

Aspectos inferenciais no Modelo de Crédito Parcial Generalizado da Teoria da Resposta ao Item

Eduardo V. Ferreira^{1†}, Caio L. N. Azevedo²

¹ *Discente do Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica da UNICAMP. Rua Sérgio Buarque de Holanda, 651 - Cidade Universitária. Barão Geraldo 13083859 - Campinas, SP ? Brasil.*

² *Professor do Departamento de Estatística do Instituto de Matemática, Estatística e Computação Científica da UNICAMP. E-mail: cnaber@ime.unicamp.br.*

Resumo: *O Modelo de Crédito Parcial Generalizado (MCPG) pertence à família dos modelos para as respostas politômicas graduais da Teoria de Resposta ao Item (TRI). É apropriado na modelagem de itens (questões) cujas respostas pertencem a alguma escala gradual. Embora ele seja de grande importância, não tem recebido a devida atenção, em particular, na literatura em Português. Mesmo na internacional, alguns pontos não têm sido estudados em detalhes. Com isto em mente, o presente projeto apresentou os principais aspectos inferenciais (sob o enfoque frequentista) do MCPG. Especificamente, avaliou-se graficamente como mudanças nos valores dos parâmetros influenciam o comportamento da curva característica do item (CCI). Estudou-se as condições necessárias para assegurar a identificabilidade do modelo. Tais discussões serviram como base para a consecução do principal objetivo: desenvolver e implementar computacionalmente os métodos de estimação frequentista do MCPG. Através de estudos de simulação, avaliou-se a acurácia das estimativas considerando diversas situações de interesse prático. Os resultados indicaram que todos os métodos produzem resultados satisfatórios. Além disso, verificou-se os fatores de maior impacto na precisão das estimativas dos traços latentes (número de itens e categorias), dos parâmetros de discriminação (número de indivíduos e categorias) e parâmetros de dificuldade (número de itens, categorias e indivíduos).*

Palavras-chave: Teoria da Resposta ao Item; Identificabilidade; Inferência frequentista.

Abstract: *The Generalized Partial Credit Model (GPCM) belongs to the family of models of gradual polytomous response of Item Response Theory (IRT). It is appropriate for modeling items (questions) which belong to some gradual scale. Although it is very important, has not received proper attention, in particular, in Portuguese literature. Even in the international, some points have not been studied in details. With this in mind, in this project we presented main inferential aspects (under the frequentist approach) of GPCM. Specifically, we studied graphically how changes in parameters values influencing a behavior of the item characteristic curve (ICC). We studied necessary conditions to ensure the identifiability of the model. These discussions served as basis for the main goal achievement: to develop and implement computationally a frequentist estimation method of GPCM. In simulation studies we evaluated an estimates accuracy considering different situations of practical interest. The results indicated that all methods produced reasonable results. In addition, we found the factors with the greatest impact on the estimates of latent traits (number of items and categories), the discrimination parameters (number of individuals and categories) and difficulty parameters (number of items, categories and individuals).*

Keywords: Item Response Theory; Identifiability; Frequentist inference.

[†] *Autor correspondente: eduardozenus@yahoo.com.br*

Introdução

Em muitos instrumentos de medida (como provas, questionários de avaliação de produtos etc), os itens (questões) que os compõem apresentam alternativas que pertencem a alguma escala gradual (ordinal). Por exemplo, num questionário de avaliação de um produto por parte de consumidores, as categorias de resposta poderiam variar de zero (ruim) a cinco (excelente), de modo discreto. Ou, numa avaliação cognitiva, como no vestibular da Unicamp, questões dissertativas são apresentadas aos candidatos, e às suas respostas são atribuídos escore que varia de zero (erro total) à quatro (acerto total), de forma discreta. Nessas situações, tem-se interesse em avaliar os traços latentes (dos indivíduos) envolvidos. No primeiro, seria a avaliação do produto por parte dos consumidores enquanto na segunda seria o conhecimento dos vestibulandos. Em ambos os casos, o Modelo de Crédito Parcial Generalizado (MCPG), veja Muraki (1992) e Muraki (1997), poderia ser utilizado para se analisar tais dados. O MCPG pertence à família de modelos para as respostas politômicas graduais da Teoria de Resposta ao Item (TRI). Embora ele seja de grande importância, não tem recebido a devida atenção na literatura, em particular, na literatura em Português; mesmo na internacional, alguns pontos não têm sido estudados em detalhes como: interpretação dos parâmetros, estudo sobre a falta de identificabilidade e avaliação do desempenho de estimadores. Com isto em mente, o presente projeto tem por objetivo apresentar os principais aspectos inferenciais (sob o enfoque frequentista) do MCPG. Tais aspectos consistem em: interpretação dos parâmetros, comportamento dos estimadores de máxima verossimilhança (marginal - perfilada) e aspectos ligados a identificabilidade, viabilizando a compreensão de um tema que por vezes é abordado de modo casual e inadequado.

Modelo de Crédito Parcial Generalizado (MCPG)

O Modelo de Crédito Parcial Generalizado, formulado por Muraki (1992), pertence à família dos modelos para as respostas politômicas graduais, sendo, portanto, apropriado na modelagem de itens cuja resposta está relacionada à algum tipo de escala gradual, ou seja, o respondente ganha mais crédito na medida em que sua resposta se aproxima da completa. A probabilidade de um indivíduo j , escolher determinada categoria k do item i é dada por

$$P_{ijk}(\theta_j) = P(Y_{ijk} = 1 | \theta_j, \zeta_i) = \frac{\exp \left[\sum_{v=1}^k Da_i(\theta_j - b_{iv}) \right]}{\sum_{c=1}^m \exp \left[\sum_{v=1}^c Da_i(\theta_j - b_{iv}) \right]}, \quad (1)$$

em que $i = 1, 2, \dots, I$; $j = 1, 2, \dots, n$ e $k = 1, 2, \dots, m$.

- θ_j é o traço latente (proficiência ou habilidade) do j -ésimo indivíduo;
- a_i representa o parâmetro de discriminação (ou de inclinação) do i -ésimo item;
- $b_{ih} = b_i - d_h$ o parâmetro relacionado à dificuldade da h -ésima categoria do item i ;
- b_i representa o parâmetro de dificuldade do item;
- d_k representa o parâmetro de categoria, a dificuldade relativa da categoria k em comparação com as outras categorias. Em geral, espera-se que sejam ordenados;
- Y_{ijk} representa a resposta (binária) do indivíduo j à categoria k do item i ;
- D é um fator de escala, utiliza-se $D = 1,7$ quando se deseja que a função logística forneça resultados semelhantes ao da ogiva normal;
- $P_{ijk}(\theta_j)$ é chamada de Função de Resposta à Categoria do Item (FRCI).

Interpretação dos parâmetros

No MCPG, devido à complexidade do modelo, a interpretação dos parâmetros dos itens, através da visualização (análise gráfica) das CCI's, torna-se de extrema importância. Serão avaliados como os parâmetros dos itens afetam o comportamento das curvas (probabilidade de escolha) de cada categoria, bem como suas interpretações.

Parâmetro θ

Teoricamente o traço latente pode assumir qualquer valor entre $-\infty$ e ∞ , pertence a uma escala arbitrária que apenas ordena os indivíduos do menos hábil para o mais hábil, importando, portanto, somente as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. Esta escala pode ser definida de tal forma que a habilidade da maioria dos indivíduos assumam valores entre -3 e 3, com média zero. Porém, podem-se fazer transformações lineares de modo a apresentar os resultados em outra métrica. Por exemplo, o SAEB/PROVA BRASIL que utiliza a métrica (250, 50). Assim, na escala (0,1), um indivíduo com traço latente 1,5 está a 1,5 desvios-padrão acima do traço latente médio na escala (0,1); este mesmo indivíduo teria um traço latente de 325, valor 1,5 desvios-padrão acima do traço latente médio na escala do SAEB/PROVA BRASIL (ANDRADE et al., 2000).

Parâmetro a_i

O parâmetro a_i é denominado parâmetro de discriminação (ou de inclinação) do item. Para o modelo em estudo, não são esperados valores de a_i negativos, uma vez que indicaria que a probabilidade de atingir categorias “mais difíceis” diminui com o aumento do traço latente.

Como ilustrado na Figura 1, fixando os parâmetros b_{ik} , itens com maior valor do parâmetro a_i apresentam a função de resposta do item com inclinação mais acentuada. A consequência disto é que a diferença entre as probabilidades de dois indivíduos com traços latentes -2 e 0, por exemplo, de atingir uma determinada categoria de resposta é maior no item 4 do que no item 1, em outras palavras, o item 4 é mais apropriado para discriminá-los.

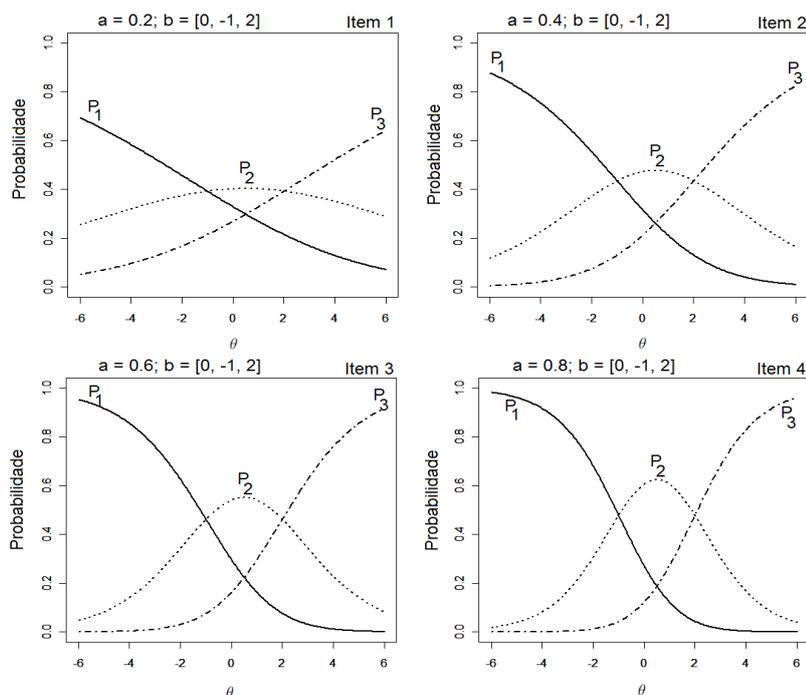


Figura 1: Representação gráfica de quatro itens variando sua discriminação

Parâmetros b_{ik}

Os parâmetros b_{ik} , $k = 1, \dots, m$, estão associados à dificuldade da categoria do item e apresentam valores na mesma escala do traço latente, com espaço paramétrico igual a \mathbb{R} . A escala comum permite o estabelecimento de significados mais objetivos sobre quais categorias o indivíduo terá mais chance de escolher. Assim, se os diferentes níveis de dificuldade estiverem vinculados a uma definição das operações cognitivas ou do conteúdo da tarefa, as informações obtidas na avaliação, na maioria das vezes, se tornam qualitativamente mais refinadas (ACÁCIA et al., 2002). A Figura 2 apresenta o gráfico da FRCI do MCPG de quatro itens com quatro categorias de resposta, em que $a_i = 1$ é mantido fixo para todos os itens, variando, somente, os valores dos parâmetros de dificuldade $b_i = (b_{i1}, b_{i2}, \dots, b_{im})$, com $b_1 = (0, -2, 0, 2)$, $b_2 = (0, -0.5, 0, 2)$, $b_3 = (0, 0, 0, 2)$ e $b_4 = (0, 1, 0, 2)$. Genericamente, a diferença entre b_{ik+1} e b_{ik} representa o intervalo de valores de θ , no qual a probabilidade de escolha da categoria b_{ik} é maior do que a probabilidade de escolha das outras $k - 1$. Observando o item 1, nota-se que um indivíduo com traço latente entre $[-2, 0]$, por exemplo, tem maior probabilidade de optar pela categoria 2; o mesmo indivíduo, no item 4, tem maior probabilidade de optar pela categoria 1.

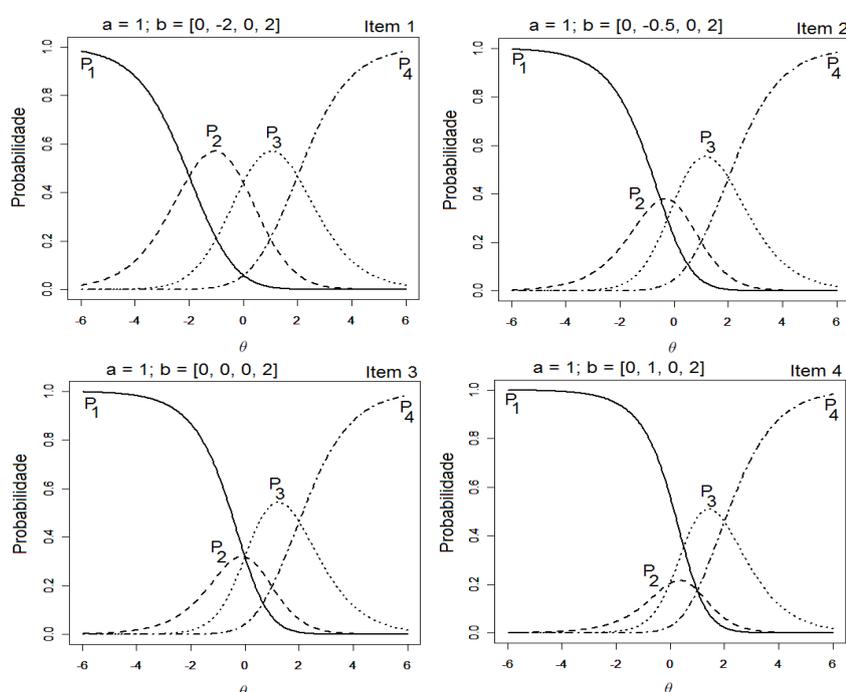


Figura 2: Representação gráfica de quatro itens variando sua dificuldade

Identificabilidade

Na TRI, o problema de falta de identificabilidade aparece com frequência, porém são escassos na literatura textos que viabilizem a compreensão de um tema que no desenvolvimento inferencial de acordo com Rivers (2003) “... é por vezes abordado de modo casual e inadequado”. E afirma também, que a identificabilidade, apesar de não ser suficiente, é condição necessária para garantir a consistência dos estimadores de máxima verossimilhança. Andrade et al. (2000) destacam que a falta de identificabilidade ocorre porque diferentes valores dos parâmetros dos itens ou da proficiência podem produzir a mesma probabilidade de um indivíduo escolher certa categoria, inviabilizando o desenvolvimento de qualquer procedimento de utilização direta da função de verossimilhança. De acordo com Matos (2008), a solução para não-identificabilidade depende basicamente do modelo em estudo e do método de estimação adotado. Na TRI, exis-

te um tipo de falta de identificabilidade devido a invariância às transformações lineares, mais conhecido como um problema de indeterminação da escala de medida, que muitas vezes é solucionada estabelecendo uma métrica (escala de medida com localização e escala, como referência de possíveis valores para os parâmetros) para um dos dois conjuntos de parâmetros, itens ou traços latentes. Porém, no MCPG, devido sua estrutura complexa, na maioria das vezes, é necessário restringir ambas as classes de parâmetros.

A indeterminação entre o conjunto de parâmetros de categoria, d_v , $v = 2, \dots, m$, com os parâmetros a_i e b_i e a variável θ_j , pode ser expressa por

$$a_i(\theta_j - b_i + d_v) = \frac{a_i}{s} [s\theta_j - (sb_i - t) + (sd_v - t)], \quad (2)$$

em que s é um fator de escala, e t é um fator de localização, ambos conhecidos.

Visando obter um único conjunto de estimativas para os parâmetros, a restrição, chamada restrição de localização, deve ser imposta à estimação dos parâmetros de categoria. Tal imposição é aplicada para que a média dos parâmetros de categoria, dentro de uma escala categórica, seja constante. Uma escolha natural é 0, ou seja

$$\sum_{v=2}^m d_v = 0. \quad (3)$$

Além disso, deve-se atribuir no processo de estimação (frequentista) uma distribuição de probabilidade para o traço latente, em geral com média e variância conhecidos, por exemplo,

$$\theta_1, \dots, \theta_n \stackrel{\text{iid}}{\sim} N(0, 1).$$

Estimação frequentista

Para estimação dos parâmetros dos itens e dos traços latentes, utilizar-se-á a variável aleatória Y_{ijk} , que representa a resposta (binária) do indivíduo j à categoria k do item i

$$Y_{ijk} = \begin{cases} 1, & \text{se o indivíduo } j \text{ no item } i \text{ escolhe a } k\text{-ésima categoria} \\ 0, & \text{caso contrário,} \end{cases}$$

com $j = 1, 2, \dots, n$, $i = 1, 2, \dots, I$ e $k = 1, 2, \dots, m$.

Seja $\theta = (\theta_1, \theta_2, \dots, \theta_n)'$ o vetor com os traços latentes dos n indivíduos e $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_I)'$ o conjunto de parâmetros dos itens. Será usada, em todo restante do texto, duas imprescindíveis suposições para o processo de estimação

(S1) Fixados os traços latentes e os parâmetros dos itens, as respostas oriundas de indivíduos diferentes são independentes;

(S2) Os itens são respondidos de forma independente por cada indivíduo (Independência Local), fixado seu traço latente e os parâmetros dos itens.

Estimação dos Parâmetros dos Itens

Como ocorre na prática, muitas das vezes não se conhece os traços latentes dos indivíduos quando se deseja estimar os parâmetros associados ao banco de itens. Diante disso, faz-se necessário lançar mão de algum artifício de forma que a função de verossimilhança não seja mais função das proficiências. Para tanto, Bock e Aitkin (1981) desenvolveram um método em que ao se marginalizar a função de verossimilhança (integrando-a em relação à distribuição dos traços latentes), esta não dependeria mais de θ . Esse artifício, denominado de Máxima Verossimilhança Marginal (MVM), é ainda hoje um dos métodos mais utilizados na prática devido sua facilidade computacional. Em geral considera-se uma densidade de $g(\theta|\eta)$, duplamente diferenciável, com

as componentes de η conhecidas e finitas (vetor de parâmetros associados a $g(\cdot)$). Para o caso mais explorado, em que θ tem distribuição normal, tem-se que $\eta = (\mu, \sigma^2)$, em que μ é média e σ^2 é a variância dos traços latentes dos indivíduos. Para indivíduos amostrados de uma população com distribuição normal (não sendo esta a única opção), denotada por $g(\theta|\eta)$, a probabilidade marginal de respostas para um determinado indivíduo j é igual a

$$P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \zeta, \eta) = \int_R P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta) d\theta \tag{4}$$

em que

$$P(\mathbf{Y}_{.j} | \theta, \zeta) = \prod_{i=1}^I \prod_{k=1}^m P_{ijk}^{y_{ijk}}$$

Portanto, usando (4), tem-se que verossimilhança marginal é dada por

$$L(\zeta, \eta) = \prod_{j=1}^n P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \zeta, \eta)$$

Segue que a log-verossimilhança pode ser escrita como

$$\begin{aligned} l(\zeta, \eta) &= \sum_{j=1}^n \ln \int_R P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta) d\theta \\ &= \sum_{j=1}^n \ln P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \zeta, \eta) \end{aligned}$$

Os Estimadores de Máxima Verossimilhança (EMV) de $\zeta_i, i = 1, 2, \dots, n$, são os valores que maximizam a verossimilhança, ou equivalentemente, são as soluções da equação

$$\frac{\partial \ln L}{\partial \zeta_i} = 0, \quad i = 1, 2, \dots, n.$$

Seja v_i representando um parâmetro do item em particular. De forma geral, a derivada de primeira ordem do logaritmo da função de verossimilhança com respeito a v_h , que é o parâmetro v_i para um específico valor $i = h$, pode ser expresso por

$$\begin{aligned} \frac{\partial \ln L}{\partial v_h} &= \frac{\partial}{\partial v_h} \left\{ \sum_{j=1}^n \ln P(\mathbf{Y}_{.j} = \mathbf{y}_{.j} | \zeta, \eta) \right\} \\ &= \sum_{j=1}^n \frac{1}{P(\mathbf{Y}_{.j} | \zeta, \eta)} \frac{\partial P(\mathbf{Y}_{.j} | \zeta, \eta)}{\partial v_h} \end{aligned} \tag{5}$$

Mas

$$\begin{aligned} \frac{\partial P(\mathbf{Y}_{.j} | \zeta, \eta)}{\partial v_h} &= \frac{\partial}{\partial v_h} \left\{ \int_R P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta) d\theta \right\} \\ &= \int_R P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta) \left\{ \sum_{k=1}^m \frac{y_{hjk}}{P_{hjk}} \left(\frac{\partial P_{hjk}}{\partial v_h} \right) \right\} d\theta \end{aligned}$$

Assim, a função de verossimilhança (5) pode ser escrita como

$$\frac{\partial \ln L}{\partial v_h} = \sum_{j=1}^n \left\{ \int_R \frac{P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta)}{\int_R P(\mathbf{Y}_{.j} | \theta, \zeta) g(\theta | \eta) d\theta} \left\{ \sum_{k=1}^m \frac{y_{hjk}}{P_{hjk}} \left(\frac{\partial P_{hjk}}{\partial v_h} \right) \right\} d\theta \right\}, \tag{6}$$

Pode-se notar que a Equação (6) envolve uma integral que não possui solução analítica. Por conta disso, algum meio deve ser encontrado para solução (aproximação) numérica da mesma. Embora existam muitos métodos de aproximação, na TRI têm sido frequente a aplicação do método *Hermite-Gauss*, usualmente denominado de *Método de quadratura gaussiana*.

Métodos de quadratura

Se $g(\theta|\eta)$ é uma função contínua (densidade) com integral finita (momentos finitos), ela pode ser aproximada, com qualquer grau de precisão, por uma outra função (discreta) que assume um número finito de pontos (*i.e.*, por um histograma). Dessa forma, o problema de obter a integral de uma função contínua é substituído pela obtenção da soma das áreas de um número finito, digamos q , de retângulos. Os pontos médios de cada retângulo, X_l , $l = 1, 2, \dots, q$, são denominados de nós (ou pontos de quadratura). Cada nó tem um peso $A_l \equiv A(X_l)$ associado que leva em conta a altura $g(X_l|\eta)$ e a largura Δ_l do respectivo intervalo, tal como $A_l = g(X_l|\eta) \times \Delta_l$ (ANDRADE et al., 2000; STROUD; SECREST, 1966). Convém dizer que não é necessário que $g(X_l|\eta)$ tenha distribuição normal, nem necessariamente contínua, em geral, ela pode ser definida empiricamente (MISLEVY; BOCK, 1982). A primeira derivada da função de verossimilhança na Equação (6), usando a quadratura de Gauss-Hermite, pode ser aproximada por

$$\frac{\partial \ln L}{\partial v_h} \approx \sum_{l=1}^q \sum_{k=1}^m \frac{\bar{r}_{hlk}}{P_{hlk}(X_l)} \frac{\partial [P_{hlk}(X_l)]}{\partial v_h},$$

em que \bar{r}_{hlk} é a frequência esperada provisória da k -ésima categoria de resposta do item h no l -ésimo ponto de quadratura, sendo descrito como

$$\bar{r}_{hlk} = \sum_{j=1}^n \frac{y_{hjk} L_j(X_l) A(X_l)}{\tilde{P}_j},$$

em que

$$\tilde{P}_j = \sum_{l=1}^q L_j(X_l) A(X_l),$$

e

$$L_j(X_l) = \prod_{i=1}^I \prod_{k=1}^m P_{ilk}^{y_{ijk}}.$$

$A(X_l)$ tem densidade de probabilidade aproximadamente normal padrão no ponto X_l ,

$$\sum_{l=1}^q A(X_l) = 1.$$

Contudo, as equações de estimação dos parâmetros dos itens em forma de quadratura não apresentam solução explícita. Para aplicação dos procedimentos Newton-Raphson ou Escore de Fisher, deve-se notar que as derivadas segundas de $\ln L(\zeta, \eta)$ com relação a ζ_u e ζ_v , para $u \neq v$, não são nulas. Portanto, a estimação dos parâmetros deve feita simultaneamente e, muito embora produza EMV com propriedades assintóticas desejáveis, devido à necessidade de inversão de matrizes, para um número alto de itens, este processo se torna inviável (BOCK; AITKIN, 1981).

Abordagem de Bock e Aitkin

Em 1981, Bock & Aitkin apresentaram uma reformulação conveniente nas equações de estimação dos parâmetros dos itens, de modo que as derivadas cruzadas da log-verossimilhança com relação aos parâmetros de diferentes itens se tornassem naturalmente nulas, isto é

$$\frac{\partial^2 \ln L(\zeta, \eta)}{\partial \zeta_u \partial \zeta_v} = 0, \quad \text{para } u \neq v.$$

Com isso, a matriz Hessiana, $H(\zeta)$, torna-se bloco diagonal, sendo possível que os mesmos sejam estimados individualmente (ANDRADE et al., 2000). Bock e Aitkin (1981) aplicaram o algoritmo pseudo-EM (DEMPSTER et al., 1977) para estimar os parâmetros de cada item, e assim repetiram o processo de iteração para todo os I itens, até que as estimativas dos mesmos se estabilizassem, segundo um critério de convergência.

Aplicação do algoritmo pseudo-EM

O algoritmo pseudo-EM é um processo iterativo para determinação de estimativas de ζ sem a presença de variáveis aleatórias não observadas, θ . Cada iteração é feita em dois passos: Esperança (E) e Maximização (M). No primeiro passo (E), a frequência esperada provisória da k -ésima categoria de resposta do item h no l -ésimo ponto de quadratura, \bar{r}_{hlk} , e o tamanho esperado provisório da amostra, \bar{f}_l (\bar{f}_{hl} se o tamanho da amostra difere para cada item), são calculados. Em seguida, no segundo passo (M) a estimação de máxima verossimilhança é obtida através do algoritmo Newton-Raphson ou Escore de Fisher (KENDALL; STUART, 1973, p.51). Estes passos compõem cada iteração do algoritmo pseudo-EM, e se repetirá até que um critério de convergência seja satisfeito. Pode-se resumir o algoritmo pseudo-EM para o MCPG da seguinte maneira:

1. Passo da Esperança (E):

Cálculo do tamanho esperado provisório da amostra e da frequência esperada provisória

$$\bar{f}_l = \sum_{j=1}^n \frac{L_j(X_l)A(X_l)}{\bar{P}_j} \quad \text{e} \quad \bar{r}_{hlk} = \sum_{j=1}^n \frac{y_{hjk}L_j(X_l)A(X_l)}{\bar{P}_j}$$

2. Passo da Maximização (M):

A próxima etapa consiste no cálculo dos parâmetros dos itens. Esses parâmetros são calculados um item por vez, tendo por base as quantidades obtidas no Passo (E), utilizando o algoritmo Newton-Raphson ou Escore de Fisher. Para tanto, considere $\hat{\zeta}^{(t)}$ a estimativa de ζ na iteração t , então na iteração $t + 1$ tem-se que

$$\hat{\zeta}^{(t+1)} = \hat{\zeta}^{(t)} + [\mathbf{V}(\hat{\zeta}^{(t)})]^{-1} \mathbf{S}(\hat{\zeta}^{(t)}),$$

em que $\hat{\zeta}^{(t+1)}$ e $\hat{\zeta}^{(t)}$ são as estimativas do parâmetro do $(t + 1)$ -ésimo e t -ésimo ciclo, respectivamente, $[\mathbf{V}(\hat{\zeta}^{(t)})]^{-1}$ é a inversa da matriz de informação, e $\mathbf{S}(\hat{\zeta}^{(t)})$ é o vetor escore, descrito como

$$\mathbf{S}(\zeta) = \begin{bmatrix} S(a_1) \\ S(b_1) \\ \vdots \\ S(a_I) \\ S(b_I) \\ S(d_2) \\ \vdots \\ S(d_m) \end{bmatrix} = \begin{bmatrix} a_1^{-1} \sum_{l=1}^q \sum_{k=1}^m Z_{1lk}^+ [\bar{r}_{1lk} - \bar{f}_{1l} P_{1lk}(\bar{\theta}_l)] \\ a_1 \sum_{l=1}^q \sum_{k=1}^m k [-\bar{r}_{1lk} + \bar{f}_{1l} P_{1lk}(\bar{\theta}_l)] \\ \vdots \\ a_I^{-1} \sum_{l=1}^q \sum_{k=1}^m Z_{Ilk}^+ [\bar{r}_{Ilk} - \bar{f}_{Il} P_{Ilk}(\bar{\theta}_l)] \\ a_I \sum_{l=1}^q \sum_{k=1}^m k [-\bar{r}_{Ilk} + \bar{f}_{Il} P_{Ilk}(\bar{\theta}_l)] \\ \sum_{h=1}^I a_h \sum_{l=1}^q \sum_{k=2}^m [\bar{r}_{hlk} - \bar{f}_{hl} P_{hlk}(\bar{\theta}_l)] \\ \vdots \\ \sum_{h=1}^I a_h \sum_{l=1}^q [\bar{r}_{hlm} - \bar{f}_{hl} P_{hlm}(\bar{\theta}_l)] \end{bmatrix}$$

Já os elementos da matriz de informação são dados por

$$\begin{aligned} V_{a_h a_h} &= \sum_{l=1}^q \bar{f}_{hl} \sum_{k=1}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial a_h} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial a_h} \right) \\ V_{b_h b_h} &= \sum_{l=1}^q \bar{f}_{hl} \sum_{k=1}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial b_h} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial b_h} \right) \\ V_{a_h b_h} &= \sum_{l=1}^q \bar{f}_{hl} \sum_{k=1}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial b_h} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial a_h} \right) \end{aligned}$$

Para $g' \leq g$

$$\begin{aligned} V_{d_g d_{g'}} &= \sum_{h=1}^I \sum_{l=1}^q \bar{f}_{hl} \sum_{k=g}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial d_{g'}} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial d_{g'}} \right) \\ V_{d_g a_h} &= \sum_{l=1}^q \bar{f}_{hl} \sum_{k=g}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial d_g} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial a_h} \right) \\ V_{d_g b_h} &= \sum_{l=1}^q \bar{f}_{hl} \sum_{k=g}^m \left(\frac{\partial Z_{hlk}^+(\bar{\theta}_l)}{\partial d_g} \right) \left(\frac{\partial P_{hlk}(\bar{\theta}_l)}{\partial b_h} \right). \end{aligned}$$

em que

$$\mathbf{Z}_{h,j}^+(\theta) = \{\mathbf{K}_h \mathbf{B}_h + \mathbf{Y}_h \mathbf{D}\},$$

com $\mathbf{B}_{h,j} = (\theta_j, -b_h)'$,

$$\mathbf{K} = \begin{bmatrix} a_h & 2a_h & \dots & ma_h \\ a_h & 2a_h & \dots & ma_h \end{bmatrix}',$$

$$\mathbf{Y} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ a_h & 0 & 0 & \dots & 0 \\ a_h & a_h & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ a_h & a_h & a_h & \dots & a_h \end{bmatrix},$$

e

$$\mathbf{D} = [d_2 \quad d_3 \quad \dots \quad d_m]'$$

Estimação dos Traços Latentes

A estimação dos traços latentes dos indivíduos é feita pelo método de máxima verossimilhança perfilada (MVP), considerando os parâmetros dos itens conhecidos. O processo consiste em introduzir as estimativas dos parâmetros dos itens, $\hat{\zeta}$, na verossimilhança original, $L(\zeta, \theta)$, obtendo assim uma verossimilhança do tipo perfilada $L(\theta, \hat{\zeta}) \equiv L(\theta)$, a qual possui propriedades semelhantes as da verossimilhança genuína (CORDEIRO, 1992). Sendo assim, pela independência entre as respostas de diferentes indivíduos (S1) e a independência local (S2), a verossimilhança é dada por

$$L(\theta) = \prod_{j=1}^n \prod_{i=1}^I \prod_{k=1}^m P_{ijk}^{y_{ijk}},$$

O estimador de máxima verossimilhança dos traços latentes é fornecido pela solução do sistema de equações de estimação

$$\begin{aligned} S(\theta_j) &\equiv \frac{\partial \ln L(\theta_j)}{\partial \theta_j} \\ &= \sum_{i=1}^I \sum_{k=1}^m \left\{ \frac{y_{ijk}}{P_{ijk}} \left(\frac{\partial P_{ijk}}{\partial \theta_j} \right) \right\}. \end{aligned}$$

Que resulta em

$$S(\theta_j) = \sum_{i=1}^I a_i \sum_{k=1}^m k [y_{ijk} - P_{ijk}(\theta_j)] \quad (7)$$

Novamente, a Equação (7) não apresenta solução explícita para θ_j e, por isso, precisa-se de algum método numérico para obter as estimativas desejadas. Para tanto, considere a matriz Hessiana dada por

$$\begin{aligned} H(\theta_j) &\equiv \frac{\partial^2 \ln L(\theta)}{\partial \theta_j^2} \\ &= \sum_{i=1}^I \sum_{k=1}^m - \left(\frac{\partial Z_{ijk}^+}{\partial \theta_j} \right) \left(\frac{\partial P_{ijk}}{\partial \theta_j} \right). \end{aligned} \quad (8)$$

Note que, a matriz Hessiana descrita em (8) é não-estocástica e, portanto a Informação de Fisher é definida como

$$\begin{aligned} V(\theta_j) &\equiv -E(H(\theta_j)) \\ &= -H(\theta_j) \end{aligned}$$

Newton-Raphson

Considerando $\hat{\theta}_j^{(t)}$ uma estimativa de θ_j na iteração t , então na iteração $t + 1$, tem-se

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} - [H(\hat{\theta}_j^{(t)})]^{-1} S(\hat{\theta}_j^{(t)})$$

Escore de Fisher

Para aplicação do método Escore de Fisher, basta substituir os componentes da matriz de derivadas segundas, Equação (8), usadas no processo iterativo de Newton-Raphson pelos seus valores esperados. Repare que nesse caso, ambos os processos iterativos são equivalentes

$$\hat{\theta}_j^{(t+1)} = \hat{\theta}_j^{(t)} + [V(\hat{\theta}_j^{(t)})]^{-1} S(\hat{\theta}_j^{(t)})$$

$t = 0, 1, 2, \dots$

Análises Envolvendo Dados Simulados

Conduzir-se-á o estudo para um conjunto de $R = 10$ réplicas, considerando três fatores, e seus níveis em parênteses, quais sejam: número de itens (NI) (20, 30, 40), número de categoria por item (NC) (4, 5) e número de examinandos (NE) (500, 1000, 2000). Portanto, um total de $3 \times 2 \times 3 = 18$ combinações. Os valores dos traços latentes dos indivíduos foram gerados a partir de uma distribuição normal com parâmetros de localização $\mu = 0$ e escala $\sigma = 1$, denotado por $N(0, 1)$, o parâmetro de discriminação variando de $[0.4, 1.2]$, parâmetro de dificuldade de $[-1, 1]$ e parâmetro de categoria $\mathbf{D}_4 = (d_2, d_3, d_4) = (-1.5, 0, 1.5)$ e $\mathbf{D}_5 = (d_2, d_3, d_4, d_5) =$

(-1.9, -0.8, 0.8, 1.9), para 4 e 5 categorias, respectivamente. A partir desses valores foi gerado um conjunto de respostas politômicas (DAVERY et al., 1997; AZEVEDO, 2002), supondo-se que este mecanismo produz respostas mais próximas daquelas que seriam geradas por respondentes quando submetidos a um conjunto de itens cujas respostas pertençam a alguma escala gradual. Com as estimativas obtidas em cada réplica, os resultados foram analisados utilizando estatísticas adequadas. Para tanto, seja ϑ um parâmetro de interesse e $\hat{\vartheta}_r$ a estimativa associada à réplica r . Defina também $\hat{\vartheta} = \frac{1}{R} \sum_{r=1}^R \hat{\vartheta}_r$. As estatísticas supracitadas são

- Vício: vício das estimativas: $(\vartheta - \hat{\vartheta})$;
- Var: variância das estimativas: $\frac{1}{R} \sum_{r=1}^R (\hat{\vartheta} - \hat{\vartheta}_r)^2$;
- REQM: raiz quadrada do erro quadrático médio: $\sqrt{\frac{1}{R} \sum_{r=1}^R (\vartheta - \hat{\vartheta}_r)^2}$;
- VAVR: valor absoluto do vício relativo: $\frac{|\vartheta - \hat{\vartheta}|}{|\vartheta|}$.

É digno de nota que, por haver vários traços latentes e parâmetros dos itens, cada estatística será obtida como a média dos diferentes valores dos parâmetros.

A fim de explorar o efeito do NE (número de examinandos), NI (número de itens) e NC (número de categoria) na precisão das estimativas dos parâmetros dos itens, empregou-se o planejamento fatorial completo tendo como resposta $\ln(\text{REQM})$, veja DeMars (2003). Note que, com esta variável, alguns pressupostos do modelo ANOVA como normalidade e homogeneidade não podem ser assegurados. Portanto, a estatística ω^2 foi considerada, sem a realização de qualquer teste estatístico. Sendo assim, para um específico fator, ω^2 é dado por

$$\omega^2 = \frac{SS_{fator} - df_{fator}MS_{erro}}{MS_{erro} + SS_{total}},$$

em que SS é a soma dos quadrados, MS é a média dos quadrados e df são os graus de liberdade. Isso nos permite avaliar a contribuição de cada efeito principal na diferença de precisão das estimativas.

Tabela 1: ANOVA para as estimativas dos parâmetros dos itens e traço latente

Fonte de Variação	Parâmetro de discriminação				Parâmetro de dificuldade				Traço latente			
	SS	df	MS	ω^2	SS	df	MS	ω^2	SS	df	MS	ω^2
NI	0,081	2	0,040	<0,01	66,098	2	33,049	0,577	434,286	2	150,309	0,122
NC	3,211	1	3,003	0,018	1,844	1	1,712	0,015	145,411	1	105,842	0,040
NE	59,748	2	28,126	0,363	7,115	2	5,293	0,061	8,643	2	4,322	<0,01
NI \times NC	0,034	2	0,017	<0,01	0,155	2	0,077	<0,01	0,220	2	0,072	<0,01
NI \times NE	0,573	4	0,143	<0,01	6,757	4	1,689	0,057	1,470	4	0,368	<0,01
NC \times NE	0,244	2	0,103	<0,01	0,345	2	0,121	<0,01	0,035	2	0,018	<0,01
NI \times NC \times NE	0,176	4	0,044	<0,01	0,468	4	0,117	<0,01	1,937	4	0,484	<0,01
Erro	98,926	522	0,189	-	31,348	522	0,060	-	2952,214	20982	0,141	-
Total	162,996	539	-	-	114,132	539	-	-	3544,218	20999	-	-

De acordo com a Tabela 1, no que concerne aos parâmetros de discriminação, os resultados apontam que o NE é o fator que contribui de modo mais eficaz para explicar a variabilidade da resposta, seguido pelo NC. Considerando os parâmetros de dificuldade, os três fatores são significativos, e, com exceção do NI \times NE, todas as interações apresentaram contribuições desprezíveis para explicar o $\ln(\text{REQM})$. Os resultados relativos ao VAVR são mostrados na Figura 3. Eles indicam que o aumento nos níveis dos fatores produzem melhores resultados, apesar de nos parâmetros de dificuldade esta melhora não ser tão evidente.

A partir da Figura 4 e de uma breve comparação entre as estatísticas constantes nas Tabelas 2, 3, 4 e 5, é possível perceber que existe uma variação menor entre as estimativas de cada

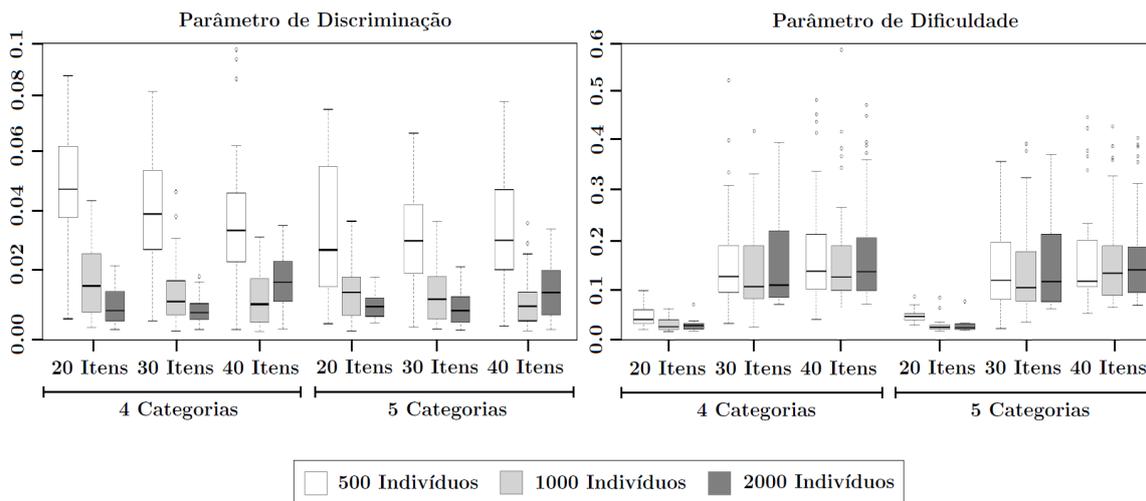


Figura 3: VAVR dos parâmetros dos itens

réplica do que em considerando suas médias. No entanto, ambas as variabilidades mostraram-se pequenas. Além disso, no que diz respeito ao REQm, na medida em que aumenta o número de examinandos e itens, de maneira geral, tal estatística tende a diminuir, sendo que para o parâmetro de dificuldade esse decréscimo acontece com maior vagar.

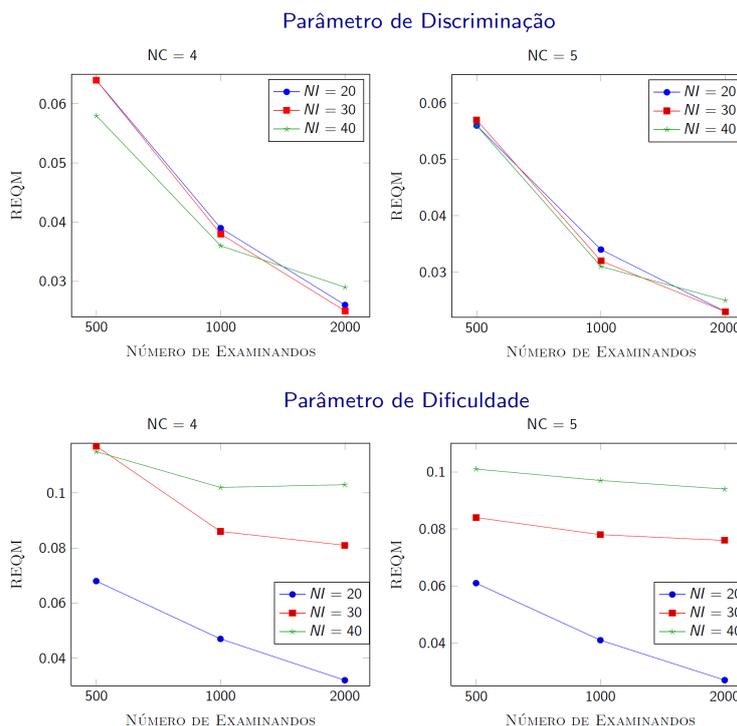


Figura 4: REQm dos parâmetros dos itens com traços latentes desconhecidos

Tabela 2: Estatísticas para os parâmetros de discriminação

NC	NI	Número de Examinandos											
		500				1000				2000			
		Média	Var*	REQM	Vício	Média	Var*	REQM	Vício	Média	Var*	REQM	Vício
4	20	0,842	2,691	0,064	-0,042	0,810	1,609	0,039	-0,010	0,794	0,780	0,026	0,005
	30	0,833	3,589	0,064	-0,033	0,803	1,595	0,038	-0,003	0,798	0,733	0,025	0,001
	40	0,836	2,737	0,058	-0,031	0,807	1,607	0,036	-0,002	0,790	0,754	0,029	0,014
5	20	0,828	3,087	0,056	-0,028	0,807	1,291	0,034	-0,007	0,795	0,605	0,023	0,004
	30	0,827	3,423	0,057	-0,027	0,806	1,360	0,032	-0,006	0,794	0,652	0,023	0,005
	40	0,833	2,818	0,056	-0,028	0,801	1,170	0,031	0,003	0,793	0,622	0,025	0,011

*valores ($\times 10^{-3}$).

Tabela 3: Estatísticas para os parâmetros de categoria de itens com 5 categorias

Parâmetro	NI	Número de Examinandos											
		500				1000				2000			
		Média	Var*	REQM	Vício	Média	Var*	REQM	Vício	Média	Var*	REQM	Vício
$d_1 = -1,9$	20	-1,840	6,217	0,095	-0,059	-1,861	1,317	0,051	-0,038	-1,917	1,497	0,040	0,017
	30	-1,839	1,954	0,073	-0,060	-1,894	5,552	0,070	-0,005	-1,917	0,967	0,034	0,017
	40	-1,846	3,294	0,076	-0,053	-1,928	0,574	0,036	0,028	-1,931	0,677	0,040	0,031
$d_2 = -0,8$	20	-0,811	18,189	0,128	0,011	-0,795	4,674	0,065	-0,004	-0,825	0,004	0,066	0,025
	30	-0,774	2,301	0,052	-0,025	-0,784	0,907	0,032	-0,015	-0,763	4,405	0,072	-0,036
	40	-0,763	4,405	0,072	-0,036	-0,789	2,312	0,046	-0,010	-0,814	1,786	0,042	0,014
$d_3 = 0,8$	20	0,813	19,563	0,133	-0,013	0,783	3,932	0,061	0,016	0,839	3,532	0,068	-0,039
	30	0,765	7,531	0,089	0,034	0,783	0,838	0,032	0,016	0,787	2,744	0,051	0,012
	40	0,796	5,504	0,070	0,003	0,808	3,244	0,054	-0,008	0,825	0,907	0,038	-0,025
$d_4 = 1,9$	20	1,838	4,177	0,087	0,061	1,873	2,644	0,055	0,026	1,903	3,534	0,056	-0,003
	30	1,846	2,485	0,071	0,053	1,884	2,818	0,052	0,015	1,914	0,950	0,032	-0,014
	40	1,813	3,167	0,101	0,086	1,909	0,552	0,024	-0,009	1,921	0,989	0,036	-0,021

*valores ($\times 10^{-3}$).

Tabela 4: Estatísticas para os parâmetros de categoria de itens com 4 categorias

Parâmetro	NI	Número de Examinandos											
		500				1000				2000			
		Média	Var*	REQM	Vício	Média	Var*	REQM	Vício	Média	Var*	REQM	Vício
$d_1 = 1,5$	20	-1,416	3,665	0,101	-0,083	-1,490	2,174	0,045	-0,009	-1,513	1,754	0,042	0,013
	30	-1,438	1,849	0,073	-0,061	-1,515	1,705	0,042	0,015	-1,513	0,886	0,031	0,013
	40	-1,446	2,373	0,070	-0,053	-1,496	1,302	0,034	-0,003	-1,528	0,899	0,039	0,028
$d_2 = 0$	20	-0,006	0,981	0,030	0,006	0,019	2,217	0,048	-0,019	-0,002	1,045	0,030	0,002
	30	0,003	7,890	0,084	-0,003	0,023	2,736	0,054	-0,023	0,001	0,634	0,023	-0,001
	40	0,016	2,161	0,046	-0,016	-0,008	1,785	0,041	0,008	0,000	0,617	0,023	0,000
$d_3 = 1,5$	20	1,422	2,633	0,091	0,077	1,470	1,178	0,043	0,029	1,516	1,982	0,045	-0,016
	30	1,435	7,005	0,102	0,064	1,491	0,931	0,030	0,008	1,511	0,720	0,028	-0,011
	40	1,430	2,130	0,082	0,069	1,505	1,034	0,030	-0,005	1,528	0,432	0,034	-0,028

*valores ($\times 10^{-3}$).

Tabela 5: Estatísticas para os parâmetros de dificuldade

NC	NI	Número de Examinandos											
		500				1000				2000			
		Média	Var*	REQM	Vício	Média	Var*	REQM	Vício	Média	Var*	REQM	Vício
4	20	0,000	4,641	0,068	0,000	0,001	2,460	0,047	-0,001	-0,007	1,275	0,032	0,007
	30	-0,011	6,926	0,117	0,084	0,004	2,962	0,086	0,069	-0,001	1,254	0,081	0,074
	40	0,001	5,478	0,115	0,091	0,003	2,900	0,102	0,089	-0,004	1,411	0,103	0,097
5	20	-0,003	3,634	0,061	0,003	0,000	1,898	0,041	0,000	-0,005	0,795	0,027	0,005
	30	0,004	2,624	0,084	0,068	0,005	1,625	0,078	0,067	0,002	0,880	0,076	0,071
	40	0,007	3,265	0,101	0,084	0,005	2,172	0,097	0,087	0,003	1,008	0,094	0,089

*valores ($\times 10^{-3}$).

Por meio da Figura 5, pode-se observar como se comporta a estatística VAVR nos diferentes cenários. Destaca-se que, em geral, para tamanhos maiores de teste, o valor absoluto do vício relativo tende a diminuir, o mesmo acontecendo quando aumenta-se o número de categorias do item. De maneira similar, a partir da Figura 6, verifica-se uma diminuição da REQM com o incremento dos níveis dos fatores, em especial, NI e NC, responsáveis por explicar a variabilidade

do $\ln(\text{REQM})$ dos traços latentes, veja Tabela 1.

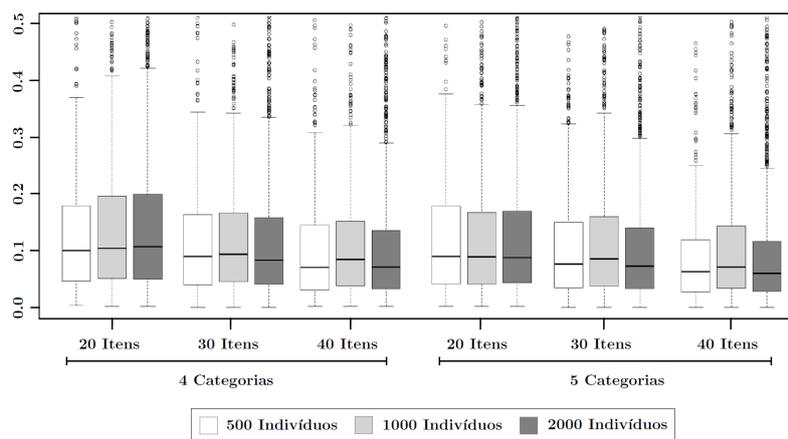


Figura 5: *VAVR dos traços latentes*

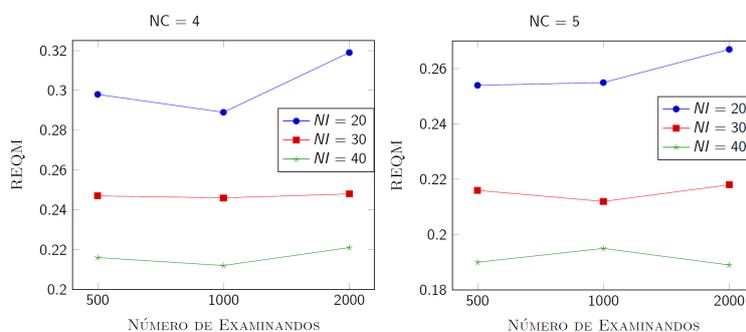


Figura 6: *REQM dos traços latentes*

Conclusões

No presente trabalho, num primeiro momento, a fim de melhor entender o significado dos parâmetros dos itens, avaliou-se, graficamente, como mudanças em seus valores influenciam o comportamento da CCI relacionada a cada categoria. Em relação à identificabilidade, condição necessária para garantir a estimabilidade do modelo, apresentou-se as condições para assegurar tal característica, com relação a aplicar restrições tanto nos parâmetros dos itens quanto nos traços latentes. Tais discussões serviram como base para a consecução do principal objetivo: desenvolver e implementar computacionalmente os métodos de estimação frequentista do MCPG. Através de estudos de simulação, avaliou-se a acurácia das estimativas considerando diversas situações de interesse prático. Os resultados indicaram que todos os métodos recuperam bem os parâmetros. Além disso, depreendeu-se os fatores de maior impacto na precisão das estimativas dos traços latentes (NI e NC), dos parâmetros de discriminação (NE e NC) e finalmente para os parâmetros de dificuldade (NI, NC, NE e NI×NE).

Acknowledgements

Os autores gostariam de agradecer ao CNPq através do Programa Institucional de Bolsas de Iniciação Científica - PIBIC/UNICAMP, concedida ao primeiro autor, sob a orientação do se-

gundo. Os autores agradecem também aos pareceristas anônimos pelos comentários perspicazes que melhorou substancialmente o artigo.

Referências

- ACÁCIA, A. et al. *O teste de cloze na avaliação da compreensão em leitura*. Psicologia: Reflexão e Crítica, 15(3), pp. 549-560, 2002.
- ANDRADE, D. F.; TAVARES, H. R.; VALLE, R.C. *Teoria de Resposta ao Item: conceitos e aplicações*. Associação Brasileira de Estatística - ABE, 14º Sinape, 2000.
- AZEVEDO, C. L. N. Métodos de estimação na teoria da resposta ao item: Máxima verossimilhança versus métodos bayesianos. *Trabalho apresentado ao 15º SINAPE: Águas de Lindóia, SP, 2002*.
- BOCK, R. D. & AITKIN, M. Marginal maximum likelihood estimation of item parameters: An application of an em algorithm. *Psychometrika*, 46(4):433 - 459, 1981
- CORDEIRO, G. M. *Introdução à Teoria de Verossimilhança*. 2º SINAPE, ABE, 1992.
- DAVERY, T., NERING, M. L., & THOMPSON, T. Realistic simulation of item response data. *ACT Research Report Series (www.act.org)*, 1997.
- DEMARS, C. E. Sample size and the recovery of nominal response model item parameters. *Applied Psychological Measurement*, 27(4):275-288, 2003.
- DEMPSTER, A. P., LAIRD, N. M., & RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39, (Series B), 1-38, 1977.
- KENDALL, M. G. & STUART, A. *The Advanced Theory of Statistics*. New York, NY: Hafner Publishing Company, vol.2, 1973.
- MATOS, G. S. *Modelos multidimensionais da TRI com distribuições assimétricas para os traços latentes*. Tese de doutorado. USP, 2008.
- MISLEVY, R. J. & BOCK, R. D. Biweight estimates of latent ability. *Educational and Psychological Measurement*, 42, 725-737, 1982.
- MURAKI, E. A generalized partial credit model: Application of an em algorithm. *Applied psychological measurement*, 16(2): 159-176, 1992.
- MURAKI, E. *A Generalized Partial Credit Model*. Wim J. van der Linden and Ronald K. Hambleton eds, Springer-Verlag, New York, NY, first edition, 1997.
- RIVERS, D. Identification of multidimensional spatial voting models. *Typescript, Department of political science, Stanford University, 2003*.
- STROUD, A. H. & SECREST, D. Gaussian quadrature formulas. *Englewood Cliffs Nj: Prentice Hall, 1966*.