mathbf{Y}_k = \mathbf{G}_k [\mathbf{T}_k (\mathbf{L}_k \boldsymbol{\beta} + \boldsymbol{\vartheta}_k) + \boldsymbol{\xi}_k] + \boldsymbol{\epsilon}_k = \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k \boldsymbol{\beta} + \mathbf{G}_k \mathbf{T}_k \boldsymbol{\vartheta}_k + \mathbf{G}_k \boldsymbol{\xi}_k + \boldsymbol{\epsilon}_k$$

Assim, temos o modelo dado por

$$\mathbf{Y}_k = \mathbf{X}_k \boldsymbol{\beta} + \mathbf{Z}_k \boldsymbol{b}_k + \boldsymbol{\epsilon}_k,$$

em que  $\mathbf{X}_k = \mathbf{G}_k \mathbf{T}_k \mathbf{L}_k$ ,  $\mathbf{Z}_k = [\mathbf{G}_k \mathbf{T}_k | \mathbf{G}_k]$  e  $\boldsymbol{b}_k = (\boldsymbol{\vartheta}_k^t, \boldsymbol{\xi}_k^t)^t$ . Para o caso de alguns dos parâmetros em  $\boldsymbol{\gamma}_k$  e/ou em  $\boldsymbol{\delta}_k$  serem fixos, as matrizes  $\mathbf{Z}_k^{**}$  ou  $\mathbf{Z}_k^*$  são obtidas através das matrizes  $\mathbf{Z}_{jk}^{**}$  e  $\mathbf{Z}_{jk}^*$  respectivamente, bem como os vetores  $\boldsymbol{b}_k^{**}$  ou  $\boldsymbol{b}_k^*$  através dos vetores  $\boldsymbol{b}_{jk}^{**}$  e  $\boldsymbol{b}_{jk}^*$  respectivamente, para  $j = 1, 2, \dots, m_k$  e para  $k = 1, 2, \dots, K$ .

Para obter a notação matricial considerando os K "clusters" referentes ao nível 3, temos

$$\mathbf{Y} = (\mathbf{Y}_{1}^{t}, \mathbf{Y}_{2}^{t}, \dots, \mathbf{Y}_{K}^{t})^{t} (\sum_{k=1}^{K} \sum_{j=1}^{m_{k}} n_{jk}) \times 1;$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_{K} \end{bmatrix}_{(\sum_{k=1}^{K} \sum_{j=1}^{m_{k}} n_{jk}) \times (P \sum_{k=1}^{K} m_{k})};$$

$$\gamma = (\gamma_{1}^{t}, \gamma_{2}^{t}, \dots, \gamma_{K}^{t})^{t} (P \sum_{k=1}^{K} m_{k}) \times 1;$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{T}_{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{T}_{K} \end{bmatrix}_{(P \sum_{k=1}^{K} m_{k}) \times (K \sum_{p=1}^{P} Q_{p})};$$

$$\boldsymbol{\delta} = (\boldsymbol{\delta}_{1}^{t}, \boldsymbol{\delta}_{2}^{t}, \dots, \boldsymbol{\delta}_{K}^{t})^{t} (K \sum_{p=1}^{P} Q_{p}) \times 1;$$

$$\mathbf{L} = [\mathbf{L}_{1}^{t} | \mathbf{L}_{2}^{t} | \dots | \mathbf{L}_{K}^{t} ]^{t} (K \sum_{p=1}^{P} Q_{p}) \times (\sum_{p=1}^{P} \sum_{q=1}^{Q} S_{pq});$$

$$\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_{1}^{t}, \boldsymbol{\vartheta}_{2}^{t}, \dots, \boldsymbol{\vartheta}_{K}^{t})^{t} (K \sum_{p=1}^{P} Q_{p}) \times 1;$$

MANGHI, R. F.

Assim, obtemos

$$\mathbf{Y} = \mathbf{G}\boldsymbol{\gamma} + \boldsymbol{\epsilon}; \tag{A.11}$$

$$\boldsymbol{\gamma} = \mathbf{T}\boldsymbol{\delta} + \boldsymbol{\xi}; \tag{A.12}$$

$$\boldsymbol{\delta} = \mathbf{L}\boldsymbol{\beta} + \boldsymbol{\vartheta}. \tag{A.13}$$

Substituindo (A.13) em (A.12) e posteriormente (A.12) em (A.11), obtemos

$$\mathbf{Y} = \mathbf{G}[\mathbf{T}(\mathbf{L}\boldsymbol{\beta} + \boldsymbol{\vartheta}) + \boldsymbol{\xi}] + \boldsymbol{\epsilon} = \mathbf{GTL}\boldsymbol{\beta} + \mathbf{GT}\boldsymbol{\vartheta} + \mathbf{G}\boldsymbol{\xi} + \boldsymbol{\epsilon}.$$

Assim, temos o modelo dado por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\epsilon},$$

em que  $\mathbf{X} = \mathbf{GTL}$ ,  $\mathbf{Z} = [\mathbf{GT}|\mathbf{G}] \in \boldsymbol{b} = (\boldsymbol{\vartheta}^t, \boldsymbol{\xi}^t)^t$ . Também no caso de parâmetros considerados fixos em  $\boldsymbol{\gamma} \in \boldsymbol{\delta}$ , podemos obter as matrizes  $\mathbf{Z}^{**}$  ou  $\mathbf{Z}^*$  a partir das matrizes  $\mathbf{Z}_k^{**}$  e  $\mathbf{Z}_k^*$  respectivamente e os vetores  $\boldsymbol{b}^{**}$  ou  $\boldsymbol{b}^*$  podem ser obtidos a partir dos vetores  $\boldsymbol{b}_k^{**} \in \boldsymbol{b}_k^*$ respectivamente, para  $k = 1, 2, \dots, K$ .

# **Referências Bibliográficas**

- Atkinson, C. A. (1985), Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis, Oxford: Oxford University Press.
- [2] Casella, G.; Berger, R. L. (2002), *Statistical Inference*, Pacific Grove: Duxbury/Thomson.
- [3] Cooper, D. M.; Thompson, R. (1977), A Note on the Estimation of the Parameters of the Autoregressive - Moving Average Process, *Biometrika*, 64, 625.
- [4] Copt, S.; Victoria-Feser, M. (2006), High Breakdown Inference in the Mixed Linear Model, Journal of the American Statistical Association, 101, 292–300.
- [5] Crawley, M. J. (2009), The R Book, Chichester: Wiley.
- [6] Crump, S. L. (1946), The Estimation of Variance Components in Analysis of Variance, *Biometrics Bulletin*, 2, 7–11.
- [7] Cysneiros, F. J. A.; Paula, G. A. (2004), One-sided Tests in Linear Models With Multivariate t-Distribution, *Communications in Statistics - Simulation and Computacion*, 33, 747–771.
- [8] Dempster, A. P.; Laird, N. M.; Rubin, D. B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, Series B (Methodological), 39, 1–38.
- [9] Dempster, A. P.; Rubin, D. B.; Tsutakawa, R. K. (1981), Estimation in Covariance Components Models, *Journal of the American Statistical Association*, 76, 341–353.

- [10] Fang, K. T.; Kotz, S; Ng, K. W. (1990), Symmetric Multivariate and Related Distributions, New York: Chapman & Hall.
- [11] Fernández, C.; Steel, M. F. J. (1999), Multivariate Student-t Regression Models:
   Pitfalls and Inference, *Biometrika*, 86, 153–167.
- [12] Galea, M.; Paula, G. A.; Bolfarine, H. (1997), Local Influence in Elliptical Linear Regression Models, Journal of the Royal Statistical Society, Series D (The Statistician), 46, 71–79.
- [13] Galea, M.; Paula, G. A.; Uribe-Opazo, M. (2003), On Influence Diagnostic in Univariate Elliptical Linear Regression Models, *Statistical Papers*, 44, 23–45.
- [14] Gelman, A.; Hill, J. (2007), Data Analysis Using Regression and Multilevel/Hierarchical Models, New York: Cambridge University Press.
- [15] Giampaoli, V.; Singer, J. M. (2009), Likelihood Ratio Tests for Variance Components in Linear Mixed Models, *Journal of Statistical Planning and Inference*, 139, 1435–1448.
- [16] Goldstein, H. (2009), Multilevel Statistical Models, Cambridge: Cambridge University Press.
- [17] Gómez, E.; Gómez-Villegas, M. A.; Marín, J. M. (1998), A Multivariate Generalization of the Power Exponential Family of Distributions, *Communications in Statistics* - Theory and Methods, 27, 589-600.
- [18] Harville, D. A. (1976), Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects, *The Annals of Statistics*, 4, 384–395.
- [19] Harville, D. A. (1977), Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems, *Journal of the American Statistical Association*, **72**, 320–338.
- [20] Hilden-Minton, J. A. (1995), Multilevel Diagnostics for Mixed and Hierarchical Models, PhD Thesis, University of California, Los Angeles.

- [21] Hill, H. C.; Rowan, B.; Ball, D. L. (2005), Effect of Teacher's Mathematical Knowledge for Teaching on Student Achievement, American Educational Research Journal, 42, 371–406.
- [22] Kenward, M. G.; Roger, J. H. (1997), Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, *Biometrics*, 53, 983–997.
- [23] Kowalsky, J.; Mendoza-Blanco, J. R.; Tu, X. M.; Gleser, L. J. (1999), On the Diference in Inference and Prediction Between the Joint and Independent t-Error Models for Seemingly Unrelated Regressions, *Communications in Statistics - Theory* and Methods, 28, 2219–2140.
- [24] Laird, N. M. (1982), Computation of Variance Components Using the EM Algorithm, Journal of Statistical Computation and Simulation, 14, 295–303.
- [25] Laird N. M.; Ware, J. H. (1982), Random-Effects Models for Longitudinal Data, Biometrics, 38, 963-974.
- [26] Lange, K. L.; Little, J. A.; Taylor, M. G. (1989), Robust Statistical Modeling Using the t Distribution, Journal of The American Statistical Association, 84, 881–896.
- [27] Lee, Y.; Nelder, J. A. (1996), Hierarchical Generalized Linear Models, Journal of the Royal Statistical Society, Series B (Methodological), 58, 619–678.
- [28] Lindley, D. V.; Smith, A. F. M. (1972), Bayes Estimates for the Linear Model, Journal of the Royal Statistical Society, Series B (Methodological), 34, 1–41.
- [29] Little, R. J. A. (1988), Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values, Journal of the Royal Statistical Society, Series C (Applied Statistics), 37, 23–38.
- [30] Liu, C.; Rubin, D. B. (1995), ML Estimation of the t Distribution Using EM and its Extensions, ECM and ECME, *Statistica Sinica*, 5, 19–39.

- [31] Longford, N. T. (1987), A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects, *Biometrika*, 74, 817–827.
- [32] McCullagh, P.; Nelder, J. A. (1999), Generalized Linear Models, Boca Raton: Chapman & Hall/CRC.
- [33] McLean, R. A.; Sanders, W. L.; Stroup, W. W. (1991), A Unified Approach to Mixed Linear Models, *The American Statistician*, 45, 54–64.
- [34] Meng, X. L.; Rubin, D. B. (1993), Maximum Likelihood Estimation Via the ECM Algorithm: A General Framework, *Biometrika*, 80, 267–278.
- [35] Meng, X. L.; van Dyk, D. (1997), The EM Algorithm An Old Folk Song Sung to A Fast New Tune (With Discussion), Journal of the Royal Statistical Society, Series B, 59, 511-567.
- [36] Morrell, C. H. (1998), Likelihood Ratio Testing of Variance Components in the Linear Mixed–Effects Model Using Restricted Maximum Likelihood, *Biometrics*, 54, 1560–1568.
- [37] Natis, L. (2000), Modelos Lineares Hierárquicos, Dissertação de Mestrado, São Paulo: IME/USP.
- [38] Nobre, J. S. (2003), Métodos de Diagnóstico para Modelos Lineares Mistos, Dissertação de Mestrado, São Paulo: IME/USP.
- [39] Nobre, J. S.; Singer, J. M. (2007), Residual Analysis for Linear Mixed Models, Biometrical Journal, 49, 863-875.
- [40] Osorio, F.; Paula, G. A.; Galea, M. (2007), Assessment of Local Influence in Elliptical Linear Models with Longitudinal Structure, *Computational Statistics and Data Analysis*, **51**, 4354–4368.
- [41] Patterson, H. D.; Thompson, R. (1971), Recovery of Inter-block Information when Block Sizes are Unequal, *Biometrika*, 58, 545.

- [42] Paula, G. A. (2010), Modelos de Regressão com Apoio Computacional, São Paulo: IME/USP. http://www.ime.usp.br/~giapaula/texto\_2010.pdf.
- [43] Pinheiro, J. C.; Bates, D. M. (2000), Mixed-Effects Models in S and S-PLUS, Berlim: Springer-Verlag.
- [44] Pinheiro, J. C.; Liu, C.; Wu, Y. W. (2001), Efficient Algorithms for Robust Estimation in Linear Mixed–Effects Models Using the Multivariate t Distribution, Journal of Computational and Graphical Statistics, 10, 249–276.
- [45] Pires, J. F. (2009), Influência Local Através da Curvatura Normal em Modelos Multiníveis, Dissertação de Mestrado, Recife: DE/UFPE.
- [46] Raudenbush, S. W.; Bryk, A. S. (2002), Hierarchical Linear Models: Applications and Data Analysis Methods, Thousand Oaks: Sage Publications.
- [47] Satterthwaite, F. E. (1946), An Approximate Distribution of Estimates of Variance Components, *Biometrics Bulletin*, 2, 110–114.
- [48] Savalli, C. (2005), Testes do Tipo Escore para Componentes de Variância em Modelos Elípticos Lineares Mistos, Tese de Doutorado, São Paulo: IME/USP.
- [49] Savalli, C.; Paula, G. A.; Cysneiros, F. J. A. (2006), Assessment of Variance Components in Elliptical Linear Mixed Models, *Statistical Modelling*, 6, 59–76.
- [50] Searle, S. R.; Casella, G.; McCulloch, C. E. (1992), Variance Components, New York: Wiley.
- [51] Self, S. G.; Liang, K. (1987), Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions, Journal of the American Statistical Association, 82, 605.
- [52] Silvapulle, M. J.; Silvapulle, P (1995), A Score Test Against One-Sided Alternatives, Journal of the American Statistical Association, 90, 342–349.
- [53] Snijders, T. A. B.; Bosker, R. J. (1999), Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, Newbury Park: Sage Publications.

- [54] Stram, D. O.; Lee, J. W. (1994), Variance Components Testing in the Longitudinal Mixed Effects Model, *Biometrics*, 50, 1171.
- [55] Valle, R. B. A. (1994), Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão, Tese de Doutorado, São Paulo: IME/USP.
- [56] Verbeke, G.; Molenberghs, G. (2000), Linear Mixed Models for Longitudinal Data, New York: Springer-Verlag.
- [57] Verbeke, G.; Molenberghs, G. (2003), The Use of Score Tests for Inference on Variance Components, *Biometrics*, 59, 254–262.
- [58] Verbyla, A. P. (1990), A conditional Derivation of Residual Maximum Likelihood, The Australian Journal of Statistics, 32, 227.
- [59] Waternaux, C.; Laird N. M.; Ware, J. H. (1989), Methods for Analysis of Longitudinal Data: Blood-Lead Concentrations and Cognitive Development, *Journal of the American Statistical Association*, 84, 33–41.
- [60] West, B. T.; Welch, K. B.; Galecki, A. T. (2007), Linear Mixed Models: A Practical Guide Using Statistical Software, Boca Raton: Chapman & Hall/CRC.
- [61] Zellner, A. (1976), Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms, *Journal of the American Statistical Association*, **71**, 400–405.