



Revista Brasileira de Educação

ISSN: 1413-2478

rbe@anped.org.br

Associação Nacional de Pós-Graduação e

Pesquisa em Educação

Brasil

Machado Soares, Tufi

Modelo de três níveis hierárquicos para a proficiência dos alunos de 4a série avaliados no teste de
língua portuguesa do SIMAVE/PROEB-2002

Revista Brasileira de Educação, núm. 29, maio-ago., 2005, pp. 73-87
Associação Nacional de Pós-Graduação e Pesquisa em Educação
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=27502907>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais artigos
- ▶ Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Modelo de três níveis hierárquicos para a proficiência dos alunos de 4^a série avaliados no teste de língua portuguesa do SIMAVE/PROEB-2002*

Tufi Machado Soares

Universidade Federal de Juiz de Fora, Faculdade de Educação e Departamento de Estatística

Introdução

Os modelos hierárquicos, largamente empregados em todo o mundo, vêm se consolidando no Brasil como um importante instrumento na análise e na interpretação dos resultados provenientes dos questionários, especialmente os utilizados em programas de avaliação educacional em larga escala, que avaliam o contexto socioeconômico, cultural, a trajetória acadêmica dos alunos e ainda o contexto socioeconômico e cultural dos professores e dos diretores, além de aspectos de suas atividades acadêmicas e administrativas. Esses modelos permitem avaliar a influência que as variáveis contextuais apresentam sobre a proficiência dos alunos. Por exemplo, Fletcher (1998)

analisou em sua pesquisa educacional os efeitos das características do ambiente escolar e do ambiente familiar no rendimento dos alunos. Nesse estudo o autor concluiu que as médias dos rendimentos dos alunos por escola, sem o ajuste das diferenças na composição social do alunado, distorcem os resultados das análises. Os dados utilizados por Fletcher foram os dados do Sistema Nacional de Avaliação da Educação Básica (SAEB) de 1995, colhidos na prova de matemática e de questionário aplicados a alunos de 8^a série. Barbosa e Fernandes (2000) utilizaram em sua pesquisa dados do SAEB-1997, colhidos em testes e questionários com o objetivo de estabelecer uma relação entre as variáveis explicativas de dois níveis (alunos e escolas) e o rendimento escolar dos alunos de 8^a série. O objetivo do estudo de Soares, Alves e Oliveira (2001) foi conhecer o efeito das escolas de nível médio no vestibular da Universidade Federal de Minas Gerais (UFMG) nos anos de 1998, 1999 e 2000, e, ao mesmo tempo, apresentar uma forma alternativa de avaliar os efeitos dessas escolas. Albernaz, Ferreira e Franco (2002) avaliaram o efeito de variáveis escolares, tais como a escolaridade do

* O autor agradece ao Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (UFJF) e à Secretaria Estadual de Educação do Estado de Minas Gerais pela cessão dos dados e o apoio a este trabalho. Agradece, ainda, as sugestões e comentários do parecerista, que contribuíram para a melhoria na apresentação dos resultados.

professor e a qualidade da infra-estrutura física no desempenho dos estudantes; esses autores também utilizam os dados provenientes das avaliações do SAEB-1997. Franco, Mandarino e Ortigão (2001) analisaram o efeito do projeto pedagógico sobre a qualidade e eqüidade das escolas. Soares (2003) avaliou a influência de características do professor e do ambiente em sala de aula sobre a proficiência em língua portuguesa dos alunos de 4^a série que participaram do Programa de Avaliação da Educação Básica (PROEB) do Estado de Minas Gerais em 2002.

Este estudo consiste na construção e interpretação de modelos de regressão hierárquicos em três níveis (aluno, turma, escola), aprofundando os resultados de Soares (2003) para explicar a proficiência dos alunos da 4^a série do ensino fundamental que participaram da avaliação em língua portuguesa realizada no ano de 2002 pelo PROEB. Foi possível obter informações provenientes de questionários de um total de 2.438 escolas, 6.286 turmas e 161.904 alunos. Do universo total foram excluídas as escolas cujo número de alunos presentes no teste foi inferior a dez. A principal contribuição desse trabalho foi construir um modelo que pôde ser estruturado em três níveis hierárquicos – pois, em geral, os estudos publicados têm considerado apenas dois níveis –, o que permitiu avaliar a proporção da variabilidade das proficiências dos alunos devida às diferenças entre os alunos, entre as turmas e entre as escolas. Esse tipo de modelo, em três níveis, apresenta dificuldades de construção, principalmente porque grande parte das escolas apresenta baixo número de turmas de uma mesma série, o que torna muitas vezes impossível distinguir o efeito relativo à turma do efeito relativo à escola, além de dificultar a estimativa dos parâmetros do modelo, por falta de informação. Tanto é assim que, num primeiro modelo produzido a partir do universo de todas as escolas, o efeito aleatório do 3^º nível não foi significativo. Felizmente, pela característica de avaliação universal do PROEB, identificou-se um conjunto de 395 escolas da rede estadual de Minas Gerais, com quatro turmas ou mais de 4^a série, e para essas escolas o efeito aleatório de 3^º nível não só é significati-

vo, mas substancialmente relevante (cerca de 11,5%). Portanto, neste trabalho, o universo final para o qual os modelos são construídos corresponde ao universo das escolas com quatro turmas ou mais de 4^a série.

Em primeiro lugar, será descrita a origem e as principais características do PROEB/Sistema Mineiro de Avaliação da Educação Pública (SIMAVE). No tópico seguinte serão apresentadas as variáveis consideradas no estudo para os níveis dos alunos, das turmas e das escolas. Depois o leitor será introduzido, sucintamente, nos modelos de regressão hierárquicos, e, em seguida, os modelos construídos serão apresentados e interpretados. As conclusões e as sugestões para estudos futuros são apresentadas ao fim.

O PROEB/SIMAVE

Em 2000, o governo do Estado de Minas Gerais, por meio da Secretaria Estadual de Educação, instituiu o Sistema Mineiro de Avaliação da Educação Pública (SIMAVE). Dentre as atividades que compõem o SIMAVE está sendo implementado, anualmente, o Programa de Avaliação da Rede Pública de Educação Básica (PROEB). Até 2002, o PROEB foi coordenado pelo Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (UFJF), Minas Gerais, e atualmente vem sendo diretamente coordenado pela Secretaria Estadual de Educação e operacionalizado com apoio do CAEd. O PROEB é um programa de avaliação que tem por objetivo avaliar as escolas da rede estadual. O ciclo de avaliação completa-se a cada dois ou três anos. Em novembro de 2000 foram aplicados testes com o objetivo de avaliar as competências dos alunos em língua portuguesa e matemática. Em 2001 foram avaliadas as competências em ciências humanas e naturais. Em 2002 foram aplicados novamente testes de língua portuguesa e em 2003, testes de matemática. Os testes são aplicados a todos os alunos da quarta e oitava séries do ensino fundamental e da terceira série do ensino médio da rede estadual. Além dos testes que avaliam as competências nessas disciplinas, o processo de avaliação inclui um questionário apli-

cado aos alunos, com o objetivo de obter dados sobre o perfil socioeconômico e a trajetória escolar dos estudantes, e, ainda, informações relevantes sobre o professor e características da turma. A metodologia empregada para avaliação pelo PROEB foi, em geral, a mesma utilizada pelo SAEB, implementado pelo Instituto Nacional de Estudos e Pesquisa (INEP) do Ministério da Educação (MEC), e é adequada para avaliação em larga escala, com precisão muito maior na avaliação das proficiências médias de grupos de alunos do que, propriamente, na avaliação individual. Os itens (no caso, questões de múltiplas escolhas) desses testes foram construídos por especialistas tendo como base a proposta curricular de Minas Gerais e as matrizes de competências utilizadas pelo SAEB. Utiliza-se uma metodologia de construção de testes que é denominada Blocos Incompletos Balanceados (BIB). Em 2002, a partir de um conjunto inicial de cerca de quinhentos itens para cada série avaliada, foram selecionados de acordo com suas características estatísticas um total de 169 itens. Esse total de itens foi disposto em 13 blocos com 13 itens cada. Seguindo uma combinação apropriada dos blocos, foram construídos 26 cadernos de teste, constituídos de três blocos cada. Cada aluno respondeu a um caderno e, consequentemente, a 39 itens. A forma de construção dos cadernos faz com que haja blocos comuns entre eles, produzindo, assim, o que se denomina de equalização horizontal dos escores estimados pela calibração simultânea dos 169 itens – isto é, a estimativa dos parâmetros dos modelos que são impostos aos itens –, o que permite a comparação dos resultados originados de respostas atribuídas aos diferentes cadernos de prova. O modelo que geralmente vem sendo utilizado na produção das proficiências é um modelo logístico de três parâmetros, cuja construção é baseada na Teoria da Resposta ao Item (TRI) (cf. Hambleton, Swaminathan & Rogers, 1991; Andrade, Tavares & Valle, 2000; Lord, 1980; Klein, 2003). Esse tipo de modelo permite produzir “escores” de proficiências que sejam independentes dos testes aplicados e da população de examinandos (*idem*), permitindo, além da aplicação de um maior

número de itens numa mesma avaliação, a comparação longitudinal dos resultados e comparações com outros sistemas de avaliação. Essas são as principais razões pelas quais seu emprego vem se universalizando em avaliações educacionais. Assim, após análises estatísticas para verificar propriedades dos itens e dos testes, tais como unidimensionalidade do teste aplicado, ausência de comportamento diferencial do item entre as principais regiões do estado e as séries avaliadas (no caso dos itens em comum), ajuste dos dados empíricos ao modelo do item proposto – o que às vezes conduz, criteriosamente, à eliminação de alguns itens –, produz-se a chamada calibração definitiva dos itens (isto é, a estimativa dos parâmetros dos modelos dos itens) e a estimativa dos escores da proficiência (ou habilidade) de cada aluno. Todo esse processo exige um certo grau de tecnicidades, que não é objetivo desse trabalho abordar, e que podem ser encontradas nas referências citadas. Para o presente estudo, as proficiências foram fornecidas pelo CAEd/UFJF.

As variáveis do estudo

Variáveis de nível dos alunos (1º nível hierárquico)

Inicialmente, são consideradas quatro variáveis básicas de grande importância, quais sejam: uma variável indicadora da condição socioeconômica do aluno (E-SOCIO), uma variável indicadora do sexo do aluno, masculino contrastando com o feminino (S-MASC), uma variável que mede, em anos, a defasagem de idade do aluno (DEFAS), calculada pela diferença entre a idade do aluno e a idade correta para a série considerada (no caso da 4^a série, 10 anos de idade, admitindo-se uma tolerância para aqueles que completam 11 anos no segundo semestre) e uma variável indicadora da raça negra do aluno (R-NEGRA). Essas variáveis são apontadas em estudos anteriores como muito importantes na explicação da proficiência e, além disso, na análise de impacto direto foram as mais explicativas dentre aquelas avaliadas nos ques-

tionários do PROEB-2002. Por isso, essas variáveis são consideradas em praticamente todos os estudos, permitindo, em certo aspecto, controlar o efeito líquido de outras variáveis sobre a proficiência.

O índice que representa a condição socioeconômica dos alunos, E-SOCIO, foi obtido de um escore construído com modelos para respostas graduadas, a partir de uma série de variáveis indicadoras de sua situação, como, por exemplo, se a residência da criança tem água encanada ou não, número de televisores, escolaridade do pai, e assim por diante. Detalhes sobre a construção desse tipo de índice são encontrados em Soares (2003) e nas referências citadas nesse estudo. Esse tipo de índice mostra-se bastante útil, principalmente como variável de controle em estudos desse tipo. Existem outros critérios e outras variáveis que têm sido utilizados como indicadores da condição socioeconômica, por exemplo, a renda familiar ou o nível educacional dos pais. Porém, um grande número dos entrevistados – no caso presente, alunos da 4ª série do ensino fundamental – desconhece esses dados familiares, o que torna a abordagem direta de se inquirir sobre a renda familiar menos precisa. Além disso, espera-se que a construção de um índice a partir de diversas variáveis indicadoras produza maior quantidade de informação do que apenas a utilização de uma variável indicadora. As demais variáveis de nível dos alunos consideradas neste estudo na tabela 1.

Essas variáveis são habitualmente empregadas nesse tipo de investigação, e o efeito que apresentam já foi extensivamente analisado. No entanto, serão utilizadas na construção do modelo final apresentado neste trabalho como uma forma de corroborar estudos anteriores e avaliar o efeito que produzem sobre a proficiência, relativamente às demais variáveis de outros níveis do modelo.

Variáveis de nível de turmas (2º nível hierárquico)

Infelizmente, não é possível, na base de dados do PROEB-2002, por uma série de razões técnicas, e

Tabela 1
Variáveis do nível do aluno

Variável	Descrição
IDADE	Idade do aluno em anos.
S-MASC	Se o aluno é do sexo masculino.
R-NEGRA	Se o aluno é de raça negra.
R-BRANCA	Se o aluno é de raça branca.
R-PARDA	Se o aluno é de raça parda.
E-SOCIO	Escore socioeconômico do aluno.
DEFAS	Defasagem escolar.
N-ABAND	Número de anos de abandono da escola pelo aluno.
N-REPET	Número de anos de repetência do aluno.
M-MAE	Se o aluno mora com a mãe.
M-PAI	Se o aluno mora com o pai.
M-AVOS	Se o aluno mora com os avós.
M-QUANTOS	Com quantas pessoas o aluno mora.
P-LER	Se o pai sabe ler.
M-LER	Se a mãe sabe ler.
E-MAE	Escolaridade em anos da mãe.
N-LIVROS	Número de livros na residência do aluno.
BOLSAE	Se a família do aluno recebe bolsa escola.
F-CULTOSM	Freqüência a cultos religiosos por parte da mãe.
F-CULTO	Freqüência a cultos religiosos por parte do aluno.
F-ALUNO	Número de dias aproximado de faltas do aluno.
F-PREESC	Número de anos de freqüência à pré-escola pelo aluno.
F-BIBLIO	Taxa de freqüência à biblioteca da escola.
F-ESTUD	Taxa de estudos por parte do aluno.
A-DEVER	Se os pais ajudam nos deveres escolares.

ainda para garantir o sigilo das respostas fornecidas pelo professor, associar diretamente a resposta dada pelo professor à sua turma e, consequentemente, à proficiência de seus alunos. A alternativa habitual consiste em tentar medir a influência dessas variáveis indiretamente, por meio da agregação das respostas dos professores de toda uma escola para as séries de ensino estudadas. Portanto, as variáveis originadas desse questionário estão associadas ao nível de escola, e não, como seria o ideal, ao nível de turma. O artifício utilizado no PROEB-2002 para produzir informações referentes a aspectos da turma e do professor foi o de introduzir perguntas caracterís-

ticas no questionário dos alunos. Particularmente no caso da 4^a série esse expediente é muito interessante, pois possibilita uma informação razoavelmente precisa sobre algumas questões específicas.

As variáveis associadas ao nível de turma, que foram obtidas a partir do questionário apresentado aos alunos, totalizaram nove variáveis e são descritas na seqüência. A primeira é o número aproximado de faltas do professor (F-PROF) durante o ano. No questionário, essa questão foi avaliada a partir de uma escala ordinal [(A) Nunca, (B) Raramente, (C) Algumas vezes, (D) Muitas vezes], e, para efeito de análise, neste trabalho foi aproximada para uma escala intervalar levando em consideração o número de dias letivos do ano, e agregada, por meio de sua média, para cada turma. Uma segunda variável é a que mede se o professor passa ou não deveres de casa (DEVER); portanto, sendo originalmente dicotômica, torna-se, quando ela é agregada ao nível de turma, um índice que indica o percentual de alunos da turma que afirmam que o seu professor passa deveres de casa. O turno de funcionamento da turma, se manhã (TURNO_M), tarde (TURNO_T) ou noite (TURNO_N), foram também considerados para análise do modelo construído. Além dessas, no questionário respondido pelos alunos dez questões foram introduzidas para avaliar aspectos como os de dedicação e empenho do professor, e ainda aspectos do ambiente em sala de aula. Essas questões, enumeradas de Q61 a Q70, foram apresentadas da seguinte forma: admitindo-se uma escala ordinal [(A) Em todas as aulas, (B) Na maioria das aulas, (C) Em algumas aulas, (D) Nunca], os alunos foram perguntados sobre a freqüência com que os fatos da tabela 2 ocorriam na sala de aula.

A escala ordinal foi aproximada para uma escala intervalar e, nesse caso também, adotando-se um critério de aproximação baseado no número de dias do ano letivo. Assim, à opção A (Em todas as aulas) foi atribuído o escore 200 (aproximadamente 200 dias letivos), à opção B (Na maioria das aulas) o escore 160, à opção C (Em algumas aulas) o escore 40 e, finalmente, à opção D (Nunca) o escore 0. Produziu-se, dessa forma, dez variáveis intervalares (Q61R –

Tabela 2
Questões sobre o ambiente em sala de aula

Questão	Descrição
Q61	O(a) professor(a) tem que esperar muito tempo até que os alunos façam silêncio.
Q62	O(a) professor(a) exige que os alunos estudem e prestem atenção nas aulas.
Q63	O(a) professor(a) exige que os trabalhos dos alunos sejam bem feitos.
Q64	O(a) professor(a) mostra interesse no aprendizado de todos os alunos.
Q65	O(a) professor(a) dá oportunidade de os alunos expressarem suas opiniões.
Q66	O(a) professor(a) está disponível para esclarecer as dúvidas dos alunos.
Q67	O(a) professor(a) continua a explicar até que todos os alunos entendam a matéria.
Q68	O(a) professor(a) se esforça para ajudar os alunos
Q69	Os alunos não prestam atenção ao que o professor fala.
Q70	Há barulho e desordem na sala de aula.

Q70R) a partir das quais as análises subseqüentes foram elaboradas. Após a construção dessas dez novas variáveis, procedeu-se ao estudo das dimensões latentes a elas associadas, por meio da análise fatorial (cf. Timm, 2002) produzida com o software SPSS®. Utilizou-se o método de extração por componentes principais a partir da matriz de correlação das variáveis, produzindo em seguida uma solução por meio do método de rotação VARIMAX (*idem*), muito embora, pelos resultados, esse procedimento de rotação seja praticamente desnecessário, tendo em vista que a solução obtida sem rotação já seria suficientemente conclusiva. Para a melhor solução encontrada, três dimensões latentes foram identificadas, e juntas explicam cerca de 71% da variância total dos dados. Para uma verificação da adequação da análise fatorial foi utilizada a medida KMO (Kaiser-Meyer-Olkin), que apresentou um valor estimado de 0,85, indicando uma adequação muito satisfatória. O coeficiente *alfa de Cronbach*, uma medida da fidedignidade dos índices

associados a cada dimensão latente, foi, respectivamente, 0,73, 0,61, 0,59. Para uma melhor explicação dessas técnicas e seu emprego nesse tipo de problema, além da referência citada o leitor pode consultar Soares e Mendonça (2003). A tabela 3 apresenta as chamadas *cargas* de cada variável na explicação de cada uma das dimensões, o que permite que se faça uma análise e interpretação dessas dimensões.

Foram reforçados em negrito na tabela 3 os valores das maiores cargas encontradas na explicação de cada dimensão, pois as variáveis que estão associadas a essas maiores cargas são usadas para a interpretação de cada fator. Assim, nota-se que a primeira dimensão latente é aquela que, num certo sentido, representa e mede o grau de interesse e dedicação do professor. A segunda dimensão representa, num certo sentido, o grau de motivação do aluno e, por consequência, o ambiente mais ou menos comportado em sala de aula. A terceira dimensão mede o grau de exigência do professor com relação a trabalhos e atenção às aulas. Os escores produzidos a partir da análise fatorial para essas três dimensões foram agregados, em valores médios, para cada turma, e posteriormente padronizados, de tal forma que seus efeitos sobre a proficiência possam ser diretamente comparados por intermédio dos coeficientes de regressão do modelo construído. Portanto, com esse procedimento foram

Tabela 3
Matriz de componentes rotacionados

Variável	Componentes		
	1	2	3
Q61R	-0.108	0.726	0.146
Q62R	0.253	0.077	0.804
Q63R	0.251	-0.017	0.805
Q64R	0.672	-0.012	0.405
Q65R	0.742	0.001	0.096
Q66R	0.792	-0.016	0.125
Q67R	0.790	0.013	0.183
Q68R	0.793	0.021	0.229
Q69R	0.175	0.646	-0.094
Q70R	-0.066	0.814	0.014

geradas três novas variáveis associadas ao nível de turma, representadas por (FAC1_T), (FAC2_T) e (FAC3_T), cada uma delas medindo uma das dimensões latentes associadas à sala de aula: respectivamente, o interesse e a dedicação do professor, o ambiente em sala de aula e o grau de exigência do professor.

Além das variáveis apresentadas, foi utilizado ainda o escore socioeconômico dos alunos médio por turma (E-SOCIOM) como uma variável associada ao nível de turma. Essa variável tem importância sobre tudo quando se analisa o efeito de interações com outras variáveis.

Variáveis de nível de escolas
(3º nível hierárquico)

Foram utilizadas como variáveis de nível de escola todas aquelas variáveis originadas do questionário aplicado ao professor no PROEB (2002) e agregadas para o nível de escola. E, também, outras variáveis da escola obtidas a partir dos dados do Censo Escolar de 2002. Todas são apresentadas na tabela 4.

Os índices de motivação dos professores, indisciplina dos alunos e de prejuízo produzidos pelas instalações da escola foram construídos a partir de quatro questões pertinentes (identificadas na análise fatorial) apresentadas aos professores com nível ordinal de medida. Para produção desses índices foi utilizado o modelo para respostas graduadas (Andrade, Tavares & Valle, 2000) da TRI através do software Parcale®. Na produção dos índices de leitura de livros e de revistas pelos professores foi utilizado o modelo de crédito parcial. Não se fará maiores observações sobre esses índices, pois não se encontrou influência estatisticamente significativa sobre a proficiência dos alunos nos modelos finais construídos e aqui apresentados.

Os índices construídos a partir das variáveis provenientes do censo escolar foram identificados através de análise fatorial e construídos a partir de extração por componentes principais e, de modo geral, são baseados no número de equipamentos/dependências dividido pelo número de alunos da escola.

Modelo de três níveis hierárquicos para a proficiência dos alunos de 4^a série...

Tabela 4
Variáveis do nível da escola

Variável	Descrição
BIBLI-E	Se a escola tem biblioteca.
E-SOCIOE	Nível socioeconômico médio dos alunos da escola.
NLESCOLA*	Número de alunos da escola.
ANO.PROE*	Média de anos dos professores da escola na profissão.
ANO.ESCE*	Média de anos trabalhados na escola pelos professores.
SALAE*	Salário médio dos professores da escola.
HORATE*	Horas médias trabalhadas pelos professores da escola.
NESC-E*	Número médio de escolas trabalhadas pelos professores da escola.
REDAF-E*	Renda familiar média dos professores da escola.
E-PROF*	Média de escolaridade dos professores da escola.
IDADE-E*	Idade média em anos dos professores da escola.
CONTIN-E*	Percentual dos professores que participaram de alguma formação continuada no ano.
MOTIVA-E*	Índice médio de motivação dos professores da escola.
INDISC-E*	Índice médio de indisciplina dos alunos avaliada pelos professores da escola.
INSTAL-E*	Índice médio do nível de prejuízo produzido pelas instalações da escola.
I-LIVE*	Média da escola do índice de leitura de livros do professor.
I-REVI*	Média da escola do índice de leitura de revistas do professor.
PEDAE*	Percentual de professores com formação em pedagogia.
LOCE**	Se a escola situa-se em região urbana.
PROPE**	Se a escola funciona em imóvel próprio.
INDDEPE**	Índice de existência de dependências físicas da escola (salas de aulas, laboratórios etc.) ¹
IMIDIAE**	Índice de posses de equipamentos de mídia da escola (TV, vídeo etc.) ¹
IBASI**	Índice de posses de equipamentos básicos da escola (fax, retroprojetor etc.) ¹
ICOMPE**	Índice de posses de equipamentos de informática da escola (computadores, impressoras etc.) ¹

*dados provenientes do SIMAVE

**dados provenientes do Censo Escolar 2002

¹Relativos ao número de alunos da escola

Modelos hierárquicos de três níveis

Modelos de regressão são usados em pesquisas educacionais, por exemplo, para explicar a proficiência alcançada pelos alunos a partir de variáveis contextuais tais como condição socioeconômica, sexo, experiência do professor e aparelhamento da escola. Basicamente, os modelos tradicionais admitem quatro hipóteses fundamentais para as características dos dados: linearidade, normalidade, independência para os elementos amostrais e homocedasticidade. Em pesquisa educacional, geralmente, a população de alunos está organizada em turmas de salas de aula, e estas, em escolas. Nesse contexto, alunos são ditos estarem aninhados dentro das turmas, e estas, dentro das escolas. Logo, a estrutura dos dados na população é naturalmente hierárquica. Essa estrutura hierárquica torna pouco razoável admitir a independência para as observações individuais (alunos, por exemplo), pois se estaria desprezando o efeito de agregação: alunos de uma mesma turma tendem a ser mais parecidos do que alunos de turmas diferentes, mesmo que apresentem uma série de características semelhantes como, por exemplo, o fato de advirem do mesmo estrato social.

Os modelos hierárquicos levam em consideração a estrutura de agrupamento dos dados, admitindo que cada turma de alunos, por exemplo, tenha um modelo de regressão próprio. Em contrapartida, tais modelos podem ser influenciados por variáveis do mesmo nível de agregação e variáveis de níveis de agregação superiores. Por exemplo, a experiência ou o método de ensino empregado pelo professor poderia afetar o desempenho do aluno de formas diferentes: diretamente, com o impacto sobre a proficiência esperada para o aluno, mas também, talvez, indiretamente, potencializando ou atenuando o efeito da variável sexo. Assim, com o emprego dos modelos hierárquicos, cada nível de agregação pode ser formalmente representado pelo seu modelo específico, e esses “submodelos” expressam o relacionamento entre as variáveis contextuais dentro do nível de agregação e, ainda, influenciam o efeito de outras

variáveis nos níveis inferiores. Não é objetivo aqui fazer uma apresentação abrangente sobre modelos hierárquicos. No anexo apresentam-se algumas formalizações técnicas, mas, para o leitor interessado, sugere-se como texto inicial o de Lee (2001).

A construção e a análise dos modelos

Introdução

O processo básico mais utilizado na construção de um modelo hierárquico é do tipo *bottom-up*, isto é, parte-se do modelo nulo e vai-se incluindo as variáveis segundo uma heurística definida pelo especialista, que neste estudo se baseia na verificação da significância dos coeficientes (parâmetros fixos e aleatórios) para cada modelo. Pode-se utilizar a estatística *deviance*, que é uma medida do grau de ajustamento dos dados ao modelo construído, para produzir um critério de escolha entre dois modelos aninhados. Isto é, utiliza-se a *deviance* para comparar um modelo mais simples com um modelo mais geral. Normalmente, os modelos com a *deviance* mais baixa são melhores; no entanto, para testar se a diferença entre as *deviances* de dois modelos é significativa ou não, emprega-se o teste de significância χ^2 com o número de graus de liberdade igual à diferença de parâmetros entre os dois modelos. Outro critério empregado para inclusão ou não de variáveis no modelo é o critério *Akaike Information Criterion* (AIC – Akaike *apud* Hox, 2001). Esse critério é utilizado para comparar-se modelos diferentes (normalmente aninhados), e é calculado a partir do valor da *deviance* adicionado a um fator que penaliza o número de parâmetros estimados. Segundo Hox (2001), esse critério pode ser empregado para decidir entre dois modelos hierárquicos. O critério AIC é dado por:

$$AIC = d + 2q$$

sendo que d é a *deviance* e q é o número de parâmetros estimados no modelo.

Normalmente, recomenda-se (*idem*) verificar as

interações entre as variáveis após a última variável ter sido incluída no modelo.

Grosso modo, a construção dos modelos aqui apresentados seguiu esses passos. Inicialmente analisou-se o modelo nulo, com o objetivo de avaliar a proporção da variância devida a cada nível hierárquico. Em seguida, introduziu-se quatro das principais variáveis de nível de aluno para produzir um modelo que se convencionou chamar de modelo básico – ou modelo de referência –, e em seguida, com a introdução das demais variáveis de aluno, produziu-se o modelo de nível de aluno (modelo 1). Numa segunda etapa, sempre seguindo a mesma heurística, foram introduzidas as variáveis de nível de turma (modelo 2) e, finalmente, variáveis de nível de escola (modelo 3). Dessa forma, pôde-se analisar a evolução da explicação alcançada produzida após a introdução de cada variável.

Modelo nulo

O primeiro modelo ajustado foi o modelo nulo. A partir dele é mais fácil estimar a proporção da variabilidade, calculada a partir das variâncias estimadas para os erros, entre os alunos, entre as turmas dentro das escolas e entre as escolas, utilizando-se as relações de (4) a (6). A estimativa obtida para o coeficiente fixo é apresentada a seguir. O valor entre parênteses representa o erro-padrão para essa estimativa, e também entre parênteses são apresentadas as estimativas para os parâmetros aleatórios (isto é, as variâncias dos erros):

$$y_{ijk} = \beta_{0jk} + e_{ijk} (\sigma_e^2 = 1311.40)$$

$$\beta_{0jk} = \gamma_{00k} + u_{0jk} (\sigma_{u_0}^2 = 732.44)$$

$$\gamma_{00k} = 198.45 (1.09) + r_{00k} (\sigma_{r_{00}}^2 = 266.19)$$

A partir desses resultados e utilizando as relações (4) a (6), chega-se às proporções de 56,7%, 31,9% e 11,5%, respectivamente, para a explicação da variabilidade dos resultados dos alunos, que é devida a características dos alunos, das turmas e das escolas na 4^a série. O mesmo resultado não foi encontrado na 8^a sé-

rie e na 3^a série do ensino médio. Por exemplo, na 8^a série, realizando-se o mesmo exercício de cálculo das proporções para a explicação da variabilidade devida a cada nível, encontrou-se, respectivamente, 79,9%, 15,4% e 5,7%, indicando maior variabilidade entre as turmas e entre as escolas nos resultados da 4^a série do que nas demais séries. A *deviance* calculada para esse modelo foi de 197967,5 (com quatro parâmetros estimados). Novamente, ressalta-se que só é possível alcançar essa partição significativa para a variância explicada entre os três níveis porque se está considerando o universo das escolas com quatro turmas ou mais. Numa amostra aleatória, de mesmo tamanho, de todas escolas, o efeito aleatório do terceiro nível é não-significativo, indicando que, como se esperava, fica difícil nesse caso distinguir o efeito relativo à variabilidade entre as turmas do efeito relativo à variabilidade entre as escolas.

Esse resultado é muito importante e muito significativo. Por um lado, ele traz consigo a expectativa de que se pode encontrar explicação para as diferenças de resultados encontrados para as proficiências dos alunos que sofram influência em razão das características associadas ao trabalho e ao perfil do professor, da composição da turma, e que sofram, ainda, influência por causa das características associadas à escola, como influência do projeto pedagógico, participação da comunidade, aspectos administrativos, entre outras. A identificação desses fatores abre a perspectiva de que o próprio sistema apresente em seu contexto parte das soluções que podem ser empregadas para uma melhoria dos níveis de proficiência. Possivelmente, em muitos casos essa melhoria carece apenas de uma atuação dos gestores que não implique grandes custos. Em resumo, parece que há bastante espaço para atuação dos gestores estaduais a partir da própria experiência do sistema de ensino com a 4^a série do ensino fundamental.

Modelo com variáveis de aluno

Esse modelo foi construído introduzindo-se, inicialmente, quatro variáveis referentes ao aluno, ex-

tremamente importantes na explicação da proficiência. Essas quatro variáveis, já extensivamente apontadas em diferentes estudos como importantes na explicação da proficiência, são: o E-SOCIO, o escore socioeconômico do aluno, a DEFAS, a defasagem em anos do aluno, a variável S-MASC, indicadora do sexo masculino do aluno, e a variável R-NEGRO, indicadora do aluno declarar-se de raça negra. Posteriormente, as demais variáveis de alunos foram sendo introduzidas, até a estabilização do modelo. Na tabela 5 apresentam-se os resultados estimados para os parâmetros do modelo.

Tabela 5
Modelo 1 (incluindo as variáveis do aluno)

Parâmetros	Estimativas (erro-padrão)*
FIXOS	
(Intercepto, β_{0jk} /Intercepto, γ_{00k} /Intercepto, π_{000})	- 191,04 (1,40)
(SEXO, β_{1jk} /Intercepto, γ_{10k} /Intercepto, π_{100})	-7,29 (0,54)
(R-NEGRO, β_{2jk} /Intercepto, γ_{20k} /Intercepto, π_{200})	-8,43 (0,90)
(E-SOCIO, β_{3jk} /Intercepto, γ_{30k} /Intercepto, π_{300})	3,16 (0,39)
(DEFAS, β_{4jk} /Intercepto, γ_{40k} /Intercepto, π_{400})	-4,77 (0,34)
(BOLSAE, β_{5jk} /Intercepto, γ_{50k} /Intercepto, π_{500})	-8,02 (0,70)
(F-PREES, β_{6jk} /Intercepto, γ_{60k} /Intercepto, π_{600})	3,51 (0,28)
(F-ESTUD, β_{7jk} /Intercepto, γ_{70k} /Intercepto, π_{700})	0,15 (0,001)
ALEATÓRIOS	
Nível 1: σ_e^2	1.210,24
Nível 2: (Intercepto, β_{0jk}) $\sigma_{u_0}^2$	514,12
Nível 2: (R-NEGRO, β_{2jk}) $\sigma_{u_2}^2$	66,05
Nível 2: (E-SOCIO, β_{3jk}) $\sigma_{u_3}^2$	4,74
Nível 2: (DEFAS, β_{4jk}) $\sigma_{u_4}^2$	13,02
Nível 2: (BOLSAE, β_{5jk}) $\sigma_{u_5}^2$	30,31
Nível 3: (Intercepto, β_{0jk} /Intercepto, γ_{00k}) $\sigma_{r_{00}}^2$	137,23
Nível 3: (E-SOCIO, β_{3jk} /Intercepto, γ_{30k}) $\sigma_{r_{30}}^2$	5,98
Nível 3: (DEFAS, β_{4jk} /Intercepto, γ_{40k}) $\sigma_{r_{40}}^2$	6,88
Nível 3: (BOLSAE, β_{5jk} /Intercepto, γ_{50k}) $\sigma_{r_{50}}^2$	18,54

Deviance: 195806,84

* Todos os valores apresentados são significativos para $p > 0,01$

** Todos os valores não apresentados foram não-significativos para $p < 0,05$

Observa-se que o valor de p (*p-value*) é uma medida da significância estatística dos parâmetros do

modelo e, consequentemente, da correlação entre a variável e a proficiência do aluno. Valores pequenos de p indicam que a variável é fortemente correlacionada com a proficiência. Em contrapartida, à medida que o valor de p aumenta, diminui a certeza de que existe essa correlação. Na literatura especializada, sugere-se um valor de p no máximo igual 0,05 para se afirmar que existe a correlação.

Como se pode notar pelo modelo estimado o, o escore socioeconômico apresenta um efeito positivo sobre a proficiência, enquanto as variáveis referentes a DEFAS (defasagem escolar do aluno), a R-NEGRO (raça negra) e SEXO (alunos do sexo masculino) apresentam efeito negativo, o que se mostra consistente com todas as análises realizadas, não só com os dados do PROEB, em amostras diferentes, mas também com estudos realizados com dados provenientes de outras avaliações, como as do SAEB. Note-se, ainda, que há evidentemente uma redução do valor da *deviance* em relação àquela encontrada para o modelo nulo, indicando grande aumento na explicação da proficiência. Esse fato é ainda reforçado pela queda na estimativa da variância do erro de nível 1, de cerca de 1311, no caso do modelo nulo, para 1210 no caso desse primeiro modelo. Ressalta-se, ainda, que as influências das variáveis R-NEGRO, E-SOCIO e DEFAS são dependentes da turma em que o aluno está inserido. Além disso, os efeitos das variáveis E-SOCIO e DEFAS são também dependentes da escola. Assim, os efeitos negativos da defasagem escolar e da condição socioeconômica do aluno sobre sua proficiência sofrem variações em função de características da turma e da escola. Em contrapartida, a associação negativa de R-NEGRO é dependente da turma, mas não se pode dizer que seja da escola.

Já a influência da variável SEXO não apresenta variação significativa que dependa da turma ou da escola. Assim, o pior resultado, em média, observado para os meninos, em língua portuguesa, é praticamente o mesmo, independente da turma ou da escola. Esse pior desempenho dos meninos é um fenômeno observado universalmente e, por uma série de estudos, é devido, em parte, por diferenças existentes entre o

desenvolvimento cognitivo dos meninos e das meninas. Aqui, caberia conjecturar se seria o caso de se pensar em adotar estratégias diferentes para a alfabetização de ambos os grupos.

Com respeito à influência das demais variáveis do aluno, chama a atenção a influência observada para as variáveis BOLSAE (se o aluno possui bolsa escola) e F-PREES (freqüência em anos da pré-escola). Como se pode observar, a freqüência à pré-escola tem influência significativa, e de grande impacto, sobre a proficiência. Assim, os alunos que possuem bolsa escola tendem a apresentar proficiência bem inferior aos que não a possuem. Claro que esse fato, em grande parte, está associado à baixa condição social desse aluno, e reproduz, portanto, um fenômeno já observado, mas não se deve esquecer que uma medida da condição social já está sendo considerada no modelo, o que tenderia a atenuar parte desse impacto. Pode ser, então, que existam fatores adicionais associados aos alunos para os quais a família recebia a bolsa escola, além daqueles associados à medida da condição socioeconômica, que também exercem influência no rendimento do aluno. Note-se ainda que o efeito da variável BOLSAE pode ser dependente da escola, mas não se encontrou evidência de que dependa da turma do aluno. Também não se encontraram evidências de que o efeito de F-PREES dependa da turma ou escola.

Modelo com variáveis de turma

Este modelo foi construído considerando-se as variáveis de turma descritas no segundo nível hierárquico, além de todas as variáveis de aluno consideradas na construção do modelo anterior. Não se vai repetir aqui os coeficientes correspondentes às variáveis de aluno, estimados no novo modelo, pois são estatisticamente equivalentes aos apresentados no modelo anterior. Esse novo modelo permite avaliar a evolução na explicação do modelo multinível a partir da inclusão das variáveis de turma (ver tabela 6).

Novamente, note-se a queda obtida para a estatística *deviance* desse modelo em relação àquela obtida para o modelo apresentado anteriormente, assim

Tabela 6
Modelo 2 (incluindo variáveis da turma)

Parâmetro	Estimativa (erro-padrão)	p-value
FIXOS		
(Intercepto, β_{0jk} / Intercepto, γ_{00k} / Intercepto, π_{000})	146,59 (6,06)	0,000
(Intercepto, β_{0jk} / E-SOCIO, γ_{01k} / Intercepto, π_{010})	20,87 (1,27)	0,000
(Intercepto, β_{0jk} / F-PROF, γ_{02k} / Intercepto, π_{020})	-1,29 (0,22)	0,000
(Intercepto, β_{0jk} / DEVER, γ_{03k} / Intercepto, π_{030})	33,34 (6,03)	0,000
(Intercepto, β_{0jk} / FAC1_T, γ_{04k} / Intercepto, π_{040})	12,52 (1,75)	0,000
(Intercepto, β_{0jk} / FAC2_T, γ_{05k} / Intercepto, π_{050})	-13,35 (0,98)	0,000
(Intercepto, β_{0jk} / FAC3_T, γ_{06k} / Intercepto, π_{060})	16,33 (1,87)	0,000
(DEFAS, β_{4jk} / E-SOCIO, γ_{41k} / Int., π_{410})	-2,04 (0,75)	0,006
ALEATÓRIOS		
Nível 1: σ^2_e	1.212,74	0,000
Nível 2: (Intercepto, β_{0jk}) $\sigma^2_{u_0}$	227,10	0,000
Nível 3: (Intercepto, β_{0jk} / Intercepto, γ_{00k}) $\sigma^2_{r_{00}}$	63,13	0,026
Nível 3: (Intercepto, β_{0jk} / F-PROF, γ_{02k}) $\sigma^2_{r_{40}}$	3,41	0,003
Nível 3: (Intercepto, β_{0jk} / FAC2_T, γ_{05k}) $\sigma^2_{r_{40}}$	65,42	0,000

Deviance: 19480,88

como a queda nas variâncias dos erros estimadas para o nível de turma, comparáveis às quedas observadas do modelo nulo para o modelo básico (com variáveis do aluno somente), indicando que as variáveis associadas à sala de aula produziram um aumento substancial na explicação da proficiência. Os resultados são coerentes com resultados anteriores (Soares, 2003) obtidos para toda a população, incluindo as escolas com três turmas ou menos. De fato, ao aumento na frequência de faltas do professor de uma turma (F-PROF) associa-se uma menor proficiência do aluno; e, da mesma forma, a um aumento na falta de motivação dos alunos, refletida no comportamento da turma (FAC2_T), associa-se uma queda na proficiência. Além disso, quanto maior a dedicação e a disponibilidade por parte do professor (FAC1_T), maior a proficiência do aluno. Aos professores mais exigentes (FAC3_T) e aos professores que passam deveres (DEVER) está associado, também, maior rendimento por parte dos alunos, sendo que essa última variável apresenta grande influência sobre a proficiência. Confir-

mando os resultados já observados para o efeito das escolas, note-se que um aumento no escore socioeconômico médio da turma (E-SOCIO) produz maior proficiência do aluno. Quanto à influência da escola sobre os efeitos das variáveis de turma, encontraram-se evidências de que a escola influencia no efeito correspondente às faltas do professor (F-PROF) e no efeito devido ao comportamento da turma (FAC2_T). Finalmente, não foram encontradas evidências de que o horário da turma (manhã, tarde ou noite) tenha impacto sobre a proficiência em português do aluno da 4^a série.

Na análise final das interações entre as variáveis dos diferentes níveis do modelo encontrou-se interação entre as variáveis E-SOCIO (o escore socioeconômico médio da turma) e a defasagem escolar do aluno, e a tendência é a de um maior nível socioeconômico médio da turma potencializar a influência negativa da condição socioeconômica do aluno.

Modelo com variáveis da escola

No modelo apresentado nesta seção foram incluídas as variáveis de nível de escola apresentadas no terceiro nível hierárquico. Esse modelo foi construído a partir do modelo apresentado na seção anterior, procurando-se incluir aquelas variáveis que, na ordem de inclusão no modelo, proporcionassem maior redução na estatística de *deviance*. Novamente, não serão apresentados os coeficientes estimados associados às variáveis de nível inferior, por apresentarem valores similares aos anteriores. Os parâmetros do modelo final são apresentados na tabela 7.

Assim, encontrou-se influência significativa, e relevante, das variáveis IDADE (idade média dos professores da escola), CONTIN (percentual de professores que participaram de formação continuada durante o ano) e ECOMP (índice de posses de equipamentos de informática da escola). Neste estudo, a primeira variável apresentou influência negativa, e as duas outras, influência positiva. No entanto, como já era esperado, o ganho na explicação da variabilidade devido à inclusão dessas variáveis de escola

Tabela 7
Modelo 2 (incluindo variáveis da escola)

Parâmetro	Estimativa (erro-padrão)	p-value
FIXOS		
(Int., β_{0jk} /Int., γ_{00k} /Int., π_{000})	162,1(8.75)	0,000
(Int., β_{0jk} /Int., γ_{00k} /IDADE, π_{000})	-0,37 (0.14)	0,010
(Int., β_{0jk} /Int., γ_{00k} /CONTIN, π_{000})	8,27 (3.53)	0,019
(Int., β_{0jk} /Int., γ_{00k} /ECOMP, π_{000})	1,30 (0.59)	0,030
ALEATÓRIOS		
Nível 1 – σ^2_e	1.213,2	
Nível 2 – (Intercepto, β_{0jk}) $\sigma^2_{u_{00}}$	226,3	0,001
Nível 3 – (Int., β_{0jk} /Int., γ_{00k}) $\sigma^2_{r_{00}}$	64,8	0,017

Deviance: 194781,60

foi relativamente pequeno quando comparado ao ganho obtido com a introdução de variáveis de turma e aluno.

Conclusão

Este estudo corroborou muitas conclusões de trabalhos anteriores, obtidas em outras populações. Além disso, introduziu novas variáveis explicativas ainda não estudadas em estudos nacionais. Por considerar uma estrutura de modelo em três níveis hierárquicos, permitiu que fosse analisada a participação da variabilidade devida às diferenças entre os alunos, entre as turmas e entre as escolas, para o sistema estadual de educação do ensino básico do Estado de Minas Gerais. Espera-se que o mesmo sirva como orientação a estudos futuros e auxilie os gestores da educação brasileira na tomada de decisões que afetem o sistema. Pode-se perceber que, pelo menos para a 4^a série, o sistema apresenta grande parcela da variabilidade dos resultados encontrados para a proficiência que é devida à turma (englobando o efeito do professor) e grande parcela devida à escola. Isso significa que, dentro do sistema estadual de educação de Minas Gerais, muito se pode ganhar com experiências e resultados que o próprio sistema já produz. O aproveitamento de experiências e padrões que fazem com que deter-

minadas escolas, que atendem a alunos de mesmo perfil, apresentem resultados bem superiores à média esperada para elas, pode indicar um caminho a ser seguido para se alcançar rapidamente melhorias nos resultados gerais. Pode também apontar aquelas escolas que necessitam mais rapidamente de apoio pedagógico (inclusive material) para alcançar o desempenho das demais. Estudos adicionais para avaliar o valor agregado de cada escola do sistema estão sendo realizados com o objetivo de nortear ações pedagógicas e administrativas por parte dos diversos atores responsáveis, que não se constituem apenas nos governos.

Os resultados suscitam diversos questionamentos, que podem dar origem a novos focos de investigação, tais como: a grande variabilidade observada para os resultados das turmas é devida à seletividade dos alunos na construção das turmas dentro das escolas, ou é apenas influenciada pela variabilidade no desempenho do professor, que, como nos parece claro, tem grande importância? A influência do efeito racial sobre a proficiência, dependente de características da turma do aluno, pode indicar algum tipo de discriminação, ou ainda existiriam outras explicações para esse efeito?

TUFİ MACHADO SOARES, doutor em teoria matemática de controle e estatística pelo Programa de Engenharia Elétrica Pontifícia Universidade Católica do Rio de Janeiro, é professor do mestrado em educação da Universidade Federal de Juiz de Fora (UFJF), professor adjunto do Departamento de Estatística na mesma instituição e pesquisador associado ao Centro de Políticas Públicas e Avaliação da Educação (CAEd) da UFJF. Principais trabalhos publicados sobre o tema: Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no SIMAVE-2002 (*Estudos em Avaliação Educacional*, São Paulo, nº 28, p. 103-124, 2003); Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000 (*Pesquisa Operacional*, São Paulo, nº 3, v. 23, p. 421-442, em co-autoria com Márcia Meneghin Mendonça, 2003); Estudo de adequação para modelos da TRI aplicado ao caso do ensino fundamental da microrregião de Juiz de Fora em 1999 (*Educação em Foco*, São Paulo, nº 2, v. 6, p. 91-108, em co-autoria com Danielle R. Miranda Pereira, 2001). E-mail: tufi@estatistica.ufjf.br

Referências bibliográficas

- ALBERNAZ, A., FERREIRA, F. H. G., FRANCO, C., (2002). Qualidade e eqüidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, Juiz de Fora, v. 32, nº 3, p. 45-59.
- ANDRADE, D. F., TAVARES, H. R., VALLE, R. C., (2000). *Teoria da resposta ao item: conceitos e aplicações*. São Paulo: ABE (Associação Brasileira de Estatística).
- BARBOSA, M. E., FERNANDES, C., (2000). Modelo multinível: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional*, nº 22, p. 135-153.
- BRYK, A., RAUDENBUSH, S., (1992). *Hierarchical linear models*. Newbury Park (Califórnia): Sage Publications.
- FLETCHER, P., (1998). *À procura do ensino eficaz*. Rio de Janeiro: Ministério da Educação e Cultura, Departamento da Avaliação da Educação Básica (mimeo.).
- FRANCO, C., MANDARINO, M., ORTIGÃO, M. I., (2001). Projeto pedagógico de escola promove qualidade e eqüidade em educação? *Revista UNDIME-RJ*, v. 7, nº 2, p. 30-46.
- GOLDSTEIN, H., (1995). *Multilevel statistical models*. 2^a ed. New York: John Wiley & Sons.
- HAMBLETON, R., SWAMINATHAN, H., ROGERS, J., (1991). *Fundamentals of item response theory*. Newbury Park (Califórnia) Sage Publications.
- HOX, J., (2001). Multilevel analysis of regression and structural equation models. In: HOX, J. (org.). *Hierarchical models of survey data*. Michigan: 54th Summer Institute, p. 70-125.
- KLEIN, R., (1997). Indicadores educacionais: disparidades regionais e socioeconômicas no Brasil. In: BOMENY, H. *Avaliação e determinação de padrões na educação latino-americana: realidades e desafios*. Rio de Janeiro: Fundação Getúlio Vargas.
- _____, (2003). Utilização da teoria da resposta ao item no Sistema Nacional de Avaliação da Educação Básica. *Ensaio: avaliação e políticas públicas em educação*. Rio de Janeiro, v. 11, nº 40, p. 282-296.
- LEE, V. L., (2001). What are multilevel questions, and how might we explore them with quantitative methods? *Estudos em Avaliação Educacional*, nº 24, p. 31-68.
- LORD, F., (1980). *Applications of item response theory to practical testing problems*. Hillsdale (New York): Lawrence Erlbaum.
- MINAS GERAIS (Estado), (2001). *Avaliação da educação: relatório técnico do PROEB/SIMAVE-2000*. Juiz de Fora: Secretaria Estadual de Educação.
- RAUDENBUSH, S., BRYK, A., CHEONH, Y., CONGDON, R., (2000). *HLM5 – Hierarchical linear and nonlinear modeling*. Illinois: Scientific Software International.
- SOARES, T. M., (2003). Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no SIMAVE-2002. *Estudos em Avaliação Educacional*, São Paulo, nº 28, p. 103-124.
- SOARES, T. M., MENDONÇA, M. C. M., (2003). Construção de um modelo de regressão hierárquico para os dados do SIMAVE-2000. *Pesquisa Operacional*, Rio de Janeiro, nº 3, v. 23, p. 421-442.
- SOARES, J. F., ALVES, M. T. G., OLIVEIRA, R. M., (2001). O efeito de 248 escolas de nível médio no vestibular da UFMG nos anos de 1998, 1999 e 2000. *Estudos em Avaliação Educacional*, São Paulo, nº 24, p. 69-117.
- TIMM, H., (2002). *Applied multivariate analysis*. New York: Springer Text in Statistics, Springer Verlag.

Recebido em julho de 2004

Aprovado em fevereiro de 2005

ANEXO

(Modelo com três níveis hierárquicos)

Os sistemas escolares são um exemplo típico de estrutura hierárquica, pois alunos, turmas e escolas constituem-se de uma seqüência de agrupamentos naturalmente aninhados.

O modelo hierárquico (Bryk & Raudenbush, 1992) leva em consideração a estrutura de agrupamento dos dados. Concretamente, isso se reflete na especificação do modelo, como, por exemplo, da seguinte forma: para o modelo de regressão clássico o intercepto e o coeficiente de inclinação são parâmetros fixos, enquanto para o modelo multinível o intercepto e o coeficiente de inclinação podem ser considerados parâmetros aleatórios, dependentes da influência do nível hierárquico mais alto.

As análises que consideram em seus modelos a estrutura de agrupamento dos dados têm várias vantagens, como, por exemplo: (i) baseiam-se em modelos mais flexíveis e estruturados que utilizam melhor a informação presente na amostra e, ainda, fornecem uma equação para cada escola ou turma, conforme o caso, o que permite análises individuais para cada grupo; (ii) o uso da informação do agrupamento dos dados possibilita formular e testar hipóteses relativas a efeitos entre os diferentes níveis; (iii) permitem a partição da variabilidade da variável-resposta entre a explicação que é devida aos diversos níveis.

Grosso modo, os modelos de regressão têm por objetivo descrever, por meio de um modelo estatístico, a relação entre variáveis explicativas e independentes, inerentes aos diversos níveis hierárquicos, e uma variável dependente y (ou mais de uma, no caso de modelos multivariados). Neste trabalho, os dados coletados apresentam uma estrutura com três níveis, estando as unidades do primeiro nível (alunos) agrupadas conforme as unidades do segundo nível (turmas) e as unidades do segundo nível agrupadas em unidades do terceiro nível (escolas). Cada aluno é representado pelo índice i , o índice j representa cada turma e o índice k representa cada escola. Suponha

que x represente, genericamente, uma variável do aluno, w uma variável da turma e z uma variável da escola. O modelo então (Bryk & Raudenbush, 1992) terá a seguinte expressão geral:

$$y_{ijk} = \beta_{0jk} + \sum_{f=1}^F \beta_{fjk} x_{fijk} + e_{ijk}, \quad (1)$$

$$\beta_{fjk} = \gamma_{f0k} + \sum_{s=1}^S \gamma_{fsk} w_{sjk} + u_{fjk}, \quad f = 0, \dots, F \quad (2)$$

$$\gamma_{fsk} = \pi_{fs0} + \sum_{t=1}^T \pi_{fst} z_{tk} + r_{fsk}, \quad f = 0, \dots, F \text{ e } s = 0, \dots, S. \quad (3)$$

Note-se que nessas expressões F representa o número de variáveis do primeiro nível, S o número de variáveis do segundo nível e T o número de variáveis do terceiro nível. y_{ijk} representa a proficiência do i -ésimo aluno da j -ésima turma da k -ésima escola.

As hipóteses de descorrelação dos erros dos diferentes níveis, também adotadas em modelos com dois níveis, são mantidas, isto é, $E(e_{ijk}) = 0$, $E(e_{fjk}) = 0$, e $E(r_{fsk} u_{fjk}) = 0$. Os erros do modelo são todos supostos normais. σ^2_e , $\sigma^2_{u_f}$ e $\sigma^2_{r_{fs}}$ são as variâncias dos erros do modelo nos diferentes níveis, e também são denominadas como as componentes aleatórias do modelo. Os parâmetros π_{fst} são os parâmetros fixos do modelo.

O modelo de três níveis incondicional, ou modelo nulo (isto é, sem variáveis explicativas), é, então, representado por:

$$y_{ijk} = \beta_{0jk} + e_{ijk}$$

$$\beta_{0jk} = \gamma_{00k} + u_{0jk}$$

$$\gamma_{00k} = \pi_{000} + r_{00k}$$

e, nesse caso, devido à hipótese de descorrelação entre os erros de diferentes níveis do modelo, segue que a proporção da variância explicada devida a cada nível é dada por:

$$\frac{\sigma_e^2}{\sigma_e^2 + \sigma_{u_0}^2 + \sigma_{r_{00}}^2} , \text{ para o nível 1} \quad (4)$$

$$\frac{\sigma_{u_0}^2}{\sigma_e^2 + \sigma_{u_0}^2 + \sigma_{r_{00}}^2} , \text{ para o nível 2} \quad (5)$$

$$\frac{\sigma_{r_{00}}^2}{\sigma_e^2 + \sigma_{u_0}^2 + \sigma_{r_{00}}^2} , \text{ para o nível 3} \quad (6)$$

Neste trabalho, a estimativa dos coeficientes fixos e a estimativa das componentes de variância são realizadas por meio do método de máxima verossimilhança (Bryk & Raudenbush, 1992). Para tanto, foi utilizado o software HLM5® (Raudenbush *et al.*, 2000). A medida de ajuste do modelo utilizada foi a denominada “estatística de *deviance*”, definida por:

$$D = -2 \log(L),$$

em que L é o valor da função de verossimilhança avaliada em seu valor. Essa medida será usada para avaliar o grau de explicação alcançado pelos modelos construídos a partir do modelo incondicional.