



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

ESTIMAÇÃO DE PROFICIÊNCIAS COM ITENS
PRÉ-PONDERADOS EM MODELOS DE
RESPOSTA AO ITEM

Ailton Corecha de Souza

Orientador: Prof. Dr. Héilton Ribeiro Tavares

Belém
2010

Ailton Corecha de Souza

ESTIMAÇÃO DE PROFICIÊNCIAS COM ITENS
PRÉ-PONDERADOS EM MODELOS DE
RESPOSTA AO ITEM

Dissertação de Conclusão de Curso apresentado ao Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará para a obtenção do Título de Mestre em Estatística.

Área de Concentração: Teoria da Resposta ao Item

Orientador: Prof. Dr. Héilton Ribeiro Tavares

Belém
2010

Ailton Corecha de Souza

ESTIMAÇÃO DE PROFICIÊNCIAS COM ITENS
PRÉ-PONDERADOS EM MODELOS DE
RESPOSTA AO ITEM

Dissertação apresentada para a obtenção do Título de Mestre em Estatística no Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará.

Belém, 08 de Julho de 2010

Prof. Dr. Giovany de Jesus Malcher Figueiredo
Coordenador do Programa de Pós-Graduação em Matemática e Estatística

Banca Examinadora

Prof. Dr. Héilton Ribeiro Tavares
Universidade Federal do Pará
Orientador

Prof. Dr. Marcos Pinto da Costa Rocha
Universidade Federal do Pará
Examinador

Prof. Dr. Jorge Ferreira
Universidade Federal Rural de Pernambuco
Examinador

*A minha mãe Sandra Maria Corecha de Souza,
por tudo que me ensinou.*

Aos meus irmãos Albenor Júnior, Acácio, Adenilton e Ariane.

*E em especial a meu pai Albenor Nunes de Souza (in memoriam),
pelos ensinamentos de justiça e honestidade.*

AGRADECIMENTOS

- ★ A Deus, que é meu refúgio e fortaleza contra as intempéries da vida, ponto de reflexão de meus erros, paz e renovação espiritual para a nova batalha;
- ★ À minha esposa, Karla Danielly Souza Ferreira, estudante de Ciências Econômicas, pelo companherismo, projeto em comum, por segurar sempre na minha mão e acreditar no trabalho, na justiça, em Deus, na Ciência para o bem, e aprendendo com os erros para mais na frente cometê-los em menor probabilidade;
- ★ Ao meu orientador, Prof. Dr. Héilton Ribeiro Tavares, pela oportunidade de aprendizado acadêmico neste mestrado e participação na aplicação da Teoria da Resposta ao Item durante a sua direção de avaliação da educação básica no MEC/INEP e importante direcionador desta dissertação ;
- ★ À Profa Dr. Maria Regina Madruga Tavares, com a qual aprendi, em uma de suas aulas de Inferência Estatística, que diante das situações mais adversas devemos explorar as nossas características mais fortes e não fugir da luta, além de sua rigorosa, mas inquestionável competência didática de ensino.
- ★ À Universidade Federal do Pará;
- ★ À Coordenação do Curso de Pós-Graduação em Matemática e Estatística.;
- ★ Aos meus colegas de curso;
- ★ Finalmente, a todos, que de alguma forma contribuíram para a realização deste trabalho.

RESUMO

SOUZA, Ailton Corecha de. Estimação de Proficiências com itens Pré-ponderados em Modelos de Resposta ao Item. Dissertação de Mestrado - (Mestrado em Matemática e Estatística - UFPA, Belém - PA, Brasil).

O objetivo deste trabalho é propor um Modelo de Resposta ao Item capaz de avaliar o tratamento diferenciado em relação aos itens que compõem um teste. Para tanto, considera-se que esses itens recebam uma pré-ponderação (grau de importância) segundo algum critério técnico que seja especificado pelo pesquisador, na área de interesse, para a estimação das habilidades dos indivíduos. A metodologia proposta baseia-se no modelo de Resposta Gradual de Samejima e através deste é introduzida a variável peso na função de verossimilhança marginal do vetor de respostas dos indivíduos, condicionado às habilidades e aos parâmetros dos itens e, por conveniência, utilizou-se o processo de estimação por Máxima Verossimilhança Marginal, onde as estimativas pontuais são substituídas pela Esperança a Posteriori (EAP). Foi desenvolvido um algoritmo e implementado na linguagem de programação matricial Ox, para a estimação das habilidades. Foram realizadas simulações no sentido de avaliar a qualidade do processo de estimação e, também, o comportamento do efeito-ponderação do modelo com e sem peso para os itens. Uma aplicação prática do modelo proposto foi realizada para as respostas de 65 (sessenta e cinco) propriedades rurais, subdivididas em 36 (trinta e seis) da região Sul e 29 (vinte e nove) do Norte, referentes ao índice de Competitividade Agropecuária (habilidade) gerados com base na Teoria da Resposta ao Item, onde aplicou-se um questionário com 31 itens relativos a 4 (quatro) fatores, subdivididos em 10 (dez) subfatores (itens) de Tecnologia, 10 (dez) subfatores de Gestão, 4 (quatro) subfatores de Relações de Mercado e 7 (sete) subfatores de Ambiente Institucional. Para cada conjunto de itens referente a um determinado fator, atribuiu-se pesos diferenciados pelo pesquisador da área de interesse segundo critérios específicos e, a partir daí, criou-se a escala do Índice de Competitividade Agropecuário para alocar as propriedades rurais do Sul e do Norte conforme os seus desempenhos. Verificou-se que, a análise dos resultados para a criação deste índice forneceu conclusões bastante coerentes em relação aos desempenhos dessas propriedades rurais.

Palavras Chave: TRI, MRG, MVM, habilidades, Peso, Itens.

ABSTRACT

The aim of this paper is to propose an item response model capable of evaluating the different treatment in relation to items that constitute a test. For this, it is considered that these items receive a pre-weighting (level of importance) according to some technical criteria that is specified by the researcher in the area of interest for the estimation of the abilities of individuals. The proposed methodology is based on the model of Samejima Graded Response and through this variable is introduced to weight the marginal likelihood function of the vector of responses from individuals, conditional on the abilities and the parameters of the items and, for convenience, we used the process of Marginal Maximum Likelihood estimation, where the point estimates are replaced by the Expectance a Posteriori (EAP). An algorithm was developed and implemented in the Ox matrix programming language for the estimation of abilities. Simulations were conducted in order to evaluate the quality of the estimation process and also the behavior of end-weighting of the model with and without weight for items. A practical application of the proposed model was conducted for responses of 65 (sixty five) Rural Properties are subdivided into 36 (thirty six) in the South and 29 (twenty nine) of the North, for the Agricultural Competitiveness Index (ability) generated based on Item Response Theory, where we applied a questionnaire with 31 items related to 4 (four) factors, subdivided into 10 (ten) sub-factors (items) of Technology, 10 (ten) subfactors Management, 4 (four) sub-factors of Market Relations and 7 (seven) sub-factors of the Institutional Environment. For each set of items for a given factor, was assigned different weights by the researcher's area of interest according to specific criteria and, thereafter, it created the scale of the Agricultural Competitiveness Index to allocate farms in southern and North according to their performances. It was found that the analysis of the results for the creation of this index has provided fairly consistent findings regarding the performance of these Rural Properties.

Words Keys: TRI, MRG, MVM, Abilities, Weight, Itens.

SUMÁRIO

RESUMO	vi
ABSTRACT	vii
1 Introdução	2
1.1 Introdução	2
2 Aspectos Gerais	5
2.1 Aspectos Gerais	5
2.2 Justificativa e Importância do Trabalho	6
2.3 A Hipótese Básica do Trabalho	6
2.4 Objetivos	6
2.4.1 Objetivo Geral	7
2.4.2 Objetivos Específicos	7
2.5 Estrutura do Trabalho	8
3 Os Modelos logísticos da TRI	9
3.1 Introdução	9
3.2 Modelos envolvendo um único grupo	9
3.2.1 Modelos para itens dicotômicos ou dicotomizados	10
3.3 Modelos envolvendo dois ou mais grupos de respondentes	14
3.4 Modelos para itens não-dicotômicos	15
4 O modelo de Resposta Gradual	17
4.1 Introdução	17
4.2 Os fundamentos do Modelo de Resposta Gradual	17
5 Os fundamentos da estimação por Máxima Verossimilhança Marginal	21
5.1 Introdução	21
5.2 Estimação dos parâmetros dos itens por Máxima Verossimilhança Marginal .	21
6 Estimação dos parâmetros dos itens	24
6.1 O Modelo de Resposta Gradual do item	24
6.2 Estimação dos parâmetros dos itens	24
7 Estimação das habilidades considerando pesos γ_i para os itens	27
7.1 Estimação das habilidades por Esperança à Posteriori	27

8	Resultados e conclusões do processo de estimação das Proficiências	29
8.1	Simulação do processo de estimação das proficiências com itens pré-ponderados	29
8.2	Conclusões sobre o processo de estimação das proficiências	31
9	Aplicação do Modelo de Resposta ao Item com pesos diferentes	32
9.1	Índice de Competitividade Agropecuário de propriedades rurais da região Sul e do Norte	32
9.2	Cálculo do Índice de Competitividade Agropecuário	32
9.3	Desempenho das propriedades rurais do Sul e do Norte	33
A	Algoritmo da simulação das estimativas das Proficiências	35
	BIBLIOGRAFIA	39

Capítulo 1

Introdução

1.1 Introdução

De um modo geral, a teoria da Resposta ao Item (TRI) representa um conjunto de modelos estatísticos utilizados por pesquisadores de diversas áreas (psicometria, marketing, avaliação educacional, etc.) para avaliar o comportamento das respostas de indivíduos a um conjunto de itens com duas ou mais categorias. Embora os modelos mais conhecidos da TRI tenham sido aplicados e apresentados com mais frequência em trabalhos da área de psicologia, diversas aplicações desses modelos têm fornecido resultados bastante importantes e significativos em outros campos da ciência. Neste trabalho, sempre com o objetivo de exemplificar o estudo, será dado ênfase à avaliação educacional.

Os modelos logísticos da TRI representam a probabilidade de um certo indivíduo fornecer uma resposta a um item como função dos parâmetros dos itens, fixada a sua habilidade, onde esses itens ou são todos *dicotômicos* ou são todos *politômicos*, mas não leva-se em consideração algum tratamento diferenciado dos itens que, geralmente, pode ser estabelecido por algum critério pedagógico (no contexto da avaliação educacional). No entanto, quando surge a necessidade de se avaliar a habilidade do indivíduo em uma determinada área do conhecimento, onde um certo conjunto de itens é respondido de modo objetivo (certo ou errado) e o outro conjunto de itens representa a partição de uma prova discursiva em várias competências de uma matriz curricular, sendo que cada uma dessas competências pode ser desmembrada em mais de duas categorias ordenadas em níveis crescentes de contemplação, o Modelo de Resposta Gradual (MRG) torna-se o mais adequado para esta situação. Neste caso, pode-se utilizar a combinação do Modelo Logístico de 2 Parâmetros (ML2) e o MRG, sendo o primeiro utilizado para os itens dicotômicos e, o segundo; para os itens corrigidos em mais de duas categorias (politômicos) que representam os níveis de escore gradual. Além disso, os dois conjuntos de itens podem

receber uma ponderação diferenciada caracterizada pela introdução da variável *peso* no modelo.

A principal suposição para a estimação simultânea dos parâmetros dos dois conjuntos de itens e, posteriormente, as habilidades dos indivíduos, é que ambos estejam relacionados ao mesmo construto unidimensional, visto que o ML2 pode ser considerado como caso particular do MRG com duas categorias de respostas.

Para melhor exemplificação, considera-se que este constructo unidimensional seja representado por itens objetivos de uma *prova de Língua Portuguesa* e mais uma *redação*. Portanto, podem-se estimar as habilidades (proficiência, no contexto da avaliação educacional) dos indivíduos a partir de itens corrigidos como certo ou errado (duas categorias de respostas) e as competências, abordadas na redação, corrigidas em níveis graduais de acerto, por exemplo, com 0.00, 0.25, 0.50, 0.75 e 1.00 (cinco categorias de respostas). Além disso, considera-se que os dois conjuntos de itens sejam caracterizados por graus de importância diferenciados segundo algum critério de peso, estabelecido pedagogicamente no processo de avaliação, para a construção da prova como um todo (objetiva mais a redação). E desta forma, surge a suposição inicial em relação ao modelo proposto, ou seja, a de que o conjunto de itens com maior grau de importância (maiores pesos) proporcionam impactos mais significativos na proficiência dos indivíduos.

O modelo proposto para a estimação das habilidades diferencia-se, basicamente, em relação aos demais modelos logísticos tradicionais, pela introdução da variável peso na função de verossimilhança do vetor de respostas dos indivíduos condicionados à habilidade e o vetor de parâmetros dos itens e, a partir dele, serão desenvolvidas as equações de estimação das habilidades dos indivíduos pelo método de Máxima Verossimilhança Marginal desenvolvido por Bock e Lieberman (1970). Obviamente, seguindo o mesmo raciocínio dos modelos tradicionais, tais equações não possuem soluções explícitas e por isso faz-se necessário a utilização de algum processo iterativo, como a de pontos de quadratura para obtenção das estimativas, que será explanado no desenvolvimento do trabalho.

A avaliação do impacto e da qualidade no processo de estimação dessas habilidades é de fundamental importância para o estudo, pois se espera que os resultados oriundos do processo iterativo, forneça estimativas confiáveis para as habilidades quando considera-se pesos (graus de importância) diferentes para os dois conjuntos de itens aplicados em um teste, garantido-se, desta forma, a convergência das estimativas por algum critério

de parada previamente estabelecido no algoritmo a ser desenvolvido em linguagem de programação matricial Ox.

Capítulo 2

Aspectos Gerais

2.1 Aspectos Gerais

Uma das etapas mais importantes da TRI é a estimação dos parâmetros dos itens e das habilidades dos respondentes. Conforme comentado na introdução deste trabalho, a probabilidade de uma resposta a um determinado item ou categoria deste item depende somente da habilidade do indivíduo e dos parâmetros que caracterizam o item. Mas em geral ambos são desconhecidos e geralmente, se tem apenas as respostas dos indivíduos aos itens do teste. Desta forma, nos modelos de resposta ao item, constata-se um problema de estimação que envolve dois tipos de parâmetros, os dos itens e as habilidades dos indivíduos. Por isso, do ponto de vista teórico, depara-se com três situações: primeira, quando os parâmetros dos itens são conhecidos e tem-se que estimar somente as habilidades; segunda, as habilidades já são conhecidas e deseja-se estimar os parâmetros dos itens e; terceira que é a mais comum, a estimação simultânea dos parâmetros dos itens e das habilidades dos respondentes de um teste. Para as três situações citadas acima a estimação pode ser realizada pelo método da Máxima Verossimilhança Marginal a partir da aplicação de algum processo iterativo, que no caso deste trabalho, será adotada a abordagem de *pontos de quadratura* (ver Baker, 1992) juntamente com o método de Newton-Raphson (ver Issac e Keller, 1966). Na situação em que desejamos estimar tanto os parâmetros dos itens, quanto as habilidades, há duas abordagens frequentemente utilizadas: estimação conjunta dos parâmetros dos itens e habilidades, ou em duas etapas, primeiro a estimação dos parâmetros dos itens e, posteriormente, das habilidades. No caso da estimação conjunta, o número de parâmetros envolvidos no processo de estimação pode ser extremamente grande, o que provoca enorme exigência computacional na inversão de matrizes dessa ordem. Vários pesquisadores atuantes na área da Teoria da Resposta ao Item propuseram diversos processos computacionais e métodos de estimação para solucionar tais problemas citados acima, mas neste trabalho, será dada maior ênfase à terceira

situação e estimação pelo método de Máxima Verossimilhança Marginal, que estabelece uma distribuição normal com média μ e variância σ^2 para os parâmetros populacionais dos grupos de indivíduos alocados em cada um dos padrões de respostas que caracterizam realizações de uma variável aleatória, e desta forma, calcula-se a integral em relação à habilidade. Esta última situação, citada anteriormente, será a base deste trabalho para avaliar o comportamento do modelo logístico de Resposta Gradual a partir da introdução da variável peso em sua expressão matemática.

2.2 Justificativa e Importância do Trabalho

A importância do trabalho está exatamente na proposta de um modelo que apresente condições de avaliar a habilidade dos indivíduos respondentes de uma prova como um único constructo unidimensional, composta inicialmente por um conjunto de itens corrigidos de maneira objetiva em duas categorias (certo ou errado) e; mais uma prova discursiva (redação, por exemplo) dividida em várias competências, onde cada competência pode ser desmembrada em mais de duas categorias de respostas, corrigidas em níveis crescentes de contemplação desta competência. Desta forma, o modelo proposto leva em consideração que os dois conjuntos de itens se diferenciam em grau de importância a partir da atribuição de pesos segundo algum critério técnico previamente estabelecido.

2.3 A Hipótese Básica do Trabalho

Parte-se da hipótese de que o conjunto de itens com maior grau de importância, caracterizado pela atribuição de peso no modelo proposto, proporcione impactos mais significativos em relação à habilidade (proficiência) dos indivíduos que respondem a essa prova como um todo, ou seja, objetiva mais discursiva.

2.4 Objetivos

Os objetivos deste trabalho subdividem-se em gerais, representados pela proposta de um modelo que seja capaz de avaliar as estimativas dos parâmetros dos itens e das habilidades de um conjunto de itens respondidos em duas categorias (certo ou errado) e outro conjunto, caracterizado por respostas classificadas em níveis crescentes de acerto com mais de duas categorias. Já os objetivos específicos, representam todas as etapas

subsequentes para a obtenção das equações de estimação por Máxima Verossimilhança Marginal e estimativas dos parâmetros dos itens e das habilidades por meio de métodos iterativos, já que as equações não possuem soluções explícitas.

2.4.1 Objetivo Geral

O objetivo geral, deste trabalho, é apresentar a proposta de introdução da variável peso no Modelo de Resposta Gradual da TRI, em sua expressão matemática, para determinar as equações de estimação dos parâmetros dos itens e a função de distribuição das habilidades dos respondentes, pelo método de Máxima Verossimilhança Marginal. Para tanto, considera-se a situação em que não se conhece nem os parâmetros dos itens e nem as habilidades dos indivíduos e, por isso, fixam-se as habilidades a partir de uma distribuição latente da população para construir a distribuição marginal integrando com relação à habilidade.

2.4.2 Objetivos Específicos

Os objetivos específicos, deste trabalho, organizam-se em:

- Obtenção das equações de estimação por Máxima Verossimilhança Marginal para os parâmetros dos itens do modelo de Resposta gradual;
- Obtenção da distribuição por Máxima Verossimilhança Marginal para as estimativas das habilidades dos indivíduos avaliados em um teste considerando pesos γ_i para os itens;
- Utilização da Esperança à Posteriori para a construção das estimativas médias individuais das habilidades dos indivíduos;
- verificação da qualidade no processo de estimação das habilidades dos indivíduos a partir da diferença em relação aos verdadeiros valores das habilidades, estipulados no início do processo de simulação para a geração da matriz de respostas às categorias dos itens. A simulação desses resultados são gerados em linguagem de programação matricial Ox;
- Avaliação do impacto nas habilidades dos indivíduos devido à atribuição de pesos diferentes para os dois conjuntos de itens em um teste como um todo, considerando-se a avaliação de um único constructo unidimensional.

2.5 Estrutura do Trabalho

Este trabalho encontra-se dividido em sete capítulos, a saber:

- Capítulo 1: Refere-se à introdução do trabalho, onde faz-se uma abordagem das áreas de aplicação da TRI, apresentação do modelo proposto e o método de estimação dos parâmetros dos itens e a distribuição das habilidades;
- Capítulo 2: Retrata os aspectos gerais deste trabalho subdividindo-se em introdução, justificativa e importância do trabalho, a hipótese básica do trabalho, objetivos geral e específico, e estrutura do trabalho;
- Capítulo 3: Faz uma descrição e revisão de literatura dos principais modelos logísticos da TRI e também do modelo de ogiva normal que foi um dos primeiros modelos para representar a curva característica do item;
- Capítulo 4: Descreve os fundamentos básicos do Modelo de Resposta Gradual;
- Capítulo 5: Apresenta uma idéia geral do procedimento de estimação pelo método de Máxima Verossimilhança Marginal;
- Capítulo 6: Estimação dos parâmetros dos itens do Modelo de Resposta Gradual para designar os dois conjuntos de itens que sofrem uma pré-ponderação, ou seja, grau de importância diferenciado;
- Capítulo 7: Refere-se ao processo de estimação das habilidades considerando os pesos γ_i para os itens aplicados em um teste.
- Capítulo 8: Finaliza o trabalho com a apresentação dos Resultados e Conclusões do processo de estimação das habilidades (Proficiências) considerando pesos γ_i para os itens de um teste;

Capítulo 3

Os Modelos logísticos da TRI

3.1 Introdução

A TRI é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma certa resposta a um item como função dos parâmetros do item e da proficiência (ou proficiências) do respondente. Essa relação é sempre expressa de tal forma que quanto maior a proficiência, maior a probabilidade de acerto do item. Os vários modelos propostos na literatura dependem fundamentalmente de três fatores:

1. da natureza do item - dicotômicos ou não dicotômicos;
2. do número de populações envolvidas - apenas uma ou mais de uma;
3. e da quantidade de traços latentes que está sendo medida - apenas um ou mais de um.

Nesse capítulo estaremos somente considerando modelos que avaliam apenas um traço latente ou proficiência, os chamados modelos unidimensionais. Modelos que consideram que mais de uma proficiência está sendo medida, os chamados modelos multidimensionais não serão contemplados neste trabalho.

Mais adiante mostraremos os modelos unidimensionais mais utilizados para um único grupo. Os modelos para dois ou mais grupos serão abordados na Seção 3.3.

3.2 Modelos envolvendo um único grupo

Em Andrade, Tavares & Valle (2000) pode-se ver a definição dos conceitos de grupo e população, que serão largamente utilizados neste e nos demais capítulos, da seguinte forma: Quando o termo *grupo*, for usado, se estará dando referência a uma amostra de indivíduos de uma *população*. O conceito de grupo está diretamente ligado ao processo

de amostragem — e estaremos sempre considerando o processo de amostragem aleatória simples. Portanto, quando falarmos em um único grupo de respondentes, nos referimos a uma amostra de indivíduos retirada de uma mesma população. Conseqüentemente, dois grupos – ou mais – de respondentes são dois conjuntos distintos de indivíduos, que foram amostrados de duas – ou mais – populações.

Na área de Avaliação Educacional é comum que uma população seja definida por determinadas características que podem variar, dependendo dos objetivos do estudo, e portanto, podem ou não ser relevantes para a diferenciação de populações. Por exemplo, pode-se considerar que a 3ª série do Ensino Médio do Brasil é a população alvo. Daí, toma-se uma única amostra dos alunos dessa população, composta de alunos do período diurno e do noturno. Nesse caso, temos então um único grupo de respondentes. Já em outro estudo, poderíamos considerar a 3ª série *diurna* e a 3ª série *noturna* do Ensino Médio do Brasil como duas populações de interesse. Então, seriam tomadas duas amostras: uma dos alunos do período diurno e outra dos alunos do noturno. Nessa situação, teríamos dois grupos de alunos. Portanto, é pelo próprio processo de amostragem do estudo que identifica-se quantas (e quais) populações estão envolvidas.

A seguir, serão mostrados os modelos mais utilizados quando um teste é aplicado a um único grupo de respondentes.

3.2.1 Modelos para itens dicotômicos ou dicotomizados

Atualmente para analisar testes de múltipla escolha, dicotomizados, (corrigidos como certo ou errado) os modelos logísticos para itens dicotômicos são os modelos de resposta ao item mais utilizados, sendo que há basicamente três tipos, que se diferenciam pelo número de parâmetros que utilizam para descrever o item. Eles são conhecidos como os modelos logísticos de 1, 2 e 3 parâmetros, que consideram, respectivamente somente a dificuldade do item, a dificuldade e a discriminação, a dificuldade, a discriminação e a probabilidade de resposta correta dada por indivíduos de baixa proficiência. Porém, os primeiros modelos desenvolvidos foram na forma de uma função distribuição da normal.

Nesta seção, será apresentado o modelo ogiva normal e o logístico de 3 parâmetros, na qual será dado ênfase ao segundo.

O modelo de ogiva normal

Uma vez que a distribuição normal representa a base da teoria estatística, não é surpresa que a mesma tenha sido utilizada como um modelo para representar a curva característica do item. Em geral, as proporções observadas das respostas corretas, ao longo da escala da habilidade, podem ser modeladas por uma ogiva normal. Além do mais, em conjuntos característicos de dados de resposta ao item, a ogiva normal tem-se demonstrado como um modelo viável, de forma que, a proporção de respostas corretas no item i para um determinado nível na escala da habilidade θ pode ser representada por

$$P_i(\theta) \equiv P(\mu_i, \sigma_i, \theta) = \Phi(Z_i) = \int_{-Z_i = -(\theta - \mu_i)/\sigma_i}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-Z^2/2} dz \quad (3.1)$$

onde μ_i é a média, σ_i é o desvio padrão, θ é a habilidade, e Z_i é o desvio normal.

Agora, considerando β_i o parâmetro de locação, medido na mesma escala de θ que indica o ponto na escala da habilidade onde a probabilidade de resposta correta é 0.50, ou seja, quando $\beta_i = \mu_i$ com $-\infty \leq \beta_i \leq \infty$. Além disso, α_i o parâmetro de escala que indica o poder de discriminação de um item, e a seguinte relação inversa $\alpha_i = 1/\sigma_i$ é constatada com $-\infty \leq \alpha_i \leq \infty$.

Portanto, a curva característica do item modelada por uma ogiva normal pode ser definida pela função $P_i(\theta) = P(\beta_i, \alpha_i, \theta)$, fornecendo

$$P(\beta_i, \alpha_i, \theta) = \int_{-Z_i = -\alpha_i(\theta - \beta_i)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{[-\alpha_i(\theta - \beta_i)]^2/2} dz \quad (3.2)$$

O modelo logístico de 3 parâmetros (ML3)

Definição

Dos modelos propostos pela TRI, o *modelo logístico unidimensional de 3 parâmetros (ML3)* é atualmente o mais utilizado e é dado por:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-D\alpha_i(\theta_j - b_i)}}, \quad (3.3)$$

com $i = 1, 2, \dots, I$, e $j = 1, 2, \dots, n$, onde:

U_{ij} é uma variável dicotômica que assume os valores 1, quando o indivíduo j responde corretamente o item i , ou 0 quando o indivíduo j não responde corretamente ao item i .

θ_j representa a proficiência (traço latente) do j -ésimo indivíduo.

$P(U_{ij} = 1|\theta_j)$ é a probabilidade de um indivíduo j com proficiência θ_j responder corretamente o item i e é chamada de Função de Resposta do Item – FRI.

b_i é o parâmetro de dificuldade (ou de posição) do item i , medido na mesma escala da proficiência.

a_i é o parâmetro de discriminação (ou de inclinação) do item i , com valor proporcional à inclinação da Curva Característica do Item — CCI no ponto b_i .

c_i é o parâmetro do item que representa a probabilidade de indivíduos com baixa proficiência responderem corretamente o item i (muitas vezes referido como a probabilidade de acerto casual).

D é um fator de escala, constante e igual a 1. Utiliza-se o valor 1,7 quando deseja-se que a função logística forneça resultados semelhantes ao da função ogiva normal.

Interpretação e representação gráfica

Note que $P(U_{ij} = 1|\theta_j)$ pode ser vista como a proporção de respostas corretas ao item i dentre todos os indivíduos da população com proficiência θ_j . A relação existente entre $P(U_{ij} = 1|\theta_j)$ e os parâmetros do modelo é mostrada na Figura 3.1, que é chamada de Curva Característica do Item – CCI.

O modelo proposto baseia-se no fato de que indivíduos com maior proficiência possuem maior probabilidade de acertar o item e que esta relação não é linear. De fato, pode-se perceber a partir do gráfico acima que a CCI tem forma de “S” com inclinação e deslocamento na escala de proficiência definidos pelos parâmetros do item.

A escala da proficiência é uma escala *arbitrária* onde o importante são as relações de ordem existentes entre seus pontos e não necessariamente sua magnitude. O parâmetro b é medido na mesma unidade da proficiência e o parâmetro c não depende da escala, pois trata-se de uma probabilidade, e como tal, assume sempre valores entre 0 e 1.

Na realidade, o parâmetro b representa a proficiência necessária para uma probabilidade de acerto igual a $(1 + c)/2$. Assim, quanto maior o valor de b , mais difícil é o item, e vice-versa.

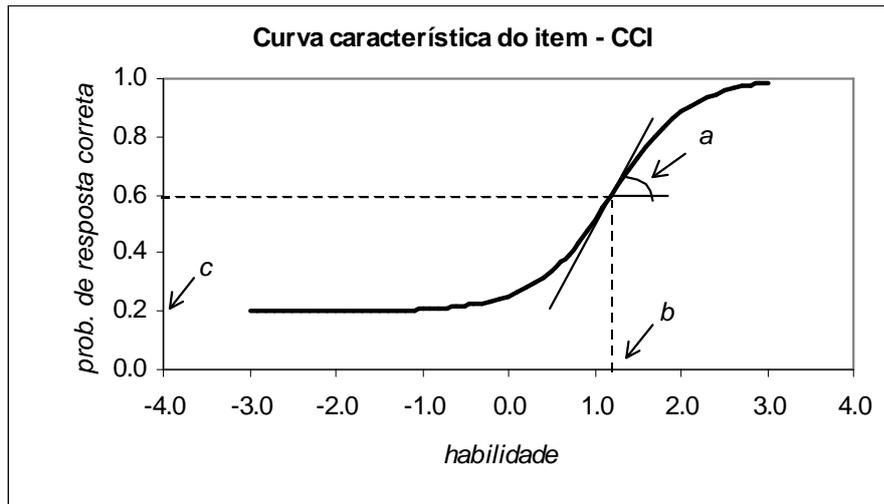


Figura 3.1 Exemplo de uma Curva Característica do Item - CCI

O parâmetro c representa a probabilidade de um aluno com baixa proficiência responder corretamente o item e é muitas vezes referido como a probabilidade de acerto ao acaso. Então, quando não é permitido “chutar”, c é igual a 0 e b representa o ponto na escala da proficiência onde a probabilidade de acertar o item é 0,5.

O parâmetro a é proporcional à inclinação da curva no ponto de inflexão. Assim, itens com a negativo não são esperados sob esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da proficiência. Baixos valores de a indicam que o item tem pouco poder de discriminação (alunos com proficiências bastante diferentes têm aproximadamente a mesma probabilidade de responder corretamente ao item) e valores muito altos indicam itens com curvas características muito “íngremes”, que discriminam os alunos basicamente em dois grupos: os que possuem proficiências abaixo do valor do parâmetro b e os que possuem proficiências acima do valor do parâmetro b Andrade, Tavares & Vale (2000).

3.3 Modelos envolvendo dois ou mais grupos de respondentes

Alguns modelos já foram desenvolvidos para serem aplicados quando um teste envolve mais de uma população, sendo basicamente, extensões dos modelos até aqui discutidos. No entanto, um dos poucos modelos que já se encontram implementados computacionalmente e que portanto, já estão sendo utilizados na prática, quando um teste é aplicado a dois ou mais grupos de respondentes, é uma generalização dos modelos logísticos unidimensionais de 1, 2 e 3 parâmetros, que foi recentemente proposta por Bock & Zimowski (1997). O modelo é dado por:

$$P(U_{ijk} = 1|\theta_{jk}) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_{jk} - b_i)}}, \quad (3.4)$$

com $i = 1, 2, \dots, I$, $j = 1, 2, \dots, n_k$, e $k = 1, 2, \dots, K$, onde:

U_{ijk} é uma variável dicotômica que assume os valores 1, quando o indivíduo j da população k responde corretamente ao item i , ou 0 quando o indivíduo não responde corretamente ao item.

θ_{jk} representa a proficiência do j -ésimo indivíduo da população k .

$P(U_{ijk} = 1|\theta_{jk})$ é a probabilidade de um indivíduo j da população k , com proficiência θ_{jk} , responder corretamente ao item i .

Os demais parâmetros já foram descritos anteriormente.

Para que seja possível a comparação entre grupos de respondentes, é sugerido que haja pelo menos vinte por cento de itens comuns entre os grupos. Assim, I representa o número total de itens *distintos* apresentados.

Andrade, Tavares & Vale (2000) revelam que a recente implementação computacional desse modelo para mais de um grupo de respondentes foi um dos maiores avanços da TRI nos últimos anos. Através dele a comparação de indivíduos de grupos distintos, submetidos a provas diferentes mas com itens comuns, passou a ser feita de uma maneira bem mais eficiente do que era feito até então, uma vez que diminui possíveis erros de modelagem que a metodologia anterior poderia vir a ter. A seguir faremos uma abordagem sobre alguns métodos de estimação dos parâmetros dos modelos para um ou mais grupos.

Maiores detalhes sobre estimação de parâmetros dos itens e das proficiências, podem ser encontrados em Andrade, Tavares & Vale (2000) e Azevedo (2003).

3.4 Modelos para itens não-dicotômicos

Os modelos que servem tanto para análise de itens abertos (de resposta livre) quanto para a análise de itens de múltipla escolha que são avaliados de forma graduada, ou seja, itens que são elaborados ou corrigidos de modo a ter-se uma ou mais categorias intermediárias ordenadas entre as categorias certo e errado. Nesse tipo de item não se considera somente se o indivíduo respondeu à alternativa correta ou não, mas também leva-se em conta qual foi a resposta dada por ele.

Dentre os modelos para itens não-dicotômicos, destacam-se:

1. o de Resposta Nominal, desenvolvido por Bock (1972) com base no modelo logístico de dois parâmetros que pode ser aplicado a todas as categorias de resposta escolhidas em um teste com itens de múltipla escolha. O propósito deste modelo de Resposta Nominal foi maximizar a precisão da habilidade estimada utilizando toda a informação contida nas respostas dos indivíduos, e não apenas se o item foi respondido corretamente ou não;
2. o modelo de Resposta Gradual de Samejima (1969), que será utilizado no modelo proposto deste trabalho, assume que as categorias de respostas de um item podem ser ordenadas entre si. Este modelo, semelhante ao de Bock, tenta obter mais informação das respostas dos indivíduos do que simplesmente se eles forneceram respostas corretas ou incorretas. Na verdade, tem-se necessariamente uma ordenação entre o nível de dificuldade das categorias de respostas de um dado item, de acordo com a classificação de seus escores;
3. o modelo de Escala Gradual, que é um caso particular do modelo de Resposta Gradual de Samejima, ou seja, este modelo também é adequado para itens com categorias de respostas ordenadas. No entanto, faz-se uma suposição a mais: a de que escores das categorias são igualmente espaçados;
4. e por fim, o modelo de Crédito Parcial, desenvolvido por Masters (1982), que também é um modelo para a análise de respostas obtidas de duas ou mais categorias ordenadas. O modelo de Crédito Parcial difere do Gradual, entretanto, por pertencer à

família de modelos de Rasch, ou seja, todos os parâmetros no modelo são de locação, sendo que o poder de discriminação é assumido ser comum para todos os itens.

As expressões matemáticas podem ser obtidas com mais detalhes em (Andrade, Tavares e Valle, 2000). Neste Trabalho daremos mais ênfase aos critérios matemáticos para o modelo de Resposta Gradual de Samejima por ser a base do modelo, aqui proposto, e abordado no capítulo 4.

Capítulo 4

O modelo de Resposta Gradual

4.1 Introdução

Na maioria dos testes, principalmente em avaliação educacional, os itens são escoreados dicotomicamente, ou seja, se os mesmos foram respondidos de maneira correta ou incorreta designados por duas categorias (certo ou errado), entretanto, outro procedimento de escore do item, tal como o caso de resposta gradual, pode ser empregado. Sob o procedimento de escore gradual, a escala da variável item pode ser dividida dentro de categorias ordenadas. Dessa forma, a categoria mais baixa contribuirá muito pouco para um escore do teste, enquanto a categoria mais alta terá maior contribuição. Itens que compõem uma escala tipo Likert tradicionalmente utilizados em questionários, atitudes espontâneas e inspeções são exemplos clássicos cujas respostas são escoreadas em uma forma gradual. Em tais itens, uma escala é utilizada, e a classificação fornece a ordem tal como, discorda plenamente, discorda, indiferente, concorda e concorda plenamente. Uma pontuação igual a 1 será assumido para discorda plenamente, 2 para discorda, 3 para indiferente e assim por diante até concorda plenamente assumindo uma pontuação igual a 5.

4.2 Os fundamentos do Modelo de Resposta Gradual

O fundamento lógico da modelagem do caso de resposta gradual será introduzido como uma extensão da justificativa de Lord da ogiva normal para itens escoreados dicotomicamente. Quando uma resposta livre de um indivíduo para um item é escoreado em uma base gradual, assume-se que uma variável item contínua hipotética fundamenta a resposta. Este item contínuo varia de $-\infty$ a $+\infty$ e subdivide-se em m categorias de resposta para um dado item. Por causa da fundamentação de continuidade, as categorias de respostas resultantes são ordenadas, com k denotando uma categoria arbitrária, $k = 1, 2, \dots, m_i$, onde m_i

é o número de categorias de resposta para o item i . É importante notar que, sem perda de generalidade, m será usado para m_i .

De um ponto de vista teórico, os indivíduos fornecem uma resposta livre para o item. A resposta é então inspecionada e alocada no intervalo de respostas incorporadas por uma categoria de resposta particular. Assim, a j -ésima resposta do indivíduo pode ser designada apenas por uma das m categorias de resposta do item, e $u_{jki} = 1$ se a resposta é designada para a categoria k do item i , e zero caso contrário. Se houver o interesse em obter o escore total do teste composto de tais itens, a categoria $k = m$ terá o maior peso associado a este item e, portanto, é a categoria mais alta. A categoria mais baixa denotada por $k = 1$ terá o menor peso, de forma que o interesse no escore total, é solucionado pela obtenção da soma dos pesos associados com as categorias de respostas do item fornecidas pelos indivíduos para cada um dos n itens na prova. O caso de resposta gradual pode ser modelado utilizando um raciocínio muito similar àquele utilizado anteriormente para o caso de respostas dicotômicas. Os valores da variável item que demarcam as categorias de resposta são denotados por γ_k , onde $\gamma_0 = -\infty$ e $\gamma_m = +\infty$. A regressão da variável item sobre a escala da habilidade é denotada por $\gamma'|\theta$. Em cada nível de habilidade, existe uma distribuição normal condicional das respostas dos indivíduos. Essas distribuições condicionais são interceptadas pelos limites das categorias de resposta. Assim, a probabilidade de uma particular resposta de um indivíduo pertencer à categoria k é denotada pela área das distribuições condicionais que localizam-se entre os limites γ_{k-1} e γ_k e é denotada por $P_{ik}(\theta)$. É importante observar que, em cada nível de habilidade,

$$\sum_{k=1}^m P_{ik} = 1 \quad (4.1)$$

O termo Curva Característica da Categoria de Resposta do Item (CCCRI) é utilizado para definir a característica operacional que é diferente daquela definida para itens escoreados de maneira dicotômica e assumi-se que a distribuição da variável item e sua distribuição condicional do item em cada θ é uma densidade normal e a única suposição necessária recai sobre a distribuição de frequência dos indivíduos sobre a escala da habilidade. Com base nessas suposições e os valores específicos para os limites da categoria γ_k a CCCRI para as diferentes categorias de resposta do item podem ser obtidas e as devidas probabilidades associadas com cada categoria sobre o intervalo de θ .

A probabilidade de uma resposta ser designada para uma determinada categoria é dada pela diferença entre dois quaisquer limites adjacentes. Se $P_{ik}^*(\theta)$ denota um limite e $P_{ik}(\theta)$ a probabilidade de uma categoria, então

$$\begin{aligned}
 P_{i4}^*(\theta) &= 0 \\
 P_{i3}^*(\theta) &= P_{i4}(\theta) \\
 P_{i2}^*(\theta) &= P_{i3}(\theta) + P_{i4}(\theta) \\
 P_{i1}^*(\theta) &= P_{i2}(\theta) + P_{i3}(\theta) + P_{i4}(\theta) \\
 P_{i0}^*(\theta) &= P_{i1}(\theta) + P_{i2}(\theta) + P_{i3}(\theta) + P_{i4}(\theta) = 1,
 \end{aligned} \tag{4.2}$$

e

$$P_{ik}^*(\theta) = \sum_{k'=k+1}^m P_{ik'}(\theta). \tag{4.3}$$

e em geral,

$$P_{ik}(\theta) = P_{i,k-1}^*(\theta) - P_{ik}^*(\theta), \tag{4.4}$$

que fornece

$$P_{ik}(\theta) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}}, \tag{4.5}$$

onde as funções $P_{ik}^*(\theta)$ representam os limites das probabilidades acumuladas que caracterizam os parâmetros α_i e $\beta_{i,k}$. Uma vez que essas ogivas representam a soma das probabilidades das categorias de respostas, diferenças negativas entre as curvas não são esperadas. Assim, essas curvas limites não podem se cruzar, e elas podem assumir um valor comum do parâmetro de discriminação α_i . Também, para manter a ordem básica das categorias de respostas, os parâmetros de locação das ogivas limites devem ser ordenados:

$$\beta_{i1} < \beta_{i,2} < \dots < \beta_{ik} < \dots < \beta_{i,m-1}, \tag{4.6}$$

onde $-\infty < \beta_{ik} < +\infty$.

Note que um item com m_i categorias, $m_i - 1$ valores de dificuldade necessitam ser estimados, além do parâmetro de discriminação para cada item. Assim, o número de parâmetros a serem estimados é dado por $\sum_{i=1}^I m_i$, onde I é o número de itens no teste.

Capítulo 5

Os fundamentos da estimação por Máxima Verossimilhança Marginal

5.1 Introdução

Os fundamentos básicos deste método devem-se a Bock e Lieberman (1970), que desenvolveram o procedimento de Máxima Verossimilhança Marginal para a estimação dos parâmetros dos itens. Não obstante, a abordagem de Bock e Lieberman promoveu uma agradável tarefa em termos computacionais. Posteriormente, houve uma reformulação deste procedimento, que sob certas condições, representa um caso particular do algoritmo EM.

5.2 Estimação dos parâmetros dos itens por Máxima Verossimilhança Marginal

A idéia fundamental da abordagem de Máxima Verossimilhança Marginal para a estimação dos parâmetros dos itens é que os indivíduos que respondem a uma determinada prova representam uma amostra aleatória de uma população onde a habilidade é distribuída de acordo com uma função densidade $g(\theta|\eta)$, onde η é o vetor de parâmetros da distribuição da habilidade da população de indivíduos. Isto equivale a considerar os itens como efeitos fixos e as habilidades como efeitos aleatórios. A essência da solução de Bock e Lieberman (1970) é integrar com relação à distribuição da habilidade, e desta forma, removendo os parâmetros de ruídos (habilidades) da função de verossimilhança. Portanto, os parâmetros dos itens são estimados por uma distribuição marginal, e essa estimação não depende da estimação da habilidade de cada indivíduo, mas sim, da distribuição da habilidade.

Na abordagem de estimação por Máxima Verossimilhança Marginal, o estimador pontual da habilidade para um indivíduo, $\hat{\theta}_j$, é substituído por uma distribuição que

aloca os dados do indivíduo j sobre a escala da habilidade que equivale à probabilidade de estar em qualquer ponto ao longo da escala da habilidade. Assim, ao invés de um único valor $\hat{\theta}_j$ para um determinado indivíduo, a probabilidade é encontrada de um indivíduo contendo cada valor possível θ_j ao longo da escala da habilidade, condicionada ao vetor de resposta do item U_j , aos parâmetros dos itens ζ_i , e a distribuição populacional da habilidade.

Para exemplificar a aplicação do método de Máxima Verossimilhança Marginal, considere a seqüência de respostas do indivíduo j aos I itens representados pelo seguinte vetor $U_j = [U_{j1}, U_{j2}, \dots, U_{jI}]$ e $U_{..} = [U_{1.}, U_{2.}, \dots, U_{n.}]$ o conjunto integral de respostas. Assumindo a independência condicional das respostas dado θ , a probabilidade da seqüência de respostas U_j . Como um produto das respectivas funções de respostas do item, de modo que

$$P(U_j|\theta, \zeta) = \prod_{i=1}^I P(u_{ji}|\theta, \zeta_i), \quad (5.1)$$

onde ζ_i é o conjunto de parâmetros do item.

Agora, suponha que θ tenha uma distribuição contínua com média finita e variância na população de indivíduos com função densidade de probabilidade $g(\theta|\eta)$. A probabilidade marginal da seqüência de respostas do vetor U_j . pode ser expressa como

$$P(U_j|\zeta, \eta) = \int_{\Theta} P(U_j|\theta, \zeta)g(\theta|\eta)d\theta, \quad (5.2)$$

onde $\eta = (\mu, \sigma^2)$ representa o vetor de parâmetros populacionais da distribuição da habilidade. E portanto, a função de Máxima Verossimilhança Marginal é dada por

$$L(\zeta, \eta) = \prod_{j=1}^n P(U_j|\zeta, \eta). \quad (5.3)$$

Pode-se, desta forma, obter a log-verossimilhança Marginal

$$\log L(\zeta, \eta) = \text{constante} + \sum_{j=1}^n \log P(U_j|\zeta, \eta), \quad (5.4)$$

e portanto, as equações de estimação por máxima verossimilhança dos parâmetros dos itens pode ser encontrada pela expressão

$$\frac{\log L(\zeta, \eta)}{\partial \zeta_i} = \sum_{j=1}^n \frac{1}{P(U_j | \zeta, \eta)} \int_{\Theta} \frac{\partial P(U_j | \theta, \zeta)}{\partial \zeta_i} g(\theta | \eta) d\theta, \quad (5.5)$$

Capítulo 6

Estimação dos parâmetros dos itens

6.1 O Modelo de Resposta Gradual do item

Nesta seção, tratar-se-á da estimação dos parâmetros dos itens pelo método da Máxima Verossimilhança Marginal. Para tanto, os cálculos serão desenvolvidos para o Modelo Resposta Gradual, pois o objetivo deste trabalho é a correção simultânea de dois conjuntos de itens de uma prova, sendo que um deles é corrigido de maneira dicotômica, e outro, desmembrado em várias competências com mais de duas categorias corrigida em nível crescente de acerto, por exemplo, 0.00, 0.25, 0.50, 0.75 e 1.00.

6.2 Estimação dos parâmetros dos itens

Seguindo este raciocínio e, considerando-se o Modelo de Resposta Gradual da TRI e usando a independência local entre as respostas dos indivíduos, podemos escrever a probabilidade associada ao vetor de respostas dos dois conjuntos de itens $U_j = [u_{j1}, u_{j2}, \dots, u_{jI}]$ como

$$P(U_j | \theta, \zeta_i) = \prod_{i=1}^I P(u_{ji} = k | \theta, \zeta_i), \quad (6.1)$$

e a probabilidade marginal associada a esse vetor de respostas pode ser descrita por

$$P(U_j) = \int P(U_j | \theta, \zeta_i) g(\theta | \eta) d\theta \quad (6.2)$$

No entanto, para facilidades e vantagens nos cálculos computacionais, utiliza-se a abordagem de padrões de respostas. Assumindo-se que, ao todo, existem K_i categorias no item i e r_j é o número de ocorrências do padrão de resposta j , ou seja, j não representará o indivíduo, mas sim, um padrão de resposta específico. E ainda, $s \leq \min(n, L)$ é o número de padrões de respostas distintos com $r_j > 0$, onde $L = \prod_{i=1}^I K_i$. Segue disso que

$$\sum_{j=1}^s r_j = n. \quad (6.3)$$

Assumindo a independência das respostas entre os diferentes indivíduos, a verossimilhança marginal pode ser escrita como

$$L(\zeta, \eta) = \frac{n!}{\prod_{j=1}^s r_j} \prod_{j=1}^s [P(U_j)]^{r_j}. \quad (6.4)$$

No objetivo de obter os estimadores de máxima verossimilhança, a log-verossimilhança marginal é maximizada com relação a ζ e η , respectivamente, onde ζ representa os parâmetros dos itens e η o vetor de parâmetros populacionais de cada grupo de indivíduos alocados no padrão de resposta j . Uma vez que cada item é respondido de forma independente, a log-verossimilhança marginal é maximizada com relação a ζ_i . Portanto, a equação de verossimilhança para os parâmetros do item i pode ser expressada como

$$\frac{\partial \text{Log}L(\zeta, \eta)}{\partial \zeta_i} = \sum_{j=1}^s \frac{r_j}{P(U_j)} \frac{\partial P(U_j)}{\partial \zeta_i} = 0 \quad (6.5)$$

Agora,

$$L(\theta) = P(U_j | \theta, \zeta_i) = \prod_{i=1}^I P(u_{ji} = k | \theta, \zeta_i) \quad (6.6)$$

e

$$g(\theta | \eta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2}\right] \quad (6.7)$$

e $u_{jik} = 1$ se $u_{ji} = k$ e $u_{jik} = 0$ caso contrário. Então,

$$\frac{\partial P(U_j)}{\partial \zeta_i} = \int \left(\frac{\partial \{\prod_{k=1}^{K_i} [P(\theta)]^{u_{jik}}\}}{\partial \zeta_i} \right) L_j(\theta) g(\theta | \eta) d\theta \quad (6.8)$$

Mas,

$$\frac{\partial \prod_{k=1}^{K_i} [P(\theta)]^{u_{jik}}}{\partial \zeta_i} = \sum_{k=1}^{K_i} \frac{u_{jik}}{P(\theta)} \frac{\partial P(\theta)}{\partial \zeta_i} \prod_{k=1}^{K_i} [P(\theta)]^{u_{jik}},$$

e portanto, a equação de verossimilhança marginal para os dois conjuntos de itens é dada por

$$\frac{\partial \log L}{\partial \zeta_i} = \sum_{j=1}^s \frac{r_j}{P(U_j)} \int \left(\sum_{k=1}^{K_i} \frac{u_{jik}}{P(\theta)} \frac{\partial P(\theta)}{\zeta_i} \right) L(\theta) g(\theta|\eta) d\theta \quad (6.9)$$

Como $\zeta_i = (a_i, b_i)$, então as equações de verossimillhanca podem ser escritas como

$$\frac{\partial \log L}{\partial a_i} = \sum_{j=1}^s \frac{r_j}{P_j} \int \left(\sum_{k=1}^{K_i} \frac{u_{jik}}{P(\theta)} [W_{i,k-1}^*(\theta) - W_{ik}^*(\theta)] \right) L(\theta) g(\theta|\eta) d\theta = 0, \quad (6.10)$$

e

$$\frac{\partial \log L}{\partial b_i} = \sum_{j=1}^s \frac{r_j}{P_j} \int \left(-\frac{u_{jik}}{P(\theta)} + \frac{u_{ji,k+1}}{P_{i,k+1}(\theta)} \right) W_{ik}^* L(\theta) g(\theta|\eta) d\theta = 0, \quad (6.11)$$

onde $W_{ik}^* = P_{ik}^*(\theta)[1 - P_{ik}^*(\theta)]$

Capítulo 7

Estimação das habilidades considerando pesos γ_i para os itens

7.1 Estimação das habilidades por Esperança à Posteriori

O procedimento de estimação por Máxima Verossimilhança Marginal fornece apenas as estimativas dos parâmetros dos itens e, portanto, a estimação das habilidades considerando pesos γ_i para os itens assume que os parâmetros dos itens são os valores verdadeiros. E para isso, deve-se considerar que a equação (6.6), do capítulo anterior, absorve os pesos γ_i em cada uma das categorias de respostas na função de verossimilhança do modelo de resposta gradual.

Como a sequência de respostas do indivíduo j aos I itens de um teste pode ser representada por um vetor $U_{jk.} = [u_{jk1}, u_{jk2}, \dots, u_{jkI}]$ e assumindo a independência condicional das respostas dada a habilidade θ_j e o vetor de parâmetros ζ_i , a probabilidade da sequência de respostas $U_{jk.}$ pode ser obtida como um produto das respectivas funções de resposta ao item i ponderado por um peso γ_i de forma que a equação pode ser reescrita da seguinte forma

$$L(\theta) = P(U_{jk.}|\theta, \zeta_i) = \prod_{i=1}^I P(u_{ji} = k|\theta, \zeta_i)^{\gamma_i}, \quad (7.1)$$

mas a seguinte dicotomização das categorias

$$u_{jik} = \begin{cases} 1, & \text{se } u_{ji} = k \\ 0, & \text{caso contrário} \end{cases}$$

fornece

$$L(\theta) = P(U_{jk}|\theta, \zeta_i) = \prod_{i=1}^{k_i} P(u_{jki} = 1|\theta, \zeta_i)^{u_{jki}\gamma_i} \quad (7.2)$$

$$(7.3)$$

Agora, utilizando-se a equação (6.2), tem-se que a probabilidade marginal associada ao vetor de respostas U_{jk} é

$$P(U_j) = \int L(\theta)g(\theta_j|\eta)d\theta \quad (7.4)$$

$$P(U_j) = \int \prod_{i=1}^{k_i} P(u_{jki} = 1|\theta, \zeta_i)^{u_{jki}\gamma_i} g(\theta_j|\eta)d\theta \quad (7.5)$$

Desta forma, as habilidades podem ser estimadas por esperança a posteriori (EAP) que corresponde à estimativa da média à posteriori da habilidade para os padrões de repostas j , $\bar{\theta}_j$ (isto é, a média da distribuição a posteriori), pode ser obtida por

$$\bar{\theta}_j = \frac{\int \theta_j L(\theta)g(\theta_j|\eta)d\theta}{P(U_j)} \quad (7.6)$$

e lembrando que $P(U_j) = \int_{\Theta} L(\theta)g(\theta_j|\eta)d\theta$ e então a equação (7.2) pode ser reescrita como

$$\bar{\theta}_j = \frac{\int \theta_j L(\theta)g(\theta_j|\eta)d\theta}{\int L(\theta)g(\theta_j|\eta)d\theta} \quad (7.7)$$

Capítulo 8

Resultados e conclusões do processo de estimação das Proficiências

8.1 Simulação do processo de estimação das proficiências com itens pré-ponderados

Para certificar-se de que o modelo proposto com pesos γ_i para os itens de um teste geram estimativas confiáveis, executou-se a simulação do processo de estimação das proficiências com itens pré-ponderados na linguagem de programação matricial OX, e para maiores detalhes, consultar Doornik (1994).

A idéia principal do algoritmo, que encontra-se disponível no Apêndice (A) deste trabalho, foi a geração de uma amostra aleatória da proficiência de 1.000 (mil) indivíduos supondo-se uma distribuição densidade de probabilidade normal com média 0 (zero) e desvio-padrão 1 (um). De posse dessa amostra, juntamente com os valores verdadeiros (pré-fixados) dos parâmetros dos itens foram construídas 100 (cem) réplicas da matriz com os dados de respostas aos itens, para estimar as proficiências dos indivíduos pelo modelo proposto com itens *pré-ponderados*. Em seguida, obteve-se as estimativas médias de proficiências dos 1.000 (mil) indivíduos com base nas 100 (cem) réplicas.

Segue abaixo o comportamento das estimativas médias por máxima verossimilhança marginal com itens pré-ponderados em relação aos verdadeiros valores das proficiências dos indivíduos gerados pela simulação.

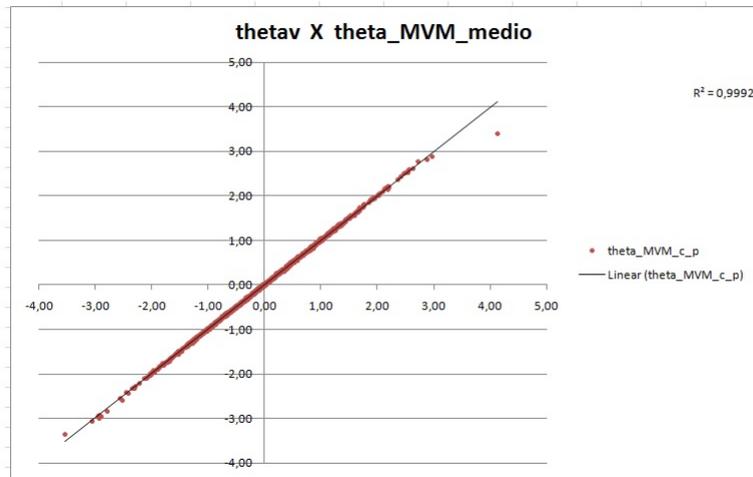


Figura 8.1 *Comportamento das estimativas verdadeiras em relação às estimativas médias*

Também foram geradas por simulação, de maneira análoga, as estimativas das proficiências desses 1.000 (mil) indivíduos sem considerar pesos γ_i para os itens para a análise do efeito ponderação. E a partir da diferença entre as estimativas médias dos itens com e sem ponderação, tem-se o seguinte comportamento gráfico:

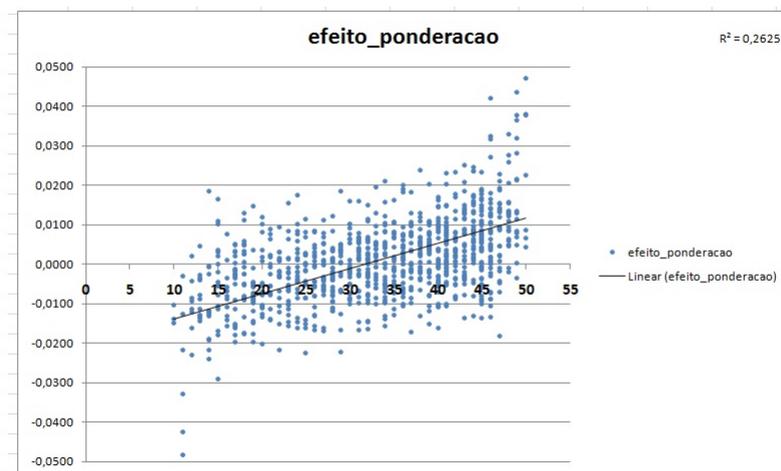


Figura 8.2 *Estudo do efeito ponderação*

8.2 Conclusões sobre o processo de estimação das proficiências

A partir da Figura 1, constata-se que o modelo proposto utilizando-se o processo de estimação por máxima verossimilhança marginal das Proficiências dos indivíduos, com itens que sofrem uma pré-ponderação, fornece estimativas médias das várias etapas da simulação muito próximas das verdadeiras proficiências estabelecidas na fase inicial do algoritmo. Já na Figura 2, pode ser observado que há um impacto considerável do efeito ponderação nas estimativas das proficiências condicionadas aos *vetores de respostas* dos indivíduos e aos verdadeiros valores dos *parâmetros dos itens*. Esse efeito é ainda mais evidente para as extremidades dos escores dos itens pré-ponderados, ou seja, os escores baixos desses itens estão diretamente relacionados à baixas proficiências e escores altos com altas proficiências.

Capítulo 9

Aplicação do Modelo de Resposta ao Item com pesos diferentes

9.1 Índice de Competitividade Agropecuário de propriedades rurais da região Sul e do Norte

A aplicação do Modelo de Resposta ao Item considerando itens pré-ponderados, tem como objetivo principal, a mensuração do estágio atual dos direcionadores e fatores de competitividade da atividade em cada uma das regiões analisadas (Sul e Norte), com ênfase na Tecnologia, Gestão, Relações de Mercado e Ambiente Institucional e, desta forma, criando-se posteriormente um Índice de Competitividade, que na TRI, é considerado como habilidades. Os dados foram cedidos pelo Prof. Msc. Ricardo Pedroso Oia-gen (UFPA/Campus de Castanhal/Faculdade de Medicina veterinária) da sua tese de doutorado intitulada "Competitividade de Sistemas de Produção de Bovinocultura de Corte nas Regiões Sul e Norte do Brasil" pelo Departamento de Zootecnia da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul (UFRGS) sob a orientação do Prof. Dr. Júlio Otávio Jardim Barcellos (UFRGS/Faculdade de Agronomia/Departamento de Zootecnia).

9.2 Cálculo do Índice de Competitividade Agropecuário

O procedimento de cálculo do Índice de Competitividade Agropecuário atende perfeitamente ao propósito do modelo enfatizado neste trabalho, uma vez que foram aplicados questionários com 31 itens; subdivididos em 10 itens de Tecnologia, 10 itens de Gestão, 4 itens de Relações de Mercado e 7 itens de Ambiente Institucional. O direcionador de Tecnologia recebe uma pré-ponderação de 3,5; o de Gestão recebe 3,5; o de Relações de Mercado recebe 2,0 e; por fim, o de Ambiente Institucional recebe 1,0. Desta forma, pode-

se criar o Índice de Competitividade levando em consideração os graus de importância de cada direcionador conforme critérios específicos estabelecidos pelo pesquisador.

Inicialmente, procedeu-se à estimação dos parâmetros dos 31 itens utilizados no cálculo do Índice de Competitividade através do software "Multilog" e, posteriormente, tais parâmetros são tomados fixos para a estimação das habilidades (Índice de Competitividade) no software de linguagem de programação matricial "OX".

9.3 Desempenho das propriedades rurais do Sul e do Norte

Gerou-se os resultados do Índice de Competitividade das 65 propriedades rurais, das quais, 36 são de procedência do Sul e 29, do Norte. Os desempenhos de todas as propriedades rurais são alocados na escala do Índice de Competitividade com média zero e desvio-padrão unitário e, portanto, os resultados individuais e por região podem ser analisados a partir da Figura 9.1.

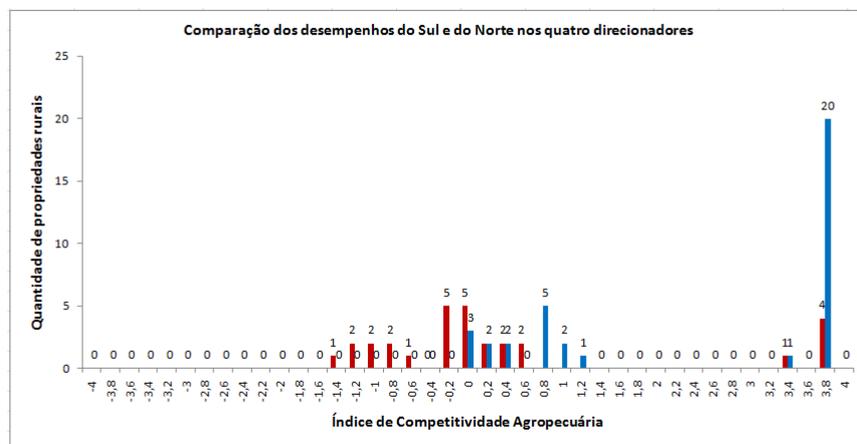


Figura 9.1 *Comparação dos desempenhos das regiões Sul e Norte em relação ao Índice de Competitividade*

Pelo que pode ser analisado, na Figura 9.1, o Modelo de Resposta ao Item, que considera uma pré-ponderação dos itens, consegue captar de forma bastante coerente as disparidades de desempenho entre as propriedades rurais do Sul e do Norte em termos do Índice de Competitividade. O desempenho médio das propriedades rurais da região Sul é ligeiramente superior ao das propriedades rurais do Norte e, ainda, embora algumas

propriedades rurais do Norte obtenham desempenho médio próximo do nível de 3,8 na escala (0,1), o número de propriedades rurais do Sul é muito maior, 4 do Norte contra 20 do Sul, caracterizando uma diferença muito grande.

Apêndice A

Algoritmo da simulação das estimativas das Proficiências

```
#include <oxstd.h> #include <oxfloat.h>
decl p1,p2,p3,p4, D, n,
a, b, av, bv1, bv2, bv3, bv4, I, u,v,
b1, b2, b3, b4,habil, theta, thetav,scorep, UsaPesos, pesos, ge, q,
theta_MVM,j,
media, habil2,dif,delta,etapas;

LerItens() {  D = 1.7; n = 1000;          // tamanho da amosta I =
50;          // número de itens UsaPesos=0; // Define se os escores
e proficiências serão calculados com pesos. av = a =
ranu(1,I)*1.5+0.5; // [0.5 ; 2.0] bv1 = b1 = ranu(1,I)*1.5-2; //
1 x I // print(rows(b1), " x ",columns(b1),"\n"); bv2 = b2 = b1 +
ranu(1,I)*1.5; bv3 = b3 = b2 + ranu(1,I)*1.5;
bv4 = b4 = b3 +
ranu(1,I)*1.5;
// print("Parametros gerados: \n Item  a      b1      b2      b3
b4", "%6.2f",range(1,I)'~a'~b1'~b2'~b3'~b4');

pesos=ones(1,I-10)^(3*ones(1,10));
if (UsaPesos==0)
pesos=ones(1,I);
else print("\nPesos associados aos Itens: ", "%4.0f", pesos);
habil2 = thetav = rann(n,1); } GerarDados() { decl v;
// ranseed[t] = .572552575;
// habil = thetav;          // habilidades dos n individuos
// print("\nHabilidades dos 20 primeiros: ",range(1,20)'~habil[:19]);
p1 = 1 ./ (1+exp(-D*av.*(thetav-bv1))); // matriz com as
probabilidades(ML2)
p2 = 1 ./ (1+exp(-D*av.*(thetav-bv2)));
//println(rows(p2)," x ",columns(p2));
p3 = 1 ./ (1+exp(-D*av.*(thetav-bv3)));
p4 = 1
./ (1+exp(-D*av.*(thetav-bv4)));
// println(habil~p1[] [0]~p2[] [0]~p3[] [0]~p4[] [0]);
// savemat("FR-Nominal.xls",habil~p1[] [0]~p2[] [0]~p3[] [0]~p4[] [0]);
// print("Imprimindo os Pki", "%6.2f", p1~p2~p3~p4);
//savemat("Nominal.xls",p1~p2~p3~p4);
// print("p(k)-p(k-1):", "%6.2f", (p1-p2), "%6.2f", (p2-p3), "%6.2f",
(p3-p4));
v = ranu(n,I);
```

```

u = 5*(v.< p4)+4*(p4.<v .&& v.<p3)+3*(p3.<v .&&
v.<p2)+2*(p2.<v .&& v.<p1)+1*(v.>p1);
print("\nDados (20 primeiros):", "%1.0f", u[:19] []);
savemat("Dados_Nominal2.xls", u);
// savemat("Dados_Nominal.dat", u, 1);
}
GeraEstatisticas() { decl cont=<>, k;
    for (k=1; k<=5; ++k) cont|=sumc(u.==k); //print("%5.0f", cont);
    print("\nProporcao em cada categoria (coluna), de cada item
    (linha)", "\n
    Item    Cat1
    Cat2    Cat3    Cat4    Cat5", "%8.1f", range(1, I) ~ (100*cont'/n));
}
// #####
#####
// ##### ESTIMAÇÃO POR MÁXIMA VEROSSIMILHANÇA MARGINAL #####
#####
// #####
#####
GeraPontosdeQuadratura() { decl max, min, delta;
    min=-4; max=4; delta=.25;
    habil=range(min, max, delta)'; // q x 1 print(habil);
    q=sizerc(habil); //print(rows(habil), " x ", columns(habil));
    println("\nEstao sendo usados ", q, " pontos de quadratura");
}
Monta_ge() { decl g, P, SP, Pj_T, Pj_Tg, Pj, Pj1, Pj2, Pj3, Pj4, Pj5,
up0, up1, P1e, P2e, P3e, P4e;
    SP=zeros(q, I);
    g=1/sqrt(M_2PI) * exp(-0.5*habil.^2);
    Pj_T=ones(q, n);

    up1 = (u.==1).*pesos; up0=(u.!=1).*pesos; // n x I //print(rows(up1)
, " x ", columns(up1), "\n");
    P1e = 1 ./ (1 + exp(-D * a .* (habil**ones(1, I) - b1))); // q x I,
matriz com as probabilidades(ML2)
    P = 1 - P1e;
    Pj_T = Pj_T .* exp(log(P)*up1' + log(1-P)*up0'); // q x n

    up1 = (u.==2).*pesos; up0=(u.!=2).*pesos; //print(rows(up1), " ",
rows(up0));
    P2e = 1 ./ (1 + exp(-D * a .* (habil**ones(1, I) - b2)));
    P = P1e - P2e;
    Pj_T = Pj_T .* exp(log(P)*up1' + log(1-P)*up0'); // q x n

    up1 = (u.==3).*pesos; up0=(u.!=3).*pesos; //print(rows(up1), " ",
rows(up0));
    P3e = 1 ./ (1 + exp(-D * a .* (habil**ones(1, I) - b3)));
    P = P2e - P3e;
    Pj_T = Pj_T .* exp(log(P)*up1' + log(1-P)*up0'); // q x n

    up1 = (u.==4).*pesos; up0=(u.!=4).*pesos; //print(rows(up1), " ",
rows(up0));
    P4e = 1 ./ (1 + exp(-D * a .* (habil**ones(1, I) - b4)));
    P = P3e - P4e;

```

```

Pj_T = Pj_T .* exp(log(P)*up1' + log(1-P)*up0');    // q x n

up1 = (u.==5).*pesos;  up0=(u.!=5).*pesos;  //print(rows(up1), " ",
rows(up0));
P = P4e;
Pj_T = Pj_T .* exp(log(P)*up1' + log(1-P)*up0');    // q x n

// -----

Pj_Tg=Pj_T.*g;    //print("Pj_Tg: ",Pj_Tg[][:4]);    // q x n
Pj=sumc(Pj_Tg);  //print("Pj: ",Pj[:4]);    // 1 x n
ge=Pj_Tg ./ Pj;    // q x n
}

EstimaHabilidadesMVM() {  GeraPontosdeQuadratura();
  Monta_ge();
  theta_MVM=sumc(habil .* ge)';
  dif = (thetav-media);
}

ImprimeHabilidades() {  decl Results;

  print("\n *****\n");
  Results=thetav~theta_MVM~(theta_MVM-thetav);
  println("Hab.Verdadeiras | Hab.Estimadas por MVM | Desvios_MVM");
  println("\n  Indiv | HabV | H_MVM | D_MVM", "%8.4f", range(1,n)'
~Results);
// println("\nResultados Medios: ", "%8.3f", meanc(Results));
// println("\nDesvios-padroao: ", "%8.3f", sqrt(varc(Results)));
// println("\nDesvios Quadratico Medio (MV e MVM): ", "%8.3f", Results[] [3]'
*Results[]
[3]/n~Results[] [5]'*Results[] [5]/n);
savemat("ResultsNominal.xls", Results);
println("\n estimativa theta : thetav | dif | media", "%8.3f", habil2~
dif~media);
Results=habil2~dif~media;
savemat("ResultsNominal.xls", Results);
print("\n *****\n");
} fim() {  print("Fim!!!"); }
main() {
  LerItens();

  for(j=0; j<100; ++j)
    { print("iteracao: ", j);
      GerarDados();
      EstimaHabilidadesMVM();
      delta = delta~theta_MVM;
      etapas = dropc(delta, 0);
      media = sumr(etapas)/columns(etapas);
    }
}

```

```
        }  
        GeraEstatisticas();  
  
        ImprimeHabilidades();  
// GuardaHabilidades();    // guardando as estimativas da habilidade.  
        fim();  
}
```

BIBLIOGRAFIA

- ANDRADE, D.F., TAVARES, H.R., VALLE, R.C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. Associação Brasileira de Estatística: São Paulo.
- Bock, R.D., & Lieberman, M. (1970). *Fitting a response model for n dichotomously scored items*. *Psychometrika*, 35, 179-197.
- AZEVEDO, C. L. N. (2003) *Métodos de Estimação na Teoria da Resposta ao Item*. UFC.
- BAKER, F. B. (1992). *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc.
- BOCK, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- BOCK, R. D. and AITKIN, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of a EM algorithm. *Psychometrika*, 46, 433-459.
- BOCK, R. D. and LIEBERMAN, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- BOCK, R. D. and ZIMOWSKI, M. F. (1997). Multiple Group IRT. In *Handbook of Modern Item Response Theory*. W.J. van der Linder e R.K. Hambleton Eds. New York: Springer-Verlag.
- DOORNIK, J. A. (1999). *Object-oriented matrix programming using ox, 3 edn*. URL: Kent: Timberlake Consultants e Oxford: <http://www.doornik.com>
- ISSAC, E. and KELLER, H. B. (1966). *Analysis of Numerical Methods*. New York: Wiley & Sons.
- MASTERS, G. N. (1982). *A Rasch model for partial credit scoring*. *Psychometrika*, 49, 149-174.
- SAMEJIMA, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17.