
INFLUÊNCIA LOCAL ATRAVÉS DA CURVATURA NORMAL EM MODELOS
MULTINÍVEIS

JULIANA FREITAS PIRES

Orientadora: Profa. Dra. Claudia Regina Oliveira de Paiva Lima

Co-orientadora: Profa. Dra. Maria Cristina Falcão Raposo

Área de Concentração: Estatística Aplicada

Dissertação submetida como requerimento parcial para obtenção do grau
de Mestre em Estatística pela Universidade Federal de Pernambuco

Recife, fevereiro de 2009

Pires, Juliana Freitas

Influência local através da curvatura normal em modelos multiníveis / Juliana Freitas Pires - Recife : O Autor, 2009.

xii, 75 folhas : il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCEN. Estatística, 2009.

Inclui bibliografia e apêndice.

1. Estatística aplicada - Modelagem. I. Título.

519.536

CDD (22.ed.)

MEI2009-024

Universidade Federal de Pernambuco Pós-Graduação em Estatística

20 de fevereiro de 2009
(data)

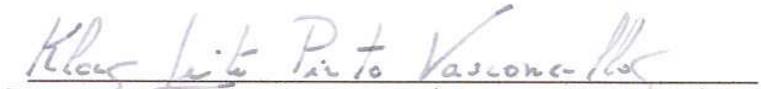
Nós recomendamos que a dissertação de mestrado de autoria de

Juliana Freitas Pires

intitulada

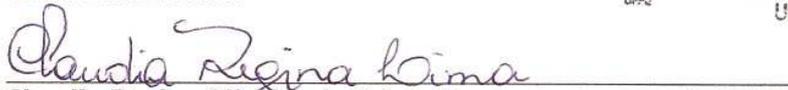
“Influência Local Baseada na Curvatura Normal em Modelos Multiníveis”

seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.



Coordenador da Pós-Graduação em Estatística
Prof. Klaus Leite P. Vasconcelos
 Coordenador de Pós-Graduação em Estatística
UFPE

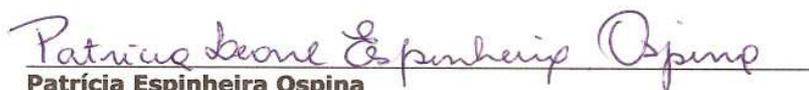
Banca Examinadora:



Claudia Regina Oliveira de Paiva Lima orientador



Viviana Giampaoli (USP)



Patricia Espinheira Ospina

Este documento será anexado à versão final da dissertação.

*Este trabalho é carinhosamente dedicado
a minha mãe, Marta.*

Agradecimentos

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.

À minha mãe, Marta, pelo amor incondicional, carinho, dedicação, confiança, paciência, amizade e por ter me ensinado os valores e princípios que até hoje norteiam a formação do meu caráter. Ao meu pai, por fazer parte da minha formação.

Ao meu companheiro Marcelo, por ser o melhor marido do mundo e ter me ajudado em todos os momentos neste período do mestrado.

À minha irmã pelo amor e carinho e Marcelinho por ter esse carinho de irmão por mim.

À minha avó, Antonina, pelo amor, carinho, confiança e apoio que sempre me deu, além de ser esse exemplo de mulher forte e cheia de vida.

Aos meus tios e tias, carinho e apoio durante todo esse tempo. Em especial a tio Augusto, Renato, Cau, e as tias Ana, Eliana e Mara pelo apoio em momentos difíceis.

Às minhas orientadoras, Claudia Regina e Maria Cristina, pela confiança, dedicação, paciência e competência durante todo o período de desenvolvimento deste trabalho.

Aos meus primos, pela convivência, amizade e confiança.

Aos meus amigos de Feira em especial às meninas (Lídia, Mona, Ari e Lane) e aos meninos (Gugga, Samuca, Tarci e Murilo) que além da convivência, amizade, confiança e muitos momentos especiais foram responsáveis por alguns valores que levo até hoje.

Aos meus amigos de Ilhéus, por terem sido minha família durante alguns anos, saudosos, da minha vida e que continuaram a estar presente nela.

Às especiais Rosas, que são as mais belas rosas do meu jardim e em especial a Juliana pois seria certamente impossível falar da minha vida sem falar dela.

Aos meus amigos do mestrado, Alice, Andrea, Lídia, Manoel, Olga, Vagner, Wilton pela amizade, carinho e pelos momentos de alegria compartilhados.

Aos meus amigos do “ano da tese”, Hemílio, Larissa, Katya, Lilian, Abraão, Valmir, Fábio Veríssimo e Raphael.

Aos meus amigos da graduação, em especial Cecilio e Lídia pela grande amizade e muitos anos de “crescimento” juntos.

À Valéria Bittencourt, pelo enorme carinho, paciência e amizade com que sempre tratou a mim e aos demais alunos do mestrado.

Aos professores do Departamento de Estatística, em especial aos professores Francisco Cribari-Neto, Francisco Cysneiros e Getúlio Amaral.

À banca examinadora, pelas valiosas críticas e sugestões.

Ao CNPq pelo apoio financeiro.

Se cada dia cai

*Se cada dia cai, dentro de cada noite,
há um poço
onde a claridade está presa.*

*Há que sentar-se na beira
do poço da sombra
e pescar luz caída
com paciência.*

Pablo Neruda.

Resumo

Os modelos lineares hierárquicos foram desenvolvidos para análise de dados que possuem uma estrutura de grupo, ou seja, uma estrutura de hierarquia, por levarem em consideração a dependência dos dados existente dentro de cada nível hierárquico. As estimativas dos parâmetros dos modelos hierárquicos são apresentadas separando os efeitos fixos dos aleatórios. O modelo linear hierárquico com dois níveis é apresentado e ilustrado através de um exemplo, juntamente com a sua teoria de estimação e testes de hipóteses. Para avaliar este modelo com respeito a sensibilidade sob pequenas perturbações, uma técnica de influência local sob a curvatura normal foi desenvolvida. Os dados do SAEPE 2005 foram estudados sob este enfoque.

Palavras-chave: Modelo linear hierárquico, efeitos fixos, efeitos aleatórios e influência local.

Abstract

The hierarchical linear models were developed for analysis of data that possess a group structure, in other words, a structure of hierarchy, for they take inside into account the variability of the existent data of each hierarchical level. The estimates of the parameters of the hierarchical models are presented separating the fixed effects from the random effects. Two level hierarquical linear models are introduced and illustrated by one example, including the estimation theory and the hypothesis testing. To assess this model with respect sensitivity under small disturbances, one local influence technique under normal curvature was proposal. The SAEPE 2005 data was studied under this approach.

Keywords: Hierarchical lineal models, fixed effects, random effects and local influence.

Lista de Figuras	xi
Lista de Tabelas	xii
1 Introdução	1
1.1 Introdução	1
2 Modelo Linear Hierárquico	6
2.1 Exemplo	6
2.2 Modelo Linear Hierárquico para Dois Níveis	9
2.3 Métodos de Estimação	11
2.3.1 BLUE e BLUP para MLH	12
2.3.2 Métodos de Máxima Verossimilhança	15
2.3.3 Algoritmo EM	18
2.4 Testes de Hipóteses.	19
2.4.1 Testes de Hipóteses para Efeitos Fixos	19
2.4.2 Testes de Hipóteses para Efeitos Aleatórios	20
2.4.3 Testes de Hipóteses para Componentes de Variância e Covariância	21

3	Influência Local	23
3.1	Influência Local em Modelos Lineares Hierárquicos	26
4	Aplicação do MLH na Área de Educação	30
4.1	Análise Descritiva	31
4.2	Especificação do Modelo	38
4.3	Análise de Influência Local para os Modelos Ajustados	40
5	Considerações Finais	46
5.1	Conclusões	46
5.2	Direções para Pesquisas Futuras	46
5.3	Plataforma Computacional	47
A	Expressões do Capítulo 2	48
A.1	BLUE	48
A.2	BLUP	49
A.3	Identities (2.29) e (2.31)	50
A.4	Identity (2.35)	50
B	Expressões do Capítulo 3	52
B.1	Perturbação na matriz de covariâncias dos erros (nível 1)	52
B.2	Perturbação na variável resposta	54
B.3	Perturbação na matriz de covariâncias dos efeitos aleatórios (nível 2)	55
B.4	Matriz Hessiana	56
C	Rotinas em Linguagem R	58
	Referências Bibliográficas	71

Lista de Figuras

4.1	Histograma das habilidades em matemática: (a) População, (b) Amostra.	32
4.2	Histograma das habilidades em português: (a) População, (b) Amostra.	33
4.3	Gráficos de influência local para o modelo ajustado para Matemática.	42
4.4	Gráficos de influência local para o modelo ajustado para Português.	43

Lista de Tabelas

4.1	Estatísticas descritivas das habilidades de português e matemática.	32
4.2	Estatísticas descritivas e p -valores dos testes de igualdade de médias para a Habilidade em Português.	36
4.3	Estatísticas descritivas e p -valores dos testes de igualdade de médias para a Habilidade em Matemática.	37
4.4	Parâmetros estimados do modelo final para as habilidades de matemática e níveis descritivos.	39
4.5	Parâmetros estimados do modelo final para as habilidades de português e níveis descritivos.	40
4.6	AIC, $\sqrt{\hat{\tau}_{00}}$ e $IC95\%$ para $\sqrt{\hat{\tau}_{00}}$ para o modelo de matemática completo e para os modelos retirando as escolas mais influentes.	44
4.7	AIC, $\sqrt{\hat{\tau}_{00}}$ e $\sqrt{\hat{\tau}_{11}}$ com seus respectivos $IC95\%$ para o modelo de português completo e para os modelos retirando as escolas mais influentes.	44

1.1 Introdução

Estruturas hierárquicas são facilmente encontradas nos dados em diversas áreas de pesquisa, sendo caracterizadas pela presença de grupos. Uma situação na qual se observa este tipo de estrutura referida por Goldstein (1999) são os estudos sobre animais e humanos que têm uma ordem natural dada pela relação de herança, onde a descendência se agrupa dentro de famílias. Descendentes tendem a ser mais semelhantes em suas características físicas e mentais que indivíduos escolhidos de forma aleatória em uma grande população. Crianças de uma mesma família podem ter tendência a ter baixa estatura, talvez porque seus pais tenham baixa estatura ou porque vivam em um ambiente comum que favoreça essa característica.

Em muitas situações da vida real tem-se estruturas de agrupamento nos dados, como uma propriedade intrínseca da população. Porém, muitas vezes algumas abordagens estatísticas tradicionais aplicadas para as análises dos dados não levam em conta tal estrutura e, deixando de considerar tal estrutura podem, como consequência, gerar alguns tipos de erros.

O modelo de regressão simples é um exemplo de um método estatístico clássico que trata os indivíduos independentes da origem ou da estrutura grupal em que estão inseridos e, desta forma, os possíveis efeitos dessas estruturas que foram ignorados podem conduzir a conclusões incorretas sobre significância estatística de um parâmetro, por exemplo, quando na verdade

este não é significativo. Esse tipo de erro pode acontecer porque indivíduos pertencentes a um mesmo grupo tendem a ser mais “parecidos” que indivíduos de diferentes grupos, assim, suposições clássicas de independência não são válidas. Em consequência, métodos estatísticos de modelagem que ignoram a estrutura hierárquica dos dados vão, possivelmente, superestimar a variância dos estimadores dos parâmetros (Noortgate et al. 2005).

Para as situações em que devemos levar em conta a estrutura de agrupamento dos dados, o modelo multinível ou hierárquico possui suas vantagens, pois respeita esse tipo de estrutura de forma parcimoniosa, permitindo que cada um dos níveis de estrutura hierárquica seja especificado separadamente e posteriormente sejam reunidos em um único modelo. Além disso, permitem a incorporação de efeitos aleatórios associados a cada um dos seus níveis de hierarquia. Essa abordagem evita erros ligados a interpretação dos resultados, erros que consistem em analisar os dados de um nível, e emitir as conclusões em outro nível.

Na avaliação educacional é muito comum encontrar aplicações dos modelos multiníveis ou hierárquicos, porque a estrutura hierárquica pode ser facilmente identificada, sendo os alunos agrupados em turmas e as turmas agrupadas em escolas e assim por diante. A utilização do modelo linear hierárquico (MLH) na área de educação permite que variáveis individuais como por exemplo gênero do aluno, nível sócio-econômico, grau de escolaridade da mãe, possam ser incorporados ao primeiro nível do modelo. Variáveis associadas à escola, como por exemplo, se ela é pública ou privada, se possui ou não biblioteca, podem ser inseridas no segundo nível do modelo. Em resumo, cada nível do MLH pode apresentar variáveis associadas às unidades experimentais que o representam, com o objetivo de tentar explicar as diferentes fontes de variabilidade da variável resposta e de estudar as possíveis relações entre cada uma destas variáveis explicativas e a resposta (Natis 2000).

Em estudos como o de Soares (2005), os modelos de regressão hierárquicos foram utilizados considerando três níveis de hierarquia (alunos, turmas e escolas). O interesse de Soares (2005) foi explicar a proficiência dos alunos da 4^a série do ensino fundamental alcançada na avaliação de língua portuguesa do Programa Mineiro de Avaliação da Educação Básica (PROEB/SIMAVE-2002). O estudo permitiu avaliar a proporção da variabilidade das proficiências dos alunos devida

às diferenças entre os alunos, entre as turmas e entre as escolas.

Recentemente Laros & Marciano (2008) utilizaram a modelagem hierárquica no banco de dados de acesso público NELS:88 (*National Educational Longitudinal Study*), este banco é o resultado de uma pesquisa conduzida pelo Ministério da Educação dos Estados Unidos com o propósito de identificar o efeito de variáveis sócio-econômicas e escolares sobre o desempenho educacional. Neste trabalho os autores mostram como estes dados são adequados para ilustrar, de forma prática, como funciona uma análise multinível.

A utilização dos modelos hierárquicos tem crescido bastante nos últimos anos também na área de saúde, este crescimento se deve ao interesse em determinar variáveis que, ao nível do grupo e/ou do indivíduo, influenciam potencialmente no desenvolvimento de determinada doença. Por exemplo, pesquisadores da área de saúde podem estar interessados em avaliar, em uma determinada empresa, se a idade, uma característica do nível do indivíduo, e o setor de trabalho, que pode ser uma característica do nível do grupo, influenciam nos níveis pressóricos dos mesmos.

Na área financeira é possível encontrar aplicações onde a utilização dos modelos hierárquicos pode ser mais atraente, como o interesse em estudar o volume de aplicação na caderneta de renda fixa em função das características dos clientes de agências bancárias e, em seguida, explicar as diferenças entre as agências (isto é, estrutura técnica, atendimento, localização, etc.). Evidentemente, nestes estudos não seria adequado que os dados fossem analisados usando métodos padrões de regressão como, por exemplo, Modelos Lineares Generalizados (McCullagh & Nelder 1989), pois estes, como dito acima, ignoram a natureza hierárquica dos dados.

Segundo Nobre (2004), assim como a escolha de um modelo estatístico que seja mais apropriado aos dados, uma avaliação do mesmo faz-se necessária, visto que o ajuste desses modelos é uma aproximação de processos complexos, e portanto sucessíveis a erros. Um item de grande importância na análise de tais modelos é a sua “validação”, usualmente concretizada por meio da **análise de diagnóstico** que consiste em duas etapas: **avaliação da qualidade 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 varia-

ção é “substancial” no sentido 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.

Um método bastante conhecido na análise de sensibilidade é a deleção de pontos, o qual avalia o impacto da retirada de uma observação particular nas estimativas dos parâmetros nas mesmas, podendo então detectar pontos que exerçam um peso desproporcional nas estimativas dos parâmetros. Outros métodos que generalizam essa idéia são propostos por Cook (1977) e Belsley et al. (1980), essa generalização baseia-se na avaliação da influência a partir de esquemas de perturbação por ponderação, em modelos de regressão normal linear, de forma que os pesos são associados às observações e assumem valores 0 ou 1, indicando se a mesma foi excluída ou mantida no conjunto de dados, respectivamente.

Contudo, uma proposta interessante na área foi apresentada por Cook (1986), consistindo em um método bastante geral para avaliar a influência conjunta das observações sob pequenas mudanças no modelo (no caso de regressão linear), ao invés da avaliação individual ou conjunta dos pontos. Essa metodologia, denominada **influência local** teve grande aceitação sobre os pesquisadores e usuários de regressão (Paula 2004).

Rocke (1983) e Fellner (1986) desenvolveram métodos robustos de estimação dos parâmetros 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) estudam este tipo de influência em modelos lineares mistos semiparamétricos.

A proposta desse trabalho é apresentar o modelo de regressão multinível de dois níveis e apresentar o método de diagnóstico da influência local através da abordagem de Cook (1986) sob os esquemas de perturbação de Beckman et al. (1987). Uma aplicação aos dados do SAEPE

(Sistema de Avaliação Educacional de Pernambuco) de 2005 é apresentada.

A presente dissertação está dividida em cinco capítulos e dois apêndices. No Capítulo 2 apresentamos uma revisão bibliográfica do modelo em estudo. No Capítulo 3 apresentamos um estudo sobre técnicas de avaliação de influência local para o modelo de interesse. No Capítulo 4 apresentamos uma aplicação aos dados de educação considerando as análises de diagnósticos. Finalmente, no Capítulo 5 apresentamos as principais conclusões, incluindo sugestões para trabalhos futuros. Nos apêndices apresentamos os desenvolvimentos de algumas expressões expostas nos Capítulos 2 e 3.

O modelo de regressão multinível como é denominado por Hox (1995) e Goldstein (1999), assume uma variedade de nomes tal como modelo linear hierárquico chamado por Bryk & Raudenbush (1992). Este modelo permite que diferentes níveis sejam especificados em modelos separados e depois combinados em um único modelo.

Este tipo de modelagem assume que há um conjunto de dados hierárquicos, que possui uma variável resposta (Y) que é medida no nível individual, e variáveis explicativas que podem residir no nível do indivíduo (X) e/ou do grupo (Z), que é um nível mais elevado. Conceitualmente o modelo pode ser visto como um sistema hierárquico de equações de regressão (Pinheiro 2005).

Neste capítulo introduziremos o conceito de modelo linear hierárquico (MLH). Iniciaremos com um exemplo simples, extraído de Bryk & Raudenbush (1992), para ajudar a familiarização e compreensão do MLH. Daí, seguiremos com expressões gerais definindo o modelo para dois níveis.

2.1 Exemplo

Vamos considerar uma população de escolas onde foram amostradas J unidades. Em cada escola foram selecionados n_j alunos, sendo observadas as seguintes variáveis quantitativas:

Y_{ij} : Aproveitamento em matemática (AM) do i -ésimo indivíduo da escola j (variável res-

posta),

X_{ij} : Nível sócio-econômico (NSE) do i -ésimo aluno da escola j (variável explicativa), $i = 1, 2, \dots, n_j$; $j = 1, 2, \dots, J$

Através do MLH com dois níveis podemos estudar a relação entre NSE e AM dentro de uma população de escolas, levando em conta a possível variação de interceptos e/ou inclinações entre as escolas. Desta forma, nós podemos descrever o relacionamento dentro da escola j pela equação

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + \epsilon_{ij}, \quad (2.1)$$

onde

β_{0j} : AM esperado de um aluno da escola j com $X_{ij} = \bar{X}_{.j}$;

β_{1j} : mudança esperada no AM quando X_{ij} aumenta em uma unidade em relação a $\bar{X}_{.j}$;

ϵ_{ij} : erro aleatório associado ao i -ésimo aluno da escola j ;

$\bar{X}_{.j}$: média do NSE dos alunos da escola j .

Com a suposição que $\epsilon_{ij} \sim N(0, \sigma^2)$ e ϵ'_{ij} s independentes.

Observe na equação (2.1), que é chamada de modelo do nível 1 (nível do aluno), a variável NSE não foi utilizada em seu valor original. Ela foi centrada na sua média para cada escola, para tornar mais apropriado o significado do intercepto. Para mais detalhes sobre esta mudança de escala veja Bryk & Raudenbush (1992).

O intercepto β_{0j} e a inclinação β_{1j} podem ser modelados da seguinte forma:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \delta_{0j}, \\ \beta_{1j} &= \gamma_{10} + \delta_{1j}, \end{aligned} \quad (2.2)$$

onde

γ_{00} : valor esperado dos interceptos na população de escolas;

γ_{10} : valor esperado das inclinações na população de escolas;

δ_{0j} : efeito aleatório da escola j no intercepto β_{0j} ;

δ_{1j} : efeito aleatório da escola j na inclinação β_{1j} .

Assumimos que δ_{qj} são variáveis aleatórias independentes que têm distribuição normal com média zero e matriz de variâncias e covariâncias τ_j . Representamos a matriz de dispersão de

efeitos aleatórios do segundo nível como

$$\tau_j = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}, \quad (2.3)$$

em que

τ_{00} : variância populacional dos interceptos;

τ_{11} : variância populacional das inclinações;

$\tau_{01} = \tau_{10}$: covariância entre β_{0j} e β_{1j} .

As equações em (2.2) formam o modelo do nível 2 (nível da escola). Ele retrata que as escolas não possuem o mesmo intercepto e a mesma inclinação, e os componentes aleatórios δ_{0j} e δ_{1j} ajudam a explicar essas diferenças.

Podemos também incluir variáveis relacionadas às escolas no modelo do nível 2 para tentar explicar a variabilidade dos interceptos e das inclinações, em outras palavras, para tentar explicar a variabilidade entre as escolas. Considere, por exemplo, a variável

$$Z_j = \begin{cases} 1 & \text{se a escola } j \text{ é privada,} \\ 0 & \text{se a escola } j \text{ é pública.} \end{cases} \quad (2.4)$$

O modelo para o nível 2 com a variável adicionada, relacionada à escola, pode ser escrito como:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}Z_j + \delta_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + \delta_{1j}. \end{aligned} \quad (2.5)$$

Neste novo modelo temos

γ_{00} : valor esperado dos interceptos nas escolas públicas;

γ_{01} : diferença entre os valores esperados dos interceptos de escolas públicas e privadas;

γ_{10} : valor esperado das inclinações nas escolas públicas;

γ_{11} : diferença entre os valores esperados das inclinações de escolas públicas e privadas;

δ_{0j} : efeito aleatório da escola j no intercepto β_{0j} ;

δ_{1j} : efeito aleatório da escola j na inclinação β_{1j} .

Assumimos que δ_{0j} e δ_{1j} são variáveis aleatórias que têm distribuição normal com média zero, variâncias τ_{00} e τ_{11} , respectivamente, e covariância τ_{01} . Ou seja,

τ_{00} : variância populacional dos interceptos corrigida pela variável Z_j ;

τ_{11} : variância populacional das inclinações corrigida pela variável Z_j ;

τ_{01} : covariância entre β_{0j} e β_{1j} .

Note agora que essas variâncias e covariâncias são componentes de variância e covariância condicionais (ou residuais), pois a variável Z_j está sendo considerada.

Substituindo a equação (2.5) em (2.1) teremos a seguinte forma combinada:

$$Y_{ij} = \gamma_{00} + \gamma_{01}Z_j + \gamma_{01}(X_{ij} - \bar{X}_{.j}) + \gamma_{11}Z_j(X_{ij} - \bar{X}_{.j}) + \delta_{0j} + \delta_{1j}(X_{ij} - \bar{X}_{.j}) + \epsilon_{ij}. \quad (2.6)$$

O erro aleatório da equação (2.6), $\delta_{0j} + \delta_{1j}(X_{ij} - \bar{X}_{.j}) + \epsilon_{ij}$, é bastante complexo. Dentro de uma mesma escola, tais erros são dependentes, já que δ_{0j} e δ_{1j} são comuns a todos os alunos desta escola. Além disso estes erros possuem variâncias desiguais, pois dependem dos valores de $(X_{ij} - \bar{X}_{.j})$ que variam entre os estudantes.

2.2 Modelo Linear Hierárquico para Dois Níveis

Agora descrevemos o MLH mais geral para dois níveis de hierarquia. Suponha que temos N sujeitos naturalmente agrupados em J unidades, onde existem n_j sujeitos na j -ésima unidade e $N = \sum_{j=1}^J n_j$. Além disso, suponha que para cada unidade j o modelo pode ser escrito da forma

$$\underline{Y}_j = \mathbf{X}_j \underline{\beta}_j + \underline{\epsilon}_j, \quad (2.7)$$

onde \underline{Y}_j é o vetor da variável resposta do grupo j que tem dimensão $n_j \times 1$, $j = 1, 2, \dots, J$; \mathbf{X}_j , com dimensão $n_j \times Q$, é a matriz de variáveis preditoras do nível 1; $\underline{\beta}_j$ é o vetor de parâmetros desconhecidos ($Q \times 1$); $\underline{\epsilon}_j$ é o vetor de erros aleatórios o qual assumimos ser normalmente distribuído com $E(\underline{\epsilon}_j) = \mathbf{0}$ e $Var(\underline{\epsilon}_j) = \sigma_j^2 \mathbf{R}_j$ e também fazemos a suposição de independência entre $\underline{\epsilon}_j$ e $\underline{\epsilon}_{j'}$ para $j \neq j'$. \mathbf{R}_j é uma matriz conhecida, positiva definida. Frequentemente, $\mathbf{R}_j = \mathbf{I}_{n_j}$, neste caso o modelo (2.7) é denominado de modelo de independência condicional homoscedástico, o qual será considerado no presente trabalho.

No segundo nível nós temos $\underline{\beta}_j$ variando no modelo de um grupo para outro. Desta forma o modelo geral para $\underline{\beta}_j$ no nível 2 é

$$\underline{\beta}_j = \mathbf{Z}_j \underline{\gamma} + \underline{\delta}_j \quad (2.8)$$

onde, $\mathbf{Z}_j = \begin{bmatrix} Z_{0j} & 0 & \cdots & 0 \\ 0 & Z_{1j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{Qj} \end{bmatrix}$ é a matriz de variáveis preditoras do nível 2,
sendo $\underline{Z}_{qj} = [1 \quad Z_{1j} \quad Z_{2j} \quad \cdots \quad Z_{Pj}]$ o vetor de preditores de $\underline{\beta}_{qj}$ e $\underline{0}$ é um vetor $(P + 1)$
de zeros, $q = 1, \dots, Q$ e $j = 1, \dots, J$;

$\underline{\gamma}^\top = [\gamma_0 \quad \gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_Q]$ é o vetor de efeitos fixos,

sendo $\gamma_q = [\gamma_{q0} \quad \gamma_{q1} \quad \gamma_{q2} \quad \cdots \quad \gamma_{qP}]$;

$\underline{\delta}_j^\top = [\delta_{0j} \quad \delta_{1j} \quad \delta_{2j} \quad \cdots \quad \delta_{Qj}]$ é o vetor de efeitos aleatórios, cuja média é um vetor nulo e com matriz de variâncias $\boldsymbol{\tau}_j$, que pode ser expressa na seguinte forma matricial:

$$\boldsymbol{\tau}_j = \sigma^2 \begin{bmatrix} \tau_{00} & \tau_{01} & \cdots & \tau_{0Q} \\ \tau_{10} & \tau_{11} & \cdots & \tau_{1Q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{Q0} & \tau_{Q1} & \cdots & \tau_{QQ} \end{bmatrix}.$$

Combinando os modelos (2.7) e (2.8) obtemos a equação do modelo combinado

$$\underline{Y}_j = \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} + \mathbf{X}_j \underline{\delta}_j + \underline{\epsilon}_j. \quad (2.9)$$

Temos ainda que

$$E(\underline{Y}_j) = \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \quad (2.10)$$

e

$$\mathbf{V}_j = \text{Var}(\underline{Y}_j) = \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \mathbf{R}_j. \quad (2.11)$$

Em geral, supõe-se que $\text{Var}(\underline{\gamma}_j)$ e $\text{Var}(\underline{\delta}_j)$ são funções de poucos parâmetros θ que independem dos parâmetros de locação $\underline{\gamma}$. Às vezes é comum colocar um parâmetro de dispersão em evidência, ou seja, fazer $\text{Var}(\underline{\gamma}_j) = \sigma^2 \mathbf{R}_j(\theta)$ e $\text{Var}(\underline{\delta}_j) = \sigma^2 \boldsymbol{\tau}_j(\theta)$ com \mathbf{R}_j e $\boldsymbol{\tau}_j$ denotando matrizes positivas definidas, e então

$$\mathbf{V}_j = \text{Var}(\underline{Y}_j) = \sigma^2 \left(\mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \mathbf{R}_j \right). \quad (2.12)$$

É possível também reescrevermos o MLH englobando os J grupos em um único modelo

$$\underline{Y} = \mathbf{XZ}\underline{\gamma} + \mathbf{X}\underline{\delta} + \underline{\epsilon}. \quad (2.13)$$

onde

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_J \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_J \end{bmatrix}, \quad (2.14)$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_J \end{bmatrix}, \quad \underline{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_J \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_J \end{bmatrix} \quad (2.15)$$

e γ como definido anteriormente.

Para este modelo temos as suposições que $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ e $\underline{\delta} \sim N(\mathbf{0}, \mathbf{T})$ onde

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_J \end{bmatrix} \text{ e } \mathbf{T} = \begin{bmatrix} \tau_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tau_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tau_J \end{bmatrix} \quad (2.16)$$

2.3 Métodos de Estimação

Em uma análise hierárquica com dois níveis há três tipos de parâmetros que podem ser estimados: os efeitos fixos, os efeitos aleatórios do nível 1 e as componentes de variância e covariância (Bryk & Raudenbush 1992).

Diversos métodos de estimação estão disponíveis na literatura, dentre eles convém destacar os métodos Bayesianos (Lindley & Smith 1972), os métodos de máxima verossimilhança restrita e completa (Bryk & Raudenbush 1992) e o método de Mínimos Quadrados. Estes métodos produzirão resultados semelhantes em grandes amostras, mas, em pequenas amostras, esses resultados podem ser diferentes.

Na escolha de teorias que envolvem equações não lineares para estimação dos parâmetros, métodos estatísticos computacionais serão necessários. Dempster et al. (1977) desenvolveram o enfoque mais comumente usado do algoritmo EM; podemos citar ainda outros métodos iterativos tais como Newton-Raphson (Lindstrom & Bates 1988) e Scoring de Fisher (Longford 1987).

O objetivo desta seção é descrever alguns processos de estimação dos parâmetros do MLH com dois níveis. Inicialmente apresentamos os estimadores obtidos através do teorema de Gauss-Markov, BLUE (Best Linear Unbiased Estimator) o melhor estimador linear não viesado para γ

e o BLUP (Best Linear Unbiased Predictor) o melhor preditor linear não viesado para o vetor de efeitos aleatórios $\underline{\delta}$. Seguimos apresentando os métodos de Máxima Verossimilhança (MV) e finalizamos a seção descrevendo o algoritmo EM.

2.3.1 BLUE e BLUP para MLH

Na sua origem o BLUE e o BLUP foram descritos como os “Estimadores de Máxima Verossimilhança” (EMV) de $\underline{\gamma}$ e $\underline{\delta}$ obtidos através da densidade conjunta do vetor aleatório $(\underline{Y}, \underline{\delta})$, sob a suposição de normalidade de $\underline{\gamma}$ e $\underline{\delta}$, 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.

Os BLUE e BLUP para modelos hierárquicos podem ser obtidos pela metodologia dos modelos mistos como visto em Hilden-Minton (1995). Nesse trabalho o autor comenta que existe uma série de vantagens em se utilizar os casos com restrição (“constraint-cases”) para a obtenção desses estimadores. A idéia básica é reexpressar o modelo linear hierárquico na forma de um modelo linear geral através da inclusão de “casos artificiais” com variâncias desconhecidas. Em Hodges (1998) podemos encontrar também vantagens na utilização desse método, das quais destacam-se a obtenção “imediate” das equações de estimação utilizadas para determinar o BLUE e o BLUP. Vamos adotar esse enfoque e descrevê-lo a seguir.

Seja $\tau_j = \tau$ para todo $j = 1, \dots, J$ e

$$\underline{Y}_j = \underline{Q}_j \underline{\gamma} + \mathbf{X}_j \underline{\beta}_j + \underline{\gamma}_j, \quad (2.17)$$

e consideremos os seguintes J “casos artificiais”

$$\underline{Q} = \mathbf{Z}_j \underline{\gamma} - \underline{\beta}_j + \underline{\delta}_j. \quad (2.18)$$

A partir daí, podemos escrever o modelo como

$$\begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_J \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{X}_1 & & \\ \vdots & & \ddots & \\ \mathbf{0} & & & \mathbf{X}_J \\ \mathbf{Z}_1 & & & \\ \vdots & & \ddots & \\ \mathbf{Z}_J & & & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \underline{\gamma} \\ \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_J \end{bmatrix} + \begin{bmatrix} \underline{\gamma}_1 \\ \vdots \\ \underline{\gamma}_J \\ \underline{\delta}_1 \\ \vdots \\ \underline{\delta}_J \end{bmatrix}. \quad (2.19)$$

Pré-multiplicando (2.19) por:

$$\begin{bmatrix} \mathbf{R}_1^{-1/2} & & & & & \\ & \ddots & & & & \\ & & \mathbf{R}_J^{-1/2} & & & \\ \mathbf{0} & & & \boldsymbol{\tau}^{-1/2} & & \\ & & & & \ddots & \\ & & & & & \boldsymbol{\tau}^{-1/2} \end{bmatrix} \quad (2.20)$$

tem-se

$$\underline{\mathbf{Y}}^* = \mathbf{X}^* \underline{\beta}^* + \underline{\zeta}, \quad (2.21)$$

$$\text{em que, } \underline{\mathbf{Y}}^* = \begin{bmatrix} \mathbf{R}_1^{-1/2} \mathbf{Y}_1 \\ \vdots \\ \mathbf{R}_J^{-1/2} \mathbf{Y}_J \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{X}^* = \begin{bmatrix} \mathbf{0} & \mathbf{R}_1^{-1/2} \mathbf{X}_1 & & \\ \vdots & & \ddots & \\ \mathbf{0} & & & \mathbf{R}_J^{-1/2} \mathbf{X}_J \\ \boldsymbol{\tau}^{-1/2} \mathbf{Z}_1 & & & \\ \vdots & & \ddots & \\ \boldsymbol{\tau}^{-1/2} \mathbf{Z}_J & & & -\boldsymbol{\tau}^{-1/2} \end{bmatrix}, \quad \underline{\beta}^* = \begin{bmatrix} \underline{\gamma} \\ \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_J \end{bmatrix} \text{ e}$$

$$\underline{\zeta} = \begin{bmatrix} \mathbf{R}_1^{-1/2} \underline{\gamma}_1 \\ \vdots \\ \mathbf{R}_J^{-1/2} \underline{\gamma}_J \\ \boldsymbol{\tau}^{-1/2} \underline{\delta}_1 \\ \vdots \\ \boldsymbol{\tau}^{-1/2} \underline{\delta}_J \end{bmatrix}, \text{ tal que } \text{Var}(\underline{\zeta}) = \sigma^2 \mathbf{I}_{JQ+N}. \text{ Desta forma (2.21) pode ser considerado um$$

modelo linear “homoscedástico”. Portanto, o BLUE para $\underline{\gamma}$ e o BLUP para $\underline{\delta}$ podem ser obtidos por meio da equação (2.21), usando o método de Mínimos Quadrados (MQ), sob o qual se obtém

$$(\mathbf{X}^*)^\top \underline{\mathbf{Y}}^* = (\mathbf{X}^*)^\top \mathbf{X}^* \hat{\underline{\beta}}^*, \quad (2.22)$$

ou seja,

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{X}_1^\top \mathbf{R}_1^{-1} \underline{Y}_1 \\ \vdots \\ \mathbf{X}_J^\top \mathbf{R}_J^{-1} \underline{Y}_J \end{bmatrix} = \begin{bmatrix} \sum \mathbf{Z}_j^\top \tau^{-1} \mathbf{Z}_j & -\mathbf{Z}_1^\top \tau^{-1} & \cdots & -\mathbf{Z}_J^\top \tau^{-1} \\ -\tau^{-1} \mathbf{Z}_1 & \mathbf{C}_1 & & \\ \vdots & & \ddots & \\ -\tau^{-1} \mathbf{Z}_J & & & \mathbf{C}_J \end{bmatrix} \begin{bmatrix} \hat{\gamma} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_J \end{bmatrix} \quad (2.23)$$

onde $\mathbf{C}_j = \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j + \tau^{-1}$.

A equação (2.22) é conhecida como equação normal e sua solução independe da distribuição de δ e γ . Resolvendo-as são obtidos os seguintes estimadores de MQ para γ e δ (Ver Apêndice A).

$$\hat{\gamma} = \left[\sum \mathbf{Z}_j^\top \left(\tau^{-1} - \tau^{-1} \mathbf{C}_j^{-1} \tau^{-1} \right) \mathbf{Z}_j \right]^{-1} \sum \mathbf{Z}_j^\top \tau^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j, \quad (2.24)$$

$$\hat{\delta}_j = \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \hat{\gamma} \right). \quad (2.25)$$

Observando que,

$$(\mathbf{V}_j)^{-1} \sigma^2 = \mathbf{M}_j = \mathbf{R}_j^{-1} - \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1}, \quad (2.26)$$

podemos obter a seguinte expressão para o BLUE:

$$\hat{\gamma} = \mathbf{G}^{-1} \sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \underline{Y}_j, \quad (2.27)$$

onde

$$\mathbf{G} = \sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j, \quad (2.28)$$

Usando os fatos que (para mais detalhes sobre as expressões abaixo, veja Apêndice A)

$$\mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} = \tau \mathbf{X}_j^\top \mathbf{M}_j \quad (2.29)$$

e

$$\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \hat{\gamma} = \mathbf{M}_j^{-1} \left(\mathbf{M}_j \underline{Y}_j - \sum_{k=1}^J \mathbf{P}_{jk} \underline{Y}_k \right) \quad (2.30)$$

podemos reescrever $\hat{\delta}_j$ como

$$\hat{\delta}_j = \boldsymbol{\tau} \mathbf{X}_j^\top \left(\mathbf{M}_j \underline{Y}_j - \sum_{k=1}^J \mathbf{P}_{jk} \underline{Y}_k \right) \quad (2.31)$$

ou, equivalentemente

$$\hat{\delta}_j = \boldsymbol{\tau} \mathbf{X}_j^\top \sum_{k=1}^J \mathbf{Q}_{jk} \underline{Y}_k. \quad (2.32)$$

onde

$$\mathbf{P}_{jk} = \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j \mathbf{G}^{-1} \mathbf{Z}_k^\top \mathbf{X}_k^\top \mathbf{M}_k. \quad (2.33)$$

e

$$\mathbf{Q}_{jk} = \begin{cases} \mathbf{M}_j - \mathbf{P}_{jj} & \text{se } k = j \\ -\mathbf{P}_{jk} & \text{se } k \neq j \end{cases}, \quad (2.34)$$

Através das expressões (2.31) e (2.32) observamos mais claramente como \underline{Y}_k influencia $\hat{\delta}_j$, o que não acontece com a expressão (2.25), mostrando assim, como as observações dos diversos grupos influenciam em $\hat{\delta}_j$.

Adicionalmente, nós obtemos (ver Apêndice A)

$$\text{Var}(\hat{\gamma}) = \sigma^2 \mathbf{G}^{-1}, \quad (2.35)$$

as expressões abaixo são encontradas em Hilden-Minton (1995).

$$\text{Var}(\hat{\delta}_j - \delta_j) = \sigma^2 \left[\boldsymbol{\tau} - \boldsymbol{\tau} \mathbf{X}_j^\top \mathbf{Q}_{jj} \mathbf{X}_j \boldsymbol{\tau} \right], \quad (2.36)$$

$$\text{Cov}(\hat{\gamma}, \hat{\delta}_j - \delta_j) = \mathbf{G}^{-1} \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \mathbf{X}_j \boldsymbol{\tau}, \quad (2.37)$$

e se $j \neq k$

$$\text{Cov}(\hat{\delta}_j - \delta_j, \hat{\delta}_k - \delta_k) = \sigma^2 \left[\boldsymbol{\tau} \mathbf{X}_j^\top \mathbf{Q}_{jk} \mathbf{X}_k \boldsymbol{\tau} \right]. \quad (2.38)$$

2.3.2 Métodos de Máxima Verossimilhança

No método de Máxima Verossimilhança as estimativas dos parâmetros são os valores que maximizam a probabilidade (variável discreta) ou a densidade de probabilidade (variável contínua) de ser obtida a amostra observada. É necessário conhecer a distribuição da variável em estudo para obter Estimadores de Máxima Verossimilhança.

Os softwares estatísticos normalmente usam dois procedimentos diferentes de estimação por máxima verossimilhança na análise de modelo linear hierárquico. Um é o procedimento de máxima verossimilhança completo, que trata as estimativas dos coeficientes de regressão como quantidades que são conhecidas quando estimam as componentes de variância, e o outro procedimento é de máxima verossimilhança restrito o qual trata estas estimativas como quantidades que carregam uma certa incerteza. De acordo com Bryk & Raudenbush (1992), para MLH's com 2 níveis, as duas abordagens em geral produzem resultados similares para σ^2 . Porém, podem ocorrer diferenças importantes na estimação da matriz τ_j .

Método de Máxima Verossimilhança Completo

O modelo linear hierárquico descreve a variável dependente Y em que assume-se que esta variável possui distribuição normal, com média dependendo dos coeficientes de regressão e o parâmetro de dispersão dependendo das componentes de variância. Estes parâmetros são estimados pela técnica correspondente, que é simplesmente chamada de máxima verossimilhança, mas às vezes também chamada de máxima verossimilhança completa (Kreft & Leeuw 1998).

De acordo com Bryk & Raudenbush (1992) para um conjunto de possíveis valores dos parâmetros γ , τ_j e σ^2 nas equações (2.7) e (2.8), existe alguma probabilidade de observar uma amostra de valores de \underline{Y} onde \underline{Y} é um vetor $N \times 1$ de valores observados da variável resposta estudada, das J unidades do nível 1. Veja que $N = \sum n_j$, onde n_j é o tamanho da amostra do j -ésimo grupo. A idéia central do método de máxima verossimilhança é escolher estimativas de γ , τ_j e σ^2 para as quais a probabilidade de observar os valores reais de \underline{Y} seja máxima.

Estimadores baseados no método de máxima verossimilhança, sob certas suposições, são consistentes (isto é, eles podem ser bastante próximos do verdadeiro valor do parâmetro com alta probabilidade, se muitos dados forem coletados) e assintoticamente eficientes (isto é, para um tamanho de amostra grande, o estimador de máxima verossimilhança é aproximadamente não viesado com variância mínima).

Outra característica do estimador de máxima verossimilhança é que sua distribuição amostral converge assintoticamente para uma distribuição normal com uma variância que pode ser

estimada facilmente.

Método de Máxima Verossimilhança Restrita

O princípio de máxima verossimilhança aplicado aos resíduos de mínimos quadrados é conhecido como máxima verossimilhança restrita. Como a distribuição dos resíduos não depende das estimativas dos efeitos fixos, só depende das componentes de variância, então, aplicar o princípio de máxima verossimilhança aos resíduos implica que não se pode estimar os coeficientes de regressão. Desta forma, é aplicado um outro princípio para estimar os coeficientes de regressão que é o de mínimos quadrados ponderados, usado para estimar as componentes de variância para construir a matriz ponderada (Kreft e Leeuw, 1998).

O método de máxima verossimilhança restrito requer a integração da verossimilhança com respeito a $\underline{\gamma}$. Esta integração da verossimilhança é uma idéia bayesiana que faz sentido apenas se a verossimilhança for observada como a densidade conjunta à posteriori de $\underline{\gamma}$, τ_j e σ^2 (para maiores detalhes ver Dempster et al. (1981)).

Tratando-se do modelo linear hierárquico, a diferença entre estimativas de variância e covariância baseada no método de máxima verossimilhança completo e máxima verossimilhança restrito, não é expressa em forma algébrica simples. Porém, as estimativas de máxima verossimilhança restrita das componentes de variância são ajustadas por conta da incerteza sobre os efeitos fixos e os resultados obtidos pelo método de máxima verossimilhança completo não sofrem este ajuste.

No caso do modelo linear hierárquico de dois níveis quando o número de grupos do nível 2, J , é grande, os dois métodos de estimação podem produzir resultados muito similares. Entretanto, se J é pequeno o estimador da variância pelo método de máxima verossimilhança completo, $\hat{\tau}_j$, pode produzir resultado menor que o encontrado pelo método de máxima verossimilhança restrita, por um fator de aproximadamente $(J - Q)/J$, onde Q é o número total de elementos do vetor de efeitos fixos, $\underline{\gamma}$ (Pinheiro 2005).

2.3.3 Algoritmo EM

Em muitos casos, o cálculo para obtenção dos estimadores de MV não é possível analiticamente, fazendo-se necessário o uso de métodos iterativos de aproximação numérica, tais como os algoritmos de Newton Raphson, Scoring de Fisher ou o algoritmo EM, como uma forma alternativa para solução deste problema.

O algoritmo EM é amplamente aplicado em uma variedade de problemas, principalmente com dados incompletos, e em muitos casos este método é notavelmente simples, tanto conceitualmente quanto computacionalmente. Diante desses atrativos o EM se torna uma boa opção ao uso dos algoritmos de Newton Raphson e Scoring de Fisher, pois estes em determinadas situações podem se tornar complicados, por requerem o cálculo da matriz de segundas derivadas da função de verossimilhança (McLachlan & Krishnan 1946).

Como relata Sousa (2002) existem na literatura, há muito tempo, algoritmos semelhantes ao EM, entretanto o termo EM foi introduzido por Dempster, Lair e Rubin em 1977, no artigo que se tornou a principal referência sobre o algoritmo EM (Dempster et al. 1977). Nesse trabalho foram desenvolvidas e enunciadas formalmente as propriedades deste método além de se estabelecer um critério de convergência.

Mais especificamente denotamos \underline{Y}_o os dados observados e \underline{Y}_f os dados faltantes. Os dados completos são denotados por $\underline{Y} = (\underline{Y}_o, \underline{Y}_f)$. O logaritmo da função de verossimilhança dos dados completos é denotado por $l_c(\underline{\psi}|\underline{Y}_o, \underline{Y}_f)$, $\underline{\psi} \in \Theta$, $Q(\underline{\psi}, \underline{\psi}^{(m-1)})$ denota o valor esperado de l_c e $\underline{\psi}^{(m-1)}$ os parâmetros estimados. Temos que

$$Q(\underline{\psi}, \underline{\psi}^{(m-1)}) = E \left[l_c(\underline{\psi}|\underline{Y}_o, \underline{Y}_f) | \underline{Y}_o, \underline{\psi}^{(m-1)} \right], \quad (2.39)$$

em que $\underline{\psi}^{(m-1)}$ são os parâmetros estimados que são usados para avaliar a esperança e $\underline{\psi}$, são estimativas atualizadas, obtidas por maximização de Q .

Cada iteração do algoritmo consiste em apenas dois passos, o passo E (obtenção das esperanças) e o passo M (estimação de máxima verossimilhança de dados completos), ou seja,

Passo-E: Calcular $Q(\underline{\psi}, \underline{\psi}^{(m-1)})$ como uma função de $\underline{\psi}$;

Passo-M: Encontrar $\underline{\psi}^{(m)}$, tal que, $Q(\underline{\psi}^{(m)}, \underline{\psi}^{(m-1)}) = \max_{\underline{\psi} \in \Theta} Q(\underline{\psi}, \underline{\psi}^{(m-1)})$.

Assim, os dois passos são repetidos até satisfazer algum critério de convergência. Cada iteração do algoritmo EM aumenta o valor do logaritmo da função de verossimilhança $l(\psi)$ e, geralmente, converge a um máximo local ou global do logaritmo da função de verossimilhança (Montenegro 2006).

Recomenda-se repetir várias vezes as iterações do algoritmo EM com diferentes valores iniciais a fim de verificar se o verdadeiro máximo foi identificado.

2.4 Testes de Hipóteses.

Respostas afirmativas ou negativas para questões importantes são necessárias em atividades nas ciências, indústrias e na vida de um modo geral. Para tentar obter essas respostas, são construídos experimentos cujos resultados têm alguma referência nas questões de interesse. O processo que determina se as respostas desses experimentos são afirmativas ou negativas é chamado de teste de hipótese (Bickel & Doksum 1976).

Neste capítulo apresentamos brevemente as hipóteses que podem ser formuladas e testadas sobre os três tipos de parâmetros do modelo linear hierárquico: os efeitos fixos, as componentes de variância e covariância e os efeitos aleatórios.

2.4.1 Testes de Hipóteses para Efeitos Fixos

Levando em consideração o modelo de regressão apresentado em (2.7) e (2.8) onde supõe-se que existem q ($q = 1, \dots, Q$) variáveis explicativas no nível 1 e p ($p = 1, \dots, P$) variáveis explicativas no nível 2, uma hipótese a ser testada é que $\underline{\gamma}$ é nulo. Desta forma, a hipótese nula pode ser escrita como:

$$H_0 : \mathbf{C}^\top \underline{\gamma} = 0. \quad (2.40)$$

Sendo a variância do vetor $\mathbf{C}^\top \underline{\hat{\gamma}}$ obtida da seguinte forma:

$$Var(\mathbf{C}^\top \underline{\hat{\gamma}}) = \mathbf{C}^\top \mathbf{V}_{\underline{\hat{\gamma}}} \mathbf{C} = \mathbf{V}_C. \quad (2.41)$$

Uma estatística de teste aproximada para a hipótese nula (2.40) é:

$$U = \hat{\underline{\boldsymbol{\gamma}}}^\top \mathbf{C} \mathbf{V}_C^{-1} \mathbf{C}^\top \hat{\underline{\boldsymbol{\gamma}}}, \quad (2.42)$$

que assintoticamente tem distribuição χ^2 sob H_0 , com graus de liberdade igual ao número de linhas em \mathbf{C}^\top .

O teste de razão de verossimilhança é uma outra abordagem para o teste multiparamétrico que pode ser usado quando se trabalha com o método de máxima verossimilhança completo, não sendo aplicável para o método de máxima verossimilhança restrita (Bryk & Raudenbush 1992).

2.4.2 Testes de Hipóteses para Efeitos Aleatórios

Sendo $\underline{\beta}$ o vetor de parâmetros aleatórios com dimensão $J(Q+1) \times 1$, então a hipótese linear geral associada a $\underline{\beta}$ é:

$$H_0 : \mathbf{C}^\top \underline{\beta} = 0. \quad (2.43)$$

Se o vetor de parâmetros pode ser estimado por mínimos quadrados ordinários (MQO), então a hipótese linear geral (2.43) pode ser testada através da estatística:

$$T_{MQO} : \hat{\underline{\beta}}^\top \mathbf{C} \left(\mathbf{C}^\top \hat{\mathbf{V}} \mathbf{C} \right)^{-1} \mathbf{C}^\top \hat{\underline{\beta}}, \quad (2.44)$$

onde $\hat{\mathbf{V}}$ é uma matriz bloco diagonal com cada bloco de dimensão $(Q+1) \times (Q+1)$ e igual a $\hat{\mathbf{V}}_j$. $\hat{\mathbf{V}}_j$ é a estimativa da matriz de dispersão dada em (2.12).

No caso de um único parâmetro, a hipótese de interesse é da forma:

$$H_0 : \beta_{qj} = 0, \quad (2.45)$$

ou mais precisamente:

$$H_0 : \delta_{qj} = 0. \quad (2.46)$$

A estatística de teste para efeitos aleatórios é obtida tomando-se a razão do efeito aleatório estimado pela estimativa do seu erro padrão, como dado abaixo:

$$t = \frac{\hat{\delta}_{qj}}{\sqrt{\widehat{Var}(\hat{\delta}_{qj})}}. \quad (2.47)$$

Esta estatística de teste segue uma distribuição t -Student com $(Q-1)$ graus de liberdade para dados balanceados e para algumas situações de dados não balanceados (Sullivan et al. 1999). O valor amostral da estatística de teste acima deve ser comparado com o ponto crítico $t_{\frac{\alpha}{2},(Q-1)}$ da distribuição t -Student.

2.4.3 Testes de Hipóteses para Componentes de Variância e Covariância

A forma mais geral de teste de hipótese para componentes de variância e covariância é o teste da razão de verossimilhança (RV), onde as hipótese nula e alternativa a serem testadas são respectivamente:

$$H_0 : \mathbf{T} = \mathbf{T}_0, \quad (2.48)$$

$$H_1 : \mathbf{T} = \mathbf{T}_1. \quad (2.49)$$

A estatística de teste do teste da RV é baseada na diferença das deviances do modelo, d_0 (supondo H_0) e d_1 (supondo H_1), sendo a deviance uma medida de ajuste do modelo que corresponde a -2 vezes o logaritmo da sua função de verossimilhança avaliada em seu máximo. A estatística de teste é dada por:

$$T_{RV} = d_0 - d_1 \quad (2.50)$$

e tem distribuição Qui-quadrado com m graus de liberdade, onde m corresponde à diferença do número de parâmetros dos dois modelos.

Uma alternativa para testar se um componente de variância é igual a zero é a estatística

$$z = \frac{\hat{\tau}_{qq}}{[\widehat{Var}(\hat{\tau}_{qq})]^{1/2}}, \quad (2.51)$$

onde a estimativa do erro padrão de $\hat{\tau}_{qq}$ é calculada através da inversa da matriz de informação de Fisher. Segundo Bryk & Raudenbush (1992) em muitos casos esta aproximação normal é muito pobre, especialmente quando τ_{qq} está próximo de zero, além do que, testes baseados em intervalos de confiança simétricos para $\hat{\tau}_{qq}$ podem levar a conclusões equivocadas. Trata-se de um assunto ainda em aberto, cuja discussão pode ser encontrada em Giampaoli (1999).

Uma forma para comparar modelos com os mesmos efeitos fixos, porém com diferentes estruturas de covariâncias é utilizando o critério de informação de Akaike (AIC) (Jones 1993). Diferentemente do teste da RV, este critério pode ser utilizado quando a estrutura de covariâncias de um dos modelos não constitui uma forma reduzida do outro.

Em geral os métodos de diagnóstico visam a verificação de possíveis afastamentos das suposições feitas para o modelo, bem como a identificação da existência de observações extremas com alguma interferência desproporcional dos resultados do ajuste (Paula 2004).

O método de influência local é considerado um dos mais modernos dentro da análise de diagnóstico (Paula 2004). A idéia de influência local foi introduzida por Cook (1986) e consiste em verificar a existência de pontos que, sob pequenas modificações no modelo, causam variações “desproporcionais” nos resultados do ajuste.

O método sugerido por Cook (1986) para avaliar mudanças nos resultados da análise quando incorporamos “pequenas perturbações” ao modelo baseia-se na função denominada *afastamento da verossimilhança* (“likelihood displacement”) (Cook & Weisberg (1982), Cook (1987) e Cook et al. (1988)), a qual é definida por

$$LD(\underline{w}) = 2 \left\{ L(\hat{\underline{\psi}}) - L(\hat{\underline{\psi}}|\underline{w}) \right\}, \quad (3.1)$$

onde $L(\cdot)$ é o logaritmo da função de verossimilhança do modelo postulado (o modelo assumido correto), $\underline{\psi}$ é um vetor de parâmetros desconhecidos $p \times 1$ e \underline{w} um vetor de perturbações $q \times 1$ assumindo valores em um conjunto aberto $\Omega \subset R^q$. Denotamos por $L(\cdot|\underline{w})$ o logaritmo da função de verossimilhança do modelo perturbado. Finalmente, $\hat{\underline{\psi}}$ e $\hat{\underline{\psi}}_{\underline{w}}$ denotam, respectivamente, os

EMV's baseados em $L(\cdot)$ e $L(\cdot|\underline{w})$. Assumimos que o modelo postulado está encaixado no modelo perturbado, isto é, existe $\underline{w}_0 \in \Omega$ tal que $L(\underline{\psi}|\underline{w}_0) = L(\underline{\psi})$, $\forall \underline{\psi} \in \Theta$, e que $L(\underline{\psi}|\underline{w})$ é duas vezes diferenciável em $(\hat{\underline{\psi}}^\top, \underline{w}_0^\top)^\top$.

Informações essenciais da influência do esquema de perturbações são encontrados em um gráfico de $LD(\underline{w})$ versus \underline{w} . Esse gráfico pode ser interpretado como uma superfície em R^{q+1} formada pelos valores do vetor

$$\alpha(\underline{w}) = \begin{pmatrix} \underline{w} \\ LD(\underline{w}) \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_q \\ LD(\underline{w}) \end{pmatrix}, \quad (3.2)$$

com \underline{w} variando em Ω . Essa superfície definida acima é chamada de *gráfico de influência*.

O estudo de influência local consiste em analisar o comportamento do *gráfico de influência* na vizinhança de \underline{w}_0 . Essa análise pode ser feita através do estudo das curvaturas das seções normais (Araújo 1998) da superfície $\alpha(\underline{w})$ ao redor de \underline{w}_0 na direção de um vetor \underline{d} ($q \times 1$) de norma unitária, tal curvatura doravante será denotada por $C_{\underline{d}}$.

Para este caso, Cook (1986) mostrou que a curvatura normal na direção \underline{d} toma a forma (detalhes ver Lobato Jr. (2005), Cook (1986))

$$C_{\underline{d}} = 2 \left| \underline{d}^\top \mathbf{H}^\top \ddot{\mathbf{L}}^{-1} \mathbf{H} \underline{d} \right|, \quad (3.3)$$

em que $-\ddot{\mathbf{L}} = \partial^2 L(\underline{\psi}) / \partial \underline{\psi}^\top \partial \underline{\psi}$ é a matriz de informação de Fisher observada para o modelo ($\underline{w} = \underline{w}_0$) e $\mathbf{H} = \partial^2 L(\underline{\psi}|\underline{w}) / \partial \underline{\psi}^\top \partial \underline{w}$ avaliadas em $\underline{\psi} = \hat{\underline{\psi}}$ e $\underline{w} = \underline{w}_0$. A curvatura da normal (3.3) é essencialmente equivalente à segunda derivada de $LD(\underline{w})$ ao redor de \underline{w}_0 (Billor & Loynes 1993, Wu & Luo 1993, Araújo 1998).

Valores grandes para $C_{\underline{d}}$ indicam sensibilidade para as perturbações introduzidas na direção \underline{d} . A direção que produz maior influência local, \underline{d}_{max} , é o autovetor normalizado correspondente ao maior autovalor C_{max} da matriz $\ddot{F} = -\mathbf{H}^\top \ddot{\mathbf{L}}^{-1} \mathbf{H}$. Desta forma, \underline{d}_{max} pode ser utilizado como ferramenta útil na análise de diagnóstico, como por exemplo a partir do gráfico dos elementos de $|\underline{d}_{max}|$ contra a ordem das observações, o qual pode revelar quais observações têm a maior

influência em $LD(\underline{w})$ em torno de w_0 . Cook (1986) propõe examinar as componentes de \underline{d}_{max} , independente do valor de C_{max} , uma vez que pode indicar observações que são conjuntamente influentes.

Em algumas situações existe o interesse em avaliar a influência parcial sobre um subconjunto de $\underline{\psi} = (\underline{\psi}_1^\top, \underline{\psi}_2^\top)^\top$, e suponhamos que nosso interesse reside em $\underline{\psi}_1^\top$. Nesse caso o gráfico de influência é uma superfície

$$\alpha_s(\underline{w}) = \begin{pmatrix} \underline{w} \\ LD_s(\underline{w}) \end{pmatrix}, \quad (3.4)$$

onde

$$LD_s(\underline{w}) = 2 \left\{ L(\hat{\underline{\psi}}_1, \hat{\underline{\psi}}_2) - L[\hat{\underline{\psi}}_{1w}, g(\hat{\underline{\psi}}_{1w})] \right\}. \quad (3.5)$$

Nessa situação, $L[\underline{\psi}_1, g(\underline{\psi}_1)]$ representa o logaritmo da função de verossimilhança perfilada para $\underline{\psi}_1$ (Cordeiro 1992). A curvatura normal do gráfico de influência na direção de um vetor \underline{d} (de norma unitária) associado a (3.3) é

$$C_{\underline{d}}(\underline{\psi}_1) = 2 \left| \underline{d}^\top \mathbf{H}^\top (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \mathbf{H} \underline{d} \right|, \quad (3.6)$$

com

$$\mathbf{B}_{22} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22}^{-1} \end{bmatrix} \quad (3.7)$$

e $\ddot{\mathbf{L}}_{22}^{-1}$ é obtido da partição de $\ddot{\mathbf{L}}$, segundo a partição de $\underline{\psi}$ (Montenegro 2006).

Fung & Kwan (1997) mostraram que a curvatura normal é invariante com relação a escala quando a derivada da medida de influência avaliada no EMV é nula; desta forma, eles sugerem a aplicação dessa metodologia apenas quando a referida derivada é nula.

Neste trabalho usaremos o método tradicional de influência local avaliado através da curvatura normal, introduzido por Cook (1986), para modelos hierárquicos. No entanto, outra forma de avaliar a influência local para modelos hierárquicos foi proposta recentemente por Shi & Chen (2008), neste artigo eles utilizaram uma forma generalizada da estatística de influência proposta por Cook (1977), a qual pode ser definida como:

$$D(\underline{w}) = \frac{\|\hat{\underline{y}} - \hat{\underline{y}}_{\underline{w}}\|^2}{p\sigma^2}, \quad (3.8)$$

onde $\|\cdot\|$ denota a norma euclidiana de um vetor, $\hat{\underline{y}}$ e $\hat{\underline{y}}_{\underline{w}}$ são vetores baseados no conjunto total das observações e no conjunto dos dados com o vetor de perturbação w , respectivamente.

Existe uma relação entre $D(w)$, conhecida como distância de Cook generalizada, e $LD(w)$, e nesta relação é possível encontrar que a distância de Cook é uma aproximação do *afastamento da verossimilhança*. Então Shi & Chen (2008) encontram os autovetores correspondentes ao maior e ao segundo maior autovalores dessa aproximação, o que é útil para detecção de observações influentes.

3.1 Influência Local em Modelos Lineares Hierárquicos

Nesta seção apresentamos as propostas de Beckman et al. (1987) de influência local para modelos mistos, no contexto dos modelos hierárquicos. Sua proposta baseia-se na verossimilhança marginal de $\underline{Y}_j \sim N(\mathbf{X}_j \mathbf{Z}_j \underline{\gamma}, \mathbf{V}_j)$ sob o modelo (2.9), onde $\mathbf{V}_j = \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \sigma^2 \mathbf{I}_{n_j}$. Por conveniência, não colocamos um parâmetro de dispersão em evidência, ou seja consideramos $Var(\underline{\delta}_j) = \boldsymbol{\tau}_j(\underline{\theta})$ com $\underline{\theta}$ representando um vetor de dimensão $l \times 1$ contendo os $l \leq \binom{q}{2} + q$ parâmetros de covariância. Neste caso, definimos o vetor de parâmetros como $\underline{\psi}_j^\top = (\underline{\gamma}^\top, \sigma^2, \underline{\theta}^\top) = (\underline{\gamma}^\top, (\underline{\theta}^*)^\top)$, logo, $(\underline{\theta}^*)^\top = (\sigma^2, \underline{\theta}^\top)$. O logaritmo da função de verossimilhança a menos de um parâmetro é

$$L(\underline{\psi}_j) = \lambda_j = -(1/2) \left\{ \ln |\mathbf{V}_j| + (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma})^\top \mathbf{V}_j^{-1} (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}) \right\}. \quad (3.9)$$

Nas duas próximas subseções calculamos a respectiva matriz \mathbf{H} , para os esquemas de perturbação propostos por Beckman et al. (1987), baseado no modelo (2.9).

Perturbação na matriz de covariância do erro (nível 1)

A avaliação da sensibilidade do modelo sob suposição de homoscedasticidade pode ser feita incorporando um vetor $n_j \times 1$ de perturbações, de tal forma que $Var(\underline{\epsilon}_j) = \sigma^2 \boldsymbol{\Lambda}(w_j)$, com $\boldsymbol{\Lambda}(w_j)$ representando uma matriz $(n_j \times n_j)$ diagonal e w_{ji} denotando o i -ésimo elemento dessa diagonal. Isto corresponde a perturbar a estrutura de variância dos erros do nível 1. Neste caso, $w_0 = 1_{n_j}$ representa um vetor de dimensão $n_j \times 1$ com todos os elementos iguais a um. O logaritmo da

função de verossimilhança perturbada pode então ser escrito como

$$L(\underline{\psi}_j, \underline{w}_j) = \lambda_j(\underline{w}_j) = -(1/2) \left\{ \ln |\mathbf{V}_j(\underline{w}_j)| + (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma})^\top \mathbf{V}_j^{-1}(\underline{w}_j) (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}) \right\}, \quad (3.10)$$

com $\mathbf{V}_j(\underline{w}_j) = \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \sigma^2 \boldsymbol{\Lambda}(\underline{w}_j)$. A i -ésima coluna da matriz \mathbf{H}_j é dada por

$$\mathbf{H}_{ji} = \left\{ \left(\frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \underline{\gamma}} \right)^\top, \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \sigma^2}, \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \theta_1}, \dots, \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \theta_l} \right\}^\top. \quad (3.11)$$

Avaliando as respectivas derivadas em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ com $\hat{\underline{\psi}}_j$ representando o EMV de $\underline{\psi}_j$. Então para $i = 1, \dots, n_j$ temos

$$\left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \underline{\gamma}} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} = \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{D}_i \hat{\underline{\tau}}_j, \quad (3.12)$$

$$\left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \theta_k} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} = -\frac{1}{2} \left\{ \text{tr} \left[\mathbf{D}_i \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \right] - 2 \hat{\underline{\tau}}_j^\top \mathbf{D}_i \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\underline{\tau}}_j \right\}, \quad (3.13)$$

$$\left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \sigma^2} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} = -\frac{1}{2} \left\{ \sigma^{-2} \text{tr} \left[\mathbf{D}_i \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \right] - 2 \hat{\underline{\tau}}_j^\top \mathbf{D}_i \hat{\mathbf{V}}_j^{-1} \hat{\underline{\tau}}_j + \sigma^{-2} \hat{\underline{\tau}}_j^\top \mathbf{D}_i \hat{\underline{\tau}}_j \right\}, \quad (3.14)$$

com $\hat{\underline{\tau}}_j = \underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}$, $\mathbf{D}_i = \partial \mathbf{V}_j^{-1}(\underline{w}_j) / \partial w_{ji} |_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j}$ e $\mathbf{D}_k = \partial \boldsymbol{\tau}_j / \partial \theta_k |_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j}$. Para esse esquema de perturbação mostra-se que (Apêndice B)

$$\mathbf{D}_i = -\sigma^2 \hat{\mathbf{V}}_j^i (\hat{\mathbf{V}}_j^i)^\top \quad (i = 1, \dots, n_j), \quad (3.15)$$

com \mathbf{V}^i representando a i -ésima coluna de \mathbf{V}^{-1} . Portanto, juntamente com $-\hat{\underline{\mathbf{L}}}$, através das expressões (3.12) a (3.14) pode-se obter o máximo de (3.3) e avaliar a influência local referente ao particular esquema de perturbação.

Perturbação na variável resposta

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

$$\underline{Y}_j(\underline{w}_j) = \underline{Y}_j + s \underline{w}_j, \quad (3.16)$$

com s representando um fator de escala e \underline{w} um vetor $n_j \times 1$ de perturbações. Nesse caso $\underline{w}_0 = \underline{0}$ e o logaritmo da função de verossimilhança perturbada é dado por

$$\lambda_j(\underline{w}_j) = -(1/2) \left((\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma})^\top \mathbf{V}_j^{-1} (\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}) \right). \quad (3.17)$$

Considerando (3.17) tem-se (ver Apêndice B)

$$\mathbf{H}_j^\top = s \hat{\mathbf{V}}_j^{-1} \left\{ \mathbf{X}_j \mathbf{Z}_j, \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j, \mathbf{X}_j \mathbf{D}_1 \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j, \dots, \mathbf{X}_j \mathbf{D}_l \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j \right\}. \quad (3.18)$$

O esquema de perturbação acima tem forte conexão com o conceito de alavancagem em modelos não lineares (St. Laurent & Cook 1993, Wei et al. 1998).

Perturbação na matriz de covariância dos efeitos aleatórios (nível 2)

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 $\boldsymbol{\gamma}_j$ da seguinte forma: $Var(\hat{\boldsymbol{\delta}}_j) = w_j \boldsymbol{\tau}_j$. Neste caso, o vetor de perturbações \boldsymbol{w} é de dimensão $J \times 1$ e $\boldsymbol{w}_0 = \mathbf{1}_J$. Considerando o modelo perturbado, tem-se que $\mathbf{V}_j(w_j) = w_j \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \sigma^2 \mathbf{I}_{n_j}$, com o logaritmo da função de verossimilhança perturbada sendo dado por

$$\lambda(\boldsymbol{w}) = \sum_{j=1}^J \lambda_j(w_j) = \sum_{j=1}^J (-1/2) \left\{ \ln |\mathbf{V}_j(w_j)| + (\boldsymbol{Y}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma})^\top \mathbf{V}_j^{-1}(w_j) (\boldsymbol{Y}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma}) \right\}. \quad (3.19)$$

A j -ésima coluna da matriz \mathbf{H} é dada por

$$\mathbf{H}_j = \left\{ \left(\frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \boldsymbol{\gamma}} \right)^\top, \frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \sigma^2}, \frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \theta_1}, \dots, \frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \theta_l} \right\}^\top, \quad (3.20)$$

com as respectivas derivadas calculadas em $\boldsymbol{w} = \boldsymbol{w}_0$ e $\boldsymbol{\psi}_j = \hat{\boldsymbol{\psi}}_j$. Para $j = 1, \dots, J$ e $k = 1, \dots, l$ (veja Apêndice B), temos

$$\left. \frac{\partial^2 \lambda_j(\boldsymbol{w})}{\partial w_j \partial \boldsymbol{\gamma}} \right|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}_j=\hat{\boldsymbol{\psi}}_j} = -\mathbf{Z}_j^\top \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j, \quad (3.21)$$

$$\begin{aligned} \left. \frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \sigma^2} \right|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}_j=\hat{\boldsymbol{\psi}}_j} &= \frac{1}{2} \text{tr} \left\{ \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \right\} \\ &- \hat{\boldsymbol{\tau}}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j, \end{aligned} \quad (3.22)$$

$$\begin{aligned} \left. \frac{\partial^2 \lambda(\boldsymbol{w})}{\partial w_j \partial \theta_k} \right|_{\boldsymbol{w}=\boldsymbol{w}_0; \boldsymbol{\psi}_j=\hat{\boldsymbol{\psi}}_j} &= -\frac{1}{2} \text{tr} \left\{ \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \right\} \\ &+ \frac{1}{2} \text{tr} \left\{ \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2} \hat{\boldsymbol{\ell}}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\ell}}_j \\
& - \hat{\boldsymbol{\ell}}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\ell}}_j.
\end{aligned} \tag{3.23}$$

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 (3.9) avaliada em $\hat{\boldsymbol{\psi}}_j$ não é nula. Tal preocupação deve-se ao fato de que estimativas negativas das componentes de variância não correspondem aos EMV's das mesmas.

Aplicação do MLH na Área de Educação

Este capítulo apresenta uma aplicação do MLH na área de Educação. Foram utilizados os dados do SAEPE (Sistema de Avaliação Educacional de Pernambuco) do ano 2005, gentilmente fornecidos pela Secretaria de Educação e Cultura do Estado de Pernambuco.

Alguns trabalhos aplicando MLH na área de Educação também foram feitos anteriormente como o realizado por Natis (2000) que aplicou MLH aos dados do SARESP (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo), do ano de 1997, estudando a habilidade (obtida a partir da TRI) em Língua Portuguesa dos alunos da 4^a série de escolas públicas do Estado de São Paulo, como também o estudo feito por Pinheiro (2005) que aplicou MLH aos dados do SAEPE 2002 para analisar o desempenho dos alunos da 4^a e 8^a série do ensino fundamental e da 3^a série do ensino médio através das notas (percentual de acertos) que eles obtiveram nas provas de português e matemática.

Neste trabalho estudamos as habilidades em Língua Portuguesa e em Matemática dos alunos da 3^a série do ensino médio de escolas públicas do estado de Pernambuco, segundo características dos alunos e das escolas que eles pertenciam. A habilidade ou proficiência é uma variável contínua obtida a partir da teoria da resposta ao item (TRI) (para maiores detalhes sobre a TRI, veja Andrade et al. (2000)). Através da TRI, os escores dos alunos nas provas são convertidos para uma outra escala denominada habilidade, de maneira que dois alunos com o mesmo escore podem

não ter a mesma habilidade, uma vez que os graus de dificuldade, que representa a proporção de alunos que acertou um determinado item, e de discriminação, que indica a capacidade do item discriminar entre alunos com alta habilidade e alunos com baixa habilidade, de cada questão da prova são levados em conta. As análises aqui apresentadas trazem apenas algumas variáveis que foram efetivamente avaliadas no SAEPE.

O SAEPE foi implantado em Pernambuco no ano 2000 com objetivo de subsidiar uma estratégia de monitoria e de incentivos permanentes, centrados na melhoria da qualidade e do desempenho do ensino básico no estado. O procedimento consistiu da aplicação de uma prova do domínio da língua (leitura e escrita) aos alunos da 2^a série do ensino fundamental, e de duas provas (português e matemática) aos alunos da 4^a e 8^a série do ensino fundamental e da 3^a série do ensino médio. Além do desempenho escolar o SAEPE avalia as condições de funcionamento da escola, o nível de eficiência das escolas na promoção de seus alunos, o perfil da direção das unidades escolares, as modalidades de gestão escolar, o perfil dos professores, os recursos pedagógicos utilizados em sala de aula e as características sócio-culturais dos alunos (Secretaria de Educação e Cultura de Pernambuco 2003).

4.1 Análise Descritiva

No ano de 2005, foram 48789 alunos da 3^a série do ensino médio distribuídos em 7142 escolas públicas do estado de Pernambuco que fizeram as provas de português e matemática. No entanto, para o presente estudo foram excluídas as escolas em que menos de 30 alunos participaram da avaliação por dificultarem o processo de convergência das estimativas dos parâmetros na análise do modelo linear hierárquico. Após esse procedimento as escolas que ficaram passaram aqui a serem consideradas população, então a partir daí foi selecionada uma amostra de 10% das escolas, visto que a realização de testes estatísticos em amostras muito grandes conduzem sempre à não-rejeição da hipótese nula.

O estudo descritivo das habilidades foi realizado tanto para população quanto para a amostra, com o intuito de verificar a representatividade desta. Para se ter uma melhor visão da distribuição das habilidades estudadas foram construídos os histogramas (Figuras 4.1 e 4.2). Algumas

medidas estatísticas de interesse são apresentadas na Tabela 4.1. Vale ressaltar que a escala da habilidade varia de $-4,0$ a $+4,0$. Os histogramas dos valores das habilidades ilustram claramente a assimetria das duas distribuições analisadas, sendo em matemática mais acentuada e, quanto às médias, os valores são muito baixos e os dados mostram que a média da habilidade em português foi maior que a de matemática.

Tabela 4.1: Estatísticas descritivas das habilidades de português e matemática.

		número de escolas	número de alunos	média	desvio padrão
matemática	População	555	45634	$-0,162$	$1,425$
	Amostra	55	4333	$-0,148$	$1,419$
português	População	555	45629	$-0,115$	$1,366$
	Amostra	55	4333	$-0,082$	$1,330$

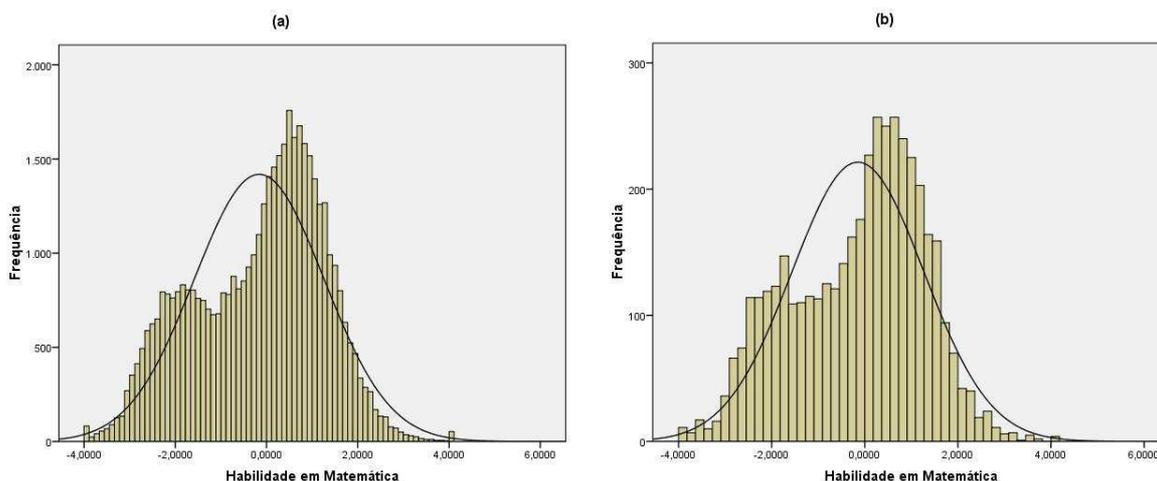


Figura 4.1: Histograma das habilidades em matemática: (a) População, (b) Amostra.

Para a modelagem estatística, observando estudos anteriores, diversas variáveis disponíveis no banco de dados no nível do aluno foram destacadas como possíveis candidatas a explicar as habilidades. As variáveis selecionadas para o primeiro nível foram:

- Gênero do aluno, definida por

$$\text{gênero} = \begin{cases} 1, & \text{se masculino,} \\ 0, & \text{se feminino;} \end{cases} \quad (4.1)$$

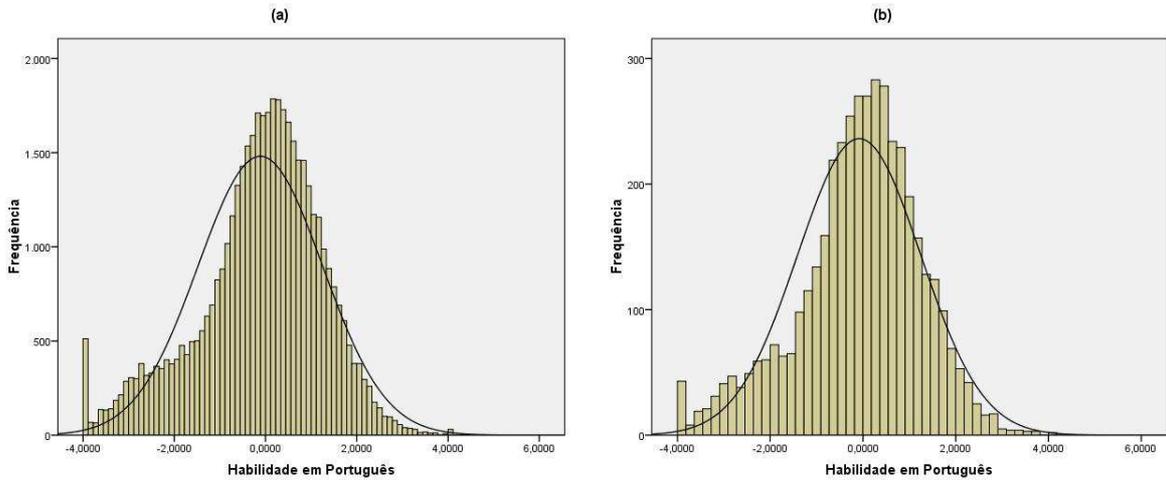


Figura 4.2: Histograma das habilidades em português: (a) População, (b) Amostra.

- Se o aluno trabalha fora de casa, definida por

$$\text{trabalha} = \begin{cases} 1, & \text{se não trabalha fora de casa,} \\ 0, & \text{se trabalha fora de casa;} \end{cases} \quad (4.2)$$

- Quantas horas, por dia e fora da sala de aula, o aluno se dedica às atividades escolares, definida por

$$\text{horaest} = \begin{cases} 0, & \text{se nenhuma,} \\ 1, & \text{se no máximo duas horas,} \\ 2, & \text{se mais de duas horas} \end{cases} \quad (4.3)$$

- Se o aluno já repetiu de ano, definida por,

$$\text{repetiu} = \begin{cases} 1, & \text{se no máximo uma vez,} \\ 0, & \text{se duas ou mais vezes;} \end{cases} \quad (4.4)$$

- Índice Sócio-cultural (isc): variável contínua obtida a partir das normas da ABIPEME (Associação Brasileira de Institutos de Pesquisa de Mercado), que faz uma classificação sócio-econômica a partir de variáveis como a posse de televisão, geladeira, máquina de lavar, automóvel, entre outras e grau de instrução do chefe de família. Para maiores detalhes acerca dos critérios de classificação econômica ver Associação Brasileira de Empresas de Pesquisa (2008).

Para a escolha das variáveis do segundo nível, além do tipo de escola: estadual ou municipal, foram construídos indicadores que qualificam o estado de conservação da escola e das salas. A

construção desses indicadores baseou-se em algumas variáveis ordinais do questionário das escolas considerando dois tipos de escala, a primeira desde: A [bom (valor = 4)], B [regular (valor = 3)], C [ruim (valor = 2)] até: D [não se aplica/não existe (valor = 1)]; e a segunda desde: A (bom) ou B (regular) [valor = 1] até: C (ruim) ou D (não se aplica/não existe) [valor = 0].

Foram definidos quatro indicadores que qualificam o estado de conservação das escolas e as instalações das salas de aula. Para a conservação da escola foram definidos dois indicadores, um para cada escala considerada, correspondentes à soma das respostas referentes a conservação do: telhado, alvenaria (paredes), piso, esquadilhas (portas e janelas), instalações hidráulica e elétrica, pintura, muros (fechamentos) e vidros. Ou seja,

$$ICE_{ij} = \sum_{k=1}^K t_k, \quad (4.5)$$

onde t_k é o valor referente a k -ésima questão considerada para a construção do índice i para a escola j , com $i = 1, 2$, referentes à primeira e segunda escalas, respectivamente, $j = 1, \dots, J$, referentes às escolas e K o número de questões consideradas para o índice.

Da mesma forma, para caracterizar a situação das condições de funcionamento das salas também foram construídos dois indicadores, somadas as respostas das perguntas em relação aos aspectos: iluminação natural e artificial, ventilação, quadro de giz ou branco, bancas de alunos e mesa do professor, na mesma escala do estado de conservação do prédio. Ou seja,

$$IIS_{ij} = \sum_{k=1}^K t_k, \quad (4.6)$$

onde t_k é o valor referente a k -ésima questão considerada para a construção do índice i para a escola j , com $i = 1, 2$, referentes à primeira e segunda escalas, respectivamente, $j = 1, \dots, J$, referentes às escolas e K o número de questões consideradas para o índice.

Dessa forma as variáveis consideradas no nível da escola são:

- Tipo de escola, definida por

$$\text{tipoesc} = \begin{cases} 1, & \text{se estadual,} \\ 0, & \text{se municipal;} \end{cases} \quad (4.7)$$

- Conservação da Escola 1, definida por

$$\text{conservesc1} = \begin{cases} 1, & \text{para os valores do } ICE_{1j} \text{ acima da mediana,} \\ 0, & \text{para os valores do } ICE_{1j} \text{ abaixo da mediana;} \end{cases} \quad (4.8)$$

- Conservação da Escola 2, definida por

$$\text{conservesc2} = \begin{cases} 1, & \text{para os valores do } ICE_{2j} \text{ acima da mediana,} \\ 0, & \text{para os valores do } ICE_{2j} \text{ abaixo da mediana;} \end{cases} \quad (4.9)$$

- Instalações da Sala 1, definida por

$$\text{conservsala1Q1} = \begin{cases} 1, & \text{para os valores do } IIS_{1j} \text{ acima do 1}^\circ \text{ quartil,} \\ 0, & \text{para os valores do } IIS_{1j} \text{ abaixo do 1}^\circ \text{ quartil;} \end{cases} \quad (4.10)$$

- Instalações da Sala 2, definida por

$$\text{conservsala2Q1} = \begin{cases} 1, & \text{para os valores do } IIS_{2j} \text{ acima do 1}^\circ \text{ quartil,} \\ 0, & \text{para os valores do } IIS_{2j} \text{ abaixo do 1}^\circ \text{ quartil;} \end{cases} \quad (4.11)$$

- Instalações da Sala 3, definida por

$$\text{conservsala2Q2} = \begin{cases} 1, & \text{para os valores do } IIS_{2j} \text{ acima da mediana,} \\ 0, & \text{para os valores do } IIS_{2j} \text{ abaixo da mediana.} \end{cases} \quad (4.12)$$

Um resumo de algumas medidas estatísticas das habilidades em matemática e português em relação às variáveis candidatas a serem explicativas estão apresentadas nas Tabelas 4.2 e 4.3, juntamente com os p -valores dos testes de Mann-Whitney ou Kruskal-Wallis para as variáveis consideradas, com o intuito de averiguar alguma possível diferença estatística nas habilidades, entre os grupos formados por essas variáveis.

Tabela 4.2: Estatísticas descritivas e p -valores dos testes de igualdade de médias para a Habilidade em Português.

variável	categoria	número de alunos	média	desvio padrão	p -valor
gênero	feminino	2676	0,012	1,272	0,000 ⁽¹⁾
	masculino	1528	-0,188	1,374	
trabalha	sim	1562	-0,222	1,313	0,000 ⁽¹⁾
	não	2494	0,065	1,288	
repetiu	no maximo uma vez	3310	0,021	1,321	0,000 ⁽¹⁾
	duas ou mais	794	-0,364	1,217	
horaest	nenhuma	547	-0,389	1,321	0,000 ⁽²⁾
	no maximo duas	2555	-0,030	1,246	
	mais que duas	984	0,113	1,401	
tipoescc	estatual	3762	-0,074	1,343	0,256 ⁽¹⁾
	municipal	571	-0,131	1,243	
conservsala1Q1	$IIS_{1j} \leq Q1$	1954	-0,141	1,344	0,002 ⁽¹⁾
	$IIS_{1j} > Q1$	2278	-0,016	1,317	
conservesc1	$ICE_{1j} \leq Q2$	2492	-0,132	1,338	0,000 ⁽¹⁾
	$ICE_{1j} > Q2$	1740	0,010	1,316	
conservesc2	$ICE_{2j} < Q2$	1954	-0,065	1,340	0,907 ⁽¹⁾
	$ICE_{2j} \geq Q2$	2200	-0,071	1,319	
conservsala2Q1	$IIS_{2j} \leq Q1$	1976	-0,155	1,344	0,000 ⁽¹⁾
	$IIS_{2j} > Q1$	2138	-0,002	1,316	
conservsala2Q2	$IIS_{2j} < Q2$	3052	-0,154	1,336	0,000 ⁽¹⁾
	$IIS_{2j} \geq Q2$	1062	0,149	1,293	

(1) Teste de Mann-Whitney (2) Teste de Kruskal-Wallis

$Q1$ e $Q2$ representam, respectivamente, o primeiro quartil e a mediana

Os testes de comparação de médias revelaram que as diferenças são estatisticamente significativas. Pode-se concluir, no nível do aluno, que:

- O desempenho médio dos alunos do gênero masculino em português foi inferior aos alunos do gênero feminino. Contudo, em matemática, a habilidade média dos homens foi superior a das mulheres;
- A habilidade média em português dos alunos que não trabalham fora de casa é maior do que a dos que trabalham;
- Os estudantes que dedicam mais tempo aos estudos possuem uma habilidade média maior, tanto em português quanto em matemática;
- Os alunos que repetiram de ano no máximo uma vez apresentam, em média, maior habilidade em português.

Tabela 4.3: Estatísticas descritivas e p -valores dos testes de igualdade de médias para a Habilidade em Matemática.

variável	categoria	número de alunos	média	desvio padrão	p -valor
gênero	feminino	2676	-0,182	1,412	0,012 ⁽¹⁾
	masculino	1528	-0,065	1,423	
trabalha	sim	1562	-0,168	1,399	0,239 ⁽¹⁾
	não	2494	-0,128	1,428	
repetiu	no maximo uma vez	3310	-0,134	1,420	0,275 ⁽¹⁾
	duas ou mais	794	-0,196	1,403	
horaest	nenhuma	547	-0,271	1,470	0,032 ⁽²⁾
	no maximo duas	2555	-0,147	1,405	
	mais que duas	984	-0,079	1,417	
tipoescc	estadual	3762	-0,130	1,418	0,025 ⁽¹⁾
	municipal	571	-0,266	1,419	
conservsala1Q1	$IIS_{1j} \leq Q1$	1954	-0,191	1,417	0,035 ⁽¹⁾
	$IIS_{1j} > Q1$	2278	-0,109	1,425	
conservesc1	$ICE_{1j} \leq Q2$	2492	-0,222	1,405	0,000 ⁽¹⁾
	$ICE_{1j} > Q2$	1740	-0,040	1,439	
conservesc2	$ICE_{2j} < Q2$	1954	-0,213	1,409	0,001 ⁽¹⁾
	$ICE_{2j} \geq Q2$	2200	-0,083	1,430	
conservsala2Q1	$IIS_{2j} \leq Q1$	1976	-0,176	1,400	0,153 ⁽¹⁾
	$IIS_{2j} > Q1$	2138	-0,120	1,439	
conservsala2Q2	$IIS_{2j} < Q2$	3052	-0,190	1,418	0,000 ⁽¹⁾
	$IIS_{2j} \geq Q2$	1062	-0,022	1,421	

(1) Teste de Mann-Whitney (2) Teste de Kruskal-Wallis

$Q1$ e $Q2$ representam, respectivamente, o primeiro quartil e a mediana

Para o nível da escola, os testes permitem concluir que:

- A habilidade em matemática dos alunos das escolas estaduais é, em média, maior que a dos alunos das escolas municipais;
- Tanto para matemática quanto para português, podemos observar os alunos das escolas cujo IIS_{1j} é maior que o primeiro quartil apresentam maiores habilidades, em média. Quando consideramos o IIS_{2j} , apenas para português os alunos das escolas cujo IIS_{2j} é maior que o primeiro quartil apresentam habilidades maiores, em média;
- Ao compararmos o IIS_{2j} com a mediana, observamos que, tanto em português quanto em matemática, os alunos das escolas cujo IIS_{2j} é maior que a mediana apresentam, em média, maiores habilidades;
- Com relação ao estado de conservação da escola, os alunos de escolas cujo ICE_{1j} é maior

que a mediana apresentam, em média, habilidades maiores, tanto em português quanto em matemática. No entanto, os alunos de escolas cujo ICE_{2j} é maior que a mediana apresentam, em média, habilidades maiores somente para matemática.

4.2 Especificação do Modelo

Na análise descrita a seguir foram considerados, inicialmente, modelos apenas com o intercepto para verificar se o mesmo poderia ser considerado como aleatório verificando o p -valor. Em seguida, foram incluídas variáveis explicativas apenas no nível 1 para verificar a significância das mesmas, sendo retiradas, uma por vez, aquelas que não apresentaram significância. Por fim, as variáveis significativas no nível 1 foram mantidas no modelo e aos parâmetros que poderiam ser considerados como aleatórios foram acrescentadas, uma após outra, variáveis explicativas no nível 2, mantendo-se aquelas com nível de significância abaixo de 10%. Devido a dados faltantes, apenas 51 escolas permaneceram nos modelos.

O modelo selecionado para explicar as habilidades em matemática dos alunos da 3ª série do ensino médio foi:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(gênero_{ij}) + \beta_{2j}(horaest_{ij}) + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(consevesc1_j) + \delta_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

Através do resultado do teste para as componentes de variância os parâmetros β_{1j} e β_{2j} foram considerados como fixos. A Tabela 4.4 mostra os resultados das estimativas do modelo hierárquico apresentado acima, o qual obteve o melhor ajuste para as habilidades em matemática dos alunos da 3ª série do ensino médio.

Para alunos do gênero feminino, com nenhuma hora de estudo em casa, que estudam em escolas com nível de conservação abaixo da mediana, a habilidade média em matemática é de $-0,4489$. Nas mesmas condições, esta média é aumentada de $0,135$ para o gênero masculino, aumentada de $0,094$ para os que estudam no máximo duas horas por dia e de $0,188$ no caso de estudar mais de duas horas. Por outro lado, se o estado de conservação da escola é acima da

mediana, a habilidade média em matemática tem um acréscimo de 1,1646.

Observando os resultados apresentados na Tabela 4.4 constatamos que, para a modelagem da habilidade em matemática o desvio padrão das notas dentro da escola é de 1,399, enquanto que o desvio padrão dos interceptos é de 0,2197.

Tabela 4.4: Parâmetros estimados do modelo final para as habilidades de matemática e níveis descritivos.

Efeitos fixos	estimativas	Erro padrão	<i>p</i> -valor
γ_{00}	-0,4489	0,0984	0,0000
γ_{01}	0,1646	0,0796	0,0441
γ_{10}	0,1350	0,0484	0,0053
γ_{20}	0,0940	0,0385	0,0147
Efeitos aleatórios	desvio padrão	IC95%	
δ_{0j}	0,2197	(0,1581; 0,3052)	
ϵ_{ij}	1,3999	(1,3686; 1,4319)	

O modelo selecionado para explicar as habilidades em português dos alunos da 3ª série do ensino médio foi:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{repetiu}_{ij}) + \beta_{2j}(\text{gênero}_{ij}) + \beta_{3j}(\text{horaest}_{ij}) + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{conservsala2Q1}_j) + \delta_{0j}$$

$$\beta_{1j} = \gamma_{10} + \delta_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

Através do resultado do teste para as componentes de variância os parâmetros β_{2j} e β_{3j} foram considerados como fixos. As variáveis *gênero* e *horaest* também foram selecionadas para a modelagem das habilidades em português. O resultado das estimativas para este modelo hierárquico, que melhor se ajustou aos dados estão apresentadas na Tabela 4.5.

Para alunos que repetiram duas ou mais vezes, do gênero feminino, com nenhuma hora de estudo em casa, que estudam em salas com nível de conservação abaixo do primeiro quartil, a habilidade média em português é de -0,7577. Mantidas as condições, esta média é aumentada de 0,370 para alunos que repetiram de ano no máximo uma vez, diminuída de 0,1145 para alunos do sexo masculino, aumentada de 0,1779 para os que estudam no máximo duas horas por dia e de 0,3558 no caso de estudar mais de duas horas. Por outro lado, se a condição de

funcionamento da sala é acima do primeiro quartil, a habilidade média em português tem um acréscimo de 0,1554.

Tabela 4.5: Parâmetros estimados do modelo final para as habilidades de português e níveis descritivos.

Efeitos fixos	estimativas	Erro padrão	p -valor
γ_{00}	-0,7577	0,1033	0,0000
γ_{01}	0,1554	0,0818	0,0635
γ_{10}	0,3704	0,0661	0,0000
γ_{20}	-0,1145	0,0437	0,0089
γ_{30}	0,1779	0,0346	0,0000
Efeitos aleatórios	desvio padrão	IC95%	
δ_{0j}	0,1617	(0,0779; 0,3355)	
δ_{1j}	0,2733	(0,1712; 0,4360)	
ϵ_{ij}	1,2518	(1,2237; 1,2806)	

Observando os resultados apresentados constatamos que, para o modelagem das habilidades em português, o desvio padrão das notas dentro da escola é de 1,2518, enquanto que o desvio padrão dos interceptos é de 0,1617 e dos que repetiram é de 0,2733.

4.3 Análise de Influência Local para os Modelos Ajustados

Nesta seção consideramos todos os tipos de perturbação discutidos no Capítulo 3, sendo as Figuras 4.3 e 4.4 referentes aos modelos para matemática e português, respectivamente. Em ambas as figuras, a parte (a) corresponde ao gráfico dos elementos $|d_{max}|$ versus observações (ou alunos) quando perturbamos a matriz de covariâncias do erro. A parte (b) das figuras corresponde ao gráfico dos elementos $|d_{max}|$ versus observações (ou alunos) quando perturbamos a variável resposta. Por fim, a parte (c) das figuras corresponde aos gráficos dos elementos $|d_{max}|$ versus as escolas quando perturbamos a matriz de covariâncias dos efeitos aleatórios.

Pela Figura 4.3 podemos concluir, como um todo, que as escolas #16 e #38 são as mais influentes no modelo ajustado para as habilidades em matemática, principalmente no que tange às estimativas dos parâmetros de covariância. Considerando o par (i, j) , onde i representa a escola e j o aluno, podemos notar pela Figura 4.3(a) que as observações mais sensíveis a pequenas perturbações na matriz de covariâncias do erros são $(8, 23)$, $(8, 25)$, $(8, 36)$, $(16, 72)$, $(38, 28)$, $(38, 60)$ que correspondem às observações que apresentam, em módulo, o maior erro. Através da Figura

4.3(b) também podemos concluir que não existem observações sensíveis a pequenas perturbações na variável resposta. Por fim, a Figura 4.3(c) evidencia que as escolas #16 e #38 são mais influentes à suposição de homogeneidade entre as matrizes de covariância dos efeitos aleatórios; como neste exemplo, $\sqrt{\widehat{\tau}_{00}} = 0,2197$, essas escolas devem apresentar uma alta influência na estimação deste valor.

As escolas mais influentes no modelo ajustado para as habilidades em português, pela Figura 4.4, são as #8 e #40, principalmente no que concerne às estimativas dos parâmetros de covariância. Através da Figura 4.4(a), concluímos que as observações mais sensíveis a pequenas perturbações na matriz de covariâncias do erros são $(8, 8)$, $(13, 69)$, $(28, 122)$, $(36, 1)$, $(40, 79)$, $(40, 81)$, $(40, 111)$, $(40, 116)$ que correspondem às observações que apresentam, em módulo, o maior erro. Como em matemática, observamos que não existem observações sensíveis a pequenas perturbações na variável resposta, como pode ser observado na Figura 4.4(b). As escolas #8 e #40 são mais influentes à suposição de homogeneidade entre as matrizes de covariância dos efeitos aleatórios, como evidencia a parte (c) da Figura 4.4.

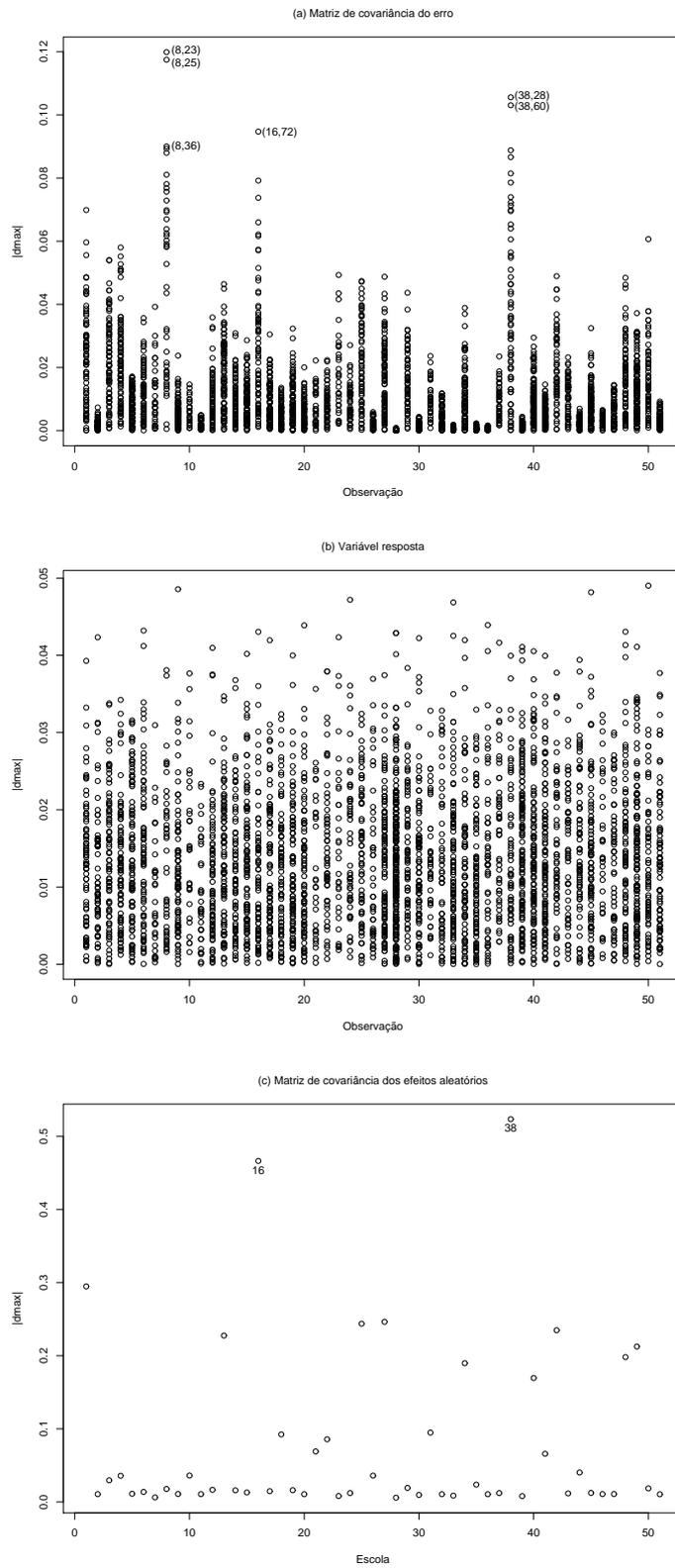


Figura 4.3: Gráficos de influência local para o modelo ajustado para Matemática.

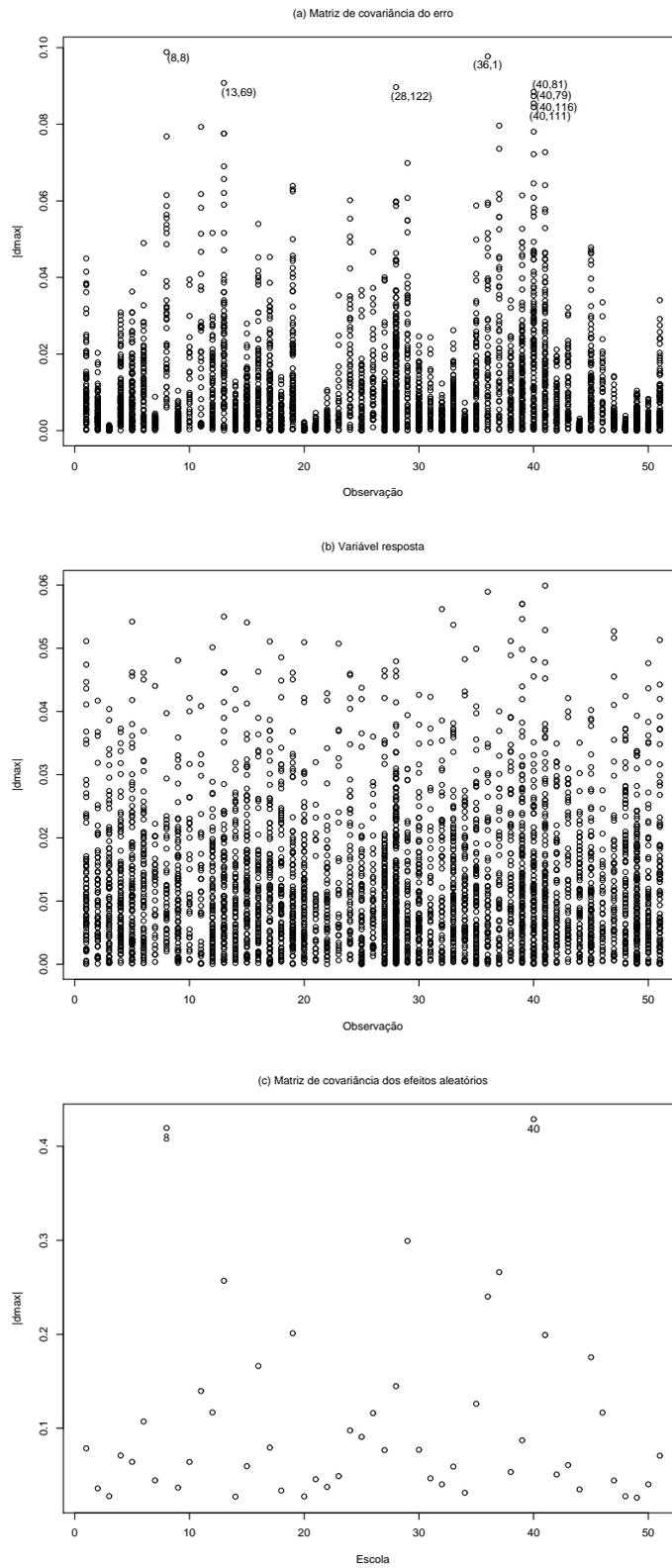


Figura 4.4: Gráficos de influência local para o modelo ajustado para Português.

Para avaliar a influência dessas escolas nas estimativas dos parâmetros de covariância dos efeitos aleatórios, foram ajustados modelos excluindo essas escolas individualmente e conjuntamente e observados os valores dos AIC e das estimativas dos desvios-padrão. Observando a Tabela 4.6, podemos concluir que, para matemática, essas escolas de fato influenciam essas estimativas, visto que a retirada dessas escolas ocasiona uma diminuição tanto nos valores dos AIC quanto nas estimativas dos desvios padrão. Para português, como podemos notar inspecionando a Tabela 4.7, há também uma diminuição nos valores dos AIC e dos desvios padrão dos interceptos ($\sqrt{\hat{\tau}_{00}}$), entretanto, não há motivos para acreditarmos que essas escolas influenciam as estimativas dos desvios-padrão das inclinações observando as alterações em $\sqrt{\hat{\tau}_{11}}$.

Uma análise semelhante foi realizada para avaliar a influência dos alunos na estimativa da matriz de covariâncias dos erros, no entanto, não se pôde observar diferenças significativas nessas estimativas, embora a análise gráfica indicasse a existência de uma possível influência. Acreditamos que isto se deva ao fato da existência de muitas observações.

Tabela 4.6: AIC, $\sqrt{\hat{\tau}_{00}}$ e $IC95\%$ para $\sqrt{\hat{\tau}_{00}}$ para o modelo de matemática completo e para os modelos retirando as escolas mais influentes.

Modelos de matemática	AIC	$\sqrt{\hat{\tau}_{00}}$	$IC95\%$
todas as escolas	13440	0,2197	(0,1581; 0,3052)
retirando a escola 16	13179	0,2050	(0,1437; 0,2923)
retirando a escola 38	13206	0,1964	(0,1366; 0,2826)
retirando as escola 16 e 38	12944	0,1756	(0,1165; 0,2649)

Tabela 4.7: AIC, $\sqrt{\hat{\tau}_{00}}$ e $\sqrt{\hat{\tau}_{11}}$ com seus respectivos $IC95\%$ para o modelo de português completo e para os modelos retirando as escolas mais influentes.

Modelos de português	AIC	$\sqrt{\hat{\tau}_{00}}$	$IC95\%$	$\sqrt{\hat{\tau}_{11}}$	$IC95\%$
todas as escolas	12634	0,1617	(0,0779; 0,3355)	0,2733	(0,1712; 0,4360)
retirando a escola 8	12495	0,1584	(0,0802; 0,3127)	0,2633	(0,1766; 0,3926)
retirando a escola 40	12253	0,1505	(0,0584; 0,3875)	0,2843	(0,1685; 0,4797)
retirando as escola 8 e 40	12114	0,1480	(0,0207; 1,0572)	0,2733	(0,0850; 0,8786)

Analisando as médias, com o intuito de averiguar a existência de alguma discrepância entre as escolas influentes e as demais escolas, concluímos que os alunos das escolas #8, #16 e #38 apresentam uma habilidade média em matemática bem acima da média global (0,8124; 0,5090; 0,7274, respectivamente). Da mesma forma, nas escolas mais influentes para o modelo ajustado

para português (#8 e #40), os alunos apresentam uma habilidade média bem acima da média global nesta disciplina (0,8613 e 0,6123, respectivamente).Analisando os dados observamos que todas as escolas consideradas influentes são estaduais e localizadas em diferentes regiões do estado.

5.1 Conclusões

Os modelos lineares hierárquicos constituem uma nova formulação para os modelos de efeitos aleatórios que permitem uma descrição e análise mais apropriada das diferentes fontes de variação. Neste trabalho fica claro como o tratamento dos dados em dois níveis facilita a construção do modelo desejado e a interpretação dos seus parâmetros.

Um aspecto importante neste trabalho foi a proposta de avaliar, a partir da influência local, o modelo ajustado. A avaliação da influência das observações no modelo ajustado é importante, visto que pontos muito influentes podem causar variações desproporcionais nos resultados do ajuste.

Uma aplicação para avaliar o desempenho escolar, utilizando o banco de dados do SAEPE 2005, ilustra o quanto é adequada a utilização de uma análise multinível. E a análise de influência feita nesta aplicação mostra escolas que indicam sensibilidade para o modelo ajustado.

5.2 Direções para Pesquisas Futuras

Neste trabalho, apresentamos uma técnica de diagnóstico para modelos lineares hierárquicos. Para pesquisas futuras, sabido que pouco foi desenvolvido para este modelo, existem muitos tópicos sobre o tema a serem explorados, como: análise de resíduos, gráfico de variáveis adici-

onadas para efeitos fixos e aleatórios e estudar valores que sirvam de referência para avaliar a sensibilidade para este enfoque. Além disso, torna-se importante o desenvolvimento de rotinas computacionais que implementem as técnicas desenvolvidas para este tipo de modelagem.

5.3 Plataforma Computacional

O ambiente de programação, análise de dados e gráficos R (R Development Core Team 2008) e o sistema tipográfico L^AT_EX (Lamport 1994) constituem a plataforma computacional utilizada no desenvolvimento desta dissertação.

Todos os resultados numéricos e todos gráficos apresentados neste trabalho foram obtidos utilizando as bibliotecas `lme4` e `nlme` do ambiente de programação, análise de dados e gráficos R em sua versão 2.7.0 para sistema operacional Microsoft Windows, que se encontra disponível gratuitamente através do site <http://www.R-project.org>.

O R é de uma linguagem caracterizada pelo compromisso entre a flexibilidade oferecida por algumas linguagens compiladas (como C, C++ e FORTRAN) e a conveniência dos tradicionais pacotes estatísticos. Foi criada por Ross Ihaka e Robert Gentleman na Universidade de Auckland com o objetivo de produzir um ambiente de programação parecido com S, uma linguagem desenvolvida no AT&T Bell Laboratories, mas sem sofrer dos mesmos problemas de demanda de memória e desempenho. Maiores detalhes sobre esta linguagem de programação podem ser encontrados em Cribari Neto & Zarkos (1999).

Neste apêndice apresentamos alguns desenvolvimentos das expressões expostas no Capítulo 2.

A.1 BLUE

Considerando a equação (2.23), obtemos as seguintes expressões

$$\begin{aligned}\mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j &= -\boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} + \mathbf{C}_j \hat{\boldsymbol{\beta}}_j \\ \hat{\boldsymbol{\beta}}_j &= \mathbf{C}_j^{-1} \left(\mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \right)\end{aligned}\quad (\text{A.1})$$

e

$$\begin{aligned}0 &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} - \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \hat{\boldsymbol{\beta}}_j \\ \left(\sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j \right) \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \hat{\boldsymbol{\beta}}_j.\end{aligned}\quad (\text{A.2})$$

Substituindo (A.1) em (A.2), temos

$$\begin{aligned}\sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \left(\mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \right) \\ \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\ \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} - \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j \\ \left(\sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{Z}_j - \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \mathbf{Z}_j \right) \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j\end{aligned}$$

$$\begin{aligned}
\left[\sum \mathbf{Z}_j^\top \left(\boldsymbol{\tau}^{-1} - \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \right) \mathbf{Z}_j \right] \hat{\boldsymbol{\gamma}} &= \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{\mathbf{Y}}_j \\
\hat{\boldsymbol{\gamma}} &= \left[\sum \mathbf{Z}_j^\top \left(\boldsymbol{\tau}^{-1} - \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \right) \mathbf{Z}_j \right]^{-1} \times \\
&\quad \sum \mathbf{Z}_j^\top \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{\mathbf{Y}}_j,
\end{aligned} \tag{A.3}$$

que é o resultado (2.24).

Como $\mathbf{C}_j = \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j + \boldsymbol{\tau}^{-1}$, temos que

$$\begin{aligned}
\boldsymbol{\tau}^{-1} - \boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} &= \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) - \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \mathbf{C}_j^{-1} \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \\
&= \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) - \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \left(\mathbf{I} - \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \\
&= \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \left[\mathbf{I} - \left(\mathbf{I} - \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \right] \\
&= \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \left(\mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \\
&= \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \\
&= \mathbf{X}_j^\top \left(\mathbf{R}_j^{-1} - \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \right) \mathbf{X}_j \\
&= \mathbf{X}_j^\top \mathbf{M}_j \mathbf{X}_j
\end{aligned} \tag{A.4}$$

e

$$\begin{aligned}
\boldsymbol{\tau}^{-1} \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} &= \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \\
&= \mathbf{X}_j^\top \mathbf{R}_j^{-1} - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \\
&= \mathbf{X}_j^\top \left(\mathbf{R}_j^{-1} - \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \right) \\
&= \mathbf{X}_j^\top \mathbf{M}_j.
\end{aligned} \tag{A.5}$$

Substituindo (A.4) e (A.5) em (A.3) obtemos,

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= \left(\sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j \right)^{-1} \sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \underline{\mathbf{Y}}_j \\
&= \mathbf{G}^{-1} \sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \underline{\mathbf{Y}}_j,
\end{aligned} \tag{A.6}$$

que é o resultado (2.27).

A.2 BLUP

De $\hat{\boldsymbol{\beta}}_j = \mathbf{Z}_j \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\delta}}_j$ temos que $\hat{\boldsymbol{\delta}}_j = \hat{\boldsymbol{\beta}}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}$. Substituindo nessa equação o $\hat{\boldsymbol{\beta}}_j$ encontrado na expressão (A.1) tem-se

$$\hat{\boldsymbol{\delta}}_j = \mathbf{C}_j^{-1} \left(\mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{\mathbf{Y}}_j + \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \right) - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}$$

$$\begin{aligned}
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} \mathbf{Z}_j \hat{\boldsymbol{\gamma}} - \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \left(\mathbf{C}_j^{-1} \boldsymbol{\tau}^{-1} - \mathbf{I} \right) \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \left[\mathbf{C}_j^{-1} \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) - \mathbf{I} \right] \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j + \left(\mathbf{I} - \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j - \mathbf{I} \right) \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \underline{Y}_j - \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \hat{\boldsymbol{\gamma}} \right),
\end{aligned} \tag{A.7}$$

que é a expressão (2.25).

A.3 Identidades (2.29) e (2.31)

$$\begin{aligned}
\boldsymbol{\tau} \mathbf{X}_j^\top \mathbf{M}_j &= \boldsymbol{\tau} \mathbf{X}_j^\top \left(\mathbf{R}_j^{-1} - \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \right) \\
&= \boldsymbol{\tau} \left(\mathbf{X}_j^\top \mathbf{R}_j^{-1} - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \right) \\
&= \boldsymbol{\tau} \left(\mathbf{C}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \\
&= \boldsymbol{\tau} \left(\boldsymbol{\tau}^{-1} + \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j - \mathbf{X}_j^\top \mathbf{R}_j^{-1} \mathbf{X}_j \right) \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1} \\
&= \mathbf{C}_j^{-1} \mathbf{X}_j^\top \mathbf{R}_j^{-1}.
\end{aligned} \tag{A.8}$$

$$\begin{aligned}
\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \hat{\boldsymbol{\gamma}} &= \underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \mathbf{G}^{-1} \sum_{k=1}^J \mathbf{Z}_k^\top \mathbf{X}_k^\top \mathbf{M}_k \underline{Y}_k \\
&= \mathbf{M}_j^{-1} \left(\mathbf{M}_j \underline{Y}_j - \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j \mathbf{G}^{-1} \sum_{k=1}^J \mathbf{Z}_k^\top \mathbf{X}_k^\top \mathbf{M}_k \underline{Y}_k \right) \\
&= \mathbf{M}_j^{-1} \left(\mathbf{M}_j \underline{Y}_j - \sum_{k=1}^J \mathbf{P}_{jk} \underline{Y}_k \right).
\end{aligned} \tag{A.9}$$

A.4 Identidade (2.35)

$$\begin{aligned}
\text{Var} \left(\hat{\boldsymbol{\gamma}} \right) &= \mathbf{G}^{-1} \left(\sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j^\top \text{Var} \left(\underline{Y}_j \right) \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j \right) \mathbf{G}^{-1} \\
&= \mathbf{G}^{-1} \left(\sum \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j^\top \sigma^2 \mathbf{M}_j^{-1} \mathbf{M}_j \mathbf{X}_j \mathbf{Z}_j \right) \mathbf{G}^{-1}
\end{aligned}$$

$$\begin{aligned} &= \sigma^2 \mathbf{G}^{-1} \left(\sum \mathbf{z}_j^\top \mathbf{X}_j^\top \mathbf{M}_j \mathbf{X}_j \mathbf{z}_j \right) \mathbf{G}^{-1} \\ &= \sigma^2 \mathbf{G}^{-1}. \end{aligned} \tag{A.10}$$

Neste apêndice serão apresentados alguns desenvolvimentos das expressões apresentadas no Capítulo 3.

B.1 Perturbação na matriz de covariâncias dos erros (nível 1)

Lembrando que, se \mathbf{A} uma matriz $n \times n$ simétrica positiva definida e \underline{t} um escalar, então (Searle et al. 1992)

$$\frac{\partial \mathbf{A}^{-1}}{\partial \underline{t}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \underline{t}} \mathbf{A}^{-1} \quad (\text{B.1})$$

e

$$\frac{\partial \ln |\mathbf{A}|}{\partial \underline{t}} = \text{tr} \left\{ \mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \underline{t}} \right\}. \quad (\text{B.2})$$

Derivando o logaritmo da função de verossimilhança perturbada (3.10) com respeito a $\underline{\gamma}$ obtemos

$$\frac{\partial \lambda_j(\underline{w}_j)}{\partial \underline{\gamma}} = \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}_j) (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}), \quad (\text{B.3})$$

derivando (B.3) com respeito a w_{ji} e calculando em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ obtém-se

$$\left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \underline{\gamma}} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} = \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{D}_i \hat{\underline{\xi}}_j, \quad (\text{B.4})$$

que é o resultado (3.12).

Usando os resultados (B.1) e (B.2) tem-se que ($\forall k = 1, \dots, l$)

$$\begin{aligned} \frac{\partial \lambda_j(\underline{w}_j)}{\partial \theta_k} &= -\frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}_j) \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \mathbf{V}_j^{-1}(\underline{w}_j) \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}_j) \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right). \end{aligned} \quad (\text{B.5})$$

Derivando (B.5) com relação a w_{ji} temos

$$\begin{aligned} \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \theta_k} &= -\frac{1}{2} \text{tr} \left\{ \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}_j) \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right) \\ &+ \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \mathbf{V}_j^{-1}(\underline{w}_j) \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right) \\ &= -\frac{1}{2} \text{tr} \left\{ \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \mathbf{X}_j \frac{\partial \boldsymbol{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}_j) \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right). \end{aligned} \quad (\text{B.6})$$

Calculando (B.6) em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ obtêm-se a expressão (3.13) ($\forall k = 1, \dots, l$).

Também, utilizando os resultados (B.1) e (B.2) para derivar (3.10) com relação a σ^2 foi obtido

$$\begin{aligned} \frac{\partial \lambda_j(\underline{w}_j)}{\partial \sigma^2} &= -\frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j) \right\} \\ &+ \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j) \mathbf{V}_j^{-1}(\underline{w}_j) \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right), \end{aligned} \quad (\text{B.7})$$

e em seguida derivando com respeito a w_{ji} obtemos

$$\begin{aligned} \left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \sigma^2} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} &= -\frac{1}{2} \text{tr} \left\{ \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} \right\} \\ &+ \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j) \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right). \end{aligned} \quad (\text{B.8})$$

Uma vez que $\mathbf{V}_j(\underline{w}_j) = \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top + \sigma^2 \boldsymbol{\Lambda}(\underline{w}_j)$, pré-multiplicando por $\mathbf{V}_j^{-1}(\underline{w}_j)$ e derivando com relação a w_{ji} é facilmente encontrado

$$\frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} = -\sigma^{-2} \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top. \quad (\text{B.9})$$

Por outro lado temos também

$$\frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j) \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} = \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \boldsymbol{\Lambda}(\underline{w}_j) \mathbf{V}_j^{-1}(\underline{w}_j)$$

$$\begin{aligned}
& + \mathbf{V}_j^{-1}(\underline{w}_j) \boldsymbol{\Lambda}(\underline{w}_j) \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} + \mathbf{V}_j^{-1}(\underline{w}_j) \frac{\partial \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} \mathbf{V}_j^{-1}(\underline{w}_j) \\
& = 2 \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} \boldsymbol{\Lambda}(\underline{w}_j) \mathbf{V}_j^{-1}(\underline{w}_j) \\
& + \mathbf{V}_j^{-1}(\underline{w}_j) \frac{\partial \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} \mathbf{V}_j^{-1}(\underline{w}_j). \tag{B.10}
\end{aligned}$$

Calculando (B.8) em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \underline{\hat{\psi}}_j$, considerando (B.9) e (B.10) e percebendo que $\frac{\partial \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} = \sigma^{-2} \frac{\partial \mathbf{V}_j(\underline{w}_j)}{\partial w_{ji}}$ e $\boldsymbol{\Lambda}(\underline{w}_0) = \mathbf{I}_{n_j}$, obtemos

$$\frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial w_{ji} \partial \sigma^2} \Big|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \underline{\hat{\psi}}_j} = -\frac{1}{2} \left\{ \sigma^{-2} \text{tr} \left[\mathbf{D}_i \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \right] - 2 \hat{\boldsymbol{\tau}}_j^\top \mathbf{D}_i \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j + \sigma^{-2} \hat{\boldsymbol{\tau}}_j \mathbf{D}_i \hat{\boldsymbol{\tau}}_j \right\}, \tag{B.11}$$

que é a expressão (3.14).

Temos também

$$\mathbf{D}_i = \frac{\partial \mathbf{V}_j^{-1}(\underline{w}_j)}{\partial w_{ji}} = -\mathbf{V}_j^{-1}(\underline{w}_j) \frac{\partial \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}} \mathbf{V}_j^{-1}(\underline{w}_j). \tag{B.12}$$

Notando que $\frac{\partial \boldsymbol{\Lambda}(\underline{w}_j)}{\partial w_{ji}}$ avaliadas em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \underline{\hat{\psi}}_j$ é igual a \mathbf{I}_{n_j} , daí segue o resultado (3.15).

B.2 Perturbação na variável resposta

Derivando (3.17) com relação ao vetor de parâmetros $\boldsymbol{\gamma}$ têm-se

$$\frac{\partial \lambda_j(\underline{w}_j)}{\partial \boldsymbol{\gamma}} = \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1} (\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma}). \tag{B.13}$$

Derivando (B.13) com respeito a \underline{w}_j^\top e calculando essa derivada em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \underline{\hat{\psi}}_j$ obtemos

$$\frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial \underline{w}_j^\top \partial \boldsymbol{\gamma}} \Big|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \underline{\hat{\psi}}_j} = s \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{Z}_j. \tag{B.14}$$

Considerando o resultado resultado (B.1) tem-se que ($\forall k = 1, \dots, l$)

$$\frac{\partial \lambda_j(\underline{w}_j)}{\partial \theta_k} = \frac{1}{2} \left(\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma} \right)^\top \mathbf{V}_j^{-1} \mathbf{X}_j \frac{\partial \tau_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1} \left(\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma} \right). \tag{B.15}$$

Derivando (B.15) com relação a \underline{w}_j e calculando essa derivada em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \underline{\hat{\psi}}_j$ obtemos

$$\frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial \underline{w}_j \partial \theta_k} \Big|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \underline{\hat{\psi}}_j} = s \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j. \tag{B.16}$$

Analogamente ao caso anterior temos

$$\frac{\partial \lambda_j(\underline{w}_j)}{\partial \sigma^2} = \frac{1}{2} \left(\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right)^\top \mathbf{V}_j^{-1} \mathbf{V}_j^{-1} \left(\underline{Y}_j + s \underline{w}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma} \right). \quad (\text{B.17})$$

Derivando (B.17) com relação a \underline{w}_j e calculando essa derivada em $\underline{w}_j = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ obtêm-se

$$\left. \frac{\partial^2 \lambda_j(\underline{w}_j)}{\partial \underline{w}_j \partial \sigma^2} \right|_{\underline{w}_j = \underline{w}_0; \underline{\psi}_j = \hat{\underline{\psi}}_j} = s \hat{\mathbf{V}}_j^{-1} \hat{\mathbf{V}}_j^{-1} \hat{\underline{r}}_j. \quad (\text{B.18})$$

Por (B.14), (B.16) e (B.18) segue o resultado (3.18).

B.3 Perturbação na matriz de covariâncias dos efeitos aleatórios (nível 2)

Derivando o logaritmo da função de verossimilhança perturbada (3.19) com respeito a $\underline{\gamma}$ obtemos

$$\frac{\partial \lambda_j(\underline{w})}{\partial \underline{\gamma}} = \sum_{j=1}^J \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}), \quad (\text{B.19})$$

derivando (B.19) com respeito a w_j temos

$$\frac{\partial^2 \lambda_j(\underline{w})}{\partial w_j \partial \underline{\gamma}} = -\mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \underline{\tau}_j \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \underline{\gamma}), \quad (\text{B.20})$$

e calculando em $\underline{w} = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ obtêm-se a expressão (3.21).

Usando os resultados (B.1) e (B.2) tem-se ($\forall k = 1, \dots, l$)

$$\begin{aligned} \frac{\partial \lambda(\underline{w})}{\partial \theta_k} &= - \sum_{j=1}^J \frac{w_j}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \sum_{j=1}^J \frac{w_j}{2} \underline{r}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \underline{r}_j, \end{aligned} \quad (\text{B.21})$$

derivando (B.21) com relação a w_j obtêm-se

$$\begin{aligned} \frac{\partial^2 \lambda(\underline{w})}{\partial w_j \partial \theta_k} &= -\frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \frac{w_j}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \underline{\tau}_j \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \right\} \\ &+ \frac{1}{2} \underline{r}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \underline{r}_j \\ &- w_j \underline{r}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \underline{\tau}_j \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \frac{\partial \underline{\tau}_j}{\partial \theta_k} \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \underline{r}_j \end{aligned} \quad (\text{B.22})$$

e calculando (B.22) em $\underline{w} = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$ temos

$$\begin{aligned}
\left. \frac{\partial^2 \lambda(\underline{w})}{\partial w_j \partial \theta_k} \right|_{\underline{w}=\underline{w}_0; \underline{\psi}_j=\hat{\underline{\psi}}_j} &= -\frac{1}{2} \text{tr} \left\{ \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \right\} \\
&+ \frac{1}{2} \text{tr} \left\{ \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \right\} \\
&+ \frac{1}{2} \hat{\boldsymbol{\tau}}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j \\
&- \hat{\boldsymbol{\tau}}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \hat{\boldsymbol{\tau}}_j \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \mathbf{D}_k \mathbf{X}_j^\top \hat{\mathbf{V}}_j^{-1} \hat{\boldsymbol{\tau}}_j
\end{aligned} \tag{B.23}$$

que é o resultado (3.23).

Também, utilizando os resultados (B.1) e (B.2) para derivar (3.19) com relação a σ^2 foi obtido

$$\begin{aligned}
\frac{\partial \lambda(\underline{w})}{\partial \sigma^2} &= -\sum_{j=1}^J \frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}) \right\} \\
&+ \sum_{j=1}^J \frac{1}{2} \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma} \right)^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{V}_j^{-1}(\underline{w}) \left(\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma} \right).
\end{aligned} \tag{B.24}$$

e em seguida derivando com respeito a w_j obtemos

$$\begin{aligned}
\frac{\partial^2 \lambda(\underline{w})}{\partial w_j \partial \sigma^2} &= \frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \right\} \\
&- \boldsymbol{r}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{X}_j \boldsymbol{\tau}_j \mathbf{X}_j^\top \mathbf{V}_j^{-1}(\underline{w}) \mathbf{V}_j^{-1}(\underline{w}) \boldsymbol{r}_j.
\end{aligned} \tag{B.25}$$

Calculando (B.25) em $\underline{w} = \underline{w}_0$ e $\underline{\psi}_j = \hat{\underline{\psi}}_j$, obtemos a expressão (3.22).

B.4 Matriz Hessiana

Temos que

$$\frac{\partial^2 L(\underline{\psi}_j)}{\partial \underline{\psi}_j^\top \partial \underline{\psi}_j} = \begin{bmatrix} \frac{\partial^2 L(\underline{\psi}_j)}{\partial \boldsymbol{\gamma}^\top \partial \boldsymbol{\gamma}} & \frac{\partial^2 L(\underline{\psi}_j)}{\partial (\boldsymbol{\theta}^*)^\top \partial \boldsymbol{\gamma}} \\ \frac{\partial^2 L(\underline{\psi}_j)}{\partial \boldsymbol{\gamma}^\top \partial \boldsymbol{\theta}^*} & \frac{\partial^2 L(\underline{\psi}_j)}{\partial (\boldsymbol{\theta}^*)^\top \partial \boldsymbol{\theta}^*} \end{bmatrix}, \tag{B.26}$$

com o logaritmo da função de verossimilhança $L(\underline{\psi}_j)$ dado por (3.9). Tem-se que ($k, r = 1, \dots, l+$

1)

$$\frac{\partial L(\underline{\psi}_j)}{\partial \boldsymbol{\gamma}} = \mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1} (\underline{Y}_j - \mathbf{X}_j \mathbf{Z}_j \boldsymbol{\gamma}) \tag{B.27}$$

$$\frac{\partial L(\underline{\psi}_j)}{\partial \boldsymbol{\theta}_k^*} = -\frac{1}{2} \text{tr} \left\{ \left(\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \boldsymbol{\theta}_k^*} \right) - \boldsymbol{r}_j^\top \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \boldsymbol{\theta}_k^*} \mathbf{V}_j^{-1} \boldsymbol{r}_j \right\}, \tag{B.28}$$

$$\frac{\partial^2 L(\psi_j)}{\partial \gamma^\top \partial \gamma} = -\mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1} \mathbf{X}_j \mathbf{Z}_j \quad (\text{B.29})$$

$$\frac{\partial^2 L(\psi_j)}{\partial \theta_k^* \partial \gamma} = -\mathbf{Z}_j^\top \mathbf{X}_j^\top \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k^*} \mathbf{V}_j^{-1} \mathcal{L}_j \quad (\text{B.30})$$

$$\begin{aligned} \frac{\partial^2 L(\psi_j)}{\partial \theta_r^* \partial \theta_k} &= -\frac{1}{2} \text{tr} \left\{ \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_r^*} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k^*} - \mathbf{V}_j^{-1} \frac{\partial^2 \mathbf{V}_j}{\partial \theta_r^* \partial \theta_k^*} \right\} \\ &+ \frac{1}{2} \mathcal{L}_j^\top \mathbf{V}_j^{-1} \frac{\partial^2 \mathbf{V}_j}{\partial \theta_r^* \partial \theta_k^*} \mathbf{V}_j^{-1} \mathcal{L}_j \\ &- \mathcal{L}_j^\top \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_r^*} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k^*} \mathbf{V}_j^{-1} \mathcal{L}_j. \end{aligned} \quad (\text{B.31})$$

Para encontrar $\ddot{\mathbf{L}}$ basta calcular todas as derivadas acima no EMV de ψ_j .

Rotinas em Linguagem R

Neste capítulo são apresentadas as rotinas em linguagem R, utilizadas para o ajuste dos modelos selecionados e para a análise de influência local destes.

```
#####  
#                                     #  
#           Ajuste dos Modelos       #  
#                                     #  
#####  
  
## Bibliotecas...  
library(lme4)  
library(nlme)  
  
## Limpando a memória...  
rm(list=ls(all=TRUE))  
  
## Definindo o diretório de trabalho...  
setwd("C:\\Users\\Juliana\\Documents\\dissertação\\Banco de dados")  
  
## Importando o arquivo de dados...  
dados = read.table("SAEPE R.dat", header = T, dec = ",")  
  
mat2 <- na.omit(dados[, c(1,2,3,4,7,8,10,12,18)])  
  
attach(mat2)  
  
genero = factor(genero)  
horaest = factor(horaest)  
repetiu = factor(repetiu)  
conservesc = factor(conservesc)  
conservsala2_Q1 = factor(conservsala2_Q1)  
  
## Modelagem das habilidades em MATEMÁTICA...  
mod2 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest, mat2)  
mat1 <- na.omit(dados[, c(1,2,3,7,8,12)])  
mod1 = lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc, data = mat2,  
na.action = na.omit)
```

```

## Escolas influentes: 8, 16, 38

gruposmat <- as.numeric(levels(as.factor(mat2$codigoesc))) # Identificadores das escolas...
gruposmat <- sort(gruposmat)

## Descritivas das escolas influentes...

# Escola 8
mediamat8 <- mean(mat1$habilidade_mat[mat1$codigoesc == gruposmat[8]])
varmat8 <- var(mat1$habilidade_mat[mat1$codigoesc == gruposmat[8]])
cat("Escola 8:",gruposmat[8],"\nHabilidade média em Matemática:",mediamat8,"\nVariância:",
varmat8,"\nDesvio padrão:",sqrt(varmat8),"\n")

# Escola 16
mediamat16 <- mean(mat1$habilidade_mat[mat1$codigoesc == gruposmat[16]])
varmat16 <- var(mat1$habilidade_mat[mat1$codigoesc == gruposmat[16]])
cat("Escola 16:",gruposmat[16],"\nHabilidade média em Matemática:",mediamat16,"\nVariância:",
varmat16,"\nDesvio padrão:",sqrt(varmat16),"\n")

# Escola 38
mediamat38 <- mean(mat1$habilidade_mat[mat1$codigoesc == gruposmat[38]])
varmat38 <- var(mat1$habilidade_mat[mat1$codigoesc == gruposmat[38]])
cat("Escola 38:",gruposmat[38],"\nHabilidade média em Matemática:",mediamat38,"\nVariância:",
varmat38,"\nDesvio padrão:",sqrt(varmat38),"\n")

# Modelos retirando as escolas influentes
mod2.8 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[8], ])
mod2.16 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[16], ])
mod2.38 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[38], ])
mod2.8.16 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[16], ])
mod2.8.38 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[38], ])
mod2.16.38 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[16] & mat2$codigoesc != gruposmat[38], ])
mod2.8.16.38 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[16] &
mat2$codigoesc != gruposmat[38], ])

# Modelos retirando as escolas influentes usando nlme
#mod3.8 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[8], ], na.action = na.omit)
mod3.16 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[16], ], na.action = na.omit)
intervals(mod3.16)
mod3.38 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[38], ], na.action = na.omit)
intervals(mod3.38)
#mod3.8.16 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[16], ], na.action = na.omit)
#mod3.8.38 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[38], ], na.action = na.omit)
mod3.16.38 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[16] & mat2$codigoesc != gruposmat[38], ], na.action = na.omit)
intervals(mod3.16.38)
#mod3.8.16.38 <- lme(habilidade_mat ~ conservesc + genero + horaest , random = ~ 1 | codigoesc,
mat2[mat2$codigoesc != gruposmat[8] & mat2$codigoesc != gruposmat[16] & mat2$codigoesc !=
gruposmat[38], ], na.action = na.omit)

#####

```

```

## Modelagem das habilidades em PORTUGUÊS...
mod2 <- lmer(habilidade_por ~ (repetiu|codigoesc) + conservsala2_Q1 + repetiu + genero
+ horaest, dados)
mod3 = lme(habilidade_por ~ repetiu + conservsala2_Q1 + horaest + genero,
random = ~ 1 + repetiu | codigoesc, data = mat2, na.action = na.omit)
intervals(mod3)
por1 <- na.omit(dados[, c(1,2,4,7,8,10,18)])

## Escolas influentes: 8, 40

grupospor <- as.numeric(levels(as.factor(por1$codigoesc))) # Identificadores das escolas...
grupospor <- sort(grupospor)

# Escola 8
mediapor8 <- mean(por1$habilidade_por[por1$codigoesc == grupospor[8]])
varpor8 <- var(por1$habilidade_por[por1$codigoesc == grupospor[8]])
cat("Escola 8:",grupospor[8],"\nHabilidade média em Português:",mediapor8,"\nVariância:",
varpor8,"\nDesvio padrão:",sqrt(varpor8),"\n")

# Escola 40
mediapor40 <- mean(por1$habilidade_por[por1$codigoesc == grupospor[40]])
varpor40 <- var(por1$habilidade_por[por1$codigoesc == grupospor[40]])
cat("Escola 40:",grupospor[40],"\nHabilidade média em Português:",mediapor40,"\nVariância:",
varpor40,"\nDesvio padrão:",sqrt(varpor40),"\n")

# Modelos retirando as escolas influentes
mod2.8 <- lmer(habilidade_por ~ (repetiu|codigoesc) + conservsala2_Q1 + repetiu + genero +
horaest, mat2[mat2$codigoesc != grupospor[8], ])
mod2.40 <- lmer(habilidade_por ~ (repetiu|codigoesc) + conservsala2_Q1 + repetiu + genero +
horaest, mat2[mat2$codigoesc != grupospor[40], ])
mod2.8.40 <- lmer(habilidade_por ~ (repetiu|codigoesc) + conservsala2_Q1 + repetiu + genero +
horaest, mat2[mat2$codigoesc != grupospor[8] & mat2$codigoesc != grupospor[40], ])

# Modelos retirando as escolas influentes nlme
mod3.8 <- lme(habilidade_por ~ genero + conservsala2_Q1 + repetiu + horaest,
random = ~ 1 + repetiu | codigoesc, mat2[mat2$codigoesc != grupospor[8], ], na.action = na.omit)
intervals(mod3.8)
mod3.40 <- lme(habilidade_por ~ genero + conservsala2_Q1 + repetiu + horaest,
random = ~ 1 + repetiu | codigoesc, mat2[mat2$codigoesc != grupospor[40], ], na.action = na.omit)
intervals(mod3.40)
mod3.8.40 <- lme(habilidade_por ~ repetiu + conservsala2_Q1 + horaest + genero,
random = ~ 1 + repetiu | codigoesc, mat2[mat2$codigoesc != grupospor[8] & mat2$codigoesc !=
grupospor[40], ], na.action = na.omit)
intervals(mod3.8.40)

#####
# #
# Influência Local para o modelo ajustado para matemática #
# #
#####

## Bibliotecas...
library(lme4)

## Limpando a memória...
rm(list=ls(all=TRUE))

## Definindo o diretório de trabalho...
#setwd("C:\\Users\\Juliana\\Documents\\dissertação\\Banco de dados")

## Importando o arquivo de dados...
dados = read.table("SAEPE R.dat", header = T, dec = ",")

mat1 <- na.omit(dados[, c(1,2,3,4,7,8,10,12,18)])

```

```

##
attach(mat1)

genero = factor(genero)
horaest = factor(horaest)
repetiu = factor(repetiu)
conservesc = factor(conservesc)
conservsala2_Q1 = factor(conservsala2_Q1)

## Modelagem das habilidades em MATEMÁTICA...
mod2 <- lmer(habilidade_mat ~ (1|codigoesc) + conservesc + genero + horaest, mat1)

## Quantidades para medidas de influência local...

# Quantidades básicas...
grupos <- as.numeric(levels(as.factor(mat1$codigoesc))) # Identificadores das escolas...
J <- length(grupos) # Número de escolas...
njs <- numeric(J) # Número de alunos nas escolas...
for (j in 1 : J) {
  njs[j] <- dim(mat1[mat1$codigoesc == grupos[j],,])[1]
}
n <- sum(njs) # Número total de alunos

aux1 <- cumsum(njs)
aux1 <- c(0, aux1)
aux2 <- length(mod2@ranef)
aux2 <- seq(0, aux2, aux2/J)

# Matriz X...
mX <- as.matrix(t(mod2@Zt))
mXjs <- list()
for (j in 1 : J) {
  mXjs[[j]] <- mX[(aux1[j] + 1) : aux1[(j + 1)], (aux2[j] + 1) : aux2[(j + 1)]]
}

# Matriz XZ...
mXZ <- as.matrix(mod2@X)
mXZjs <- list()
for (j in 1 : J) {
  mXZjs[[j]] <- mXZ[(aux1[j] + 1) : aux1[(j + 1)], ]
}

# Matriz tau...
sumario <- summary(mod2)
tau00 <- as.numeric(sumario@REmat[1,3])
mtauj <- tau00
mtau <- diag(tau00, J)

# sigma^2...
sigma2 <- sumario@sigma^2

# Matriz R...
R <- diag(sigma2, n)
Rjs <- list()
for (j in 1 : J) {
  Rjs[[j]] <- R[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
}

# Matriz V...
V <- mX%*%mtau%*%t(mX) + R
Vjs <- list()
for (j in 1 : J) {
  Vjs[[j]] <- V[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
}

```

```

}

# Dimensões da matriz Z...
dimZ <- c(dim(mX)[2], length(sumario@fixef))

# Matrizes Di's...
iV <- solve(V)
iVjs <- list()
D <- list()
for (j in 1 : J) {
  iVjs[[j]] <- iV[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
  D[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  for (i in 1 : njs[j]) {
    D[[j]][, , i] <- -sigma2 * (iVjs[[j]][, i])%*%t(iVjs[[j]][, i])
  }
}

# Matrizes Dk's...
Dks <- 1

# Resíduos...
r <- mod2@y - mXZ%*%sumario@fixef
rjs <- list()
for (j in 1 : J) {
  rjs[[j]] <- r[(aux1[j] + 1) : aux1[(j + 1)]]
}

# Matrizes H...

# Esquema 1: Perturbação na matrix de covariância do erro...

d11 <- NULL # Derivada 1...
aux3 <- list()
for (j in 1 : J) {
  aux3[[j]] <- array(matrix(), dim = c(dimZ[2], 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux3[[j]][, , i] <- t(mXZjs[[j]])%*%(D[[j]][, , i])%*%(rjs[[j]])
  }
  d11 <- rbind(d11, aux3[[j]][, , i])
}

d12 <- NULL # Derivada 2...
aux4 <- list()
aux5 <- list()
aux6 <- list()
for (j in 1 : J) {
  aux4[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux5[[j]] <- mtau_j * mXjs[[j]]%*%t(mXjs[[j]])
  aux6[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux4[[j]][, , i] <- D[[j]][, , i]%*%aux5[[j]]
    aux6[[j]][, , i] <- (-.5/sigma2)*sum(diag(aux4[[j]][, , i])) +
      t(rjs[[j]])%*%D[[j]][, , i]%*%iVjs[[j]]%*%rjs[[j]] - (.5/sigma2)*
      t(rjs[[j]])%*%D[[j]][, , i]%*%rjs[[j]]
  }
  d12 <- rbind(d12, aux6[[j]][, , i])
}

d13 <- NULL # Derivada 3...
aux7 <- list()
aux8 <- list()
aux9 <- list()
for (j in 1 : J) {
  aux7[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux8[[j]] <- mXjs[[j]]%*%t(mXjs[[j]])
}

```

```

aux9[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
for (i in 1 : njs[j]) {
aux7[[j]][ , , i] <- D[[j]][ , , i]*aux8[[j]]
aux9[[j]][ , , i] <- (-.5)*sum(diag(aux7[[j]][ , , i])) + t(rjs[[j]])**%
D[[j]][ , , i]**aux8[[j]]**rjs[[j]]
d13 <- rbind(d13, aux9[[j]][ , , i])
}
}

H1 <- cbind(d11, d12, d13)

# Esquema 2: Perturbação na variável resposta...

s = 1

d21 <- NULL # Derivada 1...
d22 <- NULL # Derivada 2...
d23 <- NULL # Derivada 3...
aux16 <- list()
aux17 <- list()
aux18 <- list()
for (j in 1 : J) {
aux16[[j]] <- s*iVjs[[j]]**mXZjs[[j]]
d21 <- rbind(d21, aux16[[j]])
#
aux17[[j]] <- s*iVjs[[j]]**iVjs[[j]]**rjs[[j]]
d22 <- rbind(d22, aux17[[j]])
#
aux18[[j]] <- s*iVjs[[j]]**mXjs[[j]]**t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d23 <- rbind(d23, aux18[[j]])
}

H2 <- cbind(d21, d22, d23)

# Esquema 3: Perturbação nos efeitos aleatórios...

d31 <- NULL # Derivada 1...
d32 <- NULL # Derivada 2...
d33 <- NULL # Derivada 3...
aux21 <- list()
aux22 <- list()
aux23 <- list()
for (j in 1 : J) {
aux21[[j]] <- (-mtauj) * t(mXZjs[[j]])**iVjs[[j]]**mXjs[[j]]**%
t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d31 <- rbind(d31, t(aux21[[j]]))
#
aux22[[j]] <- mtauj * ((.5) * sum(diag(iVjs[[j]]**mXjs[[j]]**t(mXjs[[j]])**%
iVjs[[j]])) - t(rjs[[j]])**iVjs[[j]]**mXjs[[j]]**t(mXjs[[j]])**iVjs[[j]]**%
iVjs[[j]]**rjs[[j]])
d32 <- rbind(d32, aux22[[j]])
#
aux23[[j]] <- (-.5) * sum(diag(iVjs[[j]]**mXjs[[j]]**t(mXjs[[j]])) +
mtauj * .5 *sum(diag(iVjs[[j]]**mXjs[[j]]**t(mXjs[[j]])**iVjs[[j]]**%
mXjs[[j]]**t(mXjs[[j]])) + .5 * t(rjs[[j]])**iVjs[[j]]**mXjs[[j]]**%
t(mXjs[[j]])**iVjs[[j]]**rjs[[j]] -
mtauj * t(rjs[[j]])**iVjs[[j]]**mXjs[[j]]**%
t(mXjs[[j]])**iVjs[[j]]**mXjs[[j]]**%
t(mXjs[[j]])**iVjs[[j]]**rjs[[j]])
d33 <- rbind(d33, aux23[[j]])
}

H3 <- cbind(d31, d32, d33)

# Matriz Hessiana...

```

```

auxh1 <- diag(1, J)

h1 <- -t(mXZ)%*%iV%*%mXZ
h21 <- -t(mXZ)%*%iV%*%diag(1, n)%*%iV%*%r
h22 <- -t(mXZ)%*%iV%*%mX%*%auxh1%*%t(mX)%*%iV%*%r
h31 <- .5 * sum(diag(iV%*%diag(1, n)%*%iV%*%diag(1, n))) - t(r)%*%iV%*%diag(1, n)%*%
iV%*%diag(1, n)%*%iV%*%r
h32 <- .5 * sum(diag(iV%*%mX%*%auxh1%*%t(mX)%*%iV%*%mX%*%auxh1%*%t(mX))) -
t(r)%*%iV%*%mX%*%auxh1%*%t(mX)%*%iV%*%mX%*%auxh1%*%t(mX)%*%iV%*%r

L2 <- -rbind(cbind(h1, h21, h22), cbind(t(cbind(h21, h22)), diag(c(h31, h32), 2)))

# Gráficos de Influência...

# Pertubação na matriz de covariância do erro...

auxc1 <- H1%*%solve(L2)%*%t(H1)
dmax1 <- eigen(auxc1)$vec[, 1]
postscript("Graf1Mat2.ps")
plot(rep(1 : J, njs), abs(dmax1), xlab = "Observacao", ylab = "|dmax|")
for (j in 1 : J) {
for (i in (aux1[j]+1) : aux1[j+1]) {
if (j == 1) {
if (abs(dmax1[i]) > .09)
text(j, abs(dmax1[i]) - .0015, i)
}
else
if (abs(dmax1[i]) > .09)
text(j, abs(dmax1[i]) - .0015, i - sum(njs[1:j-1]))
}
}
dev.off()

# Pertubação na variável resposta...

#auxc2 <- H2%*%solve(L2)%*%t(H2)
#dmax2 <- eigen(auxc2)$vec[, 1]
#postscript("Graf2Mat1.ps")
#plot(rep(1 : J, njs), abs(dmax2), xlab = "Observacao", ylab = "|dmax|")
#dev.off()

# Pertubação na matriz de covariância dos efeitos aleatórios...

auxc3 <- H3%*%solve(L2)%*%t(H3)
dmax3 <- eigen(auxc3)$vec[, 1]
postscript("Graf3Mat2.ps")
plot(1 : J, abs(dmax3), xlab = "Escola", ylab = "|dmax|")
for (i in 1 : J) {
if (abs(dmax3[i]) > .4) {
text(i, (abs(dmax3[i]) - 0.02), i)
# Comando usado para identificar os pontos influentes
}
}
dev.off()

#####
#
# Influência Local para o modelo ajustado para português #
#
#####

## Bibliotecas...
library(lme4)

```

```

## Limpando a memória...
rm(list=ls(all=TRUE))

## Definindo o diretório de trabalho...
#setwd("C:\\Users\\Juliana\\Documents\\dissertação\\Banco de dados")

## Importando o arquivo de dados...
dados = read.table("SAEPE R.dat", header = T, dec = ",")

##
attach(dados)

tipoes = factor(tipoesc)
grupo = factor(grupo)
genero = factor(genero)
horaest = factor(horaest)
trabalha = factor(trabalha)
repetiu = factor(repetiu)
conserve = factor(conserve)
conserve_sala = factor(conserve_sala)
Sala_Quartil1 = factor(Sala_Quartil1)
Conserve_Quartil1 = factor(Conserve_Quartil1)
conserve2_Q1 = factor(conserve2_Q1)
conserve2_Q2 = factor(conserve2_Q2)
conserve_sala2_Q1 = factor(conserve_sala2_Q1)
conserve_sala2_Q2 = factor(conserve_sala2_Q2)
instalacoes2_Q1 = factor(instalacoes2_Q1)
instalacoes2_Q2 = factor(instalacoes2_Q2)

## Modelagem das habilidades em PORTUGUÊS...
mod2 <- lmer(habilidade_por ~ (repetiu|codigoesc) + conserve_sala2_Q1 + repetiu +
genero + horaest, dados)
por1 <- na.omit(dados[, c(1,2,4,7,8,10,18)])

## Quantidades para medidas de influência local...

# Quantidades básicas...
grupos <- as.numeric(levels(as.factor(por1$codigoesc))) # Identificadores das escolas...
J <- length(grupos) # Número de escolas...
njs <- numeric(J) # Número de alunos nas escolas...
for (j in 1 : J) {
njs[j] <- dim(por1[por1$codigoesc == grupos[j],,])[1]
}
n <- sum(njs) # Número total de alunos

aux1 <- cumsum(njs)
aux1 <- c(0, aux1)
aux2 <- length(mod2@ranef)
aux2 <- seq(0, aux2, aux2/J)

# Matriz X...

mX <- as.matrix(t(mod2@Zt))
mXjs <- list()
for (j in 1 : J) {
mXjs[[j]] <- mX[(aux1[j] + 1) : aux1[(j + 1)], (aux2[j] + 1) : aux2[(j + 1)]]
}

# Matriz XZ...

mXZ <- as.matrix(mod2@X)
mXZjs <- list()
for (j in 1 : J) {
mXZjs[[j]] <- mXZ[(aux1[j] + 1) : aux1[(j + 1)], ]
}

```

```

# Matriz tau...
sumario <- summary(mod2)
tau00 <- as.numeric(sumario@REmat[1,3])
tau11 <- as.numeric(sumario@REmat[2,3])
tau01 <- as.numeric(sumario@REmat[2,5])*sqrt(tau00*tau11)
mtauj <- matrix(c(tau00, tau01, tau01, tau11), 2, 2) # Matriz tau_j
mtau <- kronecker(diag(1, J), mtauj, fun = "*") # Matriz tau

# sigma^2...
sigma2 <- sumario@sigma^2

# Matriz R...
R <- diag(sigma2, n)
Rjs <- list()
for (j in 1 : J) {
Rjs[[j]] <- R[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
}

# Matriz V...
V <- mX%*%mtau%*%t(mX) + R
Vjs <- list()
for (j in 1 : J) {
Vjs[[j]] <- V[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
}

# Dimensões da matriz Z...
dimZ <- c(dim(mX)[2], length(sumario@fixef))

# Matrizes Di's...
iV <- solve(V)
iVjs <- list()
D <- list()
for (j in 1 : J) {
iVjs[[j]] <- iV[(aux1[j] + 1) : aux1[(j + 1)], (aux1[j] + 1) : aux1[(j + 1)]]
D[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
for (i in 1 : njs[j]) {
D[[j]][[ , i] <- -sigma2 * (iVjs[[j]][[ , i]])%*%t(iVjs[[j]][[ , i]])
}
}

# Matrizes Dk's...
Dks <- list()
for (i in 1 : dimZ[2]) {
if (i == 1)
Dks[[i]] <- matrix(c(1, 0, 0, 0), 2, 2)
else if (i == 2)
Dks[[i]] <- matrix(c(0, 0, 0, 1), 2, 2)
else
Dks[[i]] <- matrix(c(0, 1, 1, 0), 2, 2)
}

# Resíduos...
r <- mod2@y - mXZ%*%sumario@fixef
rjs <- list()
for (j in 1 : J) {
rjs[[j]] <- r[(aux1[j] + 1) : aux1[(j + 1)]]
}

# Matrizes H...

# Esquema 1: Perturbação na matrix de covariância do erro...

d11 <- NULL # Derivada 1...
aux3 <- list()

```

```

for (j in 1 : J) {
  aux3[[j]] <- array(matrix(), dim = c(dimZ[2], 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux3[[j]][ , , i] <- t(mXZjs[[j]])%*%D[[j]][ , , i]%*%(rjs[[j]])
    d11 <- rbind(d11, aux3[[j]][ , , i])
  }
}

d12 <- NULL # Derivada 2...
aux4 <- list()
aux5 <- list()
aux6 <- list()
for (j in 1 : J) {
  aux4[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux5[[j]] <- mXjs[[j]]%*%mtauj%*t(mXjs[[j]])
  aux6[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux4[[j]][ , , i] <- D[[j]][ , , i]%*%aux5[[j]]
    aux6[[j]][ , , i] <- (-.5/sigma2)*sum(diag(aux4[[j]][ , , i])) + t(rjs[[j]])%*%
    D[[j]][ , , i]%*%iVjs[[j]]%*%rjs[[j]] -
    (.5/sigma2)*t(rjs[[j]])%*%D[[j]][ , , i]%*%rjs[[j]]
    d12 <- rbind(d12, aux6[[j]][ , , i])
  }
}

d13 <- NULL # Derivada 3...
aux7 <- list()
aux8 <- list()
aux9 <- list()
for (j in 1 : J) {
  aux7[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux8[[j]] <- mXjs[[j]]%*%Dks[[1]]%*t(mXjs[[j]])
  aux9[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux7[[j]][ , , i] <- D[[j]][ , , i]%*%aux8[[j]]
    aux9[[j]][ , , i] <- (-.5)*sum(diag(aux7[[j]][ , , i])) + t(rjs[[j]])%*%
    D[[j]][ , , i]%*%aux8[[j]]%*%rjs[[j]]
    d13 <- rbind(d13, aux9[[j]][ , , i])
  }
}

d14 <- NULL # Derivada 4...
aux10 <- list()
aux11 <- list()
aux12 <- list()
for (j in 1 : J) {
  aux10[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux11[[j]] <- mXjs[[j]]%*%Dks[[2]]%*t(mXjs[[j]])
  aux12[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
  for (i in 1 : njs[j]) {
    aux10[[j]][ , , i] <- D[[j]][ , , i]%*%aux11[[j]]
    aux12[[j]][ , , i] <- (-.5)*sum(diag(aux10[[j]][ , , i])) + t(rjs[[j]])%*%
    D[[j]][ , , i]%*%aux11[[j]]%*%rjs[[j]]
    d14 <- rbind(d14, aux12[[j]][ , , i])
  }
}

d15 <- NULL # Derivada 5...
aux13 <- list()
aux14 <- list()
aux15 <- list()
for (j in 1 : J) {
  aux13[[j]] <- array(matrix(), dim = c(njs[j], njs[j], njs[j]))
  aux14[[j]] <- mXjs[[j]]%*%Dks[[3]]%*t(mXjs[[j]])
  aux15[[j]] <- array(matrix(), dim = c(1, 1, njs[j]))
}

```

```

for (i in 1 : njs[j]) {
aux13[[j]][ , , i] <- D[[j]][ , , i]*aux14[[j]]
aux15[[j]][ , , i] <- (-.5)*sum(diag(aux13[[j]][ , , i])) + t(rjs[[j]])**
D[[j]][ , , i]*aux14[[j]]**rjs[[j]]
d15 <- rbind(d15, aux15[[j]][ , , i])
}
}

H1 <- cbind(d11, d12, d13, d14, d15)

# Esquema 2: Pertubação na variável resposta...

s = 1

d21 <- NULL # Derivada 1...
d22 <- NULL # Derivada 2...
d23 <- NULL # Derivada 3...
d24 <- NULL # Derivada 4...
d25 <- NULL # Derivada 5...
aux16 <- list()
aux17 <- list()
aux18 <- list()
aux19 <- list()
aux20 <- list()
for (j in 1 : J) {
aux16[[j]] <- s*iVjs[[j]]**mXZjs[[j]]
d21 <- rbind(d21, aux16[[j]])
#
aux17[[j]] <- s*iVjs[[j]]**iVjs[[j]]**rjs[[j]]
d22 <- rbind(d22, aux17[[j]])
#
aux18[[j]] <- s*iVjs[[j]]**mXjs[[j]]**Dks[[1]]**t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d23 <- rbind(d23, aux18[[j]])
#
aux19[[j]] <- s*iVjs[[j]]**mXjs[[j]]**Dks[[2]]**t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d24 <- rbind(d24, aux19[[j]])
#
aux20[[j]] <- s*iVjs[[j]]**mXjs[[j]]**Dks[[3]]**t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d25 <- rbind(d25, aux20[[j]])
}

H2 <- cbind(d21, d22, d23, d24, d25)

# Esquema 3: Pertubação nos efeitos aleatórios...

d31 <- NULL # Derivada 1...
d32 <- NULL # Derivada 2...
d33 <- NULL # Derivada 3...
d34 <- NULL # Derivada 4...
d35 <- NULL # Derivada 5...
aux21 <- list()
aux22 <- list()
aux23 <- list()
aux24 <- list()
aux25 <- list()
for (j in 1 : J) {
aux21[[j]] <- (-1) * t(mXZjs[[j]])**iVjs[[j]]**mXjs[[j]]**mtauj**
t(mXjs[[j]])**iVjs[[j]]**rjs[[j]]
d31 <- rbind(d31, t(aux21[[j]]))
#
aux22[[j]] <- .5 * sum(diag(iVjs[[j]]**mXjs[[j]]**mtauj**
t(mXjs[[j]])**iVjs[[j]])) -
t(rjs[[j]])**iVjs[[j]]**mXjs[[j]]**mtauj**
t(mXjs[[j]])**iVjs[[j]]**iVjs[[j]]**rjs[[j]]
d32 <- rbind(d32, aux22[[j]])

```

```

#
aux23[[j]] <- (-.5) * sum(diag(iVjs[[j]]%*mXjs[[j]]%*Dks[[1]]%*%
t(mXjs[[j]]))) +
.5 * sum(diag(iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*iVjs[[j]]%*%
mXjs[[j]]%*Dks[[1]]%*t(mXjs[[j]]))) +
.5 * t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*Dks[[1]]%*%
t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]] -
t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*%
iVjs[[j]]%*mXjs[[j]]%*Dks[[1]]%*t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]]
d33 <- rbind(d33, aux23[[j]])
#
aux24[[j]] <- (-.5) * sum(diag(iVjs[[j]]%*mXjs[[j]]%*Dks[[2]]%*t(mXjs[[j]]))) +
.5 * sum(diag(iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*iVjs[[j]]%*%
mXjs[[j]]%*Dks[[2]]%*t(mXjs[[j]]))) +
.5 * t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*Dks[[2]]%*%
t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]] -
t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*iVjs[[j]]%*%
mXjs[[j]]%*Dks[[2]]%*t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]]
d34 <- rbind(d34, aux24[[j]])
#
aux25[[j]] <- (-.5) * sum(diag(iVjs[[j]]%*mXjs[[j]]%*Dks[[3]]%*t(mXjs[[j]]))) +
.5 * sum(diag(iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*%
iVjs[[j]]%*mXjs[[j]]%*Dks[[3]]%*t(mXjs[[j]]))) +
.5 * t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*Dks[[3]]%*%
t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]] -
t(rjs[[j]])%*iVjs[[j]]%*mXjs[[j]]%*mtauj%*t(mXjs[[j]])%*iVjs[[j]]%*%
mXjs[[j]]%*Dks[[3]]%*t(mXjs[[j]])%*iVjs[[j]]%*rjs[[j]]
d35 <- rbind(d35, aux25[[j]])
}

H3 <- cbind(d31, d32, d33, d34, d35)

# Matriz Hessiana...

auxh1 <- kronecker(diag(1, J), Dks[[1]], fun = "")
auxh2 <- kronecker(diag(1, J), Dks[[2]], fun = "")
auxh3 <- kronecker(diag(1, J), Dks[[3]], fun = "")

h1 <- -t(mXZ)%*iV%*mXZ
h21 <- -t(mXZ)%*iV%*diag(1, n)%*iV%*r
h22 <- -t(mXZ)%*iV%*mX%*auxh1%*t(mX)%*iV%*r
h23 <- -t(mXZ)%*iV%*mX%*auxh2%*t(mX)%*iV%*r
h24 <- -t(mXZ)%*iV%*mX%*auxh3%*t(mX)%*iV%*r
h31 <- .5 * sum(diag(iV%*diag(1, n)%*iV%*diag(1, n))) -
t(r)%*iV%*diag(1, n)%*iV%*diag(1, n)%*iV%*r
h32 <- .5 * sum(diag(iV%*mX%*auxh1%*t(mX)%*iV%*mX%*auxh1%*t(mX))) -
t(r)%*iV%*mX%*auxh1%*t(mX)%*iV%*mX%*auxh1%*t(mX)%*iV%*r
h33 <- .5 * sum(diag(iV%*mX%*auxh2%*t(mX)%*iV%*mX%*auxh2%*t(mX))) -
t(r)%*iV%*mX%*auxh2%*t(mX)%*iV%*mX%*auxh2%*t(mX)%*iV%*r
h34 <- .5 * sum(diag(iV%*mX%*auxh3%*t(mX)%*iV%*mX%*auxh3%*t(mX))) -
t(r)%*iV%*mX%*auxh3%*t(mX)%*iV%*mX%*auxh3%*t(mX)%*iV%*r

L2 <- -rbind(cbind(h1, h21, h22, h23, h24), cbind(t(cbind(h21, h22, h23, h24)),
diag(c(h31, h32, h33, h34), 4)))

# Gráficos de Influência...

# Perturbação na matriz de covariância do erro...

auxc1 <- H1%*solve(L2)%*t(H1)
dmax1 <- eigen(auxc1)$vec[, 1]
postscript("Graf1Port.ps")
plot(rep(1 : J, njs), abs(dmax1), xlab = "Observacao", ylab = "|dmax|")
for (j in 1 : J) {
for (i in (aux1[j]+1) : aux1[j+1]) {

```

```

if (j == 1) {
if (abs(dmax1[i]) > .08)
text(j, abs(dmax1[i]) - .001, i)
}
else
if(abs(dmax1[i]) > .08)
text(j, abs(dmax1[i]) - .001, i - sum(njs[1:j-1]))
}
}
dev.off()

# Pertubação na variável resposta...

#auxc2 <- H2%*%solve(L2)%*%t(H2)
#dmax2 <- eigen(auxc2)$vec[, 1]
#postscript("Graf2Port.ps")
#plot(rep(1 : J, njs), abs(dmax2), xlab = "Observacao", ylab = "|dmax|")
#dev.off()

# Pertubação na matriz de covariância dos efeitos aleatórios...

auxc3 <- H3%*%solve(L2)%*%t(H3)
dmax3 <- eigen(auxc3)$vec[, 1]
postscript("Graf3Port.ps")
plot(1 : J, abs(dmax3), xlab = "Escola", ylab = "|dmax|")
for (i in 1 : J) {
if (abs(dmax3[i]) > .4) {
text(i, (abs(dmax3[i]) - 0.01), i) # Comando usado para identificar os pontos influentes
}
}
dev.off()

```

Referências Bibliográficas

- Andrade, D. F., Tavares, H. R. & Valle, R. C. (2000), Teoria da resposta ao item: Conceitos e aplicações, 14º SINAPE, ABE.
- Araújo, P. V. (1998), *Geometria Diferencial*, Rio de Janeiro: IMPA.
- Associação Brasileira de Empresas de Pesquisa (2008), '<http://www.abep.org>'.
- Banerjee, M. & Frees, E. W. (1997), 'Influence diagnostics for linear longitudinal models', *Journal of the American Statistical Association* **92**, 999–1005.
- Beckman, R. J., Nachtsheim, C. J. & Cook, R. D. (1987), 'Diagnostics for mixed-model analysis of variance.', *Technometrics* **29**, 413–426.
- Belsley, D., Kugh, E. & Welsch, R. E. (1980), *Regression Diagnostics*, New York: John Wiley and Sons.
- Bickel, P. J. & Doksum, K. A. (1976), *Mathematical Statistics*, Holden-Day.
- Billor, N. & Loynes, R. M. (1993), 'Local influence: A new approach', *Communications in Statistics, Theory and Methods* **22**, 1595–1611.
- Bryk, A. S. & Raudenbush, S. W. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications.

- Christensen, R. & Pearson, L. M. (1992), 'Case-deletion diagnostics for mixed models', *Technometrics* **34**, 38–45.
- Cook, R. D. (1977), 'Detection of influential observations in linear regressions', *Technometrics* **15**, 15–18.
- Cook, R. D. (1986), 'Assessment of local influence', *Journal of the Royal Statistical Society* **48**, 133–169.
- Cook, R. D. (1987), 'Influence assessment', *Journal Applied Statistics* **14**, 117–131.
- 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.
- Cook, R. D. & Weisberg, S. (1982), *Residuals and Influence Regression*, New York: Chapman and Hall.
- Cordeiro, G. M. (1992), *Introdução à teoria da verossimilhança*, 10^o SINAPE, ABE, Rio de Janeiro.
- Cribari Neto, F. & Zarkos, S. G. (1999), 'R: Yet another econometric programming environment', *Journal of Applied Econometrics* **14**, 319–329.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood for incomplete data via the EM algorithm', *Journal of the Royal Statistical Society, Series B* **39**, 1–8.
- Dempster, A. P., Rubin, D. B. & Tsutakawa, R. K. (1981), 'Estimation in covariance components models', *Journal of the American Statistical Association* **76**, 341–353.
- Doganaksoy, N. & Balakrishnan, N. (1997), 'A useful property of best linear unbiased predictors with applications to life-testing', *The American Statistician* **51**, 22–28.
- Fellner, W. H. (1986), 'Robust estimation of variance components', *Technometrics* **28**, 51–60.
- Fung, W. K. & Kwan, C. W. (1997), 'A note on local influence based on normal curvature', *Journal of the Royal Statistical Society, Series B* **59**, 839–843.

- 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* **64**, **B**, 565–579.
- Giampaoli, V. (1999), *Inferência em Modelos Lineares com Restrições nos Parâmetros em Condições Regulares e Não Regulares*, PhD thesis, Universidade de São Paulo.
- Goldstein, H. (1999), *Multilevel Statistical Models*, 1st edn, Institute of Education, London.
- Grenander, U. (1981), *Abstract Inference*, John Wiley & Sons, New York.
- Hilden-Minton, J. A. (1995), *Multilevel Diagnostics for Mixed and Hierarchical Linear Models*, PhD thesis, University of California.
- Hodges, J. S. (1998), ‘Some algebra and geometry for hierarchical models, applied to diagnostics (with discussion)’, *Journal of The Royal Statistical Society, Series B* **60**, 197–536.
- Hox, J. J. (1995), *Applied Multilevel Analysis*, TT - Publikaties.
- Jiang, J. (1997), ‘A derivation of blup - best linear unbiased predictor’, *Statistics & Probability Letters* **32**, 321–324.
- Jones, R. H. (1993), *Longitudinal Data with Serial Correlation: A State-space Approach*, Chapman & Hall, London.
- Kreft, I. & Leeuw, D. J. (1998), *Introducing Multilevel Modeling*, Sage Publications.
- Lamport, L. (1994), *A Document Preparation System*, Addison-Wesley.
- Laros, J. A. & Marciano, J. L. P. (2008), ‘Análise multinível aplicada aos dados do NELS:88’, *Estudos em avaliação educacional* **19**, 40.
- Lesaffre, E. & Verbeke, G. (1998), ‘Local influence in linear mixed models’, *Biometrics* **54**, 570–582.
- Lindley, D. V. & Smith, A. F. M. (1972), ‘Bayes estimates for the linear model’, *Journal of the Royal Statistical Society, Series B* **34**, 1–41.

- Lindstrom, M. J. & 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.
- Lobato Jr., D. (2005), Influência local em modelos de regressão, Master’s thesis, Universidade Federal de Campina Grande.
- Longford, N. T. (1987), ‘A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects’, *Biometrika* **74**, 817–827.
- McCullagh, P. & Nelder, J. A. (1989), *Generalized linear models*, Chapman & Hall, London.
- McCulloch, C. E. & Searle, S. R. (2001), *Generalized, Linear, and Mixed Models*, John Wiley & Sons, New York.
- McLachlan, J. G. & Krishnan, T. (1946), *The EM Algorithm and Extensions*, John Wiley & Sons, New York.
- Montenegro, L. C. C. (2006), Modelo de efeito aleatório e erros de medida, PhD thesis, USP.
- Natis, L. (2000), Modelos lineares hierárquicos, Master’s thesis, Universidade de São Paulo.
- Nobre, J. S. (2004), Métodos de diagnóstico para modelos lineares mistos, Master’s thesis, Universidade de São Paulo.
- Noortgate, W., Opdnakker, M. C. & Onghena, P. (2005), ‘The effects of ignoring a level in multilevel analysis’, *School Effectiveness and School Improvement* **16**, 281–303.
- Paula, G. A. (2004), *Modelos de Regressão - Com apoio computacional*, IME - USP.
- Pinheiro, S. M. C. (2005), Modelo linear hierárquico: Um método alternativo para análise de desempenho escolar, Master’s thesis, Universidade Federal de Pernambuco.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

- Robinson, G. K. (1991), 'That BLUP is a good thing: The estimation of random effects', *Statistical Science* **6**, 15–51.
- Rocke, D. M. (1983), 'Robust statistical analysis of interlaboratory studies', *Biometrics* **70**, 421–431.
- Searle, S. R., Casela, G. & McCulloch, C. E. (1992), *Variance Components*, Wiley, New York.
- Secretaria de Educação e Cultura de Pernambuco (2003), 'Sistema de avaliação educacional de Pernambuco – SAEPE: Relatório 2002'.
- Shi, L. & Chen, G. (2008), 'Local influence in multilevel models', *The Canadian Journal of Statistics* **36**, 259–275.
- Soares, T. M. (2005), 'Modelo de três níveis hierárquicos para a proficiência dos alunos de 4a série avaliados no teste de língua portuguesa do SIMAVE/PROEB-2002', *Revista Brasileira de Educação* **29**, 73–87.
- Sousa, S. O. (2002), Estimação robusta no modelo de calibração, Master's thesis, Universidade Federal de Pernambuco.
- St. Laurent, R. T. & Cook, R. D. (1993), 'Leverage influence and curvature in non-linear regression', *Biometrika* **80**, 99–106.
- Sullivan, L., Dukes, K. & Losina, E. (1999), 'Tutorial in biostatistics an introduction to hierarchical linear modelling', *Statistics in Medicine* **18**, 855–888.
- Tan, F. E. S., Ouwens, M. & Berger, M. P. F. (2001), 'Detection of influential observations in longitudinal mixed effects regression model', *The Statistician* **50**, 271–284.
- Wei, B., Hu, Y. & Fung, W. (1998), 'Generalized leverage and its applications', *Scandinavian Journal of Statistics* **25**, 25–37.
- Wu, X. & Luo, Z. (1993), 'Second-order approach to local influence.', *Journal of Royal Statistical Society* **55**, 929–936.