Modelos Lineares Hierárquicos

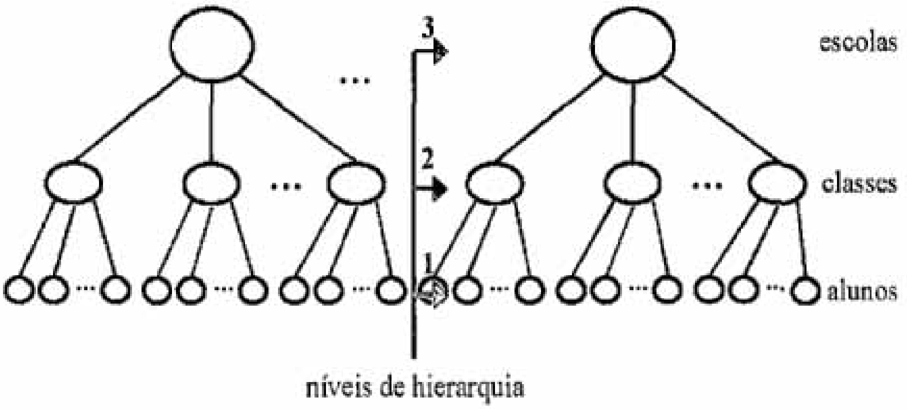
Lílian Natis
Mestre em Estatística pela Universidade de São Paulo - USP

Resumo
O artigo mostra que estruturas hierárquicas de dados são encontradas freqüentemente em diversos estudos e que 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 (MLH), que 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 desses níveis. O modelo hierárquico com dois níveis é apresentado e ilustrado. É apresentada, ao final, uma aplicação na área da educação, utilizando dados do SARESP 97 (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo)

1. Introdução

Estruturas hierárquicas de dados são encontradas com freqüência 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. Este tipo de estrutura é bastante comum em sistemas educacionais, onde temos estudantes agrupados em classes, que por sua vez estão agrupadas em escolas. As escolas podem ainda estar agrupadas em delegacias de ensino e assim por diante (veja Figura 1.1).

Figura 1.1
Representação de uma estrutura hierárquica com 3 níveis



Estas estruturas de dados podem ser estudadas pelos chamados *modelos lineares hierárquicos* (Bryk e Raudenbush 1992, Goldstein 1995), que permitem que cada um dos níveis de hierarquia seja especificado separadamente e que, posteriormente, sejam reunidos em um único modelo. Além disso, estes modelos permitem a incorporação de efeitos aleatórios associados a cada um de seus níveis de hierarquia. Estes efeitos aleatórios são na verdade erros aleatórios que representam as diferenças existentes entre as unidades de cada nível quanto à variável de interesse do estudo, mesmo após o controle de outras variáveis.

Teoricamente, cada nível de um modelo linear hierárquico (MLH) é formado por uma amostra aleatória das unidades consideradas. No caso do sistema educacional, um modelo linear hierárquico com 2 níveis poderia ter o primeiro nível composto por uma amostra de estudantes em cada escola e o segundo nível, por uma amostra de escolas.

Existem variáveis que estão diretamente associadas a cada nível de hierarquia. Por exemplo: em um sistema educacional onde se deseja avaliar a habilidade (ou proficiência) em Língua Portuguesa dos estudantes do primeiro grau, podemos considerar que os estudantes selecionados em cada uma das escolas compõem o primeiro nível do MLH, enquanto que as escolas selecionadas formam o segundo nível. 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 (nível aluno). Já as 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 (nível escola). 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 (neste caso, a habilidade em Língua Portuguesa) e de estudar as possíveis relações entre cada uma das variáveis explicativas (demais variáveis mencionadas) e a variável resposta.

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, Pfeffermann et. al. (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. 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. 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 pelo uso de mínimos quadrados generalizados reponderados iterativamente (Goldstein 1986) e pelo 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, entre eles os programas HLM (Bryk, Raudenbush e Congdon 1996), MLwiN (Rasbash et al. 2000) e SAS (SAS Institute Inc. 1997).

Os modelos lineares hierárquicos constituem portanto uma nova formulação para os modelos de efeitos aleatórios, motivada pela própria estrutura dos dados, permitindo a descrição e análise mais apropriada das diferentes fontes de variação e maior facilidade de interpretação dos parâmetros do modelo. Além disso, os modelos lineares hierárquicos podem ser empregados em inúmeras situações e áreas de aplicação.

2. Introdução ao Modelo Linear Hierárquico

Esta seção aborda algumas idéias de análise de regressão e de análise de variância (Neter et al. 1996) para introduzir o conceito de modelo linear hierárquico, apresentando o MLH com 2 níveis e alguns de seus principais submodelos.

2.1 Um exemplo utilizando Regressão Linear Simples

Suponha que em uma determinada escola foram selecionados n alunos e observadas as variáveis:

- Y_i: aproveitamento em Matemática (AM) do i-ésimo aluno (variável resposta),

- X_i: nível sócio-econômico (NSE) do i-ésimo aluno (variável explicativa), $i = 1,2,...,n$.

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

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

onde

β_0: AM esperado para um aluno com NSE igual a 0 ($X_i = 0$),

β_1: mudança esperada no AM quando o NSE (X_i) aumenta em uma unidade,

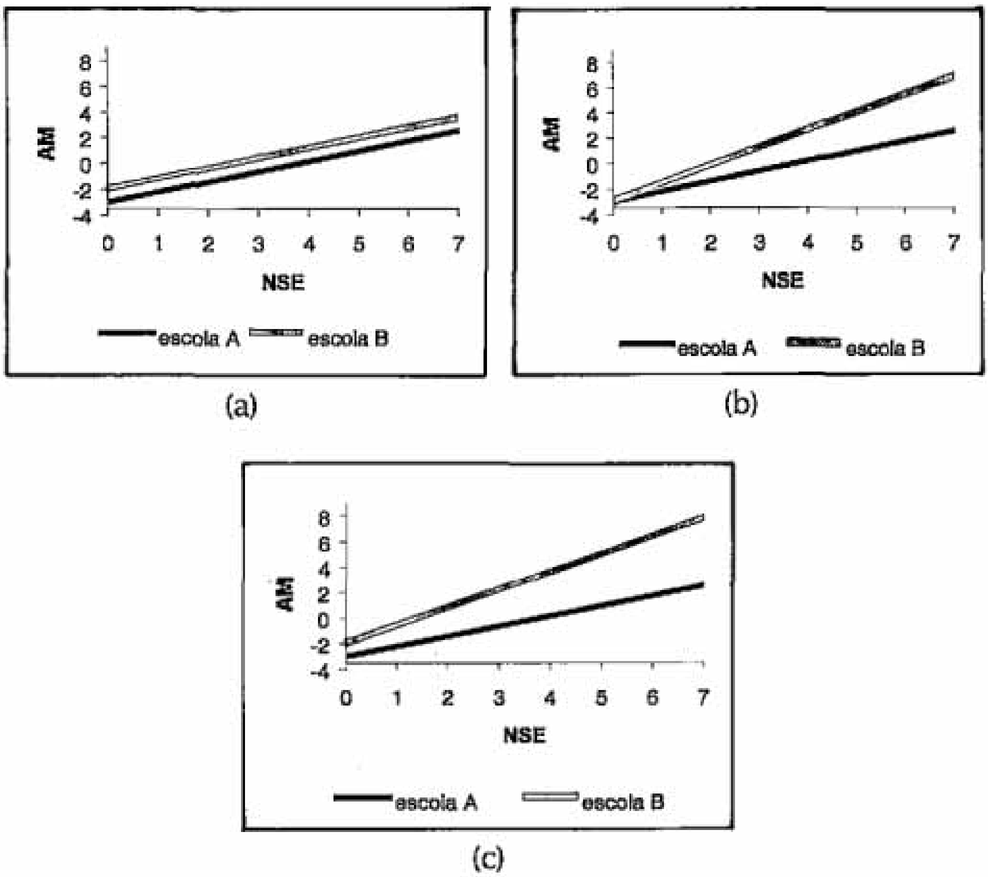
r_i : erro aleatório associado ao i-ésimo aluno,

suposições: os r_i 's possuem distribuição normal com média 0 e variância σ^2

(representado por $r_i \sim N(0, \sigma^2)$) e são 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 (β_0) é a que proporciona um AM esperado maior para alunos com NSE igual a zero. A que tem a menor inclinação (β_1) 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? Na Figura 2.1 estão ilustradas 3 diferentes situações para melhor compreensão da interpretação dos coeficientes de um modelo de regressão linear simples.

Figura 1
Comparação entre as retas médias ajustadas para
duas diferentes escolas



(a) (b) (c)

(a) As escolas A e B possuem diferentes interceptos e mesma inclinação, ou seja, a escola B proporciona maior AM esperado para alunos de menor NSE e a diferença entre as duas escolas se mantém à medida que o NSE aumenta (retas paralelas). A influência do NSE no AM é a mesma para ambas as escolas.

(b) As escolas A e B possuem mesmo intercepto e diferentes inclinações, ou seja, as duas escolas proporcionam mesmo AM esperado para alunos de menor NSE, porém a diferença entre as duas escolas cresce à medida que o NSE aumenta, devido às diferentes inclinações. A influência do NSE no AM é menor na escola A, ou seja, ela trata seus alunos de forma mais igualitária.

(c) As escolas A e B possuem diferentes interceptos e diferentes inclinações, ou seja, a escola B tem maior AM esperado para alunos de menor NSE e a diferença entre as duas escolas cresce à medida que o NSE aumenta. A influência do NSE no AM é menor na escola A.

Em muitos casos, como em um sistema educacional, 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 então ajustar um único modelo que leve em conta a variação entre todas as escolas selecionadas e que incorpore os diferentes aspectos de cada uma delas e dos alunos que as compõem, fornecendo estimativas mais apropriadas e resultados mais interessantes.

2.2 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.

Pelo MLH com 2 níveis, todas as escolas selecionadas serão incorporadas em um único modelo, levando em conta a possível variação de interceptos e inclinações entre as escolas.

Temos então:

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

$i = 1,2,\ldots,n_j$ e $j = 1,2,\ldots, J$, onde

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

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

$\overline{X}_{..}$: média amostral global da variável NSE (X_{ij})

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

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

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. Mais adiante serão comentados alguns 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_0 e a inclinação β_1 podem ser modelados da seguinte forma:

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

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 r_{ij}'s.

Em resumo:

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

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,

$Cov(u_{0j}, u_{ij}) = \tau_{01}$ 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), obtemos uma equação cujo erro aleatório $(u_{0j} + u_{1j}(X_{ij} - \overline{X}_{..}) + r_{ij})$ 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 $(X_{ij} - \overline{X}_{..})$ que variam entre os estudantes.

A aplicação de mínimos quadrados ordinários para estimação dos parâmetros 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 (Andreoni 1989). Observe que se u_{0j} e u_{1j} fossem nulos para todo j, os parâmetros poderiam ser estimados por 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. Este submodelo não possui nenhuma variável preditora em nenhum de seus níveis e, portanto, a inclinação β_{1j} no modelo do nível 1 é nula para todo j.

Temos então o modelo do nível 1:

$$Y_{ij} = \beta_{0j} + r_{ij} \qquad (2.5)$$

com

β_{0j}: resposta esperada para a j-ésima escola,

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

E o modelo do nível 2:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (2.6)$$

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.

A variância da resposta é dada por:

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

Note que esta variância é composta pela variação entre as unidades do nível 1 (alunos) e entre as unidades do nível 2 (escolas).

Este modelo hierárquico é chamado *totalmente não condicional*, pois tanto o nível 1 quanto o nível 2 não possuem nenhuma variável preditora.

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, ρ representa a correlação entre dois indivíduos distintos pertencentes ao mesmo grupo. 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. Estas 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 (2.5). 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:

$$Y_{ij} = \beta_{0j} + r_{ij}, \qquad (2.7)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \qquad (2.8)$$

onde as equações (2.7) e (2.8) 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, ou seja, o efeito da variável do nível 1 é o mesmo para todas elas. 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:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{..}) + r_{ij}, \qquad (2.9)$$

$$\left.\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10}. \end{aligned}\right. \qquad (2.10)$$

Agora β_{0j} é a resposta esperada para um indivíduo da j-ésima unidade do nível 2 com $X_{ij} = \overline{X}_{..}$. Temos ainda que

$$\beta_{0j} = \dot{\mu}_j - \beta_{1j}(\overline{X}_{.j} - \overline{X}_{..}),$$

ou seja, 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:

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

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.

2.3.3.1 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 que haja características do grupo que possam ajudar a explicar esta variação. Neste caso, o modelo do nível 1 e o modelo do nível 2 são idênticos a (2.2) e (2.3), respectivamente:

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

$$\beta_{0j} = \gamma_{00} + u_{0j},$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

2.3.3.2 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:

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

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j},$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}.$$

2.3.3.3 Modelo com Inclinações Variando Não Aleatoriamente

Imagine que, após incluir uma variável de grupo W_j como, preditora 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 τ_{11} e τ_{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 da variável de grupo 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 no nível 2 para cada coeficiente β_{qj} da regressão no nível 1 são dadas por:

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

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

$$i = 1, 2, \ldots, n_j,\ j = 1, 2, \ldots, J \text{ e } q = 0, 1, \ldots, Q.$$

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. Isto significa que uma característica do grupo pode modificar o efeito de uma característica individual na variável resposta, potencializando ou reduzindo este efeito.

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 $X_{ij} = \overline{X}_{..}$, ou uma média ajustada, como já mencionado 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 $X_{ij} = \overline{X}_{.j}$, ou seja, com preditor igual à média amostral de seu grupo. Nesta situação, o intercepto β_{0j} pode ser visto como a verdadeira média do grupo j, ou seja, $\beta_{0j} = \mu_j$.

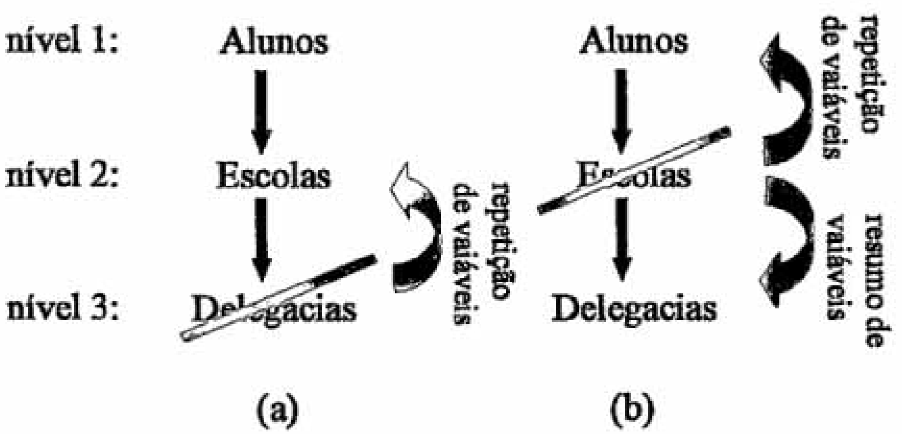
A locação pode ser feita em qualquer outro valor, desde que resulte numa interpretação adequada de β_{0j}. O mesmo cuidado deve ser tomado com relação aos preditores do nível 2, pois a mudança de locação nestas variáveis irá modificar a interpretação dos interceptos deste nível. A mudança de locação algumas vezes ajuda a solucionar possíveis problemas numéricos no processo de estimação. Estas e outras questões envolvendo a locação de preditores são discutidas em Kreft e De Leew (1995) e em Natis (2000).

Este artigo 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 construída de maneira análoga ao MLH com 2 níveis. Entretanto, 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 considerá-la em um dos demais níveis (Figura 2.2), seja repetindo os mesmos valores para diversas unidades experimentais ou utilizando uma medida resumo desta variável, o que podemos chamar de variável agregada.

Figura 2

Exemplos de eliminação de um dos 3 níveis de um MLH



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.

3. Aplicação na Área de Educação

Esta seção 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).

Neste trabalho estudou-se a habilidade em Língua Portuguesa dos alunos da 4ª 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). Pela 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 pelo programa HLM (Bryk, Raudenbush e Congdon 1996), utilizando máxima verossimilhança restrita (MVR) (veja Andreoni (1989)). Em seguida, a mesma análise será feita pelo PROC MIXED, um dos procedimentos do programa SAS (SAS Institute Inc. 1997), também sob MVR.

3.1 Análise no HLM

Será adotado inicialmente o seguinte modelo:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Idade_{ij} + r_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \end{aligned} \tag{3.1}$$

com as suposições:

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

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 3.1
Estimativas dos parâmetros do modelo inicial e níveis descritivos

Coeficiente	Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
β_{0j}	γ_{00}	0,2644	0,0241	< 0,001
β_{1j}	γ_{10}	-0,1129	0,0193	< 0,001
β_{2j}	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, pelo teste da razão de verossimilhanças, com um nível de significância de 5%. Por este 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 3.2
Seleção da estrutura de covariâncias

Hipótese	*Deviance* [1]	Parâmetros	Nível Descritivo [2]
H_a: modelo (3.1)	**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 [3]	19796,24	4	
H_0: β_{0j} fixo	20277,63	2	< 0,001
H_0: β_{2j} fixo	19809,21	2	0,002

[1] Cordeiro (1986);
[2] teste da razão de verossimilhanças;
[3] modelo selecionado

Testar a hipótese H0: 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 3.2, 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:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Idade_{ij} + r_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01} Tipo_j + \gamma_{02} Diretor_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21} Tipo_j + \gamma_{22} Diretor_j + u_{2j} \end{aligned} \tag{3.2}$$

A Tabela 3.3 traz os resultados deste ajuste e os níveis descritivos referentes aos testes de significância dos efeitos fixos:

Tabela 3.3
Estimativas dos parâmetros do modelo e níveis descritivos

Coeficiente	Parâmetro	Estimativa	Erro Padrão	Nível Descritivo
	γ_{00}	0,1500	0,0402	< 0,001
β_{0j}	γ_{01}	0,0967	0,0464	0,037
	γ_{02}	0,1295	0,0460	0,005
β_{1j}	γ_{10}	-0,1140	0,0187	< 0,001
	γ_{20}	-0,1336	0,0151	< 0,001
β_{2j}	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 (Sen e Singer 1993), os parâmetros correspondentes às variáveis *Tipo* (γ_{21}) e *Diretor* (γ_{22}), no coeficiente β_{2j}, podem ser considerados simultaneamente nulos com p=0,716.

Temos então o modelo final:

$$\begin{aligned}
Y_{ij} &= \beta_{0j} + \beta_{1j} Sexo_{ij} + \beta_{2j} Idade_{ij} + r_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01} Tipo_j + \gamma_{02} Diretor_j + u_{0j}, \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} + u_{2j}
\end{aligned} \qquad (3.3)$$

A Tabela 3.4 traz os resultados do ajuste deste modelo e os níveis descritivos referentes aos testes de significância dos efeitos fixos:

Tabela 3.4
Estimativas dos parâmetros do modelo final e níveis descritivos

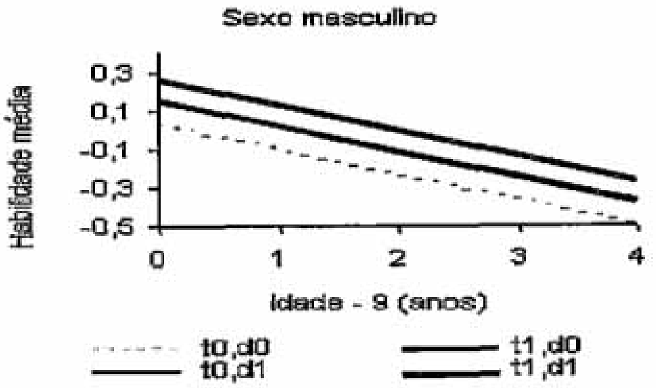
Coeficiente	**Parâmetro**	**Estimativa**	**Erro Padrão**	**Nível Descritivo**
	γ_{00}	0,1440	0,0350	< 0,001
β_{0j}	γ_{01}	0,1197	0,0365	0,001
	γ_{02}	0,1248	0,0360	0,001
β_{1j}	γ_{10}	-0,1137	0,0187	< 0,001
β_{2j}	γ_{20}	-0,1301	0,0089	< 0,001
	τ_{00}	0,1595		
	τ_{22}	0,0072		
	σ^2	0,6420		

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 magnitude.

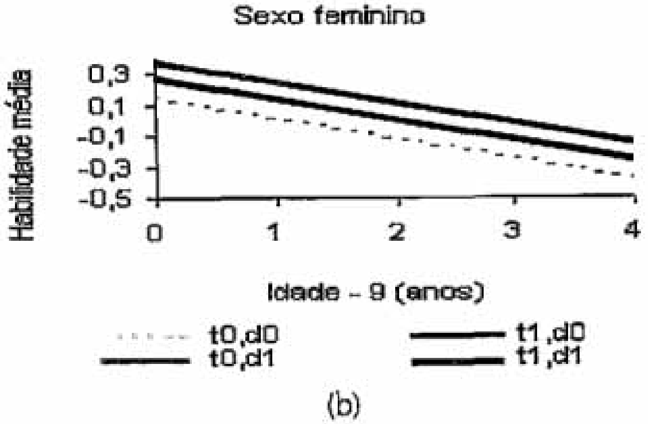
Concluindo, à medida que o aluno se afasta da idade adequada da série em estudo (9 anos), piora o seu desempenho em Língua Portuguesa. A reta média que descreve a relação desempenho-idade para os meninos está abaixo da reta média das meninas, quando se considera o mesmo tipo de escola e a mesma atuação 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 (veja Figura 3.1).

Figura 3.1

Comparação entre as retas médias ajustadas para cada tipo de escola e atuação do diretor, de acordo com o sexo



(a)



(b)

Onde:
t0,d0: escola do tipo 1ª a 8ª série e diretor menos atuante;
t0,d1: escola do tipo 1ª a 8ª série e diretor mais atuante;
t1,d0: escola do tipo 1ª a 4ª série e diretor menos atuante;
t1,d1: escola do tipo 1ª a 4ª série e diretor mais atuante.

3.2 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.

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. Vale lembrar que o processo de estimação feito no HLM utiliza o algoritmo EM. 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 programas utilizados na análise de modelos lineares hierárquicos são brevemente comentados em Kreft e De Leeuw (1998), como o MLwiN (http://multilevel.ioe.ac.uk/index.html), o VARCL e outros.

Considerações Finais

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 artigo fica claro como o tratamento dos dados em 2 níveis de hierarquia facilita a construção do modelo desejado e a interpretação de seus parâmetros.

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 artigo, toda teoria e a aplicação dos modelos hierárquicos foram apresentadas considerando modelos lineares e com distribuição normal. Entretanto, existem outras alternativas, como, por exemplo, modelos para dados binários ou, ainda, dados com distribuição de Poisson, dependendo da distribuição da variável resposta que se deseja estudar. O HLM permite trabalhar com estes modelos, assim como o PROC GENMOD do SAS.

Agradecimentos

Este trabalho é resultado da dissertação de mestrado em Estatística, defendida pela autora no Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), sob orientação do Prof. Dalton Francisco de Andrade.

A autora gostaria de agradecer à Secretaria de Estado da Educação do Estado de São Paulo por disponibilizar os dados do SARESP 97, à Fundação Carlos Chagas, em especial à Profª Yara Lúcia

Esposito, pelas discussões que tanto motivaram este trabalho, ao CNPq, pelo apoio financeiro, e ao Prof. Dalton Francisco de Andrade, pelas inúmeras contribuições ao longo de toda trajetória.

Referências Bibliográficas

ANDRADE, Dalton F. e VALLE, Raquel C. Introdução à Teoria da Resposta ao Item: conceitos e aplicações. *Estudos em Avaliação*, São Paulo : Fundação Carlos Chagas, (18), p. 13-32, 1998.

ANDREONI, S. *Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo*. Dissertação de Mestrado apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo. 1989.

BRYK, A. S.; RAUDENBUSH, S. W. *Hierarquical Linear Models: applications and data analysis methods*. Newbury Park: Sage. 1992.

BRYK, A. S.; RAUDENBUSH, S. W.; CONGDON, R. T. *HLM ™. Hierarquical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software International. 1996.

CORDEIRO, G. M. Modelos Lineares Generalizados. *VII SINAPE*, Campinas: Unicamp, 1986.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum Likelihood for Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society*, B.39, p.1-8, 1977.

DEMPSTER, A. P.; RUBIN, D. B.; TSUTAKAWA, R. K. Estimation in Covariance Components Models. *Journal of the American Statistical Association*, 76, p. 341-353, 1981.

GOLDSTEIN, H. Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares. *Biometrika* 73 (1), p. 43-56, 1986.

GOLDSTEIN, H. *Multilevel Statistical Models*, London: Edward Arnold, 1995.

KREFT, I. G. G.; DE LEEUW, J. *Introducing Multilevel Modeling*, London: Sage, 1998.

KREFT, I. G. G.; DE LEEUW, J.; AIKEN, L. S. The Effect of Different Forms of Centering in Hierarquical Linear Models. *Multivariate Behavioral Research* (30) p.1-21, 1995.

LINDLEY, D. V. e SMITH A. F. M. Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society*, B 34, p.1-41, 1972.

LONGFORD, N. T. A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects. *Biometrika* 74 (4), p.817-827, 1987.

LORD, F. M. *Applications of Item Response Theory to Practical Testing Problems*, Hillsdade: Lawrence Erlbaum Associates, 1980.

NATIS, L. Modelos Lineares Hierárquicos. Dissertação de Mestrado apresentado ao Instituto de Matemática e Estatística da Universidade de São Paulo, 2000.

NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAN, W. *Applied Linear Statistical Models*. Chicago: Irwin, 4ª edição, 1996.

PFEFFERMANN, D.; SKINNER, C. J.; HOLMES, D. J.; GOLDSTEIN, H.; RASBASH, J. Weighting for Unequal Selection Probabilities in Multilevel Models. *Journal of the Royal Statistical Society* B 60, p. 23-40, 1998.

PESSOA, D. G. C.; SILVA, P. L. N. (1998). Análise de Dados Amostrais Complexos, *XIII SINAPE*, São Paulo: Associação Brasileira de Estatística, 1998.

RASBASH, J.; BROWNE, W.; GOLDSTEIN. H.; YANG, M. et al. (2000). *A user's guide to MLwiN*, London: Institute of Education. 2ª edição, 2000.

SAS INSTITUTE INC. *SAS/STAT Software: Changes and Enhancements through Release 6.12*, Cary, NC: SAS Institute Inc., 1997.

SEN, P. K.; SINGER, J. M. *Large Sample Methods in Statistics. An Introduction with Applications*. New York: Chapman and Hall, 1993.

SINGER, J. D. Using SAS PROC MIXED to Fit Multilevel Models, Hierarquical Models and Individual Growth Models. *Journal of Educational and Behavioral Statistics* 24 (4), p. 323-355, 1998.

SMITH, A. F. M. (1973). A General Bayesian Linear Model. *Journal of the Royal Statistical Society* B 35, p. 61-75, 1973.