Métodos de Diagnóstico para Modelos Lineares Mistos

Juvêncio Santos Nobre

DISSERTAÇÃO APRESENTADA AO INSTITUTO DE MATEMÁTICA E ESTATÍSTICA DA UNIVERSIDADE DE SÃO PAULO PARA OBTENÇÃO DO GRAU DE MESTRE EM ESTATÍSTICA

Área de Concentração: Estatística Orientador: Julio da Motta Singer

Durante a elaboração deste trabalho o autor recebeu apoio financeiro do CNPq

– São Paulo, Março de 2004 –

Métodos de Diagnóstico para Modelos Lineares Mistos

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Juvêncio Santos Nobre e aprovada pela comissão julgadora.

São Paulo, 04 de Março de 2004.

Comissão Julgadora:

- Prof. Dr. Julio da Motta Singer (Orientador)- IME/USP
- Prof. Dr. Dalton Francisco de Andrade UFSC
- Prof. Dr. Geraldo da Silva e Souza UNB

"All models are wrong, but some are useful".

Box

"Wir müssen wissen, Wir werden wissen".

David Hilbert

"Mesmo as noites, totalmente sem estrelas podem anunciar a aurora de uma grande realização".

Martin Luther King

Dedico este trabalho

À Deus acima de tudo

À minha mãe Gracilene, por ser a melhor pãe (mãe e pai) do mundo À minha noiva Jacqueline, por existir na minha vida

Aos meus mestres e amigos Prof. Maurício, Prof. Julio e Prof. Dalton, obrigado por tudo

À Terezinha de Campos Modesto, In Memoriam.

Agradecimentos

Gostaria de agradecer:

A Deus, por me oferecer saúde, disposição, discernimento e por colocar várias pessoas maravilhosas na minha vida, além de me fornecer inúmeras oportunidades. A Ele só tenho a agradecer por tudo.

A meu orientador, mestre e amigo, professor Julio da Motta Singer, pela ótima recepção e enorme solicitude prestadas quando cheguei em São Paulo. Obrigado pela grande oportunidade de ser seu orientando, agradeço pela sua competente orientação, pelo grande entusiasmo, apoio, paciência, segurança, tranqüilidade e força transmitidos durante todo o meu mestrado e principalmente durante a elaboração deste trabalho.

Ao professor Maurício, *mais do que um professor*: um grande mestre, amigo e pai. Obrigado pelos inúmeros conselhos, palavras de conforto, o grande incentivo dado para eu fazer o mestrado, pelo *incomensurável* apoio (desde do período da graduação até os dias de hoje) e por sempre acreditar em mim.

Ao professor Dalton, um grande mestre e amigo. Obrigado pelo incentivo, apoio e colaboração, principalmente no que tange a decisão de fazer o mestrado.

Ao professor Welliandre, meu primeiro orientador, agradeço pela paciência, pelo entusiasmo, incentivo e ensinamentos transmitidos no decorrer de toda a minha vida acadêmica.

Agradeço à pessoa a quem tenho plena consciência de dever tudo que sou hoje: Gracilene, meu pai e mãe reunidos em uma única pessoa; grande responsável por tudo de bom que acontece na minha vida, meu grande ídolo. Obrigado, por todo amor destinado, pelos carinhos, as palavras de conforto, os inúmeros ensinamentos, a força dada e por sempre incentivar, confiar e acreditar em mim, mesmo nos piores momentos. Obrigado **Mãezinha**, este momento é a realização do nosso "sonho", que por inúmeras vezes pareceu impossível, por este motivo, dedico esse trabalho à senhora.

À minha noiva Jacqueline (meu Teorema Central do Limite), obrigado pela paciência, compreensão, ternura, pelos sonhos, companherismo, força, amor e apoio destinados nos momentos mais difíceis, nunca esqueça que você é muito importante para mim e principalmente de que o mundo nos pertence.

Aos meus padrinhos *Fátima* e *Tácito* pela grande ajuda dada a minha mãe e a mim, principalmente durante a infância. A Maria Medeiros, Terezinha de Campos Modesto e a tia Jack por nos ajudar nos momentos em que mais precisávamos.

Aos componentes da minha banca, Prof. Dalton e Prof. Geraldo da Silva, e aos professores Gilberto Alvarenga e Francisco Cysneiros, pelas sugestões e comentários valiosos para o melhoramento desta dissertação.

À todos que fazem parte do Departamento de Estatística e Matemática Aplicada da Universidade Federal do **Ceará**. Dentre os professores, gostaria de agradecer: João Maurício, Rosa Salani, João Welliandre, Nelson Braga, Júlio Barros, Sílvia Maria, Ana Maria, Robson Medeiros, André Jalles e Manoel Campelo; as funcionárias: Margarida, Margéri, Luisa e Mariluse; aos grandes amigos que formei durante meus quatro anos de graduação: Caio, Carlos, Dhavynci, Agnaldo, Cledinaldo e Adriana; aos amigos que ingressaram na turma de **1998.1**, que me ajudaram por demais nos primeiros semestres; todos os alunos da turma de Probabilidade I do primeiro semestre de 2000, em especial: Jacqueline, Saulo, Josemar, Erivan, Ricardo, Robério, Fábio, Cynthia, Velma, Cleudimar, Marcos, etc; a todos os alunos do curso de Probabilidade III ministrado no verão de 2001. Certamente todos ficaram torcendo por mim nesse desafio em São Paulo. Obrigado pela amizade e força que vocês me deram e tenham certeza que vocês fazem parte da minha família.

A todos os professores e funcionários do Departamento de Estatística do IME-USP, em especial aos professores no qual tive o prazer de manter um contato maior: Antônio Carlos, Caio Dantas, Carlos Alberto Bragança, Chang Chiann, Denise Botter, Elisabeth Kira, Fábio Prates, Gilberto Alvarenga, Márcia Branco, Serguei Popov, Sílvia Ferrari, Vanderley Bueno e Wagner Borges e as funcionárias: Simone, Cecília, Helena e Elaine (CEA) que sempre estão dispostas a ajudar com simpatia e presteza. Agradeço também a todos os funcionários da biblioteca do IME/USP e do serviço de xerox, por sempre atenderem com prontidão.

Aos meus amigos de Pós-Graduação (não ousarei enumerar todos), em especial para meu grande amigo (leia irmão) Caio, pela *descomensurável* ajuda dada no período do curso de verão e no decorrer de todo o meu primeiro ano em São Paulo, pelos momentos de força, apoio, pelas idas ao rodízio, pelas piadas sem graça, pelos jogos aos Sábados de madrugada, em especial o "quebra" e pelas "pouquissímas" ($\rightarrow 0$) vezes que tive que escutar: A mim, pouco se-me-dá que as êmulas claudiquem, o que me apraz é acicatálas. Aos três amigos no qual tenho um carinho muito especial: Diana ("chegou cedo cara ?"), Elier ("Desculpa professor") e Perseverando (pelas brincadeiras, idas ao estádio, mesmo nos jogos sem graça como SP 3 x 1 For em 17/04/03 e COR 2 x 0 For em 13/09/03), obrigado pela sincera amizade. Gostaria de agradecer também a Adrilayne, Rogério & Família (SP), Marcelo & Lane (CE-SP), Waldemar (Virgulino-PE), Michel (PE), Francisco Cysneiros & Audrey Cysneiros (PE), Michelle & Horácio (PB), Patrícia (BA), Diana & Gustavo, Regina & Olímpio (SP), Fred (SP), Cléber (SP), Iracema (SP), Kelly (SP), Paulão ("Cadê a distribuição Juvêncio?"), Gissela (DF), Márcio (CE), Marcelo (RJ), Edvaldo (MG), Alberto (PE), Jorge (Peru), Victor Hugo (Peru), Lourdes (Peru), Romeu ("Zamorano?") e a todos os colegas da minha turma de mestrado e aos amigos do futsal aos sábados dentre muitos outros que ajudaram a transformar minha vida em São Paulo infinitamente mais agradável.

Aos meus amigos cruspianos: Álvaro (SP), Ana (DF), Caio (CE), Edvaldo (MG), Juan (Peru), Marcelo (RJ) e Márcio (CE) obrigado pela excelente convivência e pelo clima de camaradagem. Agradeço também a grande ajuda prestada pela minha assistente social Neusa.

E a todos que não mencionei que me ajudaram (direta ou indiretamente) na realização desse sonho.

Resumo

Muitos fenômenos podem ser representados por meio de modelos estatísticos de forma satisfatória. Para validar tais modelos é necessário verificar se as suposições envolvidas estão satisfeitas e se o modelo é sensível a pequenas perturbações; este é o objetivo da análise de diagnóstico. Neste trabalho apresentamos, discutimos e propomos técnicas de diagnóstico em modelos lineares mistos e as ilustramos com um exemplo prático.

Abstract

Many phenomena can be represented through statistical models in a satisfactory way. To validate such models it is necessary to verify whether the assumptions are satisfied and whether the model is sensitive to small deviations; this constitutes the objective of diagnostic analysis. In this work we present, discuss and propose diagnostic techniques for mixed linear models and illustrate them with a practical example.

Lista de Tabelas

1.1	Índice de placa bacteriana	3
4.1	Estimativas (± EP) dos parâmetros dos modelos (4.2), (4.9) e (4.10) com	
	estrutura de covariâncias (4.8). \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots	48
4.2	Estimativas dos parâmetros do modelo $\left(4.10\right)$ a o eliminar as unidades ex-	
	perimentais #12 e #29	55

Lista de Figuras

1.1	Diagrama de dispersão entre os índices de placa bacteriana pré-escovação						
	e pós-escovação para escova convencional	2					
1.2	Diagrama de dispersão entre os índices de placa bacteriana pré-escovação						
	e pós-escovação para escova monobloco	2					
4.1	Ajuste do modelo final	49					
4.2	Resíduo marginal e EBLUP do modelo final (4.10). \ldots \ldots \ldots	50					
4.3	3 Resíduo condicional padronizado e envelope simulado com 95% para o						
	resíduo com confundimento mínimo	50					
4.4	Alavancagem generalizada	51					
4.5	Influência local.	52					
4.6	Distância de Cook condicional por observação	53					
4.7	Distância de Cook condicional por unidade experimental	54					

Índice

Α	gradecimentos	\mathbf{v}
\mathbf{R}	esumo	viii
A	bstract	ix
Li	ista de Tabelas	x
Li	ista de Figuras	xi
1	Modelos Lineares Mistos	1
	1.1 Introdução e motivação	1
	1.2 Especificação do modelo	5
	1.3 Inferência Estatística	7
	1.4 Testes de hipóteses e critérios de informação	11
2	Análise de Resíduos	14
	2.1 Tipos de resíduos	14
	2.2 Utilização do resíduo condicional	15
	2.3 Utilização do EBLUP	18
3	Análise de Sensibilidade em Modelos Lineares Mistos	20
	3.1 Introdução	20
	3.2 Inclusão de efeitos fixos	20
	3.3 Gráfico da variável adicionada para efeitos fixos	22
	3.4 Decomposição do gráfico da variável adicionada	23
	3.5 Pontos alavanca	24
	3.6 Eliminação de observações	26
	3.7 Influência local	33

	3.8	Influência local em modelos lineares mistos	35 35 36 36 37
4	Ap	olicação	44
	4.1	Introdução	44
	4.2	Especificação do modelo	45
	4.3	Diagnóstico do modelo ajustado	49
5	Co	mentários	56
	5.1	Recursos computacionais	56
	5.2	Pesquisas futuras	56
•	Ð		r 0
Α	Ex	pressoes do Capitulo I $(1, 0)$ $(1, 10)$ $(1, 00)$ $(1, 02)$	98 70
	A.1	$\begin{array}{c} Identifiades (1.9), (1.19), (1.20) e (1.23) \dots \dots$	08 50
		A.1.1 Identidade (1.9)	50
		A.1.2 Identidade (1.19)	59
		A.1.5 Identidade (1.20)	59
	1.9	$\mathbf{PI} \text{ IIE } \mathbf{PI} \text{ IID}$	60
	А.2 А.3	Propriedades do BLUE e BLUP	61
	11.0		01
В	\mathbf{E}	xpressões do Capítulo 2	62
	B.1	Identidades (2.5), (2.6) e (2.7) $\dots \dots \dots$	62
	B.2	Identidades (2.16) e (2.17) \ldots	62
С	xpressões do Capítulo 3	64	
	C.1	BLUE (3.4)	64
	C.2	Fórmula de atualização do BLUP (3.9)	64
	C.3	Identidades (3.10) e (3.11)	65
	C.4	BLUE e BLUP do modelo (3.22)	65
	C.5	Identidades $(3.26), (3.27), (3.28), (3.29) \in (3.30)$	66
	C.6	Decomposição da medida de Cook condicional (3.37)	68
	C.7	Identidade (3.43)	68

C.8 Derivadas (3.55), (3.57) e (3.57) $\dots \dots \dots$	68
C.9 Identidade (3.58)	70
C.10 Matriz (3.61)	70
C.11 Derivadas (3.64), (3.65) e (3.66) $\dots \dots \dots$	71
C.12 Identidade (3.70)	72
C.13 Matriz Hessiana \ldots	72

Referências Bibliográficas

 $\mathbf{74}$

Capítulo ${\bf 1}$

Modelos Lineares Mistos

1.1 Introdução e motivação

Muitas técnicas estatísticas são fundamentadas sobre a hipótese de independência entre as observações. Tal hipótese é razoável em muitos estudos do tipo **transversal** ("crosssectional"), em que apenas uma observação é considerada para cada unidade experimental. Estudos com **medidas repetidas** se referem a casos nos quais cada unidade experimental é observada pelo menos duas vezes. Por essa razão, espera-se uma dependência entre as observações referentes à mesma unidade experimental. Esses estudos abrangem, entre outros, os delineamentos com **parcelas sub-divididas** ("split-plot") e **delineamentos com intercâmbio** ("crossover"), além dos estudos **longitudinais**. A característica que distingue os estudos longitudinais é a ordenação (ao longo do tempo, por exemplo) com que os dados são coletados. Para maiores detalhes sobre estudos longitudinais, veja por exemplo Singer & Andrade (2000) ou Diggle *et al.* (2002).

Estudos longitudinais são comuns em pesquisas de diversas áreas, como Ciências Sociais, Economia, Educação, Medicina, etc. Como ilustração consideramos um estudo realizado na Faculdade de Odontologia da Universidade de São Paulo, que visa comparar dois tipos de escova: monobloco e convencional [Parizzoto (1999)]. Uma avaliação da eficácia dos dois tipos de escova na remoção de placa bacteriana, utilizando ou não dentifrício, está apresentada em Singer *et al.* (2004). Outro objetivo do estudo é comparar os tipos de escova quanto à manutenção da capacidade de remoção da placa bacteriana (durabilidade) sob uso diário. Com esta finalidade, foram observadas 32 crianças em 4 sessões quinzenais, uma das quais correspondente à avaliação inicial. As crianças foram alocadas a dois grupos de tamanhos iguais, cada um submetido ao tratamento com uma das escovas. Durante o período de observação, cada criança utilizou a mesma escova que lhe foi dada na primeira sessão. Em cada sessão de avaliação, mediu-se um índice de placa bacteriana antes (prétratamento) e depois (pós-tratamento) da escovação. Os dados encontram-se na Tabela 1.1. Nas Figuras 1.1 e 1.2 estão apresentados gráficos de dispersão entre os índices de placa bacteriana pré-tratamento (x) e pós-tratamento (y). O que caracteriza os dados desse estudo como longitudinais é a observação das mesmas unidades experimentais (crianças) ao longo das quatro sessões de avaliação. Conforme nomenclatura indicada em Singer & Andrade (2000) consideramos tal estudo como longitudinal e balanceado com respeito ao tempo.





Figura 1.2 Diagrama de dispersão entre os índices de placa bacteriana pré-escovação e pósescovação para escova monobloco.



		1 ^a s	^a sessão 2 ^a sessão		essão	3 ^a s	essão	4 ^a sessão	
Criança	Escova	Antes	Depois	Antes	Depois	Antes	Depois	Antes	Depois
1	Convencional	1.05	1.00	1.13	0.84	1.15	0.86	1.13	0.94
2	Convencional	1.07	0.62	0.92	0.62	1.02	0.57	1.15	0.85
3	Convencional	0.82	0.62	1.52	1.07	1.39	0.97	1.78	1.39
4	Convencional	1.37	0.90	1.65	1.20	1.75	1.40	1.92	1.67
5	Convencional	1.97	1.52	1.30	1.07	1.50	1.15	1.65	1.37
6	Convencional	1.30	0.82	1.17	0.70	0.75	0.50	1.47	1.12
7	Convencional	1.61	1.19	1.52	1.13	1.22	1.00	1.63	1.22
8	Convencional	1.02	0.73	1.08	0.64	0.94	0.73	1.14	0.97
9	Convencional	1.62	1.25	1.45	1.10	1.10	0.75	1.70	1.32
10	Convencional	1.65	1.22	1.57	1.22	1.47	1.10	1.62	1.17
11	Convencional	1.02	0.78	0.60	0.47	0.88	0.75	1.36	1.08
12	Convencional	0.71	0.60	1.13	0.39	0.84	0.65	1.65	1.31
13	Convencional	1.70	1.55	1.85	1.37	1.87	1.55	1.60	1.30
14	Convencional	1.30	1.02	1.65	0.97	1.72	1.20	1.37	1.22
15	Convencional	1.40	0.80	1.83	1.03	1.76	1.38	1.96	1.15
16	Convencional	1.40	1.12	1.25	0.67	1.50	1.10	1.50	1.22
17	Monobloco	1.66	1.63	1.36	1.16	1.52	0.88	1.41	1.20
18	Monobloco	1.02	0.80	0.92	0.82	1.10	0.76	1.28	1.15
19	Monobloco	0.75	0.67	1.00	0.92	1.00	0.87	1.15	1.10
20	Monobloco	1.29	1.23	0.91	0.76	1.14	0.94	1.35	0.97
21	Monobloco	1.27	1.20	1.20	0.95	1.10	1.00	1.37	1.17
22	Monobloco	1.07	0.85	1.39	1.25	1.39	1.25	1.28	1.21
23	Monobloco	1.35	1.21	1.42	1.17	1.42	1.19	1.42	1.23
24	Monobloco	1.32	1.02	1.60	1.40	1.35	1.02	1.50	1.25
25	Monobloco	1.66	1.61	1.50	1.36	1.72	1.41	1.69	1.44
26	Monobloco	1.30	1.07	0.84	0.61	0.88	0.61	0.96	0.57
27	Monobloco	1.57	1.20	1.50	1.07	1.15	1.00	1.25	1.05
28	Monobloco	1.67	1.50	1.47	1.32	1.07	0.97	1.50	1.37
29	Monobloco	0.91	0.67	0.96	0.62	1.09	0.53	1.12	0.37
30	Monobloco	1.06	0.70	1.00	0.85	1.15	0.93	1.12	1.00
31	Monobloco	2.30	2.00	1.37	1.25	1.40	1.32	2.15	1.90
32	Monobloco	1.15	1.00	1.23	1.11	1.15	1.07	1.26	1.00

Tabela 1.1 Índice de placa bacteriana.

Grande parte do esforço empregado na análise de dados com medidas repetidas está relacionada com a modelagem da estrutura de correlação intra-unidades amostrais. Com essa finalidade, Laird & Ware (1982) e Ware (1985) propõem a utilização de **modelos lineares mistos**. McCulloch & Searle (2001) sugerem a inclusão de variáveis latentes (não observáveis) em modelos lineares (ou não lineares) como alternativa para modelar a estrutura de correlação intra-unidades experimentais. Os modelos lineares mistos têm como casos particulares o **modelo linear clássico**, o **modelo de componentes de variância** e também os **modelos hierárquicos** (multiníveis) [Natis (2000)].

Outras alternativas para análise de dados com medidas repetidas consideram modelos lineares generalizados com a inclusão de uma matriz de correlação de "trabalho" para modelar a matriz de covariâncias intra-unidades amostrais [Liang & Zeger (1986)]. A análise sob esses modelos utiliza as chamadas equações de estimação generalizadas (EEG). Para detalhes e aplicações, veja por exemplo, Heyde (1997), Artes (1997), Hardin & Hilbe (2003) e Venezuela (2003). Uma terceira alternativa é utilizar os modelos lineares generalizados mistos [McCulloch & Searle (2001)].

Assim como os demais modelos estatísticos, esta classe de modelos é utilizada como aproximação para processos complexos. Dentro desse contexto é preciso avaliar se tal aproximação é aceitável. Um item de suma importância na análise de tais modelos é sua "validação", usualmente concretizada por meio da **análise de diagnóstico**, que consiste de duas etapas: **avaliação do ajuste** e **análise de sensibilidade**. A primeira etapa corresponde à verificação de possíveis afastamentos das suposições adotadas. A segunda etapa tem por objetivo estudar a variação dos resultados da análise quando se modifica discretamente a formulação considerada inicialmente. Se esta variação é "substancial" no sentindo de mudar conclusões, diz-se que o modelo não é **robusto**, pois sob pequenas modificações leva a resultados significativamente distintos. Neste caso, as conclusões devem ser tomadas (se tomadas) de forma cautelosa, ou então deve-se decidir pelo uso de outro modelo.

No caso de regressão linear, existe uma gama de propostas de medidas e testes para avaliar o ajuste do modelo. Para detalhes, veja, por exemplo, Cook (1977), Hoaglin & Welsch (1978), Belsley *et al.* (1980), Cook & Weisberg (1982), Atkinson (1985), Chatterjee & Hadi (1986, 1988), Johnson & McCulloch (1987), Gray (1989) e Besley (1991). Paula (2003) traz um resumo das técnicas de diagnóstico utilizadas no caso linear e nos modelos lineares generalizados; Venezuela (2003, Cap.3) apresenta técnicas de diagnóstico utilizadas em modelos lineares generalizados para análise de dados com medidas repetidas.

Rocke (1983) e Fellner (1986) desenvolveram métodos robustos de estimação dos parâme-

tros do modelo de componentes de variância; este último autor mostrou como seu método pode ser utilizado para identificar dados discrepantes ("outliers"). Beckman *et al.* (1987) e Lesaffre & Verbeke (1998) desenvolveram métodos de diagnóstico para modelos lineares mistos com base no conceito de influência local. Christensen & Pearson (1992), Hilden-Minton (1995), Banerjee & Frees (1997) e Tan *et al.* (2001) estudam a influência em modelos lineares mistos, causada pela eliminação de observações enquanto que Fung *et al.* (2002) estuda este tipo de influência em **modelos lineares mistos semiparamétricos**.

O objetivo deste trabalho é apresentar métodos de diagnóstico em modelos lineares mistos visando sua utilização prática. Nesse capítulo fazemos uma revisão da teoria de modelos lineares mistos. No Capítulo 2, são discutidas formas para avaliar se as suposições do modelo são satisfeitas, por intermédio da análise de resíduos, enquanto que no Capítulo 3, daremos ênfase à análise de sensibilidade. Uma aplicação a dados reais está apresentada no Capítulo 4.

1.2 Especificação do modelo

Um modelo linear misto pode ser escrito na forma

$$\boldsymbol{Y}_{i} = \boldsymbol{X}_{i}\boldsymbol{\beta} + \boldsymbol{Z}_{i}\boldsymbol{\gamma}_{i} + \boldsymbol{\varepsilon}_{i}, \quad i = 1, ..., c,$$

$$(1.1)$$

em que \boldsymbol{Y}_i representa um vetor $(n_i \times 1)$ de respostas da *i-ésima* unidade experimental, $\boldsymbol{\beta}$ é um vetor $(p \times 1)$ de parâmetros (efeitos fixos), \boldsymbol{X}_i é uma matriz $(n_i \times p)$ de especificação (conhecida e de posto completo) dos efeitos fixos, $\boldsymbol{\gamma}_i$ é um vetor $(q \times 1)$ de variáveis latentes, comumente denominadas efeitos aleatórios, que refletem o comportamento individual da *i-ésima* unidade experimental, \boldsymbol{Z}_i é uma matriz $(n_i \times q)$ de especificação (conhecida e de posto completo) dos efeitos aleatórios e $\boldsymbol{\varepsilon}_i$ é um vetor $(n_i \times 1)$ de erros aleatórios. Fazendo $\boldsymbol{Y} = (\boldsymbol{Y}_1^{\top}, \cdots, \boldsymbol{Y}_c^{\top})^{\top}, \ \boldsymbol{X} = (\boldsymbol{X}_1^{\top} \cdots \boldsymbol{X}_c^{\top})^{\top}, \ \boldsymbol{Z} = \text{diag}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_c), \ \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\top}, \cdots, \boldsymbol{\gamma}_c^{\top})^{\top}$ e $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^{\top}, \cdots, \boldsymbol{\varepsilon}_c^{\top})^{\top}$, podemos reescrever o modelo (1.1) compactamente como

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{1.2}$$

Em geral, assume-se que $I\!\!E[\gamma] = 0, I\!\!E[\varepsilon] = 0$ com

$$\operatorname{Cov} \begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \Delta & \mathbf{0}_{cq \times n} \\ \mathbf{0}_{n \times cq} & \Sigma \end{bmatrix}, \qquad (1.3)$$

em que $\mathbf{0}_{k_1 \times k_2}$ representa uma matriz nula de ordem $k_1 \times k_2$, $\Delta \in \Sigma$ são matrizes quadradas de ordens $cq \in n = \sum_{i=1}^{c} n_i$, positivas definidas, que correspondem respectivamente, às matrizes de covariâncias dos vetores aleatórios $\gamma \in \boldsymbol{\varepsilon}$. No modelo (1.2), os efeitos fixos são usados para modelar o valor esperado da variável resposta \boldsymbol{Y} , enquanto que os efeitos aleatórios são utilizados para modelar sua estrutura de covariância. Usualmente, assume-se que $\gamma (\boldsymbol{\varepsilon})$ segue distribuição normal cq (n)-variada, com $\gamma_1, ..., \gamma_c \stackrel{\text{i.i.d.}}{\sim} N_q(\boldsymbol{0}, \boldsymbol{G})$, implicando que $\boldsymbol{\Delta} = \boldsymbol{I}_c \bigotimes \boldsymbol{G}$, com \boldsymbol{I}_c representando a matriz identidade de ordem $c \in \bigotimes$ o produto de Kronecker. Quando se atribui uma distribuição a priori para $\boldsymbol{\gamma}$, o modelo (1.2) é denominado **modelo linear geral de Bayes** [Lindley & Smith (1972)]. Fazendo $\boldsymbol{\xi} = \boldsymbol{Z} \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, obtém-se

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\xi},\tag{1.4}$$

e essas especificações implicam que $\boldsymbol{\xi}$ tem distribuição normal *n*-variada com vetor de médias $\mathbf{0}_n$ e matriz de covariâncias

$$V = Z\Delta Z^{\top} + \Sigma.$$

Em geral, supõe-se que Δ e Σ são funções de poucos parâmetros (desconhecidos) $\boldsymbol{\theta}$ que independem dos parâmetros de localização $\boldsymbol{\beta}$. Às vezes é comum colocar um parâmetro de dispersão σ^2 em evidência, ou seja, fazer $\Delta = \sigma^2 \boldsymbol{D}(\boldsymbol{\theta})$ e $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}(\boldsymbol{\theta})$, com \boldsymbol{D} e \boldsymbol{R} denotando matrizes positivas definidas, e então

$$\boldsymbol{V} = \sigma^2 \left(\boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^\top + \boldsymbol{R} \right). \tag{1.5}$$

Diferentes estruturas para $D \in \mathbf{R}$ podem ser encontradas na literatura; veja por exemplo, Rao & Kleffe (1991), Searle *et al.* (1992), Verbeke & Molenberghs (1997), Singer & Andrade (2000), Littell *et al.* (2000), Pinheiro & Bates (2000) e Rocha (2004). Quando \mathbf{R} é uma matriz diagonal, o modelo (1.2) é denominado **modelo de independência condicional**; se além disso, $\mathbf{R} = \mathbf{I}_n$ e $\Delta = \mathbf{0}_{cq \times cq}$, o modelo (1.2) corresponde ao modelo linear homocedástico usual. No presente trabalho, daremos enfâse ao **modelo de independência condicional homocedástico** [$\mathbf{R} = \mathbf{I}_n$].

Os modelos lineares mistos podem ser generalizados da mesma forma com que o modelo linear geral foi generalizado por Nelder & Weddeburn (1972). Em particular, podemos citar os modelos lineares generalizados mistos (MLGM) ou modelos lineares generalizados latentes (MLGL), em que se inclui um vetor de efeitos aleatórios γ no preditor linear; tais modelos são muito utilizados em análise de dados com medidas repetidas quando a variável resposta pertence à família exponencial. Nessa classe, modelase uma função φ do vetor de médias condicionais $\boldsymbol{\mu} = \boldsymbol{I}\!\!E[\boldsymbol{Y}|\boldsymbol{\gamma}]$ por meio de um preditor

Nobre, Juvêncio S.

linear da forma

$$\varphi(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma}; \tag{1.6}$$

além disso assume-se que a distribuição condicional de \mathbf{Y}_i dado $\boldsymbol{\gamma}$ pertence à família exponencial e que a função φ é diferenciável e monótona. Para detalhes referentes a esses modelos, veja, por exemplo, Schall (1991), Breslow & Clayton (1993), McGilchrist (1994), Kuk (1995) e McCulloch & Searle (2001, Cap. 8). Sob o enfoque Bayesiano, Lee & Nelder (1996) propuseram **modelos lineares generalizados hierárquicos** em que não é necessário supor uma distribuição normal para $\boldsymbol{\gamma}$, mas sim uma distribuição conjugada da distribuição da variável resposta \mathbf{Y} . McCullogh & Searle (2001, p. 224) mostraram como a inclusão de efeitos aleatórios no preditor linear pode ser útil em casos com superdispersão, quando a distribuição condicional de \mathbf{Y}_i dado $\boldsymbol{\gamma}$ é Poisson. Costa (2003) apresenta aplicações dos MLGM para dados longitudinais. Outras aplicações desses modelos podem ser encontradas em Breslow (1984), Williams (1992) e Tempelman & Gianola (1996), por exemplo.

Algumas notas históricas a respeito de modelos lineares mistos, podem ser encontradas em Rao & Kleffe (1991), Searle *et al.* (1992) e Pinheiro (1994).

1.3 Inferência Estatística

Vários métodos de estimação dos parâmetros do modelo (1.2) estão disponíveis na literatura; dentre eles convém destacar os métodos Bayesianos [Tountenburg (1982), Maritz & Lwin (1989) e Searle *et al.* (1992)], o método de Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR) [Patterson & Thompson (1971), Harville (1977), Robinson (1991), Searle *et al.* (1992) e Jiang (1996)] e o método de Mínimos Quadrados (MQ) [Searle *et al.*(1992), Draper & Smith (1998) e Hoffman & Vieira (1998)].

Supondo que Σ (\mathbf{R}) e Δ (\mathbf{D}) são conhecidas todos esses métodos são equivalentes, desde que no método Bayesiano seja atribuída uma distribuição a priori não informativa para γ [Hilden-Minton (1995) e Jiang (1997)]. Através do teorema de Gauss-Markov para efeitos aleatórios Harville (1976) obtém o **melhor estimador linear não viesado** ("**BLUE**best linear unbiased estimator") para β e o **melhor preditor não viesado** ("**BLUP**best linear unbiased predictor") para o vetor de efeitos aleatórios γ .

Na prática Σ (\mathbf{R}) e Δ (\mathbf{D}) são desconhecidas, e neste caso existe uma série de divergências sobre os procedimentos de "estimação" [Searle *et al.* (1992) e Hilden-Minton (1995)]. O método mais utilizado para estimar o vetor dos parâmetros de covariância $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^{\top}, \sigma^2)^{\top}$ é o de MVR, pois o correspondente viés dos estimadores dos parâmetros de covariância é menor do que aquele obtido sob os demais métodos.

A estimação de θ^* geralmente envolve equações de estimação não lineares sendo necessá-ria a utilização de métodos iterativos, tais como o **EM** [Dempster *et al.* (1977)], Newton-Raphson [Lindstrom & Bates (1988)], entre outros, para obter as estimativas. Nesses procedimentos alternam-se iterações para a estimação dos parâmetros de covariância e parâmetros de localização, β . Dempster *et al.* (1977, 1981), Laird & Ware (1982) e McLachlan & Krishnan (1997, p. 191) utilizam uma abordagem unificada, via algoritmo EM, para estimar todos os parâmetros de interesse do modelo (1.2). Propostas de implementações alternativas do algoritmo EM podem ser encontradas em Liu & Rubin (1994), McLachlan & Krishnan (1997) e Meng & van Dyk (1998). Propriedades assintóticas dos estimadores de MV e MVR dos parâmetros dos modelos (1.2) são discutidas em Miller (1977), Harville (1977), Pinheiro (1994), Jiang (1996) e Verbeke & Lesaffre (1996b), por exemplo.

A seguir apresentaremos um resumo dos principais resultados envolvendo estimadores e preditores obtidos sob o modelo (1.2) dada sua importância para as técnicas de diagnóstico.

Sejam $\widehat{\gamma} \in \widehat{\beta}$, respectivamente, o BLUP e o BLUE de $\gamma \in \beta$ então:

- $\widehat{\gamma} \in \widehat{\beta}$ são funções lineares de Y;
- $I\!\!E[\widehat{\gamma} \gamma] = 0$ e $I\!\!E[\widehat{\beta} \beta] = 0$, ou seja, $\widehat{\gamma} \in \widehat{\beta}$ são não viesados, respectivamente, para $\gamma \in \beta$;
- $\hat{\gamma}$ é o **melhor** preditor de γ e $\hat{\beta}$ é o melhor estimador de β dentro da classe dos preditores (estimadores) lineares, no sentido de que minimizam o erro quadrático médio (**EQM**) de previsão (estimação) $\mathbb{E}[(\hat{\gamma} \gamma)^{\top}(\hat{\gamma} \gamma)]$ ($\mathbb{E}[(\hat{\beta} \beta)^{\top}(\hat{\beta} \beta)]$).

Na sua gênese o BLUE e o BLUP foram descritos como os "EMV" de β e γ obtidos através da densidade conjunta do vetor aleatório (Y, γ), sob a suposição de normalidade de γ e ε , tratada como uma "verossimilhança" [Robinson (1991)]. Grenander (1981) define o BLUE e o BLUP dentro de um contexto bem mais abstrato e apresenta condições suficientes para que eles sejam unicamente definidos. Diferentes formas de obtenção, tanto do ponto de vista clássico como Bayesiano, e aplicações do BLUP e BLUE podem ser encontradas em Robinson (1991), Searle *et al.* (1992), Hilden-Minton (1995), Doganaksoy & Balakrishnan (1997), Jiang (1997) e McCulloch & Searle (2001), por exemplo.

Hilden-Minton (1995) e Hodges (1998) comentam que existe uma série de vantagens

em se utilizar os **casos com restrição** ("constraint-cases") para a obtenção do BLUE e BLUP. A idéia básica é reexpressar o modelo linear misto (1.2) na forma de um modelo linear geral através da inclusão de "casos artificiais" com variâncias desconhecidas [Hodges (1998, seção 2.2)]. Dentre as vantagens citadas em Hodges (1998), destacam-se a obtenção "imediata" das equações de estimação utilizadas para determinar o BLUE e o BLUP e a conseqüente analogia que pode ser feita com as técnicas de diagnóstico existentes para modelos lineares. Vamos adotar esse efoque a seguir. Consideremos o modelo (1.2) com a inclusão do seguinte "caso artificial",

$$\mathbf{0}_{cq\times 1} = \mathbf{0}_{cq\times 1} - \boldsymbol{I}_{cq}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{1.7}$$

simultaneamente com Var $[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{R}$, Var $[\boldsymbol{\eta}] = \sigma^2 \boldsymbol{D}$ e Cov $(\boldsymbol{\varepsilon}, \boldsymbol{\eta}^{\top}) = \mathbf{0}_{n \times cq}$, em que $\boldsymbol{\eta}$ é um vetor de dimensão $cq \times 1$ que faz o papel do "erro" na segunda equação [Hilden-Minton (1995)], e reescrevendo as equações (1.2) e (1.7) em forma matricial, temos

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\eta} \end{bmatrix}.$$
(1.8)

Pré-multiplicando (1.8) por $\mathbf{R}^{-1/2} \bigoplus \mathbf{D}^{-1/2}$, (para detalhes, veja Apêndice A.1) tem-se

$$\boldsymbol{Y}^* = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{\zeta}, \qquad (1.9)$$

em que, $\mathbf{Y}^* = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{Y} \\ \mathbf{0} \end{bmatrix}$, $\mathbf{X}^* = \begin{bmatrix} \mathbf{R}^{-1/2}\mathbf{X} & \mathbf{R}^{-1/2}\mathbf{Z} \\ \mathbf{0} & -\mathbf{D}^{-1/2} \end{bmatrix}$, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^{\top}, \boldsymbol{\gamma}^{\top})^{\top}$, tal que $\operatorname{Var}[\boldsymbol{\zeta}] = \sigma^2 \mathbf{I}_{cq+n}$. Desta forma, (1.9) pode ser considerado como um modelo linear "homocedástico". Portanto, o BLUE para $\boldsymbol{\beta}$ e o BLUP para $\boldsymbol{\gamma}$ podem ser obtidos por meio da equação (1.9), usando o método de MQ, sob o qual se obtém

$$(\boldsymbol{X}^*)^{\top} \boldsymbol{Y}^* = (\boldsymbol{X}^*)^{\top} \boldsymbol{X}^* \widehat{\boldsymbol{\beta}^*},$$

ou seja,

$$\begin{bmatrix} \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix}.$$
(1.10)

Essas equações são conhecidas na literatura como **Equações de Henderson**. Note que, se $D^{-1} \equiv 0$ (o que implica que γ é um efeito fixo) então (1.10) coincide com a equação de estimação obtida via método de **mínimos quadrados generalizados** (MQG) [Hoffman & Vieira (1998, cap. 7)]. O BLUP e o BLUE são obtidos resolvendo-se as equações (1.10), que independem da distribuição de $\gamma \in \epsilon$. Definindo

$$M = \sigma^{2} V^{-1} = (R + Z D Z^{\top})^{-1} = R^{-1} - R^{-1} Z C^{-1} Z^{\top} R^{-1}, \qquad (1.11)$$

 com

$$C = D^{-1} + Z^{\top} R^{-1} Z, \qquad (1.12)$$

mostra-se (ver Apêndice A.2) que o BLUE de $\boldsymbol{\beta}$ é

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{Y} = \left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{Y}.$$
(1.13)

Também pode-se observar que

$$\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{M}^{-1}\boldsymbol{Q}\boldsymbol{Y}, \qquad (1.14)$$

com $\boldsymbol{Q} = \boldsymbol{M} - \boldsymbol{M} \boldsymbol{X} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\top} \boldsymbol{M}$. A matriz \boldsymbol{Q} é simétrica de ordem n, e tal que

$$\boldsymbol{Q}\boldsymbol{M}^{-1}\boldsymbol{Q} = \boldsymbol{Q} \tag{1.15}$$

е

$$\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{0}.\tag{1.16}$$

Além disto, posto(Q)=n - p. O BLUP de γ (Apêndice A.2) é dado por

$$\widehat{\boldsymbol{\gamma}} = (\boldsymbol{Z}^{\top} \boldsymbol{R}^{-1} \boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}^{\top} \boldsymbol{R}^{-1} (\boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}})$$
(1.17)

$$= \boldsymbol{C}^{-1}\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}).$$
(1.18)

O BLUP e o BLUE satisfazem (veja Apêndice A.1)

$$\boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Y} = \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z}\widehat{\boldsymbol{\gamma}}.$$
 (1.19)

Uma outra identidade útil é (veja Apêndice A.1)

$$\boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{M} = \boldsymbol{C}^{-1}\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}.$$
 (1.20)

A identidade (1.20) fornece uma fórmula alternativa para o cálculo de $\hat{\gamma}$, pois considerando simultaneamente (1.14), (1.18) e (1.20), tem-se

$$\widehat{\gamma} = \boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{M}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{\Delta}\boldsymbol{Z}^{\top}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{Q}\boldsymbol{Y}.$$
(1.21)

Propriedades de $\hat{\beta}$ e $\hat{\gamma}$, são dadas em Henderson (1975), McLean *et al.* (1991), Robinson

(1991), Searle *et al.* (1992), McCulloch & Searle (2001). Algumas delas são apresentadas no Apêndice A.3.

Henderson (1975), mostrou que

$$\operatorname{Cov}\left[\begin{array}{c}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\end{array}\right] = \sigma^{2}\left[\begin{array}{cc} \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z}\\ \boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{D}^{-1}\end{array}\right]^{-1}.$$
(1.22)

Utilizando os resultados clássicos de regressão obtemos (1.22) diretamente no modelo (1.9). Discussões a respeito do processo de estimação quando X não tem posto completo e as matrizes D e R não são positivas definidas, podem ser encontradas em Henderson (1975) e Harville (1976).

No Apêndice A.1, mostra-se que

$$\mathbb{E}[\mathbf{Y}^{\top}\mathbf{Q}\mathbf{Y}] = \sigma^2(n-p), \qquad (1.23)$$

ou seja, que $\mathbf{Y}^{\top} \mathbf{Q} \mathbf{Y} / (n-p)$ é um estimador não viesado para σ^2 . Esse estimador coincide com o EMVR de σ^2 no caso linear homocedástico, uma vez que

$$\frac{\boldsymbol{Y}^{\top}\boldsymbol{Q}\boldsymbol{Y}}{n-p} = \frac{\boldsymbol{Y}^{\top}\left[\boldsymbol{I}_{n}-\boldsymbol{H}\right]\boldsymbol{Y}}{n-p} = \frac{\mathrm{SQRes}}{n-p} = \mathrm{QMRes}.$$

Como $D \in R$ dependem de um vetor de parâmetros de covariância θ^* desconhecido, é razoável calcular o BLUE e BLUP com base num estimador $\hat{\theta}^*$ de θ^* ; esses "estimadores" são denominados BLUE e BLUP empíricos (**EBLUE** e **EBLUP**). Se $\hat{\theta}^*$ é o EMV de θ^* , então o EBLUE e EBLUP, são respectivamente, o EMV de β e o preditor empírico de Bayes dos efeitos aleatórios sob a hipótese de normalidade de $\gamma \in \varepsilon$. Sob algumas condições, tanto o BLUP quanto o BLUE empíricos continuam não viesados [Kackar & Harville (1984) e Jiang (1999)]. Harville & Jeske (1992) fornecem expressões aproximadas para o EQM nesse caso. Verbeke & Lesaffre (1996b) mostram que o EBLUE e o EMV de θ^* são assintoticamente normais, mesmo quando a distribuição de γ é incorretamente especificada. Jiang (1998) obtém algumas propriedades assintóticas do EBLUP e EBLUE nos modelos de componentes de variância, considerando o EMVR para θ^* sem supor normalidade dos efeitos aleatórios e do "erro". Entre outras propriedades, ele mostra que, sob certas condições de regularidade, os EBLUP são assintoticamente independentes, o que é muito útil para diagnóstico do modelo [Jiang (1998)].

1.4 Testes de hipóteses e critérios de informação

Em geral os testes de interesse são baseados no modelo marginal $\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \boldsymbol{V})$ e utilizam estatísticas de Wald ou da Razão de Verossimilhanças (RV).

Assintoticamente, sob a hipótese nula, a estatística de Wald tem uma distribuição χ_r^2 , com r representando a correspondente redução no número de parâmetros; tal estatística não é apropriada para casos em que o tamanho da amostra é pequeno, a distribuição dos efeitos aleatórios é assimétrica ou a hipótese a ser testada encontra-se na fronteira do espaço paramétrico. Quando o interesse é testar contrastes do tipo $C\beta = 0$, com Crepresentando uma matriz de dimensão $k_1 \times p$, a estatística do teste é

$$\xi_W = (\boldsymbol{C}\widehat{\boldsymbol{\beta}})^\top [\boldsymbol{C} Var(\widehat{\boldsymbol{\beta}})\boldsymbol{C}^\top]^{-1} \boldsymbol{C}\widehat{\boldsymbol{\beta}}, \qquad (1.24)$$

e sua distribuição aproximada é $\chi^2_{posto(c)}$. Dividindo-se (1.24) por posto(C), obtém-se uma estatística com distribuição aproximada $F_{(posto(c),k)}$, com o número de graus de liberdade do denominador k sendo obtido através de aproximação. Diferentes aproximações para k são discutidas em Fai & Cornelius (1996) e Verbeke & Molenberghs (1997), por exemplo.

O teste da RV pode ser utilizado para testar a hipótese nula de que o modelo com mais parâmetros não se ajusta significativamente melhor do que um modelo restrito (com um número reduzido de parâmetros). A estatística da RV é dada por

$$\xi_{RV} = -2(L_1 - L_2), \tag{1.25}$$

com L_1 representando o máximo da log-verossimilhança sob o modelo restrito (encaixado) e L_2 a respectiva função correspondente do modelo com r parâmetros adicionais. Quando o modelo reduzido não se situa na fronteira do espaço paramétrico, tem-se que $\xi_{RV} \sim \chi_r^2$. Self & Liang (1987) mostram que quando o modelo reduzido se situa na fronteira do espaço paramétrico, então a distribuição assintótica de (1.25) é uma mistura de distribuições χ^2 . O teste da RV não é apropriado para testar hipóteses referentes aos efeitos fixos quando se utiliza a log-verossimilhança restrita, uma vez que ela exclui tais efeitos. Recentemente Verbeke & Molenberghs (2003) utilizaram o teste "Score" e observaram os mesmos "problemas" dos testes de Wald e da RV.

Quando os modelos não são encaixados ou quando a hipótese de interesse situa-se na fronteira do espaço paramétrico, podem-se utilizar alguns critérios de informação fundamentados na teoria da decisão que penalizam os modelos com um grande número de parâmetros. Alguns desses critérios são baseados nas estatísticas AIC (Akaike Information Criterion), o BIC (Bayesian Information Criterion) e o CAIC (Consistent Akaike's Information Criterion) definidos como

$$AIC = -2l + 2d \tag{1.26}$$

$$BIC = -2l + d\ln n \tag{1.27}$$

CAIC =
$$-2l + d(\ln n + 1),$$
 (1.28)

com l representando o máximo da log-verossimilhança (completa ou restrita), d o número de parâmetros do modelo e n o número de observações. Quanto menor for o valor dessas estatísticas, maior evidência favorável ao modelo em questão.

Detalhes sobre testes de hipóteses e critérios de seleção para modelos lineares mistos podem ser encontrados em Bozdogan (1987), Andreoni (1989), Öfversten (1993), Stram & Lee (1994), Suyama (1995), Christensen (1996), Verbeke & Molenberghs (1997), Pinheiro & Bates (2000), dentre outros.

Análise de Resíduos

Resíduos são utilizados para avaliar a validade das suposições de modelos estatísticos. Por exemplo, no caso linear normal, utilizam-se os resíduos padronizados para verificar homocedasticidade, existência de pontos discrepantes, normalidade e independência dos erros. Cox & Snell (1968) apresentam uma forma geral para definir resíduos para modelos com uma única fonte de variação. Como no modelo linear misto, existe mais de uma fonte de variação, e conseqüentemente mais de um tipo de resíduo, tal definição não pode ser utilizada. No presente capítulo discutiremos algumas propostas de utilização dos diferentes tipos de resíduos associados ao ajuste do modelo (1.1) para avaliar possíveis afastamentos das suposições e detectar a existência de observações e/ou unidades experimentais discrepantes.

2.1 Tipos de resíduos

Sob o modelo (1.1) podemos definir três tipos de vetores de erros:

- Erros condicionais: $\boldsymbol{\varepsilon} = \boldsymbol{Y} \mathbb{E}[\boldsymbol{Y}|\boldsymbol{\gamma}] = \boldsymbol{Y} \boldsymbol{X}\boldsymbol{\beta} \boldsymbol{Z}\boldsymbol{\gamma};$
- Effitos aleatórios: $Z\gamma = \mathbb{E}[Y|\gamma] \mathbb{E}[Y];$
- Erros marginais: $\boldsymbol{\xi} = \boldsymbol{Y} \mathbb{E}[\boldsymbol{Y}] = \boldsymbol{Y} \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}.$

Os correspondentes valores preditos, denominados resíduos, são dados respectivamente por $\hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\gamma}, \mathbf{Z}\hat{\gamma} \in \hat{\boldsymbol{\xi}} = \hat{\boldsymbol{r}} = \mathbf{Y} - \mathbf{X}\hat{\beta}, \operatorname{com} \hat{\beta} \in \hat{\gamma}$ representando, respectivamente, o BLUE de β e o BLUP de γ . Cada tipo de resíduo é útil para avaliar algum tipo de suposição do modelo (1.1). Por exemplo, para avaliar a suposição de linearidade da relação entre $\mathbb{E}[\mathbf{Y}]$ e as covariáveis \mathbf{X} , Hilden-Minton (1995) sugere construir um gráfico dos resíduos $\hat{\boldsymbol{\xi}}$ contra os valores das covariáveis. Espera-se que os elementos de $\hat{\boldsymbol{\xi}}$ variem aleatoriamente em torno de zero sob a veracidade dessa suposição. Como $\operatorname{Var}[\boldsymbol{\xi}] = \mathbf{V}$, então o resíduo $\hat{\boldsymbol{\xi}}$ também pode ser útil para avaliar a validade da estrutura de covariâncias [Weiss (1995) e Rocha (2004)]. Lesaffre & Verbeke (1998) utilizaram $\hat{\boldsymbol{\xi}}$ para esse fim, como veremos adiante.

Utilizando (1.11) e (1.14) podemos concluir que

$$RQY = RM(Y - X\widehat{\beta})$$

= $(I_n - ZC^{-1}Z^{\top}R^{-1})(Y - X\widehat{\beta}) = Y - X\widehat{\beta} - Z\widehat{\gamma} = \hat{\varepsilon}$ (2.1)

e que $\hat{\boldsymbol{\xi}} = \boldsymbol{M}^{-1} \boldsymbol{Q} \boldsymbol{Y}$. Utilizando (1.15), (1.16) e (1.21) tem-se também que

$$\operatorname{Var}[\hat{\boldsymbol{\varepsilon}}] = \sigma^2 \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R}, \qquad (2.2)$$

$$\operatorname{Var}[\hat{\boldsymbol{\xi}}] = \sigma^2 \boldsymbol{M}^{-1} \boldsymbol{Q} \boldsymbol{M}^{-1}, \qquad (2.3)$$

$$\operatorname{Var}[\boldsymbol{Z}\widehat{\boldsymbol{\gamma}}] = \sigma^2 \boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^\top \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^\top.$$
(2.4)

Para Hilden-Minton (1995) o resíduo *puro*, é aquele que depende apenas das componentes fixas do modelo e do respectivo erro do qual ele é preditor. Já um resíduo que depende de dois ou mais erros é denominado resíduo *confundido*. Note que (Apêndice B.1) sob a validade do modelo, temos

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{R}\boldsymbol{Q}\boldsymbol{\varepsilon} + \boldsymbol{R}\boldsymbol{Q}\boldsymbol{Z}\boldsymbol{\gamma},$$
(2.5)

$$Z\widehat{\gamma} = ZDZ^{\top}QZ\gamma + ZDZ^{\top}Q\varepsilon, \qquad (2.6)$$

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} = -\boldsymbol{X} (\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{\xi}.$$
(2.7)

De (2.5) e (2.6) concluímos que $\hat{\boldsymbol{\varepsilon}} \in \boldsymbol{Z} \hat{\boldsymbol{\gamma}}$ são resíduos *confundidos* pela presença de $\boldsymbol{\gamma}$ e $\boldsymbol{\varepsilon}$, respectivamente. Se $\boldsymbol{Z} \in \mathbb{C}(\boldsymbol{X})$, com $\mathbb{C}(\boldsymbol{X})$ representando o subespaço gerado pelas colunas da matriz \boldsymbol{X} , então $\boldsymbol{Q} \boldsymbol{Z} = \boldsymbol{0}$ e nesse caso os resíduos são *puros*. Quando o interesse é verificar a suposição de normalidade para o erro $\boldsymbol{\varepsilon}$, não é aconselhável utilizar $\hat{\boldsymbol{\varepsilon}}$, porque ele é confundido por $\boldsymbol{\gamma}$; logo, quando $\boldsymbol{\gamma}$ se afasta muito da normalidade, $\hat{\boldsymbol{\varepsilon}}$ pode não apresentar características de normalidade, mesmo quando $\boldsymbol{\varepsilon}$ segue uma distribuição normal.

2.2 Utilização do resíduo condicional

Pinheiro & Bates (2000, p.175) sugerem o uso de gráficos de $\hat{\boldsymbol{\varepsilon}}$ versus $\hat{\boldsymbol{Y}}$ e Q-Q para avaliar as suposições de homocedasticidade e normalidade do erro condicional. O resíduo $\hat{\boldsymbol{\varepsilon}}$ também pode ser utilizado para identificar observações *discrepantes*. Propostas semelhantes para avaliar homocesdaticidade por meio do resíduo condicional são dadas em Weiss & Lazaro (1992) e Oman (1995). Como os elementos de $\hat{\boldsymbol{\varepsilon}}$ podem ter variâncias distintas, sugerimos padronizá-los, ou seja, considerar

$$\hat{\boldsymbol{\varepsilon}}_{i}^{*} = \frac{\hat{\boldsymbol{\varepsilon}}_{i}}{\sigma\sqrt{q_{ii}}},\tag{2.8}$$

com $\hat{\boldsymbol{\varepsilon}}_i$ representando o *i-ésimo* componente de $\hat{\boldsymbol{\varepsilon}}$, q_{ii} o *i-ésimo* elemento da diagonal principal de $\boldsymbol{Q} \in \boldsymbol{R} = \boldsymbol{I}_n$. Para motivar o uso de (2.8) na identificação de observações discrepantes, considere o estimador não viesado para σ^2 , obtido quando eliminamos da amostra um conjunto $I = \{i_1, i_2, ..., i_k\}$ $(1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq n)$, denotado por $\hat{\sigma}^2_{(I)}$. Com base em (1.23), obtém-se

$$\hat{\sigma}_{(I)}^2 = \frac{\boldsymbol{Y}^{\top} (\boldsymbol{Q} - \boldsymbol{Q} \boldsymbol{U}_I (\boldsymbol{U}_I^{\top} \boldsymbol{Q} \boldsymbol{U}_I)^{-1} \boldsymbol{U}_I^{\top} \boldsymbol{Q}) \boldsymbol{Y}}{n - p - k}, \qquad (2.9)$$

com $U_I = (u_{ij})_{n \times k} = (U_{i_1}, U_{i_2}, ..., U_{i_k})$ em que U_i denota a *i-ésima* coluna da matriz I_n . Quando eliminamos a *i-ésima* observação, lembrando que $\mathbf{R} = I_n$, tem-se $(\hat{\boldsymbol{\varepsilon}}_i)^2/q_{ii} = \mathbf{Y}^{\top} \mathbf{Q} U_I (\mathbf{U}_I^{\top} \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^{\top} \mathbf{Q} \mathbf{Y}$, e por (2.9)

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} = \frac{(n-p-1)\hat{\sigma}_{(i)}^2}{\sigma^2} + \frac{(\hat{\boldsymbol{\varepsilon}}_i)^2}{\sigma^2 q_{ii}}$$

implicando

$$\frac{\hat{\sigma}_{(i)}^2}{\hat{\sigma}^2} = \left(\frac{n-p-\hat{\varepsilon}_i^2/q_{ii}}{n-p-1}\right),\tag{2.10}$$

que é uma função decrescente de $|\hat{\boldsymbol{\varepsilon}}_i^*|$. Assim, os resíduos condicionais padronizados (2.8) são úteis para identificar observações com alta influência na estimativa de σ^2 . No caso linear normal tal resíduo serve para testar se a *i-ésima* observação é discrepante [Cook & Weisberg (1982)]. Para o caso em que eliminamos um conjunto I com k observações, tem-se que

$$\frac{\hat{\sigma}_{(I)}^2}{\hat{\sigma}^2} = \left(\frac{n-p-M_I}{n-p-k}\right),\tag{2.11}$$

com $M_I = \mathbf{Y}^{\top} \mathbf{Q} \mathbf{U}_I (\mathbf{U}_I^{\top} \mathbf{Q} \mathbf{U}_I)^{-1} \mathbf{U}_I^{\top} \mathbf{Q} \mathbf{Y}$. Equivalentemente ao caso anterior, um valor grande de M_I sugere a existência de ao menos uma observação aberrante no conjunto I. Uma vez que não conhecemos $\sigma^2 \in \mathbf{Q}$, os resíduos acima definidos são calculados com base nas suas respectivas estimativas.

Podemos avaliar a suposição de homocedasticidade por meio do gráfico dos elementos (2.8) versus os correspondentes valores preditos. Para avaliar a hipótese de normalidade

de ε a partir de (2.8) o problema é mais complicado, dado que ele é confundido por γ . Considerando (2.5), Hilden-Minton (1995) comenta que a habilidade para avaliar a normalidade de ε diminui quando Var $[RQZ^{\top}\gamma] = \sigma^2 RQZDZ^{\top}QR$ cresce em relação a Var $[RQ\varepsilon] = \sigma^2 RQRQR$. Esse autor define a fração de confudimento para $\hat{\varepsilon}_i$ como

$$0 \le CF_i = \frac{\boldsymbol{U}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{U}_i}{\boldsymbol{U}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{U}_i} = 1 - \frac{\boldsymbol{U}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{U}_i}{\boldsymbol{U}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{U}_i} \le 1,$$
(2.12)

que representa a proporção da variabilidade de $\hat{\boldsymbol{\varepsilon}}_i$ devida ao confundimento com o BLUP. Quanto maior for (2.12) maior é o grau de confundimento de $\hat{\boldsymbol{\varepsilon}}_i$.

Hilden-Minton (1995) sugere utilizar uma transformação linear $\boldsymbol{L}^{\top} \hat{\boldsymbol{\varepsilon}}$, que minimize o confundimento em algum sentindo. Denotando as linhas de \boldsymbol{L} por \boldsymbol{l}_i (i = 1, ..., n), uma sugestão é minimizar o confundimento de $\boldsymbol{l}_i^{\top} \hat{\boldsymbol{\varepsilon}}$, ou seja maximizar

$$\lambda_i = \frac{\boldsymbol{l}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{l}_i}{\boldsymbol{l}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{l}_i}, \qquad (2.13)$$

sujeito à restrição $\operatorname{Var}[\boldsymbol{l}_i^{\top} \hat{\boldsymbol{\varepsilon}}] \propto \boldsymbol{l}_i^{\top} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{l}_i > 0$. Como a matriz \boldsymbol{R} tem posto completo, o foco é o espaço não-nulo da matriz semi-positiva definida \boldsymbol{Q} . Considerando a decomposição espectral [Harville (1997, p. 515)]

$$\boldsymbol{R}^{1/2}\boldsymbol{Q}\boldsymbol{R}^{1/2} = \boldsymbol{K}\boldsymbol{\Pi}\boldsymbol{K}^{\top},$$

com \boldsymbol{K} denotando uma matriz $n \times (n-p)$; $\boldsymbol{K}^{\top} \boldsymbol{K} = \boldsymbol{I}_{n-p}$ e $\boldsymbol{\Pi}$ denotando uma matriz quadrada diagonal de ordem (n-p). Além disso, seja $\boldsymbol{l}_i = \boldsymbol{R}^{-1/2} \boldsymbol{K} \boldsymbol{\Pi}^{-1/2} \boldsymbol{v}_i$, para algum vetor \boldsymbol{v}_i de dimensão $(n-p) \times 1$; então (2.13) pode ser escrita como

$$\lambda_i = \frac{\boldsymbol{v}_i^\top \boldsymbol{\Pi} \boldsymbol{v}_i}{\boldsymbol{v}_i^\top \boldsymbol{v}_i},\tag{2.14}$$

que implica [Graybill (1983, p. 409)] $\pi_{n-p} \leq \lambda_i \leq \pi_1$, com $\pi_{n-p} \leq \cdots \leq \pi_1 \leq 1$ representando os elementos ordenados de Π (auto-valores não nulos de $\mathbf{R}^{1/2}\mathbf{Q}\mathbf{R}^{1/2}$). Considerando \mathbf{v}_i igual a *i-ésima* coluna de \mathbf{I}_{n-p} tem-se que

$$\boldsymbol{l}_{i} = \boldsymbol{R}^{-1/2} \boldsymbol{K} \boldsymbol{\Pi}^{-1/2} \boldsymbol{v}_{i} = \pi_{i}^{-1/2} \boldsymbol{R}^{-1/2} \boldsymbol{K}_{i} \quad (i = 1, ..., n - p),$$
(2.15)

em que \mathbf{K}_i representa a *i-ésima* coluna de \mathbf{K} . Note que $(\mathbf{l}_1, ..., \mathbf{l}_{n-p})$ formam uma base ortogonal do espaço não nulo de \mathbf{Q} . Pode-se mostrar que (Apêndice B.2)

$$\boldsymbol{l}_{i}^{\mathsf{T}} \hat{\boldsymbol{\varepsilon}} = \sqrt{\pi_{i}} \boldsymbol{K}_{i}^{\mathsf{T}} \boldsymbol{R}^{-1/2} \boldsymbol{Y}, \qquad (2.16)$$

$$\operatorname{Cov}[\boldsymbol{l}_i^{\top}\hat{\boldsymbol{\varepsilon}}, \boldsymbol{l}_j\hat{\boldsymbol{\varepsilon}}] = \sigma^2 \delta_{ij} \quad (i, j = 1, ..., n - p), \qquad (2.17)$$

Nobre, Juvêncio S.

IME-USP

com $\delta_{ij} = 1$ se i = j e zero em caso contrário. Assim, $(\boldsymbol{l}_i^{\top} \hat{\boldsymbol{\varepsilon}})/\sigma$ são resíduos padronizados, não correlacionados com fração de confundimento (2.12) igual a $1 - \pi_i$. Denominaremos esses resíduos por **resíduos com confundimento mínimo**. Hilden-Minton (1995) sugere avaliar a hipótese de normalidade do erro condicional através dos resíduos com confundimento mínimo $(\boldsymbol{l}_i^{\top} \hat{\boldsymbol{\varepsilon}})/\sigma$ (i = 1, ..., n - p) por intermédio do gráfico Q-Q com envelope simulado [Atkinson (1985)].

2.3 Utilização do EBLUP

Considerando o modelo (1.1), $\mathbf{Z}_i \hat{\boldsymbol{\gamma}}_i$ reflete a diferença entre o valor predito e o valor médio populacional predito para a *i-ésima* unidade experimental; logo pode-se utilizálo para encontrar possíveis unidades experimentais discrepantes, conforme sugerido em Waternaux *et al.* (1989), Verbeke (1995), Verbeke & Lesaffre (1996a), Pinheiro & Bates (2000) e Longford (2001), por exemplo. Pinheiro & Bates (2000), por exemplo, sugerem utilizar os gráficos dos elementos dos $\hat{\boldsymbol{\gamma}}_i$, digamos $\hat{\gamma}_{ij}$ (j = 1, ..., q) versus o índice das unidades experimentais. Tal gráfico é útil para identificar unidades experimentais que apresentam um valor discrepante para o *j-ésimo* elemento do seu respectivo BLUP. Levando em consideração que os $\hat{\boldsymbol{\gamma}}_i$ (i = 1, ..., c) são comparáveis apenas quando as covariáveis de \mathbf{Z}_i são iguais para todas as unidades experimentais [Verbeke & Lesaffre (1996a)], podemos utilizar o gráfico dos elementos de $\mathbf{Z}_i \hat{\boldsymbol{\gamma}}_i$, ou então utilizar a **distância de Mahalanobis**, $\zeta_i = \hat{\boldsymbol{\gamma}}_i^{\top} \widehat{\mathrm{Var}}[\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i] \hat{\boldsymbol{\gamma}}_i$ proposta por Waternaux *et al.* (1989) para encontrar unidades experimentais discrepantes. Sob a validade do modelo, tem-se $\zeta_i \approx \chi_{n_i}^2$ para n_i suficientemente grande.

Para verificar a plausibilidade da estrutura adotada para a matriz de covariâncias G dos efeitos aleatórios, Pinheiro & Bates (2000, p.187) sugerem utilizar o gráfico de dispersão múltiplo dos elementos dos BLUP. No Capítulo 3, é proposto um meio para avaliar quais unidades experimentais são sensíveis à hipótese de homogeneidade entre as matrizes de covariâncias dos efeitos aleatórios.

Os EBLUP também servem para avaliar a suposição de normalidade do vetor de efeitos aleatórios γ . Lange & Ryan (1989) sugerem a utilização de um gráfico Q-Q ponderado pelas variâncias dos elementos de γ para avaliar a hipótese de normalidade dos efeitos aleatórios. Algumas críticas a respeito da proposta de Lange & Ryan (1989) são feitas em Hilden-Minton (1995), Verbeke (1995) e Verbeke & Molenberghs (1997). Jiang (2001) propõe um teste de aderência para avaliar a hipótese de que as distribuições de γ e ε são como especificadas; ele mostra que a distribuição nula assintótica do teste é

19

uma mistura de distribuições qui-quadrado. As propostas dos dois artigos supracitados são válidas assintoticamente. Para obtenção do BLUE (1.13) e BLUP (1.17) não utilizamos a suposição de normalidade; tal suposição só é utilizada para encontrar o EMV dos parâmetros de covariância e seus respectivos erros-padrão. Uma alternativa é utilizar a função *score* obtida sob a suposição de normalidade de γ e ε para obter o respectivo EMV; tal procedimento é utilizado no método de MVR [Jiang (1996)].

Considerando que o vetor de médias $X\beta$ e a matriz de covariâncias V estão corretamente especificados, Butler & Louis (1992), mostraram via simulação, que o BLUE não é afetado pela má especificação da distribuição de γ . Tal resultado foi confirmado teoricamente por Verbeke & Lesaffre (1996b) que mostraram que as estimativas do modelo (1.1) obtidas sob hipótese de normalidade são assintoticamente consistentes mesmo quando a distribuição de γ não é normal mas tem terceiro momento absoluto finito, sendo necessário apenas uma correção na matriz de covariâncias; essa condição é válida para as distribuições gama, log-normal, Weibull, *t*-Student (se o número de graus de liberdade for maior que 3), Poisson, dentre outras.

Denotando por $L(\boldsymbol{\psi})$ a log-verossimilhança do modelo (1.1) sob a hipótese de normalidade, $\boldsymbol{\psi}$, o respectivo vetor de parâmetros, $\boldsymbol{U}(\boldsymbol{\psi}) = \partial L(\boldsymbol{\psi})/\partial \boldsymbol{\psi}$, o vetor score, $\boldsymbol{A}(\boldsymbol{\psi}) = \partial^2 L(\boldsymbol{\psi})/\partial \boldsymbol{\psi}^\top \partial \boldsymbol{\psi}$ e $\boldsymbol{B}(\boldsymbol{\psi}) = \boldsymbol{U}(\boldsymbol{\psi})\boldsymbol{U}(\boldsymbol{\psi})^\top$, então um estimador robusto da matriz de covariâncias do EMV $\hat{\boldsymbol{\psi}}$ [Verbeke & Lesaffre (1996b)] é $\widehat{\operatorname{Var}}[\hat{\boldsymbol{\psi}}] = \boldsymbol{A}(\hat{\boldsymbol{\psi}})^{-1}\boldsymbol{B}(\hat{\boldsymbol{\psi}})\boldsymbol{A}(\hat{\boldsymbol{\psi}})^{-1}$. Esse estimador é conhecido como "estimador sanduíche".

Se o modelo é corretamente especificado, tem-se que $\mathbf{A}(\hat{\psi}) \approx \mathbf{B}(\hat{\psi})$, implicando $\widehat{\operatorname{Var}}[\hat{\psi}] \approx \mathbf{A}(\hat{\psi})^{-1}$ que é a estimativa usual da matriz de covariâncias de $\hat{\psi}$. Assim, se $\lambda_{\min} \approx \lambda_{\max} \approx 1$, com λ_{\min} e λ_{\max} denotando, respectivamente, o menor e o maior autovalor da matriz $\mathbf{B}(\hat{\psi})\mathbf{A}(\hat{\psi})^{-1}$, temos indício de que o vetor de efeitos aleatórios tem distribuição normal. Em geral os erros-padrão robusto e não robusto (não corrigido) são muito similares para os BLUE, o que não ocorre para os erros-padrão dos estimadores dos parâmetros de covariância que tendem a serem subestimados pelos erros-padrão não corrigidos [Verbeke & Lesaffre (1997)]. Outras aproximações para os erros-padrão das estimativas das componentes de variância, obtidas sem a suposição de normalidade, estão implementadas no procedimento MIXED do SAS [SAS Institute Inc. (1997)].

No contexto de EEG, o estimador $A(\hat{\psi})^{-1}$ é conhecido como estimador "baseado no modelo" (model-based) ou "ingênuo" (naive) e é consistente apenas quando o modelo está corretamente especificado; já o estimador robusto (sanduíche) é sempre consistente, porém pode apresentar um alto vício quando o número de unidades experimentais é pequeno.

Capítulo ${\bf 3}$

Análise de Sensibilidade em Modelos Lineares Mistos

3.1 Introdução

A análise de sensibilidade consiste em estudar o comportamento do modelo ajustado quando ele está sujeito a algum tipo de perturbação, ou seja, sob alguma mudança nas hipóteses ou nos dados. Avaliar a influência das observações no modelo ajustado é importante; todavia sabe-se que uma observação não tem a mesma influência em todos os resultados. Uma pergunta natural é "avaliar influência em que?". Esta pergunta deve ser respondida por meio da definição do objetivo da pesquisa; por exemplo, se o objetivo é fazer previsões, então é razoável medir a influência nos valores preditos e não nos parâmetros de localização [Chatterjee & Hadi (1986, 1988)].

Existem medidas de influência baseadas nos resíduos, na curva de influência, na verossimilhança, no volume dos elipsóides de confiança, em um subconjunto do vetor de parâmetros de localização (influência parcial) e nos pontos remotos do espaço vetorial gerado pelas colunas da matriz de especificação X. Para detalhes e exemplos, dentro do contexto de modelos lineares, veja Belsley *et al.* (1980), Cook & Weisberg (1982) e Chatterjee & Hadi (1986,1988).

Dentre as abordagens mais utilizadas na prática, para medir influência em modelos lineares mistos, destacam-se as análises baseadas em **influência local** [Cook (1986)] e aquelas obtidas via eliminação de observações (**influência global**). Nas próximas seções discutimos algumas propostas de análise de sensibilidade no contexto dos modelos estudados aqui.

3.2 Inclusão de efeitos fixos

Considere o modelo (1.1), com as matrizes $D \in \mathbf{R}$ conhecidas e contendo inicialmente

apenas os efeitos fixos $\boldsymbol{\beta}_1$, isto é

$$\boldsymbol{Y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\xi}, \qquad (3.1)$$

em que $\xi = Z\gamma + \varepsilon$. Há interesse em adicionar novos efeitos fixos β_2 ao modelo (3.1), ou seja, ajustar o modelo

$$Y = X\beta + \xi$$

= $X_1\beta_1 + X_2\beta_2 + \xi$, (3.2)

 $\operatorname{com} \boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathsf{T}}, \boldsymbol{\beta}_2^{\mathsf{T}})^{\mathsf{T}} \operatorname{e} \boldsymbol{X} = [\boldsymbol{X}_1; \boldsymbol{X}_2] \operatorname{de} \operatorname{posto} \operatorname{completo} \operatorname{e} \operatorname{tal} \operatorname{que} \operatorname{posto}(\boldsymbol{X}) = p = \operatorname{posto}(\boldsymbol{X}_1) + \operatorname{posto}(\boldsymbol{X}_2) = p_1 + p_2. \operatorname{De} (1.13), \operatorname{tem-se} \operatorname{que} \operatorname{o} \operatorname{BLUE} \operatorname{de} \boldsymbol{\beta}_1, \operatorname{relativamente} \operatorname{ao} \operatorname{modelo} (3.1),$ é dado por

$$\widehat{\boldsymbol{\beta}}_{1}^{0} = \left(\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\right)^{-1}\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{Y}, \qquad (3.3)$$

enquanto que o BLUE de $\pmb{\beta}_2$ (veja Apêndice C.1) referente ao modelo (3.2) é

$$\widehat{\boldsymbol{\beta}}_{2} = \left(\boldsymbol{X}_{2}^{\top}\boldsymbol{Q}_{1}\boldsymbol{X}_{2}\right)^{-1}\boldsymbol{X}_{2}^{\top}\boldsymbol{Q}_{1}\boldsymbol{Y}, \qquad (3.4)$$

 $\operatorname{com}\,\boldsymbol{Q}_1 = \boldsymbol{M} - \boldsymbol{M}\boldsymbol{X}_1 \left(\boldsymbol{X}_1^\top \boldsymbol{M}\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1^\top \boldsymbol{M} \text{ tal que}$

$$Q_1 X_1 = 0 \ e \ Q_1 M^{-1} Q_1 = Q_1.$$
 (3.5)

Considerando a expressão (1.19) sob o modelo (3.2) e as propriedades (3.5), obtemos

$$\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{Y} = \boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\widehat{\boldsymbol{\beta}}_{1} + \boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{2}\widehat{\boldsymbol{\beta}}_{2}, \qquad (3.6)$$

com $\hat{\boldsymbol{\beta}}_1$ representando o BLUE de $\boldsymbol{\beta}_1$ sob o modelo (3.2). Pré-multiplicando (3.6) por $(\boldsymbol{X}_1^{\top}\boldsymbol{M}\boldsymbol{X}_1)^{-1}$ e considerando (3.3), temos

$$\widehat{\boldsymbol{\beta}}_{1}^{0} = \widehat{\boldsymbol{\beta}}_{1} + \left(\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\right)^{-1}\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{2}\widehat{\boldsymbol{\beta}}_{2}, \qquad (3.7)$$

de onde se obtém a seguinte fórmula de *atualização* para o BLUE de β_1 ,

$$\widehat{\boldsymbol{\beta}}_{1} - \widehat{\boldsymbol{\beta}}_{1}^{0} = -\left(\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\right)^{-1}\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{2}\widehat{\boldsymbol{\beta}}_{2}.$$
(3.8)

Denotando por $\hat{\gamma}^0$ e $\hat{\gamma}$ os BLUP referentes ao vetor de efeitos aleatórios dos modelos (3.1) e (3.2), respectivamente, mostra-se (vide Apêndice C.2) que

$$\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}^0 = -\boldsymbol{D}\boldsymbol{Z}^\top \boldsymbol{Q}_1 \boldsymbol{X}_2 \widehat{\boldsymbol{\beta}}_2.$$
(3.9)

Nobre, Juvêncio S.

IME-USP

As fórmulas de atualização (3.8) e (3.9) do BLUE e BLUP, respectivamente, para inclusão de novas variáveis com "efeitos fixos" no modelo linear misto foram obtidas por Hilden-Minton (1995, cap.3).

Para incluirmos um efeito aleatório, o procedimento é bem mais complicado, uma vez que devemos atualizar (aumentar) a matriz de covariâncias dos efeitos aleatórios D e para isso é necessário conhecer sua estrutura [Rocha (2004)]. Além do mais, este processo de *atualização* envolve equações não lineares, ao contrário do processo já discutido. No presente trabalho não trataremos desse caso.

3.3 Gráfico da variável adicionada para efeitos fixos

É comum utilizar gráficos de variáveis adicionadas [Johnson & McCulloch (1987)] para se ter idéia sobre a sua inclusão no modelo, estudar o tipo de relação (linear/não linear) existente com a variável resposta, avaliar se o respectivo coeficiente é significativo devido a influência de poucas observações, etc.

Considereremos o caso em que X_2 sob (3.2) tem uma única coluna, ou seja em que há interesse em incluir um único efeito fixo. Mostra-se que (Apêndice C.3)

$$M^{-1/2}Q_{1}Y = M^{1/2}(Y - X_{1}\hat{\beta}_{1}) = M^{-1/2}R^{-1}(Y - X_{1}\hat{\beta}_{1}^{0} - Z\hat{\gamma}^{0})$$
(3.10)

е

$$M^{-1/2}Q_{1}X_{2} = M^{-1/2}(X_{2} - X_{1}\widehat{\beta}_{1}^{*}) = M^{-1/2}R^{-1}(X_{2} - X_{1}\widehat{\beta}_{1}^{*} - Z\widehat{\gamma}^{*}), \quad (3.11)$$

com $\widehat{\boldsymbol{\beta}}_{1}^{*}$ e $\widehat{\boldsymbol{\gamma}}^{*}$ representando, respectivamente, o BLUE e o BLUP de $\boldsymbol{\beta}_{1}$ e $\boldsymbol{\gamma}$ sob o modelo $\boldsymbol{X}_{2} = \boldsymbol{X}_{1}\boldsymbol{\beta}_{1}^{*} + \boldsymbol{Z}\boldsymbol{\gamma}^{*} + \boldsymbol{\varepsilon}^{*}$, com $\boldsymbol{\gamma}^{*}$ e $\boldsymbol{\varepsilon}^{*}$ distribuídos como $\boldsymbol{\gamma}$ e $\boldsymbol{\varepsilon}$ sob o modelo (3.2). Considerando (3.4), (3.5), (3.10) e (3.11) temos

$$\widehat{\boldsymbol{\beta}}_{2} = \frac{\boldsymbol{X}_{2}^{\top} \boldsymbol{Q}_{1} \boldsymbol{Y}}{\boldsymbol{X}_{2}^{\top} \boldsymbol{Q}_{1} \boldsymbol{X}_{2}} = \frac{\boldsymbol{X}_{2}^{\top} \boldsymbol{Q}_{1} \boldsymbol{M}^{-1} \boldsymbol{Q}_{1} \boldsymbol{Y}}{\boldsymbol{X}_{2}^{\top} \boldsymbol{Q}_{1} \boldsymbol{M}^{-1} \boldsymbol{Q}_{1} \boldsymbol{X}_{2}} = \frac{(\boldsymbol{M}^{-1/2} \boldsymbol{Q}_{1} \boldsymbol{X}_{2})^{\top} (\boldsymbol{M}^{-1/2} \boldsymbol{Q}_{1} \boldsymbol{Y})}{(\boldsymbol{M}^{-1/2} \boldsymbol{Q}_{1} \boldsymbol{X}_{2})^{\top} (\boldsymbol{M}^{-1/2} \boldsymbol{Q}_{1} \boldsymbol{X}_{2})} = \frac{\boldsymbol{r}_{1}^{\top} \boldsymbol{r}_{2}}{\boldsymbol{r}_{1}^{\top} \boldsymbol{r}_{1}},$$
(3.12)

com $\mathbf{r}_1 = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{X}_2 = \mathbf{M}^{-1/2} \mathbf{R}^{-1} (\mathbf{X}_2 - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^* - \mathbf{Z} \widehat{\boldsymbol{\gamma}}^*)$ e $\mathbf{r}_2 = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{Y} = \mathbf{M}^{-1/2} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}_1 \widehat{\boldsymbol{\beta}}_1^0 - \mathbf{Z} \widehat{\boldsymbol{\gamma}}^0)$. Portanto, (3.12) pode ser interpretado como o coeficiente da regressão linear sem intercepto do resíduo \mathbf{r}_2 sobre o resíduo \mathbf{r}_1 . Se existir alguma relação linear entre \mathbf{r}_1 e \mathbf{r}_2 , a variável \mathbf{X}_2 deve ser incluída no modelo de forma linear.
Quando X_2 tem uma única coluna, Hilden-Minton (1995) sugere utilizar o gráfico de $M^{-1/2}Q_1Y$ versus $M^{-1/2}Q_1X_2$ como gráfico da variável adicionada.

Se $\gamma = 0$ e $\mathbf{R} = \mathbf{I}_n$ o gráfico de $\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{Y} = \mathbf{M}^{1/2}(\mathbf{Y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1) = (\mathbf{Y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1)$ versus $\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{X}_2 = \mathbf{M}^{1/2}(\mathbf{X}_2 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1^*) = (\mathbf{X}_2 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1^*)$ coincide com o gráfico da variável adicionada utilizado no caso linear clássico [Johnson & McCulloch (1987)]. Após a decisão sobre as variáveis a serem consideradas no modelo, Besley *et al.* (1980, p.30) e O'Hara-Hines & Carter (1983) sugerem, dentro do contexto de modelos lineares e MLG, respectivamente, construir gráficos da variável adicionada para cada variável do modelo; tais gráficos são denominados **gráficos de alavancagem parcial** ("partial leverage plots") e têm por objetivo avaliar a relação existente entre a variável resposta e as variáveis explicativas (no caso do modelo linear normal, tal gráfico fornece uma idéia sobre o coeficiente de correlação parcial entre a variávei resposta e cada variável explicativa). Para modelos lineares mistos tais gráficos podem ser utéis para avaliar a possível relação existente entre a variável resposta variável explicativa (ma vez que se espera uma relação linear entre os resíduos considerados quando a respectiva variável explicativa tem uma relação linear com a variável resposta. Outra utilidade seria identificar possíveis observações influentes nos coeficientes das variáveis explicativas.

3.4 Decomposição do gráfico da variável adicionada

Hilden-Minton (1995) propõe decompor o gráfico da variável adicionada discutido anteriormente para avaliar o efeito da inclusão de um efeito fixo na predição dos efeitos aleatórios do modelo (3.1). A idéia básica consiste em substituir $\boldsymbol{M}^{-1/2}$ por uma matriz \boldsymbol{A} de dimensão $(n+q) \times n$ tal que $\boldsymbol{M}^{-1} = \boldsymbol{A}^{\top} \boldsymbol{A}$ nas expressões $\boldsymbol{M}^{-1/2} \boldsymbol{Q}_1 \boldsymbol{Y} \in \boldsymbol{M}^{-1/2} \boldsymbol{Q}_1 \boldsymbol{X}_2$ utilizadas no gráfico da variável adicionada. Em particular, ele sugere utilizar

$$oldsymbol{A} = \left[egin{array}{c} oldsymbol{R}^{1/2} \ oldsymbol{D}^{1/2}oldsymbol{Z}^{ op} \end{array}
ight]$$

O gráfico da variável adicionada decomposto corresponde ao gráfico das componentes de AQ_1Y versus AQ_1X_2 . Considerando (1.21), (C.6) e (3.11) temos

$$\boldsymbol{A}\boldsymbol{Q}_{1}\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{R}^{1/2}\boldsymbol{Q}_{1}\boldsymbol{Y} \\ \boldsymbol{D}^{1/2}\boldsymbol{Z}^{\top}\boldsymbol{Q}_{1}\boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}^{-1/2}(\boldsymbol{Y} - \boldsymbol{X}_{1}\widehat{\boldsymbol{\beta}}_{1} - \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}) \\ \boldsymbol{D}^{-1/2}\widehat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{11} \\ \boldsymbol{r}_{12} \end{bmatrix}$$
(3.13)

е

$$\boldsymbol{A}\boldsymbol{Q}_{1}\boldsymbol{X}_{2} = \begin{bmatrix} \boldsymbol{R}^{1/2}\boldsymbol{Q}_{1}\boldsymbol{X}_{2} \\ \boldsymbol{D}^{1/2}\boldsymbol{Z}^{\top}\boldsymbol{Q}_{1}\boldsymbol{X}_{2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}^{-1/2}(\boldsymbol{X}_{2} - \boldsymbol{X}_{1}\widehat{\boldsymbol{\beta}}_{1}^{*} - \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{*}) \\ \boldsymbol{D}^{-1/2}\widehat{\boldsymbol{\gamma}}^{*} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{21} \\ \boldsymbol{r}_{22} \end{bmatrix}. \quad (3.14)$$

O gráfico de r_{11} versus r_{21} é chamado por Hilden-Minton (1995) de gráfico da variável adicionada intra-unidades ("within-unit"), pois leva em consideração as mudanças no BLUE e BLUP. Tal gráfico é similar ao gráfico da variável adicionada proposto na subseção 3.3. O segundo gráfico, de r_{12} versus r_{22} é denominado gráfico da variável adicionada entre-unidades ("between-unit"), uma vez que considera apenas a relação existente entre os BLUP.

A principal vantagem da decomposição proposta é que ela possibilita avaliar o efeito da inclusão de um efeito fixo no BLUP, enquanto que o gráfico da variável adicionada "puro" fornece uma idéia a respeito da inclusão do efeito fixo e de observações influentes no coeficiente do efeito fixo a ser incluído.

3.5 Pontos alavanca

Define-se como observação (ponto) alavanca ("high leverage") aquela que tem uma forte influência no correspondente valor predito. No caso linear, uma observação é dita ser alavanca se o *i-ésimo* elemento da diagonal principal da matriz $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}$, h_{ii} , é "grande". Nesse caso, a respectiva observação representa um ponto discrepante ("outlier") no subespaço gerado pelas colunas da matriz \boldsymbol{X} . Desta forma destacam-se aqueles pontos em que os valores das variáveis explicativas são mais atípicos [Cook & Weisberg (1982) e Wei *et al.* (1998)]. Para maiores detalhes sobre as propriedades da matriz \boldsymbol{H} , veja Cook (1977), Hoaglin & Welsch (1978), Besley *et al.* (1980), Cook & Weisberg (1982), Atkinson (1985) e Chatterjee & Hadi (1988).

Para modelos lineares mistos, Christensen & Pearson (1992) sugerem avaliar a alavancagem da *i-ésima* observação através do valor $h_i^* = \tilde{h}_i/s_i$, em que

$$\widetilde{h}_i = \widetilde{\boldsymbol{x}}_i^\top (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \widetilde{\boldsymbol{x}}_i, \qquad (3.15)$$

$$\widetilde{\boldsymbol{x}}_{i} = \boldsymbol{x}_{i} - \boldsymbol{X}_{(i)}^{\top} \boldsymbol{V}_{(i)}^{-1} \boldsymbol{v}_{i}, \qquad (3.16)$$

$$s_i = v_{ii} - \boldsymbol{v}_i^\top \boldsymbol{V}_{(i)}^{-1} \boldsymbol{v}_i, \qquad (3.17)$$

com $\boldsymbol{x}_i \in \boldsymbol{v}_i$ representando, respectivamente, a *i-ésima* coluna das matrizes $\boldsymbol{X} \in \boldsymbol{V}$, enquanto $\boldsymbol{X}_{(i)} \in \boldsymbol{V}_{(i)}$ representam, respectivamente, as matrizes $\boldsymbol{X} \in \boldsymbol{V}$ com a *i-ésima* coluna removida e v_{ii} representa o *i-ésimo* elemento da diagonal principal de \boldsymbol{V} . No caso linear normal, tem-se que $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i, s_i = \sigma^2, \tilde{h}_i = \sigma^2 h_{ii} \in h_i^* = h_{ii}$. A grande desvantagem de tal abordagem é a necessidade de calcular $\hat{\boldsymbol{V}}_{(i)}^{-1}$ para os *n* valores amostrados.

Recentemente, Wei et al. (1998) definiram a matriz de alavancagem generalizada

("generalized leverage matrix") para uma gama de modelos estatísticos. Quando $\widehat{Y} = \widehat{I\!\!E[Y]} = \mu(\widehat{\beta})$, a matriz de alavancagem generalizada definida por

$$GL(\widehat{\boldsymbol{\beta}}) = \frac{\partial \widehat{\boldsymbol{Y}}}{\partial \boldsymbol{Y}^{\top}} = \left(\frac{\partial \hat{y}_i}{\partial y_j}\right)_{n \times n},\tag{3.18}$$

reflete a taxa de mudança instantânea no respectivo valor predito quando a variável resposta é acrescida de um infinitésimo. No caso linear normal, $GL(\hat{\beta}) = H$. A alavancagem generalizada da *i-ésima* observação corresponde ao termo $GL(\hat{\beta})_{ii} = \partial \hat{y}_i / \partial y_i$, ou seja, o *i-ésimo* elemento da diagonal principal da matriz (3.18).

Considerando V conhecida no modelo (1.4) e lembrando (1.13), mostra-se que (3.18) se reduz a

$$GL(\widehat{\boldsymbol{\beta}}) = \frac{\partial \widehat{\boldsymbol{Y}}}{\partial \boldsymbol{Y}^{\top}} = \frac{\partial \widehat{\boldsymbol{E}[\boldsymbol{Y}]}}{\partial \boldsymbol{Y}^{\top}} = \frac{\partial \boldsymbol{X}\widehat{\boldsymbol{\beta}}}{\partial \boldsymbol{Y}^{\top}} = \frac{\partial \boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{Y}}{\partial \boldsymbol{Y}^{\top}} = \boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}.$$
(3.19)

Tal matriz fornece informações sobre a alavancagem das observações com relação às estimativas dos efeitos fixos [Fung *et al.* (2002)]. Puterman (1988), Martin (1992) e Banerjee & Frees (1997), sugeriram utilizar $H^* = V^{-1/2} X (X^{\top} V^{-1} X)^{-1} X^{\top} V^{-1/2}$, com $V^{-1} = (V^{-1/2})^{\top} V^{-1/2}$, como matriz de alavancagem. Uma vez que $V^{-1/2}$ não é unicamente determinada, fica a dúvida referente à sensibilidade do resultado em relação à decomposição utilizada para obter $V^{-1/2}$.

Dentro do contexto de medidas repetidas, Banerjee & Frees (1997) definem uma matriz de alavancagem para cada unidade experimental. De forma semelhante, podemos definir a matriz de alavancagem generalizada referente a i-*ésima* unidade experimental como $\boldsymbol{H}_i = \boldsymbol{X}_i (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1}$, com \boldsymbol{V}_i representando a matriz de covariâncias (1.5) para a unidade experimental em questão.

Considerando (3.19), tem-se

$$\operatorname{tr}\left[GL(\widehat{\boldsymbol{\beta}})\right] = \operatorname{tr}\left[\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\right]$$
$$= \operatorname{tr}\left[\left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}^{\top}\right] = \operatorname{tr}\left[\boldsymbol{I}_{p}\right] = p. \quad (3.20)$$

Definindo $h_{ii}^* = GL(\hat{\beta})_{ii}$, segue que $\bar{h}^* = n^{-1} \sum_{i=1}^n h_{ii}^* = p/n$. Como Hoaglin & Welsch (1978), no contexto do modelo linear normal, consideraremos o *i-ésimo* ponto como "alavanca" se $h_{ii}^* \ge 2p/n$. Usando a abordagem de Banerjee & Frees (1997) podemos definir uma unidade experimental como alavanca se

$$\frac{\operatorname{tr}(\boldsymbol{H}_i)}{n_i} = \frac{\sum_{j \in I} h_{jj}^*}{n_i} \ge 2p/n,$$

com I representando o conjunto das n_i observações da *i-ésima* unidade experimental. Proposta semelhante foi apresentada por Venezuela (2003, p.27).

As propostas comentadas anteriormente referem-se apenas a pontos alavanca relativos às estimativas dos parâmetros fixos. Uma vez, que no modelo (1.2) uma observação pode influenciar tanto as estimativas dos parâmetros fixos como as predições dos efeitos aleatórios, é aconselhável medir esta influência de forma conjunta. Desta forma, para incorporar informações a respeito dos efeitos aleatórios, sugerimos considerar $\widehat{Y}^* = \widehat{I\!E[Y|\gamma]} =$ $\widehat{X}\widehat{\beta} + \widehat{Z}\widehat{\gamma}$. Derivando \widehat{Y}^* com relação a Y^{\top} obtemos

$$GL(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}}) = \frac{\partial \widehat{\boldsymbol{Y}}^*}{\partial \boldsymbol{Y}^\top} = \frac{\widehat{\boldsymbol{Y}}}{\partial \boldsymbol{Y}^\top} + \frac{\partial \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}}{\partial \boldsymbol{Y}^\top} = GL(\widehat{\boldsymbol{\beta}}) + \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^\top\boldsymbol{Q}.$$
(3.21)

O termo $GL(\hat{\beta})$ leva em consideração apenas os pontos alavanca referentes às estimativas dos efeitos fixos e $ZDZ^{\top}Q$ leva em consideração a estrutura de covariância D e a matriz de especificação Z dos efeitos aleatórios. Nesse caso, definimos a alavancagem generalizada da *i-ésima* observação relativamente às estimativas dos parâmetros fixos (BLUE) e às predições dos efeitos aleatórios (BLUP) como o *i-ésimo* elemento da diagonal principal de (3.21), ou seja, $GL(\hat{\beta}, \hat{\gamma})_{ii} = h_{ii}^* + (ZDZ^{\top}Q)_{ii}$. De forma semelhante ao caso anterior, definimos uma observação como alavanca se $GL(\hat{\beta}, \hat{\gamma})_{ii} \ge 2\text{tr}(GL(\hat{\beta}, \hat{\gamma}))/n$ e uma unidade experimental como alavanca se $(n_i)^{-1}\sum_{j\in I} GL(\hat{\beta}, \hat{\gamma})_{jj} \ge 2\text{tr}(GL(\hat{\beta}, \hat{\gamma}))/n$. Como não se conhecem $D \in V$, devemos avaliar todas as medidas obtidas com base em estimativas $\hat{D} \in \hat{V}$; como conseqüência todos os resultados obtidos são sensíveis a sua má especificação.

3.6 Eliminação de observações

Uma prática simples e de fácil interpretação consiste em avaliar a influência de uma particular observação, ou de um conjunto de observações, por intermédio dos efeitos provocados por sua eliminação do conjunto de dados. Tal prática foi introduzida por Cook (1977) e pode ser adaptada a uma diversidade de modelos. Diversas medidas de influência baseiam-se no conceito de eliminação de observações. Para o caso linear normal, veja Besley *et al.* (1980), Cook & Weisberg (1982), Atkinson (1985) e Chatterjee & Hadi (1986, 1988), por exemplo.

Sob essa abordagem, é essencial obter expressões que relacionem o estimador do parâmetro de interesse obtido com base em toda amostra com o respectivo estimador obtido quando se elimina um conjunto de observações, sem a necessidade de reajustar o modelo. Quando as estimativas são obtidas iterativamente como na classe dos modelos lineares mistos o procedimento apresenta inconveniências. Algumas propostas de aproximação para esses casos são dadas em Pregibon (1981)[aproximação por 1 passo], Jorgensen (1993), Mak (1993), Tsai (1994) e Haslett & Dillane (2004). Uma alternativa muito utilizada [Christensen *et al.* (1992), Hilden-Minton (1995), Banerjee & Free (1997), Haslett (1999), Fung *et al.* (2002) e Fei & Pan (2003)] é considerar a estrutura de covariância conhecida, a menos de um parâmetro de escala σ^2 , de forma que o processo de estimação seja linear, permitindo encontrar a relação existente entre os estimadores. Então, os estimadores são avaliados com base nas estimativas dos parâmetros de covariância. Por conseguinte, torna-se essencial especificar tal estrutura de forma correta [Fei & Pan (2003)].

Geralmente as fórmulas de atualização são obtidas quando se exclui uma única observação. Existem situações em que há interesse em obter tais fórmulas quando se exclui um conjunto de observações, principalmente quando esses conjuntos são definidos pela estrutura dos dados. Em estudos com medidas repetidas, por exemplo, esses conjuntos podem ser definidos pelas observações associadas a uma mesma unidade experimental. Um outro exemplo envolve a avaliação da influência conjunta de duas ou mais observações, pois a eliminação de uma única observação pode mascarar o efeito de observações que são conjuntamente influentes (*masking effect*) [Cook & Weisberg (1980, fig.1)]. Hilden-Minton (1995) e Fung *et al.* (2002) consideram o **modelo de deslocamento médio para pontos discrepantes** ("mean shift outlier") e usam a sua equivalência com o modelo de eliminação de observações [Cook & Weisberg (1982)] para encontrar as fórmulas de atualização do BLUE e BLUP quando eliminamos um conjunto de observações. Apresentaremos aqui a formulação dada por Hilden-Minton (1995).

Se supusermos que D e V são conhecidas e que temos interesse em eliminar os casos indexados pelo conjunto $I = \{i_1, i_2, ..., i_k\}$ $(1 \le i_1 \le i_2 \le ... \le i_k \le n)$, o modelo considerado é

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{U}_{I}\boldsymbol{\phi}_{I} + \boldsymbol{\varepsilon}, \qquad (3.22)$$

com ϕ_I representando um vetor de parâmetros de dimensão $k \times 1$ e U_I definida como em (2.9). Hilden-Minton (1995) e Fung *et al.* (2002) mostram que o BLUE de β e o BLUP de γ sob o modelo (3.22) são equivalentes ao BLUE de β e ao BLUP de γ sob o modelo (1.2) quando eliminamos as observações do conjunto I (veja Apêndice C.4).

Usando os resultados da seção 3.3, podemos encontrar o BLUE de β e o BLUP de γ sob o modelo (3.22). Nesse contexto $X, U_I, \beta \in \phi_I$ fazem o papel de $X_1, X_2, \beta_1 \in \beta_2$,

respectivamente. Portanto, o BLUE de ϕ_I obtido via (3.4) é dado por

$$\widehat{\boldsymbol{\phi}}_{I} = \left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{Y}.$$
(3.23)

Similarmente, pelas equações (3.8) e (3.9) obtém-se

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)} = \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{U}_{I} \widehat{\boldsymbol{\phi}}_{I}$$
(3.24)

е

$$\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)} = \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \widehat{\boldsymbol{\phi}}_{I}, \qquad (3.25)$$

com $\widehat{\boldsymbol{\beta}}_{(I)}$ ($\widehat{\boldsymbol{\gamma}}_{(I)}$) representando o BLUE (BLUP) de $\boldsymbol{\beta}$ ($\boldsymbol{\gamma}$) quando se eliminam as observações do conjunto *I*. O resultado (3.24) apresentado por Hilden-Minton (1995) generaliza os resultados (3.3) e (4.1) obtidos por Fung *et al.* (2002) [quando desconsideramos a função não-paramétrica *f* e o processo estocástico U(t)], obtidos quando se exclui uma única observação.

A expressão (3.25) está relacionada com o segundo termo de (3.21); a expressão (3.24) mantém uma relação com (3.19). Se o interesse se concentra na mudança em $Z\hat{\gamma} \in X\hat{\beta}$, quando eliminamos as observações do conjunto I, pré-multiplicamos (3.25) por Z, obtendo $ZDZ^{\top}QU_{I}\hat{\phi}_{I}$ que é uma submatriz de $ZDZ^{\top}Q$ multiplicada pelas componentes de (3.23). Um resultado analógo é válido para (3.24).

Por (3.8) e (3.9) tem-se diretamente que

$$\operatorname{Var}[\widehat{\boldsymbol{\phi}}_{I}] = \sigma^{2} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1}, \qquad (3.26)$$

$$\operatorname{Var}[\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}] = \sigma^{2} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{M} \boldsymbol{X} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} (3.27)$$

е

$$\operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)}] = \sigma^2 \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_I \left(\boldsymbol{U}_I^{\top} \boldsymbol{Q} \boldsymbol{U}_I \right)^{-1} \boldsymbol{U}_I^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D}.$$
(3.28)

Além disso, mostra-se que

$$\operatorname{Var}[\widehat{\boldsymbol{\beta}}_{(I)}] = \operatorname{Var}[\widehat{\boldsymbol{\beta}}] + \operatorname{Var}[\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}]$$
(3.29)

е

$$\operatorname{Var}[\widehat{\boldsymbol{\gamma}}_{(I)} - \boldsymbol{\gamma}] = \operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}] + \operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)}].$$
(3.30)

As demonstrações dos resultados (3.26), (3.27), (3.28), (3.29) e (3.30) encontram-se no Apêndice C.5.

Uma das medidas mais utilizadas para avaliar a influência de um conjunto de observações, via eliminação, é a distância de Cook [Cook (1977), Cook & Weisberg (1982) e Chatterjee & Hadi (1986, 1988)] definida como

$$D_{I} = \frac{\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(I)}\right)^{\top} \boldsymbol{U}\left(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_{(I)}\right)}{c}, \qquad (3.31)$$

com $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\theta}}_{(I)}$ representando, respectivamente, a estimativa do vetor $\boldsymbol{\theta}$ com todos os dados da amostra e com a eliminação do conjunto de observações I, U denotando uma matriz positiva definida e c um parâmetro de escala. A medida D_I mede a influência das observações do conjunto I na estimativa de $\boldsymbol{\theta}$, segundo a métrica definida por U e c. No caso linear normal, costuma-se utilizar $U = \mathbf{X}^{\top} \mathbf{X}$ e $c = p \widehat{\sigma}^2$ [Cook (1977)]. Para modelos lineares mistos, uma proposta é utilizar

$$D_{I} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)})^{\top} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X}) (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)})}{c}$$
$$= \frac{(\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(I)})^{\top} \boldsymbol{V}^{-1} (\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(I)})}{c}, \qquad (3.32)$$

para medir a influência das observações do conjunto I nas estimativas dos parâmetros fixos. Christensen & Pearson (1992) sugerem utilizar (3.32) para medir a influência de uma única observação, enquanto Banerjee & Frees (1997), dentro do contexto de análise de dados longitudinais, utilizam (3.32) para medir a influência das unidades experimentais (nesse caso, I refere-se às n_i observações da unidade experimental em questão). Fung et al. (2002) consideram ambos os casos. As abordagens são fundamentalmente distintas, uma vez que a primeira tem por objetivo encontrar observações influentes, enquanto que a segunda destina-se a encontrar o impacto das unidades experimentais sem se preocupar com as observações particulares. Desta forma, Banerjee & Frees (1997) sugerem utilizar a segunda abordagem para detecção de unidades experimentais *influentes*. Banerjee (1998) e Tan et al. (2001) mostram como a distância de Cook (3.32) tem uso limitado em modelos lineares mistos, pois ela pode falhar na detecção de observações que tenham grande impacto em $\hat{\gamma}$ [Tan *et al.* (2001), Teorema 1]. Tan *et al.* (2001) consideram um modelo semelhante a (1.1), no qual as matrizes Z_i são consideradas submatrizes de X_i . Levando em consideração que uma mudança no BLUP causada pela eliminação de um conjunto de observações I não tem a mesma influência em todos os elementos de $\hat{\beta}$, esses autores sugeriram abordar o problema condicionalmente aos BLUP $\hat{\gamma}_i$ (i = 1, ..., c), pois o efeito de eliminar uma observação na estrutura de covariância é equivalente aos efeitos causados nas componentes de $\hat{\gamma}_j$ (j = 1, ..., c)[Tan *et al.* (2001)]. O modelo condicional utilizado é da forma

$$\boldsymbol{Y} = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*, \tag{3.33}$$

em que

$$\boldsymbol{X}^* = [\boldsymbol{X} : \boldsymbol{Z}], \tag{3.34}$$

com $X \in Z$ matrizes de posto completo definidas como em (1.1) e $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. Além do mais, $\boldsymbol{\varepsilon}^* = \boldsymbol{\varepsilon}$ uma vez que usualmente $\boldsymbol{\varepsilon} \in \boldsymbol{\gamma}$ são considerados independentes sob o modelo (1.1).

O modelo (3.33) não é identificável, dado que a matriz X^* não tem posto completo. Desta forma, os autores sugerem o uso do BLUE e BLUP não condicionais de β e γ , res-pectivamente. Assim, a **distância de Cook condicional** é comparável a (3.32). A distância de Cook condicional a γ foi definida por Tan *et al.* (2001) como

$$D_{i}^{cond} = \sum_{j=1}^{c} \frac{\boldsymbol{P}_{j(i)}^{\top} \operatorname{Var}[\boldsymbol{Y}|\boldsymbol{\gamma}]^{-1} \boldsymbol{P}_{j(i)}}{(n-1)c+p}$$
$$= \sum_{j=1}^{c} \frac{\boldsymbol{P}_{j(i)}^{\top} \boldsymbol{P}_{j(i)}}{k}, \qquad (3.35)$$

com

$$\boldsymbol{P}_{j(i)} = \widehat{\boldsymbol{Y}}_{j} - \widehat{\boldsymbol{Y}}_{j(i)} = (\boldsymbol{X}_{j}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_{j}\widehat{\boldsymbol{\gamma}}_{j}) - (\boldsymbol{X}_{j}\widehat{\boldsymbol{\beta}}_{(i)} + \boldsymbol{Z}_{j}\widehat{\boldsymbol{\gamma}}_{j(i)}), \qquad (3.36)$$

em que $\widehat{\boldsymbol{\beta}}_{(i)}$ e $\widehat{\boldsymbol{\gamma}}_{j(i)}$ representam, respectivamente, $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\gamma}}_{j}$ obtidos quando eliminamos da amostra a *i-ésima* observação e $k = \sigma^2([n-1]c+p)$. Podemos decompor (3.35) da seguinte forma (Apêndice C.6)

$$D_i^{cond} = D_{1i}^{cond} + D_{2i}^{cond} + D_{3i}^{cond}, ag{3.37}$$

em que

$$D_{1i}^{cond} = \frac{(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})^{\top} (\boldsymbol{X}^{\top} \boldsymbol{X}) (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})}{k} = \frac{(\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(i)})^{\top} (\widehat{\boldsymbol{Y}} - \widehat{\boldsymbol{Y}}_{(i)})}{k}, \quad (3.38)$$

$$D_{2i}^{cond} = \frac{(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(i)})^{\top} \boldsymbol{Z}^{\top} \boldsymbol{Z}(\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(i)})}{k}, \qquad (3.39)$$

е

$$D_{3i}^{cond} = \frac{2(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)})^{\top} \boldsymbol{X}^{\top} \boldsymbol{Z} (\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(i)})}{k}.$$
(3.40)

O primeiro termo, D_{1i}^{cond} , é uma medida útil para avaliar a influência da *i-ésima* observação em $\hat{\beta}$ e é comparável a (3.32), não incluindo a padronização pela matriz de covariâncias de \mathbf{Y} . O segundo termo, D_{2i}^{cond} , é uma medida útil para avaliar a influência da *i-ésima* observação em $\hat{\gamma}$. Já o terceiro termo, D_{3i}^{cond} , é uma medida de covariância, entre a mudança nas estimativas do BLUE e BLUP, quando eliminamos a *i-ésima* observação. Geralmente (3.40) tem um valor desprezível [Tan *et al.* (2001)], que de certa forma, reproduz a propriedade de "independência" entre o BLUE e o BLUP (veja Apêndice A.3, propriedade 6). A grande vantagem de (3.37) é que podemos avaliar a influência de uma observação por intermédio de sua influência nos efeitos fixos e aleatórios. Tan *et al.* (2001) sugerem utilizar (3.39), para avaliar a influência da *i-ésima* observação nos parâmetros da estrutura de covariância $\boldsymbol{\theta}^*$, ao invés da proposta de Christensen & Pearson (1992), que sugeriram avaliar a influência em $\hat{\boldsymbol{\beta}}$ e no EMVR de $\boldsymbol{\theta}^*$ separadamente, uma vez que os mesmos só são assintoticamente independentes [Miller (1977), Harville (1977) e Lesaffre & Verbeke (1998)]. Desta forma, é mais sensato avaliar a influência por meio da abordagem proposta por Tan *et al.* (2001).

Dentro do contexto de análise de dados longitudinais, se o interesse é avaliar a influência de um conjunto de observações, basta calcular (3.35), (3.38) e (3.39) com relação ao conjunto de interesse, ao invés de uma simples observação. No entanto, quando eliminamos todas as observações de uma unidade experimental, alguns BLUP não podem ser obtidos. Desta forma, nós propomos avaliar a influência da *i-ésima* unidade experimental através da média das distâncias (3.35) referentes a todas as suas observações, ou seja,

$$D_{i.}^{cond} = (n_i)^{-1} \sum_{j \in I} D_j^{cond}, \qquad (3.41)$$

com I representando o conjunto das n_i observações da *i-ésima* unidade experimental.

Com base na expressão (3.29), Hilden-Minton (1995) sugere avaliar a influência das observações pertecentes ao conjunto I a partir do **volume dos elipsóides de confiança** [Chatterjee & Hadi (1988, p.134)] da matriz de covariâncias de $\hat{\beta}$. Considerando a propriedade 4 do Apêndice A.3 em conjunto com (3.27) e (3.29) tem-se

$$\frac{\left|\widehat{\operatorname{Var}}[\widehat{\boldsymbol{\beta}}_{(I)}]\right|}{\left|\widehat{\operatorname{Var}}[\widehat{\boldsymbol{\beta}}]\right|} = \frac{\left|\sigma^{2}\left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{I}_{n}+\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{M}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\right)\right|}{\left|\sigma^{2}\left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\right|}$$
$$= \left|\boldsymbol{I}_{n}+\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{M}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\right| = \operatorname{COVRATIO}_{(I)}^{*}.$$

Tal medida coincide com aquela denominada COVRATIO proposta por Beslev *et al.*

(1980) quando se conhecem $\sigma^2 \in V$. Uma outra proposta, seria considerar conhecida apenas a matriz V e avaliar o efeito da eliminação do conjunto I na matriz de covariâncias do BLUE por meio de

$$COVRATIO_{(I)} = \frac{|\widehat{Var}[\widehat{\boldsymbol{\beta}}_{(I)}]|}{|\widehat{Var}[\widehat{\boldsymbol{\beta}}]|}$$
$$= \frac{|\widehat{\sigma}_{(I)}^{2}(\boldsymbol{X}_{(I)}^{\top}\boldsymbol{M}_{(I)}\boldsymbol{X}_{(I)})^{-1}|}{|\widehat{\sigma}^{2}(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X})^{-1}|}$$
$$= \left(\frac{n-p-M_{I}}{n-p-k}\right)^{p}COVRATIO_{(I)}^{*}, \qquad (3.42)$$

com $M_I = \boldsymbol{Y}^\top \boldsymbol{Q} \boldsymbol{U}_I (\boldsymbol{U}_I^\top \boldsymbol{Q} \boldsymbol{U}_I)^{-1} \boldsymbol{U}_I^\top \boldsymbol{Q} \boldsymbol{Y}.$

Christensen & Pearson (1992) sugerem avaliar a mudança na variância total da matriz de covariâncias dos efeitos fixos, quando se elimina uma única observação, através de

$$t_i^* = \left| \operatorname{tr} \{ \operatorname{Var}[\widehat{\boldsymbol{\beta}}]^{-1} \operatorname{Var}[\widehat{\boldsymbol{\beta}}_{(i)}] \} - p \right| = \left| \frac{h_i^*}{1 - h_i^*} \right|.$$
(3.43)

que é uma função crescente de h_i^* (veja Apêndice C.7). Desta forma, h_i^* é útil para dar idéia a respeito da mudança ocorrida na matriz de covariâncias de $\hat{\beta}$, quando eliminamos a *i-ésima* observação.

Recentemente, Fei & Pan (2003), baseados no artigo de Zhu *et al.* (2001), propuseram medir a influência da eliminação das observações do conjunto I por meio de uma **distância de Cook generalizada** baseada na função $Q(\psi/\tilde{\psi}) = \mathbb{E}[\ln f(\mathbf{Y}, \boldsymbol{\gamma})|\mathbf{Y}, \tilde{\psi}]$ utilizada no algoritmo EM, com $\tilde{\psi}$ sendo a solução atualizada de $\boldsymbol{\psi} = (\boldsymbol{\beta}^{\top}, (\boldsymbol{\theta}^*)^{\top})$. A medida proposta por Fei & Pan (2003) é

$$D_{I}^{*} = \left[\dot{Q}_{(I)}(\widehat{\psi}/\widehat{\psi})\right]^{\top} \left\{-I\!\!E[\ddot{Q}(\widehat{\psi}/\widehat{\psi})]\right\}^{-1} \left[\dot{Q}_{(I)}(\widehat{\psi}/\widehat{\psi})\right], \qquad (3.44)$$

com $\dot{Q}(\hat{\psi}/\hat{\psi})$ e $\ddot{Q}(\hat{\psi}/\hat{\psi})$ representando, respectivamente, a primeira e segunda derivada de *Q* calculada no EMV $\hat{\psi}$, enquanto que $Q_{(I)}$ representa a função *Q* quando eliminamos as observações referentes ao conjunto *I*. Analogamente à (3.35), (3.44) pode ser decomposta da seguinte forma:

$$D_I^* = D_{I\beta}^* + D_{I\theta^*}^*, (3.45)$$

com $D_{I\beta}^*$ e $D_{I\theta^*}^*$ tendo interpretação semelhante a (3.38) e (3.39), respectivamente [Fei & Pan (2003)]. Esta medida de influência também é condicional a γ . A vantagem de se utilizar (3.35) é que não é preciso supor uma distribuição específica para $\gamma \in \boldsymbol{\varepsilon}$.

3.7 Influência local

O conceito de influência local foi proposto por Cook (1986) com o objetivo de avaliar a mudança nos resultados da análise quando incorporamos "pequenas perturbações" ao modelo. Dentro desse contexto, pode-se perturbar a matriz de covariâncias, os parâmetros, a variável resposta, etc. A abordagem original baseia-se na análise do **afastamento da verossimilhança** ("likelihood displacement") [Cook & Weisberg (1982), Cook (1987) e Cook *et al.* (1988)]

$$LD(\boldsymbol{w}) = 2\left\{L(\widehat{\boldsymbol{\theta}}) - L(\widehat{\boldsymbol{\theta}}_{\boldsymbol{w}})\right\}, \qquad (3.46)$$

em que $L(\cdot)$ é a log-verossimilhança do modelo postulado, $\boldsymbol{\theta}$ é um vetor $p \times 1$ de parâmetros desconhecidos, $L(\cdot|\boldsymbol{w})$ é a log-verossimilhança do modelo "perturbado", \boldsymbol{w} representa um vetor $q \times 1$ de perturbações, restrito a um intervalo aberto $\Omega \subset \mathbb{R}^q$, $\boldsymbol{\hat{\theta}} \in \boldsymbol{\hat{\theta}}_{\boldsymbol{w}}$ são respectivamente, os EMV baseados em $L(\cdot)$ e $L(\cdot|\boldsymbol{w})$. Assume-se que $\boldsymbol{w}_0 \in \Omega$ (ausência de perturbação) é tal que $L(\boldsymbol{\theta}|\boldsymbol{w}_0) = L(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \Theta$ e que $L(\boldsymbol{\theta}|\boldsymbol{w})$ possua as duas primeiras derivadas contínuas na vizinhança de $(\boldsymbol{\hat{\theta}}^{\top}, \boldsymbol{w}_0^{\top})$. Nesse contexto $LD(\boldsymbol{w})$ é utilizada para comparar $\boldsymbol{\hat{\theta}} \in \boldsymbol{\hat{\theta}}_{\boldsymbol{w}}$ com respeito aos contornos da log-verossimilhança $L(\cdot)$.

Cook (1986) considerou o gráfico de influência $(LD(\boldsymbol{w})$ vs. $\boldsymbol{w})$ como uma superfície em $I\!\!R^{q+1}$ formada pelos valores do vetor

$$\boldsymbol{\alpha}(\boldsymbol{w}) = \left(\boldsymbol{w}^{\top}, LD(\boldsymbol{w})\right)^{\top}, \qquad (3.47)$$

com \boldsymbol{w} variando em Ω . Tal gráfico contém informação essencial da influência do esquema de perturbação em questão. O método proposto por Cook (1986) consiste no estudo do comportamento local (influência local) do gráfico de influência na vizinhança de \boldsymbol{w}_0 . Para medir a sensibilidade do afastamento da verossimilhança, ele utilizou a curvatura normal [Araujo (1998)] de (3.47) ao redor de \boldsymbol{w}_0 na direção de um vetor \boldsymbol{d} ($q \times 1$) de norma unitária, que doravante será denominada $C_{\boldsymbol{d}}$.

A curvatura normal [para detalhes veja Souza (1999)], nesse caso é dada por [Cook (1986, eq.16)]

$$C_{\boldsymbol{d}} = 2 \left| \boldsymbol{d}^{\mathsf{T}} \boldsymbol{H}^{\mathsf{T}} \ddot{\boldsymbol{L}}^{-1} \boldsymbol{H} \boldsymbol{d} \right|, \qquad (3.48)$$

com $\ddot{\boldsymbol{L}} = \left\{ \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta} \right\} |_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}$ e $\boldsymbol{H} = \left\{ \partial^2 L(\boldsymbol{\theta} | \boldsymbol{w}) / \partial \boldsymbol{\theta}^\top \partial \boldsymbol{w} \right\} |_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}$. A curvatura normal (3.48) é essencialmente equivalente à segunda derivada de $LD(\boldsymbol{w})$ ao redor de \boldsymbol{w}_0 [Billor & Loynes (1993), Wu & Luo (1993b) e Araujo (1998)]. Pode-se mostrar que $C_{min} \leq C_d \leq C_{max}$, com $C_{min} \leq C_2 \dots \leq C_{q-1} \leq C_{max}$ representando os q autovalores da

matriz $\ddot{\boldsymbol{F}} = -\boldsymbol{H}^{\top} \ddot{\boldsymbol{L}}^{-1} \boldsymbol{H}$. O autovetor normalizado \boldsymbol{d}_{max} associado com C_{max} é extremamente útil, uma vez que indica o tipo de perturbação no modelo postulado que produz maior troca em $LD(\boldsymbol{w})$, ou seja, indica que combinação dos elementos de \boldsymbol{w} são mais influentes na direção de maior curvatura (contorno) de $LD(\boldsymbol{w})$.

Desta forma, d_{max} pode ser utilizado como uma ferramenta útil na análise de diagnóstico. O gráfico dos elementos de $| d_{max} |$ pode revelar qual o tipo de perturbação tem a maior influência em LD(w) na "vizinhança" de w_0 [Cook (1986)], é importante também investigar quais as causas específicas desta sensibilidade. Na literatura, outros tipos de gráficos são sugeridos para diagnóstico. Cook (1986) propõe inspecionar as componentes de d_{max} , independentemente do valor de C_d , uma vez que ele pode indicar observações que são conjuntamente influentes.

Pode-se usar o conceito de influência local utilizando outras medidas de influência, veja por exemplo, Cook (1986), McCulloch (1989), Wu & Luo (1993a) e Lee & Zhao (1996) ou outras abordagens, vide Billor & Loynes (1993). Por exemplo, quando o interesse é avaliar a influência parcial em um subconjunto de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\top}, \boldsymbol{\theta}_2^{\top})^{\top}$, digamos $\boldsymbol{\theta}_1$, Cook (1986) sugere utilizar

$$LD_{s}(\boldsymbol{w}) = 2\left\{L(\widehat{\boldsymbol{\theta}}_{1}, \widehat{\boldsymbol{\theta}}_{2}) - L(\widehat{\boldsymbol{\theta}}_{1\boldsymbol{w}}, g(\widehat{\boldsymbol{\theta}}_{1\boldsymbol{w}}))\right\}, \qquad (3.49)$$

ao invés de (3.46). Nesse caso $L(\widehat{\theta}_1, g(\widehat{\theta}_1))$ representa a log-verossimilhança perfilada de θ_1 [Cordeiro (1992)]. A curvatura normal do gráfico de influência na direção de um vetor d (de norma unitária) associado a (3.49) é

$$C_{\boldsymbol{d}}(\boldsymbol{\theta}_1) = 2 \left| \boldsymbol{d}^{\mathsf{T}} \boldsymbol{H}^{\mathsf{T}} (\ddot{\boldsymbol{L}}^{-1} - \boldsymbol{B}_{22}) \boldsymbol{H} \boldsymbol{d} \right|, \qquad (3.50)$$

com

$$\boldsymbol{B}_{22} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddot{\boldsymbol{L}}_{22}^{-1} \end{pmatrix}, \qquad (3.51)$$

e $\ddot{\boldsymbol{L}}_{22} = \left\{ \partial^2 L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_2^\top \partial \boldsymbol{\theta}_2 \right\} |_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$. Fung & Kwan (1997) mostram que a curvatura normal é invariante com relação à escala quando a derivada da medida de influência avaliada no EMV é nula (válido, para (3.46), por exemplo); desta forma eles sugerem a aplicação da metodologia de influência local, baseada na curvatura normal, apenas quando a referida derivada é nula. Expressões dos elementos da matriz $\ddot{\boldsymbol{L}}$ são apresentadas no Apêndice (C.13).

3.8 Influência local em modelos lineares mistos

Nesta seção iremos discutir as propostas apresentadas em Beckman *et al.* (1987) e Lesaffre & Verbeke (1998). Ambas baseiam-se na verossimilhança marginal de $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$ sob o modelo (1.1). Por conveniência, não colocaremos um parâmetro de dispersão em evidência, ou seja, consideraremos $\operatorname{Var}[\boldsymbol{\gamma}] = \boldsymbol{D}(\boldsymbol{\theta}) \operatorname{com} \boldsymbol{\theta}$ representando um vetor de dimensão $l \times 1$ contendo os $l \leq \binom{q}{2} + q$ parâmetros de covariância. Nesse caso, o vetor de parâmetros é $\boldsymbol{\psi}^{\top} = (\boldsymbol{\beta}^{\top}, \sigma^2, \boldsymbol{\theta}^{\top}) = (\boldsymbol{\beta}^{\top}, (\boldsymbol{\theta}^*)^{\top})$. A log-verossimilhança, a menos de uma constante, é

$$L(\boldsymbol{\psi}) = \lambda = -(1/2) \left\{ \ln |\boldsymbol{V}| + (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right\}.$$
(3.52)

Beckman *et al.* (1987) consideram o modelo de componentes de variância. Obteremos nas duas próximas subseções a respectiva matriz H, para os esquemas de perturbação propostos por eles, baseado no modelo geral (3.52).

3.8.1 Perturbação na matriz de covariâncias do erro

Podemos avaliar a sensibilidade do modelo de independência condicional homocedástico com relação à essa suposição, incorporando um vetor $n \times 1$ de perturbações, de tal forma que Var $[\varepsilon] = \sigma^2 \Lambda(w)$, com $\Lambda(w)$ representando uma matriz $(n \times n)$ diagonal e w_k denotando o k-ésimo elemento dessa diagonal. Neste caso, $w_0 = \mathbf{1}_n$ representa um vetor de dimensão $n \times 1$ com todos os elementos iguais a um. A log-verossimilhança perturbada é

$$L(\boldsymbol{\psi}_{\boldsymbol{w}}) = \lambda(\boldsymbol{w}) = -(1/2) \left\{ \ln |\boldsymbol{V}(\boldsymbol{w})| + (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top} \boldsymbol{V}(\boldsymbol{w})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right\}, \quad (3.53)$$

com $V(w) = ZDZ^{\top} + \sigma^2 \Lambda(w)$. A k-ésima coluna da matriz H é dada por

$$\boldsymbol{H}^{k} = \left\{ \left(\frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \boldsymbol{\beta}} \right)^{\top}, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \sigma^{2}}, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \theta_{1}}, \cdots, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \theta_{l}} \right\}^{\top},$$
(3.54)

com as respectivas derivadas calculadas em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$, com $\hat{\boldsymbol{\psi}}$ representando o EMV de $\boldsymbol{\psi}$. Então para k = 1, ..., n (veja Apêndice C.8), temos

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}} = \boldsymbol{X}^\top \boldsymbol{D}_k \hat{\boldsymbol{r}}, \qquad (3.55)$$

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \theta_i} \bigg|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}=\boldsymbol{\hat{\psi}}} = -\frac{1}{2} \left\{ \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^{\top} \right] - 2 \hat{\boldsymbol{r}}^{\top} \boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^{\top} \hat{\boldsymbol{r}} \right\}, \quad (i = 1, ..., l) \quad (3.56)$$

е

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \sigma^2} \Big|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}=\boldsymbol{\widehat{\psi}}} = -\frac{1}{2} \left\{ \hat{\sigma}^{-2} \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \boldsymbol{\widehat{D}} \boldsymbol{Z}^\top \right] - 2 \hat{\boldsymbol{r}}^\top \boldsymbol{D}_k \boldsymbol{\widehat{V}}^{-1} \hat{\boldsymbol{r}} + \hat{\sigma}^{-2} \hat{\boldsymbol{r}}^\top \boldsymbol{D}_k \hat{\boldsymbol{r}} \right\}, \quad (3.57)$$

com $\hat{\boldsymbol{r}} = \boldsymbol{Y} - \boldsymbol{X} \hat{\boldsymbol{\beta}}, \ \boldsymbol{D}_k = \partial \boldsymbol{V}(\boldsymbol{w})^{-1} / \partial w_k |_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}} e \dot{\boldsymbol{D}}_i = \partial \boldsymbol{D} / \partial \theta_i |_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}}$. Para esse esquema de perturbação mostra-se que (Apêndice C.9)

$$\boldsymbol{D}_{k} = -\hat{\sigma}^{2} \widehat{\boldsymbol{V}}^{k} (\widehat{\boldsymbol{V}}^{k})^{\top} \quad (k = 1, 2, ..., n),$$
(3.58)

com V^k representando a k-ésima coluna de V^{-1} . Portanto, juntamente com \ddot{L} (Apêndice C.13), através das expressões (3.55) a (3.58) pode-se obter o máximo de (3.48) e avaliar a influência local referente ao particular esquema de perturbação.

3.8.2 Perturbação na variável resposta

Beckman et al. (1987) sugerem perturbar o vetor da variável resposta da seguinte forma

$$\boldsymbol{Y}(\boldsymbol{w}) = \boldsymbol{Y} + s\boldsymbol{w},\tag{3.59}$$

com s representando um fator de escala e \boldsymbol{w} um vetor $n \times 1$ de perturbações. Nesse caso $\boldsymbol{w}_0 = \boldsymbol{0}$ e a log-verossimilhança perturbada é dada por

$$\lambda(\boldsymbol{w}) = -(1/2)(\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})^{\top}\boldsymbol{V}^{-1}(\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta}).$$
(3.60)

Considerando (3.60) tem-se (vide Apêndice C.10)

$$\boldsymbol{H}^{\top} = s \boldsymbol{\widehat{V}}^{-1} \left[\boldsymbol{X}, \boldsymbol{\widehat{V}}^{-1} \boldsymbol{\hat{r}}, \boldsymbol{Z} \boldsymbol{\dot{D}}_1 \boldsymbol{Z}^{\top} \boldsymbol{\widehat{V}}^{-1} \boldsymbol{\hat{r}}, \cdots, \boldsymbol{Z} \boldsymbol{\dot{D}}_l \boldsymbol{Z}^{\top} \boldsymbol{\widehat{V}}^{-1} \boldsymbol{\hat{r}} \right].$$
(3.61)

No caso linear normal Schwarzmann (1991) mostrou que observações sensíveis a esse esquema de perturbação apresentam um valor alto para o erro de predição $|y_i - \hat{y}_i|$. O esquema de perturbação acima tem uma forte conexão com o conceito de alavancagem em modelos não lineares [St. Laurent & Cook (1993) e Wei *et al.* (1998)].

3.8.3 Perturbação na matriz de covariâncias dos efeitos aleatórios

Para avaliar a sensibilidade do modelo com relação à suposição de homogeneidade entre as matrizes de covariâncias dos efeitos aleatórios, sugerimos perturbar a matriz de covariâncias de γ_i da seguinte forma: $\operatorname{Var}[\gamma_i] = w_i G$. Nesse caso, o vetor de perturbações \boldsymbol{w} é de dimensão $(c \times 1)$ e $\boldsymbol{w}_0 = \mathbf{1}_c$. Considerando o modelo perturbado, tem-se que $V_i(w) = \text{Var}[Y_i] = w_i Z_i G Z_i^{\top} + \sigma^2 I_{n_i}$, com a log-verossimilhança perturbada sendo dada por

$$\lambda(\boldsymbol{w}) = L(\boldsymbol{\psi}|\boldsymbol{w}) = \sum_{i=1}^{c} (-1/2) \left\{ \ln |\boldsymbol{V}_i(\boldsymbol{w})| + \boldsymbol{r}_i^{\top} \boldsymbol{V}_i(\boldsymbol{w})^{-1} \boldsymbol{r}_i \right\},$$
(3.62)

com $\mathbf{r}_i = \mathbf{\xi}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$ representando o vetor de erros marginais referente a *i-ésima* unidade experimental. A *k-ésima* coluna da matriz \mathbf{H} é dada por

$$\boldsymbol{H}^{k} = \left\{ \left(\frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \boldsymbol{\beta}} \right)^{\top}, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \sigma^{2}}, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \theta_{1}}, \cdots, \frac{\partial^{2} \lambda(\boldsymbol{w})}{\partial w_{k} \partial \theta_{l}} \right\}^{\top},$$
(3.63)

com as respectivas derivadas calculadas em $\boldsymbol{w} = \boldsymbol{w}_0 \in \boldsymbol{\psi} = \boldsymbol{\widehat{\psi}}$. Para $k = 1, ..., c \in j = 1, ..., l$ (veja Apêndice C.11), temos

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}} = \boldsymbol{X}_k^\top \widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \widehat{\boldsymbol{G}} \boldsymbol{Z}_k^\top \widehat{\boldsymbol{V}}_k^{-1} \widehat{\boldsymbol{r}}_k, \qquad (3.64)$$

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \theta_j} \bigg|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}=\boldsymbol{\widehat{\psi}}} = \operatorname{tr} \left[\widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \widehat{\boldsymbol{G}} \boldsymbol{Z}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \dot{\boldsymbol{G}}_j \boldsymbol{Z}_k^{\top} \right] \\ - \hat{\boldsymbol{r}}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \widehat{\boldsymbol{G}} \boldsymbol{Z}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \dot{\boldsymbol{G}}_j \boldsymbol{Z}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \hat{\boldsymbol{r}}_k \qquad (3.65)$$

е

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \sigma^2} \Big|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}=\hat{\boldsymbol{\psi}}} = -\frac{1}{2} \left\{ \operatorname{tr} \left[\widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \widehat{\boldsymbol{G}} \boldsymbol{Z}_k^{\top} \right] - 2 \widehat{\boldsymbol{r}}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \boldsymbol{Z}_k \widehat{\boldsymbol{G}} \boldsymbol{Z}_k^{\top} \widehat{\boldsymbol{V}}_k^{-1} \widehat{\boldsymbol{V}}_k^{-1} \widehat{\boldsymbol{v}}_k^{-1} \widehat{\boldsymbol{r}}_k \right\}, \quad (3.66)$$

 $\operatorname{com} \hat{\boldsymbol{r}}_k = \boldsymbol{Y}_k - \boldsymbol{X}_k \widehat{\boldsymbol{\beta}} \, \mathrm{e} \, \dot{\boldsymbol{G}}_j = \partial \boldsymbol{G} / \partial \theta_j |_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}}.$

Outros esquemas de perturbação são sugeridos em Beckman *et al.* (1987). Esses autores comentam sobre a inadequabilidade da abordagem de influência local, quando utilizada sem modificação no caso em que a primeira derivada de (3.52) avaliada em $\hat{\psi}$ não é nula. Tal preocupação deve-se ao fato de que estimativas negativas das componentes de variância não correspondem aos EMV das mesmas.

3.8.4 Caso ponderado

Recentemente, Verbeke (1995) e Lesaffre & Verbeke (1998) abordaram o problema de avaliar a sensibilidade no ajuste de um modelo linear misto via influência local. Considere o modelo (1.1) com G representando uma matriz simétrica não estruturada. Sob a hipótese de normalidade, a log-verossimilhança do modelo marginal é dada por

$$L(\boldsymbol{\psi}) = \sum_{i=1}^{c} L_{i}(\boldsymbol{\psi}) = \sum_{i=1}^{c} (-1/2) \left\{ \ln |\boldsymbol{V}_{i}| + \boldsymbol{r}_{i}^{\top} \boldsymbol{V}_{i}^{-1} \boldsymbol{r}_{i} \right\},$$
(3.67)

com $L_i(\boldsymbol{\psi})$ representando a log-verossimilhança referente a *i-ésima* unidade experimental. Uma vez que em (3.67) todos os $L_i(\boldsymbol{\psi})$ (i = 1, ..., c) têm a mesma importância, Verbeke (1995) e Lesaffre & Verbeke (1998) surgeriram perturbar esta log-verossimilhança da seguinte forma,

$$L_i(\boldsymbol{\psi}|\boldsymbol{w}) = \sum_{i=1}^c w_i L_i(\boldsymbol{\psi}), \qquad (3.68)$$

em que \boldsymbol{w} é um vetor $c \times 1$ de perturbações. Nesse caso tem-se $\boldsymbol{w}_0 = \mathbf{1}_c$. Esta abordagem é adequada para avaliar a sensibilidade referente a uma **unidade experimental** (indivíduo).

Considerando a curvatura normal (3.48) do afastamento da verossimilhança (3.46) na direção do *i-ésimo* indivíduo, ou seja, com d_i representando um vetor com valor 1 na *i-ésima* posição e zero nas demais, então a curvatura normal calculada nessa direção é

$$C_{i} = 2 \left| \boldsymbol{d}_{i}^{\top} \boldsymbol{H}^{\top} \ddot{\boldsymbol{L}}^{-1} \boldsymbol{H} \boldsymbol{d}_{i} \right| = 2 \left| \boldsymbol{H}_{i}^{\top} \ddot{\boldsymbol{L}}^{-1} \boldsymbol{H}_{i} \right|, \qquad (3.69)$$

com \boldsymbol{H}_i representando a *i-ésima* coluna da matriz \boldsymbol{H} . C_i é denominada influência local referente ao *i-ésimo* indivíduo[Verbeke (1995) e Lesaffre & Verbeke (1998)]. Verbeke (1995) mostrou que C_i é assintoticamente $(c \to \infty)$ igual a $2\rho_i$, com $\rho_i = -(\widehat{\boldsymbol{\psi}} - \widehat{\boldsymbol{\psi}}_{(i)}^1)^\top \ddot{\boldsymbol{L}}(\boldsymbol{\psi})^{-1}(\widehat{\boldsymbol{\psi}} - \widehat{\boldsymbol{\psi}}_{(i)}^1)$ representando a proposta de Pregibon (1981) para medir a influência da *i-ésima* observação, via aproximação por 1 passo de $\widehat{\boldsymbol{\psi}}_{(i)}$. Nesse sentindo, um alto valor de C_i indica que o *i-ésimo* indivíduo tem um grande impacto na estimativa de $\boldsymbol{\psi}$, tanto no sentindo de influência local como global.

Uma escolha muito utilizada consiste em considerar a direção de maior curvatura (\boldsymbol{d}_{max}) . Os componentes de \boldsymbol{d}_{max} e a medida de influência (3.69) podem conter informações distintas, uma vez que (vide Apêndice C.12)

$$C_i = 2\sum_{j=1}^c \lambda_j \boldsymbol{v}_{ji}^2, \qquad (3.70)$$

com $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_c$ denotando os c autovalores de $-\mathbf{H}^\top \ddot{\mathbf{L}}^{-1}\mathbf{H}$ e $\mathbf{d}_{max} \equiv \mathbf{v}_1, \cdots, \mathbf{v}_c$ os autovetores ortogonais normalizados correspondentes, com \mathbf{v}_{ji} representando o *i-ésimo* componente do vetor \mathbf{v}_j . Os casos individuais podem apresentar um valor alto para C_i sem que o *i-ésimo* componente de \mathbf{d}_{max} seja elevado. Assim, é razoável investigar os autovalores intermediários de $-\mathbf{H}^\top \ddot{\mathbf{L}}^{-1}\mathbf{H}$, quando eles não são muito menores que λ_1 .

Se o interesse é avaliar a influência local do i-ésimo indivíduo apenas nas estimativas dos parâmetros fixos ou somente na estimativa dos parâmetros de covariância, então (3.50)

fica dada por

$$C_{i}(\boldsymbol{\beta}) = 2 \left| \boldsymbol{H}_{i}^{\top} \left\{ \ddot{\boldsymbol{L}}^{-1} - \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddot{\boldsymbol{L}}_{22}^{-1} \end{pmatrix} \right\} \boldsymbol{H}_{i} \right|$$
(3.71)

ou

$$C_{i}(\boldsymbol{\theta},\sigma^{2}) = 2 \left| \boldsymbol{H}_{i}^{\top} \left\{ \ddot{\boldsymbol{L}}^{-1} - \begin{pmatrix} \ddot{\boldsymbol{L}}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \right\} \boldsymbol{H}_{i} \right|.$$
(3.72)

Diferentemente da regressão linear clássica, o BLUE de β depende da estrutura de covariância envolvida, e conseqüentemente, a influência de um indivíduo na estimativa dos efeitos fixos envolve aspectos referentes aos parâmetros de covariância. Lesaffre & Verbeke (1998) argumentam que para avaliar a influência das observações nesse caso, é preciso considerar características distintas dos métodos utilizados na regressão clássica. Esses autores reparametrizam os elementos da diagonal principal de G, denotados por g_{kk} , substituindo-os por $\sqrt{2}g_{kk}$, com o objetivo de simplificar as expressões das derivadas da log-verossimilhança. Podemos escrever C_i como

$$C_i = 2||\ddot{\boldsymbol{L}}^{-1}||\cos\phi_i||\boldsymbol{H}_i||^2, \qquad (3.73)$$

em que ϕ_i representa o ângulo entre $\operatorname{vec}(-\ddot{\boldsymbol{L}}^{-1})$ e $\operatorname{vec}(\boldsymbol{H}_i\boldsymbol{H}_i^{\top})$ e $||\boldsymbol{A}|| = |\operatorname{vec}(\boldsymbol{A})|$ denota a norma de Frobenius da matriz \boldsymbol{A} [Graybill (1983, p. 94)]). A idéia de Lesaffre & Verbeke (1998) foi decompor $||\boldsymbol{H}_i||^2$ como a soma dos quadrados das normas da contribuição da *i-ésima* unidade experimental para o vetor *score* de $\boldsymbol{\beta}, \boldsymbol{\theta} \in \sigma^2$. Assim, tem-se que $C_i = a_i + b_i + d_i$ com

$$a_i = 2||\ddot{\boldsymbol{L}}^{-1}||\cos\phi_i||\boldsymbol{X}_i^{\top}\widehat{\boldsymbol{V}}_i^{-1}\widehat{\boldsymbol{r}}_i||^2, \qquad (3.74)$$

$$b_{i} = ||\ddot{\boldsymbol{L}}^{-1}||\cos\phi_{i}||\boldsymbol{Z}_{i}^{\top}\widehat{\boldsymbol{V}}_{i}^{-1}\boldsymbol{Z}_{i} - \boldsymbol{Z}_{i}^{\top}\widehat{\boldsymbol{V}}_{i}^{-1}\hat{\boldsymbol{r}}_{i}\widehat{\boldsymbol{r}}_{i}^{\top}\widehat{\boldsymbol{V}}_{i}^{-1}\boldsymbol{Z}_{i}||^{2}$$
(3.75)

е

$$d_{i} = \frac{1}{2} ||\ddot{\boldsymbol{L}}^{-1}||\cos\phi_{i}||\operatorname{tr}\{\widehat{\boldsymbol{V}}_{i}^{-1}\} - \hat{\boldsymbol{r}}_{i}^{\top}\widehat{\boldsymbol{V}}_{i}^{-1}\widehat{\boldsymbol{V}}_{i}^{-1}\hat{\boldsymbol{r}}_{i}||^{2}.$$
(3.76)

Uma vez que $\frac{\partial L_i(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i, \left\| \frac{\partial L_i(\boldsymbol{\psi})}{\partial \boldsymbol{\theta}} \right\|^2 = \frac{1}{2} \left\| \boldsymbol{Z}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i - \boldsymbol{Z}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i \boldsymbol{r}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \right\|^2$ e $\frac{\partial L_i(\boldsymbol{\psi})}{\partial \sigma^2} = -\frac{1}{2} \left(\operatorname{tr} \{ \boldsymbol{V}_i^{-1} \} - \boldsymbol{r}_i^{\top} \boldsymbol{V}_i^{-1} \boldsymbol{V}_i^{-1} \boldsymbol{r}_i \right) \text{ tem-se}$ $\| \boldsymbol{H}_i \|^2 = \| \boldsymbol{X}_i^{\top} \boldsymbol{\hat{V}}_i^{-1} \hat{\boldsymbol{r}}_i \|^2 + \frac{1}{2} \| \boldsymbol{Z}_i^{\top} \boldsymbol{\hat{V}}_i^{-1} \boldsymbol{Z}_i - \boldsymbol{Z}_i^{\top} \boldsymbol{\hat{V}}_i^{-1} \hat{\boldsymbol{r}}_i \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{Z}_i \|^2$ $+ \frac{1}{4} \| \operatorname{tr} \{ \boldsymbol{\hat{V}}_i^{-1} \} - \hat{\boldsymbol{r}}_i^{\top} \boldsymbol{\hat{V}}_i^{-1} \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{r}}_i \|^2, \qquad (3.77)$

por (3.77), garante-se a validade de (3.74), (3.75) e (3.76). Definindo $\mathcal{R}_i = \hat{V}_i^{-1/2} \hat{r}_i$, $\mathcal{X}_i = \hat{V}_i^{-1/2} X_i$ e $\mathcal{Z}_i = \hat{V}_i^{-1/2} Z_i$, Lesaffre & Verbeke (1998) reescrevem (3.74), (3.75) e (3.76) como

$$a_{i} = 2\left\{\cos\phi_{i}\cos\psi_{i}||\ddot{\boldsymbol{L}}^{-1}||\right\}||\boldsymbol{\mathcal{X}}_{i}\boldsymbol{\mathcal{X}}_{i}^{\top}||^{2}||\boldsymbol{\mathcal{R}}_{i}||^{2}, \qquad (3.78)$$

$$b_{i} = \left\{ \cos \phi_{i} \cos \kappa_{i} || \ddot{\boldsymbol{L}}^{-1} || \right\} || \boldsymbol{\mathcal{Z}}_{i} \boldsymbol{\mathcal{Z}}_{i}^{\top} ||^{2} || \boldsymbol{I}_{n_{i}} - \boldsymbol{\mathcal{R}}_{i} \boldsymbol{\mathcal{R}}_{i}^{\top} ||^{2}$$
(3.79)

е

$$d_{i} = \frac{1}{2} \left\{ \cos \phi_{i} \cos^{2} \nu_{i} || \ddot{\boldsymbol{L}}^{-1} || \right\} || \widehat{\boldsymbol{V}}_{i}^{-1} ||^{2} || \boldsymbol{I}_{n_{i}} - \boldsymbol{\mathcal{R}}_{i} \boldsymbol{\mathcal{R}}_{i}^{\top} ||^{2},$$
(3.80)

com ψ_i representando o ângulo entre $\operatorname{vec}(\boldsymbol{\mathcal{X}}_i \boldsymbol{\mathcal{X}}_i^{\top})$ e $\operatorname{vec}(\boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^{\top}), \nu_i$ o ângulo entre $\operatorname{vec}(\boldsymbol{\widehat{V}}_i^{-1})$ e $\operatorname{vec}(\boldsymbol{I}_n - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^{\top})$ enquanto que κ_i representa o ângulo entre $\operatorname{vec}(\boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^{\top} \otimes \boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^{\top})$ e $\operatorname{vec}\left[\operatorname{vec}(\boldsymbol{I}_{n_i} - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^{\top})\operatorname{vec}(\boldsymbol{I}_{n_i} - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^{\top})^{\top}\right]$ [Verbeke (1995)], que constituem os termos não interpretáveis de a_i, b_i e d_i , respectivamente; $||\boldsymbol{\overset{-}{L}}^{-1}||$ é a parte comum a todas as componentes. Os ângulos em questão não são utilizados, não pelo fato de que os mesmos sejam neglegeciáveis, porém pelo fato de não possuírem uma interpretação clara [Lesaffre & Verbeke (1998)].

Desta forma, Lesaffre & Verbeke (1998) sugerem utilizar os termos interpretáveis de a_i , $b_i \in d_i$

$$||\boldsymbol{\mathcal{X}}_{i}\boldsymbol{\mathcal{X}}_{i}^{\top}||^{2}, ||\boldsymbol{\mathcal{R}}_{i}||^{2}, ||\boldsymbol{\mathcal{Z}}_{i}\boldsymbol{\mathcal{Z}}_{i}^{\top}||^{2}, ||\boldsymbol{I}_{n_{i}} - \boldsymbol{\mathcal{R}}_{i}\boldsymbol{\mathcal{R}}_{i}^{\top}||^{2}, ||\boldsymbol{\widehat{V}}_{i}^{-1}||^{2},$$
(3.81)

para avaliar a influência do *i-ésimo* indivíduo para o modelo linear misto considerado.

Os termos interpretáveis de a_i consistem de $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^{\top}||^2$, o comprimento das covariáveis padronizadaspara os efeitos fixos, e $||\boldsymbol{\mathcal{R}}_i||^2$, o comprimento ao quadrado dos resíduos padronizados. Se C_i é alto devido ao a_i , então a influência da unidade experimental em questão pode ser causada por um valor alto de $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^{\top}||^2$ e/ou de $||\boldsymbol{\mathcal{R}}_i||^2$. Um alto valor de a_i pode estar associado por uma unidade experimental que tem muitas observações ou que não é bem predita pelo modelo. Quando, se tem um estudo balanceado, $||\boldsymbol{\mathcal{X}}_i\boldsymbol{\mathcal{X}}_i^{\top}||^2$ é diretamente comparável para todos os indivíduos.

O termo b_i tende a assumir um valor elevado quando $||\boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^\top||^2 e/ou ||\boldsymbol{I}_{n_i} - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^\top||^2$ assumem valores altos. Similarmente, ao caso anterior, $||\boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^\top||^2$ tende a assumir um valor elevado para uma unidade experimental com muitas observações. Por outro lado Lesaffre & Verbeke (1998) argumentam que $||\boldsymbol{I}_{n_i} - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^\top||^2$ tende a ser próximo de zero quando \boldsymbol{V}_i é "próxima" de $\boldsymbol{r}_i \boldsymbol{r}_i^\top$. Uma vez que Var $[\boldsymbol{Y}_i]$ pode ser estimada por $\boldsymbol{r}_i \boldsymbol{r}_i^\top$ quando o vetor de médias é corretamente modelado por $\mathbf{X}_i \boldsymbol{\beta}$, podemos interpretar $||\mathbf{I}_{n_i} - \mathcal{R}_i \mathcal{R}_i^{\top}||^2$ como um resíduo que mede a validade da estrutura de covariância adotada. Portanto, b_i tende a assumir um valor alto para uma unidade experimental com muitas observações com a respectiva matriz de covariâncias mal ajustada.

Por outro lado, d_i tende a ser alto quando $||I_{n_i} - \mathcal{R}_i \mathcal{R}_i^{\top}||^2$ e/ou $||\widehat{V}_i^{-1}||^2$ assument valores altos. Porém, $||\widehat{V}_i^{-1}||^2$ tende a assumir um valor alto quando V_i tem pequenos autovalores, ou seja, quando a variabilidade referente a *i-ésima* unidade experimental é baixa [Lesaffre & Verbeke (1998)]. Portanto, d_i tende a assumir um valor alto, para um indivíduo com pequena variabilidade e com respectiva matriz de covariâncias mal ajustada.

Verbeke (1995) mostrou que $\sum_{i=1}^{c} C_i = -2 \operatorname{tr} \left\{ \ddot{\boldsymbol{L}}^{-1} \sum_{i=1}^{c} \boldsymbol{H}_i \boldsymbol{H}_i^{\mathsf{T}} \right\}$ converge para 2s (s: número de parâmetros). Desta forma, ele considera a *i-ésima* unidade experimental influente se $C_i > 2\bar{C} \approx 4s/c$, quando tem um número suficientemente grande de unidades experimentais. Já para os termos em (3.81), é difícil definir pontos de corte, então sugerese a comparação dos termos supracitados associados às unidades experimentais e define-se um ponto de corte de forma totalmente descritiva.

Notando que as quantidades em (3.81) podem ser afetadas pelo valor de n_i , Lesaffre & Verbeke (1998) sugerem cautela ao analisá-las. Em estudos não-balanceados, os autores sugerem o uso da proposta de Waternaux *et al.* (1989), que comparam $||\mathcal{R}_i||^2$ com os quantis de uma distribuição $\chi^2_{n_i}$. Uma sugestão dada pelos autores é construir gráficos das quantidades em (3.81), juntamente com o gráfico de n_i , devido à incerteza com relação ao tipo de correção a ser feita com relação a dimensionalidade.

De forma similar, Lesaffre & Verbeke (1998) decompõem $C_i(\boldsymbol{\beta}) \in C_i(\boldsymbol{\theta}, \sigma^2)$. O processo é análogo ao anterior, trocando apenas a matriz \boldsymbol{L} por uma matriz \boldsymbol{A}^{-1} apropriada [veja (3.71) e (3.72)] e ϕ_i por ϕ_i^* , que é o ângulo entre vec $(-\boldsymbol{A}^{-1})$ e vec $(\boldsymbol{H}_i \boldsymbol{H}_i^{\top})$. Ao contrário de Christensen *et al.* (1992), Lesaffre & Verbeke (1998), sugerem avaliar a influência conjuntamente nas estimativas dos efeitos fixos e das componentes de covariância, pois eles só são assintoticamente independentes. Para $c \to \infty$, $\boldsymbol{\beta} \in \theta^*$ são ortogonais, e $C_i \approx$ $C_i(\boldsymbol{\beta}) + C_i(\boldsymbol{\theta}, \sigma^2)$, com

$$C_{i}(\boldsymbol{\beta}) \approx -2\boldsymbol{H}_{1i}^{\top} \boldsymbol{\ddot{L}}_{11}^{-1} \boldsymbol{H}_{1i} \approx \boldsymbol{\mathcal{R}}_{i}^{\top} \boldsymbol{\mathcal{X}}_{i} \left(\sum_{i=1}^{c} \boldsymbol{X}_{i}^{\top} \boldsymbol{\widehat{V}}_{i}^{-1} \boldsymbol{X}_{i} \right)^{-1} \boldsymbol{\mathcal{X}}_{i}^{\top} \boldsymbol{\mathcal{R}}_{i}$$
$$= \boldsymbol{\widehat{r}}_{i}^{\top} \boldsymbol{\widehat{V}}_{i}^{-1} \boldsymbol{X}_{i} \left(\boldsymbol{X}^{\top} \boldsymbol{\widehat{V}}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}_{i}^{\top} \boldsymbol{\widehat{V}}_{i}^{-1} \boldsymbol{\widehat{r}}_{i}, \qquad (3.82)$$

e $C_i(\boldsymbol{\theta}, \sigma^2) \approx -2\boldsymbol{H}_{2i}^{\top} \boldsymbol{L}_{22}^{-1} \boldsymbol{H}_{2i}$, com \boldsymbol{H}_{1i} (\boldsymbol{L}_{11}) e \boldsymbol{H}_{2i} (\boldsymbol{L}_{22}) representando as partições da matriz \boldsymbol{H} (\boldsymbol{L}) referentes a $\boldsymbol{\beta} \in \boldsymbol{\theta}^*$, respectivamente. A decomposição de $C_i(\boldsymbol{\beta})$ apresenta

somente $||\boldsymbol{\mathcal{X}}_{i}\boldsymbol{\mathcal{X}}_{i}^{\top}||^{2}$ e $||\boldsymbol{\mathcal{R}}_{i}||^{2}$ como termos interpretáveis, enquanto que a decomposição de $C_{i}(\boldsymbol{\theta},\sigma^{2})$ inclui os demais termos de (3.81). Na prática, se o interesse maior é a influência em relação a $\hat{\boldsymbol{\beta}}$, condicionada à estimação de $\boldsymbol{\theta}^{*}$, então é suficiente analisar $||\boldsymbol{\mathcal{X}}_{i}\boldsymbol{\mathcal{X}}_{i}^{\top}||^{2}$ e $||\boldsymbol{\mathcal{R}}_{i}||^{2}$. Por outro lado, se o objetivo principal é analisar a influência nos parâmetros de covariância, basta concentrar a atenção nos termos interpretáveis de $C_{i}(\boldsymbol{\theta},\sigma^{2})$. Para definir os pontos de corte para $C_{i}(\boldsymbol{\beta})$ e $C_{i}(\boldsymbol{\theta},\sigma^{2})$ basta utilizar uma idéia semelhante a usada anteriormente, considerando que $\sum_{i=1}^{c} C_{i}(\boldsymbol{\beta}) \approx 2p$ e $\sum_{i=1}^{c} C_{i}(\boldsymbol{\theta},\sigma^{2}) \approx q(q+1)+2$, quando $c \to \infty$ [Verbeke (1995)].

Tendo em vista que (3.82) é uma combinação dos resíduos $\hat{\boldsymbol{r}}_i$, da alavancagem do *i*ésimo indivíduo e da matriz de covariâncias $\hat{\boldsymbol{V}}_i$ para c grande $C_i(\boldsymbol{\beta})$ tem característica semelhante ao da regressão clássica, no sentindo de ser uma combinação da alavancagem e do resíduo [Cook (1986) e Beckman *et al.* (1987)].

A principal vantagem da abordagem utilizada por Lesaffre & Verbeke (1998) é a decomposição de C_i em termos interpretáveis, facilitando assim a identificação das razões da alta influência. Como d_{max} não tem expressão analítica para este problema, esse termo não fornece idéia a respeito da alta influência de uma observação. Além disso, Lesaffre & Verbeke (1998) comentam que a abordagem proposta por eles difere da proposta apresentada em Beckman *et al.* (1987) e deve equivaler à abordagem de Christensen *et al.* (1992), quando aplicada dentro do contexto de medidas repetidas, para um número muito grande de unidades experimentais.

Todos os termos de (3.81) dependem diretamente de V_i e por conseguinte devem ser especificadas de forma **correta** para garantir a validade do processo de diagnóstico proposto. No presente trabalho, assumimos que as covariáveis e a matriz de covariâncias já tenham sido previamente especificadas e não nos atentaremos a este fato. Para detalhes referentes a métodos de especificação dos efeitos (fixos e aleatórios) e da matriz de covariâncias, veja, por exemplo, Wolfinger (1993), Pinheiro (1994), Verbeke (1995), Keselman *et al.* (1998) e Rocha (2004).

Todos os resultados obtidos por Lesaffre & Verbeke (1998) baseiam-se no método de máxima verossimilhança. Esses autores comentam que tal decomposição não pode ser feita, quando consideramos o método de máxima verossimilhança restrita, uma vez que a log-verossimilhança restrita nem sempre pode ser escrita como uma soma de contribuições individuais independentes. Eles também consideram a matriz G não estruturada. Porém decomposição similar é válida quando admitimos uma estrutura particular para a referida matriz.

Outras propostas são dadas em Ouwens et al. (2001) e Zhu & Lee (2001, 2003) que

discutem a aplicação da metodologia de influência local em MLGM. Dentro do contexto do modelo linear geral de Bayes, pode-se utilizar a proposta de McCulloch (1989) para analisar a sensibilidade do BLUE e BLUP sob o modelo (1.2) quando se perturbam as distribuições a priori do erro e dos efeitos aleatórios.

Capítulo $\mathbf{4}$

Aplicação

4.1 Introdução

No presente capítulo aplicamos as propostas de diagnóstico discutidas nos capítulos 2 e 3 ao modelo ajustado aos dados do estudo descrito no capítulo 1. Pelas Figuras 1.1 e 1.2 temos indicação de existência de associação entre os índices de placa bacteriana pré-escovação (x) e pós-escovação (y), para os dois tipos de escova.

Singer & Andrade (1997) analisaram um problema semelhante e apontaram as seguintes características que o modelo para esse tipo de dados deve apresentar:

- (i) Um índice pré-tratamento nulo implica um índice pós-tratamento também nulo;
- (ii) Os índices pré-tratamento e pós-tratamento são não-negativos;
- (iii) Os dados são possivelmente heterocedásticos (pois são não-negativos e satisfazem a desigualdade $y \le x$);
- (iv) A relação entre os índices pré-tratamento e pós-tratamento é possivelmente nãolinear;
- (v) As observações realizadas numa mesma unidade experimental são possivelmente correlacionadas.

Os autores propuseram o seguinte modelo:

$$y = \beta x^{\delta} \xi, \tag{4.1}$$

em que x é o índice de placa bacteriana pré-escovação, y é o índice de placa bacteriana pós-escovação, $\beta > 0$ é um coeficiente de placa bacteriana residual, δ é um coeficiente de uniformidade da taxa de placa bacteriana residual esperada e ξ é um erro aleatório não-negativo. Sob a validade do modelo (4.1), $\mathbb{E}[Y]/x = \beta x^{\delta-1}\mathbb{E}[\xi]$ representando a taxa esperada residual de placa bacteriana pós-escovação por unidade de índice de placa bacteriana pré-escovação. Se $\delta = 1$ essa taxa é constante, por outro lado, se $\delta > 1$ ($\delta < 1$) a taxa é crescente (decrescente) com o índice de placa bacteriana pré-tratamento, indicando uma menor (maior) eficácia da escova. Quanto maior for o coeficiente de β menor será a eficácia da escova na remoção do índice de placa bacteriana.

Note que, sob o modelo inicialmente proposto, as condições (i) e (ii) são automaticamente satisfeitas. Além disso, tem-se que $\operatorname{Var}[Y] = (\beta x^{\delta})^2 \operatorname{Var}[\xi]$, satisfazendo a possível heterocedasticidade mencionada em (iii). Se $\delta \neq 1$, tem-se uma relação não linear, satisfazendo a suposição (iv). Já a possível correlação existente entre observações de uma mesma unidade experimental, pode ser imposta no modelo através da especificação de uma estrutura de covariância adequada para ξ . O ajuste do modelo (4.1) pode ser feito via modelos lineares para medidas repetidas se considerarmos a transformação logarítmica. Vantagens e desvantagens desta transformação são discutidas em Singer *et al.* (2004), que apresentam também modelos alternativos ajustados via metodologia de equações de estimação generalizadas.

4.2 Especificação do modelo

Utilizando as propostas de Singer *et al.* (2004) sugerimos o seguinte modelo para os dados apresentados na Tabela (1.1):

$$y_{ijd} = \beta_{jd} x_{ijd}^{\delta_{jd}} \xi_{ijd}, \tag{4.2}$$

com $\beta_{jd} > 0, i = 1, 2, ..., 32, j = 0, 1, d = 1, 2, 3, 4$, em que y_{ijd} (x_{ijd}) é o índice de placa bacteriana pós-tratamento (pré-tratamento) relativo à *i-ésima* criança com a *j-ésima* escova (j = 0: escova convencional e j = 1: escova monobloco) na *d-ésima* sessão de avaliação, β_{jd} é um coeficiente de placa bacteriana residual relativo à *j-ésima* escova e à *d-ésima* sessão de avaliação, δ_{jd} é um coeficiente de uniformidade da taxa de placa residual esperada relativo à *j-ésima* escova e a *d-ésima* sessão de avaliação e ξ_{ijd} é um erro aleatório não-negativo relativo à *i-ésima* criança com a *j-ésima* escova na *d-ésima* sessão de avaliação. O modelo (4.2) pode ser linearizado por intermédio da seguinte transformação

$$\ln y_{ijd} = \ln \beta_{jd} + \delta_{jd} \ln x_{ijd} + \ln \xi_{ijd}$$

$$(4.3)$$

que pode ser reparametrizado por

$$y_{ijd}^* = \lambda_{jd} + \delta_{jd} x_{ijd}^* + \xi_{ijd}^*, \tag{4.4}$$

Nobre, Juvêncio S.

IME-USP

em que $y_{ijd}^* = \ln y_{ijd}$, $\lambda_{jd} = \ln \beta_{jd}$, $x_{ijd}^* = \ln x_{ijd}$. Assumimos que $\ln \xi_{ijd} = \xi_{ijd}^*$ tem distribuição normal com parâmetros a serem especificados. Para satisfazer a característica (v), consideramos

$$\xi_{ijd}^* = \psi_i + \varepsilon_{ijd},\tag{4.5}$$

com $\psi_i \sim N(0, \tau^2)$ e $\varepsilon_{ijd} \sim N(0, \sigma^2)$, denotando respectivamente, o efeito aleatório da *i-ésima* criança e o erro de medida. Com essas especificações, podemos escrever o modelo (4.4) na forma matricial

$$\ln \boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{\psi}_i + \boldsymbol{\varepsilon}_i, \qquad (4.6)$$

em que $\boldsymbol{\beta} = (\ln \beta_{01}, \ln \beta_{02}, \cdots, \ln \beta_{13}, \ln \beta_{14}, \delta_{01}, \delta_{02}, \cdots, \delta_{13}, \delta_{14})^{\top}$ e $\boldsymbol{Z}_i = \boldsymbol{1}_4$. Para i = 1, ..., 16, temos

$$\begin{aligned} \boldsymbol{Y}_{i} &= (y_{i01}, y_{i02}, y_{i03}, y_{i04})^{\top}, \\ \boldsymbol{\varepsilon}_{i} &= (\varepsilon_{i01}, \varepsilon_{i02}, \varepsilon_{i03}, \varepsilon_{i04})^{\top}, \\ \boldsymbol{X}_{i} &= \left[\boldsymbol{I}_{4} \bigotimes (1, 0) \stackrel{:}{:} \bigoplus_{d=1}^{4} \ln x_{i0d} \stackrel{:}{:} \boldsymbol{0}_{4 \times 4} \right], \end{aligned}$$

com $\bigoplus_{d=1}^{n} \ln x_{i0d} = \text{diag}(\ln x_{i01}, \ln x_{i02}, \ln x_{i03}, \ln x_{i04})$. Para i = 17, ..., 32, temos

$$\begin{aligned} \boldsymbol{Y}_{i} &= (y_{i11}, y_{i12}, y_{i13}, y_{i14})^{\top}, \\ \boldsymbol{\varepsilon}_{i} &= (\varepsilon_{i11}, \varepsilon_{i12}, \varepsilon_{i13}, \varepsilon_{i14})^{\top}, \\ \boldsymbol{X}_{i} &= \left[\boldsymbol{I}_{4} \bigotimes (0, 1, 0) \stackrel{:}{:} \bigoplus_{d=1}^{4} \ln x_{i1d} \right]. \end{aligned}$$

Adotamos uma estrutura auto-regressiva de primeira ordem, AR(1) para a matriz de covariâncias associada ao vetor de erros ε_i , visando contemplar a expectativa de correlações maiores para observações adjacentes. Com as suposições supracitadas tem-se que

$$\boldsymbol{V}_{i} = \boldsymbol{Z}_{i} \tau^{2} \boldsymbol{Z}_{i}^{\top} + \boldsymbol{\Sigma}_{i} = \tau^{2} \boldsymbol{1}_{4} \boldsymbol{1}_{4}^{\top} + \sigma^{2} \begin{pmatrix} 1 & \rho & \rho^{2} & \rho^{3} \\ \rho & 1 & \rho & \rho^{2} \\ \rho^{2} & \rho & 1 & \rho \\ \rho^{3} & \rho^{2} & \rho & 1 \end{pmatrix}, \qquad (4.7)$$

em que ρ representa o coeficiente de correlação entre duas observações adjacentes. Tentouse simplificar a estrutura de covariâncias (4.7) sob o modelo saturado (4.2) conforme sugerido em Diggle *et al.* (2002). Primeiramente, não se rejeita a hipótese de que $\rho = 0$ dado que $\hat{\rho} = 0$. Portanto, é razoável reduzir (4.7) para

$$\boldsymbol{V}_{i} = \begin{pmatrix} \tau^{2} + \sigma^{2} & \tau^{2} & \tau^{2} & \tau^{2} \\ \tau^{2} & \tau^{2} + \sigma^{2} & \tau^{2} & \tau^{2} \\ \tau^{2} & \tau^{2} & \tau^{2} + \sigma^{2} & \tau^{2} \\ \tau^{2} & \tau^{2} & \tau^{2} & \tau^{2} + \sigma^{2} \end{pmatrix}.$$
(4.8)

Neste caso, estamos sob o modelo de independência condicional, ou seja $\Sigma_i = \sigma^2 I_4$. Os logaritmos das log-verossimilhanças maximizadas correspondentes ao modelo (4.2) sob as estruturas de covariância (4.7) e (4.8) são ambos iguais a 65.5. Quando não estruturamos Σ_i , o logaritmo da verossimilhança maximizada é 67.2, indicando assim a plausibilidade da redução. A seguir descrevemos a estratégia da análise utilizada para simplificar o modelo saturado (4.2) com estrutura de covariância (4.8):

(i) Testar a homogeneidade entre os coeficientes de uniformidade para as duas escovas nas quatro sessões de avaliação ($\delta_{jd} = \delta, j = 0, 1 \in d = 1, ..., 4$;), ou seja, reduzir o modelo (4.2) para

$$y_{ijd} = \beta_{jd} x_{ijd}^{\delta} \xi_{ijd}; \tag{4.9}$$

 (ii) Testar a significância da interação e dos efeitos principais dos tipos de escova com relação aos coeficientes de placa bacteriana residual, ou seja,

$$\beta_{01}/\beta_{11} = \beta_{02}/\beta_{12} = \beta_{03}/\beta_{13} = \beta_{04}/\beta_{14},$$

ou equivalentemente,

$$\lambda_{01} - \lambda_{11} = \lambda_{02} - \lambda_{12} = \lambda_{03} - \lambda_{13} = \lambda_{04} - \lambda_{14}$$

е

$$\beta_{jd} = \beta_j,$$

ou equivalentemente,

$$\lambda_{jd} = \lambda_j, \quad d = 1, 2, 3, 4, \quad j = 0, 1.$$

 (iii) Ajustar o modelo que incorpora as conclusões obtidas em (i) e (ii), ou seja, reduzir o modelo (4.9) para

$$y_{ijd} = \beta_j x_{ijd}^{\delta} \xi_{ijd}, \tag{4.10}$$

IME-USP

Os testes utilizados basearam-se nas razões de verossimilhanças e teste F aproximado, ambos implementados no procedimento MIXED do SAS e discutidos na seção 1.4.

Na Tabela 4.1, estão apresentadas as estimativas de MV dos parâmetros dos modelos (4.2), (4.9) e (4.10) considerando a estrutura de covariância (4.8). Para encontrar os errospadrão de $\hat{\beta}_{ij}$ utilizou-se o método Delta (Sen & Singer, 1993). Para os parâmetros de covariância, foram construidos intervalos com 95% de confiança.

Parâmetros	Modelo (4.2)	Modelo (4.9)	Modelo (4.10)
$\beta_{01} = \exp(\lambda_{01})$	0.76 ± 0.04	0.74 ± 0.03	0.72 ± 0.02
$\beta_{02} = \exp(\lambda_{02})$	0.65 ± 0.03	0.65 ± 0.03	-
$\beta_{03} = \exp(\lambda_{03})$	0.74 ± 0.04	0.74 ± 0.03	-
$\beta_{04} = \exp(\lambda_{04})$	0.86 ± 0.08	0.78 ± 0.03	-
$\beta_{11} = \exp(\lambda_{11})$	0.82 ± 0.04	0.84 ± 0.04	0.81 ± 0.02
$\beta_{12} = \exp(\lambda_{12})$	0.83 ± 0.04	0.83 ± 0.03	-
$\beta_{13} = \exp(\lambda_{13})$	0.79 ± 0.04	0.79 ± 0.03	-
$\beta_{14} = \exp(\lambda_{14})$	0.71 ± 0.05	0.79 ± 0.03	-
δ_{01}	0.88 ± 0.13	1.01 ± 0.07	1.06 ± 0.06
δ_{02}	1.00 ± 0.13	-	-
δ_{03}	1.02 ± 0.13	-	-
δ_{04}	0.79 ± 0.21	-	-
δ_{11}	1.11 ± 0.14	-	-
δ_{12}	1.02 ± 0.17	-	-
δ_{13}	0.97 ± 0.21	-	-
δ_{14}	1.40 ± 0.20	-	-
$ au^2$	[0.004; 0.021]	[0.004; 0.022]	[0.004; 0.022]
σ^2	[0.013; 0.022]	[0.013; 0.023]	[0.016; 0.028]

Tabela 4.1 Estimativas (\pm EP) dos parâmetros dos modelos (4.2), (4.9) e (4.10) com estrutura de covariâncias (4.8).

Tendo em vista que $\hat{\beta}_0 < \hat{\beta}_1$, o modelo (4.10) sugere que a escova convencional é mais eficaz do que a escova monobloco na manutenção da capacidade de remoção de placa bacteriana. Como $\hat{\delta} > 1$ podemos concluir que a taxa esperada residual de placa bacteriana pós-escovação por unidade de índice de placa bacteriana pré-escovação é uma função crescente do índice de placa bacteriana pré-escovação. A Figura 4.1 representa o ajuste do modelo final.



Figura 4.1 Ajuste do modelo final.

As observações representadas por • referem-se as crianças que utilizaram escova do tipo monobloco.

4.3 Diagnóstico do modelo ajustado

Os resíduos definidos no Capítulo 2, referentes ao modelo ajustado estão apresentados nas Figuras 4.2 e 4.3.

Analisando a Figura 4.2 (a), temos indicações da validade da hipótese de linearidade, pois não se observa nenhum tipo de tendência do resíduo marginal conforme o valor de $\ln x_{ijd}$. Uma análise da Figura 4.2 (b) mostra que a unidade experimental #29 apresenta um comportamento atípico comparado com as demais; tal comportamento produz uma acentuada assimetria na distribuição observada dos EBLUP e deve também influenciar a estimativa de τ^2 .

Pela Figura 4.3, não se percebe nenhum afastamento da normalidade por parte do resíduo com confundimento mínimo, indicando a plausibilidade da suposição de normalidade por parte do erro condicional; duas observações 12.2, a observação referente a segunda sessão da criança 12, e 29.4, a observação referente a quarta sessão da criança 29, destacam-se perante as demais no que tange ao valor do resíduo condicional padro-



Figura 4.2 Resíduo marginal e EBLUP do modelo final (4.10).

Figura 4.3 Resíduo condicional padronizado e envelope simulado com 95% para o resíduo com confundimento mínimo.



nizado; por conseguinte, tais observações são consideradas como possíveis observações discrepantes e influentes com relação à estimativa de σ^2 .

Na Figura 4.4 mostramos a alavancagem generalizada por observação e por unidade experi-mental para os efeitos fixos e para os efeitos fixos e aleatórios. Com base nas referidas figuras, as unidades experimentais #11 e #12 são consideradas unidades experimentais alavanca nos dois casos.

A seguir consideramos todos os tipos de perturbação discutidos no Capítulo 3. As Figuras 4.5 (a), (b) e (c) correspondem, respectivamente, aos gráficos dos elementos $|\mathbf{d}_{max}|$ versus observações (ou unidades experimentais) quando perturbamos a matriz de covariâncias do erro, a variável resposta e a matriz de covariâncias dos efeitos aleatórios. Já as Figuras 4.5 (d) e (e) correspondem respectivamente, ao gráfico de C_i , definido em (3.69),



Figura 4.4 Alavancagem generalizada.

e o gráfico dos elementos de $|\boldsymbol{d}_{max}|$ versus unidades experimentais quando consideramos a perturbação proposta por Lesaffre & Verbeke (1998) (subseção 3.8.4). As Figuras 4.5 (f), (g) e (h) representam, respectivamente, os gráficos dos termos interpretáveis $||\boldsymbol{\mathcal{X}}_i \boldsymbol{\mathcal{X}}_i^{\top}||^2$, $||\boldsymbol{\mathcal{R}}_i||^2$ e $||\boldsymbol{I}_{n_i} - \boldsymbol{\mathcal{R}}_i \boldsymbol{\mathcal{R}}_i^{\top}||^2$. Uma vez que no modelo ajustado, a matriz de covariâncias intraunidades experimentais e a matriz de planejamento dos efeitos aleatórios são iguais para todas as unidades experimentais, então os gráficos de $||\boldsymbol{\widehat{V}}_i^{-1}||^2$ e $||\boldsymbol{\mathcal{Z}}_i \boldsymbol{\mathcal{Z}}_i^{\top}||^2$ não fornecem informações a respeito da influência das unidades experimentais.



Figura 4.5 Influência local.

Pela Figura 4.5 conclui-se que as unidades experimentais #12 e #29 são as mais influentes no modelo ajustado, principalmente no que tange às estimativas dos parâmetros de covariância. As observações mais sensíveis a pequenas perturbações na variável resposta e na matriz de covariâncias do erro condicional são #12.2 e #29.4 que correspondem às observações que apresentam, em módulo, o maior erro condicional, concordando com o resultado obtido por Schwarzmann (1991) no caso linear normal. A unidade experimental

#29 é a mais influente à suposição de homogeneidade entre as matrizes de covariância dos efeitos aleatórios; como neste exemplo, $Var[\delta_i] = \tau^2$, essa unidade experimental deve apresentar uma alta influência na estimação de τ^2 . Na Figura 4.6 apresentamos os valores da distância de Cook condicional por observação e seus respectivos valores decompostos. Na Figura 4.7 mostram-se os valores da distância de Cook condicional, e sua respectiva decomposição, referente às unidades experimentais, conforme sugerido no Capítulo 3.







Figura 4.7 Distância de Cook condicional por unidade experimental.

Nas duas abordagens utilizadas (influência local e eliminação de observações) destacaramse como influentes as unidades experimentais #12 e #29 principalmente com relação às estimativas dos parâmetros de covariância, uma vez que o segundo termo da decomposição da medida de Cook condicional é que dá a maior contribuição. A seguir descrevemos suas características, visando entender os motivos dessa alta influência.

- # 11: Essa criança utilizou a escova convencional e apresentou o menor índice de placa bacteriana pré-escovação (0.60) e o terceiro menor valor do índice de placa bacteriana pós-escovação (0.47), ambas na segunda sessão;
- # 12: Essa criança utilizou a escova convencional e apresentou o segundo menor índice de placa bacteriana pré-escovação (0.71) e pós-escovação (0.39), ambas na segunda sessão; apresenta também um alto índice de placa bacteriana pós-escovação, entre as 25% contradizendo o modelo ajustado, que prever maiores índices de placa bacteriana pós-escovação para crianças que utilizam a escova monobloco; essa criança também apresenta a maior variabilidade entre os índices de placa bacteriana pós-escovação e a segunda maior variabilidade entre os valores de índice de placa bacteriana pré-escovação para as quatro sessões de avaliação além de apresentar o segundo menor índice de redução de placa bacteriana (y/x);

• # 29: Essa criança apesar de ter utilizado a escova monobloco, apresentou todos seus índices de placa bacteriana pós-escovação entre os 25% menores índices, inclusive o menor (0.37) obtido na quarta sessão; este resultado contraria o esperado sob o modelo ajustado, que prever menores índices para crianças que utilizaram a escova convencional; apresentou também dois entre os três menores, incluindo o menor, índices de redução de placa bacteriana (y/x).

A seguir, está apresentada a análise confirmatória, obtida reajustando o modelo sem as unidades experimentais $\#12 \ e \ \#29$ para avaliar o impacto nas estimativas dos parâmetros do modelo (4.10). Entre parênteses apresentamos o impacto percentual na estimativa do parâmetro, quando eliminamos a respectiva unidade experimental.

Tabela 4.2 Estimativas dos parâmetros do modelo (4.10) ao eliminar as unidades experimentais $\#12 \ e \ \#29.$

Parâmetros	λ_0	λ_1	δ	$ au^2$	σ^2
Modelo Completo	$0.72{\pm}0.02$	$0.81{\pm}0.02$	$1.06 {\pm} 0.06$	$0.006 {\pm} 0.003$	$0.021 {\pm} 0.03$
Excluindo $\#12$	$0.72{\pm}0.02$	$0.80{\pm}0.02$	$1.06{\pm}0.06$	$0.007 {\pm} 0.003$	$0.015{\pm}0.02$
	(0.00)	(-1.23)	(0.00)	(16.67)	(-28.57)
Excluindo $#29$	$0.72{\pm}0.02$	$0.83{\pm}0.02$	$1.07{\pm}0.05$	$0.001 {\pm} 0.001$	$0.017 {\pm} 0.02$
	(0.00)	(2.47)	(0.94)	(-83.33)	(-19.05)
Excluindo #12 e #29	$0.72{\pm}0.02$	$0.83{\pm}0.02$	$1.07{\pm}0.05$	$0.003 {\pm} 0.001$	$0.012{\pm}0.01$
	(0.00)	(2.47)	(0.94)	(-50.00)	(-42.86)

Pela Tabela 4.2 percebe-se que as unidades experimentais #12 e #29 exercem uma alta influência nas estimativas dos parâmetros de covariância, por outro lado não se detecta nenhum tipo de influência na estimativa dos parâmetros fixos, conforme foi indicado pelas medidas de diagnóstico anteriormente utilizadas. A influência esperada na retirada de uma unidade experimental é de $(1/32)\times100=\pm 3,13\%$. Quando eliminamos a unidade experimental #29 a estimativa de τ^2 decresce 83.33\%, implicando uma alta influência da respectiva unidade experimental na estimativa da variância do efeito aleatório da criança, conforme esperávamos segundo a Figura 4.5 (c).

No exemplo considerado, quando se excluem as unidades experimentais influentes, toda inferência realizada com base na amostra completa, continua válida, indicando que o modelo ajustado é **robusto**.

Comentários

5.1 Recursos computacionais

A difusão do uso das técnicas de diagnóstico está intimamente relacionada com a facilidade em que elas são implementadas computacionalmente. Por exemplo, para se obter os diferentes tipos de resíduos definidos no Capítulo 2 pode-se utilizar o PROC MIXED do SAS ou a biblioteca NLME (http://nlme.stat.wisc.edu) desenvolvida em linguagem S-Plus. Pinheiro & Bates (2000) indicam, através de exemplos práticos, como utilizar a referida biblioteca para ajustar modelos lineares (não-lineares) mistos e como extrair os diferentes tipos de resíduos. Com relação ao caso ponderado (influência local), Verbeke & Molenberghs (1997) disponibilizam uma macro desenvolvida no SAS para calcular C_i , $C_i(\boldsymbol{\beta}), C_i(\boldsymbol{\theta}, \sigma^2)$ e as respectivas quantidades intepretáveis (3.81) para cada unidade experimental. Tal macro está disponível na página: www.springer-ny.com.

Uma sub-rotina em linguagem R para calcular o resíduo com confudimento mínimo, distância condicional de Cook e respectiva decomposição, matrix de lavancagem generalizada e alguns gráficos referentes a influência local está sendo desenvolvida. A sub-rotina poderá ser obtida brevemente nas páginas **www.ime.usp.br**/~**juvencio** e **www.ime.usp.br**/~**jmsinger**. Os programas e o banco de dados utilizados nesta dissertação podem ser solicitados via e-mail: **juvencio@ime.usp.br**.

5.2 Pesquisas futuras

Neste trabalho, apresentamos, discutimos e propomos algumas técnicas de diagnóstico para modelos lineares mistos. Para pesquisas futuras, ainda existem tópicos a serem explorados, como:

• Propor um gráfico da variável adicionada para efeitos aleatórios.

- Utilizar o EBLUP com confundimento mínimo como ferramenta para avaliar a suposição de normalidade dos efeitos aleatórios.
- Estender as técnicas de diagnóstico aqui apresentadas para os modelos lineares mistos sem se restringir ao modelo de independência condicional, modelos não-lineares mistos e para os modelos lineares generalizados mistos.
- Estudar a sensibilidade das medidas de diagnóstico apresentadas, com relação à má especificação das matrizes $R \in D$.

Apêndice \mathbf{A}

Expressões do Capítulo 1

A.1 Identidades (1.9), (1.19), (1.20) e (1.23)

A.1.1 Identidade (1.9)

Pré-multiplicando (1.8)

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\eta} \end{bmatrix}, \quad (A.1)$$

por $\mathbf{R}^{-1/2} \bigoplus \mathbf{D}^{-1/2} = \begin{bmatrix} \mathbf{R}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1/2} \end{bmatrix}$, com \bigoplus representando a soma direta [Searle *et al.* (1992) e Magnus & Neudecker (1988)], obtém-se:

em que,

$$oldsymbol{Y}^* = \left[egin{array}{c} oldsymbol{R}^{-1/2}oldsymbol{Y} & oldsymbol{R}^{-1/2}oldsymbol{Z} & oldsymbol{R}^{-1/2}oldsymbol{Z} & oldsymbol{R}^* = (oldsymbol{eta},oldsymbol{\gamma})^{ op}, ext{ com} \ \mathbf{Var}[oldsymbol{\zeta}] = \left[egin{array}{c} oldsymbol{R}^{-1/2} & \mathbf{0} \ \mathbf{0} & oldsymbol{D}^{-1/2} \end{array}
ight] \sigma^2 \left[egin{array}{c} oldsymbol{R} & \mathbf{0} \ \mathbf{0} & oldsymbol{D} \end{array}
ight] \left[egin{array}{c} oldsymbol{R}^{-1/2} oldsymbol{Z} & oldsymbol{\Omega} \\ \mathbf{0} & oldsymbol{D}^{-1/2} \end{array}
ight] \sigma^2 \left[egin{array}{c} oldsymbol{R} & \mathbf{0} \ \mathbf{0} & oldsymbol{D} \end{array}
ight] \left[egin{array}{c} oldsymbol{R}^{-1/2} oldsymbol{Z} \\ \mathbf{0} & oldsymbol{D}^{-1/2} \end{array}
ight] = \sigma^2 oldsymbol{I}_{cq+n} \end{array}$$
A.1.2 Identidade (1.19)

Considerando que
$$\mathbf{R}^{-1} - \mathbf{M} = \mathbf{R}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^{\top} \mathbf{R}^{-1}$$
, (1.14), (1.16) e (1.18), então
 $\mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{X} \widehat{\boldsymbol{\beta}} + \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Z} \widehat{\boldsymbol{\gamma}} = \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{X} \widehat{\boldsymbol{\beta}} + \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^{\top} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}})$
 $= \mathbf{X}^{\top} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{M}^{-1} \mathbf{Q} \mathbf{Y})$
 $+ \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{M}^{-1} \mathbf{Q} \mathbf{Y}$
 $= \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{M}^{-1} (\mathbf{M} - \mathbf{Q}) \mathbf{Y}$
 $+ \mathbf{X}^{\top} (\mathbf{R}^{-1} - \mathbf{M}) \mathbf{M}^{-1} \mathbf{Q} \mathbf{Y}$
 $= \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{M}^{-1} (\mathbf{M} - \mathbf{Q} + \mathbf{Q}) \mathbf{Y} - \mathbf{X}^{\top} \mathbf{Q} \mathbf{Y}$
 $= \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{M}^{-1} (\mathbf{M} - \mathbf{Q} + \mathbf{Q}) \mathbf{Y} - \mathbf{X}^{\top} \mathbf{Q} \mathbf{Y}$

A.1.3 Identidade (1.20)

$$DZ^{\top}M = DZ^{\top} \left(R^{-1} - R^{-1}ZC^{-1}Z^{\top}R^{-1} \right)$$

= $D\left(C - Z^{\top}R^{-1}Z \right)C^{-1}Z^{\top}R^{-1}$
= $D\left(D^{-1} \right)C^{-1}Z^{\top}R^{-1}$ (A.2)
= $C^{-1}Z^{\top}R^{-1}$.

Note que
$$(A.2)$$
 segue de (1.12) .

A.1.4 Identidade (1.23)

Lembrando que

$$\mathbb{E}[\mathbf{Y}^{\top}\mathbf{Q}\mathbf{Y}] = \mathbb{E}\left[\operatorname{tr}(\mathbf{Y}^{\top}\mathbf{Q}\mathbf{Y})\right] \\
 = \mathbb{E}\left[\operatorname{tr}(\mathbf{Q}\mathbf{Y}\mathbf{Y}^{\top})\right] = \operatorname{tr}(\mathbb{E}\left[\mathbf{Q}\mathbf{Y}\mathbf{Y}^{\top}\right]). \quad (A.3)$$

Levando em consideração (1.2) e (1.16),

$$\begin{bmatrix} \boldsymbol{Q}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}} \end{bmatrix} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{E} \begin{bmatrix} \boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}} \end{bmatrix} \boldsymbol{X}^{\mathsf{T}} \\
 + \boldsymbol{Q} \left(\boldsymbol{Z}\boldsymbol{E} \begin{bmatrix} \boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathsf{T}} \end{bmatrix} \boldsymbol{Z}^{\mathsf{T}} + \boldsymbol{E} \begin{bmatrix} \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}} \end{bmatrix} \right) \\
 = \sigma^{2}\boldsymbol{Q} \left(\boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^{\mathsf{T}} + \boldsymbol{R} \right) = \sigma^{2}\boldsymbol{Q}\boldsymbol{M}^{-1} \\
 = \sigma^{2} \left[\boldsymbol{I}_{n} - \boldsymbol{M}\boldsymbol{X} \left(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{M}\boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\mathsf{T}} \right]$$
(A.4)

Nobre, Juvêncio S.

IME-USP

-

Substituindo (A.4) em (A.3), temos

$$\begin{split} E[\mathbf{Y}^{\top} \mathbf{Q} \mathbf{Y}] &= \sigma^{2} \mathrm{tr} \left[\mathbf{I}_{n} - \mathbf{M} \mathbf{X} \left(\mathbf{X}^{\top} \mathbf{M} \mathbf{X} \right)^{-1} \mathbf{X}^{\top} \right] \\ &= \sigma^{2} \left(\mathrm{tr} [\mathbf{I}_{n}] - \left[\mathrm{tr} \left[\mathbf{M} \mathbf{X} \left(\mathbf{X}^{\top} \mathbf{M} \mathbf{X} \right)^{-1} \mathbf{X}^{\top} \right] \right] \right) \\ &= \sigma^{2} \left(n - \left[\mathrm{tr} \left[\left(\mathbf{X}^{\top} \mathbf{M} \mathbf{X} \right)^{-1} \mathbf{X}^{\top} \mathbf{M} \mathbf{X} \right] \right] \right) \\ &= \sigma^{2} \left(n - \mathrm{tr} (\mathbf{I}_{p}) \right) = \sigma^{2} (n - p). \end{split}$$

A.2 BLUE e BLUP

Por (1.10), tem-se:

$$\boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z}\widehat{\boldsymbol{\gamma}} = \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Y}$$
(A.5)

$$\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + (\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1})\widehat{\boldsymbol{\gamma}} = \boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Y}.$$
 (A.6)

Por (A.6),

$$\widehat{\boldsymbol{\gamma}} = (\boldsymbol{Z}^{\top} \boldsymbol{R}^{-1} \boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}^{\top} \boldsymbol{R}^{-1} (\boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}).$$
(A.7)

Substituindo (A.7) em (A.6),

$$\boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z}(\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1}\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{X}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Y}.$$

Reescrevendo à equação acima e considerando (1.11) temos

$$\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{Y}. \tag{A.8}$$

Uma vez que X é posto coluna completo e M é positiva definida, implicando que $X^{\top}MX$ é não singular; por conseguinte, o BLUE de β é dado por

$$\widehat{oldsymbol{eta}} = \left(oldsymbol{X}^ op oldsymbol{M} oldsymbol{X}
ight)^{-1} oldsymbol{X}^ op oldsymbol{M} oldsymbol{Y} = \left(oldsymbol{X}^ op oldsymbol{V}^{-1} oldsymbol{X}
ight)^{-1} oldsymbol{X}^ op oldsymbol{V}^{-1} oldsymbol{Y}.$$

Substituindo (1.13) em (A.7)

$$\begin{split} \widehat{\boldsymbol{\gamma}} &= (\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1})^{-1}\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \boldsymbol{C}^{-1}\boldsymbol{Z}^{\top}\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \end{split}$$

A.3 Propriedades do BLUE e BLUP

Propriedades de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\gamma}}$, são dadas em Henderson (1975), McLean *et al.* (1991), Robinson (1991), Searle *et al.* (1992), McCulloch & Searle (2001), etc. Dentre elas convém destacar:

- 1. Na classe dos estimadores lineares, o BLUP $\hat{\gamma}$ maximiza a correlação entre γ^{\top} e qualquer outro preditor γ^* , cujo valor máximo é $\rho_{(\gamma^{\top},\hat{\gamma})}$;
- 2. Se $\boldsymbol{K}^{\top}\boldsymbol{\beta}$ é estimável sendo \boldsymbol{K} um vetor $n \times 1$ conhecido, então o BLUE de $\boldsymbol{K}^{\top}\boldsymbol{\beta}$ é $\boldsymbol{K}^{\top}\boldsymbol{\hat{\beta}}$;
- 3. $I\!\!E[\widehat{\gamma}/\gamma] = \gamma;$
- 4. $\operatorname{Var}[\widehat{\boldsymbol{\beta}}] = \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X} \right)^{-1};$
- 5. $\operatorname{Var}[\widehat{\gamma} \gamma] = \sigma^2 \left[\boldsymbol{D} \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \right];$
- 6. $\operatorname{Cov}[\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}] = \mathbf{0};$
- 7. $\operatorname{Cov}[\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}} \boldsymbol{\gamma}] = -\operatorname{Cov}[\widehat{\boldsymbol{\beta}}, \boldsymbol{\gamma}] = -\sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{Z} \boldsymbol{D}.$

Expressões do Capítulo 2

B.1 Identidades (2.5), (2.6) e (2.7)

Sob a validade do modelo (1.2) e considerando a propriedade (1.16), temos

$$\begin{split} \boldsymbol{\xi} &= \boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}} &= (\boldsymbol{I}_n - \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}) \boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}) (\boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\xi}) \\ &= (\boldsymbol{I}_n - \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}) \boldsymbol{\xi} \\ &= \boldsymbol{\xi} - \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{\xi}, \end{split}$$

$$egin{array}{rll} Z \widehat{m{\gamma}} = Z D Z^ op Q Y &=& Z D Z^ op Q (Xeta + Zm{\gamma} + arepsilon) \ &=& Z D Z^ op Q Z m{\gamma} + Z D Z^ op Q arepsilon \end{array}$$

е

$$egin{array}{rcl} \hat{arepsilon} = RQY &= RQ(Xeta + Zeta + arepsilon) \ &= RQZeta + RQarepsilon. \end{array}$$

B.2 Identidades (2.16) e (2.17)

Considerando a decomposição (2.14) e (2.15), temos

$$egin{aligned} m{l}_i^ op \hat{m{arepsilon}} &= & \pi_i^{-1/2} m{K}_i^ op m{R}^{-1/2} \hat{m{arepsilon}} &= & \pi_i^{-1/2} m{K}_i^ op m{R}^{-1/2} m{R} m{Q} m{Y} \ &= & \pi_i^{-1/2} m{K}_i^ op m{K}^ op m{K}^ op m{K}^{-1/2} m{Y} \ &= & \sqrt{\pi_i} m{K}_i^ op m{R}^{-1/2} m{Y} \end{aligned}$$

 \mathbf{e}

$$\begin{aligned} \operatorname{Cov}[\boldsymbol{l}_i^{\mathsf{T}} \hat{\boldsymbol{\varepsilon}}, \boldsymbol{l}_j \hat{\boldsymbol{\varepsilon}}] &= \sigma^2 \boldsymbol{l}_i^{\mathsf{T}} \boldsymbol{R} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{l}_j \\ &= \sigma^2 \boldsymbol{l}_i^{\mathsf{T}} \boldsymbol{R}^{1/2} \boldsymbol{K} \boldsymbol{\Pi} \boldsymbol{K}^{\mathsf{T}} \boldsymbol{R}^{1/2} \boldsymbol{l}_j \\ &= \sigma^2 \pi_i^{-1/2} \boldsymbol{K}_i^{\mathsf{T}} \boldsymbol{R}^{-1/2} \boldsymbol{R}^{1/2} \boldsymbol{K} \boldsymbol{\Pi} \boldsymbol{K}^{\mathsf{T}} \pi_j^{-1/2} \boldsymbol{R}^{-1/2} \boldsymbol{K}_j \\ &= \frac{\sigma^2}{\sqrt{\pi_i \pi_j}} \boldsymbol{K}_i^{\mathsf{T}} \boldsymbol{K} \boldsymbol{\Pi} \boldsymbol{K}^{\mathsf{T}} \boldsymbol{K}_j \\ &= \sigma^2 \delta_{ij}. \end{aligned}$$

63

Apêndice \mathbf{C}

Expressões do Capítulo 3

C.1 BLUE (3.4)

Considerando $\mathbf{X} = [\mathbf{X}_1; \mathbf{X}_2]$, no modelo (3.2), tal que posto $(\mathbf{X}) = p = p_1 + p_2 = \text{posto}(\mathbf{X}_1) + \text{posto}(\mathbf{X}_2)$, tem-se [Searle *et al.* (1992, p.450)]

$$\left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1} = \begin{bmatrix} \boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1} & \boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{2} \\ \boldsymbol{X}_{2}^{\top}\boldsymbol{M}\boldsymbol{X}_{1} & \boldsymbol{X}_{2}^{\top}\boldsymbol{M}\boldsymbol{X}_{2} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^{\top} \\ \boldsymbol{B} & \left(\boldsymbol{X}_{2}^{\top}\boldsymbol{Q}_{1}\boldsymbol{X}_{2}\right)^{-1} \end{bmatrix}, \quad (C.1)$$

com

$$\boldsymbol{B} = -\left(\boldsymbol{X}_{2}^{\top}\boldsymbol{Q}_{1}\boldsymbol{X}_{2}\right)^{-1}\boldsymbol{X}_{2}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\left(\boldsymbol{X}_{1}^{\top}\boldsymbol{M}\boldsymbol{X}_{1}\right)^{-1}.$$
 (C.2)

Por (1.13) tem-se que

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = \left(\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^\top \\ \boldsymbol{B} & \left(\boldsymbol{X}_2^\top \boldsymbol{Q}_1 \boldsymbol{X}_2 \right)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \end{bmatrix} \boldsymbol{M} \boldsymbol{Y}$$

$$= \begin{bmatrix} \boldsymbol{A} \boldsymbol{X}_1^\top + \boldsymbol{B} \boldsymbol{X}_2^\top \\ \boldsymbol{B} \boldsymbol{X}_1^\top + \left(\boldsymbol{X}_2^\top \boldsymbol{Q}_1 \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2^\top \end{bmatrix} \boldsymbol{M} \boldsymbol{Y}.$$
(C.3)

Portanto, considerando (C.2) e (C.3) simultaneamente, obtemos

$$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2^{\top} \boldsymbol{Q}_1 \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2^{\top} \left(\boldsymbol{M} - \boldsymbol{M} \boldsymbol{X}_1 \left(\boldsymbol{X}_1^{\top} \boldsymbol{M} \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1^{\top} \boldsymbol{M} \right) \boldsymbol{Y}$$

$$= \left(\boldsymbol{X}_2^{\top} \boldsymbol{Q}_1 \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2^{\top} \boldsymbol{Q}_1 \boldsymbol{Y}.$$

C.2 Fórmula de atualização do BLUP (3.9)

Por (1.21) temos que $\hat{\gamma}$ é dado por:

$$\begin{aligned} \widehat{\boldsymbol{\gamma}} &= \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{M} \left(\boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}} \right) \\ &= \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{M} \left(\boldsymbol{Y} - \boldsymbol{X}_{1} \widehat{\boldsymbol{\beta}}_{1} - \boldsymbol{X}_{2} \widehat{\boldsymbol{\beta}}_{2} \right) \\ &= \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{M} \left(\boldsymbol{Y} - \boldsymbol{X}_{1} \left(\widehat{\boldsymbol{\beta}}_{1} - \widehat{\boldsymbol{\beta}}_{1}^{0} \right) - \boldsymbol{X}_{1} \widehat{\boldsymbol{\beta}}_{1}^{0} - \boldsymbol{X}_{2} \widehat{\boldsymbol{\beta}}_{2} \right), \end{aligned}$$
(C.4)

considerando (3.7) em (C.4),

$$egin{array}{rcl} \widehat{m{\gamma}} &=& m{D}m{Z}^{ op}m{M}\left(m{Y}-m{X}_1\widehat{m{eta}}_1^0
ight) \ &+& m{D}m{Z}^{ op}m{M}\left(m{X}_1\left(m{X}_1^{ op}m{M}m{X}_1
ight)^{-1}m{X}_1^{ op}m{M}m{X}_2\widehat{m{eta}}_2-m{X}_2\widehat{m{eta}}_2
ight) \ &=& \widehat{m{\gamma}}_0+m{D}m{Z}^{ op}m{M}\left(m{X}_1\left(m{X}_1^{ op}m{M}m{X}_1
ight)^{-1}m{X}_1^{ op}m{M}-m{I}_n
ight)m{X}_2\widehat{m{eta}}_2 \ &=& \widehat{m{\gamma}}_0-m{D}m{Z}^{ op}m{Q}_1m{X}_2\widehat{m{eta}}_2. \end{array}$$

Desta forma,

$$\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_0 = -\boldsymbol{D}\boldsymbol{Z}^\top \boldsymbol{Q}_1 \boldsymbol{X}_2 \widehat{\boldsymbol{\beta}}_2. \tag{C.5}$$

C.3 Identidades (3.10) e (3.11)

Por (1.11), (1.14) e (1.18) temos

$$QY = M(Y - X\widehat{\beta}) = (R^{-1} - R^{-1}ZC^{-1}Z^{\top}R^{-1})(Y - X\widehat{\beta})$$

= $R^{-1}(Y - X\widehat{\beta}) - R^{-1}ZC^{-1}Z^{\top}R^{-1}(Y - X\widehat{\beta})$
= $R^{-1}(Y - X\widehat{\beta}) - R^{-1}Z\widehat{\gamma} = R^{-1}(Y - X\widehat{\beta} - Z\widehat{\gamma}).$ (C.6)

Analogamente, tem-se que

$$\boldsymbol{M}^{-1/2}\boldsymbol{Q}_{1}\boldsymbol{Y} = \boldsymbol{M}^{1/2}(\boldsymbol{Y} - \boldsymbol{X}_{1}\widehat{\boldsymbol{\beta}}_{1}) = \boldsymbol{M}^{-1/2}\boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{1}\widehat{\boldsymbol{\beta}}_{1}^{0} - \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{0})$$
(C.7)

е

$$M^{-1/2}Q_1X_2 = M^{-1/2}(X_2 - X_1\hat{\beta}_1^*) = M^{-1/2}R^{-1}(X_2 - X_1\hat{\beta}_1^* - Z\hat{\gamma}^*).$$
 (C.8)

C.4 BLUE e BLUP do modelo (3.22)

Fung *et al.* (2002) [Teorema 2, quando desconsideramos a função não-paramétrica f e o processo estocástico U(t)] demonstraram este resultado quando o conjunto I tem uma única observação. Hilden-Minton (1995) apresenta a seguinte prova heurística.

Nobre, Juvêncio S.

Considere que $[\mathbf{X}: \mathbf{U}_I]$ tem posto completo no modelo (3.22). Pertubando o vetor de variável resposta do conjunto I da seguinte forma

$$\boldsymbol{Y}(\boldsymbol{w}) = \boldsymbol{Y} + \boldsymbol{U}_{I}\boldsymbol{w},$$

obtemos o seguinte modelo

$$\boldsymbol{Y}(\boldsymbol{w}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{U}_{I}\boldsymbol{\phi}_{I}^{*} + \boldsymbol{\varepsilon}, \quad \forall \boldsymbol{w},$$
(C.9)

em que $\boldsymbol{\beta} \in \boldsymbol{\gamma}$, bem como seus respectivos BLUE e BLUP, são idênticos a do modelo (3.22), pois reparametrizando $\boldsymbol{\phi}_I^* = \boldsymbol{\phi}_I + \boldsymbol{w}$ em (C.9) obtém-se (3.22). Deste modo, o BLUE de $\boldsymbol{\beta}$ e o BLUP de $\boldsymbol{\gamma}$ do modelo (3.22) independem das observações $(y_i)_{i \in I}$, ou seja, podemos interpretar as estimativas de $\boldsymbol{\beta} \in \boldsymbol{\gamma}$ do modelo (3.22) como sendo as respectivas estimativas do (1.2) quando eliminamos as observações do conjunto I.

C.5 Identidades (3.26), (3.27), (3.28), (3.29) e (3.30)

Por (3.23) tem-se diretamente que

$$\operatorname{Var}[\widehat{\boldsymbol{\phi}}_{I}] = \left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}$$
$$= \sigma^{2}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{M}^{-1}\boldsymbol{Q}\boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}$$
$$= \sigma^{2}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}$$
$$= \sigma^{2}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}, \qquad (C.10)$$

que prova (3.26).

De forma semelhante, para provar (3.27), basta considerar (3.24) e (C.10) pois

$$\begin{aligned} \operatorname{Var}[\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(I)}] &= (\boldsymbol{X} \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{U}_{I} \operatorname{Var}[\widehat{\boldsymbol{\phi}}_{I}] \boldsymbol{U}_{I}^{\top} \boldsymbol{M} \boldsymbol{X} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right) \\ &= \sigma^{2} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{M} \boldsymbol{X} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1}. \end{aligned}$$

Analogamente,

$$\operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)}] = \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \operatorname{Var}[\widehat{\boldsymbol{\phi}}_{I}] \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D}$$
$$= \sigma^{2} \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D}. \quad (C.11)$$

Por (1.13), (3.23) e (3.24) temos

$$\widehat{\boldsymbol{\beta}}_{(I)} = \left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{M}\left(\boldsymbol{I}_{n} - \boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}\boldsymbol{Q}\right)\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{Y},$$

com $\boldsymbol{W} = \left(\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}\boldsymbol{M}\left(\boldsymbol{I}_{n} - \boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}\boldsymbol{Q}\right).$ Portanto,
$$\operatorname{Var}[\widehat{\boldsymbol{\beta}}_{(I)}] = \boldsymbol{W}\boldsymbol{V}\boldsymbol{W}^{\top}.$$
 (C.12)

Considerando $V = \sigma^2 M$ e as propriedades (1.15) e (1.16), temos por (C.12) que

$$\operatorname{Var}[\widehat{\boldsymbol{\beta}}_{(I)}] = \sigma^{2} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} + \sigma^{2} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{M} \boldsymbol{X} \left(\boldsymbol{X}^{\top} \boldsymbol{M} \boldsymbol{X} \right)^{-1} = \operatorname{Var}[\widehat{\boldsymbol{\beta}}] + \operatorname{Var}[\widehat{\boldsymbol{\beta}}_{(I)} - \widehat{\boldsymbol{\beta}}].$$
(C.13)

Por (3.25) tem-se que

$$\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)} = \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{Y}, \qquad (C.14)$$

por outro lado, considerando (1.21)

$$\widehat{\gamma} - \gamma = DZ^{\top}QY - \gamma.$$
 (C.15)

Como $\widehat{\gamma}_{(I)} - \gamma = \widehat{\gamma}_{(I)} - \widehat{\gamma} + (\widehat{\gamma} - \gamma)$, temos $\operatorname{Var}[\widehat{\gamma}_{(I)} - \gamma] = \operatorname{Var}[\widehat{\gamma}_{(I)} - \widehat{\gamma}] + \operatorname{Var}[\widehat{\gamma} - \gamma]$ $+ \operatorname{Cov}[\widehat{\gamma}_{(I)} - \widehat{\gamma}, \widehat{\gamma} - \gamma] + \operatorname{Cov}[\widehat{\gamma}_{(I)} - \widehat{\gamma}, \widehat{\gamma} - \gamma]^{\top}. \quad (C.16)$

Mas, por (C.14) e (C.15)

$$\operatorname{Cov}[\widehat{\boldsymbol{\gamma}}_{(I)} - \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}] = \operatorname{Cov}[-\boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I}\left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}_{I}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\boldsymbol{Y}, \boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{Q}\boldsymbol{Y} - \boldsymbol{\gamma}], (C.17)$$

usando o fato de que o operador Cov[,] é bilinear tem-se que (C.17) fica dado por

$$\operatorname{Cov}[\widehat{\boldsymbol{\gamma}}_{(I)} - \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}] = -\boldsymbol{D}\boldsymbol{Z}^{\top}\boldsymbol{Q}\boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}_{I}\boldsymbol{U}_{I}\right)^{-1}\boldsymbol{U}_{I}^{\top}\boldsymbol{Q}\{\boldsymbol{V}(\boldsymbol{Q}\boldsymbol{Z}\boldsymbol{D}) - \operatorname{Cov}[\boldsymbol{Y},\boldsymbol{\gamma}]\}.$$
(C.18)

Considerando as suposições referentes ao modelo (1.2) tem-se que $\text{Cov}[\boldsymbol{Y}, \gamma] = \sigma^2 \boldsymbol{Z} \boldsymbol{D}$. Logo, lembrando (1.15), (C.18) pode ser escrita como

$$\operatorname{Cov}[\widehat{\gamma}_{(I)} - \widehat{\gamma}, \widehat{\gamma} - \gamma] = -\sigma^{2} \{ \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} (\boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{U}_{I})^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{M}^{-1} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} - \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} (\boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{U}_{I})^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \} = -\sigma^{2} \{ \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} (\boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{U}_{I})^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} - \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} (\boldsymbol{U}_{I}^{\top} \boldsymbol{Q}_{I} \boldsymbol{U}_{I})^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \} = \mathbf{0}.$$
(C.19)

Nobre, Juvêncio S.

Usando (C.19), a propriedade 5 do BLUP (veja apêndice A.3) e (C.12), temos

$$\operatorname{Var}[\widehat{\boldsymbol{\gamma}}_{(I)} - \boldsymbol{\gamma}] = \operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}] + \operatorname{Var}[\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{(I)}]$$
$$= \sigma^{2} \{ \boldsymbol{D} - \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} + \boldsymbol{D} \boldsymbol{Z}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \left(\boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{U}_{I} \right)^{-1} \boldsymbol{U}_{I}^{\top} \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{D} \} (C.20)$$

C.6 Decomposição da medida de Cook condicional (3.37)

Considerando $\boldsymbol{P}_{j(i)}$ definido em (3.36) tem-se que

$$\boldsymbol{P}_{j(i)}^{\top}\boldsymbol{P}_{j(i)} = (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{i})^{\top}\boldsymbol{X}_{j}^{\top}\boldsymbol{X}_{j}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{i}) + (\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_{j(i)})^{\top}\boldsymbol{Z}_{j}^{\top}\boldsymbol{Z}_{j}(\widehat{\boldsymbol{\gamma}}_{j} - \widehat{\boldsymbol{\gamma}}_{j(i)}) + 2(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{i})^{\top}\boldsymbol{X}_{j}^{\top}\boldsymbol{Z}_{j}(\widehat{\boldsymbol{\gamma}}_{j} - \widehat{\boldsymbol{\gamma}}_{j(i)}). \quad (C.21)$$

Somando (C.21) para j = 1, 2, ...c, obtêm-se (3.37).

C.7 Identidade (3.43)

Considerando o item (2) da Proposição (2) de Christensen et al. (1992), tem-se que

$$(\boldsymbol{X}_{(i)}^{\top}\boldsymbol{V}_{(i)}^{-1}\boldsymbol{X}_{(i)})^{-1} = (\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \left[\boldsymbol{I}_{p} + \widetilde{\boldsymbol{x}}_{i}\widetilde{\boldsymbol{x}}_{i}^{\top}(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}/(s_{i}-\widetilde{h}_{i}) \right], \quad (C.22)$$

com as respectivas matrizes e vetores definidos na seção 3.5. Desta forma,

$$t_{i}^{*} = \left| \sigma^{-2} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X}) \sigma^{2} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} (\boldsymbol{I}_{p} + \widetilde{\boldsymbol{x}}_{i} \widetilde{\boldsymbol{x}}_{i}^{\top} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} / (s_{i} - \widetilde{h}_{i})) - p \right|$$

$$= \left| \operatorname{tr} \{ \boldsymbol{I}_{p} \} \operatorname{tr} \{ \widetilde{\boldsymbol{x}}_{i} \widetilde{\boldsymbol{x}}_{i}^{\top} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \} / (s_{i} - \widetilde{h}_{i})) p \right|$$

$$= \left| \operatorname{tr} \{ \widetilde{\boldsymbol{x}}_{i}^{\top} (\boldsymbol{X}^{\top} \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \widetilde{\boldsymbol{x}}_{i} \} / (s_{i} - \widetilde{h}_{i}) \right|$$

$$= \left| \frac{\widetilde{h}_{i}}{s_{i} - \widetilde{h}_{i}} \right| = \left| \frac{h_{i}^{*}}{1 - h_{i}^{*}} \right|. \quad (C.23)$$

C.8 Derivadas (3.55), (3.57) e (3.57)

Seja \mathbf{A} uma matriz $n \times n$ simétrica positiva definida e t um escalar, então [Searle et al. (1992)]

$$\frac{\partial \mathbf{A}^{-1}}{\partial t} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial t} \mathbf{A}^{-1}$$
(C.24)

Nobre, Juvêncio S.

IME-USP

е

$$\frac{\partial \ln |\mathbf{A}|}{\partial t} = \operatorname{tr} \left\{ \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial t} \right\}, \qquad (C.25)$$

Derivando a log-verossimilhança perturbada (3.60) com respeito a β obtemos

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^{\top} \boldsymbol{V}(\boldsymbol{w})^{-1} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta}), \qquad (C.26)$$

derivando (C.26) com respeito a \boldsymbol{w}_k e calculando em $\boldsymbol{w}=\boldsymbol{w}_0$ e $\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}$ obtém-se

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial \boldsymbol{w}_k \partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}} = \boldsymbol{X}^\top \boldsymbol{D}_k \hat{\boldsymbol{r}}.$$
(C.27)

Usando os resultados (C.24) e (C.25) tem-se que $(\forall i = 1, ..., l)$

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \theta_{i}} = -\frac{1}{2} \operatorname{tr} \left\{ \boldsymbol{V}(\boldsymbol{w})^{-1} \boldsymbol{Z} \frac{\partial \boldsymbol{D}}{\partial \theta_{i}} \boldsymbol{Z}^{\top} \right\} + \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta})^{\top} \boldsymbol{V}(\boldsymbol{w})^{-1} \boldsymbol{Z} \frac{\partial \boldsymbol{D}}{\partial \theta_{i}} \boldsymbol{Z}^{\top} \boldsymbol{V}(\boldsymbol{w})^{-1} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta})$$
(C.28)

е

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \sigma^2} = -\frac{1}{2} \operatorname{tr} \{ \boldsymbol{V}(\boldsymbol{w})^{-1} \boldsymbol{\Lambda}(\boldsymbol{w}) \} + \frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta})^\top \boldsymbol{V}(\boldsymbol{w})^{-1} \boldsymbol{\Lambda}(\boldsymbol{w}) \boldsymbol{V}(\boldsymbol{w})^{-1} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta}).$$
(C.29)

Derivando (C.28) com relação a w_k temos

$$\frac{\partial^{2}\lambda(\boldsymbol{w})}{\partial w_{k}\partial\theta_{i}} = -\frac{1}{2}\mathrm{tr}\left\{\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_{k}}\boldsymbol{Z}\frac{\partial \boldsymbol{D}}{\partial\theta_{i}}\boldsymbol{Z}^{\top}\right\}$$

$$+ (1/2)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top}\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_{k}}\boldsymbol{Z}\frac{\partial \boldsymbol{D}}{\partial\theta_{i}}\boldsymbol{Z}^{\top}\boldsymbol{V}(\boldsymbol{w})^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$+ (1/2)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top}\boldsymbol{V}(\boldsymbol{w})^{-1}\boldsymbol{Z}\frac{\partial \boldsymbol{D}}{\partial\theta_{i}}\boldsymbol{Z}^{\top}\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_{k}}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$= \frac{1}{2}\mathrm{tr}\left\{\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_{k}}\boldsymbol{Z}\frac{\partial \boldsymbol{D}}{\partial\theta_{i}}\boldsymbol{Z}^{\top}\right\}$$

$$+ (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top}\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_{k}}\boldsymbol{Z}\frac{\partial \boldsymbol{D}}{\partial\theta_{i}}\boldsymbol{Z}^{\top}\boldsymbol{V}(\boldsymbol{w})^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}). \quad (C.30)$$

Calculando (C.30) em ${\boldsymbol w}={\boldsymbol w}_0$
e ${\boldsymbol \psi}=\widehat{\boldsymbol \psi}$ obtêm-se $(\forall i=1,...,l)$

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \theta_i} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \boldsymbol{\hat{\psi}}} = -\frac{1}{2} \left\{ \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^\top \right] - 2 \hat{\boldsymbol{r}}^\top \boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^\top \hat{\boldsymbol{r}} \right\}.$$
(C.31)

Nobre, Juvêncio S.

IME-USP

Uma vez que $\boldsymbol{V}(\boldsymbol{w}) = \sigma^2 \Lambda(\boldsymbol{w}) + \boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^{\top}$. Portanto,

$$\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1} \Lambda(\boldsymbol{w})}{\partial w_k} = -\sigma^{-2} \frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_k} \boldsymbol{Z} \boldsymbol{D} \boldsymbol{Z}^{\top}.$$
 (C.32)

Por outro lado temos também (considerando C.24 e C.32)

$$\frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1} \Lambda(\boldsymbol{w}) \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_k} = \frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_k} \Lambda(\boldsymbol{w}) \boldsymbol{V}(\boldsymbol{w})^{-1} + \boldsymbol{V}(\boldsymbol{w})^{-1} \Lambda(\boldsymbol{w}) \frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_k} + \boldsymbol{V}(\boldsymbol{w})^{-1} \frac{\partial \Lambda(\boldsymbol{w})}{\partial w_k} \boldsymbol{V}(\boldsymbol{w})^{-1}$$
(C.33)

Derivando (C.29) com relação a w_k , considerando (C.32) e (C.33) e calculando em $\boldsymbol{w} = \boldsymbol{w}_0 \in \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$, percebendo que $\Lambda(\boldsymbol{w}_0) = \boldsymbol{I}_n$, obtemos

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_k \partial \sigma^2} \Big|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}=\boldsymbol{\hat{\psi}}} = -\frac{1}{2} \left\{ \hat{\sigma}^{-2} \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \boldsymbol{\widehat{D}} \boldsymbol{Z}^{\top} \right] - 2 \hat{\boldsymbol{r}}^{\top} \boldsymbol{D}_k \boldsymbol{\widehat{V}}^{-1} \hat{\boldsymbol{r}} + \hat{\sigma}^{-2} \hat{\boldsymbol{r}}^{\top} \boldsymbol{D}_k \hat{\boldsymbol{r}} \right\}.$$
(C.34)

Por (C.27), (C.31) e (C.34) mostram-se os resultados (3.55), (3.56) e (3.57).

C.9 Identidade (3.58)

Por (C.24) temos

$$\boldsymbol{D}_k = \frac{\partial \boldsymbol{V}(\boldsymbol{w})^{-1}}{\partial w_k} = -\boldsymbol{V}^{-1} \frac{\partial \Lambda(\boldsymbol{w})}{\partial w_k} \boldsymbol{V}^{-1},$$

avaliada em $\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$. Notando que $\frac{\partial \Lambda(\boldsymbol{w})}{\partial w_k}$ calculada em $\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$ é igual a uma matriz $\boldsymbol{A} = (\delta_{ik} \delta_{kj})_{i,j}$, com $\delta_{ik} = 1$ se i = k e zero em caso contrário, o resultado segue.

C.10 Matriz (3.61)

Derivando (3.60) com relação ao vetor de parâmetros β tem-se

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^{\top} \boldsymbol{V}^{-1} (\boldsymbol{Y} + s \boldsymbol{w} - \boldsymbol{X} \boldsymbol{\beta}).$$
(C.35)

Derivando (C.35) com respeito a \boldsymbol{w}^{\top} e calculando essa derivada em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}$ obtemos

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial \boldsymbol{w}^\top \partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} = s \widehat{\boldsymbol{V}}^{-1} \boldsymbol{X}.$$
(C.36)

Considerando o resultado (C.25) tem-se que $(\forall i = 1, ..., l)$

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \theta_i} = \frac{1}{2} (\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} \boldsymbol{Z} \frac{\partial \boldsymbol{D}}{\partial \theta_i} \boldsymbol{Z}^\top \boldsymbol{V}^{-1} (\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta}) + k.$$
(C.37)

Derivando (C.37) com relação a \boldsymbol{w} e calculando essa derivada em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}$ obtemos

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial \boldsymbol{w} \partial \theta_i} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}} = s \widehat{\boldsymbol{V}}^{-1} \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^\top \boldsymbol{V}^{-1} \hat{\boldsymbol{r}}.$$
(C.38)

Analogamente ao caso anterior temos

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \sigma^2} = \frac{1}{2} (\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} \boldsymbol{V}^{-1} (\boldsymbol{Y} + s\boldsymbol{w} - \boldsymbol{X}\boldsymbol{\beta}) + k.$$
(C.39)

Derivando (C.39) com relação a \boldsymbol{w} e calculando essa derivada em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}$ obtém-se

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial \boldsymbol{w} \partial \sigma^2} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}} = s \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{V}}^{-1} \widehat{\boldsymbol{r}}.$$
(C.40)

Por (C.36), (C.38) e (C.40) segue o resultado (3.61).

C.11 Derivadas (3.64), (3.65) e (3.66)

Derivando a log-verossimilhança perturbada (3.62) com respeito a β obtemos

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{c} \boldsymbol{X}_{i}^{\top} \boldsymbol{V}_{i}(\boldsymbol{w})^{-1} \boldsymbol{r}_{i}.$$
 (C.41)

Derivando (C.41) com respeito a \boldsymbol{w}_k e calculando em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}$ obtém-se

$$\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial \boldsymbol{w}_k \partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{w} = \boldsymbol{w}_0; \boldsymbol{\psi} = \boldsymbol{\hat{\psi}}} = \boldsymbol{X}_k^\top \boldsymbol{\hat{V}}_k^{-1} \boldsymbol{Z}_k \boldsymbol{\hat{G}} \boldsymbol{Z}_k^\top \boldsymbol{\hat{V}}_k^{-1} \boldsymbol{\hat{r}}_k.$$
(C.42)

Usando os resultados (C.24) e (C.25) tem-se que $(\forall j=1,...,l)$

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \theta_{j}} = -\frac{w_{j}}{2} \sum_{i=1}^{c} \operatorname{tr} \left\{ \boldsymbol{V}_{i}(\boldsymbol{w})^{-1} \boldsymbol{Z}_{i} \frac{\partial \boldsymbol{G}}{\partial \theta_{j}} \boldsymbol{Z}_{i}^{\top} \right\}$$

+
$$\frac{1}{2} \sum_{i=1}^{c} \boldsymbol{r}_{i}^{\top} \boldsymbol{V}_{i}(\boldsymbol{w})^{-1} \boldsymbol{Z}_{i} \frac{\partial \boldsymbol{G}}{\partial \theta_{j}} \boldsymbol{Z}^{\top} \boldsymbol{V}_{i}(\boldsymbol{w})^{-1} \boldsymbol{r}_{i}$$
(C.43)

е

$$\frac{\partial \lambda(\boldsymbol{w})}{\partial \sigma^2} = -\frac{1}{2} \left\{ \sum_{i=1}^c \operatorname{tr} \left\{ \boldsymbol{V}_i(\boldsymbol{w})^{-1} \right\} - 2\boldsymbol{r}_i^\top \boldsymbol{V}_i(\boldsymbol{w})^{-1} \boldsymbol{V}_i(\boldsymbol{w})^{-1} \boldsymbol{r}_i \right\}.$$
(C.44)

Derivando (C.43) e (C.44) com relação a w_k e calculando em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}$ obtém-se (3.65) e (3.66).

C.12 Identidade (3.70)

Note que $C_i = -2d_i^{\top} H^{\top} \ddot{L}^{-1} H d_i$. Como $d_i = Pa$, com $P = \begin{bmatrix} v_1^{\top} \\ \vdots \\ v_c^{\top} \end{bmatrix}$ representando

uma matriz ortogonal e $\boldsymbol{a} = (\boldsymbol{v}_{1i}, \cdots, \boldsymbol{v}_{ci})$. Usando resultados matriciais [Morrison (1976)], temos

$$C_{i} = -2\boldsymbol{a}^{\top}\boldsymbol{P}^{\top}\boldsymbol{H}^{\top}\ddot{\boldsymbol{L}}^{-1}\boldsymbol{H}\boldsymbol{P}\boldsymbol{a}$$

$$= 2\boldsymbol{a}^{\top}\operatorname{diag}(\lambda_{1},\cdots,\lambda_{c})\boldsymbol{a}$$

$$= 2\sum_{j=1}^{c}\lambda_{j}\boldsymbol{v}_{ji}^{2}.$$
 (C.45)

C.13 Matriz Hessiana

Temos que

$$\frac{\partial^2 L(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^{\top}} = \begin{pmatrix} \frac{\partial^2 L(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}^{\top} \partial \boldsymbol{\beta}} & \frac{\partial^2 L(\boldsymbol{\psi})}{\partial (\boldsymbol{\theta}^*)^{\top} \boldsymbol{\beta}} \\ \frac{\partial^2 L(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}^{\top} \partial \boldsymbol{\theta}^*} & \frac{\partial^2 L(\boldsymbol{\psi})}{\partial (\boldsymbol{\theta}^*)^{\top} \partial \boldsymbol{\theta}^*} \end{pmatrix},$$

com a log-verossimilhança $L(\boldsymbol{\psi})$ dada por (3.52). Tem-se que (i,j=1,...,l+1)

$$\frac{\partial L(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^{\top} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta}), \qquad (C.46)$$

$$\frac{\partial L(\boldsymbol{\psi})}{\partial \theta_i^*} = -\frac{1}{2} \left\{ \operatorname{tr}(\boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_i) - (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_i \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \right\},$$
(C.47)

$$\frac{\partial^2 L(\boldsymbol{\psi})}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} = -\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X}, \qquad (C.48)$$

$$\frac{\partial^2 L(\boldsymbol{\psi})}{\partial \theta_i^* \partial \boldsymbol{\beta}} = -\boldsymbol{X}^t \boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_i \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X} \boldsymbol{\beta}), \qquad (C.49)$$

$$\frac{\partial^2 L(\boldsymbol{\psi})}{\partial \theta_j^* \partial \theta_i^*} = \frac{1}{2} [\operatorname{tr}(\boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_j \boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_i - \boldsymbol{V}^{-1} \ddot{\boldsymbol{V}}_{ij}) - 2(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_j \boldsymbol{V}^{-1} \dot{\boldsymbol{V}}_i \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

+
$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\top} \boldsymbol{V}^{-1} \ddot{\boldsymbol{V}}_{ij} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})],$$
 (C.50)

(C.51)

 $\begin{array}{l} \operatorname{com} \dot{\boldsymbol{V}}_{i} = \frac{\partial \boldsymbol{V}}{\partial \boldsymbol{\theta}_{i}^{*}} \; \mathrm{e} \; \ddot{\boldsymbol{V}}_{ij} = \frac{\partial^{2} \boldsymbol{V}}{\partial \boldsymbol{\theta}_{i}^{*} \partial \boldsymbol{\theta}_{j}^{*}}. \\ \mathrm{Para \; encontrar} \; \ddot{\boldsymbol{L}} \; \mathrm{basta \; calcular \; todas \; as \; derivadas \; acima \\ \mathrm{no \; EMV} \; \hat{\boldsymbol{\psi}} \; \mathrm{e \; em \; } \boldsymbol{w} = \boldsymbol{w}_{0}. \end{array}$

Considerando a reparametrização utilizada por Lesaffre & Verbeke (1998) para o caso em que \boldsymbol{G} é não estruturada, temos $l = \begin{pmatrix} q \\ 2 \end{pmatrix} + q$. Denotando os parâmetros da diagonal principal de \boldsymbol{G} por g_{kk} , todas as derivadas parciais de primeira ordem que envolvem g_{kk} , ou as derivadas parciais de segunda ordem que envolvem exatamente um elemento da diagonal de \boldsymbol{G} , devem ser multiplicadas por $\sqrt{2}$. Já as derivadas de segunda ordem da log-verossimilhança com respeito a g_{kk} e $g_{jj}(j, k = 1, ..., l)$ devem ser multiplicadas por 2.

Referências Bibliográficas

- Andreoni, S. (1989). Modelos de Efeitos Aleatórios para Análise de Dados Longitudinais não Balanceados em Relação ao Tempo. Dissertação de Mestrado. IME/USP, São Paulo.
- [2] Araujo, P.V. (1998). Geometria Diferencial. Rio de Janeiro: IMPA (Coleção Universitária).
- [3] Artes, R. (1997). Extensões da Teoria das Equações de Estimação Generalizadas a Dados Circulares e Modelos de Dispersão. Tese de Doutorado. IME/USP, São Paulo.
- [4] Atkinson, C.A. (1985). Plots, Transformations, and Regression: An Introduction to graphical methods of diagnostic regression analysis. Oxford: Oxford University Press.
- [5] Banerjee, M. (1998). Cook's Distance in Linear Longitudinal Models. Communications in Statistics, Theory and Methods 27, 2973-2983.
- [6] Banerjee, M. & Frees, E.W. (1997). Influence Diagnostics for Linear Longitudinal Models. Journal of the American Statistical Association 92, 999-1005.
- [7] Beckman, R.J., Nachtsheim, C.J. & Cook, R.D. (1987). Diagnostics for Mixed-Model Analysis of Variance. *Technometrics* 29, 413-426.
- [8] Beckman, R.J., Nachtsheim, C.J. & Cook, R.D. (1990). Correction. *Technometrics* 32, 241.
- Belsley, D.A. (1991). Conditioning Diagnostics: Collinearity and Weak Data in Regression. New York: John Wiley & Sons.
- [10] Belsley, D.A., Kuh, E. & Welsch, R.E. (1980). Regression Diagnostics: Identifying influential data and Sources of collinearity. New York: John Wiley & Sons.
- [11] Billor, N. & Loynes, R.M. (1993). Local Influence: A New Approach. Communications in Statistics, Theory and Methods 22, 1595-1611.

- Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika* 52, 345-370.
- [13] Breslow, N.E. (1984). Extra-Poisson Variation in Log-linear Models. Applied Statistics 33, 38-44.
- [14] Breslow, N.E. & Clayton, D.G. (1993). Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 48, 9-25.
- [15] Butler, S.M. & Louis, T.A. (1992). Random Effects Models with non-parametric priors. *Statistics in Medicine* 11, 1981-2000.
- [16] Chatterjee, S. & Hadi, A.S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression (with discussion). *Statistical Science* 1, 379-393.
- [17] Chatterjee, S. & Hadi, A.S. (1988). Sensitivity Analysis in Linear Regression. New York: John Wiley & Sons.
- [18] Christensen, R. & Pearson, L.M. (1992). Case-Deletion Diagnostics for Mixed Models. *Technometrics* 34, 38-45.
- [19] Christensen, R. (1996). Exact Tests for Variance Components. Biometrics 52, 309-314.
- [20] Cook, R.D. (1977). Detection of Influential Observation in Linear Regression. Technometrics 19, 15-18.
- [21] Cook, R.D. (1986). Assessment of Local Influence (with discussion). Journal of the Royal Statistical Society B 48, 133-169.
- [22] Cook, R.D. (1987). Influence Assessment. Journal of Applied Statistics 14, 117-131.
- [23] Cook, R.D. & Weisberg, S. (1980). Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression. *Technometrics* 22, 495-508.
- [24] Cook, R.D. & Weisberg, S. (1982). Residuals and Influence Regression. New York: Chapman & Hall.
- [25] Cook, R.D., Peña, D. & Weisberg, S. (1988). The Likelihood Displacement: A Unifying Principle for Influence Measures. *Communications in Statistic, Theory and methods* 17, 623-640.

- [26] Cordeiro, G. (1992). Introdução à Teoria da Verossimilhança. 10° SINAPE. Rio de Janeiro: ABE.
- [27] Costa, S.C.D. (2003). Modelos Lineares Generalizados Mistos para Dados Longitudinais. Tese de Doutorado. ESALQ/USP, Piracicaba.
- [28] Cox, D.R. & Snell, E.J. (1968). A general Definition of Residuals (with discussion). Journal Royal Statistical Society B 30, 248-275.
- [29] Dempster, A.P. & Ryan, L.M. (1985). Weighted Normal Plots. Journal of the American Statistical Association 80, 845-850.
- [30] 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* B 39, 1-38.
- [31] Dempster, A.P., Rubin, D.B. & Tsutakawa, R.K. (1981). Estimation in Covariance Components Models. *Journal of the American Statistical Association* 76, 341-353.
- [32] Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002). Analysis of Longitudinal Data, 2nd edition. Oxford: Oxford University Press.
- [33] Draper, N.R. & Smith, H. (1998). Applied regression analysis, 3rd Edition. John Wiley & Sons, New York.
- [34] Doganaksoy, N. & Balakrishnan, N. (1997). A Useful Property of Best Linear Unbiased Predictors with Applications to Life-Testing. *The American Statistician* 51, 22-28.
- [35] Fai, A.H.T. & Cornelius, P.L. (1996). Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-Plot Experiments. *Journal Statistical Computing and Simulation* 54, 363-378.
- [36] Fei, Y. & Pan, J. (2003). Influence Assessments for Longitudinal Data in Linear Mixed Models. In 18 th International Workshop on Statistical Modelling. Eds. G. Verbeke, G. Molenberghs, M. Aerts and S. Fieuws. Leuven: Belgium, 143-148.
- [37] Fellner, W.H. (1986). Robust Estimation of Variance Components. *Technometrics* 28, 51-60.

- [38] Fung, W.K. (1993). Unsmasking Outliers and Leverage points: A confirmation. Journal of the American Statistical Associations 88, 515-519.
- [39] Fung, W.K. & Kwan, C.W. (1997). A Note on Local Influence Based on Normal Curvature. Journal of the Royal Statistical Society B 59, 839-843.
- [40] Fung, W.K., Zhu, Z.Y., Wei, B.C. & He, X. (2002). Influence Diagnostics and Outliers tests for Semiparametric Mixed Models. *Journal of the Royal Statistical Society B* 64, 565-579.
- [41] Gray, J.B. (1989). On the Use of Regression Diagnostics. The Statistician 38, 97-105.
- [42] Graybill, F.A. (1983). Matrices with Applications in Statistics, 2nd Edition. California: Wadsworth Publishing Company.
- [43] Grenander, U. (1981). Abstract Inference. New York: Jonh Wiley & Sons.
- [44] Hardin, J.W. and Hilbe, J.M. (2003). Generalized Estimating Equations. Chapman & Hall, New York.
- [45] Harville, D.A. (1976). Extension of The Gauss-Markov Theorem to Include the Estimation of Random Effects. The Annals of Statistics 4, 384-395.
- [46] Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association* 72, 320-340.
- [47] Harville, D.A. (1985). Decomposition of prediction error. Journal of the American Statistical Association 80, 132-138.
- [48] Harville, D.A. (1997). Matrix Algebra from a Statistician's Perspective. Springer-Verlag: New York.
- [49] Harville, D.A. & Jeske, D.R. (1992). Mean Squared Error of Estimation or Prediction Under a General Linear Model. *Journal of the American Statistical Association* 87, 724-731.
- [50] Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated erros. *Journal of the Royal Statistical Society B* 61, 603-609.

- [51] Haslett, J. & Dillane, D. (2004). Application of 'delete=replace' to deletion diagnostics for variance component estimation in the linear mixed model. *Journal of the Royal Statistical Society B* 66, 131-143.
- [52] Haslett, J. & Hayes, K. (1998). Residuals for the Linear Model with General Covariance Structure. Journal of the Royal Statistical Society B 60, 201-215.
- [53] Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics* **31**, 423-447.
- [54] Heyde, C.C. (1997). Quasi-Likelihood and its Applications: A General Approach to Optimal Parameter Estimation. New York: Springer-Verlag.
- [55] Hilden-Minton, J.A. (1995). Multilevel Diagnostics for Mixed and Hierarchical Linear Models. PhD Thesis. University of California, Los Angeles.
- [56] Hoaglin, D.C. & Welsch, R.E. (1978). The Hat Matrix in Regression and ANOVA. The American Statistician 32, 17-22.
- [57] Hodges, J.S. (1998). Some Algebra and Geometry for Hierarchical Models, applied to diagnostics(with discussion). Journal of the Royal Statistical Society B 60, 197-536.
- [58] Hoffman, R. & Vieira, S. (1998). Análise de Regressão: Uma Introdução à Econometria. 3ª Edição. São Paulo: Editora Hucitec.
- [59] Jiang, J. (1996). REML Estimation: Asymptotic Behavior and Related Topics. The Annals of Statistics 24, 255-286.
- [60] Jiang, J. (1997). A derivation of BLUP-Best Linear Unbiased Predictor. Statistics & Probability Letters 32, 321-324.
- [61] Jiang, J. (1998). Asymptotic Properties of the Empirical BLUP and BLUE in Mixed Linear Models. *Statistica Sinica* 8, 861-885.
- [62] Jiang, J. (1999). On Unbiasedeness of the Empirical BLUE and BLUP. Statistics & Probability Letters 41, 19-24.
- [63] Jiang, J. (2001). Goodness-of-fit Tests for Mixed Model Diagnostics. The Annals of Statistics 29, 1137-1164.

- [64] Johnson, B.W. & McCulloch, R.E. (1987). Added-Variable Plots in Linear Regression. *Technometrics* 29, 427-433.
- [65] Jorgensen, M.A. (1993). Influence Functions for Iteratively Defined Statistics. Biometrika 80, 253-265.
- [66] Kackar, R.N. & Harville, D.A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association* 79, 853-862.
- [67] Keselman, H.J., Algina, J., Kowalchuk, R.K. & Wolfinger, R.D. (1998). A Comparison of two Approaches for Selecting Covariance Structures in the Analysis of Repeated Measurements. *Communications in Statistics-Simulation* 27, 591-604.
- [68] Kuk, A.Y.C. (1995). Asymptotically Unbiased Estimation in Generalized Linear Models with Random Effects. *Journal of the Royal Statistical Society* B 57, 395-407.
- [69] Laird, N.M. & Ware, J.M. (1982). Random-Effects Models for Longitudinal Data. Biometrics 38, 963-974.
- [70] Lange, N. & Ryan, L. (1989). Assessing Normality in Random Effects Models. The Annals of Statistic 17, 624-642.
- [71] Lee, Y. & Nelder, J.A. (1996). Hierarchical Generalized Linear Models (with discussion). Journal of the Royal Statistical Society B 58, 619-678.
- [72] Lee, A.H. & Zhao, Y. (1996). Sensitivity of Pearson's goodness-of-fit Statistic in Generalized Linear Models. Communications in Statistic, Theory and Methods 25, 143-157.
- [73] Lesaffre, E. & Verbeke, G. (1998). Local Influence in Linear Mixed Models. *Biometrics* 54, 570-582.
- [74] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- [75] Lindley, D.V. & Smith, A.F.M. (1972). Bayes Estimates for the Linear Model. Journal of the Royal Statistical Society B 34, 1-41.

- [76] Lindstrom, M. & Bates, D.M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association* 83, 1014-1022.
- [77] Littell, R.C., Pendergast, J. & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* 19, 1793-1819.
- [78] Liu, C. & Rubin, D.B. (1994). The ECME algorithm: A simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika* 81, 633-648.
- [79] Longford, N.T. (2001). Simulation-based diagnostics in random-coefficient models. Journal of the Royal Statistical Society A 164, 259-273.
- [80] Magnus, J.R. & Neudecker, H.(1988). Matrix differential calculus with applications in Statistics and Econometrics. John Wiley & Sons, New York.
- [81] Mak, T.K. (1993). Solving Non-Linear Estimation Equations. Journal of the Royal Statistical Society B 55, 945-955.
- [82] Maritz, J.S. & Lwin, T. (1989). Empirical Bayes Methods, 2nd Edition. London: Chapman & Hall.
- [83] Martin, R.J. (1992). Leverage, Influence and Residuals in Regression Models when observations are correlated. *Communications in Statistics- Theory and Methods* 21, 1183-1212.
- [84] McCulloch, R.E. (1989). Local Model Influence. Journal of the American Statistical Association 84, 473-478.
- [85] McCulloch, C.E. & Searle, S.R. (2001). Generalized, Linear, and Mixed Models. New York: John Wiley & Sons.
- [86] McGilchrist, C.A. (1994). Estimation in Generalized Mixed Models. Journal of the Royal Statistical Society B 56, 61-69.
- [87] McLachlan, G.J. & Krishnan, T. (1997). The EM algorithm and extensions. New York: John Wiley & Sons.
- [88] McLean, A.R., William, L.S. & Stroup, W.W. (1991). A unified Approach to Mixed Linear Models. *The American Stastician* 45, 54-64.

- [89] Meng, Xiao-Li & van Dyk, D. (1998). Fast EM-type implementations for mixed effects models. *Journal of the Royal Statistical Society B* 60, 559-578.
- [90] Miller, J.J. (1977). Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance. *The Annals of Statistics* 5, 746-762.
- [91] Morrison, D.F. (1976). Multivariate Statistical Methods, 2nd edition. New York: McGraw-Hill.
- [92] Natis, L. (2000). Modelos Lineares Hierárquicos. Dissertação de mestrado. IME/USP, São Paulo.
- [93] Nelder, J.A. & Weddeburn, R.W.M. (1972). Generalized Linear Models. Journal of the Royal Statistical Society A 135, 370-384.
- [94] Ofversten, F. (1993). Exact Tests for Variance Components in Unbalanced Mixed Linear Models. *Biometrics* 49, 45-57.
- [95] O'Hara-Hines, R.J. & Carter, E.M. (1993). Improved Added Variable and Partial Residual Plots for Detections of Influential Observations in Generalized Linear Models (with discussion). Applied Statistics' 42, 3-20.
- [96] Oman, S.D. (1995). Checking the assumptions in mixed-model analysis of variance: a residual analysis approach. *Computational Statistics & Data Analysis* **20**, 309-330.
- [97] Ouwens, M.J.N.M., Tan, F.E.S. & Berger, M.P.F. (2001). Local Influence to Detect Influential Data Structures for Generalized Linear Mixed Models. *Biometrics* 57, 1166-1172.
- [98] Parizotto, S.P.C.O.L. (1999). Estudo comparativo da eficácia da escova dental utilizando-se dois tipos de escova, com ou sem, dentifrício e relação do desgate das escovas com o controle da placa bacteriana na dentição decídua. Dissertação de mestrado. Faculdade de Odontologia da Universidade de São Paulo.
- [99] Patterson, H.D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554.
- [100] Paula, G.A. (2003). Modelos de regressão com apoio computacional. São Paulo: IME/USP. (http://www.ime.usp.br/~giapaula)

- [101] Pinheiro, J.C. (1994). Topics in Mixed-Effects Models, PhD Thesis, University of Wisconsin, Madison.
- [102] Pinheiro, J.C. & Bates, D.M. (2000). Mixed-Effects in S and S-PLUS. New York: Springer.
- [103] Pregibon, D. (1981). Logistic regression diagnostics. Annals of Statistics 9, 739-750.
- [104] Puterman, M.L. (1988). Leverage and Influence in Autocorrelated Regression Models. Applied Statistics 37, 76-86.
- [105] Rao, C.R. & Kleffe, J. (1991). Estimation of Variance Components and Applications. Amsterdam: North-Holland series in Statistics and Probability, v.3.
- [106] Robinson, G.K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects(with discussion). Statistical Science 6, 15-51.
- [107] Rocha, F.M.M. (2004). Seleção de Estruturas de Covariância em Dados com Medidas Repetidas. Dissertação de Mestrado. IME/USP, São Paulo.
- [108] Rocke, D.M. (1983). Robust Statistical Analysis of Interlaboratory Studies. Biometrics 70. 421-431.
- [109] SAS Institute Inc. (1997). SAS/STAT Software: Changes and Enhancements Through Release 6.12, Cary, NC: SAS Institute Inc.
- [110] Schall, R. (1991). Estimation in Generalized Linear Models with Random Effects. Biometrika 78, 719-727.
- [111] Schawarzmann, B. (1991). A connection Between Local-Influence Analysis and Residuals Diagnostics. *Technometrics* 35, 103-104.
- [112] Searle, S.R., Cassela, G. & McCullogh, C.E. (1992). Variance Components. New York: Jonh Wiley & Sons.
- [113] Sef, H.C. (1999) Modelos de regressão multiplicativos para dados pré-teste/pós-teste em blocos. Dissertação de mestrado. IME/USP.
- [114] Self, S.G. & Liang, K.Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of* the American Statistical Association 82, 605-610.

- [115] Sen, P.K. & Singer, J.M. (1993). Large Sample Methods in Statistics: An Introduction With Applications. New York: Chapman & Hall.
- [116] Singer, J.M. & Andrade, D.F. (1997). Regression Models for the Analysis of Pretest/Posttest Data. *Biometrics* 53, 729-735.
- [117] Singer, J.M. & Andrade, D.F. (2000). Analysis of longitudinal data. In Handbook of Statistics, Volume 18: Bio-environmental and Public Health Statistics. Eds. P.K. Sen and C.R. Rao. Amsterdam: North Holland, 115-160.
- [118] Singer, J.M., Nobre, J.S. & Sef, H.C. (2004). Regression models for pretest/posttest data in blocks. *Submetido para publicação*.
- [119] Souza, F.A.M.D. (1999). Influência Local e Análise de Resíduos em Modelos de Regressão Von Mises. Tese de Doutorado. IME/USP, São Paulo.
- [120] Stram, D.O. & Lee, J.W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics* 50, 1171-1177.
- [121] St. Laurent, R.T. & Cook, R.D. (1993). Leverage Influence and Curvature in Nonlinear Regression. *Biometrika* 80, 99-106.
- [122] Suyama, E. (1995). Modelos de Efeitos Aleatórios para Dados Longitudinais. 40^a RBRAS e 6^a SEAGRO. Ribeirão Preto-SP.
- [123] Tan, F.E.S., Ouwens, M.J.N. & Berger, M.P.F. (2001). Detection of Influential Observations in Longitudinal mixed effects regression models. *The Statistician* 50, 271-284.
- [124] Tempelman, R.J. & Gianola, D. (1996). A Mixed Effects Model for Overdispersed Count Data in Animal Breeding. *Biometrics* 52, 265-279.
- [125] Tountenburg, H. (1982). Prior information in Linear Models. Chichester: John Wiley & Sons.
- [126] Tsai, C.L. (1994). A Note on Jorgensen's Iteratively Definde Statistics. *Biometrika* 81, 781-786.
- [127] Venables, W.N. & Ripley, B.D. (1999). Modern Applied Statistics with S-Plus, 3rd Edition. Springer: New York.

- [128] Venezuela, M.K. (2003). Modelos Lineares Generalizados para Análise de Dados com Medidas Repetidas. Dissertação de Mestrado. IME/USP, São Paulo.
- [129] Verbeke, G. (1995). The linear mixed model. A critical investigation in the context of longitudinal data analysis. PhD Thesis. Catholic University of Leuven, Faculty of Science, Department of Mathematics, Leuven: Belgium.
- [130] Verbeke, G. & Lesaffre, E. (1996a). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* 91, 217-221.
- [131] Verbeke, G. & Lesaffre, E. (1996b). Large Samples properties of the maximum likelihood estimators in linear mixed models with misspecified random-effects distributions. Technical report, Biostatistical Centre for Clinical Trials, Catholic University of Leuven, Belgium.
- [132] Verbeke, G. & Lesaffre, E. (1997). The effect of misspecifying the random-effects distributions in linear mixed models for longitudinal data. *Computational Statistics* & Data Analysis 23, 541-556.
- [133] Verbeke, G. & Molenberghs, G. (1997). Linear Mixed Models In Pratice: A SAS Oriented Approach. Lecture Notes in Statistics 126. New York: Springer-Verlag.
- [134] Verbeke, G. & Molenberghs, G. (2003). The Use of Score Tests for Inference on Variance Components. *Biometrics* 59, 254-262.
- [135] Ware, J.H. (1985). Linear Models for the Analysis of Longitudinal Studies. The American Statistician 39, 95-101
- [136] 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.
- [137] Wei, B.C., Hu, Y.Q. & Fung, W.K. (1998). Generalized Leverage and its Applications. Scandinavian Journal of Statistics 25, 25-37.
- [138] Weiss, R.E. (1995). Residual and Outliers in Repeated Measures Random Effects. *Technical Report*, University of California at Los Angeles, Department of Biostatistics. Models.(http://www.rem.ph.ucla.edu/~rob/papers/outlier.ps)

- [139] Weiss, R.E. & Lazaro, C.G. (1992). Residual plots for repeated measures. Statistics in Medicine 11, 115-124.
- [140] Williams, D.A. (1982). Extra-Binomial Variation in Logistic Linear Models. Applied Statistical 31, 144-148.
- [141] Wolfinger, R. (1993). Covariance Structure Selection in General Mixed Models. Communications in Statistics-Simulation 22, 1079-1106.
- [142] Wolfinger, R., Tobias, R.D., and Sall, J. (1994). Computing Gaussian Likelihoods and their Derivatives for General Linear Mixed Models. SIAM Journal on Scientific Computing, 15, 1294 -1310.
- [143] Wu, X. & Luo, Z. (1993a). Second-Order Approach to Local Influence. Journal of the Royal Statistical Society B 55, 929-936.
- [144] Wu, X. & Luo, Z. (1993b). Residual sum of Squares and Multiple potential, diagnostics by a second order local approach. *Statistical & Probability Letters* 16, 289-296.
- [145] Zhu, H.T. & Lee, S.Y. (2001). Local Influence for incomplete-data models. Journal of the Royal Statistical Society B 63, 111-126.
- [146] Zhu, H.T. & Lee, S.Y. (2003). Local Influence for generalized linear mixed models. The Canadian Journal of Statistics 31, 293-309.