SMNE UFPR
Simpósio de Métodos Numéricos em Engenharia
25 a 27 de outubro, 2017

Construção de Escalas para Avaliação em Conteúdos Matemáticos Básicos com Base na Teoria da Resposta ao Item

Tânia Robaskiewicz Coneglian Fujii
Universidade Estadual Paulista - FCT/Unesp
Presidente Prudente, Brasil
taniaconeglian@hotmail.com

Aparecida Donizete Pires de Souza
Departamento de Estatística
Universidade Estadual Paulista - FCT/Unesp
Presidente Prudente, Brasil
adps@fct.unesp.br

***Resumo*—Esta pesquisa tem por objetivo a construção e interpretação de escalas para avaliação de proficiência em conteúdos matemáticos básicos com base na Teoria da Resposta ao Item (TRI). A TRI consiste em um conjunto de modelos matemáticos, que representam a probabilidade de um indivíduo dar uma certa resposta a um item, como função dos parâmetros do item e da sua habilidade (ou traço latente) [2]. Neste trabalho, utilizou-se o modelo logístico unidimensional de três parâmetros (ML3) para estimar a proficiência em conteúdos matemáticos básicos, de ingressantes nos cursos da área de exatas. Este é um dos modelos mais utilizados para itens dicotômicos, caracterizado pelos parâmetros de discriminação, de dificuldade e de acerto ao acaso. Como instrumento de medida do conhecimento foi utilizada uma avaliação composta por trinta e seis itens de múltipla escolha, corrigidos de maneira dicotômica (certo ou errado). Devido à complexidade do modelo, para o procedimento de inferência sobre os parâmetros e habilidades, sob o paradigma clássico ou bayesiano, se faz necessário o uso de métodos numéricos. Neste caso, para a estimação dos parâmetros dos itens e das habilidades utilizou-se o amostrador de Gibbs, algoritmo da classe dos Métodos de Monte Carlo via Cadeia de Markov (MCMC) [3], implementado via software OpenBUGS (Bayesian inference Using Gibbs Sampling), direcionado para análise bayesiana de modelos complexos.**

***Palavras-chave*—Teoria de Resposta ao Item; Inferência Bayesiana; Construção de Escalas**

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

I. INTRODUÇÃO

De acordo com [2] a TRI consiste em um conjunto de modelos matemáticos, que representam a probabilidade de um indivíduo dar uma determinada resposta a um item em função dos parâmetros do item e da habilidade (ou traço latente) do respondente. É expressa de tal modo que, quanto maior a habilidade, maior será a probabilidade de resposta correta ao item. Os diversos modelos propostos na literatura dependem fundamentalmente da natureza do item (dicotômicos ou não), do números de populações envolvidas (apenas uma ou mais de uma) e da quantidade de traços latentes que está sendo medida (apenas um ou mais de um).

Neste contexto, o objetivo principal deste trabalho é a construção e interpretação de uma escala para medir a proficiência em conteúdos matemáticos básicos de alunos ingressantes nos cursos da área de exatas. Para que seja possível o desenvolvimento desta escala utilizou-se como população os alunos ingressantes nos cursos da área de exatas da FCT/Unesp. Foi utilizado como instrumento de medida do conhecimento, uma avaliação composta por trinta e seis itens de múltipla escolha, corrigidos de maneira dicotômica, elaborados com base na matriz de referência construída por [5] . O modelo matemático escolhido foi o modelo logístico unidimensional de três parâmetros (ML3), um dos modelos mais utilizados em testes educacionais. Este modelo é considerado o mais completo, pois leva em consideração a probabilidade de acerto casual. A estimação dos parâmetros e das habilidades foi feita com a utilização do software OpenBUGS, direcionado para análise bayesiana de modelos complexos.

II. O Modelo Logístico Unidimensional de Três Parâmetros (ML3)

O modelo (ML3) é dado por

$$P(U_{ij}=1 \mid \theta_j) = c_i + (1-c_i)\frac{1}{1+\exp^{-Da_i(\theta_j-b_j)}} \qquad (1)$$

com $i = 1, 2, \cdots, I$, e $j = 1, 2, \cdots, n$, em que, U_{ij} corresponde a uma variável dicotômica que assume os valores 1 quando o respondente j responde corretamente o item i ou 0 caso contrário; θ_j corresponde a habilidade (ou traço latente) do j-ésimo indivíduo; $P(U_{ij} = 1 \mid \theta_j)$ corresponde a probabilidade de um indivíduo j com traço latente θ_j responder corretamente o item i; b_i corresponde ao parâmetro de dificuldade (ou de posição) do item i, medido na mesma escala do traço latente; a_i corresponde ao parâmetro de discriminação (ou inclinação) do item i, com valor proporcional à inclinação da Curva Característica do Item - CCI no ponto b_i; c_i corresponde ao parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade (ou traço latente) responderem corretamente a um item i; D corresponde a um fator de escala, constante e igual a 1; quando se desejar que a função logística forneça resultados semelhantes aos da função ogiva normal, utiliza-se para o fator D o valor de 1,7.

No contexto bayesiano, para especificação completa do modelo, deve-se atribuir funções de distribuição a priori para os parâmetros dos itens e para a habilidade. Desta forma, assumiu-se as distribuições, $\theta_j \sim Normal(\mu_\theta, \sigma^2_\theta)$, $a_i \sim LogNormal(\mu_a, \sigma^2_a)$, $b_i \sim Normal(\mu_b, \sigma^2_b)$ e $c_i \sim Beta(\alpha, \beta)$. Assim, assumindo por exemplo $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ o vetor dos n indivíduos e $\boldsymbol{\lambda} = (\boldsymbol{\lambda_1}, \ldots, \boldsymbol{\lambda_I})'$ o vetor de parâmetros dos itens, com $\boldsymbol{\lambda}_i = (a_i, b_i, c_i)'$, $i = 1, \ldots, n$, a função densidade de probabilidade a posteriori conjunta é dada por

$$p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{u}) \propto L(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{u}) p(\boldsymbol{\theta}, \boldsymbol{\lambda}) \qquad (2)$$

em que $L(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{u}) = \prod_{j=1}^{n}\prod_{i=1}^{I} P(U_{ij} = u_{ij} \mid \theta_j, \boldsymbol{\lambda}_i)$ representa a função de verossimilhança e $p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ a função de distribuição de probabilidade a priori conjunta.

III. Método de Estimação

De acordo com equação (2) a distribuição a posteriori conjunta depende dos parâmetros e das habilidades de forma complexa e métodos numéricos se faz necessário no procedimento de inferência. Neste caso, será utilizado o amostrador de Gibbs, algoritmo da classe dos Métodos de Monte Carlo via Cadeia de Markov (MCMC) [3], já Implementado no software OpenBUGS (Bayesian Inference Using Gibbs Sampler) [4]. A ideia básica do MCMC é gerar amostras da distribuição a posteriori, a partir de distribuições condicionais completas que formam o núcleo de uma cadeia de Markov. Sob condições gerais de regularidade a distribuição da cadeia gerada pelo amostrador de Gibbs, converge para a distribuição de equilíbrio, neste caso a posteriori conjunta.

IV. A Escala de habilidades na TRI

Como já mencionado na definição do modelo, a distribuição normal padrão é associada ao parâmetro habilidade. Seguindo a metodologia da TRI, para a definição da escala é necessário estabelecer uma origem, associada ao nível zero, por exemplo 250, e o grau de espalhamento das habilidades dos indivíduos, escolhido de modo a representar o desvio-padrão, neste caso igual a 50.

Para tornar possível a interpretação pedagógica dos valores das habilidades, criou-se as escalas de conhecimento (ou escalas de proficiência). Estas escalas são definidas a partir de níveis âncoras. Estes níveis são definidos por itens âncora. Os níveis âncora são pontos na escala selecionados pelo pesquisador (analista) para serem interpretados pedagogicamente. Já os itens âncora, podem ser selecionados de acordo com os critérios sugeridos por [1]. Considerando dois níveis âncora consecutivos Y e Z com $Y < Z$, um determinado item é dito ser âncora para o nível Z se e somente se forem satisfeitas simultaneamente as seguintes condições: $P(U = 1 \mid \theta = Z) \geq 0,65$; $P(U = 1 \mid \theta = Y) < 0,50$ e $P(U = 1 \mid \theta = Z) - P(U = 1 \mid \theta = Y) \geq 0,30$.

Ou seja, para um item ser âncora de um determinado nível âncora da escala, ele precisa ser respondido corretamente por uma grande proporção de indivíduos ($\geq 65\%$) neste nível, deve ser respondido corretamente por uma proporção menor de indivíduos ($< 50\%$) com um nível de habilidade imediatamente inferior e além disso, a diferença entre essas proporção deve ser maior que 30%.

V. Considerações Finais

Os resultados mostraram que os itens propostos são eficientes para estimar a habilidade dos respondentes, sendo onze deles classificados como âncora. No entanto, adaptações precisam ser feitas de modo a garantir a interpretação em toda a amplitude da escala. Espera-se que este estudo contribua para a proposição de atividades que auxiliem os alunos no acompanhamento das disciplinas de cálculo e similares.

Agradecimentos

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo incentivo a esta pesquisa.

Referências

[1] A. E. Beaton and N. L. Allen, *Chapter 6: Interpreting Scales Through Scale Anchoring*, Journal of Educational Statistics, Sage Publications Sage CA: Los Angeles, CA, v.17, n. 2, p. 191-204, 1992.

[2] D. F. Andrade, H. R. Tavares and R. C. Valle *Teoria de Resposta ao Item: Conceitos e Aplicações*, São Paulo: 4º SINAPE, ABE, 2000.

[3] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2rd ed. Boca Raton: CRC Press, 2006.

[4] D. Lunn, D. Spiegelhalter, A. Thomas and N. Best, *The BUGS project: Evolution, critique and future directions (with discussion)*, Statistics in Medicine 28: 30493082, 2015.

[5] R. P. Rossi, *Construção de uma Escala para Avaliação da Proficiência em Conteúdos Matemáticos Básicos, 2015, 109 f.*, Dissertação (Mestrado) – Universidade Estadual Paulista, 2015.