MODELOS LINEARES HIERÁRQUICOS

Lílian Natis

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

Área de concentração: Estatística

Orientador: **Prof. Dr. Dalton Francisco de Andrade**

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

- São Paulo, Agosto de 2000 -

MODELOS LINEARES HIERÁRQUICOS

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Lílian Natis e aprovada pela comissão julgadora.

São Paulo, 11 de Agosto de 2000.

Comissão Julgadora:

- Prof. Dr. Dalton Francisco de Andrade (Orientador) - UFC
- Profa. Dra. Carmen Diva Saldiva de André - IME - USP
- Prof. Dr. Pedro Luis do Nascimento Silva - IBGE

*"Nada é impossível, algumas coisas
são só menos prováveis que outras."*

Jonathan Winters

*Aos meus pais Alcione e Sérgio
e à minha irmã Élen*

Agradecimentos

São muitas as pessoas que de alguma forma colaboraram para que este trabalho fosse realizado e às quais certamente devo agradecer.

Primeiramente quero agradecer aos meus pais e à minha irmã, pela força e constante incentivo para que eu pudesse alcançar este objetivo. Tudo que sou e conquistei devo a eles, que a vida toda lutaram para sempre me proporcionar o melhor.

Ao Prof. Dalton Francisco de Andrade, pela orientação segura e competente, pela atenção e especialmente pela agradável convivência, amizade e confiança no meu trabalho.

À Secretaria de Estado da Educação do Estado de São Paulo e ao pesquisador Guilherme Veiga Guimarães, do Laboratório de Insuficiência Cardíaca do InCor, que gentilmente cederam os dados de suas pesquisas, e à Fundação Carlos Chagas, em especial à Profa. Yara Lúcia Esposito.

À Profa. Carmen Diva Saldiva de André e ao Prof. Pedro Luis do Nascimento Silva, pela colaboração e contribuição dadas a este trabalho.

Aos meus amigos do mestrado: Adriana, Alessandra, André, Ângela, Carine, Delhi, Esteban, Liliam, Luís, Maria Paula, Mariana, Paulo, Raquel e Rodrigo. Vocês certamente me ensinaram muitas coisas e tornaram esta trajetória muito mais fácil e divertida.

Aos demais amigos e colegas do IME, que me acompanharam desde a graduação.

Agradeço ainda aos professores do IME pela oportunidade de melhorar meus conhecimentos e minha formação.

Resumo

Estruturas hierárquicas de dados são encontradas freqüentemente em diversos estudos. Elas são caracterizadas pela presença de unidades experimentais agrupadas em outras unidades maiores. Para avaliar dados desta natureza podem ser utilizados os *modelos lineares hierárquicos*. Eles constituem uma nova formulação para os modelos de efeitos aleatórios, que permite especificar cada um dos níveis de hierarquia separadamente e incorporar efeitos aleatórios associados a cada um destes níveis. O modelo linear hierárquico com 2 níveis é apresentado e ilustrado através de alguns submodelos, juntamente com a sua teoria de estimação e testes de hipóteses. São apresentadas duas aplicações: uma em Educação e a outra em Medicina, executadas em dois programas computacionais: HLM e SAS. A última aplicação mostra como um estudo de dados longitudinais pode ser tratado sob o enfoque de um modelo linear hierárquico.

Abstract

Hierarchical data structures occur in many studies. They are characterized by the presence of experimental units nested in larger ones. Hierarchical linear models can be used to analyze this type of data. These models constitute a new formulation for random effects models, allowing to specify each level of hierarchy separately as well as to incorporate random effects associated to each level. Two-level hierarchical linear models are introduced and illustrated by some submodels, including the estimation theory and the hypothesis testing. Two applications are presented: the first one in Education and the other one in Medicine, both executed using two software: HLM and SAS. The last application shows how data from a longitudinal study can be treated under the hierarchical linear model approach.

Sumário

1	Motivação para o Estudo do Modelo Linear Hierárquico - MLH	1
2	Introdução ao Modelo Linear Hierárquico	5
2.1	Regressão Linear Simples: um Exemplo	5
2.2	O Modelo Linear Hierárquico com 2 Níveis	6
2.3	Alguns Submodelos	9
2.3.1	ANOVA com 1 Fator e Efeitos Aleatórios	10
2.3.2	Regressão de Médias como Respostas	12
2.3.3	ANCOVA com 1 Fator e Efeitos Aleatórios	12
2.3.4	Modelo de Regressão com Coeficientes Aleatórios	14
2.3.5	Interceptos e Inclinações como Respostas	15
2.3.6	Modelo com Inclinações Variando Não Aleatoriamente	15
2.4	Algumas Considerações	16

3	Escolha da Locação dos Preditores	20
3.1	Caso 1: Modelos EB_1 , GL_1 e GR_1	22
3.2	Caso 2: Modelos Contextuais (EB_2 , GL_2 e GR_2)	24
3.3	Extensão de Alguns Resultados	26
3.4	Interpretação dos Coeficientes da Regressão	28
3.5	Considerações na Escolha da Locação	31
4	Estimação	33
4.1	Introdução ao Modelo Linear Geral de Bayes	33
4.2	Paralelo entre o Modelo Linear Hierárquico com 2 Níveis e o Modelo Linear Geral de Bayes	35
4.2.1	Máxima Verossimilhança Restrita (MVR)	38
4.2.2	Máxima Verossimilhança (MV)	42
5	Testes de Hipóteses	45
5.1	Testes para os Efeitos Fixos	45
5.2	Testes para os Componentes de Variância e Covariância	47
6	Aplicação na Área de Educação	50
6.1	Análise no HLM	51

6.2	Análise no SAS e Aspectos Computacionais	55
7	Aplicação em Dados Longitudinais	58
7.1	Análise no HLM	59
7.2	Análise no SAS	63
8	Conclusão	64
A		66
A.1	Cálculo de $\mathbf{D}^*$	66
A.2	Cálculo de $\boldsymbol{\theta}^*$	68
A.3	Representação de Alguns Vetores e Matrizes	69
A.3.1	Matriz $\mathbf{A}_1$	69
A.3.2	Matriz $\mathbf{A}_2$	69
A.3.3	Matriz $\mathbf{C}$	70
B		71
B.1	Tabela Referente ao Capítulo 6	71
B.2	Tabela Referente ao Capítulo 7	72
C		73

SUMÁRIO

iv

C.1 Programas no SAS, PROC MIXED	73
--	----

Capítulo 1

Motivação para o Estudo do Modelo Linear Hierárquico - MLH

Estruturas hierárquicas de dados são encontradas com frequência em diversos estudos. Elas são caracterizadas pela presença de unidades experimentais agrupadas em outras unidades ainda maiores, que por sua vez também podem ou não formar novos grupos. Muitas vezes esta hierarquia é uma propriedade intrínseca da população de interesse. Para melhor descrever este tipo de estrutura, serão apresentados alguns exemplos.

Em um estudo cujo objetivo é investigar como as características do local de trabalho influem na produtividade dos funcionários, tanto os trabalhadores quanto as empresas selecionadas constituem unidades experimentais. Existe uma estrutura hierárquica, uma vez que os indivíduos estão agrupados em empresas.

Uma determinada característica de interesse é observada em um grupo de pessoas em mais de uma ocasião ao longo do tempo. Temos então um conjunto de medidas associado a cada indivíduo. Podemos imaginar uma situação ainda mais complexa, como por exemplo um grupo de entrevistadores em que cada um deles entrevista pessoas diferentes e cada pessoa é visitada em diversas ocasiões. Nesta situação temos um conjunto de medidas para

cada indivíduo e um grupo de indivíduos associado a cada entrevistador.

Um pesquisador deseja saber como as diferenças na implementação de um tratamento, métodos de pesquisa, características dos indivíduos e aspectos relacionados ao tratamento afetam as estimativas de diversos estudos da mesma natureza (meta-análise). Neste caso, temos um grupo de indivíduos associado a cada estudo.

Em um sistema educacional, deseja-se avaliar a habilidade (ou proficiência) em Língua Portuguesa dos estudantes do primeiro grau. Temos então, um grupo de estudantes associado a cada escola. As escolas poderiam também estar agrupadas em delegacias de ensino e assim por diante.

Todas estas situações ilustram estruturas de dados que podem ser estudadas através dos chamados *modelos lineares hierárquicos* (Bryk e Raudenbush (1992), Goldstein (1995)). Estes modelos permitem que cada um dos níveis desta estrutura hierárquica seja especificado separadamente e que posteriormente sejam reunidos em um único modelo. Além disso permitem a incorporação de efeitos aleatórios associados a cada um de seus níveis de hierarquia.

Existem variáveis que estão diretamente associadas a cada nível de hierarquia. Considere novamente o exemplo do sistema educacional. O MLH permite que variáveis individuais como por exemplo sexo do aluno, nível sócio-econômico, grau de escolaridade da mãe, possam ser incorporadas ao primeiro nível do modelo. Variáveis associadas às escolas a que esses alunos pertencem, como por exemplo se a escola é pública ou privada, se possui ou não biblioteca, a verba disponível por aluno, 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 do modelo e de estudar as possíveis relações entre cada uma destas variáveis explicativas e a resposta.

Teoricamente, cada nível de um MLH é formado por uma amostra aleatória das unidades experimentais consideradas. No caso do sistema educacional, por exemplo, um modelo

linear hierárquico com 2 níveis teria o seu primeiro nível composto por uma amostra de estudantes em cada escola e o segundo nível, por uma amostra de escolas.

Os modelos lineares hierárquicos também representam uma importante alternativa para a análise de dados amostrais complexos (Pessoa e Silva 1998), pois são capazes de incorporar aspectos importantes deste tipo de amostra tais como conglomeração e estratificação. Quando as unidades em qualquer nível de hierarquia são selecionadas com diferentes probabilidades, de formas não consideradas por covariáveis ligadas à estrutura populacional ou ao plano amostral, Peffermann, Skinner, Holmes, Goldstein e Rasbash (1998) apresentam uma maneira de incluir pesos no ajuste de modelos hierárquicos para compensar essas probabilidades distintas.

Os modelos lineares hierárquicos podem ainda ser encontrados na literatura sob outros títulos como *modelos lineares multinível*, *modelos de efeitos mistos*, *modelos de efeitos aleatórios*, *modelos de regressão com coeficientes aleatórios* e *modelos de componentes de variância*. Mas o termo *modelo linear hierárquico* retrata um aspecto importante da estrutura dos dados que é comum em muitas aplicações, como já exposto anteriormente. Este termo surgiu da contribuição na estimação Bayesiana de modelos lineares feita por Lindley e Smith (1972) e Smith (1973). Neste contexto, estes autores elaboraram uma modelagem para dados hierárquicos envolvendo erros com estruturas complexas. Mas a necessidade da estimação de componentes de variância na presença de dados desbalanceados dificultou o uso dos modelos lineares hierárquicos, pois no início dos anos 70 ainda não havia um método de estimação geral aplicável a esses casos. O algoritmo EM proposto em Dempster, Laird e Rubin (1977) abriu um novo caminho para a estimação de componentes de variância e Dempster, Rubin e Tsutakawa (1981) demonstraram que ele poderia ser aplicado a estruturas hierárquicas de dados. Em seguida, outros métodos numéricos para a estimação de componentes de variância foram propostos através do uso de mínimos quadrados generalizados reponderados iterativamente (Goldstein 1986) e do algoritmo *Scoring de Fisher* (Longford 1987). Felizmente começaram a surgir programas computacionais sofisticados que viabilizaram o ajuste desses modelos. Hoje podemos contar com uma série de programas estatísticos que facilitam o uso de modelos lineares hierárquicos. Alguns destes programas serão brevemente comentados

posteriormente.

Os modelos lineares hierárquicos constituem portanto uma nova formulação para os modelos de efeitos aleatórios, incentivada pela própria estrutura dos dados, permitindo a descrição e análise mais apropriada das diferentes fontes de variação (efeitos fixos e aleatórios). Essa nova visão, a maior facilidade de interpretação dos parâmetros do modelo, bem como as inúmeras situações e áreas de aplicação em que os modelos lineares hierárquicos podem ser empregados trazem grande motivação para o estudo mais detalhado destes modelos.

No capítulo 2 os modelos lineares hierárquicos são apresentados de maneira mais formal, ilustrando diferentes formulações matemáticas para alguns casos particulares. O capítulo 3 trata de aspectos relacionados à escolha da localização dos preditores. O capítulo 4 aborda o processo de estimação dos parâmetros do modelo, sob máxima verossimilhança e também sob máxima verossimilhança restrita. O capítulo 5 ilustra como algumas hipóteses de interesse podem ser formuladas e testadas. Os capítulos 6 e 7 apresentam aplicações do MLH em duas diferentes áreas. O primeiro traz uma aplicação do MLH na área de Educação, cujo objetivo é estudar o desempenho em Língua Portuguesa dos alunos de escolas públicas do Estado de São Paulo, segundo características destes alunos e das escolas a que eles pertencem. Já o segundo aborda um problema de dados longitudinais, solucionado sob o enfoque de um MLH. Nestes dois últimos capítulos, as análises são executadas através de 2 programas estatísticos: HLM e SAS.

Capítulo 2

Introdução ao Modelo Linear Hierárquico

Este capítulo aborda algumas idéias de análise de regressão e de análise de variância (Neter, Kutner, Nachtsheim e Wasserman 1996) para introduzir o conceito de modelo linear hierárquico, apresentando o MLH com 2 níveis e alguns de seus principais submodelos, ilustrados através de exemplos.

2.1 Regressão Linear Simples: um Exemplo

Em uma determinada escola foram selecionados n alunos e observadas as seguintes variáveis quantitativas:

- Y_i : aproveitamento em Matemática (AM) do i -ésimo aluno (variável resposta),
- X_i : nível sócio-econômico (NSE) do i -ésimo aluno (variável explicativa), $i = 1, 2, \dots, n$.

O modelo de regressão linear simples para os alunos desta escola é dado por:

$$Y_i = \beta_0 + \beta_1 X_i + r_i \quad (2.1)$$

onde

β_0 : AM esperado para um aluno com $X_i = 0$,

β_1 : mudança esperada no AM quando X_i aumenta em uma unidade,

r_i : erro aleatório associado ao i -ésimo aluno,

suposições: $r_i \sim N(0, \sigma^2)$ e r_i 's independentes.

Selecionando uma outra escola cujos alunos estejam em condições similares aos alunos da escola anterior e propondo um modelo da mesma forma que o da equação (2.1), temos uma outra reta de regressão. Comparando as duas escolas, aquela que possui o maior intercepto é a que proporciona um AM esperado maior para alunos com NSE igual a zero. A que tem a menor inclinação mostra uma relação AM-NSE mais fraca, ou seja, o NSE possui menor influência sobre o AM, o que faz com que a variável NSE seja menos preditiva se comparada à outra escola. Mas o que faz com que uma escola tenha melhor aproveitamento que a outra, para alunos com NSE igual a zero? Por que a relação AM-NSE não é a mesma para as duas escolas? Em outras palavras, por que estas duas escolas apresentam diferentes coeficientes para a reta média ajustada? Se o objetivo não é estudar o AM apenas em uma determinada escola, mas sim estudar o AM dos alunos em uma população de escolas, é possível ajustar um único modelo que leve em conta a variação entre todas as escolas selecionadas e incorpore os diferentes aspectos de cada uma delas e dos alunos que as compõem, fornecendo estimativas mais apropriadas e resultados mais interessantes.

2.2 O Modelo Linear Hierárquico com 2 Níveis

Vamos agora considerar uma população de escolas onde foram amostradas J unidades. Em cada escola foram selecionados n_j alunos.

Através do MLH com 2 níveis todas as escolas selecionadas serão incorporadas ao

modelo, levando em conta a possível variação de interceptos e inclinações entre as escolas.

Temos então:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij} \quad (2.2)$$

com $i = 1, 2, \dots, n_j$ e $j = 1, 2, \dots, J$,

onde

Y_{ij} : AM do i -ésimo aluno da escola j ,

X_{ij} : NSE do i -ésimo aluno da escola j ,

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

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

r_{ij} : erro aleatório associado ao i -ésimo aluno da escola j ,

suposições: $r_{ij} \sim N(0, \sigma^2)$ e r_{ij} 's independentes.

Observe na equação (2.1) que a variável NSE foi utilizada em seu valor original, enquanto que na equação (2.2) ela foi centrada na média global. No capítulo 3 serão discutidos aspectos relacionados à mudança de locação dos preditores do nível 1.

A equação (2.2) é chamada modelo do nível 1 (nível aluno).

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

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} , \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \quad (2.3)$$

onde

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

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

u_{0j} : efeito aleatório da escola j no intercepto β_{0j} ,

u_{1j} : efeito aleatório da escola j na inclinação β_{1j} ,

τ_{00} : variância populacional dos interceptos,

τ_{11} : variância populacional das inclinações,

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

suposições: $u_{0j} \sim N(0, \tau_{00})$, u_{0j} 's independentes,

$u_{1j} \sim N(0, \tau_{11})$, u_{1j} 's independentes
 $Cov(u_{0j}, u_{1j}) = \tau_{01}$ e
 u_{0j} 's e u_{1j} 's independentes dos τ_{ij} 's.

Em resumo:

$$\beta_{0j} \sim N(\gamma_{00}, \tau_{00}), \quad \beta_{1j} \sim N(\gamma_{10}, \tau_{11}) \text{ e } Cov(\beta_{0j}, \beta_{1j}) = \tau_{01}.$$

As equações em (2.3) formam o modelo do nível 2 (nível escola). Ele retrata que as escolas não possuem o mesmo intercepto e a mesma inclinação, e os componentes aleatórios u_{0j} e u_{1j} ajudam a explicar estas diferenças.

Podemos também incluir variáveis relacionadas às escolas no modelo de 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

$$W_j = \begin{cases} 1 & \text{se a escola } j \text{ é privada,} \\ 0 & \text{se a escola } j \text{ é pública.} \end{cases}$$

Um modelo alternativo para o nível 2 pode ser escrito como:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}W_j + u_{1j}. \end{aligned} \tag{2.4}$$

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,
- u_{0j} : efeito aleatório da escola j no intercepto β_{0j} ,
- u_{1j} : efeito aleatório da escola j na inclinação β_{1j} ,
- τ_{00} : variância populacional dos interceptos corrigida pela variável W_j ,
- τ_{11} : variância populacional das inclinações corrigida pela variável W_j ,

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

suposições: $u_{0j} \sim N(0, \tau_{00})$, u_{0j} 's independentes,

$u_{1j} \sim N(0, \tau_{11})$, u_{1j} 's independentes

$Cov(u_{0j}, u_{1j}) = \tau_{01}$ e

u_{0j} 's e u_{1j} 's independentes dos r_{ij} 's.

Sob o novo modelo τ_{00} , τ_{11} e τ_{01} são agora componentes de variância e covariância condicionais (ou residuais), pois a variável W_j está sendo considerada.

Substituindo a equação (2.4) em (2.2) temos a forma combinada do modelo:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij} - \bar{X}_{..}) + \gamma_{11}W_j(X_{ij} - \bar{X}_{..}) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij} . \quad (2.5)$$

O erro aleatório da equação (2.5), $u_{0j} + u_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij}$, é bastante complexo. Dentro de uma mesma escola, tais erros são dependentes, já que u_{0j} e u_{1j} são comuns a todos os alunos desta escola. Além disso estes erros possuem variâncias desiguais, pois dependem dos valores de $(X_{ij} - \bar{X}_{..})$ que variam entre os estudantes.

A aplicação de mínimos quadrados ordinários para estimação dos parâmetros em (2.5) não é apropriada para estruturas de erros deste tipo, pois seria preciso que eles fossem independentes, normalmente distribuídos e com variância constante. Mas é possível estimar tais modelos utilizando procedimentos iterativos de máxima verossimilhança. Observe que se u_{0j} e u_{1j} fossem nulos para todo j , a equação (2.5) poderia ser estimada através de uma regressão de mínimos quadrados ordinários.

2.3 Alguns Submodelos

A partir do MLH com 2 níveis representado pelas equações (2.2) e (2.4), podem ser obtidos submodelos mais simples. Basta que alguns termos destas equações sejam considerados nulos. Nesta seção iremos ilustrar alguns desses submodelos:

- ANOVA com 1 fator e efeitos aleatórios,
- regressão de médias como respostas,
- ANCOVA com 1 fator e efeitos aleatórios,
- modelo de regressão com coeficientes aleatórios,
- interceptos e inclinações como respostas,
- modelo com inclinações variando não aleatoriamente.

Para maiores detalhes sobre regressão linear, análise de variância (ANOVA) e análise de covariância (ANCOVA) clássicas consulte Neter et al. (1996).

2.3.1 ANOVA com 1 Fator e Efeitos Aleatórios

É a estrutura mais simples possível do modelo linear hierárquico com 2 níveis, onde a inclinação β_{1j} no modelo do nível 1 é nula, para todo j .

Temos então o modelo do nível 1:

$$Y_{ij} = \beta_{0j} + r_{ij} \quad (2.6)$$

com

β_{0j} : resposta esperada para a j -ésima escola,
suposições: $r_{ij} \sim N(0, \sigma^2)$ e r_i 's independentes.

E o modelo do nível 2:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2.7)$$

onde

γ_{00} : resposta esperada para a população,
 u_{0j} : efeito aleatório associado à j -ésima escola,

suposições: $u_{0j} \sim N(0, \tau_{00})$, u_{0j} 's independentes e u_{0j} 's independentes dos r_{ij} 's.

Substituindo a equação (2.7) em (2.6) obtemos a forma combinada do modelo:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} . \quad (2.8)$$

A variância da resposta é dada por:

$$Var(Y_{ij}) = var(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2.$$

Esse modelo hierárquico é chamado *totalmente não condicional*, pois tanto o nível 1 quanto o nível 2 não possuem nenhum preditor.

Um parâmetro bastante útil associado à ANOVA com 1 fator e efeitos aleatórios é o *coeficiente de correlação intra-classe*, dado por:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} .$$

Ele representa a proporção da variância da resposta explicada pela variabilidade entre as unidades do nível 2 (escolas).

Ainda sob este modelo, a correlação entre dois indivíduos distintos pertencentes ao mesmo grupo é dada por:

$$Corr(Y_{ij}, Y_{kj}) = \frac{Cov(Y_{ij}, Y_{kj})}{\sqrt{Var(Y_{ij})Var(Y_{kj})}} = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \rho , \quad i \neq k .$$

Então ρ pode ser visto também como uma medida do grau de dependência dos indivíduos de um mesmo grupo ou ainda como uma medida de homogeneidade deste grupo. Espera-se por exemplo, que estudantes de uma mesma escola tenham mais similaridades do que estudantes de escolas diferentes, uma vez que eles dividem as mesmas experiências, o mesmo ambiente etc. Essas idéias são discutidas em Kreft e De Leeuw (1998).

2.3.2 Regressão de Médias como Respostas

Neste submodelo do MLH com 2 níveis, são incorporadas variáveis preditoras no segundo nível do modelo, buscando explicar a variabilidade dos coeficientes β_{0j} entre as unidades do nível 2 (escolas).

O modelo do nível 1 é igual ao caso da ANOVA com 1 fator e efeitos aleatórios, definido em (2.6). Já o coeficiente β_{0j} torna-se a variável resposta de uma regressão linear, cujas variáveis explicativas correspondem a características do grupo j . Por exemplo:

$$Y_{ij} = \beta_{0j} + r_{ij} , \quad (2.9)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \quad (2.10)$$

onde as equações (2.9) e (2.10) correspondem ao modelo do nível 1 e ao modelo do nível 2 respectivamente.

Agora u_{0j} corresponde ao resíduo da regressão no nível 2. Portanto a variância de u_{0j} (τ_{00}) é a variância residual, ou ainda, a variância condicional de β_{0j} corrigida pela variável W_j .

O coeficiente ρ apresentado na subseção 2.3.1 agora é chamado *coeficiente de correlação intra-classe condicional* e continua representando o grau de dependência entre indivíduos de um mesmo grupo, porém corrigido pela variável W_j .

2.3.3 ANCOVA com 1 Fator e Efeitos Aleatórios

A diferença básica entre este modelo e a ANCOVA clássica é a existência de efeitos aleatórios. No modelo do nível 1 são incluídas covariáveis, como na equação (2.2). No modelo do nível 2, temos as equações dadas por (2.3), porém considerando nulos os efeitos aleatórios u_{1j} . Desta forma supõe-se que as inclinações β_{1j} são iguais para todas as escolas. Então, na

Análise de Covariância com 1 fator e efeitos aleatórios, os modelos do nível 1 e do nível 2 podem ser escritos, respectivamente, como:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij} , \quad (2.11)$$

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} , \\ \beta_{1j} &= \gamma_{10} . \end{aligned} \quad (2.12)$$

Agora β_{0j} é a resposta esperada para um indivíduo da j -ésima unidade do nível 2 com $X_{ij} = \bar{X}_{..}$. Tomando a média de (2.11) para toda a amostra de alunos da escola j , obtemos a equação:

$$\bar{Y}_{.j} = \beta_{0j} + \beta_{1j}(\bar{X}_{.j} - \bar{X}_{..}) + \bar{r}_{.j} . \quad (2.13)$$

Considerando apenas o primeiro nível do modelo temos:

$$\begin{aligned} E(\bar{Y}_{.j}) &= \mu_j = \beta_{0j} + \beta_{1j}(\bar{X}_{.j} - \bar{X}_{..}) \Leftrightarrow \\ \beta_{0j} &= \mu_j - \beta_{1j}(\bar{X}_{.j} - \bar{X}_{..}) . \end{aligned}$$

Na verdade, β_{0j} corresponde à resposta esperada da escola j corrigida pelas diferenças em X_{ij} encontradas entre as escolas, ou seja, β_{0j} é uma média ajustada para o grupo j .

Além de covariáveis no nível 1, é possível incluir também covariáveis no nível 2, como por exemplo:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}W_j + u_{0j} , \\ \beta_{1j} &= \gamma_{10} . \end{aligned}$$

A ANCOVA clássica assume que o efeito da covariável é comum a todos os grupos. Essa suposição pode ser facilmente relaxada utilizando modelos cujas inclinações β_{1j} variam aleatoriamente ou não aleatoriamente, como acontece nos submodelos descritos a seguir.

2.3.4 Modelo de Regressão com Coeficientes Aleatórios

Todos os submodelos discutidos anteriormente são exemplos de *modelos com interceptos aleatórios*. As inclinações do nível 1, β_{1j} , só estão presentes na ANCOVA com 1 fator e efeitos aleatórios, mas têm efeito comum a todos os grupos (efeito fixo).

Vamos agora descrever um submodelo onde as inclinações do nível 1 variam aleatoriamente entre as unidades do nível 2. O caso mais simples é o modelo de regressão com efeitos aleatórios, onde interceptos e inclinações variam aleatoriamente sem a necessidade de prever esta variação.

O modelo do nível 1 e o modelo do nível 2, idênticos a (2.2) e (2.3) respectivamente, são dados por :

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

$$\beta_{0j} = \gamma_{00} + u_{0j} ,$$

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

onde

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

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

u_{0j} : efeito aleatório da j -ésima unidade do nível 2 no intercepto β_{0j} e

u_{1j} : efeito aleatório da j -ésima unidade do nível 2 na inclinação β_{1j} .

A matriz de variâncias e covariâncias dos efeitos aleatórios do nível 2 pode ser escrita como:

$$\begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}$$

onde

$\tau_{00} = Var(u_{0j})$: variância não condicional dos interceptos,

$\tau_{11} = Var(u_{1j})$: variância não condicional das inclinações,

$\tau_{01} = Cov(u_{0j}, u_{1j})$: covariância não condicional entre interceptos e inclinações.

Observe que os componentes de variância e covariância são chamados não condicionais, já que o modelo não apresenta nenhum preditor no nível 2.

Na forma combinada do modelo

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{..}) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij} , \quad (2.14)$$

Y_{ij} é uma função da equação de regressão média $\gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{..})$ mais um erro aleatório com 3 componentes: u_{0j} , $u_{1j}(X_{ij} - \bar{X}_{..})$ e r_{ij} .

2.3.5 Interceptos e Inclinações como Respostas

No modelo de interceptos e inclinações como respostas, o objetivo é tentar incorporar variáveis no modelo do nível 2 de forma que elas ajudem a explicar não só a variabilidade dos interceptos, mas também a das inclinações. Para este submodelo, as equações do nível 1 e do nível 2 coincidem com (2.2) e (2.4), respectivamente:

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

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} ,$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j} .$$

2.3.6 Modelo com Inclinações Variando Não Aleatoriamente

Imagine que após incluir uma variável W_j como preditor de β_{1j} , reste pouca ou quase nenhuma variância para ser explicada. Uma vez que a variância residual está bem próxima de zero, é sensato considerar nulos todos os resíduos u_{1j} . Desta forma γ_{11} e γ_{01} serão eliminados e portanto não será necessário estimá-los.

Neste caso a equação que descreve as inclinações β_{1j} no modelo de nível 2 pode ser

escrita como:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j .$$

Observe que as inclinações variam de grupo para grupo, porém sua variação não é aleatória, pois β_{1j} varia estritamente como função de W_j .

2.4 Algumas Considerações

A extensão para modelos com múltiplos preditores em ambos os níveis é bastante simples. As expressões gerais para modelos lineares hierárquicos com 2 níveis, considerando Q preditores no nível 1 e s_q preditores para cada coeficiente β_{qj} da regressão no nível 1 são dadas por

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \dots + \beta_{Qj}X_{Qij} + r_{ij} \quad (\text{nível 1}) \text{ e}$$

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_{1j} + \gamma_{q2}W_{2j} + \dots + \gamma_{qs_q}W_{s_qj} + u_{qj} \quad (\text{nível 2}) ,$$

com $i = 1, 2, \dots, n_j$, $j = 1, 2, \dots, J$ e $q = 0, 1, \dots, Q$.

É importante ressaltar que a inclusão de variáveis explicativas nas equações do modelo do nível 2, com exceção da que representa o coeficiente β_{0j} , resulta no aparecimento de termos de interação entre variáveis dos 2 níveis do modelo.

Outro aspecto importante é a locação dos preditores, tanto do nível 1 quanto do nível 2. Ela pode tornar adequada a interpretação do intercepto da regressão. No nível 1, por exemplo, se utilizarmos a métrica natural de um preditor X_{ij} , talvez a interpretação do intercepto β_{0j} não seja apropriada. Se X_{ij} fosse altura, por exemplo, não faria sentido interpretar β_{0j} como a resposta esperada para um indivíduo do grupo j com altura igual a zero. Seria preciso escolher outra locação, como a média amostral global da variável altura, ou ainda, a média amostral de cada unidade do nível 2. Mas desta forma estaríamos mudando a interpretação de β_{0j} .

No primeiro caso, trata-se da resposta esperada para um indivíduo da j -ésima unidade do nível 2 com $X_{ij} = \bar{X}_{..}$, ou uma média ajustada, como já discutido na subseção 2.3.3. No segundo caso, β_{0j} é a resposta esperada para um indivíduo da j -ésima unidade do nível 2 com $X_{ij} = \bar{X}_{.j}$. Tomando toda a amostra que representa o grupo j , considerando apenas o modelo do nível 1, temos:

$$\begin{aligned} E(\bar{Y}_{.j}) &= \mu_j = \beta_{0j} + \beta_{1j}(\bar{X}_{.j} - \bar{X}_{..}) \Leftrightarrow \\ \beta_{0j} &= \mu_j . \end{aligned}$$

Nesta situação, o intercepto β_{0j} pode ser visto como a verdadeira média do grupo j .

A locação pode ser feita em qualquer outro valor, desde que resulte numa interpretação adequada de β_{0j} . O mesmo cuidado deve ser tomado com relação aos preditores do nível 2, pois a mudança de locação nestas variáveis irá modificar a interpretação dos interceptos deste nível. A mudança de locação algumas vezes ajuda a solucionar possíveis problemas numéricos no processo de estimação, como pode ocorrer por exemplo em curvas de crescimento. Estes e outros aspectos da questão da locação dos preditores serão melhor discutidos no capítulo 3.

Este trabalho ilustra o MLH em 2 níveis. Mas como já citado inicialmente, existem dados que podem estar estruturados em mais de 2 níveis de hierarquia. Se estivéssemos avaliando, por exemplo, alunos agrupados em escolas e estas por sua vez também agrupadas em delegacias de ensino, teríamos um MLH com 3 níveis cuja representação poderia ser feita da seguinte forma:

$$Y_{ijk} = \sum_{p=0}^P (\pi_{pj} Z_{pijk}) + \epsilon_{ijk} \quad (\text{nível 1}),$$

$$\pi_{pj} = \sum_{q=0}^{Q_p} (\beta_{pq} X_{qjk}) + r_{pj} \quad (\text{nível 2}) \text{ e}$$

$$\beta_{pq} = \sum_{s=0}^{S_{pq}} (\gamma_{ps} W_{sk}) + u_{pq} \quad (\text{nível 3}).$$

Note que nesta representação o intercepto foi imbutido na somatória e neste caso $Z_{pijk} = 1$ quando $p = 0$, $X_{qjk} = 1$ quando $q = 0$ e $W_{sk} = 1$ quando $s = 0$.

Definindo brevemente os elementos que compõem as equações dos níveis 1, 2 e 3, temos:

- Y_{ijk} : resposta do i -ésimo aluno, pertencente à j -ésima escola da delegacia k ,
- Z_{pijk} : p -ésima característica do i -ésimo aluno, da escola j e delegacia k (preditor do nível 1), $p = 1, \dots, P$,
- π_{0jk} : intercepto da j -ésima escola da delegacia k ,
- $\pi_{pj k}$: coeficiente que representa o efeito da variável de aluno Z_{pijk} na j -ésima escola da delegacia k , $p = 1, \dots, P$,
- ϵ_{ijk} : erro aleatório que representa o desvio da resposta do aluno ijk em relação ao modelo ajustado no nível 1,

- β_{p0k} : intercepto da delegacia k na modelagem de $\pi_{pj k}$, $p = 0, \dots, P$,
- X_{qjk} : característica da j -ésima escola da delegacia k (preditor do nível 2), $q = 1, \dots, Q_p$,
- β_{pqk} : coeficiente que representa o efeito da variável de escola X_{qjk} em $\pi_{pj k}$, $p = 0, \dots, P$, $q = 1, \dots, Q_p$,
- $r_{pj k}$: efeito aleatório que representa o desvio do coeficiente $\pi_{pj k}$ da escola jk em relação ao modelo ajustado no nível 2, $p = 0, \dots, P$,

- γ_{pq0} : intercepto no modelo do nível 3 para o coeficiente β_{pqk} , $p = 0, \dots, P$, $q = 0, \dots, Q_p$,
- W_{sk} : característica da k -ésima delegacia (preditor do nível 3), $s = 1, \dots, S_{pq}$,
- γ_{pqs} : coeficiente que representa o efeito da variável de delegacia W_{sk} em β_{pqk} , $p = 0, \dots, P$, $q = 0, \dots, P$ e $s = 1, \dots, S_{pq}$,
- u_{pqk} : efeito aleatório que representa o desvio do coeficiente β_{pqk} da delegacia k em relação ao modelo ajustado no nível 3, $p = 0, \dots, P$, $q = 0, \dots, P$.

Este MLH com 3 níveis foi abordado com o intuito de mostrar como expandir o número de níveis de um MLH. Quanto maior o número de níveis de hierarquia, mais complicada se torna a interpretação dos coeficientes do modelo e sua estrutura de variância e covariância.

Algumas vezes o número de unidades experimentais de um dos níveis do modelo pode ser insuficiente para que este nível possa ser levado em conta, uma vez que o número de parâmetros que se deseja estimar cresce com o aumento do número de níveis de hierarquia. Neste caso, uma alternativa para incorporar uma variável associada a este nível é consi-

derá-la em um dos demais níveis, seja repetindo os mesmos valores para diversas unidades experimentais ou utilizando uma medida resumo desta variável, o que podemos chamar de variável agregada.

Considere o MLH com 3 níveis, anteriormente mencionado, com alunos no nível 1, escolas no nível 2 e delegacias no nível 3. Por exemplo: se o número de delegacias for insuficiente para que o nível 3 seja considerado, ou ainda, se não estivermos interessados em estimar a variabilidade de delegacias e sim apenas em estudar uma ou mais de uma de suas características, por exemplo região geográfica de localização, podemos utilizar estas variáveis no nível de escola, repetindo os mesmos valores destas variáveis para as escolas da mesma delegacia. Outra situação: pode-se utilizar uma variável de escola no nível de aluno, ou ainda utilizar uma medida resumo desta variável para as escolas pertencentes à mesma delegacia e incorporá-la ao nível de delegacia. Nestes dois casos estaríamos reduzindo o MLH de 3 para 2 níveis. Vale ressaltar que é preciso ter cuidado com a mudança de interpretação que o efeito da variável pode sofrer quando ela é avaliada em outro nível.

Capítulo 3

Escolha da Localização dos Preditores

A mudança na localização dos preditores em modelos de coeficientes aleatórios pode tornar mais adequada a interpretação dos coeficientes do modelo, bem como solucionar possíveis problemas numéricos no processo de estimação e melhorar aspectos computacionais. Porém esta é uma questão bastante discutida, uma vez que os efeitos causados pela mudança na localização em modelos de coeficientes aleatórios não são os mesmos que ocorrem em modelos de coeficientes fixos, onde a simples adição ou subtração de uma constante modifica apenas o intercepto da regressão enquanto os demais coeficientes e componentes de variância e covariância permanecem os mesmos.

O primeiro aspecto importante é a interpretação do intercepto da regressão. Num modelo cujo preditor é utilizado em sua métrica natural, ou seja, com localização igual a zero, a interpretação do intercepto se resume à resposta esperada de um indivíduo com este preditor igual a zero. Em muitas situações isto pode não fazer sentido e é necessário utilizar alguma outra localização que torne mais apropriado o significado do intercepto.

Os modelos onde os preditores são utilizados em sua métrica natural são chamados de modelos de escores brutos ("*raw scores*"). As duas localizações mais utilizadas são a média global e a média de cada grupo. Neste tópico será apresentada uma comparação entre

modelos de escores brutos (*EB*), modelos com preditores centrados na média global (*GL*) e modelos com preditores centrados na média de cada grupo (*GR*), utilizando o critério de equivalência, segundo Kreft, De Leeuw e Aiken (1995).

De acordo com estes autores, para que dois modelos com distribuição normal sejam equivalentes é necessário que eles tenham o mesmo conjunto de valores esperados e dispersões. É possível que haja equivalência quanto aos valores esperados e não equivalência quanto à matriz de variâncias e covariâncias e vice-versa. Estes modelos podem ter diferentes parametrizações, desde que exista uma transformação (um para um) de uma parametrização para outra. Além disso, ter modelos equivalentes não significa que eles irão produzir exatamente as mesmas estimativas dos parâmetros, pois isto também depende do procedimento de estimação utilizado. Mas no caso da estimação feita através de máxima verossimilhança, espera-se que estas estimativas sejam coincidentes devido à propriedade da invariância dos estimadores de máxima verossimilhança (Degroot 1998).

Em Kreft et al. (1995), foi feita a verificação da equivalência de modelos lineares hierárquicos com 2 níveis:

1. com um único preditor X_{ij} no nível 1 (identificados por EB_1 , GL_1 e GR_1) e
2. com um único preditor X_{ij} no nível 1 e a média do grupo para a variável X_{ij} reintroduzida no nível 2 como preditora do intercepto (EB_2 , GL_2 e GR_2).

No segundo caso, temos os chamados modelos contextuais. Eles podem ser definidos como modelos de regressão onde a mesma variável preditora é apresentada duas vezes: como uma medida individual (preditor do nível 1) e como uma medida do grupo (preditor do nível 2), tal como a média.

3.1 Caso 1: Modelos EB_1 , GL_1 e GR_1

Considere o modelo EB_1 :

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} , \\ \beta_{0j} &= \gamma_{00} + u_{0j} , \\ \beta_{1j} &= \gamma_{10} + u_{1j} . \end{aligned} \quad (3.1)$$

Para este modelo temos:

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00} + \gamma_{10}X_{ij} , \\ Var(Y_{ij}) &= \tau_{00} + 2\tau_{01}X_{ij} + \tau_{11}X_{ij}^2 + \sigma^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00} + \tau_{01}(X_{ij} + X_{kj}) + \tau_{11}X_{ij}X_{kj} , \quad i \neq k . \end{aligned} \quad (3.2)$$

Para simplificar, vamos adotar as notações:

$$X_{ij}^* = X_{ij} - \bar{X}_{..} \quad \text{e} \quad X_{ij}^{**} = X_{ij} - \bar{X}_{.j} . \quad (3.3)$$

De maneira análoga podem ser construídas as esperanças, variâncias e covariâncias dos modelos GL_1 e GR_1 . Temos no modelo GL_1 :

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00}^* + \gamma_{10}^*X_{ij}^* , \\ Var(Y_{ij}) &= \tau_{00}^* + 2\tau_{01}^*X_{ij}^* + \tau_{11}^*(X_{ij}^*)^2 + (\sigma^*)^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00}^* + \tau_{01}^*(X_{ij}^* + X_{kj}^*) + \tau_{11}^*X_{ij}^*X_{kj}^* , \quad i \neq k \end{aligned} \quad (3.4)$$

e no modelo GR_1 :

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00}^{**} + \gamma_{10}^{**}X_{ij}^{**} , \\ Var(Y_{ij}) &= \tau_{00}^{**} + 2\tau_{01}^{**}X_{ij}^{**} + \tau_{11}^{**}(X_{ij}^{**})^2 + (\sigma^{**})^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00}^{**} + \tau_{01}^{**}(X_{ij}^{**} + X_{kj}^{**}) + \tau_{11}^{**}X_{ij}^{**}X_{kj}^{**} , \quad i \neq k . \end{aligned} \quad (3.5)$$

Como apresentado em Kreft et al. (1995), verifica-se que os modelos EB_1 e GL_1 são equivalentes tanto na parte fixa (esperanças) como na parte aleatória (dispersões), pois é

possível expressar todos os parâmetros de um dos modelos em termos dos parâmetros do outro.

Com relação aos parâmetros que envolvem as esperanças destes dois modelos temos:

$$\begin{aligned}\gamma_{00}^* &= \gamma_{00} + \gamma_{10}\bar{X}_{..}, \\ \gamma_{10}^* &= \gamma_{10}.\end{aligned}$$

E para os componentes de variância:

$$\begin{aligned}\tau_{00}^* &= \tau_{00} + 2\tau_{01}\bar{X}_{..} - \tau_{11}\bar{X}_{..}^2, \\ \tau_{01}^* &= \tau_{01} + \tau_{11}\bar{X}_{..}, \\ \tau_{11}^* &= \tau_{11} \\ \sigma^* &= \sigma.\end{aligned}$$

Já entre os modelos EB_1 e GR_1 não há equivalência nem na parte fixa e nem na parte aleatória, pois não existe uma transformação um para um que leve aos parâmetros do outro modelo.

Se isto fosse possível, a esperança da variável resposta Y_{ij} seria a mesma em EB_1 e GR_1 , ou seja:

$$\begin{aligned}\gamma_{00} + \gamma_{10}X_{ij} &= \gamma_{00}^{**} + \gamma_{10}^{**}X_{ij}^{**} && \Leftrightarrow \\ \gamma_{00} + \gamma_{10}X_{ij} &= \gamma_{00}^{**} + \gamma_{10}^{**}X_{ij} - \gamma_{10}^{**}\bar{X}_{.j} && \Leftrightarrow \\ (\gamma_{00} - \gamma_{00}^{**}) + (\gamma_{10} - \gamma_{10}^{**})X_{ij} + \gamma_{10}^{**}\bar{X}_{.j} &= 0.\end{aligned}\tag{3.6}$$

Em geral o vetor de constantes X_{ij} e o vetor com os valores $\bar{X}_{.j}$ são linearmente independentes e então a equação (3.6) não apresenta solução. O mesmo problema ocorre com a parte aleatória, onde é necessária a solução da equação:

$$\tau_{00} + 2\tau_{01}X_{ij} + \tau_{11}X_{ij}^2 = \tau_{00}^{**} + 2\tau_{01}^{**}X_{ij}^{**} + \tau_{11}(X_{ij}^{**})^2.\tag{3.7}$$

Novamente a expansão do lado direito da equação (3.7) apresenta termos como $\bar{X}_{.j}$, $\bar{X}_{.j}^2$ e $X_{ij}\bar{X}_{.j}$, ausentes no lado esquerdo, impossibilitando expressar uma relação única entre os parâmetros dos dois modelos.

Em Kreft et al. (1995) pode ser encontrado um exemplo numérico comparando os resultados de EB_1 , GL_1 e GR_1 . Neste exemplo, as principais diferenças dessas locações em relação ao modelo EB_1 foram encontradas na magnitude do intercepto, da sua variância e da covariância entre o intercepto e a inclinação. Tanto a locação na média global quanto na média do grupo apresentaram uma redução da correlação entre intercepto e inclinação.

3.2 Caso 2: Modelos Contextuais (EB_2 , GL_2 e GR_2)

Os modelos contextuais (Kreft e De Leeuw 1998) podem ser definidos como modelos onde a mesma variável preditora é incluída duas vezes: como uma medida individual (preditor do nível 1) e como uma medida que representa o grupo (preditor do nível 2), como a média. Em algumas situações os grupos podem assumir comportamentos médios diferentes com relação a uma variável preditora X_{ij} e pode existir interesse em estudar o efeito deste comportamento médio do grupo na resposta individual. Por exemplo: o nível-sócio econômico (NSE) individual pode ser preditor do desempenho de um aluno em Matemática. Mas o NSE médio da escola em que ele estuda também pode ter influência no seu desempenho. Então a variável NSE pode estar contribuindo para explicar o desempenho dos alunos através de um componente individual e também de um outro componente que pode estar representando a influência do grupo. Um aluno talvez pudesse melhorar seu desempenho se ele estudasse numa escola cujo NSE médio é mais alto que seu NSE individual. Esta investigação pode ser feita através do uso de modelos contextuais, que além da variável individual X_{ij} como preditora no nível 1, incorporam as médias de cada grupo ($\bar{X}_{.j}$) como preditoras no segundo nível do modelo.

Temos então o modelo contextual EB_2 :

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} , \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}\bar{X}_{.j} + u_{0j} , \\ \beta_{1j} &= \gamma_{10} + u_{1j} . \end{aligned} \tag{3.8}$$

Para este modelo

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00} + \gamma_{01}\bar{X}_{.j} + \gamma_{10}X_{ij} , \\ Var(Y_{ij}) &= \tau_{00} + 2\tau_{01}X_{ij} + \tau_{11}X_{ij}^2 + \sigma^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00} + \tau_{01}(X_{ij} + X_{kj}) + \tau_{11}X_{ij}X_{kj} , \quad i \neq k . \end{aligned} \quad (3.9)$$

Analogamente, podem ser construídas as esperanças, variâncias e covariâncias dos modelos GL_2 e GR_2 , usando a notação de (3.3). Para o modelo contextual GL_2 temos:

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00}^* + \gamma_{01}^*\bar{X}_{.j} + \gamma_{10}^*X_{ij}^* , \\ Var(Y_{ij}) &= \tau_{00}^* + 2\tau_{01}^*X_{ij}^* + \tau_{11}^*(X_{ij}^*)^2 + (\sigma^*)^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00}^* + \tau_{01}^*(X_{ij}^* + X_{kj}^*) + \tau_{11}^*X_{ij}^*X_{kj}^* , \quad i \neq k \end{aligned} \quad (3.10)$$

e para o modelo GR_2 :

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00}^{**} + \gamma_{01}^{**}\bar{X}_{.j} + \gamma_{10}^{**}X_{ij}^{**} , \\ Var(Y_{ij}) &= \tau_{00}^{**} + 2\tau_{01}^{**}X_{ij}^{**} + \tau_{11}^{**}(X_{ij}^{**})^2 + (\sigma^{**})^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00}^{**} + \tau_{01}^{**}(X_{ij}^{**} + X_{kj}^{**}) + \tau_{11}^{**}X_{ij}^{**}X_{kj}^{**} , \quad i \neq k . \end{aligned} \quad (3.11)$$

Com relação aos parâmetros que envolvem as esperanças de EB_2 , GL_2 e GR_2 , temos as seguintes relações:

$$\begin{aligned} \gamma_{00} &= \gamma_{00}^* - \gamma_{10}^*\bar{X}_{..} = \gamma_{00}^{**} , \\ \gamma_{01} &= \gamma_{01}^* = \gamma_{01}^{**} - \gamma_{10}^{**} , \\ \gamma_{10} &= \gamma_{10}^* = \gamma_{10}^{**} . \end{aligned} \quad (3.12)$$

Constatamos que os modelos EB_2 , GL_2 e GR_2 são equivalentes quanto às esperanças. Porém com relação às dispersões ocorre o mesmo que no caso 1: os modelos EB_2 e GL_2 são equivalentes, mas o mesmo não ocorre com os modelos EB_2 e GR_2 . As expressões da variância e da covariância nos modelos EB_2 , GL_2 e GR_2 são idênticas às expressões dadas por (3.2), (3.4) e (3.5) respectivamente. A conclusão é que apenas EB_2 e GL_2 são equivalentes.

Kreft et al. (1995) também comparou EB_1 e EB_2 através de um exemplo, onde as estimativas dos componentes τ_{00} , τ_{11} e τ_{01} em EB_2 apresentaram uma redução em relação

às estimativas de EB_1 . A correlação entre intercepto e inclinação também foi reduzida. Mas um ponto importante foi observado neste exemplo: o coeficiente de X_{ij} em EB_2 é positivo, assim como em EB_1 , porém a média $\bar{X}_{.j}$ introduzida como preditora em EB_2 tem um coeficiente com sinal negativo. Isto revela que uma medida agregada pode produzir um resultado diferente do obtido pela variável individual, pois esta variável pode tomar diferentes significados dependendo do nível em que é analisada. Por exemplo, o NSE médio de uma escola pode ser mais importante no desempenho do aluno que seu NSE individual. A variável NSE no nível aluno pode representar uma medida de recursos financeiros e intelectuais em seu ambiente familiar. Já no nível escola pode ser uma medida de recursos da escola e ambiente normativo. Uma das vantagens do modelo linear hierárquico é que ele permite uma decomposição da relação entre a variável resposta e a variável independente em componentes separados, no primeiro e no segundo nível, se assim for desejado.

Kreft et al. (1995) ressalta que por GR_2 não ser equivalente a EB_2 (e conseqüentemente a GL_2) quanto aos seus componentes de variância e covariância, as variáveis incorporadas ao segundo nível do modelo que não possuem componentes individuais, chamadas variáveis globais, são as mais afetadas pela mudança de locação. Por modificar a interpretação dos coeficientes da regressão, a mudança de locação traz significados diferentes para os parâmetros relacionados a estas variáveis globais que estão sendo utilizadas para explicar a variabilidade dos coeficientes da regressão.

3.3 Extensão de Alguns Resultados

Algumas modificações nos modelos utilizados nos casos 1 e 2 trazem resultados interessantes.

Primeiramente vamos considerar a situação mais simples, onde o coeficiente da variável independente no nível 1 não é aleatório, tanto para o caso 1 quanto para o caso 2.

No caso 1, temos as mesmas equações de (3.1), porém com o erro aleatório $u_{1j} = 0$.

Assim $\tau_{01} = \tau_{11} = 0$ e conseqüentemente $Var(Y_{ij})$ e $Cov(Y_{ij}, Y_{kj})$ não mais dependem dos valores de X_{ij} . Para o modelo EB_1 temos então:

$$\begin{aligned} E(Y_{ij}) &= \gamma_{00} + \gamma_{10}X_{ij} , \\ Var(Y_{ij}) &= \tau_{00} + \sigma^2 , \\ Cov(Y_{ij}, Y_{kj}) &= \tau_{00} , \quad i \neq k . \end{aligned} \quad (3.13)$$

Embora a esperança do modelo GR_1 continue não sendo equivalente às esperanças de EB_1 e GL_1 , agora todos são equivalentes quanto ao seu conjunto de dispersões.

No caso 2, as equações dos modelos contextuais são análogas a (3.8) com $u_{1j} = 0$. Agora $Var(Y_{ij})$ e $Cov(Y_{ij}, Y_{kj})$ são idênticas às de (3.13) e não mais dependem de X_{ij} . Então os modelos EB_2 , GL_2 e GR_2 tornam-se equivalentes tanto na parte fixa quanto na parte aleatória.

Vamos agora abordar uma outra situação mais complexa para o caso 2, incluindo a variável $\bar{X}_{.j}$ como preditora do coeficiente aleatório β_{1j} , ou seja,

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\bar{X}_{.j} + u_{1j} .$$

A esperança para o modelo EB_2 neste caso fica:

$$E(Y_{ij}) = \gamma_{00} + \gamma_{01}\bar{X}_{.j} + \gamma_{10}X_{ij} + \gamma_{11}\bar{X}_{.j}\bar{X}_{ij} .$$

Comparando os modelos EB_2 e GL_2 temos:

$$\begin{aligned} \gamma_{00} &= \gamma_{00}^* - \gamma_{10}^*\bar{X}_{..} , \\ \gamma_{01} &= \gamma_{01}^* - \gamma_{11}^*\bar{X}_{..} , \\ \gamma_{10} &= \gamma_{10}^* , \\ \gamma_{11} &= \gamma_{11}^* . \end{aligned}$$

As expressões para a variância e covariância de EB_2 , GL_2 e GR_2 continuam as mesmas apresentadas em (3.9), (3.10) e (3.11) respectivamente. Portanto os modelos EB_2 e GL_2

continuam sendo equivalentes. Mas a esperança de GR_2 nesta nova situação é dada por:

$$\begin{aligned}
 E(Y_{ij}) &= \gamma_{00}^{**} + \gamma_{01}^{**} \bar{X}_{.j} + \gamma_{10}^{**} X_{ij}^{**} + \gamma_{11}^{**} \bar{X}_{.j} X_{ij}^{**} \\
 &= \gamma_{00}^{**} + \gamma_{01}^{**} \bar{X}_{.j} + \gamma_{10}^{**} (X_{ij} - \bar{X}_{.j}) + \gamma_{11}^{**} \bar{X}_{.j} (X_{ij} - \bar{X}_{.j}) \\
 &= \gamma_{00}^{**} + \gamma_{01}^{**} \bar{X}_{.j} + \gamma_{10}^{**} X_{ij} - \gamma_{10}^{**} \bar{X}_{.j} + \gamma_{11}^{**} \bar{X}_{.j} X_{ij} - \gamma_{11}^{**} \bar{X}_{.j}^2. \quad (3.14)
 \end{aligned}$$

A presença de um termo quadrático em (3.14) revela de imediato que GR_2 não é equivalente nem a EB_2 e nem a GL_2 quando $\bar{X}_{.j}$ também é preditor de β_{1j} .

3.4 Interpretação dos Coeficientes da Regressão

Além da mudança na interpretação do intercepto da regressão no nível 1, a mudança de locação também pode afetar a interpretação de outros coeficientes do modelo. Isto fica mais claro observando as comparações entre algumas alternativas de análises propostas por Bryk e Raudenbush (1992):

1. regressão utilizando os valores individuais da variável resposta e do preditor, sem levar em conta os grupos, que será chamada de *regressão total*;
2. regressão utilizando as médias dos grupos tanto para a variável resposta quanto para o preditor, chamada de *regressão agregada*;
3. MLH com preditor centrado na média global, com intercepto aleatório e inclinação fixa, chamado de *modelo GL*.
4. MLH com preditor centrado na média do grupo, com intercepto aleatório e inclinação fixa, chamado de *modelo GR*;

A seguir temos a tabela apresentada em Bryk e Raudenbush (1992) com o resumo destas comparações:

Tabela 3.1: Comparação do coeficiente de inclinação do nível 1 sob diferentes modelos

<i>Regressão total</i>	$Y_i = \beta_0 + \beta_1 X_i + r_i$ $i : \text{ indivíduos}$	$\beta_1 = \beta_t$
<i>Regressão agregada</i>	$\bar{Y}_{.j} = \beta_0 + \beta_1 \bar{X}_{.j} + u_j$ $j : \text{ grupos}$	$\beta_1 = \beta_b$
<i>Modelo GL</i>	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij}$ $\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10}$	$\gamma_{10} = \frac{P_1\beta_w + P_2\beta_b}{P_1 + P_2}$
<i>Modelo GR</i>	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij}$ $\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10}$	$\gamma_{10} = \beta_w$

onde

β_t : coeficiente da *regressão total*,

β_b : coeficiente da *regressão agregada* (entre grupos),

β_w : coeficiente da regressão dentro dos grupos.

As inclinações obtidas nas diferentes propostas de análise possuem significados diferentes. De acordo com Kreft e De Leeuw (1998)

$$\beta_t = \eta^2(X)\beta_b + (1 - \eta^2(X))\beta_w$$

onde $\eta^2(X)$ corresponde à proporção da variância total do preditor X que é atribuída à variabilidade entre os grupos.

Ao utilizarmos o MLH, a inclinação tem significado diferente dependendo da locação utilizada no preditor. No *modelo GL* ela corresponde a uma média ponderada de β_b e β_w . Como estamos subtraindo um valor comum ($\bar{X}_{..}$) de cada valor individual X_{ij} , estará embutida na estimativa dessa inclinação toda variabilidade de X_{ij} , ou seja, a que ocorre dentro dos grupos e entre os grupos. Já no *modelo GR* a inclinação corresponde a β_w . Ao subtrairmos $\bar{X}_{.j}$ de cada valor individual X_{ij} estaremos eliminando a variabilidade de X_{ij} entre os grupos, restando apenas a variabilidade que ocorre dentro dos grupos. Se for

desejado estimar β_w , o *modelo GR* traz uma estimativa não viciada desse coeficiente.

Se o objetivo for estimar o efeito individual do preditor sobre a variável de interesse e for razoável assumir que a relação entre a variável resposta e o preditor é a mesma tanto dentro dos grupos como entre os grupos ($\beta_w = \beta_b$), então o *modelo GL* será apropriado tanto quanto o *modelo GR*. Mas se $\beta_w \neq \beta_b$, o *modelo GL* não mais atenderá a este objetivo. Se houver interesse em estimar esses dois componentes a sugestão é que $\bar{X}_{.j}$ seja incluída como preditor de β_{0j} (modelo contextual).

Retomando então os modelos contextuais, vimos que os modelos EB_2 , GL_2 e GR_2 são equivalentes quando β_{1j} é fixo. Mas estes modelos apresentam diferenças no cálculo do chamado efeito contextual. Dizemos que o efeito contextual ocorre quando uma medida agregada ($\bar{X}_{.j}$) de características individuais está relacionada à variável resposta mesmo depois de controlar o efeito individual dessa característica. Ele é definido por:

$$\beta_c = \beta_b - \beta_w .$$

Bryk e Raudenbush (1992) ilustra, através da tabela a seguir, como identificar os componentes β_b , β_w e β_c no *modelo GL* e no *modelo GR*, agora com a variável $\bar{X}_{.j}$ incorporada aos interceptos destes dois modelos:

Tabela 3.2: Comparação dos coeficientes β_b , β_w e β_c sob diferentes modelos

<i>Modelo GL</i>	<i>Modelo GR</i>
$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij}$	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij}$
$\beta_{0j} = \gamma_{00} + \gamma_{01}\bar{X}_{.j} + u_{0j}$	$\beta_{0j} = \gamma_{00} + \gamma_{01}\bar{X}_{.j} + u_{0j}$
$\beta_{1j} = \gamma_{10}$	$\beta_{1j} = \gamma_{10}$
$\gamma_{01} = \beta_c$	$\gamma_{01} = \beta_b$
$\gamma_{10} = \beta_w$	$\gamma_{10} = \beta_w$
$\gamma_{01} + \gamma_{10} = \beta_b$	$\gamma_{01} - \gamma_{10} = \beta_c$

Através desta tabela fica ainda mais claro que os coeficientes que desejamos estimar tomam diferentes interpretações dependendo da locação utilizada, por isso é preciso muita cautela na escolha da locação, adotando a mais adequada para os objetivos e características

do estudo.

3.5 Considerações na Escolha da Locação

Não há uma escolha estatisticamente correta entre *EB*, *GL* e *GR*. Como já ressaltado anteriormente, há situações em que a mudança de locação é necessária para que a interpretação do intercepto da regressão faça sentido. Do ponto de vista de estimação, *GL* e *GR* também podem oferecer maior estabilidade e facilidade computacional, por exemplo quando os preditores possuem escalas muito diferentes. Outra vantagem é a redução da multicolinearidade que promove uma melhor estimação.

Em seu artigo Kreft et al. (1995) destaca algumas referências onde os grupos são considerados de grande importância na análise e que trazem à tona algumas considerações interessantes. Coloca-se, por exemplo, que a postura de um indivíduo pode ser resultado de uma atitude individual em direção a um determinado comportamento e também da sua percepção da pressão social e normas que podem encaminhá-lo para este tipo de comportamento. A média do grupo pode funcionar como uma representação destas normas e pressões sociais. A posição de um indivíduo em relação ao seu grupo pode ser determinante no seu comportamento.

Kreft (1998) alerta que subtrair $\bar{X}_{.j}$ do preditor sem que essa média seja reintroduzida no segundo nível do modelo resulta num ajuste mais pobre, pois uma variação importante entre os grupos pode estar sendo eliminada. Quando se emprega esse tipo de modelo, na verdade está se assumindo que a variação entre os grupos não é expressiva e tal suposição necessita de boa argumentação. Sob o modelo GR_1 , dois indivíduos com mesmo valor para o preditor X_{ij} que estejam em grupos com médias $\bar{X}_{.j}$ diferentes não terão a mesma resposta esperada. Mesmo quando se inclui $\bar{X}_{.j}$ no segundo nível do modelo, a escolha entre os modelos *EB*, *GL* ou *GR* requer considerações teóricas. Trabalhar com o desvio em relação à média do grupo faz com que a inclinação esperada do modelo represente o acréscimo (ou decréscimo) na resposta esperada de um indivíduo quando este desvio aumenta em uma

unidade. Fixando um valor X_{ij} , o valor deste desvio irá mudar quando os grupos tiverem médias $\bar{X}_{.j}$ diferentes. A influência da variável individual X_{ij} estará sendo avaliada dentro de um grupo específico.

Portanto, a escolha da locação deve ser feita não só de acordo com as considerações práticas, tais como os aspectos computacionais, mas levando em conta principalmente argumentos teóricos como as questões de interesse, o conhecimento dos dados e os objetivos do estudo.

Capítulo 4

Estimação

O objetivo deste capítulo é descrever o processo de estimação dos efeitos fixos e parâmetros de variância e covariância de um MLH com 2 níveis. Para melhor compreensão deste complexo mecanismo, o primeiro passo será introduzir o *modelo linear geral de Bayes* (Lindley e Smith 1972).

4.1 Introdução ao Modelo Linear Geral de Bayes

Qualquer MLH pode ser visto como um caso particular do modelo linear geral de Bayes (MLGB), que é dado por:

$$\mathbf{Y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{r} \quad , \quad \mathbf{r} \sim N(\mathbf{0}, \boldsymbol{\Psi}) \quad (4.1)$$

onde $\mathbf{Y}$ representa o vetor de respostas, $\mathbf{A}$ é uma matriz de preditores, $\boldsymbol{\theta}$ é o vetor de parâmetros e $\mathbf{r}$ é o vetor de erros. O vetor $\mathbf{r}$ tem como média um vetor nulo e uma matriz de dispersões $\boldsymbol{\Psi}$, positiva definida. Este modelo se diferencia de um modelo linear da estatística clássica porque também é atribuída uma distribuição ao vetor de parâmetros $\boldsymbol{\theta}$. Então temos que:

$$\boldsymbol{\theta} \sim N(\bar{\boldsymbol{\theta}}, \boldsymbol{\Omega}). \quad (4.2)$$

Mais detalhes teóricos sobre estes modelos podem ser encontrados em Lindley e Smith (1972) e em Dempster et al. (1981). Das equações (4.1) e (4.2), temos que a distribuição conjunta de $\mathbf{Y}$ e $\boldsymbol{\theta}$ dados $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$ e $\bar{\boldsymbol{\theta}}$ é uma normal multivariada, ou seja:

$$\begin{bmatrix} \mathbf{Y} \\ \boldsymbol{\theta} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mathbf{A}\bar{\boldsymbol{\theta}} \\ \bar{\boldsymbol{\theta}} \end{bmatrix}, \begin{bmatrix} \mathbf{A}\boldsymbol{\Omega}\mathbf{A}' + \boldsymbol{\Psi} & \mathbf{A}\boldsymbol{\Omega} \\ \boldsymbol{\Omega}\mathbf{A}' & \boldsymbol{\Omega} \end{bmatrix} \right\}.$$

Da teoria da distribuição normal multivariada (Morrison 1967), a distribuição condicional de $\boldsymbol{\theta}$ dados $\mathbf{Y}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$ e $\bar{\boldsymbol{\theta}}$ é uma normal com um vetor de médias $\boldsymbol{\theta}^*$ e matriz de dispersão $\mathbf{D}^*$, onde:

$$\boldsymbol{\theta}^* = \bar{\boldsymbol{\theta}} + \boldsymbol{\Omega}\mathbf{A}'(\mathbf{A}\boldsymbol{\Omega}\mathbf{A}' + \boldsymbol{\Psi})^{-1}(\mathbf{Y} - \mathbf{A}\bar{\boldsymbol{\theta}}) \quad \text{e} \quad (4.3)$$

$$\mathbf{D}^* = \boldsymbol{\Omega} - \boldsymbol{\Omega}\mathbf{A}'(\mathbf{A}\boldsymbol{\Omega}\mathbf{A}' + \boldsymbol{\Psi})^{-1}\mathbf{A}\boldsymbol{\Omega}. \quad (4.4)$$

Utilizando a identidade matricial

$$(\mathbf{A}\boldsymbol{\Omega}\mathbf{A}' + \boldsymbol{\Psi})^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Psi}^{-1}\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Psi}^{-1}$$

nas equações (4.3) e (4.4), pode-se reescrevê-las da seguinte forma:

$$\boldsymbol{\theta}^* = \mathbf{D}^*(\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{\theta}} + \mathbf{A}'\boldsymbol{\Psi}^{-1}\mathbf{Y}) \quad \text{e} \quad (4.5)$$

$$\mathbf{D}^* = (\mathbf{A}'\boldsymbol{\Psi}^{-1}\mathbf{A} + \boldsymbol{\Omega}^{-1})^{-1}. \quad (4.6)$$

Considerando $\mathbf{Y}$ e $\boldsymbol{\theta}$ como variáveis aleatórias e $\boldsymbol{\Psi}$ e $\boldsymbol{\Omega}$ como parâmetros desconhecidos a serem estimados, a função de verossimilhança é dada pela densidade marginal de $\mathbf{Y}$ dados $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$ e $\bar{\boldsymbol{\theta}}$. Assumindo convenientemente $\bar{\boldsymbol{\theta}} = \mathbf{0}$, valor apropriado para o desenvolvimento deste trabalho, esta verossimilhança pode ser escrita como:

$$f(\mathbf{Y}|\boldsymbol{\Psi}, \boldsymbol{\Omega}) = \frac{g(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\Omega})h(\boldsymbol{\theta}|\boldsymbol{\Psi}, \boldsymbol{\Omega})}{p(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\Psi}, \boldsymbol{\Omega})} \quad (4.7)$$

onde f , g , h e p são funções densidades de probabilidades.

Substituindo na equação (4.7) as densidades g , h e p já conhecidas, chega-se à expressão:

$$f(\mathbf{Y}|\boldsymbol{\Psi}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Psi}|^{-1/2}|\boldsymbol{\Omega}|^{-1/2}|\mathbf{D}^*|^{1/2}\exp\left(-\frac{1}{2}\mathbf{Q}^*\right) \quad (4.8)$$

onde $\mathbf{Q}^* = \mathbf{Y}'\boldsymbol{\Psi}^{-1}(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^*)$.

Os passos para obtenção deste resultado estão descritos em Dempster et al. (1981).

4.2 Paralelo entre o Modelo Linear Hierárquico com 2 Níveis e o Modelo Linear Geral de Bayes

Esta seção tem por objetivo mostrar como o MLH com 2 níveis pode ser reescrito na forma do MLGB, uma vez que os resultados do processo de estimação podem ser obtidos através das equações (4.5), (4.6) e (4.8).

Cada uma das J unidades do nível 2 pode ser representada através do seguinte modelo do nível 1:

$$\mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{r}_j \quad (4.9)$$

onde $\mathbf{Y}_j$ é o vetor de respostas do grupo j , $\mathbf{X}_j$ é a matriz de preditores do nível 1 para as unidades do grupo j , $\boldsymbol{\beta}_j$ é o vetor de coeficientes do nível 1 para o grupo j e $\mathbf{r}_j$ é o vetor de erros, sendo $\mathbf{r}_j \sim N(\mathbf{0}, \boldsymbol{\Psi}_j)$, $j = 1, 2, \dots, J$.

Representando na forma matricial temos:

$$\mathbf{Y}_j = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{n_j j} \end{bmatrix}, \quad \boldsymbol{\beta}_j = \begin{bmatrix} \beta_{1j} \\ \beta_{2j} \\ \vdots \\ \beta_{Qj} \end{bmatrix}, \quad \mathbf{r}_j = \begin{bmatrix} r_{1j} \\ r_{2j} \\ \vdots \\ r_{n_j j} \end{bmatrix},$$

$$\mathbf{X}_j = \begin{bmatrix} X_{11j} & X_{21j} & \cdots & X_{Q1j} \\ X_{12j} & X_{22j} & \cdots & X_{Q2j} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n_j j} & X_{2n_j j} & \cdots & X_{Qn_j j} \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Psi}_j = \sigma^2 I_{n_j}.$$

As J unidades do nível 2 podem então ser englobadas em um único modelo do nível 1:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r} \quad , \quad \mathbf{r} \sim N(\mathbf{0}, \boldsymbol{\Psi}) \quad (4.10)$$

onde

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_J \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_J \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_J \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_J \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Psi}_J \end{bmatrix}.$$

Cada vetor de coeficientes β_j pode ser escrito como:

$$\beta_j = \mathbf{W}_j \boldsymbol{\gamma} + \mathbf{u}_j \quad (4.11)$$

onde $\boldsymbol{\gamma}$ é o vetor de efeitos fixos, $\mathbf{W}_j$ é a matriz de preditores do nível 2 e $\mathbf{u}_j$ é o vetor de efeitos aleatórios, cuja média é um vetor nulo e matriz de dispersões $\boldsymbol{\tau}$, $j = 1, 2, \dots, J$.

Expressando na forma matricial:

$$\mathbf{W}_j = \begin{bmatrix} \underbrace{W_{1j}^{(1)} \ W_{2j}^{(1)} \ \cdots \ W_{s_1j}^{(1)}}_{\text{preditores de } \beta_{1j}} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \underbrace{W_{1j}^{(2)} \ W_{2j}^{(2)} \ \cdots \ W_{s_2j}^{(2)}}_{\text{preditores de } \beta_{2j}} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & & \ddots & & \vdots & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & \underbrace{W_{1j}^{(Q)} \ W_{2j}^{(Q)} \ \cdots \ W_{s_Qj}^{(Q)}}_{\text{preditores de } \beta_{Qj}} \end{bmatrix},$$

onde cada $W_{kj}^{(q)}$ representa um preditor de β_{qj} , $k = 1, \dots, s_q$, $q = 1, \dots, Q$,

$$\boldsymbol{\gamma} = \left[\underbrace{\gamma_{11} \ \gamma_{12} \ \cdots \ \gamma_{1s_1}}_{\text{parâmetros de } \beta_{1j}} \ \underbrace{\gamma_{21} \ \gamma_{22} \ \cdots \ \gamma_{2s_2}}_{\text{parâmetros de } \beta_{2j}} \ \cdots \ \underbrace{\gamma_{Q1} \ \gamma_{Q2} \ \cdots \ \gamma_{Qs_Q}}_{\text{parâmetros de } \beta_{Qj}} \right]',$$

$$\mathbf{u}_j = \begin{bmatrix} u_{1j} \\ u_{2j} \\ \vdots \\ u_{Qj} \end{bmatrix} \quad \text{e} \quad \boldsymbol{\tau} = \begin{bmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1Q} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{Q1} & \tau_{Q2} & \cdots & \tau_{QQ} \end{bmatrix}.$$

Novamente as J unidades do nível 2 podem ser representadas em um único modelo do nível 2:

$$\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\gamma} + \mathbf{u}, \quad \mathbf{u} \sim N(\mathbf{0}, \mathbf{T}) \quad (4.12)$$

onde

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_J \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_J \end{bmatrix} \quad \text{e} \quad \mathbf{T} = \begin{bmatrix} \tau & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tau & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tau \end{bmatrix}.$$

As equações (4.10) e (4.12) resultam na forma combinada do modelo

$$\mathbf{Y} = \mathbf{XW}\boldsymbol{\gamma} + \mathbf{Xu} + \mathbf{r} \quad (4.13)$$

que pode ser reescrita como:

$$\mathbf{Y} = \mathbf{A}_1\boldsymbol{\theta}_1 + \mathbf{A}_2\boldsymbol{\theta}_2 + \mathbf{r} \quad (4.14)$$

fazendo $\mathbf{A}_1 = \mathbf{XW}$, $\mathbf{A}_2 = \mathbf{X}$, $\boldsymbol{\theta}_1 = \boldsymbol{\gamma}$ e $\boldsymbol{\theta}_2 = \mathbf{u}$.

Caso alguns elementos de $\mathbf{u}$ sejam considerados nulos (quando o coeficiente β_{qj} é considerado fixo), $\boldsymbol{\theta}_2$ passa a ser o subconjunto não nulo de $\mathbf{u}$ e $\mathbf{A}_2$ o correspondente subconjunto de $\mathbf{X}$.

A teoria de estimação dos parâmetros de interesse será abordada sob a formulação de máxima verossimilhança restrita (MVR) e máxima verossimilhança (MV). Sob MVR é imposta uma distribuição a priori para $\boldsymbol{\theta}_1$, enquanto que sob MV este vetor é considerado fixo e desconhecido. De acordo com Bryk e Raudenbush (1992), para o MLH com 2 níveis, as duas abordagens em geral produzem resultados similares para σ^2 . Porém podem ocorrer diferenças importantes na estimação da matriz $\mathbf{T}$. Quando o número de unidades nível 2

(J) é grande, as duas abordagens tendem a produzir resultados parecidos. Mas quando J é pequeno, as estimativas dos componentes de variância τ_{qq} sob MV são menores que sob MVR.

4.2.1 Máxima Verossimilhança Restrita (MVR)

Para realizar o processo de estimação sob MVR, assume-se uma distribuição a priori não informativa para os efeitos fixos da equação (4.14), fazendo $\theta_1 \sim N(\mathbf{0}, \Gamma)$. Assumindo que $\Gamma^{-1} \rightarrow \mathbf{0}$, a dispersão a priori Γ será considerada arbitrariamente grande. Desta forma, é conveniente e sensato tomar $\mathbf{0}$ como média a priori de θ_1 .

Novamente temos o modelo

$$\mathbf{Y} = \mathbf{A}_1\theta_1 + \mathbf{A}_2\theta_2 + \mathbf{r},$$

onde são feitas as seguintes considerações:

$$\theta_1 \sim N(\mathbf{0}, \Gamma), \quad \theta_2 \sim N(\mathbf{0}, \mathbf{T}) \quad \text{e} \quad \mathbf{r} \sim N(\mathbf{0}, \Psi).$$

Temos claramente um caso particular do MLGB, onde:

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad \bar{\theta} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{A} = [\mathbf{A}_1 \mid \mathbf{A}_2],$$

$$\Omega = \begin{bmatrix} \Gamma & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{bmatrix} \quad \text{e} \quad \Psi = \sigma^2 \mathbf{I},$$

Logo, através das equações (4.5) e (4.6), podemos obter a esperança e a variância a posteriori de $\theta_1, \theta_2 | \mathbf{Y}, \sigma^2, \Gamma, \mathbf{T}$ expressas respectivamente por:

$$\theta^* = \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{11}\mathbf{A}_1'(\mathbf{I} - \mathbf{A}_2\mathbf{C}^{-1}\mathbf{A}_2')\mathbf{Y} \\ \mathbf{C}^{-1}\mathbf{A}_2'(\mathbf{Y} - \mathbf{A}_1\theta_1^*) \end{bmatrix}, \quad (4.15)$$

$$\mathbf{D}^* = \begin{bmatrix} \mathbf{D}_{11}^* & \mathbf{D}_{12}^* \\ \mathbf{D}_{21}^* & \mathbf{D}_{22}^* \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \quad (4.16)$$

onde

$$\begin{aligned} \mathbf{C} &= \mathbf{A}_2' \mathbf{A}_2 + \sigma^2 \mathbf{T}^{-1}, \\ \mathbf{V}_{11} &= (\mathbf{A}_1' \mathbf{A}_1 - \mathbf{A}_1' \mathbf{A}_2 \mathbf{C}^{-1} \mathbf{A}_2' \mathbf{A}_1)^{-1}, \\ \mathbf{V}_{12} &= -\mathbf{V}_{11} \mathbf{A}_1' \mathbf{A}_2 \mathbf{C}^{-1}, \\ \mathbf{V}_{21} &= \mathbf{V}_{12}' \quad \text{e} \\ \mathbf{V}_{22} &= \mathbf{C}^{-1} + \mathbf{C}^{-1} \mathbf{A}_2' \mathbf{A}_1 \mathbf{V}_{11} \mathbf{A}_1' \mathbf{A}_2 \mathbf{C}^{-1}. \end{aligned}$$

Os cálculos de $\mathbf{D}^*$ e $\boldsymbol{\theta}^*$ podem ser encontrados com mais detalhes em A.1 e A.2 do apêndice A, respectivamente.

Vamos agora reescrever algumas expressões, devolvendo o índice j , de forma que os cálculos sejam baseados nas estatísticas suficientes:

$$\begin{aligned} \mathbf{S}_{11j} &= \mathbf{A}_{1j}' \mathbf{A}_{1j}, \quad \mathbf{S}_{12j} = \mathbf{A}_{1j}' \mathbf{A}_{2j}, \quad \mathbf{S}_{21j} = \mathbf{S}_{12j}', \quad \mathbf{S}_{22j} = \mathbf{A}_{2j}' \mathbf{A}_{2j}, \\ \mathbf{S}_{1yj} &= \mathbf{A}_{1j}' \mathbf{Y}_j \quad \text{e} \quad \mathbf{S}_{2yj} = \mathbf{A}_{2j}' \mathbf{Y}_j. \end{aligned}$$

Estas expressões são dadas por:

$$\mathbf{C}_j = \mathbf{S}_{22j} + \sigma^2 \boldsymbol{\tau}^{-1}, \quad (4.17)$$

$$\mathbf{V}_{11} = \left[\sum \mathbf{S}_{11j} - \sum \mathbf{S}_{12j} \mathbf{C}_j^{-1} \mathbf{S}_{21j} \right]^{-1}, \quad (4.18)$$

$$\mathbf{D}_{11}^* = \sigma^2 \mathbf{V}_{11} \quad (4.19)$$

$$\mathbf{D}_{12j}^* = -\mathbf{D}_{11}^* \mathbf{S}_{12j} \mathbf{C}_j^{-1}, \quad (4.20)$$

$$\mathbf{V}_{22j} = \mathbf{C}_j^{-1} + \mathbf{C}_j^{-1} \mathbf{S}_{21j} \mathbf{V}_{11} \mathbf{S}_{12j} \mathbf{C}_j^{-1}, \quad (4.21)$$

$$\boldsymbol{\theta}_1^* = \mathbf{V}_{11} \left[\sum \mathbf{S}_{1yj} - \sum \mathbf{S}_{12j} \mathbf{C}_j^{-1} \mathbf{S}_{2yj} \right] \quad (4.22)$$

$$\boldsymbol{\theta}_{2j}^* = \mathbf{C}_j^{-1} (\mathbf{S}_{2yj} - \mathbf{S}_{21j}) \boldsymbol{\theta}_1^*. \quad (4.23)$$

Mais detalhes sobre a representação de alguns destes componentes estão em A.3 do apêndice A.

Retomando a equação (4.8), vamos escrever o $\log$ da função de verossimilhança para o modelo de 2 níveis no caso da MVR. Verifica-se que

$$\begin{aligned} \log|\Psi| &= N\log(\sigma^2), \quad \text{com } N = \sum_{j=1}^J n_j \quad \text{e} \\ \log|\Omega| &\propto \log|\Gamma| + \log|\mathbf{T}|. \end{aligned}$$

Assumindo que $|\Gamma|$, arbitrariamente grande, é uma constante conhecida, então

$$\log|\Omega| \propto \log|\mathbf{T}| = J\log|\tau|.$$

Calculando o determinante de uma matriz particionada, onde F é o número de efeitos fixos do modelo e R corresponde ao número de efeitos aleatórios para cada unidade j ($R \leq Q$), temos:

$$\begin{aligned} |\mathbf{D}^*| &= |\mathbf{D}_{11}^*| |\mathbf{D}_{22}^* - \mathbf{D}_{21}^* \mathbf{D}_{11}^{*-1} \mathbf{D}_{12}^*| \\ &= \sigma^{2F} |\mathbf{V}_{11}| \sigma^{2JR} |\mathbf{V}_{22} - \mathbf{V}_{21} \mathbf{V}_{11}^{-1} \mathbf{V}_{12}|, \end{aligned}$$

de forma que $\log|\mathbf{D}^*| = (F + JR)\log(\sigma^2) + \log|\mathbf{V}_{11}| + \sum \log|\mathbf{C}_j^{-1}|$,

resultando em:

$$\begin{aligned} \log[f(\mathbf{Y}|\sigma^2, \tau)] &\propto -(N - JR - F)\log(\sigma^2) - J\log|\tau| + \log|\mathbf{V}_{11}| \\ &\quad + \sum \log|\mathbf{C}_j^{-1}| - \sum Q_j^*, \end{aligned} \quad (4.24)$$

onde $Q_j^* = \mathbf{Y}_j'(\mathbf{Y}_j - \mathbf{A}_{1j}\boldsymbol{\theta}_1^* - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}^*)/\sigma^2$.

O algoritmo EM pode ser utilizado para maximizar a equação (4.24) (ver Dempster et al. (1981)). Suponha que os efeitos desconhecidos $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$, além dos dados $\mathbf{Y}$, tivessem sido observados. Neste caso $\mathbf{Y}$, $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ são então chamados *dados completos*. O log da função de verossimilhança dos dados completos, dado por

$$\begin{aligned} \log[f(\mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\sigma^2, \tau)] &\propto -N\log(\sigma^2) - J\log|\tau| \\ &\quad - \sum \mathbf{r}_j' \mathbf{r}_j / \sigma^2 - \sum \boldsymbol{\theta}_{2j}' \tau^{-1} \boldsymbol{\theta}_{2j}, \end{aligned} \quad (4.25)$$

deve ser maximizado com relação aos parâmetros σ^2 e τ computando as estimativas:

$$\hat{\sigma}^2 = \sum \mathbf{r}'_j \mathbf{r}_j / N \quad \text{e} \quad (4.26)$$

$$\hat{\tau} = J^{-1} \sum \boldsymbol{\theta}_{2j} \boldsymbol{\theta}'_{2j}. \quad (4.27)$$

Como $\mathbf{r}_j$ e $\boldsymbol{\theta}_{2j}$ não foram observados para que estas estimativas pudessem ser diretamente calculadas, então as estatísticas suficientes $\sum \mathbf{r}'_j \mathbf{r}_j$ e $\sum \boldsymbol{\theta}_{2j} \boldsymbol{\theta}'_{2j}$ são substituídas por suas esperanças condicionais, dados $\mathbf{Y}$ e valores iniciais para σ^2 e τ .

Temos então:

$$\begin{aligned} E \left(\sum \mathbf{r}'_j \mathbf{r}_j | \mathbf{Y}, \sigma^2, \tau \right) &= E[(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta})'(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}) | \mathbf{Y}, \sigma^2, \tau] \\ &= E[(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^* + \mathbf{A}\boldsymbol{\theta}^* - \mathbf{A}\boldsymbol{\theta})'(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^* + \mathbf{A}\boldsymbol{\theta}^* - \mathbf{A}\boldsymbol{\theta}) | \mathbf{Y}, \sigma^2, \tau] \\ &= (\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^*)'(\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^*) + E[(\boldsymbol{\theta} - \boldsymbol{\theta}^*)' \mathbf{A}' \mathbf{A} (\boldsymbol{\theta} - \boldsymbol{\theta}^*) | \mathbf{Y}, \sigma^2, \tau] \end{aligned} \quad (4.28)$$

onde

$$\mathbf{Y} - \mathbf{A}\boldsymbol{\theta}^* = \mathbf{Y} - \mathbf{A}_1 \boldsymbol{\theta}_1^* - \mathbf{A}_2 \boldsymbol{\theta}_2^* \quad \text{e}$$

$$\begin{aligned} E[(\boldsymbol{\theta} - \boldsymbol{\theta}^*)' \mathbf{A}' \mathbf{A} (\boldsymbol{\theta} - \boldsymbol{\theta}^*) | \mathbf{Y}, \sigma^2, \tau] &= \text{tr}(\mathbf{A}' \mathbf{A} \mathbf{D}^*) \\ &= \text{tr} \begin{bmatrix} \mathbf{A}'_1 \mathbf{A}_1 & \mathbf{A}'_1 \mathbf{A}_2 \\ \mathbf{A}'_2 \mathbf{A}_1 & \mathbf{A}'_2 \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{11}^* & \mathbf{D}_{12}^* \\ \mathbf{D}_{21}^* & \mathbf{D}_{22}^* \end{bmatrix}. \end{aligned} \quad (4.29)$$

A equação (4.29) pode ser escrita como:

$$\sigma^2 \left[F - \text{tr} \left(\mathbf{V}_{11} \sum \mathbf{S}_{12j} \mathbf{C}_j^{-1} \mathbf{S}_{21j} \right) + \text{tr} \left(\sum \mathbf{V}_{22j} \mathbf{S}_{22j} \right) \right].$$

A esperança da outra estatística suficiente mencionada pode ser calculada como:

$$E \left(\sum \boldsymbol{\theta}_{2j} \boldsymbol{\theta}'_{2j} | \mathbf{Y}, \sigma^2, \tau \right) = \sum \boldsymbol{\theta}_{2j}^* \boldsymbol{\theta}_{2j}^{*'} + \sigma^2 \sum \mathbf{V}_{22j}. \quad (4.30)$$

A partir de valores iniciais para σ^2 e τ , inicia-se o algoritmo EM, calculando as esperanças condicionais das estatísticas suficientes (o chamado passo E do algoritmo EM, ou esperança). Em seguida é realizado o passo M do algoritmo (ou maximização), através das

equações (4.26) e (4.27). Com os novos valores de σ^2 e τ , passa-se à próxima iteração do algoritmo EM, repetindo os passos E e M, até que a diferença entre a log verossimilhança das últimas duas iterações seja suficientemente pequena ou até que seja atingido algum outro critério de convergência. A teoria geral do EM garante que a verossimilhança não diminui a cada iteração.

Após a estimação dos componentes de variância e covariância sob MVR, a estimação dos efeitos fixos pode ser feita através do método de mínimos quadrados generalizados.

4.2.2 Máxima Verossimilhança (MV)

Na formulação de MVR, $\mathbf{Y}$, $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ são tratados como aleatórios, uma vez que uma distribuição é atribuída a cada um deles, enquanto que σ^2 e $\mathbf{T}$ são considerados fixos. Já na formulação de MV, somente $\mathbf{Y}$ e $\boldsymbol{\theta}_2$ são tratados como aleatórios. Desta forma a verossimilhança dependerá das quantidades fixas $\boldsymbol{\theta}_1$, σ^2 e $\mathbf{T}$.

Considere novamente o MLGB

$$\mathbf{Y} = \mathbf{A}_1\boldsymbol{\theta}_1 + \mathbf{A}_2\boldsymbol{\theta}_2 + \mathbf{r}$$

onde são feitas as suposições:

$$\boldsymbol{\theta}_2 \sim N(\mathbf{0}, \mathbf{T}) \quad \text{e} \quad \mathbf{r} \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Fazendo $\mathbf{d} = \mathbf{Y} - \mathbf{A}_1\boldsymbol{\theta}_1$, podemos reescrever este modelo como:

$$\mathbf{d} = \mathbf{A}_2\boldsymbol{\theta}_2 + \mathbf{r}. \tag{4.31}$$

A equação (4.31) é equivalente à equação (4.1), com

$$\mathbf{Y} = \mathbf{d}, \quad \mathbf{A} = \mathbf{A}_2, \quad \boldsymbol{\theta} = \boldsymbol{\theta}_2, \quad \boldsymbol{\Psi} = \sigma^2\mathbf{I} \quad \text{e} \quad \boldsymbol{\Omega} = \mathbf{T}.$$

A partir das equações (4.5) e (4.6), obtém-se a média e a variância a posteriori de $\boldsymbol{\theta}_2$, dadas respectivamente por:

$$\boldsymbol{\theta}_2^* = \mathbf{C}^{-1}\mathbf{A}_2'\mathbf{d} \quad \text{e} \tag{4.32}$$

$$\mathbf{D}_{22}^* = \sigma^2 \mathbf{C}^{-1} \quad (4.33)$$

onde $\mathbf{C}^{-1} = (\mathbf{A}_2' \mathbf{A}_2 + \sigma^2 \mathbf{T}^{-1})^{-1}$.

Retomando a equação (4.8) e procedendo de maneira análoga ao que foi feito no caso da MVR, após algumas modificações podemos escrever o $\log$ da função de verossimilhança para o modelo de 2 níveis sob a formulação de MV:

$$\begin{aligned} \log[f(\mathbf{Y}|\boldsymbol{\theta}_1, \sigma^2, \boldsymbol{\tau})] &\propto -(N - JR)\log(\sigma^2) - J\log|\boldsymbol{\tau}| \\ &+ \sum \log|\mathbf{C}_j^{-1}| - \sum \mathbf{d}_j'(\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}^*)/\sigma^2. \end{aligned} \quad (4.34)$$

Novamente o algoritmo EM pode ser empregado para maximizar a equação (4.34). Agora $\mathbf{Y}$ e $\boldsymbol{\theta}_2$ é que representam os *dados completos*. O log da função de verossimilhança dos dados completos, dado por

$$\begin{aligned} \log[f(\mathbf{Y}, \boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \sigma^2, \boldsymbol{\tau})] &\propto -N\log(\sigma^2) - J\log|\boldsymbol{\tau}| \\ &- \sum (\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j})'(\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j})/\sigma^2 - \sum \boldsymbol{\theta}_{2j}'\boldsymbol{\tau}^{-1}\boldsymbol{\theta}_{2j}, \end{aligned} \quad (4.35)$$

deve ser maximizado com relação aos parâmetros $\boldsymbol{\theta}_1$, σ^2 e $\boldsymbol{\tau}$ computando as estimativas:

$$\hat{\boldsymbol{\theta}}_1 = \left(\sum \mathbf{A}_{1j}' \mathbf{A}_{1j} \right)^{-1} \sum \mathbf{A}_{1j}' (\mathbf{Y}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}), \quad (4.36)$$

$$\hat{\sigma}^2 = \sum (\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j})'(\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j})/N \quad \text{e} \quad (4.37)$$

$$\hat{\boldsymbol{\tau}} = J^{-1} \sum \boldsymbol{\theta}_{2j}\boldsymbol{\theta}_{2j}'. \quad (4.38)$$

No passo E do algoritmo EM são então computadas as esperanças condicionais das estatísticas suficientes:

$$\begin{aligned} E \left[\left(\sum \mathbf{A}_{1j}' \mathbf{A}_{1j} \right)^{-1} \sum (\mathbf{Y}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}) | \mathbf{Y}, \boldsymbol{\theta}_1, \sigma^2, \boldsymbol{\tau} \right] \\ = \left(\sum \mathbf{A}_{1j}' \mathbf{A}_{1j} \right)^{-1} \sum (\mathbf{Y}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}^*), \end{aligned} \quad (4.39)$$

$$\begin{aligned} E \left[\sum (\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j})'(\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}) | \mathbf{Y}, \boldsymbol{\theta}_1, \sigma^2, \boldsymbol{\tau} \right] \\ = \sum (\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}^*)'(\mathbf{d}_j - \mathbf{A}_{2j}\boldsymbol{\theta}_{2j}^*) + \sigma^2 \text{tr} \left[\sum \mathbf{A}_{2j}' \mathbf{A}_{2j} \mathbf{C}_j^{-1} \right] \quad \text{e} \end{aligned} \quad (4.40)$$

$$E \left(\sum \boldsymbol{\theta}_{2j} \boldsymbol{\theta}_{2j}' | \mathbf{Y}, \boldsymbol{\theta}_1, \sigma^2, \boldsymbol{\tau} \right) = \sum \boldsymbol{\theta}_{2j}^* \boldsymbol{\theta}_{2j}^{*'} + \sigma^2 \sum \mathbf{C}_j^{-1}. \quad (4.41)$$

A partir de valores iniciais para $\boldsymbol{\theta}_1$, σ^2 e $\boldsymbol{\tau}$, inicia-se o algoritmo EM. Alternando entre o passo E e o passo M, repete-se o procedimento até que se atinja a convergência, resultando nas estimativas finais de máxima verossimilhança dos parâmetros $\boldsymbol{\theta}_1$, σ^2 e $\boldsymbol{\tau}$.

Capítulo 5

Testes de Hipóteses

Neste capítulo serão abordados testes de hipóteses para os efeitos fixos e também para os componentes de variância e covariância de um MLH. Serão apresentados testes para hipóteses que envolvem um único parâmetro e também testes para múltiplos parâmetros.

5.1 Testes para os Efeitos Fixos

Uma das primeiras coisas que se pode querer investigar em um modelo linear hierárquico é se cada um de seus efeitos fixos estimados são significativamente diferentes de zero. Para testar um único efeito fixo, comumente temos a hipótese nula:

$$H_0 : \gamma_{qs} = 0. \quad (5.1)$$

Esta hipótese pode ser testada através da estatística:

$$z = \frac{\hat{\gamma}_{qs}}{(\hat{V}_{\hat{\gamma}_{qs}})^{1/2}} \quad (5.2)$$

onde $\hat{\gamma}_{qs}$ é o estimador do efeito fixo γ_{qs} e $\hat{V}_{\hat{\gamma}_{qs}}$ é a estimativa da variância de $\hat{\gamma}_{qs}$. Na formulação de MV, $\hat{\gamma}_{qs}$ corresponde ao estimador de máxima verossimilhança e $\hat{V}_{\hat{\gamma}_{qs}}$ pode ser

obtida através da inversa da matriz de informação de Fisher. Já sob MVR, $\hat{\gamma}_{qs}$ corresponde ao estimador de mínimos quadrados generalizados (MQG) e $\hat{V}_{\hat{\gamma}_{qs}}$ pode ser obtida a partir da expressão da variância do estimador de MQG, em que os componentes de variância e covariância desconhecidos são substituídos por suas estimativas de MVR (Vonesh e Chinchilli 1997). Sob H_0 , a estatística z tem distribuição assintótica normal padrão. Porém é mais adequado considerá-la como uma estatística com distribuição t-Student, com $J - s_q - 1$ graus de liberdade.

Para testar mais de um efeito fixo simultaneamente, é necessário fazer uso de contrastes e testá-los através da estatística de Wald.

Suponha que o vetor de efeitos fixos seja dado por:

$$\gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} \end{bmatrix}'.$$

Para testar, por exemplo, a hipótese

$$H_0 : \gamma_{12} = 0 \quad \text{e} \quad \gamma_{22} = 0$$

é conveniente reescrevê-la na forma matricial

$$H_0 : C\gamma = 0$$

onde

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Seja $\hat{\gamma}$ o estimador do vetor de efeitos fixos γ e $\hat{V}_{\hat{\gamma}}$ a estimativa da variância de $\hat{\gamma}$.

Então

$$\widehat{Var}(C\hat{\gamma}) = C\hat{V}_{\hat{\gamma}}C' = \hat{V}_C$$

e sob H_0 a estatística do teste de Wald é dada por:

$$H = \hat{\gamma}'C'\hat{V}_C^{-1}C\hat{\gamma}$$

cuja distribuição assintótica é uma Qui-quadrado com número de graus de liberdade igual ao número de linhas da matriz $\mathbf{C}$.

Quando a estimação é feita através de MV, pode-se também testar os efeitos fixos através do *teste da razão de verossimilhanças*. Neste caso temos na hipótese nula todos os efeitos fixos que se deseja testar considerados iguais a zero, ou seja, a hipótese nula corresponde a uma forma reduzida da hipótese alternativa. Vale lembrar que os componentes de variância e covariância das hipóteses nula e alternativa devem ser os mesmos. A estatística do teste da razão de verossimilhanças será abordada na próxima subseção.

5.2 Testes para os Componentes de Variância e Covariância

Assim como os efeitos fixos, os componentes de variância e covariância podem também ser testados isoladamente ou simultaneamente.

Seja τ_{qq} a variância do coeficiente β_{qj} . Para avaliar se este coeficiente deve ser considerado fixo ou aleatório, podemos testar a hipótese nula:

$$H_0 : \tau_{qq} = 0. \quad (5.3)$$

Para testar esta hipótese podem ser utilizadas as estimativas de mínimos quadrados ordinários (MQO) dos coeficientes β_{qj} , representadas por $\hat{\beta}_{qj}$. Mas este teste só é possível se todos ou pelo menos a maioria dos grupos tiver dados suficientes para calcular estas estimativas de MQO. Se o coeficiente β_{qj} fosse fixo, a estimativa da variância amostral do estimador de MQO do grupo j seria dada por $\hat{V}_{qqj}$, que corresponde ao q -ésimo elemento da diagonal da matriz $\hat{\mathbf{V}}_j = \hat{\sigma}^2(\mathbf{X}'_j\mathbf{X}_j)^{-1}$, onde $\mathbf{X}_j$ corresponde à matriz de preditores do nível 1.

Sob H_0 , a variância do efeito aleatório u_{qj} é nula e então

$$\beta_{qj} = \sum_{i=1}^{s_q} \gamma_{qs} W_{sj} .$$

Este teste pode ser realizado através da estatística:

$$X = \sum_{j=1}^J \left(\hat{\beta}_{qj} - \sum_{i=1}^{s_q} \hat{\gamma}_{qs} W_{sj} \right)^2 / \hat{V}_{qqj} \quad (5.4)$$

cujas distribuição é aproximadamente uma Qui-quadrado com $J - s_q - 1$ graus de liberdade.

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

$$z = \frac{\hat{\tau}_{qq}}{[\widehat{Var}(\hat{\tau}_{qq})]^{1/2}} \quad (5.5)$$

onde a estimativa do erro padrão de $\hat{\tau}_{qq}$ é calculada através da inversa da matriz de informação de Fisher. Segundo Bryk e 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).

Para testar um ou mais componentes de variância e covariância simultaneamente, pode-se utilizar o teste da razão de verossimilhanças, já mencionado anteriormente, onde temos na hipótese nula todos os componentes que se deseja testar considerados nulos. Desta forma, a hipótese nula corresponde a uma forma reduzida da hipótese alternativa. Vale ressaltar que tanto na hipótese nula quanto na hipótese alternativa, os modelos devem ser idênticos com relação aos efeitos fixos.

Para calcular a estatística do teste da razão de verossimilhanças é preciso obter o "deviance" do modelo ajustado sob H_0 (D_0) e o "deviance" do modelo ajustado sob H_1 (D_1). Este "deviance" corresponde a -2 vezes o log da sua função de verossimilhança avaliada em seu máximo, e ele pode ser visto como uma medida de qualidade do ajuste, ou seja, quanto maior o "deviance", mais pobre o ajuste. A estatística do teste da razão de verossimilhanças

é dada por:

$$H_{RV} = D_0 - D_1$$

e tem distribuição Qui-quadrado com m graus de liberdade, onde m corresponde à diferença no número de parâmetros dos dois modelos.

Para comparar modelos com os mesmos efeitos fixos, porém com diferentes estruturas de covariâncias pode-se utilizar o critério de informação de Akaike (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. O critério de informação de Akaike (AIC) é calculado como o log da função de verossimilhança avaliada em seu máximo, menos o número de parâmetros. O modelo que apresentar o maior valor do AIC pode ser considerado mais apropriado. O AIC também pode ser utilizado para comparar modelos com os mesmos componentes de variância e covariância porém com diferentes efeitos fixos, porém neste caso o teste só é possível sob a formulação de MV.

Capítulo 6

Aplicação na Área de Educação

Este capítulo apresenta uma aplicação do MLH na área de Educação. Foram utilizados parte dos dados do SARESP 97 (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo), gentilmente fornecidos pela Secretaria de Estado da Educação do Estado de São Paulo.

Neste trabalho estudou-se a habilidade em Língua Portuguesa dos alunos da 4^a. série de escolas públicas do Estado de São Paulo, segundo características dos alunos e das escolas a que eles pertenciam. A habilidade (ou proficiência) é uma variável contínua obtida a partir da teoria da resposta ao item (TRI) (veja Lord (1980), Andrade e Valle (1998)). Através da TRI, os escores dos alunos nas provas são convertidos para uma outra escala, de maneira que dois alunos com o mesmo escore podem não ter a mesma habilidade, uma vez que os graus de dificuldade e de discriminação de cada questão da prova são levados em conta. As análises aqui apresentadas trazem apenas algumas das variáveis que foram efetivamente avaliadas no Saresp 97, a título de ilustração.

O conjunto de dados aqui considerado contém uma amostra de 7897 alunos distribuídos em 557 escolas públicas do Estado de São Paulo. O índice i identifica o aluno e o índice j identifica a escola a que ele pertence. Para cada aluno temos as seguintes variáveis:

- Y_{ij} : habilidade na prova de Língua Portuguesa,
- $Sexo_{ij}$: masculino=1, feminino=0,
- $Idade_{ij}$: (idade-9), (variação de 0 a 4 anos).

Vale ressaltar que a variável $Idade_{ij}$ representa o desvio entre a idade efetiva do aluno e a idade adequada para alunos da 4ª série no início do ano letivo (9 anos), quando foi realizada a prova.

Temos também as variáveis de escola:

- $Tipo_j$: tipo de escola - 1ª a 4ª série=1, 1ª a 8ª série=0,
- $Diretor_j$: atuação do diretor - mais atuante=1, menos atuante=0.

Primeiramente, os dados serão analisados através do programa HLM (Bryk, Raudenbush e Congdon 1996), sob MVR. Em seguida, a mesma análise será feita através do PROC MIXED, um dos procedimentos do programa SAS (SAS Institute Inc. 1997), também sob MVR.

6.1 Análise no HLM

Será adotado inicialmente o seguinte modelo:

$$\begin{aligned}
 Y_{ij} &= \beta_{0j} + \beta_{1j}Sexo_{ij} + \beta_{2j}Idade_{ij} + r_{ij}, \\
 \beta_{0j} &= \gamma_{00} + u_{0j}, \\
 \beta_{1j} &= \gamma_{10} + u_{1j}, \\
 \beta_{2j} &= \gamma_{20} + u_{2j}
 \end{aligned} \tag{6.1}$$

com as suposições:

$\tau_{ij} \sim N(0, \sigma^2)$, τ_{ij} 's independentes,
 $u_{0j} \sim N(0, \tau_{00})$, u_{0j} 's independentes,
 $u_{1j} \sim N(0, \tau_{11})$, u_{1j} 's independentes,
 $u_{2j} \sim N(0, \tau_{22})$, u_{2j} 's independentes e
 $Cov(u_{qj}, u_{kj}) = \tau_{qk}$ se $q \neq k$ e
 u_{0j} 's, u_{1j} 's e u_{2j} 's independentes dos τ_{ij} 's.

A locação da variável $Idade_{ij}$ no valor 9 faz com que o coeficiente β_{0j} seja interpretado como o valor esperado da habilidade na escola j para meninas com 9 anos de idade.

A seguir temos os resultados do ajuste do modelo e os níveis descritivos referentes aos testes de significância dos efeitos fixos:

Tabela 6.1: Estimativas dos parâmetros do modelo inicial e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	0,2644	0,0241	<0,001
γ_{10}	-0,1129	0,0193	<0,001
γ_{20}	-0,1307	0,0089	<0,001
τ_{00}	0,1637		
τ_{11}	0,0135		
τ_{22}	0,0072		
σ^2	0,6388		

Para avaliar quais dos coeficientes β_{qj} podem ser considerados fixos, este modelo inicial será comparado com outros com estrutura de covariâncias reduzida, através do teste da razão de verossimilhanças (RV), com um nível de significância de 5%. Através deste teste será avaliado, para cada coeficiente, se a sua variância e as covariâncias entre este e os demais coeficientes do modelo podem ser consideradas nulas. A seguir temos um resumo das comparações, sob MVR:

Tabela 6.2: Seleção da estrutura de covariâncias

Hipótese	Deviance	Parâmetros	Nível Descritivo*
H_a : modelo (6.1)	19794,92	7	
H_0 : β_{0j} fixo	20162,26	4	<0,001
H_0 : β_{1j} fixo	19796,24	4	0,724
H_0 : β_{2j} fixo	19807,71	4	0,005
H_a : β_{1j} fixo **	19796,24	4	
H_0 : β_{0j} fixo	20277,63	2	<0,001
H_0 : β_{2j} fixo	19809,21	2	0,002

*teste da razão de verossimilhanças; **modelo selecionado

Testar a hipótese H_0 : β_{qj} fixo, corresponde a testar se $\tau_{qq} = 0$ e $\tau_{qk} = 0$ para $q \neq k$, $q = 0, 1, 2$.

De acordo com os resultados da tabela 6.2, o modelo selecionado apresenta somente o coeficiente da variável *Sexo* como fixo. Portanto o componente τ_{11} pode ser considerado nulo e as novas estimativas de τ_{00} , τ_{22} e σ^2 são 0,1661, 0,0073 e 0,6420 respectivamente. Após a redução da estrutura de covariâncias, o próximo passo consiste em verificar se alguma das variáveis de escola ajuda a explicar a variabilidade dos coeficientes β_{0j} e β_{2j} .

Será então ajustado o modelo:

$$\begin{aligned}
 Y_{ij} &= \beta_{0j} + \beta_{1j}Sexo_{ij} + \beta_{2j}Idade_{ij} + r_{ij}, \\
 \beta_{0j} &= \gamma_{00} + \gamma_{01}Tipo_j + \gamma_{02}Diretor_j + u_{0j}, \\
 \beta_{1j} &= \gamma_{10}, \\
 \beta_{2j} &= \gamma_{20} + \gamma_{21}Tipo_j + \gamma_{22}Diretor_j + u_{2j}.
 \end{aligned} \tag{6.2}$$

A tabela 6.3 traz os resultados deste ajuste e os níveis descritivos referentes aos testes de significância dos efeitos fixos:

Tabela 6.3: Estimativas dos parâmetros do modelo (6.2) e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	0,1500	0,0402	<0,001
γ_{01}	0,0967	0,0464	0,037
γ_{02}	0,1295	0,0460	0,005
γ_{10}	-0,1140	0,0187	<0,001
γ_{20}	-0,1336	0,0151	<0,001
γ_{21}	0,0150	0,0186	0,420
γ_{22}	-0,0032	0,0181	0,861
τ_{00}	0,1598		
τ_{22}	0,0073		
σ^2	0,6419		

De acordo com a estatística de Wald, os parâmetros correspondentes às variáveis *Tipo* (γ_{21}) e *Diretor* (γ_{21}), no coeficiente β_{2j} , podem ser considerados simultaneamente nulos com $p=0,716$.

Temos então o modelo final:

$$\begin{aligned}
 Y_{ij} &= \beta_{0j} + \beta_{1j}Sexo_{ij} + \beta_{2j}Idade_{ij} + r_{ij}, \\
 \beta_{0j} &= \gamma_{00} + \gamma_{01}Tipo_j + \gamma_{02}Diretor_j + u_{0j}, \\
 \beta_{1j} &= \gamma_{10}, \\
 \beta_{2j} &= \gamma_{20} + u_{2j}.
 \end{aligned} \tag{6.3}$$

A tabela 6.4 traz os resultados do ajuste deste modelo e os níveis descritivos referentes aos testes de significância dos efeitos fixos:

Tabela 6.4: Estimativas dos parâmetros do modelo final e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	0,1440	0,0350	<0,001
γ_{01}	0,1197	0,0365	0,001
γ_{02}	0,1248	0,0360	0,001
γ_{10}	-0,1137	0,0187	<0,001
γ_{20}	-0,1301	0,0089	<0,001
τ_{00}	0,1595		
τ_{22}	0,0072		
σ^2	0,6420		

Utilizando a estatística dada em (5.4), tanto a variância condicional do intercepto (τ_{00}) como a variância do coeficiente β_{2j} (τ_{22}) são significativamente diferentes de zero, com $p < 0,001$ e $p = 0,002$ respectivamente. Observe que a variabilidade entre os alunos (σ^2) é o componente de variância de maior valor.

Concluindo, à medida que o aluno se afasta da idade adequada da série em estudo (9 anos), piora o seu desempenho em Língua Portuguesa. A reta média que descreve a relação desempenho-idade para os meninos está abaixo da reta média das meninas, quando se considera o mesmo tipo de escola e a mesma atuação do diretor. Além disso, escolas de 1ª a 4ª série e também com um diretor mais atuante, contribuem para aumentar ainda mais o desempenho médio dos alunos.

6.2 Análise no SAS e Aspectos Computacionais

Os dados foram reanalisados nos SAS, utilizando o PROC MIXED, sob MVR, seguindo a mesma estratégia adotada no HLM. Os resultados obtidos são praticamente idênticos aos do HLM. As estimativas dos parâmetros diferem a partir da 4ª casa decimal e os níveis

descritivos também estão bastante próximos.

Os resultados do ajuste do modelo final sob MV e os níveis descritivos para os testes de significância dos efeitos fixos obtidos no HLM podem ser encontrados na tabela B.1 do apêndice B. Sob MV, estes valores também são praticamente os mesmos obtidos através do SAS. Vale ressaltar que os resultados da estimação sob MV nesta aplicação são muito semelhantes aos resultados obtidos sob MVR.

A leitura dos dados nos dois programas utilizados é bem diferente. O HLM necessita de um arquivo de dados para cada nível do modelo, enquanto que o SAS trabalha com um único arquivo. Nesta aplicação por exemplo, o HLM utiliza um arquivo para os alunos e outro para as escolas. Cada linha do arquivo do primeiro nível representa um aluno, contendo uma variável que identifica a escola a que ele pertence e os preditores do nível 1. Cada linha do arquivo do segundo nível representa uma escola, contendo uma variável que identifica esta escola e os preditores do nível 2. Já o PROC MIXED do SAS utiliza um arquivo com todos os preditores, tanto do nível 1 quanto do nível 2. Assim os valores dos preditores do nível 2 são repetidos para todos os alunos que pertencem à mesma escola. Cada linha do arquivo representa um aluno e é preciso ter também uma variável que identifica a escola a que ele pertence.

Assim como na leitura dos dados, a saída com os resultados gerada pelo HLM reflete os diferentes níveis de hierarquia. Já o SAS não possui uma saída organizada desta forma, uma vez que ele não foi especificamente elaborado para análise de modelos lineares hierárquicos.

A implementação de um MLH no PROC MIXED não é tão direta como no HLM. Singer (1998) apresenta exemplos semelhantes ao desta aplicação, mostrando como implementá-los no PROC MIXED, de forma bastante didática. No apêndice C podem ser encontrados alguns dos programas utilizados na análise no SAS. Vale lembrar que o processo de estimação feito no HLM utiliza o algoritmo EM, como descrito no capítulo 4. Já o SAS utiliza como "default" o algoritmo de Newton-Raphson (veja Andreoni (1989) e SAS Institute Inc. (1997)).

O HLM tem sido largamente utilizado para aplicações na área educacional, possuindo

uma interface bastante amigável com o usuário (veja <http://www.ssicentral.com/home.htm>). Este e outros softwares utilizados na análise de modelos lineares hierárquicos são brevemente comentados em Kreft e De Leeuw (1998), como o MLwiN (<http://www.ioe.ac.uk/multilevel>), o VARCL e outros.

Capítulo 7

Aplicação em Dados Longitudinais

Estudos longitudinais (veja Singer e Andrade (2000), Singer e Andrade (1986)) investigam as alterações que ocorrem em uma determinada característica ao longo de uma dimensão específica como o tempo, a distância de uma certa origem, a dosagem de uma determinada substância e outros. Para cada indivíduo presente no estudo, essa característica é medida em duas ou mais ocasiões. A análise destes dados deve levar em conta a possível dependência entre as observações feitas no mesmo indivíduo.

Os modelos multivariados para análise de dados longitudinais (Morrison 1967) comumente descritos na literatura não são apropriados para dados desbalanceados ou incompletos. Já os modelos de efeitos aleatórios (Laird e Ware 1982) permitem que o número de observações e as ocasiões em que elas ocorreram variem de um indivíduo para outro.

Um modelo de efeitos aleatórios pode ser tratado sob o enfoque de um MLH. Considere um MLH com dois níveis, em que o primeiro nível descreve uma curva de crescimento para cada indivíduo e o segundo nível analisa individualmente cada coeficiente desta curva. Como já abordado no capítulo 2, cada coeficiente da curva pode ser considerado fixo, aleatório e até mesmo variar não aleatoriamente. Este modelo permite a inclusão de variáveis dependentes do tempo, bem como variáveis que não sofram alterações durante todo o período de estudo.

A utilização de níveis de modelagem torna mais explícita a variação intraindivíduo e entre indivíduos e muitas vezes facilita a interpretação dos parâmetros.

Este capítulo apresenta uma aplicação de um MLH a um estudo longitudinal. Os dados utilizados foram gentilmente cedidos pelo pesquisador Guilherme Veiga Guimarães, do Laboratório de Insuficiência Cardíaca do InCor (Instituto do Coração - Hospital das Clínicas da FMUSP).

Nesta pesquisa 35 crianças entre 5 e 12 anos realizaram um teste ergo-espirométrico. Entre estas crianças, 10 eram saudáveis e 25 sofriam de insuficiência cardíaca congestiva (ICC). No teste em questão, foram medidas, em diversos instantes, a ventilação pulmonar (VE), em l/min , e a produção de dióxido de carbono (VCO_2), em ml/min . O objetivo foi estudar o comportamento médio da VE ao longo de diferentes valores da VCO_2 e também verificar se este comportamento médio diferia entre crianças normais e crianças com ICC. O número de avaliações em que a VE e a VCO_2 foram observadas não foi o mesmo para todas as crianças. Além disso, estas crianças não foram todas observadas nos mesmos valores da VCO_2 .

Assim como no capítulo anterior, os dados serão analisados através do programa HLM e, em seguida, através do PROC MIXED, em ambos os casos considerando MVR.

7.1 Análise no HLM

Temos então as variáveis do nível 1 (nível avaliação):

- VE_{ij} : valor da VE na i -ésima avaliação da VCO_2 para a criança j (resposta),
- $VCO_{2,ij}$: valor da VCO_2 na i -ésima avaliação para a criança j

e a variável do nível 2 (nível criança):

$$\bullet \text{ } ICC_j = \begin{cases} 1 & \text{se a criança } j \text{ tem ICC,} \\ 0 & \text{caso contrário.} \end{cases}$$

Inicialmente será adotado o seguinte modelo:

$$\begin{aligned} VE_{ij} &= \beta_{0j} + \beta_{1j}(VCO2_{ij} - L) + r_{ij}, \\ \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + u_{1j}. \end{aligned} \tag{7.1}$$

onde L representa a média global da variável $VCO2_{ij}$,

$r_{ij} \sim N(0, \sigma^2)$, r_{ij} 's independentes,

$u_{0j} \sim N(0, \tau_{00})$, u_{0j} 's independentes,

$u_{1j} \sim N(0, \tau_{11})$, u_{1j} 's independentes e

$Cov(u_{0j}, u_{1j}) = \tau_{01}$ e

u_{0j} 's e u_{1j} 's independentes dos r_{ij} 's.

Ajustando este modelo com o preditor do nível 1 centrado na média global, estaremos estudando a mudança esperada na VE da criança j quando a VCO2 aumenta em uma unidade em relação à média global L . Esta locação foi adotada de modo a tornar o intercepto interpretável, ou seja, β_{0j} é a VE esperada na criança j quando ela atinge uma VCO2 igual a L .

A seguir temos os resultados do ajuste do modelo (7.1) e os níveis descritivos referentes ao testes de significância de cada um dos efeitos fixos:

Tabela 7.1: Estimativas dos parâmetros do modelo inicial e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	23,0824	0,8484	<0,001
γ_{10}	0,0435	0,0018	<0,001
τ_{00}	24,5871		
τ_{11}	0,0001		
σ^2	1,5086		

Para avaliar se algum dos coeficientes do modelo do nível 1 podem ser considerados fixos, foram realizados testes da razão de verossimilhanças (RV), com um nível de significância de 5%. Segue a tabela com o resumo das comparações, sob MVR:

Tabela 7.2: Seleção da estrutura de covariâncias

Hipótese	Deviance	Parâmetros	Nível Descritivo*
H_a : modelo (7.1) **	1278,35	4	
H_0 : β_{0j} fixo	1723,50	2	<0,001
H_0 : β_{1j} fixo	1497,75	2	<0,001

*teste da razão de verossimilhanças; **modelo selecionado

De acordo com os resultados da tabela 7.2, todos os coeficientes do modelo de nível 1 devem ser considerados aleatórios. Isto representa que existe variabilidade entre as retas médias de cada criança, tanto nos interceptos quanto nas inclinações.

Resta averiguar se a variabilidade destes coeficientes pode ser explicada pela variável ICC.

Será então ajustado o modelo:

$$\begin{aligned} VE_{ij} &= \beta_{0j} + \beta_{1j}(VCO2_{ij} - L) + r_{ij}, \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}ICC_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}ICC_j + u_{1j}. \end{aligned} \quad (7.2)$$

Os resultados do ajuste do modelo (7.2) e os níveis descritivos referentes ao teste de cada um dos efeitos fixos estão na tabela a seguir:

Tabela 7.3: Estimativas dos parâmetros do modelo (7.2) e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	19,7923	1,4464	<0,001
γ_{01}	4,6033	1,7173	0,012
γ_{10}	0,0374	0,0031	<0,001
γ_{11}	0,0085	0,0037	0,029
τ_{00}	20,6370		
τ_{11}	0,0001		
σ^2	1,5053		

Utilizando a estatística (5.4), tanto a variância condicional τ_{00} quanto a variância condicional τ_{11} são significativamente diferentes de zero, com $p < 0,001$ para ambos os testes.

Diferentemente do que ocorreu na outra aplicação, em que σ^2 foi o maior componente de variância, neste exemplo τ_{00} apresenta o maior valor. Nesta aplicação é τ_{00} que compõe a variabilidade entre indivíduos enquanto que na aplicação anterior é σ^2 que representa essa variabilidade.

Foi verificado através do teste da RV se as inclinações β_{1j} poderiam ser consideradas como variando apenas entre crianças que possuem e não possuem ICC. A pergunta feita é: será que a variável ICC consegue explicar praticamente toda a variabilidade do coeficiente β_{1j} ? O resultado do teste da RV indica que não ($p < 0,001$) e portanto a tabela 7.3 traz os

resultados do modelo final.

Concluindo, a relação VE-VCO₂ é mais forte em crianças que sofrem de ICC. Elas possuem um acréscimo esperado na VE de 0,0085 *l/min* por unidade de VCO₂, em relação às crianças saudáveis. Em outras palavras, a VE aumenta mais rapidamente nas crianças doentes à medida que a VCO₂ cresce. Ainda nas crianças com ICC, o valor esperado da VE no momento em que a VCO₂ atinge o valor L é maior que nas crianças normais em quase 5 *l/min*.

7.2 Análise no SAS

Os dados foram reanalisados nos SAS, utilizando o PROC MIXED, sob MVR. Novamente os resultados obtidos em cada uma das etapas das análises são praticamente os mesmos do HLM, com estimativas dos parâmetros diferindo a partir da 4^a casa decimal e níveis descritivos muito próximos.

Os resultados do ajuste do modelo final sob MV e os níveis descritivos para os testes de significância dos efeitos fixos obtidos no HLM podem ser encontrados na tabela B.2 do apêndice B. Sob MV, estes valores também são praticamente os mesmos obtidos através do SAS. Os resultados da estimação sob MV nesta aplicação são muito semelhantes aos resultados sob MVR.

Novamente alguns dos programas utilizados na análise no SAS estão incluídos no apêndice C.

Capítulo 8

Conclusão

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. Diversos submodelos do MLH com 2 níveis foram apresentados, desde os mais simples, em que não há nenhuma variável preditora, até os mais complexos, em que existem preditores nos 2 níveis do modelo. Neste trabalho fica claro como o tratamento dos dados em 2 níveis de hierarquia facilita a construção do modelo desejado e a interpretação de seus parâmetros.

Outro aspecto importante discutido foi a questão da locação dos preditores no nível 1, abordando modelos de escores brutos, modelos com locação na média global e modelos com locação na média do grupo. Muitas vezes a mudança de locação pode tornar mais adequado o significado do intercepto do modelo, mas pode também resultar na mudança de interpretação dos outros coeficientes. Uma vantagem da mudança de locação é que ela soluciona possíveis problemas numéricos no processo de estimação. Entretanto não há uma escolha estatisticamente correta entre as diferentes locações apresentadas. A escolha da locação deve levar em conta não só aspectos como o processo de estimação e vantagens computacionais, mas principalmente o conhecimento dos dados e objetivos do estudo.

Uma sugestão para a continuidade deste trabalho é realizar um estudo mais detalhado da questão da locação dos preditores do nível 1, explorando diferentes conjuntos de dados para investigar se é possível prever o que acontece com as estimativas dos componentes de variância e covariância quando se altera a locação dos preditores em determinadas situações.

Os modelos lineares hierárquicos vêm sendo largamente utilizados em diversas áreas de estudo, como por exemplo em Educação, e cada vez mais crescem suas aplicações, estudos teóricos e ferramentas computacionais para a implementação destes modelos. Neste trabalho, toda teoria e aplicações dos modelos hierárquicos foram apresentadas considerando modelos lineares e com distribuição normal. Existem outras alternativas, como por exemplo modelos para dados binários ou ainda dados com distribuição de Poisson. O HLM permite trabalhar com estes modelos, assim como o PROC GENMOD do SAS. Outra sugestão para estudos futuros é então abordar a teoria de modelos lineares hierárquicos generalizados ou ainda modelo hierárquicos não lineares, ilustrando suas aplicações.

Apêndice A

A.1 Cálculo de D^*

Para calcular a variância a posteriori de $\theta_1, \theta_2 | Y, \sigma^2, \Gamma, T$ na subseção 4.2.1, foi utilizado o método de inversão de uma matriz particionada simétrica (ver Morrison (1967)), descrito a seguir:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$$

com

$$\begin{aligned} \mathbf{V}_{11} &= (\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1}, \\ \mathbf{V}_{12} &= -\mathbf{V}_{11}\mathbf{BC}^{-1}, \\ \mathbf{V}_{21} &= \mathbf{V}_{12}' \quad \text{e} \\ \mathbf{V}_{22} &= \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}'\mathbf{V}_{11}\mathbf{BC}^{-1}. \end{aligned}$$

A variância a posteriori $\mathbf{D}^*$ é dada por:

$$\begin{aligned}
 \mathbf{D}^* &= (\mathbf{A}'\Psi^{-1}\mathbf{A} + \Omega^{-1})^{-1} \\
 &= (\sigma^{-2}\mathbf{A}'\mathbf{A} + \Omega^{-1})^{-1} \\
 &= \left\{ \begin{bmatrix} \sigma^{-2}\mathbf{A}'_1\mathbf{A}_1 & \sigma^{-2}\mathbf{A}'_1\mathbf{A}_2 \\ \sigma^{-2}\mathbf{A}'_2\mathbf{A}_1 & \sigma^{-2}\mathbf{A}'_2\mathbf{A}_2 \end{bmatrix} + \begin{bmatrix} \Gamma^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{bmatrix} \right\}^{-1} \\
 &= \left\{ \begin{array}{cc} \sigma^{-2}\mathbf{A}'_1\mathbf{A}_1 + \Gamma^{-1} & \sigma^{-2}\mathbf{A}'_1\mathbf{A}_2 \\ \sigma^{-2}\mathbf{A}'_2\mathbf{A}_1 & \sigma^{-2}\mathbf{A}'_2\mathbf{A}_2 + \mathbf{T}^{-1} \end{array} \right\}^{-1} \\
 &= \sigma^2 \left\{ \begin{array}{cc} \mathbf{A}'_1\mathbf{A}_1 + \sigma^2\Gamma^{-1} & \mathbf{A}'_1\mathbf{A}_2 \\ \mathbf{A}'_2\mathbf{A}_1 & \mathbf{A}'_2\mathbf{A}_2 + \sigma^2\mathbf{T}^{-1} \end{array} \right\}^{-1}.
 \end{aligned}$$

Utilizando o método de inversão mencionado temos, neste caso, que:

$$\mathbf{A} = \mathbf{A}'_1\mathbf{A}_1 + \sigma^2\Gamma^{-1},$$

$$\mathbf{B} = \mathbf{A}'_1\mathbf{A}_2 \quad \text{e}$$

$$\mathbf{C} = \mathbf{A}'_2\mathbf{A}_2 + \sigma^2\mathbf{T}^{-1}.$$

Lembrando que $\Gamma^{-1} \rightarrow \mathbf{0}$, então

$$\mathbf{V}_{11} = (\mathbf{A}'_1\mathbf{A}_1 - \mathbf{A}'_1\mathbf{A}_2\mathbf{C}^{-1}\mathbf{A}'_2\mathbf{A}_1)^{-1},$$

$$\mathbf{V}_{12} = -\mathbf{V}_{11}\mathbf{A}'_1\mathbf{A}_2\mathbf{C}^{-1},$$

$$\mathbf{V}_{21} = \mathbf{V}'_{12} \quad \text{e}$$

$$\mathbf{V}_{22} = \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{A}'_2\mathbf{A}_1\mathbf{V}_{11}\mathbf{A}'_1\mathbf{A}_2\mathbf{C}^{-1}.$$

Desta forma

$$\mathbf{D}^* = \begin{bmatrix} \mathbf{D}_{11}^* & \mathbf{D}_{12}^* \\ \mathbf{D}_{21}^* & \mathbf{D}_{22}^* \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}, \quad \text{ou seja,}$$

$$\mathbf{D}_{11}^* = \text{Var}(\boldsymbol{\theta}_1|\mathbf{Y}, \sigma^2, \mathbf{T}) = \sigma^2\mathbf{V}_{11},$$

$$\mathbf{D}_{12}^* = \text{Cov}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{Y}, \sigma^2, \mathbf{T}) = -\mathbf{D}_{11}^*\mathbf{A}'_1\mathbf{A}_2\mathbf{C}^{-1},$$

$$\mathbf{D}_{21}^* = \text{Cov}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_1|\mathbf{Y}, \sigma^2, \mathbf{T}) = \mathbf{D}_{12}^{*'} \quad \text{e}$$

$$\mathbf{D}_{22}^* = \text{Var}(\boldsymbol{\theta}_2|\mathbf{Y}, \sigma^2, \mathbf{T}) = \sigma^2\mathbf{V}_{22}.$$

A.2 Cálculo de θ^*

$$\theta \sim N(\bar{\theta}, \Omega)$$

Se $\bar{\theta} = 0$ então

$$\begin{aligned} \theta^* &= D^*(\Omega^{-1}\bar{\theta} + A'\Psi^{-1}Y) \\ &= D^*A'\Psi^{-1}Y \\ &= \sigma^{-2} \begin{bmatrix} D_{11}^* & D_{12}^* \\ D_{21}^* & D_{22}^* \end{bmatrix} \begin{bmatrix} A_1' \\ A_2' \end{bmatrix} Y \\ &= \sigma^{-2} \begin{bmatrix} D_{11}^*A_1'Y + D_{12}^*A_2'Y \\ D_{21}^*A_1'Y + D_{22}^*A_2'Y \end{bmatrix} = \begin{bmatrix} \theta_1^* \\ \theta_2^* \end{bmatrix} \end{aligned}$$

e portanto

$$\begin{aligned} \theta_1^* &= \sigma^{-2}D_{11}^*A_1'Y + \sigma^{-2}D_{12}^*A_2'Y \\ &= V_{11}A_1'Y - V_{11}A_1'A_2C^{-1}A_2'Y \\ &= V_{11}A_1'(I - A_2C^{-1}A_2')Y, \end{aligned}$$

$$\begin{aligned} \theta_2^* &= \sigma^{-2}D_{21}^*A_1'Y + \sigma^{-2}D_{22}^*A_2'Y \\ &= (-V_{11}A_1'A_2C^{-1})'A_1'Y + V_{22}A_2'Y \\ &= -C^{-1}A_2'A_1V_{11}A_1'Y + C^{-1}A_2'Y + C^{-1}A_2'A_1V_{11}A_1'A_2C^{-1}A_2'Y \\ &= C^{-1}A_2'Y - C^{-1}A_2'A_1V_{11}A_1'(I - A_2C^{-1}A_2')Y \\ &= C^{-1}A_2'Y - C^{-1}A_2'A_1\theta_1^* \\ &= C^{-1}A_2'(Y - A_1\theta_1^*). \end{aligned}$$

A.3 Representação de Alguns Vetores e Matrizes

A.3.1 Matriz A_1

$$\begin{aligned}
 A_1 &= XW = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_J \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_J \end{bmatrix} \\
 &= \begin{bmatrix} X_1 W_1 \\ X_2 W_2 \\ \vdots \\ X_J W_J \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{12} \\ \vdots \\ A_{1J} \end{bmatrix}.
 \end{aligned}$$

Portanto $A_{1j} = X_j W_j$, $j = 1, 2, \dots, J$.

A.3.2 Matriz A_2

Seja X_j° a matriz de preditores X_j cujas colunas correspondentes aos coeficientes β_{qj} considerados fixos foram suprimidas.

Por exemplo:

$$X_j = \begin{bmatrix} X_{11j} & X_{21j} & X_{31j} \\ X_{12j} & X_{22j} & X_{32j} \\ X_{13j} & X_{23j} & X_{33j} \\ \vdots & \vdots & \vdots \\ X_{1n_jj} & X_{2n_jj} & X_{3n_jj} \end{bmatrix} \quad e \quad \beta_j = \begin{bmatrix} \beta_{1j} \\ \beta_{2j} \\ \beta_{3j} \end{bmatrix}.$$

Suponha que apenas o coeficiente β_{3j} seja considerado fixo. Neste caso

$$\mathbf{X}_j^\bullet = \begin{bmatrix} X_{11j} & X_{21j} \\ X_{12j} & X_{22j} \\ X_{13j} & X_{23j} \\ \vdots & \vdots \\ X_{1n_jj} & X_{2n_jj} \end{bmatrix}.$$

Então

$$\mathbf{A}_2 = \mathbf{X}^\bullet = \begin{bmatrix} \mathbf{X}_1^\bullet & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^\bullet & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_J^\bullet \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{21} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{2J} \end{bmatrix}$$

e portanto $\mathbf{A}_{2j} = \mathbf{X}_j^\bullet$.

A.3.3 Matriz C

$$\begin{aligned} \mathbf{C} &= \mathbf{A}_2' \mathbf{A}_2 + \sigma^2 \mathbf{T}^{-1} \\ &= \begin{bmatrix} \mathbf{A}_{21}' \mathbf{A}_{21} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}' \mathbf{A}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{2J}' \mathbf{A}_{2J} \end{bmatrix} + \begin{bmatrix} \tau^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tau^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tau^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{21}' \mathbf{A}_{21} + \tau^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}' \mathbf{A}_{22} + \tau^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_{2J}' \mathbf{A}_{2J} + \tau^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}_J \end{bmatrix} \end{aligned}$$

e portanto $\mathbf{C}_j = \mathbf{A}_{2j}' \mathbf{A}_{2j} + \sigma^2 \tau^{-1}$.

Apêndice B

B.1 Tabela Referente ao Capítulo 6

Tabela B.1: Estimativas dos parâmetros do modelo final
sob MV e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	0,1441	0,0349	<0,001
γ_{01}	0,1196	0,0364	0,001
γ_{02}	0,1248	0,0359	0,001
γ_{10}	-0,1138	0,0187	<0,001
γ_{20}	-0,1301	0,0089	<0,001
τ_{00}	0,1583		
τ_{22}	0,0071		
σ^2	0,6419		

B.2 Tabela Referente ao Capítulo 7

Tabela B.2: Estimativas dos parâmetros do modelo final
sob MV e níveis descritivos

Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
γ_{00}	19,7882	1,4014	<0,001
γ_{01}	4,5929	1,6639	0,010
γ_{10}	0,0374	0,0030	<0,001
γ_{11}	0,0084	0,0036	0,025
τ_{00}	19,3629		
τ_{11}	0,0001		
σ^2	1,5097		

Apêndice C

C.1 Programas no SAS, PROC MIXED

Implementação do modelo 6.1

```
proc mixed data=port4_97 method=reml;  
class escola;  
model habilid=sexo idade/solution ddfm=bw;  
random int sexo idade/type=un subject=escola;  
run;
```

Implementação do modelo 6.2

```
proc mixed data=port4_97 method=reml;  
class escola;  
model habilid=sexo idade tipo diretor tipo*idade diretor*idade/solution ddfm=bw chisq;  
random int idade/type=un subject=escola;  
contrast 'tipo=0 e diretor=0 em idade'  
tipo*idade 1,  
diretor*idade 1;  
run;
```

Implementação do modelo 7.1

```
proc mixed data=VE_VCO2 method=reml;  
class nome;  
model VE = VCO2_MU/solution ddfm=bw;  
random int VCO2_MU/type=un subject=nome;  
run;
```

Implementação do modelo 7.2

```
proc mixed data=VE_VCO2 method=reml;  
class nome;  
model VE = ICC VCO2_MU ICC*VCO2_MU/solution ddfm=bw;  
random int VCO2_MU/type=un subject=nome;  
run;
```

Referências Bibliográficas

- Andrade, D. F. e Valle, R. C. (1998). Introdução à Teoria da Resposta ao Item: Conceitos e Aplicações, *Estudos em Avaliação* **18**: 13–32.
- Andreoni, S. (1989). *Modelos de Efeitos Aleatórios para Análise de Dados Longitudinais Não Balanceados em Relação ao Tempo*, Dissertação de Mestrado, Instituto de Matemática e Estatística da Universidade de São Paulo.
- Bryk, A. S. e Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*, Newbury Park: Sage.
- Bryk, A. S., Raudenbush, S. W. e Congdon, R. T. (1996). *HLMTM. Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*, Chicago: Scientific Software International.
- Degroot, M. H. (1998). *Probability and Statistics*, Reading, MA: Addison-Wesley.
- Dempster, A. P., Laird, N. M. e Rubin, D. B. (1977). Maximum Likelihood for Incomplete Data Via the EM Algorithm, *Journal of the Royal Statistical Society B* **39**: 1–8.
- Dempster, A. P., Rubin, D. B. e Tsutakawa, R. K. (1981). Estimation in Covariance Components Models, *Journal of the American Statistical Association* **76**: 341–353.
- Giampaoli, V. (1999). *Inferência Estatística para Modelos Lineares com Restrições nos Parâmetros em Condições Regulares e Não Regulares*, Tese de Doutorado, Instituto de Matemática e Estatística da Universidade de São Paulo.

- Goldstein, H. (1986). Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares, *Biometrika* **73**(1): 43–56.
- Goldstein, H. (1995). *Multilevel Statistical Models*, London: Edward Arnold.
- Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*, London: Chapman & Hall.
- Kreft, I. G. G. (1998). The Effects of Centering in Multilevel Analysis: Is the Public School the Loser or the Winner? A New Analysis of an Old Question.
- Kreft, I. G. G. e De Leeuw, J. (1998). *Introducing Multilevel Modeling*, London: Sage.
- Kreft, I. G. G., De Leeuw, J. e Aiken, L. S. (1995). The Effect of Different Forms of Centering in Hierarchical Linear Models, *Multivariate Behavioral Research* **30**: 1–21.
- Laird, N. M. e Ware, H. (1982). Random-Effects Models for Longitudinal Data, *Biometrics* **38**: 963–974.
- Lindley, D. V. e Smith, A. F. M. (1972). Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society B* **34**: 1–41.
- Longford, N. T. (1987). A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects, *Biometrika* **74**(4): 817–827.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*, Hillsdale: Lawrence Erlbaum Associates.
- Morrison, D. F. (1967). *Multivariate Statistical Methods*, New York: McGraw-Hill.
- Neter, J., Kutner, M. H., Nachtsheim, C. J. e Wasserman, W. (1996). *Applied Linear Statistical Models*, 4 edn, Chicago: Irwin.
- Peffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H. e Rasbash, J. (1998). Weighting for Unequal Selection Probabilities in Multilevel Models, *Journal of the Royal Statistical Society B* **60**: 23–40.

- Pessoa, D. G. C. e Silva, P. L. N. (1998). *Análise de Dados Amostrais Complexos*, São Paulo: Associação Brasileira de Estatística.
- SAS Institute Inc. (1997). *SAS/STAT Software: Changes and Enhancements through Release 6.12*, Cary, NC: SAS Institute Inc.
- Singer, J. D. (1998). Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models and Individual Growth Models, *Journal of Educational and Behavioral Statistics* 24(4): 323–355.
- Singer, J. M. e Andrade, D. F. (1986). *Análise de Dados Longitudinais*, São Paulo: Associação Brasileira de Estatística.
- Singer, J. M. e Andrade, D. F. (2000). *Analysis of Longitudinal Data*, Vol. 18 of *Bio-Environmental and Public Health Statistics*, Amsterdam, North Holland: C. R. Rao and P. K. Sen Eds.
- Smith, A. F. M. (1973). A General Bayesian Linear Model, *Journal of the Royal Statistical Society B* 35: 61–75.
- Vonesh, E. F. e Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker.