

UM MODELO MULTINÍVEL PARA A PROFICIÊNCIA EM MATEMÁTICA NA 4ª SÉRIE DO ENSINO FUNDAMENTAL DE MINAS GERAIS - SAEB/99

Nilton C. Santos, Cláudio Felipe Ribeiro da Silva.

CLAVES/FIOCRUZ – ENCE/IBGE

INTRODUÇÃO

Neste estudo foram utilizados dados extraídos da base de dados do Sistema Nacional de Avaliação da educação Básica – SAEB, INEP/MEC.

O Instituto Nacional de Estudos e Pesquisas Nacionais (INEP) é um órgão vinculado ao ministério da educação que em parceria com as secretarias estaduais de educação desenvolve iniciativas conjuntas na área de avaliação educacional. Esta parceria é muito importante pois traz subsídios para o desenvolvimento de programas de formação de professores, apoiados em aspectos qualitativos identificados pelas avaliações. O Sistema Nacional de Avaliação da educação Básica – SAEB 99, mostra em seus relatórios que os alunos avaliados nas 4ª e 8ª séries do ensino fundamental e na 3ª série do ensino médio mantiveram-se nos mesmos níveis de desempenho em matemática e português, em comparação com o SAEB 97.

“ O SAEB confirma não só a incorporação de novos segmentos ao sistema educacional como também a permanência dos alunos no sistema. Este continua a ser o grande desafio da educação básica no Brasil: Desmontar o mito da velha escola pública de qualidade restrita às elites, e construir a nova escola pública de qualidade, mais democrática e inclusiva capaz de incorporar de forma competente os historicamente excluídos da vida cidadã”, afirma o ministro da Educação, Paulo Renato Souza.

As informações geradas pelo SAEB visam subsidiar a definição de ações voltadas para correção das distorções e debilidades identificadas no sistemas educacional de forma que as esferas políticas possam dirigir sua atenção tanto para o desenvolvimento e aperfeiçoamento do sistema de Ensino, quanto para a redução das desigualdades nele existentes.

O trabalho discute e analisa o desempenho em matemática dos alunos da 4ª série do ensino fundamental do Estado de Minas Gerais. Neste contexto é proposta a utilização de modelos hierárquicos, segundo diferentes níveis sociais, étnicos e acadêmicos, sinalizando fatores que influenciam na proficiência do aluno para que políticas públicas possam ser implementadas no intuito de reduzir as diferenças e incluir os excluídos à uma nova realidade acadêmica.

OBJETIVOS

Este trabalho tem como objetivo principal identificar as características que tornam escolas e sistemas de ensino mais eficazes, promovendo assim a aprendizagem, identificando pontos positivos através da utilização de modelos hierárquicos na intenção de descobrir os fatores que afetam o desempenho acadêmico dos alunos da 4ª série do Ensino Fundamental do Estado de Minas Gerais na disciplina de matemática utilizando informações sócioeconômicas, étnicas e acadêmicas, definindo a proficiência média e as variáveis que contribuem de forma positiva e/ou negativa para o aumento ou decréscimo deste rendimento acadêmico.

MATERIAL E MÉTODO

A cobertura do SAEB é feita por meio de amostras aleatórias – cuja característica principal é a sua natureza probabilística - representativas do país e de cada uma das 27 unidades da federação. São avaliados os alunos das 4ª e 8ª series do Ensino Fundamental e da 3ª série do Ensino Médio. O SA EB é aplicado de dois em dois anos. Já foram realizadas 5 avaliações desde 1990. São verificados alguns aspectos contextuais e escolares que incidem na qualidade do Ensino Básico, tais como: as características de infra estrutura e de disponibilidades da unidade escolar, o perfil do diretor e os mecanismos de gestão escolar, o perfil e a prática docente, as características socioculturais e os hábitos

de estudos dos alunos. O planejamento amostral do SAEB/99 considera a população de alunos dividida em três universos segundo a série que estão matriculados. Por não existir um cadastro de alunos individuais de onde se possa retirar uma amostra diretamente de alunos, foi necessário adotar uma amostragem conglomerada de alunos. A partir dos dados do Censo Escolar de 1998 e de posse de um cadastro de escolas, adotou-se um plano amostral em duas etapas para a seleção dos alunos testados como parte na avaliação do SAEB/99. Na Segunda etapa foram selecionadas turmas dentro das escola selecionadas na primeira etapa para cada uma das séries. Uma vez selecionada a turma todos os alunos participaram da avaliação. Foram aplicados testes de desempenho dos alunos, elaborados com base em matrizes curriculares validadas nacionalmente. Esses testes medem tanto os conteúdos quanto as habilidades em seus diversos níveis de complexidade. Os testes de matemática são de resposta fechadas (múltipla escolha). Os testes para os alunos são aplicados por agentes externos à escola, com procedimentos unificados e tempo controlado. Estes mesmos agentes coletam informações sobre a escola. Os questionários do professor e do diretor são auto-administrados.

A amostra consta de 1061 alunos e 195 escolas. A modelagem dos dados foram feitas a partir do software *MLWin* 1.1, para a análise exploratória dos dados utilizamos o software SPSS 9.0 e para o fechamento do relatório utilizamos o WORD 97.

MODELO MULTINÍVEL

A partir do nosso exemplo em que temos um estudo em dois níveis, alunos (nível 1) e escolas (nível 2), Discorreremos brevemente sobre Modelo de Regressão Multinível.

O modelo multinível, com intercepto aleatório é definido assim:

$$y_{ij} = b_{0j} + b_1 x_{ij} + e_{ij}$$

onde,

$$\mathbf{b}_{0j} = \mathbf{b}_0 + u_{0j}$$

logo,

$$y_{ij} = \mathbf{b}_0 + \mathbf{b}_1 x_{ij} + u_{0j} + e_{ij}$$

sendo,

y_{ij} a resposta do i -ésimo aluno (nível 1) no j -ésimo nível 2 (escola);

\mathbf{b}_0 e \mathbf{b}_1 parâmetros fixos;

x_{ij} vetor de covariáveis tendo fixo o efeito \mathbf{b}_1 ;

e_{ij} erro associado ao nível 1, parâmetro aleatório, distribuição Normal, $N(0, \mathbf{s}_{\varepsilon}^2)$;

u_{0j} variação associada ao nível 2, parâmetro aleatório, distribuição $N(0, \mathbf{s}_{u0}^2)$.

Assumimos que as variáveis aleatórias de diferentes níveis são não correlacionadas, isto é, $\text{cov}(\mathbf{e}_{ij}, u_{ij}) = 0$. E que as observações em diferentes clusters são independentes, com $\text{cov}(y_{ij}, y_{i'j'}) = 0$. Entretanto, as observações dos mesmos clusters são correlacionadas, $\text{cov}(y_{ij}, y_{i'j}) = \text{cov}(u_{0j}, u_{0j}) = \mathbf{s}_{u0}^2$ e a correlação entre dois estudantes é portanto medida como:

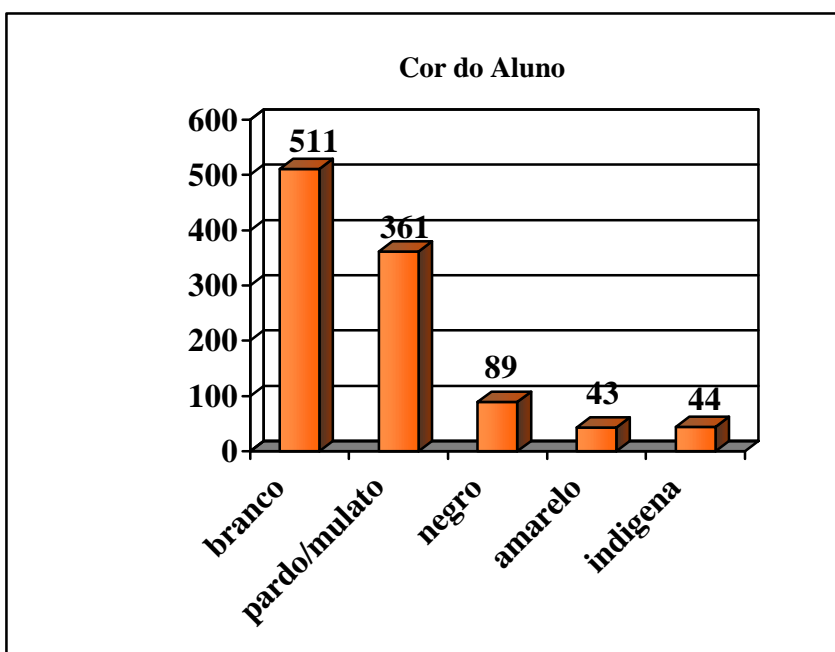
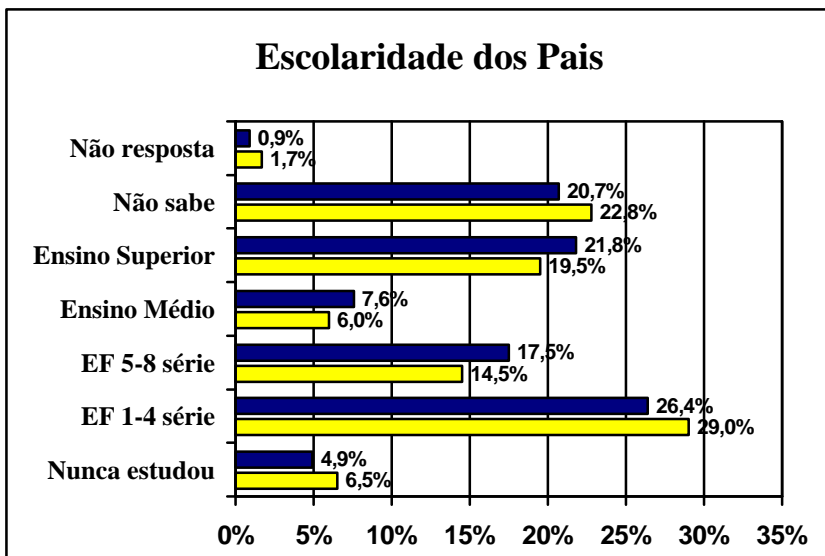
$$\mathbf{r} = \frac{\mathbf{s}_{u0}^2}{(\mathbf{s}_{u0}^2 + \mathbf{s}_{e0}^2)}$$

que é conhecida como “correlação dentro das unidades de nível 2”; neste caso a correlação intra-escola. Esta correlação mede a proporção do total da variância que é explicada entre as escolas.

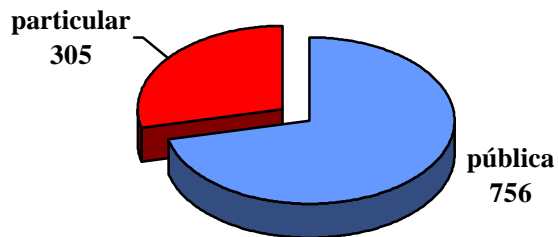
JUSTIFICATIVA DA ESCOLHA DO MODELO

Quando nos deparamos com uma população com estrutura hierárquica, faz-se necessário verificarmos se estas diferenças de níveis podem influenciar nas nossas estimações. No nosso caso em que estamos estudando proficiência do aluno cabe verificar se a escola não exerce influência nesta proficiência. Assim através de modelos hierárquicos podemos verificar os fenômenos intra-escolares e inter-escolares. A modelagem multinível consiste então neste caso em uma equação de regressão para o primeiro nível (intercepto e inclinações) e um conjunto de regressões para o segundo nível (variáveis dependentes).

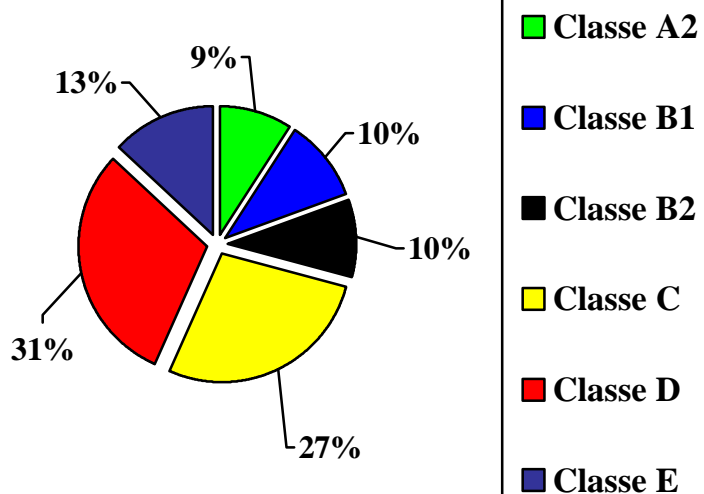
ANÁLISE EXPLORATÓRIA DOS DADOS



Distribuição de Alunos por tipo de Rede



Classe Econômica do Aluno



AQUI ENTRAM OS GRÁFICOS

ANÁLISE ESTATÍSTICA

As variáveis utilizadas no modelo são do tipo dicotômicas com exceção de **profic** (proficiência do aluno em matemática) e **a_leitur** (hábitos de leitura do aluno) que são contínuas. O modelo final foi escolhido após procedermos algumas

etapas de estimação: o primeiro modelo (incondicional) foi calculado apenas com o intercepto aleatório conforme mostramos abaixo:

$$\text{profic}_{ij} \sim N(XB, \Omega)$$

$$\text{profic}_{ij} = \beta_{0ij} \text{cons}$$

$$\beta_{0ij} = 185.180(2.443) + u_{0ij} + e_{0ij}$$

$$\begin{bmatrix} u_{0ij} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 846.633(121.072) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 1308.632(62.553) \end{bmatrix}$$

$$-2 * \loglikelihood(IGLS) = 11018.250(1061 \text{ of } 1061 \text{ cases in use})$$

O coeficiente de correlação intra-classe encontrada foi de 39%, o que justifica a aplicação de uma modelagem multinível, a média da proficiência em matemática na 4ª série do Ensino Fundamental de Minas Gerais é de 185,180 pontos com desvio padrão de 2,443, a variância no nível 2 (escola) é de 846,633 com desvio padrão de 121,072 e a variância de nível 1 (aluno) é de 1308,632 com desvio padrão de 62,553.

O modelo seguinte:

$$\text{profic}_{ij} \sim N(XB, \Omega)$$

$$\begin{aligned} \text{profic}_{ij} = & \beta_{0ij} \text{cons} + 50.592(10.053) \text{a_cla1}_{ij} + 36.630(6.801) \text{a_cla2}_{ij} + 30.925(6.139) \text{a_clb1}_{ij} + \\ & 28.354(6.043) \text{a_clb2}_{ij} + 16.550(4.364) \text{a_clc}_{ij} + 10.394(4.128) \text{a_cld}_{ij} + \\ & -1.695(6.002) \text{a_bran}_{ij} + -7.485(6.083) \text{a_pard}_{ij} + -16.403(7.094) \text{a_negr}_{ij} + \\ & 2.253(8.023) \text{a_amar}_{ij} + -7.062(1.969) \text{a_trab}_{ij} + -9.099(1.293) \text{a_repet}_{ij} + \\ & 17.435(4.417) \text{a_licao}_{ij} + -15.030(16.912) \text{d_cla1}_j + -19.177(16.673) \text{d_cla2}_j + \\ & -25.601(16.652) \text{d_clb1}_j + -23.444(16.860) \text{d_clb2}_j + -30.870(16.627) \text{d_clc}_j + \\ & -30.692(4.724) \text{rpub}_j \end{aligned}$$

$$\beta_{0ij} = 222.034(18.451) + u_{0ij} + e_{0ij}$$

$$[u_{0ij}] \sim N(0, \Omega_u) : \Omega_u = [113.391(37.043)]$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [1219.000(60.070)]$$

$$-2 * \text{loglikelihood}(IGLS) = 9835.380(756 \text{ of } 1061 \text{ cases in use})$$

O modelo foi gerado com a inclusão de todas as variáveis com o parâmetro fixo e sem as variáveis que segundo o teste de ajuste foram consideradas não significativas – *Intervals and Tests / MLWin* . ao prosseguirmos com o procedimento de estimação e agora incluindo algumas variáveis relativas ao aluno com efeito no nível 2, verificamos que apenas a variável **a_repet** (se o aluno já repetiu de ano pelo menos uma vez) tinha efeito aleatório, mas ao procedermos o teste de qualidade do ajuste verificamos que a mesma não era significativa, paramos então no modelo anterior, do qual apresentamos na tabela abaixo os resultados encontrados:

Efeito Fixo			
Variáveis		Coeficiente	S. E.
Intercepto		222,034	18,451
Classe Econômica do aluno	Classe A1	50,592	10,053
	Classe A2	36,630	6,801
	Classe B1	30,925	6,139

	Classe B2	28,354	6,043
	Classe C	16,550	4,364
	Classe D	10,394	4,128
Cor	Branca	-1,695	6,002
	Parda/mulata	-7,485	6,083
	Negra	-16,403	7,094
	Amarela	2,253	8,023
Trabalha		-7,062	1,969
Já repetiu de ano		-9,099	1,293
Faz lição de casa		17,435	4,417
Classe econômica do Diretor	Classe A1	-15,030	16,912
	Classe A2	-19,177	16,673
	Classe B1	-25,601	16,652
	Classe B2	-23,444	16,860
	Classe C	-30,870	16,627
Rede	Pública	-30,692	4,724
Efeito Aleatório			
	Variáveis	Var.	S. E.
Intercepto		113,391	37,043

CONCLUSÕES

Através do coeficiente de correlação intra-classe de 39% medido a partir do modelo incondicional podemos concluir que a proficiência em matemática na 4^a série do Ensino Fundamental do estado de Minas Gerais varia segundo as escolas. Ao ajustarmos o modelo de parâmetros fixos, obtivemos um coeficiente de correlação intra classe de 8,5%, o que significa que a inclusão destas variáveis reduziu em 78,2% a variação entre escolas.

Podemos observar neste modelo que quanto mais elevada a classe econômica do aluno maior é sua proficiência. Embora a cor do aluno tenha um impacto negativo na proficiência, com exceção da cor amarela, os alunos da cor negra tem uma perda bem maior do que os outros. O fato de o aluno trabalhar, já ter repetido de ano e estudar em escola da rede pública faz com que este tenha um decréscimo na sua proficiência que pode chegar até 46,85 pontos , já os alunos que fazem a lição de casa conseguem aumentar a sua média em 17,43 pontos. Embora o impacto da classe social do diretor seja negativo na proficiência do aluno, este é minorado a medida que o diretor se encontra em classes mais elevadas. A variância entre as escolas é reduzida em 87% e a variância intra-escola é reduzida em 6,8%.

REFERÊNCIAS BIBLIOGRÁFICAS

BARBOSA, Maria Eugênia Ferrão; FERNANDES, Cristiano. **Modelo multinível: uma aplicação a dados de avaliação educacional**. Estudos em Avaliação Educacional, Fundação Carlos Chagas. Rio de Janeiro, v. 22, p. 135-153, 2000.

KREFT, Ita, LEEUW, Jan de. **Intruducing multilevel modeling**. London, Sage Publications.

LEE, Valerie E., BRYK, Anthony S. **A multilevel model of the social distribution of high school achievement**. A Multilevel Model of The Social Distribution Sociology of Education, American Sociological Association, v. 62, p. 172-192, 1989.

MALUF, Mônica Maia Bonel. **Sistema nacional de avaliação da educação básica no Brasil: análises e proposições**. Estudos em Avaliação Educacional, Fundação Carlos Chagas. Rio de Janeiro, v. 14, p. 5-38, 1996.

PFEFFERMANN, D. Skinner C.J. and D.J. Holmes, H. Goldstein and J. Rabash
“Weighting for unequal selection probabilities in multilevel models”, J.R. Statist.
Soc. B, 1998, part1:23-40

SNIDGERS, T. A. B., Bosker, R. J. (1999). Multilevel Analysis - An Introduction to
Basic and Advanced Multilevel Modeling. JAGE Publi

YANG, M., GOLDSTEIN, H., RATH, T., HILL, N. **The use of assessment data for
school improvement purposes.** Oxford Review of Education, v. 25, n. 4, p. 469-
483, 1999.

PEREIRA, Júlio Cesar R., L. **Análise de Dados Qualitativos.** São Paulo: Editora
da Universidade de São Paulo. 1999.