

UTILIZAÇÃO DA TEORIA DA RESPOSTA AO ITEM NA PRODUÇÃO DE INDICADORES SÓCIO-ECONÔMICOS

Tufi Machado Soares

Departamento de Estatística
Centro de Avaliação Educacional (CAEd)
Universidade Federal de Juiz de Fora (UFJF)
Juiz de Fora – MG
tufi@estatistica.ufjf.br

*Recebido em 03/2004; aceito em 01/2005
Received March 2004; accepted January 2005*

Resumo

No Brasil a Teoria da Resposta ao Item (TRI) tem sido empregada principalmente na produção de índices de proficiência para alunos que respondem a testes de avaliação educacional em larga escala. No entanto, seus diferentes modelos permitem construir indicadores com as mais variadas finalidades, e este é o caso dos indicadores de condição sócio-econômica. Este estudo apresenta e compara algumas técnicas empregadas para a produção de indicadores da condição sócio-econômica, tendo como base modelos da teoria da resposta ao item. Como estudo de caso utiliza-se a produção de índices do padrão de vida das famílias dos alunos que participaram do SIMAVE/PROEB (2002) – Programa de Avaliação Educacional do Estado de Minas Gerais. Adicionalmente, o trabalho fornece sugestões para a continuidade da produção de indicadores sócio-econômicos dentro do âmbito do SIMAVE.

Palavras-chave: teoria da resposta ao item; modelo para respostas graduadas; indicadores da condição sócio-econômica.

Abstract

In Brasil, the Item Response Theory (IRT) has been used on the production of the pupil's proficience in large scale assessment. However, its different kinds of models may be used to build others indices as is the case of socio-economic status index. This study presents and compares some techniques based on IRT that can lead to these indices. As a case study, the socio-economic index of the pupil's family, who have participated of the SIMAVE-2002 – assessment program of the Minas Gerais State, is produced. In addition, it is given some suggestions and guidelines for future socio-economic status index that will be produced in SIMAVE.

Keywords: item response theory; multidimensional model; socio-economic status.

1. Introdução

Principalmente nas áreas de ciências sociais, ciências humanas e ciências da saúde, é comum o interesse em avaliar determinados *constructos* teóricos, que se caracterizam por serem variáveis latentes (não diretamente observadas) como por exemplo, a habilidade cognitiva de um aluno em língua portuguesa, a satisfação no trabalho, a ansiedade, a depressão, a condição sócio-econômica. Tais *constructos*, em geral, são abstrações teóricas definidas dentro de uma rede de relações produzidas a partir de vários outros *constructos* (cf. Spector, 1992). Por não serem diretamente medidos, a escala de valores segundo a qual esses *constructos* são avaliados é produzida a partir de instrumentos (testes, questionários, etc) que se constituem de itens (questões do teste ou perguntas específicas, devidamente estruturadas de acordo com um modelo que será empregado) que, no entender dos especialistas que os elaboram, se associam diretamente ao *constructo* (ou *constructos*) de interesse. Assim, as variáveis produzidas a partir das questões apresentadas no instrumento são variáveis indicadores do *constructo* que se deseja medir.

Existem diferentes técnicas para se obter uma medida do *constructo* e, *grosso modo*, podem ser divididas segundo as chamadas técnicas clássicas, onde o modelo para a construção da escala é diretamente baseado no resultado obtido no instrumento como um todo, e técnicas onde modelos específicos são construídos para cada item do instrumento e a construção da escala considera todos esses modelos individuais. Nesse último caso, podem ser classificadas as técnicas baseadas na teoria da resposta ao item (TRI) que, originalmente, foram empregadas na produção de escalas de proficiência em testes de avaliação educacional. Dentre as vantagens que a TRI apresenta sobre as técnicas clássicas estão: A TRI permite uma melhor análise de cada item que constitui o instrumento de avaliação (ou medida), considerando suas características estatísticas específicas na produção das escalas, como as que medem a capacidade de discriminar os indivíduos e as dificuldades dos itens; facilita, também, a interpretação da escala produzida e permite conhecer, diretamente, quais itens estão produzindo a informação gerada ao longo do *continuum* de valores, segundo o qual a escala é construída, tipicamente o conjunto dos números reais; além disso, a TRI permite a comparabilidade dos resultados produzidos para grupos de indivíduos diferentes, mesmo quando instrumentos (parcialmente) diferentes são aplicados; finalmente, A TRI permite um tratamento natural de casos com dados faltantes. Os critérios habituais exigem que todos os itens sejam preenchidos ou, então, que alguma técnica de preenchimento de dados ausentes seja empregada. No caso da TRI, na estimação da variável latente quando existem dados faltantes empregam-se apenas os itens respondidos.

No Brasil, o critério mais utilizado com a finalidade de medir a condição sócio-econômica é o critério de classificação econômica Brasil, proposto pela ANEP – Associação Nacional de Empresas de Pesquisa, baseado em um estudo realizado em 1996 numa amostra de 20000 domicílios urbanos em cidades com mais de 20000 habitantes. Esse critério tem a finalidade principal de referenciar as empresas de pesquisa e opinião para medir a capacidade de compra dos consumidores. A escala produzida se baseia numa pontuação bruta ponderada a partir de uma série de questões que avaliam a posse de determinados bens de conforto doméstico, como número de aparelhos de rádio, geladeiras, etc, além da escolaridade do chefe da família. Para se chegar aos pesos da pontuação, são realizados estudos como o mencionado e, desses estudos, modelos de regressão clássicos são construídos tendo como variável dependente a renda familiar. No presente estudo, devido às características específicas da população constituída pelas famílias dos alunos que participam do SIMAVE (rural, urbana de grandes e médios municípios, e, ainda, urbana de

pequenos municípios) optou-se pela produção de uma escala, e conseqüentemente uma classificação, própria.

O objetivo desse trabalho é, inicialmente, apresentar algumas técnicas para produção de indicadores da condição sócio-econômica do indivíduo (padrão de vida) baseadas em modelos da Teoria da Resposta ao Item. Esses estudos foram desenvolvidos a partir da experiência do autor com os dados do SIMAVE/PROEB (Sistema Mineiro de Avaliação Educacional/Programa de Avaliação da Educação Básica – Rede estadual e parte da rede municipal) com esse tipo de metodologia. Adicionalmente, este estudo tem, também, o objetivo de apoiar uma decisão sobre qual seria a melhor metodologia, com o menor número de variáveis possível, que produziria uma medida adequada às finalidades do SIMAVE/PROEB da condição sócio-econômica dos alunos que participam das avaliações educacionais que periodicamente vem sendo realizadas. A introdução do menor número possível de variáveis reserva espaço nos questionários para outras questões relevantes.

Embora natural, o que não quer dizer trivial, no contexto da TRI, não é, ainda, comum no Brasil a utilização desse tipo de ferramenta na produção de indicadores da condição sócio-econômica. Soares & Mambrini (2003) apresentam um trabalho onde comparam o efeito produzido sobre a nota dos alunos no vestibular da UFMG do escore produzido pelo critério Brasil e o escore produzido por um específico modelo da TRI. Em outros estudos, que empregam modelos multinível (*cf.* Bryk & Raudenbush, 1992) em pesquisas educacionais, escores sócio-econômicos foram construídos com a finalidade de observar a influência dessa variável sobre a proficiência do aluno. Especificamente, o emprego de modelos da TRI parece ser o caso do estudo de Fletcher (1998), onde o autor, provavelmente, utiliza o modelo de dois parâmetros. Além desse, existe o trabalho de Soares (2003).

Nas seções 2, 3, e 4 serão revistos conceitos básicos sobre as principais estatísticas e modelos utilizados na construção dos índices. Preferiu-se enfatizar o que não tem sido muito abordado na literatura brasileira evitando, por outro lado, repetir o que já foi suficientemente bem apresentado. Assim, procurou-se apresentar deduções e justificativas das estatísticas e dos processos construtivos dos modelos que, em geral, não são bem discutidos. Na seção 2 apresentam-se as principais estatísticas chamadas de “clássicas” para testes psicométricos que são importantes para análises preliminares, mesmo quando se empregam modelos da Teoria da Resposta ao Item. Na seção 3 introduz-se à teoria da resposta ao item, enfocando os modelos que são empregados neste estudo e, na seção, 4 aborda-se a questão da dimensionalidade do instrumento aplicado. Na seção 5 apresentam-se os processos construtivos de três índices, a partir de variações metodológicas, fazendo-se uma análise comparativa dos três diferentes índices produzidos.

2. Métodos Clássicos para Discriminação dos Itens e Análise da Dimensionalidade

2.1 Correlação ponto bisserial

Considere o caso de testes constituídos por itens binários ou *dicotômicos*, isto é, itens para os quais se admite duas respostas possíveis. A correlação bisserial e a correlação ponto bisserial são medidas estatísticas que medem a correlação do resultado de um item em particular do teste com o resultado do teste (isto é, o escore bruto total), sendo, portanto, uma medida da capacidade de discriminação do item em relação ao resultado do teste. A correlação ponto bisserial é a correlação de Pearson e uma expressão que considera, explicitamente, os parâmetros da distribuição pode ser obtida. Para tanto, admita que S represente o escore

bruto obtido no teste. Admita que Y represente o resultado da resposta atribuída a um item, uma variável dicotômica (no caso de testes educacionais, por exemplo, atribui-se o valor $Y=0$, se a resposta for errada, e $Y=1$, para uma resposta correta; e, no caso de um item que avalia a condição sócio-econômica $Y=1$ representa a posse de um bem, por exemplo). O índice de correlação de Pearson é definido por:

$$\rho_{SY} = \frac{E(SY) - E(S)E(Y)}{\sigma_Y \sigma_S}$$

Uma estimativa natural obtida sobre o resultado do teste é a seguinte:

$$\bar{\rho}_{pb} = \frac{\bar{S}_p - \bar{S}}{\bar{\sigma}_S} \sqrt{\frac{\bar{p}}{\bar{q}}}$$

onde \bar{S}_p é o escore médio no teste para os que acertaram o item e \bar{S} é o escore médio no teste para todos. \bar{p} é a proporção dos que acertaram o item no teste, $\bar{\sigma}_S$ é o desvio padrão dos escores obtidos nos testes pelos respondentes e, a estimativa $\bar{\rho}_{pb}$ para a correlação de Pearson é o que freqüentemente se denomina na literatura de correlação ponto bisserial.

2.2 Correlação bisserial

Seja Z uma variável aleatória (não observada), associada ao *constructo* latente do respondente, tal que $Z \sim N(0,1)$. Admita ainda que o escore bruto do respondente no teste se associa linearmente a essa variável da seguinte forma: $S = AZ + B + \varepsilon$, onde $E(\varepsilon) = 0$ e, $E(\varepsilon Z) = 0$. Note-se que $E(S) = A E(Z) + B$ e, então a correlação de Pearson para S e Z é dada por:

$$\rho_{SZ} = \frac{E(AZ^2 + BZ) - E(S)E(Z)}{\sigma_Z \sigma_S} = \frac{A}{\sigma_S} \sigma_Z = \frac{A}{\sigma_S}$$

Sejam dois conjuntos de possíveis respondentes, os que acertam o item e os que erram o item. Assim, $E(S|Y=0) = A E(Z|Y=0) + B$ e, $E(S|Y=1) = A E(Z|Y=1) + B$, tal que:

$$A = \frac{E(S|Y=0) - E(S|Y=1)}{E(Z|Y=0) - E(Z|Y=1)} \quad (1)$$

É fácil obter estimativas para os termos no numerador da equação (1), basta tomar a média dos escores em todo o teste dos que acertam e dos que erram o item. O mesmo não ocorre em relação ao denominador por se tratarem de variáveis latentes. Admite-se, então, que os respondentes que acertam o item são os que apresentam valores para Z superiores à z_p , onde

z_p é tal que $\int_{z_p}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = p$. Logo, definindo $h(z_p) \triangleq \frac{e^{-z_p^2/2}}{\sqrt{2\pi}}$, tem-se que:

$$E(Z|Y=1) = \frac{h(z_p)}{q} \quad \text{e,} \quad E(Z|Y=0) = \frac{-h(z_p)}{q}$$

Assim, de (1):

$$A = \frac{E(S | Y = 0) - E(S | Y = 1)}{\frac{h(z_p)}{q} - \frac{h(z_p)}{p}}$$

e uma estimativa para a correlação de Pearson é dada, então, por:

$$\bar{\rho}_{bis} \triangleq \bar{\rho}_{SZ} = \frac{\bar{S}_q - \bar{S}_p}{\frac{-h(z_{\bar{p}})}{\bar{q}} - \frac{h(z_{\bar{p}})}{\bar{p}}} \frac{1}{\bar{\sigma}_S}$$

onde \bar{S}_q é o escore bruto médio para os que erram o item, \bar{S}_p é o escore bruto médio para os que acertam o item, $h(z_{\bar{p}})$ é o valor da função de densidade normal padrão em $z_{\bar{p}}$, \bar{p} é a proporção dos que acertaram o item no teste e, $\bar{\sigma}_S$ é o desvio padrão dos escores brutos obtidos no teste. Finalmente, pode-se mostrar que:

$$\bar{\rho}_{bis} = \bar{\rho}_{pb} \frac{\sqrt{\bar{p} \bar{q}}}{h(z_{\bar{p}})}. \quad (2)$$

2.3 Correlação polisserial e ponto polisserial

Os conceitos de correlação ponto biserial e biserial podem ser estendidos para o caso de itens politômicos, os quais apresentam mais de duas categorias ordenadas de respostas $(T_0, T_1, \dots, T_m), T_{K+1} \geq T_K$. A correlação ponto polisserial (ρ_{ppol}) é definida, simplesmente, como sendo a correlação de Pearson entre o escore bruto do teste (S) e o escore do item, medido segundo uma escala ordenada de inteiros cujas diferenças entre dois valores sucessivos seja sempre a mesma (por exemplo, (0, 1, 2, ..., m)). A correlação polisserial é definida com base na relação (2) da seguinte forma:

$$\bar{\rho}_{pol} = \bar{\rho}_{ppol} \frac{\bar{\sigma}}{\sum_{k=0}^{m-1} h(z_{\bar{p}_{k+1}})(T_{k+1} - T_k)} \quad (3)$$

onde \bar{p}_{k+1} é a proporção dos que alcançaram o escore k e $\bar{\sigma}$ é o desvio padrão dos escores.

2.4 Correlação tetracórica

Esse índice mede a correlação entre os resultados dos itens de um teste. Suponha que $Z_1 \sim N(0,1)$ seja uma variável (e, associada a habilidade latente) relacionada à resposta correta ou não atribuída ao item 1 e $Z_2 \sim N(0,1)$ represente a variável (igualmente associada a habilidade latente do indivíduo) também associada à resposta correta ou não ao item 2. Seja p_1 a proporção dos que acertam o item 1 e p_2 dos que acertam o item 2. Admita que, os que

acertam o item 1 são os que apresentam valores para $Z_1 \geq z_{p_1}$, e, os que acertam o item 2 são os que apresentam habilidade $Z_2 \geq z_{p_2}$, onde:

$$p_1 = \int_{z_{p_1}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \quad \text{e,} \quad p_2 = \int_{z_{p_2}}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du$$

Se Z_1 e, Z_2 são normalmente conjuntamente distribuídas com coeficiente de correlação ρ então a distribuição conjunta das duas variáveis é dada por:

$$p_{Z_1, Z_2}(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(z_1^2 + z_2^2 - 2\rho z_1 z_2)}{2(1-\rho^2)}} \quad (4)$$

Então, a probabilidade de acerto de ambos os itens é dada por:

$$p_{12} = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{z_{p_1}}^{\infty} \int_{z_{p_2}}^{\infty} e^{-\frac{(u_1^2 + u_2^2 - 2\rho u_1 u_2)}{2(1-\rho^2)}} du_1 du_2 \quad (5)$$

Como a relação é uma função implícita de ρ , uma solução para a equação acima tem que ser encontrada a partir de aproximações numéricas. Um método para obter-se uma aproximação é o de Divgi (1979), que está disponível no *software TESTFACT* (cf. Wilson *et al.*, 1998).

3. Teoria da Resposta ao Item

3.1 Introdução

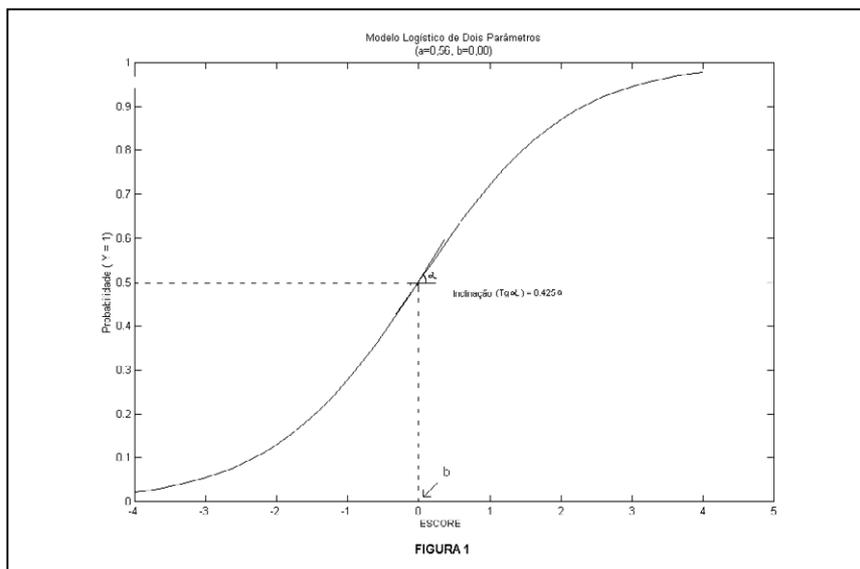
Conforme diversos autores (ver, por exemplo Baker, 1993), a teoria a resposta ao item surgiu com os trabalhos pioneiros de Lord (1952) e Rasch (1960). Esses autores foram os primeiros a propor modelos estatísticos paramétricos para itens de testes, que associavam a probabilidade de uma dada resposta (inicialmente, certa ou errada) a uma variável latente (não observada) interpretada como sendo a proficiência ou habilidade dos respondentes dentro de um contexto de testes para avaliação educacional ou avaliação psicométrica. Inicialmente, foi utilizada a distribuição normal acumulada na especificação do modelo. Birnbaum (1968) sugeriu utilizar a função logística pois, pelo fato de ser uma função explícita dos parâmetros dos itens e da proficiência, é matematicamente mais convenientemente. Desde então, essa teoria tem tido notável avanço teórico, sendo que novos modelos têm sido incorporados, o que tem trazido grande versatilidade a suas aplicações. É o caso dos modelos politômicos, que incorporam várias categorias das respostas (além das dicotômicas: certo ou errado), os modelos multidimensionais, que permitem produzir escalas para mais de uma variável latente associada, por exemplo, quando se admite à idéia de diferentes habilidades para o desenvolvimento cognitivo do aluno e, ainda, os modelos que incorporam comportamento diferenciado entre os itens em grupos diferentes – conhecidos como modelos para grupos múltiplos.

3.2 Modelo logístico de dois parâmetros (unidimensional)

Admita que Y seja uma variável aleatória dicotômica assumindo os valores 0, ou, 1. No caso específico de um teste educacional o valor 0 está associado a uma resposta errada e, o valor 1 a uma resposta certa por parte do aluno. O modelo de dois parâmetros expressa, então, a relação entre a variável latente θ e a resposta dada ao item da seguinte forma:

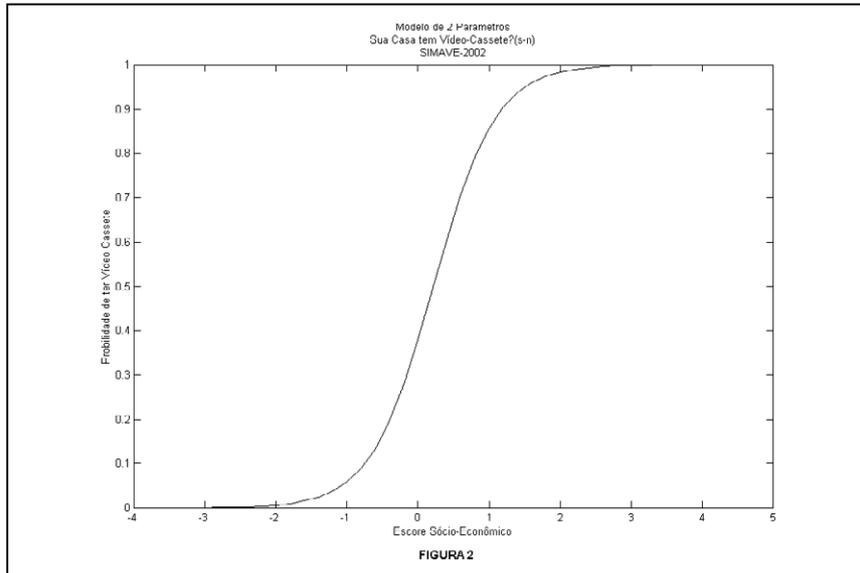
$$P_i(Y = 1; \theta, a_i, b_i) = \frac{e^{D a_i (\theta - b_i)}}{1 + e^{D a_i (\theta - b_i)}} \quad (6)$$

onde o índice i representa o item; a_i é denominado parâmetro de discriminação do item; b_i é denominado de parâmetro de dificuldade do item. D é um fator de escala, constante, e normalmente igual a 1,7, usado para aproximar a curva logística da distribuição normal. Esse fator de escala faz com que para um dado θ a probabilidade $P(Y = 1; \theta)$ seja aproximadamente a mesma nos dois tipos de modelos e, conseqüentemente, permite que os valores de θ dos indivíduos sejam estimados com valores muito próximos em ambos os casos. Como se pode notar, o parâmetro b_i representa o ponto na escala da variável latente θ , para o qual há 50% de chance de escolha da resposta representada por $Y = 1$ pelo indivíduo. É fácil observar que se (6) for derivada em relação à θ , a função resultante atinge seu máximo em $\theta = b_i$ com um valor igual a $0,425 a_i$. Portanto, quanto maior for o valor do parâmetro a_i , mais sensível torna-se o modelo a variações na habilidade em torno de seu ponto de dificuldade. Por isso, ele é conhecido como parâmetro de discriminação do item. Na Figura 1, apresenta-se a denominada curva característica de um item, isto é, a representação dos valores sob forma de gráfico de um modelo em função de θ , enfatizando as propriedades de seus parâmetros:



Podem ser consideradas como variáveis indicadoras da condição sócio-econômica, a posse de determinados bens como por exemplo eletrodomésticos, automóvel, etc. Assim, Y representará a posse ou não de um certo tipo de bem, e o modelo (6) será empregado para

associar o nível sócio-econômico (representado pela variável latente θ) com a probabilidade de o indivíduo possuir o bem. O modelo abaixo corresponde ao item Q19 (Se a família do aluno possui ou não aparelho de vídeo cassete) utilizado na construção do escore sócio-econômico do SIMAVE-2002.



O parâmetro de discriminação estimado para esse item foi $a = 1.292$, e o parâmetro de dificuldade foi $b = 0.218$. No presente caso o índice de dificuldade pode ser interpretado como o ponto da escala correspondente ao nível sócio-econômico para o qual a probabilidade de uma família possuir um aparelho de vídeo cassete é de 0.5. Note-se que esse item apresenta boa discriminação e, conseqüentemente, bom nível de informação, para os valores da condição sócio-econômica em torno de 0.218.

3.3 Modelo logístico para respostas graduadas

O modelo de respostas graduadas é uma generalização do modelo de dois parâmetros para o caso de mais de duas categorias ordenadas de respostas. Para sua formulação admita, inicialmente, que o indivíduo possa alcançar aos seguintes níveis, tendo em vista as suas respostas atribuídas ao item i , $\eta^i = \eta_0, \eta_1, \dots, \eta_{m_i}$. Admita, ainda, que tendo alcançado o nível η_ℓ , ele tenha também alcançado os níveis η_j para $j < \ell$. Assim, considere a classe de eventos $N = \{(\eta_0), (\eta_0, \eta_1), \dots, (\eta_0, \eta_1, \dots, \eta_{m_i})\}$ onde $N_\ell = (\eta_0, \eta_1, \dots, \eta_\ell)$ é o evento que representa o fato de o indivíduo ter alcançado o nível η_ℓ e, \bar{N}_ℓ é o correspondente evento complementar. A resposta dada ao instrumento será classificada segundo a escala ordinal $K_i = 0, 1, \dots, m_i$ onde $K_i = \ell$, representa que o indivíduo alcançou o nível η_ℓ e não alcançou os níveis η_j , $j \geq \ell$. Assim,

$$P_i(K = \ell; \theta) = P(N_\ell \cap \bar{N}_{\ell+1}; \theta) \quad (7)$$

Note-se inicialmente, que:

$$P(N_\ell | N_{\ell-1}; \theta) = P(N_{\ell-1} | N_\ell; \theta) \frac{P(N_\ell; \theta)}{P(N_{\ell-1}; \theta)} = \frac{P(N_\ell; \theta)}{P(N_{\ell-1}; \theta)}$$

tal que, por recorrência:

$$P(N_\ell; \theta) = \prod_{j=1}^{\ell} P(N_j | N_{j-1}; \theta) P(N_0; \theta) = \prod_{j=1}^{\ell} P(N_j | N_{j-1}; \theta) \quad (8)$$

Admitindo que $P(N_0; \theta) = 1$. O Termo $P(N_j | N_{j-1}; \theta)$ é conhecido como função de processamento, “processing function” (cf. Samejima, 1996). De (7):

$$\begin{aligned} P_i(K=\ell; \theta) &= P(N_\ell \cap \bar{N}_{\ell+1}; \theta) = P(\bar{N}_{\ell+1} | N_\ell; \theta) P(N_\ell; \theta) = \\ &= [1 - P(N_{\ell+1} | N_\ell; \theta)] P(N_\ell; \theta) = \prod_{j=1}^{\ell} P(N_j | N_{j-1}; \theta) [1 - P(N_{\ell+1} | N_\ell; \theta)] \quad (9) \end{aligned}$$

Para um fechamento da expressão (9), define-se que $P(N_{m_i+1} | N_{m_i}; \theta) = 0$ e, conseqüentemente, $P(N_{m_i+1}; \theta) = 0$. Por outro lado por (8) e (9):

$$\begin{aligned} P_i(K=\ell; \theta) &= P(N_\ell; \theta) - P(N_{\ell+1} | N_\ell; \theta) P(N_\ell; \theta) = \\ &= P(N_\ell; \theta) - P(N_{\ell+1} | N_\ell; \theta) \prod_{j=1}^{\ell} P(N_j | N_{j-1}; \theta) = P(N_\ell; \theta) - P(N_{\ell+1}; \theta) \quad (10) \end{aligned}$$

Samejima (1997), estabelece 4 condições para se definir adequadamente um modelo $P(N_\ell; \theta)$, entre elas a de que exista solução única para o estimador de máxima verossimilhança de θ (ver Samejima, 1997). De fato, essas condições são satisfeitas quando se admite que:

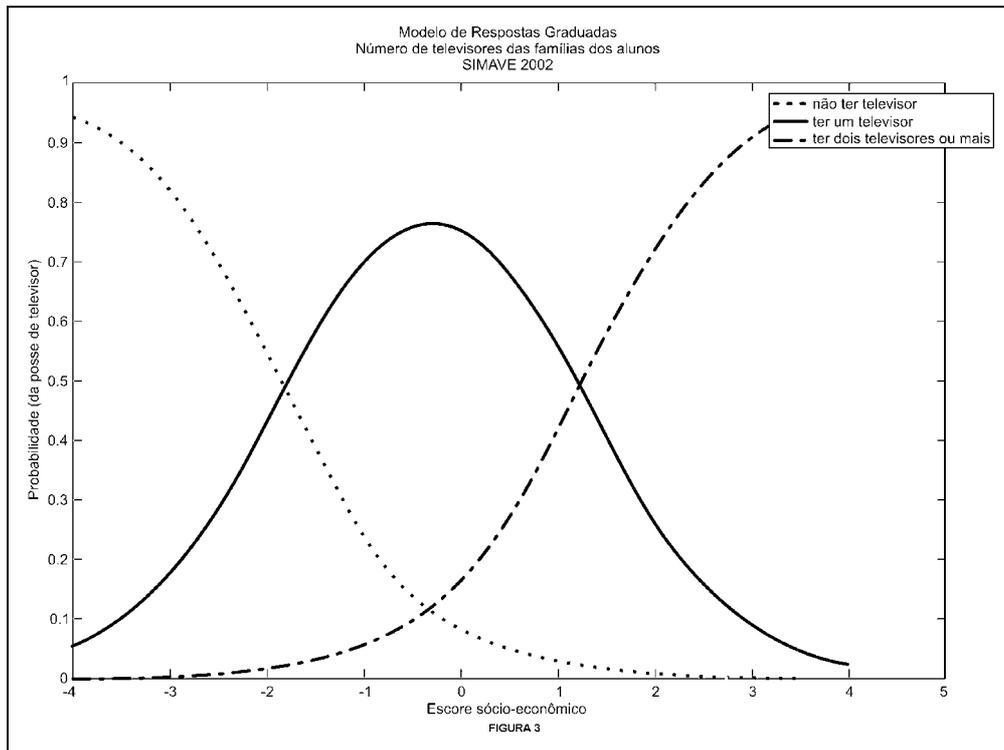
$$P(N_\ell; \theta) = \frac{e^{D a_i (\theta - b_i + c_\ell)}}{1 + e^{D a_i (\theta - b_i + c_\ell)}}$$

ou seja, a probabilidade de se alcançar pelo menos o nível η_ℓ é representado por um modelo de dois parâmetros, onde a dificuldade correspondente ao item é modulada pelo parâmetro C_ℓ , característico do nível ℓ . E, portanto:

$$P_i(K=\ell; \theta) = \frac{1}{1 + e^{-D a_i (\theta - b_i + c_\ell)}} - \frac{1}{1 + e^{-D a_i (\theta - b_i + c_{\ell+1})}}$$

onde $C_0 = \infty$ e, $b_{i,\ell} = b_i - C_\ell$ é a dificuldade de se alcançar pelo menos o nível ℓ no item i . O seguinte exemplo ilustra uma aplicação de modelos para respostas graduadas. Novamente, utilizando-se os dados do Simave-2002, o modelo abaixo foi o encontrado para a questão que

perguntava sobre o número de televisores que a família do aluno possuía (as possibilidades de respostas foram as seguintes: nenhum, 1 televisor, 2 ou mais). Para esse tipo de item, a probabilidade de escolha de cada resposta pode ser modelada através modelos de respostas graduadas. E o seguinte modelo foi obtido ($a_i = 0.767$, $b_i = -0.300$, $C_0 = \infty$, $C_1 = 1.535$, $C_2 = -1.535$), a probabilidade de cada resposta está representada no gráfico abaixo:



Note-se que a probabilidade de não ter televisor é mais elevada nas famílias de escores sócio-econômicos mais baixos, decaindo com o aumento da condição sócio-econômica. A probabilidade de a família ter apenas um televisor é mais elevada nas famílias com escores medianos e a probabilidade de se ter dois ou mais televisores é característico de famílias mais abastadas. É importante observar que:

$$a_i (\theta - b_i + c_\ell) = \frac{a_i}{s} [s\theta - s(b_i - t) + s(c_\ell - t)] \quad (s \neq 0)$$

o que, naturalmente, conduz a um problema de identificabilidade dos parâmetros se não forem admitidas restrições adicionais. Assim, admite-se que $\frac{C_1 + \dots + C_{m_i}}{m_i} = cte$ (0.0, tipicamente)

o que resolve o problema de identificabilidade nesses casos. Por último, deve-se notar que o modelo de respostas graduadas se reduz ao modelo de dois parâmetros quando $m_i = 1$.

3.4 Estimação dos escores de habilidades (θ) a partir do método de máxima verossimilhança (*maximum likelihood*)

Admita-se que um item i , tenha m_i categorias de respostas. Seja $P_i(K_i; \theta)$ o modelo que representa a probabilidade de escolha do K -ésima categoria ($K_i = 0, 1, \dots, m_i$) admita, ainda, que os parâmetros desse modelo sejam conhecidos e se deseja estimar o valor de θ para um dado padrão de (escolha ou resposta) $\underline{\Gamma}(I) = (K_1, K_2, \dots, K_I)$ de um indivíduo em I itens diferentes que constituem o teste apresentado a ele. Considerando que as respostas atribuídas aos diferentes itens são independentes quando condicionadas a habilidade θ , propriedade conhecida como independência local dos itens, a função de verossimilhança para esse padrão de escolhas é dada por:

$$\ell(\theta; \underline{\Gamma}) = \prod_{i=1}^I P_i(K_i; \theta) \quad (11)$$

e, por conveniência matemática dada as características dos modelos utilizados, o seu logaritmo Neperiano é representado por:

$$L(\theta; \underline{\Gamma}) = \sum_{i=1}^I \log P_i(K_i; \theta) \quad (12)$$

Uma estimativa para θ é obtida então maximizando (11) ou, equivalentemente, (12). Por isso, o termo verossimilhança máxima. Considerando a condição necessária para a obtenção da solução ótima, $\frac{\partial L(\theta; \underline{\Gamma})}{\partial \theta} = 0$, tem-se:

$$\sum_{i=1}^I \frac{1}{P_i(K_i; \theta)} P_i'(K_i; \theta) = 0 \quad (13)$$

A solução de (13) não pode ser obtida analiticamente para os modelos considerados, por isso deve ser resolvida através de procedimentos numéricos iterativos. O método mais usado na TRI é o conhecido método de Newton-Raphson (*cf.* Minoux, 1986). A equação de iteração desse procedimento é dada por:

$$[\hat{\theta}(\underline{\Gamma})]_{t+1} = [\hat{\theta}(\underline{\Gamma})]_t - \left[\frac{\partial^2 L(\theta; \underline{\Gamma})}{\partial \theta^2} \right]_t^{-1} \left[\frac{\partial L(\theta; \underline{\Gamma})}{\partial \theta} \right]_t \quad (14)$$

e, a solução para (13) é obtida a partir do momento que critério de convergência seja alcançado, como por exemplo: $|\hat{\theta}_{t-1}(\cdot) - \hat{\theta}_t(\cdot)| \leq \delta$ (δ pequeno). Obviamente, para se empregar

o método é necessário se obter as fórmulas para o Gradiente $\frac{\partial L(\cdot)}{\partial \theta}$ e Hessiano $\frac{\partial^2 L(\cdot)}{\partial \theta^2}$ em

cada tipo de modelo diferente. No caso dos modelos da TRI aqui empregados, a obtenção de fórmulas explícitas não representa maiores problemas e podem ser obtidas analiticamente (ver, por exemplo, Baker, 1993; Andrade *et al.*, 2001). Sabe-se que no caso dos modelos de respostas graduadas a solução existe e é única (ver, por exemplo, Samejima, 1997).

3.5 Função de informação do item e informação do teste

Muito embora sua origem tenha sido um pouco distinta, a definição da função de informação de um item proposta por Birnbaum (1968), deriva diretamente da função de informação de Fischer e, portanto, considere:

$$I(\theta) = -E\left(\frac{\partial L^2(\cdot)}{\partial \theta^2}\right) = E\left\{\sum_{i=1}^I \left[\frac{-P_i''(K_i;\theta)}{P_i(K_i;\theta)} + \frac{P_i'^2(K_i;\theta)}{P_i^2(K_i;\theta)}\right]\right\} \quad (15)$$

onde P_i' representa a derivada com respeito à habilidade θ . Assim, pelo fato de a distribuição dos valores das respostas ser discreta o valor da informação no teste é dado por:

$$I(\theta) = \sum_{i=1}^I \sum_{j=1}^{m_i} \left[\left(\frac{P_i'^2(K_j;\theta)}{P_i^2(K_j;\theta)} - \frac{P_i''(K_j;\theta)}{P_i(K_j;\theta)} \right) P_i(K_j;\theta) \right] = \sum_{i=1}^I \sum_{j=1}^{m_i} \left[\frac{P_i'^2(K_j;\theta)}{P_i(K_j;\theta)} - P_i''(K_j;\theta) \right] \quad (16)$$

e essa é a função de informação esperada de um determinado teste. Note-se que a independência das respostas atribuídas aos itens é essencial para obtenção da expressão na forma acima. Para um item específico, a informação é dada por:

$$I_i(\theta) = \sum_{j=1}^{m_i} \left[\frac{P_i'^2(K_j;\theta)}{P_i(K_j;\theta)} - P_i''(K_j;\theta) \right] \quad (17)$$

Particularmente, no caso de modelos dicotômicos ($m_i = 1$), tem-se que $P_i(0;\theta) = 1 - P_i(1;\theta)$ e então:

$$I_i(\theta) = \left[\frac{P_i'^2(1;\theta)}{P_i(1;\theta)} - P_i''(1;\theta) \right] + \left[\frac{P_i'^2(1;\theta)}{1 - P_i(1;\theta)} + P_i''(1;\theta) \right] = \frac{P_i'^2(1;\theta)}{P_i(1;\theta)(1 - P_i(1;\theta))} \quad (18)$$

que é a expressão proposta por Birnbaum para a informação de um item dicotômico. Fórmulas explícitas para os diversos modelos podem ser encontradas em Baker (1992) mas também podem ser obtidas facilmente para modelos de dois parâmetros e modelos de respostas graduadas.

3.6 Estimação dos parâmetros do modelo

Neste trabalho foi utilizado o método da Máxima Verossimilhança Marginal (ou marginalizada) para a estimação dos parâmetros dos modelos (cf. Bock & Lieberman, 1970). Admita que $P_i(K;\theta, \mathcal{P}_i)$ seja o modelo de probabilidade correspondente a escolha de K no i -ésimo item ($K=1, \dots, m_i$). A idéia original de Bock & Lieberman (1970), foi a de considerar, inicialmente uma distribuição $g(\theta, \eta)$ para as proficiências θ , onde η são os hiperparâmetros da distribuição (conhecidos ou arbitrados nesse tipo de modelo), e resolver o problema de estimar os parâmetros considerando a média segundo essa distribuição sobre o espaço de parâmetros incidentais θ . Isto é, esses autores propõem resolver, inicialmente, o problema de encontrar \mathcal{P} que maximize a verossimilhança marginal:

$$\ell_M(\mathcal{P}; \underline{\Gamma}(I)) = \int \prod_{\Theta} \prod_{j=1}^N \left[\prod_{i \in I(j)} P_i(K_i(j); \theta, \mathcal{P}_i) \right] g(\theta, \eta) d\theta \quad (19)$$

ou, equivalentemente, como as respostas dos examinandos são independentes:

$$L_M(\mathcal{P}; \underline{\Gamma}(I)) = \sum_{j=1}^N \log \int \left[\prod_{i \in I(j)} P_i(K_i(j); \theta, \mathcal{P}) \right] g(\theta, \eta) d\theta \quad (20)$$

Encontrados os parâmetros $\hat{\mathcal{P}}_{\mathcal{M}}$, os parâmetros $\hat{\theta}_j$, $j=1, \dots, N$, poderiam ser facilmente estimados numa etapa posterior. A otimização da expressão acima apresenta a dificuldade inerente à avaliação da integral e, por isso, é aproximado por um problema onde a integral é “discretizada” nos chamados pontos de quadratura. Isto é, considerando os pares $(\theta_q, A_q = g(\theta_q; \eta) \Delta \theta_q)$ com Q pontos finitos (“grande o suficiente” para se alcançar uma boa aproximação), o problema aproximado consiste em encontrar $\hat{\mathcal{P}}_{\mathcal{M}}$ que maximize:

$$L_{M,Q}(\mathcal{P}; \underline{\Gamma}(I)) = \sum_{j=1}^N \log \sum_{q=1}^Q \left[\prod_{i \in I(j)} P_i(K_i(j); \theta_q, \mathcal{P}_i) A_q \right] \quad (21)$$

A condição necessária para se obter a solução ótima é dada por:

$$\begin{aligned} & \frac{\partial}{\partial \mathcal{P}} \left\{ \sum_{j=1}^N \log \left[\sum_{q=1}^Q \prod_{i \in I(j)} P_i(K_i(j); \theta_q, \mathcal{P}_i) A_q \right] \right\} = \\ & = \sum_{j=1}^N \left[\sum_{q=1}^Q \prod_{i \in I(j)} P_i(K_i(j); \theta_q, \mathcal{P}_i) A_q \right]^{-1} \cdot \sum_{q=1}^Q A_q \frac{\partial}{\partial \mathcal{P}} \prod_{i \in I(j)} P_i(K_i(j); \theta_q, \mathcal{P}_i) = 0 \end{aligned} \quad (22)$$

Definindo,

$$\bar{r}_{q_i}(K; \mathcal{P}) := \sum_{j=1}^N u(K_i(j), K) \frac{\prod_{\ell \in I(j)} P_{\ell}(K_{\ell}(j); \theta_q, \mathcal{P}_{\ell}) A_q}{\left[\sum_{q=1}^Q \prod_{\ell \in I(j)} P_{\ell}(K_{\ell}(j); \theta_q, \mathcal{P}_{\ell}) A_q \right]}$$

onde $u(K_i(j), K) = 1$ se $K_i(j) = K$ e $u(K_i(j), K) = 0$ se $K_i(j) \neq K$, pode-se (cf. Backer, 1992) mostrar que (22) pode ser colocada na seguinte forma:

$$\sum_{q=1}^Q \sum_{K=1}^{m_i} \bar{r}_{q_i}(K; \mathcal{P}) P_i^{-1}(K; \theta_q, \mathcal{P}) \frac{\partial}{\partial \mathcal{P}_i} P_i(K; \theta_q, \mathcal{P}_i) = 0 \quad i=1, \dots, I \quad (23)$$

Se as proficiências fossem conhecidas, $\bar{r}_{q_i}(K; \mathcal{P})$ representaria a freqüência observada de proficiências no intervalo q. Aqui, como as proficiências são desconhecidas, esse termo pode ser interpretado como a freqüência esperada de proficiência no intervalo q, condicionada aos parâmetros do modelo \mathcal{P} . Bock & Aitkin (1981) propõe um procedimento iterativo onde:

- i) Admite-se um valor $\hat{\mathcal{P}}^t$ provisório para os parâmetros.
- ii) Calcula-se $\bar{r}_{q_i}(K; \hat{\mathcal{P}}^t)$, $K=1, \dots, m_i$, $i=1, \dots, I$.

iii) Resolve-se (28), calculando $\hat{\mathcal{P}}^{t+1}$ para $\bar{r}_{qi}(K; \hat{\mathcal{P}}^t)$ do passo ii).

iv) Se $\hat{\mathcal{P}}^{t+1} \cong \hat{\mathcal{P}}^t$, o algoritmo alcançou convergência e a solução deve ter sido alcançada. Caso contrário, deve-se retornar ao passo ii).

A justificativa para o emprego do algoritmo acima e, conseqüentemente, justificativas para uma provável convergência do algoritmo advém de sua interpretação dentro das condições gerais para o conhecido algoritmo EM, conforme estudado em Dempster, Laird & Rubin (1977). Note-se que a estimativa obtida depende crucialmente da função $g(\theta, \eta)$ que se admite para as proficiências. No *BILOG-MG* há, *grosso modo*, três possibilidades: fixa (normal), fixa (com valores de quadratura fornecidos pelo usuário), e, empírica, com valores de quadratura re-estimados a cada passo do algoritmo acima.

4. Análise da Dimensionalidade

4.1 Modelo de análise fatorial para variáveis dicotômicas

Tanto o método de análise da dimensionalidade associada a um conjunto de variáveis dicotômicas que consiste na inspeção dos autovalores da chamada matriz de correlação *tetracórica* quanto o método da *análise fatorial plena* (cf. Bock & Aitkin, 1981; Bock, Gibbons & Muraki, 1988; Muraki & Engelhard, 1985) foram propostos a partir de uma adaptação do modelo tradicional de análise fatorial que considera a estrutura de dimensões associadas a variáveis contínuas. Novamente, como no caso de algumas estatísticas clássicas a definição de uma variável artificial é a chave para a construção do método. Assim, definindo uma variável X_i , tal que $\sigma_X = 1$, e $E(X_i) = 0$, e o relacionamento dessa variável com a variável dicotômica Y_i que representa a resposta atribuída ao item i (assumindo os valores 0 ou 1) é tal que: se $X_i \geq \gamma_i$, então $Y_i = 1$ e, se $X_i < \gamma_i$, então $Y_i = 0$.

O modelo de análise fatorial é então definido a partir da variável X_i da seguinte forma:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1d} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nd} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \Lambda \theta + e$$

Os valores λ_{ij} são conhecidos como as cargas associadas ao fator θ_j e à variável X_i , sendo uma medida do grau de associação entre o fator e a variável. Representa-se o vetor de dimensões latentes associadas por θ e, por hipótese, admite-se que $E(\theta e) = 0$ e, $E(\theta_i \theta_j) = 0$ para $i \neq j$, e, ainda, que $e \sim N(0, \Psi)$, com Ψ diagonal. Dessa forma, sob essas hipóteses é fácil mostrar que a correlação de X é dada por:

$$\Sigma = \Lambda \Theta \Lambda' + \Psi,$$

onde Θ é a matriz de covariância de θ . Em particular, se o modelo é unidimensional, então as linhas de $\Lambda \Theta \Lambda'$ serão todas linearmente dependentes entre si e assim, portanto, os seus autovalores serão todos zeros exceto um deles. Na prática, a unidimensionalidade deve ser entendida como a predominância de uma única dimensão sobre as demais e se aceita como a dimensão associada às variáveis um certo número de autovalores cujos valores sejam razoavelmente maiores que os dos demais. Embora, esse critério seja subjetivo.

4.2 Método de análise fatorial de informação plena

Com o intuito de evitar a subjetividade inerente ao uso da inspeção dos autovalores da matriz de correlação tetracórica como método para detecção da dimensionalidade. Bock & Aitkin (1981), Bock, Gibbons & Muraki (1988), propuseram o método da análise fatorial de informação plena. Considere, novamente, o modelo de análise fatorial apresentado na seção anterior. Dessa forma então, $P(Y_i = 1) = P(X_i \geq \gamma_i) = P(\sum_{j=1}^d \lambda_{ij} \theta_j + e_i \geq \gamma_i) = P(e_i \geq \gamma_i - \sum_{j=1}^d \lambda_{ij} \theta_j)$. Lembrando que por hipótese $e \sim N(0, \Psi)$, com Ψ diagonal, tem-se então que:

$$P(Y_i = 1) = \int_{\frac{\gamma_i - \sum_{j=1}^d \lambda_{ij} \theta_j}{\sigma_{e_i}}}^{\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz \quad (24)$$

onde $Z_d := \gamma_i - \sum_{j=1}^d \lambda_{ij} \theta_j$ e, $\sigma_{e_i}^2$ é a variância de e_i . Da estrutura do modelo é fácil

verificar que $\sigma_{e_i} = \sqrt{1 - \sum_{j=1}^d \lambda_{ij}^2}$, e reparametrizando (29) da seguinte forma:

$$b_i = -\frac{\gamma_i}{\sigma_{e_i}}, \quad a_{ij} = \frac{\lambda_{ij}}{\sigma_{e_i}}, \quad (25)$$

tem-se um modelo multidimensional que utiliza a curva de ogivas normal (função de distribuição de uma normal padronizada):

$$P(Y_i = 1; \underline{\theta}) = \int_{b_i + \sum_{j=1}^d a_{ij} \theta_j}^{\infty} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz \quad (26)$$

onde b_i é interpretado como a dificuldade geral do item, e os valores a_{ij} como os parâmetros de discriminação específicos a cada dimensão. O método para estimação dos parâmetros desse modelo pode ser, *mutato mutandis*, o mesmo método de máxima verossimilhança marginal (cf. Bock & Aitkin, 1981) empregado nos modelos mais comuns e anteriormente apresentado. Nota-se que as equações (25) fornecem uma forma direta para se obter as cargas do modelo de análise fatorial, basta que se inverta as relações. Para a estimação dos parâmetros do modelo (26) empregando-se o método mencionado pode-se utilizar o *software testfact* (cf. Wilson, Wood & Gibbons, 1998). Para uma decisão quanto ao número de dimensões os autores sugerem um teste para as razões entre os valores da verossimilhança de dois modelos aninhados. Isto é, compara-se estatisticamente a diferença entre os logaritmos do valor da função de verossimilhança para de dois modelos, tendo um deles uma dimensão a mais que o anterior. Sob condições apropriadas, e sob a hipótese nula de que a dimensão correta é a do modelo com menor dimensão, essa diferença apresenta, dependendo do tamanho da amostra, uma distribuição aproximadamente *qui-quadrado* com graus de

liberdade igual à diferença entre o número de parâmetros dos dois modelos. Infelizmente, o resultado desse teste é altamente dependente do tamanho da amostra empregada e precisa ser usado com bastante cuidado. Habitualmente, recomenda-se que várias amostras, de tamanhos diferentes, sejam testadas antes de uma decisão final. Adicionalmente, a esse tipo de análise recomenda-se testar também a razão entre os autovalores da matriz de correlações estimada.

5. Índices para o Padrão de Vida

5.1 Introdução

Aos alunos da 4ª e 8ª série do ensino fundamental e da 3ª série do ensino médio das escolas estaduais (e, também, algumas municipais) do estado de Minas Gerais, que participaram do SIMAVE-2002, foi apresentado um questionário com aproximadamente 60 questões com finalidades variadas, principalmente a de buscar explicações para os resultados alcançados no teste pelos alunos. Parte dessas questões é destinada à avaliação da condição sócio-econômica familiar do aluno. Uma medida dessa condição é de grande importância nos trabalhos que buscam construir modelos que explicam os resultados alcançados no teste – como os estudos que utilizam modelos multinível, permitindo controlar pela condição sócio-econômica, por exemplo, a influência que outras variáveis exercem sobre o rendimento dos alunos. Especificamente, no caso do SIMAVE-2002 as seguintes questões, 20 ao todo, foram apresentadas aos alunos com a finalidade de se construir um índice desse tipo:

Tabela 1 – Questões indicadoras da condição sócio-econômica.

QUESTÃO	DESCRIÇÃO
Q06	• Você trabalha? (A-SIM, B-NÃO)
Q07	• Onde você mora existe água encanada? (A-SIM, B-NÃO)
Q08	• Onde você mora existe eletricidade? (A-SIM, B-NÃO)
Q09	• A rua em que você mora tem calçamento ou asfalto? (A-SIM, B-NÃO)
Q10	• A rua em que você mora tem coleta de lixo? (A-SIM, B-NÃO)
Q11	• A casa em que você mora tem cozinha? (A-SIM, B-NÃO)
Q12	• A casa em que você mora tem banheiro? (A-Um B-Dois C-Três ou mais D-NÃO)
Q13	• A casa em que você mora tem sala? ? (A-Um B-Dois C-Três ou mais D-NÃO)
Q14	• A casa em que você mora tem quarto? ? (A-Um B-Dois C-Três D-Quatro ou mais E-NÃO)
Q15	• A casa em que você mora tem televisão a cores? ? (A-Um B-Dois ou mais C-NÃO)
Q16	• A casa em que você mora tem geladeira? ? (A-Um B-Dois ou mais C-NÃO)
Q17	• A casa em que você mora tem máquina de lavar roupa? (A-Sim B-Não)
Q18	• A casa em que você mora tem aparelho de som? ? (A-Um B-Dois ou mais C-NÃO)
Q19	• A casa em que você mora tem aparelho de videocassete? ? (A-Um B-Dois ou mais C-NÃO)
Q20	• A casa em que você mora tem freezer? (A-SIM, B-NÃO)
Q21	• A casa em que você mora tem telefone? (A-SIM, B-NÃO)
Q22	• A casa em que você mora tem telefone celular? (A-SIM, B-NÃO)
Q23	• A casa em que você mora tem computador? (A-SIM, B-NÃO)
Q24	• A casa em que você mora tem automóvel? ? (A-Um B-Dois ou mais C-NÃO)
Q32-Q35	• Instrução do pai ou responsável? (1- Não sabe ler 2- Sabe ler mas não completou a 4ª série 3- Completou a 4ª série 4- Completou a 8ª série 5- Completou o ensino médio 6- Completou um curso superior)

Note-se que as questões de q32 à q35 referem-se à escolaridade do pai ou responsável. De fato, a variável resultante dessas questões pode ser uma boa *proxy* para a condição sócio-econômica da família e portanto será utilizada, também, na produção dos índices. No presente estudo, como pode ser observado pela forma como as questões foram construídas, há variáveis com escala de medida categórica que essencialmente admitem respostas dicotômicas (tem/não-tem asfalto, trabalha/não-trabalha, etc), outras com escala intervalar porém com valores inteiros (número de geladeiras, televisores, e etc), o que se caracteriza por variáveis típicas de contagem, e outras com escala ordinal de medida (as que medem a escolaridade do pai). O modelo de dois parâmetros, apresentado na seção 3.2, permite que se construa um índice quando as variáveis são todas dicotômicas. Já o modelo de respostas graduadas permite que se empregue uma escala ordinal de medidas, não necessariamente com o mesmo número de categorias, podendo, ainda, considerar itens dicotômicos. A idéia central do trabalho é a de construir e comparar indicadores da condição sócio-econômica usando esses dois tipos de modelos.

5.2 Índices obtidos a partir da dicotomização direta dos itens

Nesta seção, então, o primeiro tipo de índice analisado será produzido a partir da utilização do modelo de dois parâmetros, para itens dicotômicos e uma única dimensão latente. Portanto, o primeiro passo, nesse caso, será o de tornar dicotômicas todas as variáveis originais. Assim, as variáveis que já não são binárias terão suas respostas transformadas de tal forma que as novas variáveis produzidas admitem apenas respostas dicotômicas – é o caso, por exemplo, da variável q16 (que representa o Número de Geladeiras da família) que é transformada para uma nova variável q16d (que é indicadora de se a família possui ou não esse bem). Assim procedendo, chega-se ao um conjunto de 20 variáveis (representadas por Q06d a Q32d, conforme a questão da qual ela tenha se originado) que foram utilizadas na produção do **índice_1**. O primeiro passo na análise é o da avaliação da correlação bisserial, para verificar se uma determinada variável apresenta correlação significativa com o escore bruto produzido pelo conjunto das variáveis. Esse passo é crucial para a escolha de itens que de fato apresentam consistência interna e se associam bem ao escore que será produzido. Normalmente, aceita-se valores para a correlação bisserial superior a 0.3. Os resultados encontrados para a correlação bisserial, indicam que todas as variáveis indicadas na Tabela 2 apresentam correlação bisserial superior a 0,3 exceto a variável q06d, que apresentou um valor de 0,04. Na Tabela 2 é apresentado o percentual de posses por parte das famílias dos alunos e o valor estimado para a correlação bisserial.

Esses resultados permitem concluir que a variável q06d (se o aluno trabalha ou não) não apresenta uma boa correlação com o escore bruto produzido pelo conjunto das variáveis, portanto, não apresenta boa discriminação para o índice que se deseja produzir. Essa variável é, então, excluída de qualquer análise posterior. Se forem observados os valores das proporções de posses, verifica-se que a maioria das variáveis apresenta uma proporção elevada, o que pode traduzir-se num índice com baixa informação devida a ausência de itens com níveis de dificuldade de acesso ao bem mais elevados. As variáveis Q11d a Q14d, particularmente, apresentam uma proporção de posses muito elevada, de tal forma, que provavelmente não trarão informação para o índice que será produzido. Assim, poderiam, a critério do analista serem eliminadas. No entanto, como um objetivo do trabalho é o de comparar os diferentes índices, optou-se por mantê-las e excluí-las somente em função de características estatísticas que possam ser indesejáveis na construção do índice.

Tabela 2 – Valores para a correlação bisserial.

QUESTÃO	Proporção de Posses	Correlação Bisserial	Correlação Ponto Bisserial
Q06d	0,29	0,04	0,03
Q07d	0,92	0,64	0,35
Q08d	0,96	0,73	0,33
Q09d	0,72	0,71	0,54
Q10d	0,85	0,72	0,47
Q11d	0,99	0,30	0,08
Q12d	0,98	0,99	0,36
Q13d	0,97	0,60	0,23
Q14d	1,00	0,75	0,14
Q15d	0,90	0,86	0,50
Q16d	0,91	0,94	0,54
Q17d	0,81	0,73	0,51
Q18d	0,85	0,71	0,47
Q19d	0,40	0,74	0,58
Q20d	0,21	0,61	0,44
Q21d	0,61	0,79	0,62
Q22d	0,37	0,68	0,53
Q23d	0,16	0,70	0,46
Q24d	0,44	0,59	0,47
Q32d	0,89	0,54	0,33

Como já abordado na seção 4, uma etapa importante na construção de um índice desse tipo é a investigação das dimensões avaliadas pelas questões que serão utilizadas na construção. Como no presente caso as variáveis são dicotômicas, pode-se realizar a análise das dimensões associadas a essas variáveis a partir das ferramentas disponíveis no *software testfact*[®]. O modelo utilizado pressupõe que exista uma única dimensão latente. Nesse caso, é importante testar se de fato existe apenas uma dimensão que seja preponderante, ou, no jargão da psicometria, um único fator dominante. Se existe um único fator que seja dominante, pode-se produzir o índice diretamente a partir do conjunto de variáveis considerando o modelo (6) e utilizando-se do próprio *software testfact*, ou ainda do *software Bilogmg*, que apresenta maior versatilidade, tendo em vista que oferece opções de métodos diferentes para a estimação dos parâmetros dos modelos e para a estimação dos escores produzidos para o índice. No caso de existir mais de uma dimensão latente, dois caminhos podem ser seguidos, sendo que o primeiro consiste em remover variáveis que estejam mais fortemente associadas às outras dimensões além daquela que representa, no caso, o índice que se deseja produzir. Esse procedimento, provavelmente, tem o inconveniente de perder informações associadas a essas variáveis, porém tem a virtude de ao se reduzir o problema ao caso unidimensional apresentar procedimentos mais estáveis para a estimação dos modelos e dos escores.

Para se decidir sobre o número de dimensões pode-se valer das estatísticas apresentadas na seção 4. No presente estudo, além da inspeção dos autovalores da matriz de correlações tetracóricas, foi utilizado o método *stepwise*, disponível no *testfact*, que compara a significância da diferença obtida para as estatísticas G^2 calculadas entre dois modelos aninhados, tendo um deles uma dimensão a mais que o anterior. Como comentado na seção 4, essa diferença apresenta, aproximadamente sob condições apropriadas, particularmente dependente do tamanho da amostra, uma distribuição *qui-quadrado* com graus de liberdade igual à diferença entre o número de parâmetros dos dois modelos. Deve-se tomar o cuidado, já que se vai basear sua decisão no resultado de uma teste de significância, para se utilizar um tamanho de amostra que seja no mínimo suficiente para estimar satisfatoriamente os parâmetros do modelo mas, também, não tão grande que se torne demasiadamente sensível às pequenas diferenças observadas e conduza a interpretações equivocadas com base na significância estimada. Estudos que avaliam o tamanho da amostra mais conveniente são raros e pouco conclusivos. Em geral, esse tamanho da amostra dependerá do número de parâmetros a ser estimado no modelo, do número de respondentes para cada item e, das características específicas de cada modelo e dos métodos de estimação empregados. Recomenda-se testar vários tamanhos de amostras e usar o bom senso e a experiência para uma decisão final quanto ao número de dimensões. No presente estudo, foi utilizada uma amostra de 1000 casos para conduzir essa específica análise, o que se mostrou bastante conveniente. Além disso, como sugerido por um dos revisores deste artigo, testou-se a estabilidade da razão entre os autovalores e o maior autovalor em 5 amostras diferentes. Os sete primeiros autovalores da matriz de correlações tetracóricas, e a respectiva razão entre o autovalor e o maior autovalor, são apresentados abaixo para 5 amostras diferentes:

Tabela 3 – Autovalores da matriz de correlação (7 maiores valores).

DIMENSÃO	1	2	3	4	5	6	7
AUTOVALOR	8,34	2,45	1,60	1,08	0,84	0,80	0,71
RAZÃO: $\lambda_i / \lambda_{\max}$	0,29	0,19	0,13	0,10	0,09	0,08	0,29
AUTOVALOR	7,97	2,40	1,84	0,90	0,82	0,72	0,70
RAZÃO: $\lambda_i / \lambda_{\max}$	0,30	0,23	0,11	0,10	0,09	0,09	0,30
AUTOVALOR	7,97	2,35	1,39	1,22	1,03	0,99	0,85
RAZÃO: $\lambda_i / \lambda_{\max}$	0,29	0,17	0,15	0,13	0,12	0,10	0,29
AUTOVALOR	7,55	2,30	1,82	1,08	1,04	0,86	0,80
RAZÃO: $\lambda_i / \lambda_{\max}$	0,30	0,24	0,14	0,14	0,11	0,10	0,30
AUTOVALOR	8,24	2,55	1,59	1,05	1,00	0,77	0,73
RAZÃO: $\lambda_i / \lambda_{\max}$	0,30	0,19	0,13	0,12	0,09	0,08	0,30

Todas as análises subseqüentes de dimensionalidade foram baseadas na amostra correspondente ao primeiro conjunto de autovalores. O resultado das significâncias para as diferenças observadas das estatísticas G^2 entre os modelos estimados pelo método de máxima verossimilhança marginal, com dimensões sucessivas a partir de uma dimensão, foi o seguinte:

Tabela 4 – Análise da dimensionalidade.

Nº de Fatores	Variância Associada					Diferença de G ²	df.	Sig.
	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5			
1	41,09	-	-	-	-	-	-	-
2	36,68	10,83	-	-	-	144,04	18	0,000
3	38,23	10,67	5,20	-	-	20,47	17	0,251
4	38,02	11,31	6,41	2,93	-	16,36	16	0,428
5	41,07	10,76	4,22	3,98	2,62	15,17	15	0,439

Sendo que o percentual de variância explicada para os 5 fatores é, respectivamente, 42,0%, 11,0%, 4,2%, 4,0% e 2,6%. A “matriz de cargas” obtida para uma extração, sem e com rotação pelo método Varimax, dos dois primeiros fatores foi a seguinte:

Tabela 5 – Cargas das dimensões antes e após serem rotacionadas. Parâmetros dos modelos produzidos (Índice_1b).

QUESTÃO	Dimensão 1	Dimensão 2	Dimensão 1 após rotação	Dimensão 2 após rotação	Parâmetro de Discriminação (a)	Parâmetro de Dificuldade (b)
Q07d	0,407	0,313	0,123	0,687	*	*
Q08d	0,603	0,468	0,172	0,716	*	*
Q09d	0,542	0,212	0,337	0,614	*	*
Q10d	0,526	0,294	0,259	0,720	*	*
Q11d	0,413	0,560	-0,292	0,445	*	*
Q12d	0,813	0,311	0,626	0,576	1,232 (0,030)	-2,803 (0,037)
Q13d	0,456	0,066	0,389	0,180	0,476 (0,020)	-4,657 (0,165)
Q14d	0,596	0,260	0,138	0,666	*	*
Q15d	0,729	0,058	0,574	0,445	1,114 (0,018)	-1,803 (0,018)
Q16d	0,830	-0,009	0,641	0,445	1,857 (0,036)	-1,561 (0,013)
Q17d	0,689	-0,012	0,532	0,314	0,895 (0,013)	-1,282 (0,015)
Q18d	0,672	-0,088	0,574	0,220	0,849 (0,013)	-1,676 (0,019)
Q19d	0,686	-0,373	0,713	0,222	1,292 (0,018)	0,218 (0,007)
Q20d	0,562	-0,333	0,683	-0,055	0,762 (0,012)	1,237 (0,016)
Q21d	0,764	-0,224	0,688	0,336	1,218 (0,017)	-0,368 (0,008)
Q22d	0,587	-0,259	0,637	0,142	0,887 (0,012)	0,580 (0,010)
Q23d	0,661	-0,477	0,817	0,101	1,251 (0,021)	1,324 (0,013)
Q24d	0,458	-0,548	0,676	-0,155	0,744 (0,010)	0,329 (0,010)
Q32d	0,473	0,031	0,376	0,163	0,550 (0,012)	-2,567 (0,046)

Após análise desses resultados, decidiu-se que existem dois fatores dominantes. Note-se que a conclusão, nesse caso, depende de um certo grau de subjetividade. Interpretou-se ainda que o primeiro fator (ou dimensão) representa o escore sócio-econômico desejado. O segundo fator (fortemente associado às variáveis q07d, q08d, q09d, q10d, q11d, q12d e q14d) pode ser interpretado como uma medida do grau de disponibilidade de serviços públicos, mas pode também ser uma caracterização de uma população rural, ou ainda, fruto apenas de uma dependência local devida a algumas dessas variáveis, pois a ele se associam variáveis com altos níveis de percentuais de acesso por parte das famílias dos alunos.

Assim, um índice sócio-econômico foi diretamente construído a partir da obtenção dos escores estimados pelo *software testfact* considerando-se um modelo com duas dimensões (**Índice_1a**). Outro procedimento possível é o de eliminar-se variáveis com maiores *cargas* associadas à segunda dimensão, e menores cargas associadas à primeira dimensão, até que apenas uma dimensão seja dominante. Começou-se por excluir a variável q11d, seguindo-se das variáveis q14d, q07d, q10d, q09d e, finalmente, q08d para se alcançar a unidimensionalidade conforme o mesmo critério adotado anteriormente. Isto é, encontrou-se um valor final para a diferença entre as estatísticas G^2 , dos modelos com 2 e uma dimensão, igual 15,21 (não significativa, $p=0,173$) com proporções de variância explicada em cada dimensão iguais a 48% e 5%. O modelo com uma dimensão retém, nesse caso, cerca de 55% da variância explicada e o escore foi diretamente produzido também pelo *testfact* apenas com as 13 variáveis restantes (**Índice_1b**). Os parâmetros para os modelos encontrados foram apresentados na Tabela 5.

5.3 Índices obtidos a partir da dicotomização de todas as opções

Na construção do **Índice_2** procura-se explorar mais intensamente a informação contida nas múltiplas alternativas de respostas de cada item, produzindo não apenas uma variável *dicotomizada* a partir de cada questão, mas tantas variáveis binárias quanto for possível dada à quantidade de alternativas. Por exemplo, a questão q15 (em sua casa tem TV a cores?) apresenta três alternativas de respostas ((A) SIM, 1 TV (B) SIM, 2 TV ou mais e (C) Não), nesse caso 2 variáveis dicotômicas foram criadas (q15d, a casa tem ou não TV) como no caso dos índices obtidos como na seção anterior e, (q15d_2, se tem 2 TV ou mais). O objetivo é o de verificar se uma maior informação é obtida com o emprego da variável q15d_2 do que com a variável q15d na construção do índice. Ao todo, foram produzidas 34 variáveis a partir das mesmas questões anteriores. Naturalmente, não se deve manter duas dessas variáveis binárias assim construídas e originadas a partir do mesmo item na análise que conduz à construção dos modelos, tendo em vista que elas estão completamente correlacionadas e, portanto, produzirão “falsas” dimensões associadas. É o que se pode entender como uma dependência local entre as questões, que não é devida à existência de um fator (ou dimensão) específico. A questão então é a de decidir sobre qual dessas variáveis deverá ser mantida para a produção do índice. Um critério pode ser o de se manter aquela de maior carga no fator principal obtido a partir da análise fatorial, esse foi o critério empregado neste trabalho. Finalmente, procedendo a análise da dimensionalidade associada de forma análoga ao procedimento adotado na construção dos índices descritos na seção 5.2, chegou-se ao seguinte conjunto de 14 variáveis que foram utilizadas na produção do **Índice_2**.

Tabela 6 – Variáveis utilizadas na construção do Índice_2.

QUESTÃO	Percentual de Posses	Descrição	Parâmetro de Discriminação (a)	Parâmetro de Dificuldade (b)
Q12d_2	0,235	• A casa em que mora tem dois banheiros ou mais	0,859 (0,012)	1,066 (0,013)
Q13d_3	0,033	• A casa em que mora tem três salas ou mais	0,585 (0,019)	4,067 (0,108)
Q14d_2	0,915	• A casa em que mora tem dois quartos ou mais	0,349 (0,012)	-4,189 (0,135)
Q15d_2	0,333	• A casa em que mora tem duas televisões a cores ou mais	1,175 (0,015)	0,599 (0,009)
Q16d	0,913	• A casa em que mora tem (ou não) geladeira	1,593 (0,028)	-1,623 (0,014)
Q17d	0,796	• A casa em que você mora tem (ou não) máquina de lavar roupa	0,849 (0,012)	-1,324 (0,015)
Q18d	0,873	• A casa em que você mora tem (ou não) aparelho de som	0,829 (0,013)	-1,702 (0,020)
Q19d	0,452	• A casa em que você mora tem (ou não) aparelho de videocassete	1,343 (0,017)	0,216 (0,007)
Q20d	0,237	• A casa em que você mora tem (ou não) freezer	0,767 (0,011)	1,234 (0,015)
Q21d	0,620	• A casa em que você mora tem (ou não) telefone	1,235 (0,016)	-0,367 (0,008)
Q22d	0,356	• A casa em que você mora tem (ou não) telefone celular	0,860 (0,011)	0,589 (0,010)
Q23d	0,161	• A casa em que você mora tem (ou não) computador	1,252 (0,020)	1,327 (0,013)
Q24d	0,458	• A casa em que você mora tem (ou não) automóvel	0,764 (0,010)	0,324 (0,010)
Q33d	0,396	• O pai completou o 1º grau	0,682 (0,010)	0,529 (0,012)

Em relação às variáveis empregadas na construção do índice_1, as que aqui são usadas na construção do índice apresentam maior variação no percentual de posses alcançado pelas famílias. Os parâmetros obtidos para os modelos dos itens estão apresentados na Tabela 6.

5.4 Índice obtido a partir de modelos para respostas graduadas

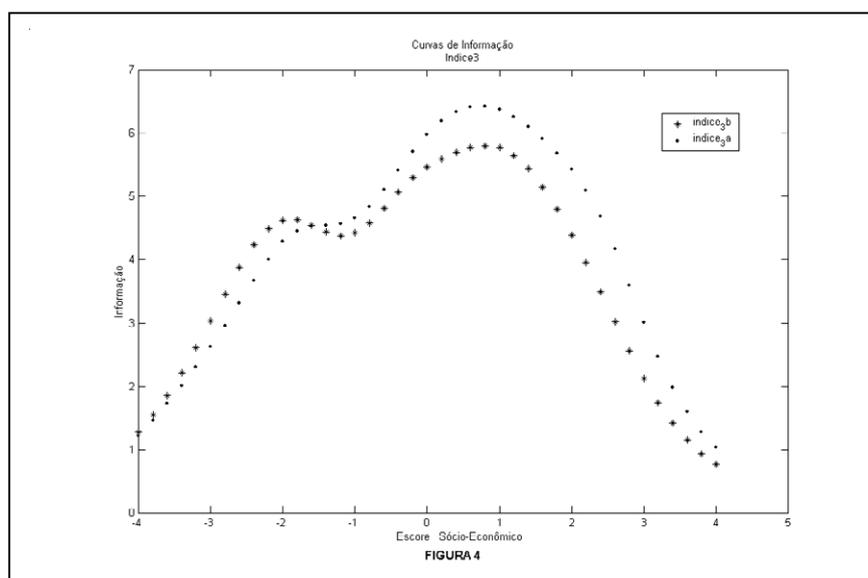
Finalmente, foram utilizados os modelos de respostas graduadas na construção do índice_3. Para a calibração dos modelos foi utilizado o *software Parscale*[®] (cf. Muraki & Bock, 1995). Inicialmente produziu-se uma análise da proporção de posses e da correlação polisserial de cada variável, valores que estão apresentados na Tabela 10 (no anexo), com a respectiva

descrição das questões. Dois índices foram construídos utilizando o modelo para respostas graduadas. Assim, no índice_3a foram consideradas 16 variáveis (as variáveis Q07, Q08 e Q11, foram excluídas por apresentarem baixo valor para a correlação polisserial). Eventualmente, apenas uma categoria de um item pode vir a ser eliminada. E, no índice_3b, foram consideradas 14 variáveis, com base na análise de dimensionalidade anteriormente produzida, tendo sido excluídas, ainda, as questões Q09 e Q10. Uma tentativa de produzir um índice incluindo todas as variáveis não foi bem sucedida, pois não se alcançou uma boa convergência no método de estimação dos parâmetros dos modelos. Provavelmente, devido à inadequação do modelo originada pela existência de uma ou mais dimensões adicionais. Os parâmetros dos modelos, estimados para os dois índices, são apresentados na Tabela 11 (no anexo). Para se decidir sobre qual desses dois índices deve ser empregado no caso de se optar por modelos de respostas graduadas fez-se, inicialmente, uma análise da correlação de Pearson para os dois índices encontrando-se o valor de 0.989 para essa correlação. Além disso, foi calculada a média dos desvios absolutos entre os dois escores, encontrando-se o valor 0.1. A distribuição dos escores, segundo alguns percentis, foi também obtida e seus valores foram os seguintes:

Tabela 7 – Valores percentis para os índices.

	5%	10%	25%	50%	75%	90%	95%
Índice_3a	-1.54	-1.14	-0.61	0.00	0.63	1.18	1.50
Índice_3b	-1.55	-1.16	-0.62	0.00	0.63	1.18	1.50

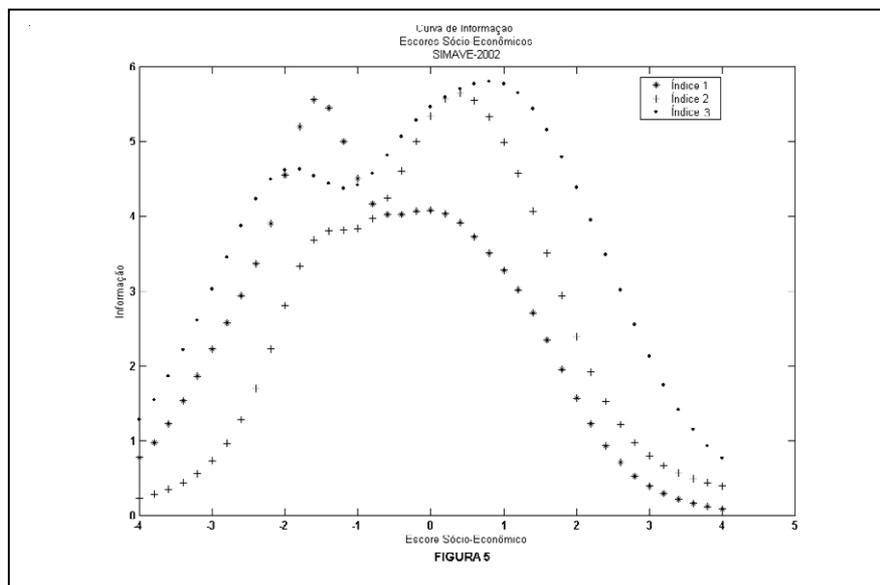
Finalmente, foi analisada a curva de informação esperada para os dois índices, construída a partir dos modelos de respostas graduadas. Um programa específico foi construído usando rotinas do *software matlab*[®], e as curvas de informação estão apresentadas na Figura 4.



A conclusão final é que, devido aos resultados muito similares, o índice_3a, por apresentar um pouco mais de informação que o índice_3b, deve ser empregado. Por outro lado, em futuros levantamentos, justamente tendo em vista a proximidade entre esses resultados, pode-se pensar em utilizar apenas os 14 itens para os quais o índice_3b foi construído, pois implica numa economia quanto ao número de questões apresentadas no instrumento de avaliação destinadas a produção do indicador sócio-econômico. Na próxima seção quando for referido índice_3, estará subentendido que se está referindo ao índice_3a.

5.5 Análise comparativa dos 3 índices produzidos

Finalmente, vai-se fazer um estudo comparativo dos índices produzidos pelos 3 métodos. Lembrando, o índice_1 foi construído a partir da dicotomização direta das questões (tem ou não tem um determinado bem) e utilizando-se modelos binários de dois parâmetros; o índice_2 foi construído a partir da dicotomização das opções de resposta de cada item, selecionando-se as mais informativas, e utilizando-se também modelos binários de dois parâmetros; e, o índice_3 foi construído utilizando-se de modelos para respostas graduadas. A primeira análise é a da informação obtida para os três índices. Como se pode observar pelas curvas de informação dos três índices, o índice_1 apresenta maiores níveis de informação nos níveis de escores menos elevados e, o índice_2 nos níveis de escores mais elevados. Ambos produzem bom nível de informação entre os níveis -2 e 2 (cerca de 95% dos casos para uma distribuição normal). Já o índice_3 apresenta boa informação ao longo de toda a escala de habilidades.



Assim, do ponto de vista da informação do índice, como era esperado, deve-se preferir utilizar o índice_3. Mas, esse argumento não invalida o emprego dos dois outros índices caso não seja possível construí-lo por ausência de *softwares* adequados, por exemplo, ou ainda, quando se deseja um instrumento mais preciso. De fato, quando se analisa a correlação de

Pearson entre os três índices, encontram-se os valores de 0,963 para a correlação entre o Índice_1 e o Índice_2, 0,944 entre o Índice_1 e o Índice_3 e, de 0,968 entre o Índice_2 e o Índice_3, indicando grande proximidade entre os escores produzidos. Por outro lado, a distribuição segundo os valores percentis para os três índices foi a seguinte:

Tabela 8 – Valores percentis para os índices.

	5%	10%	25%	50%	75%	90%	95%
Índice_1	-1,55	-1,22	-0,60	0,05	0,59	1,23	1,46
Índice_2	-1,54	-1,12	-0,62	0,00	0,60	1,26	1,58
Índice_3	-1,54	-1,14	-0,61	0,00	0,63	1,18	1,50

Novamente, a distribuição dos índices segundo os valores percentis indicam pouca diferença entre os 3 índices. Assim, se o objetivo da análise é a de produzir indicadores para agregados de indivíduos, qualquer um dos índices, provavelmente, será muito bom. No entanto, se o objetivo é produzir um indicador por indivíduo, claro que a preferência deve recair-se sobre aquele que apresenta maior informação.

Um índice sócio-econômico pode ser construído para diversas finalidades. Entre elas, no caso de avaliações educacionais, a mais comum, talvez, seja a de avaliar o efeito que este apresenta sobre a proficiência do aluno. Frequentemente, deseja-se que essa variável sirva de controle quando se está interessado em avaliar o efeito que outras variáveis possam exercer sobre a proficiência. Vai-se, então, comparar os três índices acima através do efeito que apresentam sobre a proficiência através de modelos de regressão hierárquicos (ver, por exemplo, Barbosa & Fernandes, 2001; Soares, 2003). No presente estudo, foram considerados modelos de três níveis hierárquicos, isto é, aluno turma e escola. Por esse motivo o universo em estudo consiste das escolas com três ou mais turmas de quarta série. Três modelos diferentes foram então produzidos para avaliar o efeito de cada um dos três índices sócio-econômico sobre a proficiência do aluno. O modelo utilizado (cf. Bryk & Raudenbush, 1992) pode ser representado por:

$$y_{ijk} = \beta_{0,jk} + \beta_{1,jk} x_{1ijk} + c_{ijk}$$

$$\beta_{0,jk} = \gamma_{00k} + \mu_{0,jk}, \quad \gamma_{00k} = \pi_{000} + r_{00k}$$

$$\beta_{1,jk} = \gamma_{10k} + \mu_{1,jk}, \quad \gamma_{10k} = \pi_{100} + r_{10k}$$

onde y_{ijk} representa a proficiência do i -ésimo aluno da j -ésima turma da k -ésima escola e x_{1ijk} representa o escore sócio-econômico do aluno. As hipóteses adotadas em modelo com dois níveis são mantidas, isto é, $E(e, u_f) = 0$, $E(e, r_{fs}) = 0$ e, $E(r_{fs}, u_f) = 0$. Os erros do modelo são todos supostos normais. σ_e^2 , $\sigma_{u_f}^2$, $\sigma_{r_{fs}}^2$, são as variâncias dos parâmetros aleatórios do modelo, também, são denominadas como as componentes aleatórias do modelo. Os parâmetros π_{fst} são os parâmetros fixos do modelo. Os modelos estimados para todos os índices foram os seguintes:

Tabela 9 – Modelos estimados para os diversos índices.

Índice	π_{000}	π_{100}	σ_{u0}^2	σ_{u1}^2	σ_{r00}^2	σ_{r10}^2	σ_e^2	Deviance
Índice1_a	195,14 (0,95)	2,22 (0,32)	731,4	12,6	217,4	7,88	1309	439606
Índice1_b	195,14 (0,99)	2,16 (0,31)	731,8	12,7	217,5	7,93	1309	439609
Índice_2	195,0 (1,00)	2,26 (0,31)	731,0	11,2	214,7	9,68	1309	439583
Índice_3a	195,1 (1,00)	2,11 (0,30)	732,3	9,3	216,8	8,75	1310	439618
Índice_3b	195,2 (1,00)	1,65 (0,30)	736,3	7,9	223,5	8,46	1311	439633

Conclui-se dos resultados apresentados na tabela acima que tendo em vista o objetivo de se avaliar a influência do escore sócio-econômico sobre a proficiência em teste de avaliação educacional para o aluno, praticamente, todos os itens conduzem a resultados bastante semelhantes. Como o escore sócio-econômico é estimado, idealmente, o modelo deveria levar em consideração o erro de medida dessa variável. No presente estudo, como os diversos índices apresentam níveis de informação aproximados, provavelmente, os efeitos dos erros de medidas nos diferentes modelos devem ser muito próximos. Isso pode ser observado, por exemplo, pela proximidade dos valores estimados para o coeficiente π_{100} .

6. Conclusões

Neste trabalho foram comparados três tipos de indicadores, produzidos através de modelos da teoria da resposta ao item, para a condição sócio-econômica familiar de alunos que participaram de programas de avaliação educacional no estado de Minas Gerais. A finalidade do estudo foi, primeiramente, orientar a produção desse tipo de indicador no âmbito do programa de avaliação. Segundo, o de tentar encontrar o mínimo possível de variáveis a serem introduzidas no questionário que conduziriam a um indicador satisfatório. Sob esse aspecto, o trabalho demonstrou que de um total de 20 variáveis originalmente disponíveis, apenas 13 ou 14 delas podem ser empregadas para produzir o índice com qualidade. Por outro lado, demonstrou-se que, pelo menos para o tipo de aplicação e finalidade aqui apresentado, praticamente é indiferente o fato de se empregar um modelo de dois parâmetros ou um modelo mais geral de repostas graduadas, o que permite utilizar questões mais simplificadas na elaboração do instrumento. Esteja claro, no entanto, que se for apenas uma questão de escolha, ou ainda, se o objetivo for produzir uma medida para o indivíduo em si, o modelo mais geral deve ter, naturalmente, a preferência, pois permite um melhor aproveitamento da informação dos dados e, conseqüentemente, maior precisão nas estimativas dos índices calculados para os indivíduos.

Estudos futuros devem focar a comparação do critério Brasil com os indicadores produzidos através dos modelos da TRI. Parece claro, que há muitas vantagens destes últimos, resta saber se essas vantagens justificam a maior complexidade metodológica em aplicações práticas.

Agradecimentos

O autor agradece ao Caed e à Secretaria de Educação do Estado de Minas Gerais pela cessão dos dados e apoio ao trabalho. Agradece, em especial, ao Professor Manuel Palácios, Diretor do CAEd, pelas sugestões ao trabalho. Agradece, ainda, aos comentários e sugestões dos três revisores anônimos.

Referências Bibliográficas

- (1) Andrade, D.F.; Tavares, H.R. & Valle, R.C. (2001). *Teoria da Resposta ao Item: Conceitos e Aplicações*. ABE – Associação Brasileira de Estatística, 2000, São Paulo.
- (2) Backer, F.B. (1992). *Item Response Theory*. STATISTICS: Textbooks and Monographs, vol. 129. Marcel Dekker, Inc., New York, USA.
- (3) Barbosa, M.E. & Fernandes, C. (2000). Modelo multinível: uma aplicação a dados de avaliação educacional. *Estudos em Avaliação Educacional: Fundação Carlo Chagas*, **22**, 135-153.
- (4) Birnbaum, A. (1968). Some Latent Traits Models and Their Use in Inferring na Examinee's Ability. **In: Statistical Theories of Mental Test Scores** [edited by F.M. Lord and M.R. Novick], 397-472. Addison-Wesley, Reading, MA.
- (5) Bock, R.D. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, **46**, 443-459.
- (6) Bock, R.D.; Gibbons, R.D. & Muraki, E. (1988). Full-Information Factor Analysis. *Applied Psychological Measurement*, **12**, 261-280.
- (7) Bock, R.D. & Lieberman, D. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, **35**, 179-197.
- (8) Bryk, S.A. & Raudenbush, W. (1992). *Hierarchical Linear Models*. Sage Publications, Inc., Newbury Park, Califórnia.
- (9) Dempster, A.P.; Laird, N.M. & Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Statistics Society B*, **39**, 1-38.
- (10) Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, **44**, 169-172.
- (11) Fletcher, P.R. (1998). *À Procura do Ensino Eficaz*. Ministério da Educação e Cultura, Departamento da Avaliação da Educação Básica, Rio de Janeiro.
- (12) Hambleton, R.K.; Swaminathan, H. & Rogers, H (1991). *Fundamentals of Item Response Theory*. Sage Publications, Inc., Newbury Park, Califórnia.
- (13) Lord, F.M. (1952). A Theory of Test Scores. *Psychometric Monograph*, **7**.
- (14) Minoux, M. (1986). *Mathematical Programming: Theory and Algorithms*. John Wiley & Sons Ltd., Norwich, UK.
- (15) Muraki, E. & Engelhard, G. (1985). Full Information Item Factor Analysis: applications of EAP scores. *Applied Psychological Measurement*, **9**, 417-430.

- (16) Muraki, E. & Bock, R.D. (1995). PARSCALE: Parameter Scaling of Rating Data [Computer Program]. Scientific Software, Inc., Chicago, IL.
- (17) Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Institute for Educational Research, Copenhagen, Danish.
- (18) Samejima, F. (1997). Graded Response Model. **In: Handbook of Modern Item Response Theory** [edited by W. Van der Linden and R. Hambleton]. Springer-Verlag, New York.
- (19) Soares, F.J. & Mambrini, J. (2003). Medida do Nível Socioeconômico de Estudantes em Pesquisas Educacionais. Anais da 35ª Reunião Regional da ABE e Reunião Regional da SOBRAPO. Florianópolis, SC, maio 2003.
- (20) 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*. Fundação Carlos Chagas. Jul/Dez, **28**.
- (21) Spector, P.E. (1992). *Summated Rating Scale Construction: An Introduction*. Series: Quantitative Applications in the Social Sciences. Sage Publications, London.
- (22) Thissen, D. & Wainer, H. (2001). *Test Scoring*. Lawrence Erlbaum Associates Pub, Mahwah, New Jersey.
- (23) Wilson, D.T.; Wood, R. & Gibbons, R. (1998). *Testfact: Test Scoring, and Item Factor Analysis*. Scientific Software International, Inc., Lincolnwood, Illinois.

Anexo A

Tabela 10 – Valores para a correlação polisserial para os itens do questionário.

		Correlação		Descrição
		Ponto Polisserial	Polisserial	
Q07_0 Q07_1	7,0% 91,7%	0,158	0,298	A casa em que mora não tem água encanada A casa em que mora tem água encanada
Q08_0 Q08_1	2,6% 95,8%	0,194	0,504	A casa em que mora não tem eletricidade A casa em que mora tem eletricidade
Q09_0 Q09_1	26,4% 71,4%	0,383	0,515	A rua em que mora não tem calçamento ou asfalto A rua em que mora tem calçamento ou asfalto
Q10_0 Q10_1	14,5% 83,5%	0,321	0,493	A rua em que mora não tem coleta de lixo A rua em que mora tem coleta de lixo
Q11_0 Q11_1	1,2% 97,3%	0,071	0,236	A casa em que mora não tem cozinha A casa em que mora tem cozinha
Q12_0 Q12_1 Q12_2 Q12_3	1,6% 72,9% 18,9% 4,6%	0,577	0,728	A casa em que mora não tem banheiro A casa em que mora tem 1 banheiro A casa em que mora tem 2 banheiros A casa em que mora tem 3 banheiros ou mais
Q13_0 Q13_1 Q13_2 Q13_3	2,9% 73,1% 19,4% 2,3%	0,431	0,536	A casa em que mora não tem sala A casa em que mora tem 1 sala A casa em que mora tem 2 salas A casa em que mora tem 3 salas ou mais

Q14_0	0,3%	0,384	0,410	A casa em que mora não tem quartos
Q14_1	8,1%			A casa em que mora tem 1 quarto
Q14_2	28,9%			A casa em que mora tem 2 quartos
Q14_3	42,5%			A casa em que mora tem 3 quartos
Q14_4	17,9%			A casa em que mora tem 4 quartos ou mais
Q15_0	8,9%	0,605	0,694	A casa em que mora não tem televisões a cores
Q15_1	56,7%			A casa em que mora tem 1 televisão a cores
Q15_2	32,3%			A casa em que mora tem 2 televisões a cores ou mais
Q16_0	8,2%	0,452	0,620	A casa em que mora não tem geladeira
Q16_1	83,8%			A casa em que mora tem 1 geladeira
Q16_2	6,5%			A casa em que mora tem 2 geladeiras ou mais
Q17_0	19,0%	0,373	0,535	A casa em que você mora não tem máq. de lavar roupa
Q17_1	79,2%			A casa em que mora tem 1 máquina de lavar ou mais
Q18_0	13,2%	0,518	0,593	A casa em que você mora não tem aparelho de som
Q18_1	62,5%			A casa em que você mora tem 1 aparelho de som
Q18_2	22,2%			A casa em que você mora tem 2 ap. de som ou mais
Q19_0	55,8%	0,576	0,690	A casa em que você mora não tem videocassete
Q19_1	38,9%			A casa em que você mora tem 1 videocassete
Q19_2	3,7%			A casa em que você mora tem 2 videocassetes ou mais
Q20_0	74,7%	0,445	0,619	A casa em que você mora não tem freezer
Q20_1	21,8%			A casa em que você mora tem 1 freezer ou mais
Q21_0	37,2%	0,508	0,647	A casa em que você mora não tem telefone
Q21_1	60,6%			A casa em que você mora tem 1 telefone ou mais
Q22_0	63,2%	0,465	0,599	A casa em que você mora não tem telefone celular
Q22_1	34,1%			A casa em que você mora tem telefone celular
Q23_0	81,8%	0,487	0,747	A casa em que você mora não tem computador
Q23_1	14,3%			A casa em que você mora em computador
Q24_0	56,9%	0,520	0,616	A casa em que você mora não tem automóvel
Q24_1	31,6%			A casa em que você mora tem 1 automóvel
Q24_2	9,6%			A casa em que você mora tem 2 ou mais automóveis
Q33R_0	7,2%	0,547	0,591	O pai (ou responsável) não sabe ler
Q33R_1	20,0%			O pai não completou a 4ª série do ensino fundamental
Q33R_2	36,0%			O pai (ou responsável) completou a 4ª série
Q33R_3	13,7%			O pai (ou responsável) completou o ensino fundamental
Q33R_4	12,0%			O pai (ou responsável) completou o ensino médio
Q33R_5	8,6%			O pai (ou responsável) completou um curso superior

Tabela 11 – Parâmetros estimados dos modelos (Índice_3).

QUESTÃO	Parâmetro de Discriminação (a) Índice_3a	Parâmetro de Dificuldade (b) Índice_3a		Parâmetro de Discriminação (a) Índice_3b	Parâmetro de Dificuldade (b) Índice_3b	
Q09_0 Q09_1	0,718 (0,003)	-1,043 (0,005)	∞ 0,0	*	*	*
Q10_0 Q10_1	0,716 (0,003)	-1,796 (0,007)	∞ 0,0	*	*	*
Q12_0 Q12_1 Q12_2 Q12_3	1,020 (0,002)	0,065 (0,003)	∞ 3,245 -0,921 -2,352	1,024 (0,002)	0,067 (0,003)	∞ 3,230 -0,916 -2,314

Q13_0	0,476	0,668	∞	0,501	0,640	∞
Q13_1	(0,001)	(0,005)	5,335	(0,001)	(0,005)	5,125
Q13_2			-1,067			-1,030
Q13_3			-4,268			-4,094
Q14_0	0,195	-5,564	∞	0,213	-5,125	∞
Q14_1	(0,000)	(0,009)	12,695	(0,000)	(0,008)	11,672
Q14_2			1,733			1,603
Q14_3			-4,130			-3,794
Q14_4			-10,299			-9,481
Q15_0	1,182	-0,613	∞	1,178	-0,614	∞
Q15_1	(0,003)	(0,002)	1,189	(0,003)	(0,002)	1,192
Q15_2			-1,189			-1,192
Q16_0	1,081	0,073	∞	1,099	0,071	∞
Q16_1	(0,002)	(0,003)	2,014	(0,002)	(0,003)	2,001
Q16_2			-2,014			-2,001
Q17_0	0,821	-1,361	∞	0,824	-1,359	∞
Q17_1	(0,004)	(0,005)	0,0	(0,004)	(0,005)	0,0
Q18_0	0,766	-0,300	∞	0,777	-0,298	∞
Q18_1	(0,001)	(0,003)	1,535	(0,001)	(0,003)	1,523
Q18_2			-1,535			-1,523
Q19_0	1,207	1,301	∞	1,201	-1,304	∞
Q19_1	(0,003)	(0,002)	1,065	(0,003)	(0,002)	1,068
Q19_2			-1,065			-1,068
Q20_0	0,785	1,225	∞	0,810	1,204	∞
Q20_1	(0,003)	(0,005)	0,0	(0,004)	(0,004)	0,0
Q21_0	1,315	-0,368	∞	1,200	-0,382	∞
Q21_1	(0,005)	(0,002)	0,0	(0,005)	(0,003)	0,0
Q22_0	0,826	0,609	∞	0,829	0,609	∞
Q22_1	(0,003)	(0,003)	0,0	(0,003)	(0,003)	0,0
Q23_0	1,246	1,354	∞	1,237	1,361	∞
Q23_1	(0,006)	(0,004)	0,0	(0,006)	(0,004)	0,0
Q24_0	0,677	1,368	∞	0,707	1,328	∞
Q24_1	(0,002)	(0,003)	0,994	(0,002)	(0,003)	0,964
Q24_2			-0,994			-0,964
Q33R_0	0,584	0,379	∞	∞	0,028	∞
Q33R_1	(0,001)	(0,003)	3,016	2,320	(0,002)	2,327
Q33R_2			0,033	1,980		1,982
Q33R_3				-0,419		-0,415
Q33R_4				-1,347		-1,346
Q33R_5				-2,552		-2,548