Departamento de Estatística - Universidade de Brasília

Trabalho de Conclusão de Curso 2

Aplicação da Teoria de Resposta ao Item para itens dicotômicos do Vestibular da UnB

Rafael Rezende de Moraes

Orientador: Prof^{o}. Antônio Eduardo Gomes

Brasília, 01 de Novembro de 2018

Rafael Rezende de Moraes

Aplicação da Teoria de Resposta ao Item para itens dicotômicos do Vestibular da UnB

Trabalho de conclusão de curso apresentado para obtenção do título de Bacharel em Estatística ao Departamento de Estatística da Universidade de Brasília.

Brasília,
01 de Novembro de 2018

Resumo

O presente estudo tem como objetivo analisar as notas (escores) e o efeito da não resposta nos escores dos indivíduos que realizaram o vestibular do segundo semestre de 2014 da UnB, realizado pelo CEBRASPE, de acordo com os métodos da Teoria de Resposta ao Item (TRI), que é atualmente aplicado no ENEM (Exame Nacional do Ensino Médio). Esta análise começa com uma análise descritiva das respostas dos indivíduos que realizaram a mesma, seguida por algumas comparações dos escores calculados. A primeira comparação é em relação ao Método Convencional e o Modelo de Resposta Gradual (MRG), e posteriormente a comparação entre este e o Modelo dicotômico de 3 parâmetros também da TRI, tomando a não reposta como erro no de 3 parâmetros. O modelo do MRG poderia ser um substituto do método convencional com o intuito de melhorar a avaliação do traço latente dos indivíduos, visto que a teoria da resposta ao item possui algumas vantagens. Os resultados se mostraram bem parecidos comparando às notas estimadas pelo MRG, sendo assim uma opção eficaz na avaliação. Em relação a questão do efeito da não respostas no escore estimado, foi notado que o modelo politômico (MRG) favorece os indivíduos com escore mais elevado.

Palavras Chave : Teoria da Resposta ao Item, Modelo de Resposta Gradual, Modelo dicotômico de 3 parâmetros, Vestibular, CEBRASPE, Avaliação, Não Resposta.

Abstract

The paper aims to analyze the scores and the effect of the non-response in the scores of people that participated in the UnB's admission test of the second semester of 2014, conducted by CEBRASPE, according with the Item Response Theory (IRT) methods, which is nowadays used in ENEM (a national exam of the high school). This study begins with the analysis of the responses of each individual, followed by some comparisons of the calculated scores. The first comparison is regarding the conventional method and the Gradual Response Model(GRM), and therefore the comparison between the last and the dichotomous model of 3 parameters either of the IRT, taking the non-response as an error in the second model. The model of GRM could be a substitute for the current method with the intent to improve the evaluation of the latent trace of the individuals, since in item response theory have some advantages. The results were very close compared with the scores estimated by GRM, thus is an appelaing option for this evaluation. In relation to the question of the effect of non-responses in the estimated score, it was noticed that the GRM benefits people with highers scores.

Key Words: Item Response Theory, Gradual Response Model, Dichotomous Model of 3 parameters, Admission test, CEBRASPE, Evaluation, Non-Response.

Sumário

1 Introdução

Quando vamos falar sobre algo como conhecimento, temos que levar em consideração que a variável em questão não é diretamente observável. Na área da psicometria, esse tipo de variável, comumente chamada de não observável, traço latente, construto ou até mesmo habilidade, é frequentemente citada.

Embora existam diversas técnicas de medição de conhecimento, boa parte delas não são fáceis e diretas por diversos aspectos técnicos, como é o caso da Teoria de Resposta ao Item. Nessa teoria, que foi criada como alternativa a Teoria Clássica das Medidas (TCM), o papel principal é a proposição de modelos de medição para traços latentes, como por exemplo, a proficiência em múltiplas matérias dispostas na prova do Vestibular da UnB.

Essa medida indireta é obtida a partir de respostas coletadas de um conjunto de itens, elaborados de modo a formar um instrumento de medida que possa permitir a sua quantificação de modo fidedigno, sendo mais precisa que a TCM, que compara somente o desempenho de indivíduos que realizaram a mesma prova. Sendo assim, se uma pessoa realizar duas provas diferentes, pode ser que ela obtenha notas diferentes, fato que provavelmente não ocorreria na TRI, desde que essas provas sejam feitas de acordo com os padrões de qualidade da mesma.

Nesta proposta, iremos estudar e aplicar modelos usados na Teoria de Resposta ao Item a um banco de dados real que contém informações de respondentes do Vestibular da UnB de 2014, tendo como foco o modelo de resposta gradual, no qual iremos calcular a nota do respondente considerando as seguintes categorias: “Não responder”, “Acertar” e “Errar”. Neste, considera-se que as categorias de resposta de uma questão podem ser ordenadas entre si, de modo que a categoria mais alta adicione mais a soma total do escore do respondente e a categoria mais baixa contribua menos.

2 OBJETIVOS

2.1 Objetivos Gerais

Obter os escores dos respondentes do Vestibular da UnB de 2014 via Teoria de Resposta ao Item.

2.2 Objetivos Específicos

- Estudar a Teoria de Resposta ao Item (TRI).
- Implementar o modelo usando o Software R.
- Comparar o escore obtido pela TRI com o escore convencional do vestibular (efeito da não resposta no escore).
- Comparar a estimativa do θ (proficiência) quando temos 3 categorias e quando temos 2 categorias (considerando a não resposta como erro).
- Analisar a diferença desses θs com o número de não respostas.

3 Teoria de Resposta ao Item

3.1 Contextualização

A necessidade de se medir traços latentes é bastante antiga, ou seja, muitos métodos e estudos foram estimulados e desenvolvidos em prol de obter uma solução para tal. Inicialmente foi criada a Teoria Clássica dos Testes (TCT) ou Teoria das Medidas, bastante reconhecida e com notável participação de Gulliksen (1953) e Guilford (1954).

A TCT é baseada no escore total do teste, ou seja, suas análises e interpretações estão sempre ligadas à prova como um todo e não a um item específico. Desde o início, notou-se que haviam limitações no método da TCT, sendo a inviabilidade de comparação entre indivíduos que se submeteram e que não se submeteram as mesmas provas, a mais preponderante.

Pouco depois foi desenvolvida a técnica da Teoria de Resposta ao Item (TRI), que veio com o intuito de sanar todas ou boa parte dessas limitações. Esta tem os itens como elementos centrais da análise, é composta de vários modelos de medição para traços latentes que possam preencher as falhas da TCT e consequentemente aumentar as áreas de aplicações.

Dentre as vantagens da TRI, as seguintes são de grande relevância: Possibilita a comparação de traços latentes tanto de indivíduos de populações iguais quando submetidos a testes diferentes, quanto de indivíduos de populações diferentes quando submetidos a testes iguais; Possibilita que somente os dados respondidos sejam utilizados em suas análises; Presença da invariância, em que os parâmetros dos itens não dependem do traço latente do respondente e os parâmetros dos indivíduos não dependem dos itens presentes na prova.

Após a criação de pacotes computacionais, a TRI começou a ser utilizada em larga escala, chegando no Brasil inicialmente na análise dos resultados do Sistema Nacional de Ensino Médio (SAEB) em 1995, e até hoje é utilizada amplamente em diversas áreas de estudo, sendo bastante citada na psicometria, medicina, entre outras.

3.2 Modelos da Teoria de Resposta ao Item

Segundo Andrade et al.(2000), a TRI é um conjunto de modelos matemáticos que buscam determinar a probabilidade de um indivíduo acertar um item como função dos parâmetros dos itens e dos traços latentes do respondente, sendo assim, essa probabilidade aumenta de acordo com a habilidade do indivíduo.

Por Demars(2010) segue que os modelos da TRI apresentam a relação entre o traço latente medido pelo instrumento e uma resposta a um determinado item. Sendo que os itens podem ser divididos em dicotômicos (em que há 2 categorias de resposta) e politômicos (mais de 2 categorias).

Os diversos modelos da Teoria de Resposta ao Item dependem dos seguintes atributos: Natureza do item, número de populações do modelo e a quantidade de traços latentes a serem calculados. A natureza do item se refere ao número de categorias dispostas nos itens, como já citado, são classificados em dicotômicos ou politômicos. Já o número de populações é classificado como uma ou mais de uma. O número de traços de latentes é classificado como unidimensional quando há apenas um traço a ser calculado e multidimensional quando há mais de um traço.

Este presente trabalho consisitirá na análise quando os itens são dicotômicos e quando consideramos a não resposta como uma categoria intermediária, tornando os itens politômicos.

3.2.1 Modelos para Itens Dicotômicos

Para os itens dicotômicos, os modelos comumente utilizados são os modelos logísticos que são diferenciados pelo números dos parâmetros dos itens. O modelo mais básico é o logístico unidimensional de 1 parâmetro, também citado como modelo de Rasch, que contêm somente o parâmetro de dificuldade (b_i). Outro modelo é o logístico unidimensional de 2 parâmetros, que além da presença do parâmetro de dificuldade tem também o parâmetro de discriminação (a_i). Por fim, temos o modelo com todos os parâmetros, modelo logístico unidimensional de 3 parâmetros contendo os a_i,b_i e c_i que é o parâmetro de acerto ao acaso.

A seguir estão as fórmulas dos modelos para itens dicotômicos, respectivamente segundo o número de parâmetros:

$$\mathrm{P}(U_{ij} = 1\Big|\theta_j) = \tfrac{1}{1+e^{-D(\theta_j - b_i)}}(1)$$

$$P(U_{ij} = 1\Big|\theta_j) = \tfrac{1}{1+e^{-Dai(\theta_j - b_i)}}(2)$$

$$P(U_{ij} = 1\Big|\theta_j) = c_i + (1 - c_i)\tfrac{1}{1+e^{-Dai(\theta_j - b_i)}}(3)$$

com i=1,2,...,I, e j=1,2,...,n, sendo:

• U_{ij}, uma variável dicotômica que assume o valor 1 quando o j-ésimo indivíduo acerta o i-ésimo item, ou assume 0 quando ele não acerta o mesmo;

• θ_j, representa a habilidade ou traço latente do j-ésimo indivíduo;

• P($U_{ij} = 1|\theta_j$), é a probabilidade de um indivíduo j com habilidade θ_j acertar o item i, tal probabilidade é chamada de Função de Resposta do Item (FRI);

• b_i, o parâmetro de dificuldade, ou posição, do item i, medido na mesma escala que a habilidade;

• a_i, 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, o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade acertarem o item i (comumente referido como probabilidade de acerto casual);

• D, constante igual a 1 quando queremos aproximar a curva logística, e quando o valor é igual a 1,702 se aproxima a curva logística da ogiva normal.

3.2.2 Função Característica do Item

A TRI é usada para analisar um conjunto de respostas de uma prova questionário a fim de obter as probabilidades de um indivíduo dar uma resposta correta para os itens presentes, sendo que para isso é levado em consideração o traço latente do indivíduo e algumas características dos itens. Tal probabilidade pode ser representada pela proporção dos acertos de um determinado item entre todos os respondentes e que é representada por uma curva em formato de S, chamada de Curva Característica do Item (CCI) e que está representada na Figura 2.1.

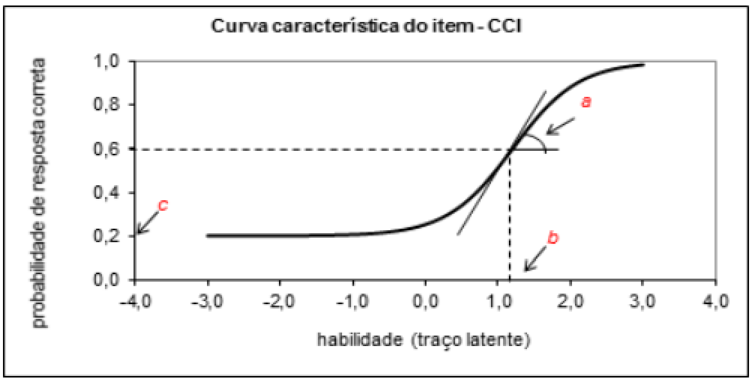


Figura 2.1: Exemplo de uma Curva Característica do Item (CCI) com $a_i = 1.3$, $b_i = 1.2$ e $c_i = 0.2$.

Fonte: Andrade et al. (2000).

Podemos notar que essa curva não apresenta comportamento linear e que quanto maior for a habilidade, maior será sua probabilidade de acerto. Devemos ressaltar também que cada curva dependerá exclusivamente dos parâmetros do respectivo item. Sendo que o valor de b é igual ao valor da habilidade no ponto em que a inclinação da função é máxima, a é proporcional a tangente da curva no ponto de inflexão e c, que muitas vezes é citado como a probabilidade de advinhação, não varia como função do nível de habilidade. Ou seja, pessoas que possuem habilidade alta ou baixa teriam a mesma probabilidade de advinhar a questão.

3.2.3 Função de Informação do Item e do Teste

A função de informação do teste é bastante usada juntamente com a curva característica do Item e ela permite a análise da quantidade de informção determinado item possui em relação ao traço latente. A função de informação do item é expressa pela seguinte expressão matemática:

$$I_i(\theta) = \frac{[\frac{\mathrm{d}}{\mathrm{d}\theta}P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

Sendo que $\mathrm{I}_i(\theta)$ é a "informação"fornecida pelo item i no traço latente θ, $\mathrm{P}_i(\theta)$=P(U_{ij}=1/θ) e $\mathrm{Q}_i(\theta)$=1-$\mathrm{P}_i(\theta)$.

Como o modelo logístico de 3 parâmetro será utilizado neste projeto, adaptamos esta função ao modelo, ficando da seguinte maneira:

$$I_i(\theta) = D^2 a_i^2 \frac{Q_i(\theta)}{P_i(\theta)}[\frac{P_i(\theta) - c_i}{1 - c_i}]^2$$

Tal equação nos mostra a representatividade dos 3 parâmeros presente no modelo em relação a toda informação do item, ou seja, a informação é maior:

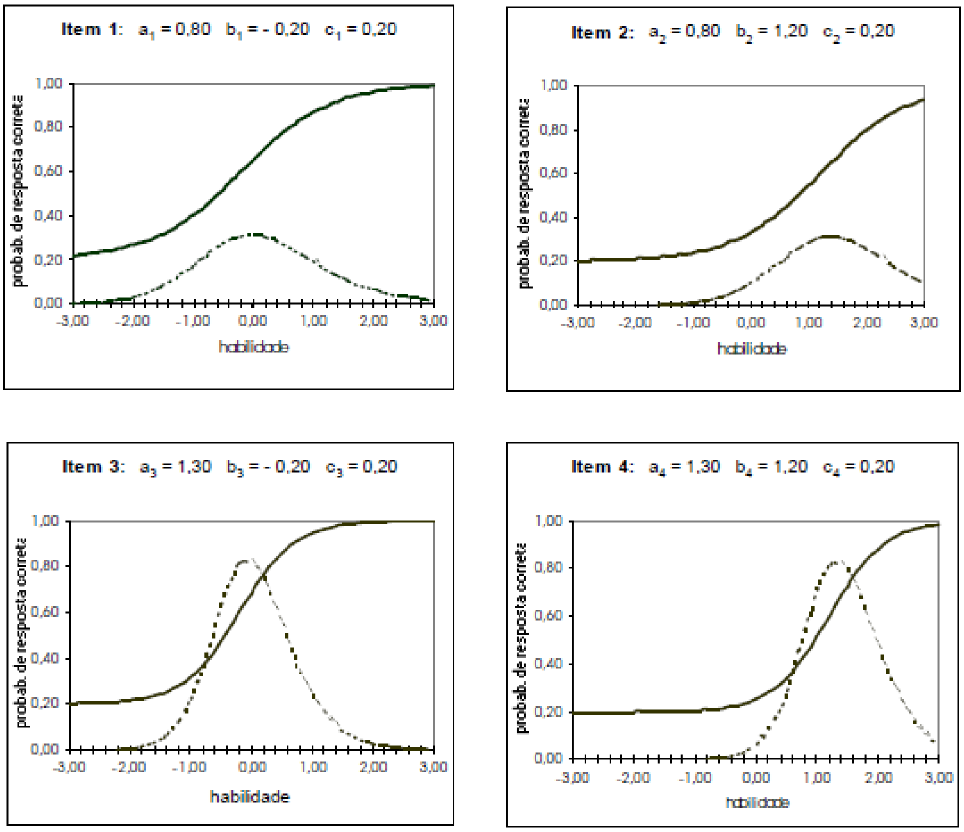
a) Quando b_i se aproxima de θ;

b) Quanto maior for o a_i;

c) Quanto mais c_i se aproximar de 0.

Como comentado anteriormente, a função de informação do Item é usada em conjunto com a CCI, sendo representada pela parábola como podemos ver na Figura 2.2

Figura 2.2 *Curvas características e de informação de vários itens*



Fonte: Andrade et al. (2000)

A Função de Informação do teste é nada mais que o somatório das informações fornecidas pelo itens presentes no teste/questionário em questão.

$$I(\theta) = \sum_{i=1}^{I} I_i(\theta)$$

3.2.4 Modelos para Itens Politômicos

Os modelos seguintes são para análise de itens de múltipla escolha e itens de resposta livre (abertos), itens esses que são graduados, ou seja, são corrigidos ou elaborados de forma que uma ou mais categorias intermediárias ordenadas entre as categorias certo e errado. É válido ressaltar que para esse tipo de item a resposta é considerada sendo ela certa ou errada.

Dentre os diversos Modelos para itens não dicotômicos (politômicos) iremos comentar sobre os mais conhecidos.

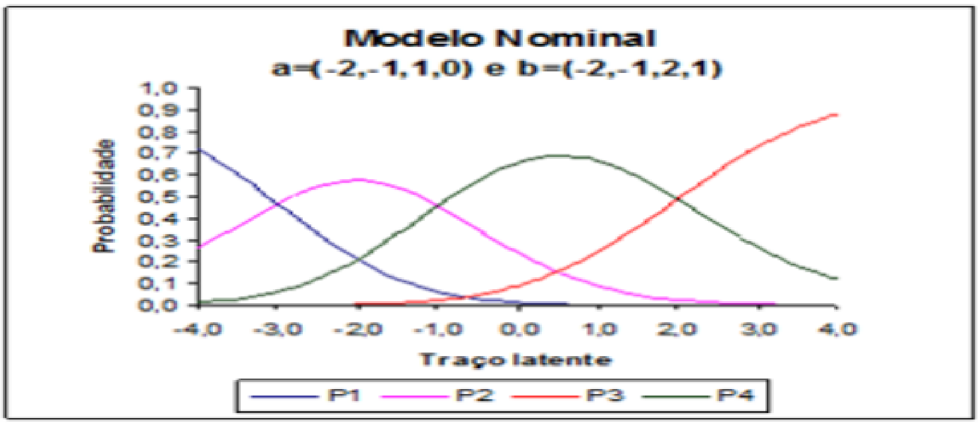
3.2.4.1 Modelo de Resposta Nominal (Nominal Categories Model)

O modelo de Resposta Nominal foi desenvolvido por Bock(1972), foi baseado 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. Esse modelo veio com o intuito de maximizar a precisão da habilidade estimada usando toda a informação presente nas repostas dos indivíduos. Tendo Bock assumido que a probabilidade com que o indivíduo (j) selecionasse determinada opção (k) seria dado da seguinte forma:

$$P_{i,k}(\theta_j) = \frac{e^{a^+_{i,k}(\theta_j - b^+_{i,k})}}{\sum_{h=1}^{m_i} e^{a^+_{i,h}(\theta_j - b^+_{i,h})}}$$

com $i = 1, \ldots, I$; $j = 1, \ldots, n$; e $k = 1, \ldots, m_i$.

Em cada θ_j , a soma das probabilidades sobre as m_i opções é 1. As quantidades ($b^+_{i,k}$; $a^+_{i,k}$) são parâmetros do item i relacionados a k-ésima opção, sendo que o modelo assume a não ordenação a priori das opções de resposta. Abaixo podemos ver a representação de um Modelo de Resposta Nominal quanto ao traço latente



Fonte: Andrade (2005)

3.2.4.2 Modelo de Resposta Gradual (Graded Response Model)

O modelo de resposta gradual foi desenvolvido por Samejima (1969) e assume que as categorias de resposta de um item podem ser ordenadas entre si, e como no modelo de Bock, tenta obter mais informação das respostas dos indivíduos do que simplesmente se eles deram respostas certas ou erradas.

A probabilidade de um indivíduo j escolher uma particular categoria ou outra mais alta do item i pode ser dada pela seguinte expressão, que é uma extensão do modelo logístico de 2 parâmetros:

$$P^{+}_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}}$$

$i = 1, \ldots, I$; $j = 1, \ldots, n$; e $k = 1, \ldots, mi$; onde: $b_{i,k}$ é o parâmetro de dificuldade da k-ésima categoria do item i e os outros parâmetros são análogos aos definidos anteriormente.

No caso dos modelos para itens não dicotômicos, a discriminação de uma categoria específica de resposta depende tanto do parâmetro de inclinação, comum a todas as categorias do item, quanto da distância das categorias de dificuldade adjacentes.

3.2.4.3 Modelo de Escala Gradual (Rating Scale Model)

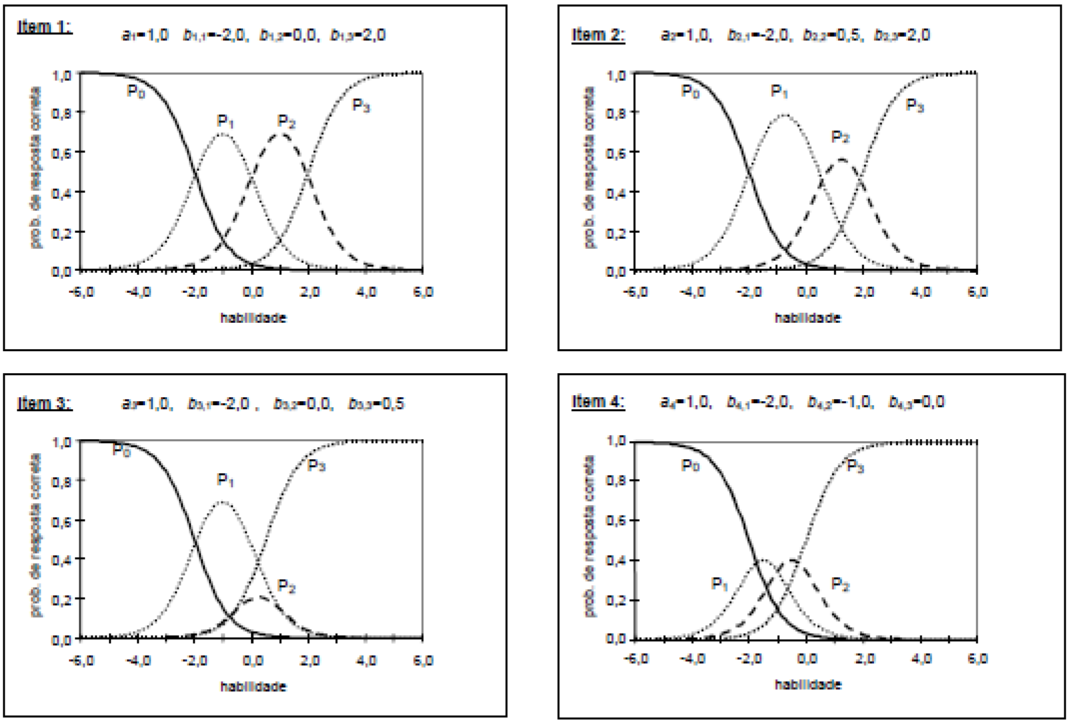
O modelo de Escala Gradual foi desenvolvido por Andrich (1978) e é um caso particular do anterior (Modelo de Resposta Gradual). Esse modelo também é adequado para itens com categorias de resposta ordenadas, mas com uma suposição a mais: a de que os escores das categorias são igualmente espaçados. Sua expressão matemática se dá da seguinte forma:

$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_k)}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_i + d_{k+1})}}$$

i=1,...I; j=1,...n; e k=1,...mi; onde: b_i é o parâmetro de locação do item i e d_k é o parâmetro de categoria.

É válido ressaltar que os parâmetros de categoria d_k não dependem do item, ou seja, são comuns a todos os itens do teste. Logo, se os itens que compõem a prova tiverem número de categorias diferentes, então este modelo não é adequado.

Figura 2.3 *Representação gráfica dos modelos de escala gradual e de resposta gradual*



Fonte: Andrade et al. (2000)

Na Figura 2.3 temos a representação gráfica do modelo de escala gradual e do modelo de resposta gradual para alguns itens com 4 categorias de resposta. Em todos os itens, o parâmetro a_i foi mantido igual a 1. Dessa maneira, podemos verificar a representatividade dos parâmetros de categoria $b_{i,k}$. O modelo de resposta gradual poderia ser representado por qualquer um dos itens acima, mas somente os itens 1 e 4, por terem os parâmetros de categoria igualmente espaçados, podem ser representantes do modelo de escala gradual.

3.3 Estimação dos parâmetros

Um dos pontos críticos da TRI é a estimação dos parâmetros que caracterizam o modelo da resposta ao item: parâmetros dos itens e as habilidades dos respondentes.

Nos modelos da TRI, como já mencionado anteriormente, a probabilidade de resposta certa a um determinado item depende exclusivamente da habilidade do examinado e dos parâmetros dos itens. Porém, majoritariamente, essas medidas são desconhecidas, sendo conhecidas somente as respostas aos itens dos testes.

Nos modelos da TRI têm-se um problema de estimação que envolve esses dois tipos de parâmetros. Podendo-se dividir o problema basicamente em três situações diferentes: Estimação das habilidades, quando já se conhecem os parâmetros dos itens; estimação dos parâmetros dos itens (processo chamado calibração) quando já se conhecem as habilidades e, por fim, a estimação conjunta das habilidades e dos parâmetros dos itens. Dessa maneira, será considerado o caso em que os dois tipos de medidas são desconhecidas.

Inicialmente, a estimação dos parâmetros em questão era feita através do método de máxima verossimilhança (MV) conjunta, o qual envolvia um grande número de parâmetros a serem estimados simultaneamente e gerava, consequentemente, inúmeros problemas computacionais e teóricos, como a possível inconsistência dos estimadores obtidos por essa técnica.

A estimação dos parâmetros dos itens e das habilidades se trata do caso mais comum, em que nem os parâmetros dos itens são conhecidos e nem as habilidades. Entretanto, devido à dificuldade de se estimar conjuntamente esse dois tipos de parâmetros, essa estimação é realizada em duas fases.

Na primeira fase, é realizada a estimação dos parâmetros dos itens e esta pode ser feita tanto pelo método de máxima verossimilhança conjunta (MVC), como por máxima verossimilhança marginal ou por métodos Bayesianos. Em seguida, são estimadas as habilidades na mesma escala dos parâmetros dos itens, uma vez que eles foram estimados na fase anterior. Esta fase pode ser realizada por métodos de máxima verossimilhança ou por métodos Bayesianos.

3.4 Métodos de Estimação

3.4.1 Estimação por Máxima Verossimilhança

A seguir, será descrito o método de máxima verossimilhança para a estimação dos parâmetros dos itens, considerando as habilidades conhecidas e desconhecidas, e também a utilização do método para a estimção das habilidades.

3.4.1.1 Estimação dos Parâmetros dos Itens

Quando se deseja estimar os parâmetros dos itens na situação em que as habilidades são conhecidas, consideramos as seguintes suposições:

i) as respostas oriundas de indivíduos diferentes são independentes;

ii) os itens são respondidos de forma independente por cada indivíduo dada a sua habilidade.

A função de verossimilhança pode ser expressada por:

$$L(\zeta) = \prod_{j=1}^{n} P(U_j = u_j \,|\theta_j, \zeta)$$

$$= \prod_{j=1}^{n} \prod_{i=1}^{I} P(U_{i,j} = u_{i,j} \,|\theta_j, \zeta_i)$$

$$P(U_{i,j} = u_{i,j} \left|\theta_j, \zeta_i\right) = P(U_{i,j} = 1 \left|\theta_j, \zeta_i\right)^{u_{ij}} P(U_{i,j} = 0 \left|\theta_j, \zeta_i\right)^{1-u_{ij}}$$

$$= P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

Sendo assim, a verossimilhança pode ser escrita como:

$$L(\zeta) = \prod_{j=1}^{n} \prod_{i=1}^{I} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

Com os procedimentos descritos em Andrade et al(2000), chegamos as seguintes equações dos parâmetros:

$$a_i : (1 - c_i) \sum_{j=1}^{n} \int_R [(u_{ij} - P_i)(\theta - b_i)W_i] g_j^*(\theta) = 0$$

$$b_i : -a_i(1 - c_i) \sum_{j=1}^{n} \int_R [(u_{ij} - P_i)W_i] g_j^*(\theta) = 0$$

$$c_i : \sum_{j=1}^{n} \int_R [(u_{ij} - P_i)\frac{W_i}{P_i^*}] g_j^*(\theta) = 0,$$

em que

$\mathrm{W}_i = \frac{P_i^* Q_i^*}{P_i Q_i}, P_i = c_i + (1 - c_i)[1 + e^{-Da_i(\theta - b_i)}]^{-1}, P_i^* = [1 + e^{-Da_i(\theta - b_i)}]^{-1},$

$\mathrm{Q}_i = 1 - P_i, Q_i^* = 1 - P_i^*, g_j^*(\theta) = \frac{P(U_{.j}|\zeta,\theta)g(\theta|\eta)}{\int P(U_{.j}|\zeta,\theta)g(\theta|\eta)\mathrm{d}\theta.}$

3.4.1.2 Estimação das Habilidades

Em relação as habilidades dos indivíduos, assumimos que os parâmetros dos itens já são conhecidos, estimados como explicado anteriormente. Sendo assim, podemos escrever a log-verossimilhança baseado em i) e ii) da seguinte maneira:

$$l(\theta) = lnL(\theta) \sum_{j=1}^{n} \sum_{i=1}^{I} [u_{ij} lnP_{ij} + (1 - u_{ij}) lnQ_{ij}]$$

De maneira simplificada as expressões referentes aos processos de estimação são:

Vetor Escore :

$$S(\theta_j) = \sum_{i=1}^{I} a_i(1-c_i)(u_{ij} - P_{ij})W_{ij},$$

Matriz Hessiana :

$$H(\theta_j) = \sum_{i=1}^{I}(u_{ij} - P_{ij})W_{ij}[H_{ij} - (u_{ij} - P_{ij})W_{ij}h_{ij}^2],$$

Informação de Fisher :

$$I(\theta_j) = \sum_{i=1}^{I} P_{ij}^* Q_{ij}^* W_{ij} h_{ij}^2,$$

com

$$h_{ij} = (P_{ij}^* Q_{ij}^*)^{-1}(\frac{\partial P_{ij}}{\partial \theta_j}) = a_i(1-c_i)$$

e

$$H_{ij} = (P_{ij}^* Q_{ij}^*)^{-1}(\frac{\partial^2 P_{ij}}{\partial \theta_j^2}) = a_i^2(1-c_i)(1-2P_{ij}^*)$$

Newton-Raphson :

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

Escore de Fisher:

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

3.5 Estimação alternativa dos parâmetros dos itens e das proficiências

A tentativa de ajuste dos modelos de TRI via pacote "mirt"apresenta problemas de convergência. Como alternativa, adotamos o seguinte procedimento:

1) obtemos, para cada item, uma estimativa não paramétrica da CCI utilizando a técnica de regressão isotônica (descrita mais abaixo);

2) como a estimativa obtida no passo (1) é uma função escada, utilizamos o método de kernel para obter uma versão suavizada da estimativa não paramétrica da CCI;

3) para ajustar o modelo paramétrico, obtemos estimativas dos parâmetros dos itens que minimizem a distância entre o modelo paramétrico e a estimativa não paramétrica suavizada da CCI via função "optim"no pacote R.

A partir das estimativas dos parâmetros dos itens, foram obtidas as estimativas das proficiências para o modelo ditocômico logístico de 3 parâmetros via função "eap"o pacote "ltm"do R.

Para ajuste do modelo de resposta gradual (Samejima, 1969) quando considera-mos itens politômicos, fizemos o ajuste das duas CCI's de forma análoga à descrita acima. O parâmetro de discriminação comum às duas curvas, e os parâmetros de locação (dificuldade) foram estimados de modo a minimizar a soma das distâncias (d_{L1}) entre as estimativas paramétricas das duas CCI's e as respectivas estimativas não paramétricas, também obtidas via regressão isotônica, considerando, para uma delas, a dicotomização "acerto"e "não acerto"(agrupando "erro"e "não resposta"numa mesma categoria) e, para a outra curva, considerando dicotomização "erro"e "não erro"(agrupando "acerto"e "não resposta"numa mesma categoria). Após a obtenção das estimativas dos parâmetros, procedeu-se a obtenção dos estimadores de máxima verossimilhança das proficiências.

Para o modelo de resposta gradual, a estimativa de θ para cada respondente foi obtida via máxima verossimilhança maximizando a função :

$$\prod_{i=1}^{I}[1 - P_{i1}^{+}(\theta)]^{I_{\{-1\}}}[P_{i1}^{+}(\theta) - P_{i2}^{+}(\theta)]^{I_{\{0\}}}[P_{i2}^{+}(\theta)]^{I_{\{1\}}}$$

3.5.1 Regressão isotônica

Considerando o conjunto ordenado $\{x_1, \ldots, x_n\}$ e os respectivos valores observados $\{g(x_1), \ldots, g(x_n)\}$ de uma função g, a regressão isotônica g^* é a função que minimiza a soma de quadrados

$$\sum_{i=1}^{n}(g(x_i) - f(x_i))^2 w(x_i)$$

entre todas as funções f não decrescentes com domínio nos pontos $\{x_1, \ldots, x_n\}$. Os valores $\{g^*(x_1), \ldots, g^*(x_n)\}$ são dados pelas derivadas à esquerda da função minorante convexa máxima do diagrama de soma acumulada formado pelos pontos $P_0 = (0, 0), P_i = (\sum_{j=1}^{i} w(x_j), \sum_{j=1}^{i} g(x_j)w(x_j))$. A função minorante convexa máxima é a maior função entre todas as funções convexas que assumem valores abaixo de todos os pontos do diagrama de soma acumulada acima.

A estimativa não paramétrica da CCI utilizada neste trabalho é dada pela regressão isotônica de $g(\widehat{\theta}_j) = 1$ se o j-ésimo respondente deu resposta positiva para o item, e $g(\widehat{\theta}_j) = 0$ se o j-ésimo respondente deu resposta negativa para o item. Os pesos $w(\widehat{\theta}_j)$ são todos iguais a 1.

Em nosso problema, fazemos x_j=$\widehat{\theta}_j$, j=1,...,n , onde assumimos $\widehat{\theta}_1 \leq \widehat{\theta}_2 \leq \ldots \leq \widehat{\theta}_n$.

No software R, a regressão isotônica pode ser obtida atráves da função *monoreg* do pacote *fdrtool*.

3.6 Contextualização e proposta de estudo

Neste estudo, consideraremos que o banco de dados será constituído com questões do tipo "A"as quais além de apresentarem as opções “certo” ou “errado”, como citado anteriormente, são corrigidas a partir do seguinte cálculo: caso a resposta do respondente esteja em concordância com o gabarito oficial definido na prova, ou seja, caso ele acerte a questão, ele tem uma pontuação +1 (um ponto positivo). Caso a resposta do candidato esteja em discordância com o gabarito oficial definido na prova, ou seja, caso ele erre a questão, ele tem uma pontuação -1 (um ponto negativo). E, por fim, caso não haja marcação por parte do candidato, ele tem pontuação 0. As informações foram extraídas do Edital N^{o}1, do 1^o vestibular da UnB de 2014, lançado no dia 22 de abril de 2014.

Em casos de incerteza do candidato, a não resposta costuma ser vantajosa. Sendo assim, no banco de dados utilizado nesta pesquisa, as respostas faltantes serão consideradas também como uma opção de resposta, ou seja, teremos três opções de resposta: “Errar”,“Não Responder” e “Acertar”.

A partir da suposição considerada acima, o modelo utilizado neste trabalho será o modelo da teoria da resposta ao item, com somente uma população envolvida no estudo; unidimensional, ou seja, com apenas um traço latente medido e de natureza dicotômica (inicialmente), fizemos uma transformação para itens não dicotômicos para comparação e averiguação do efeito da "não resposta"nos escores. O modelo para itens dicotômicos considerado será o modelo logístico de 3 parâmetros e para itens politômicos será o modelo de Resposta Gradual, em que as categorias de resposta de uma questão podem ser ordenadas entre si, de tal forma que a categoria mais baixa contribua menos para o escore do respondente e a categoria mais alta contribua mais.

4 Metodologia

O desenvolvimento do presente trabalho constitui-se de pesquisa bibliográfica, com abordagem quantitativa, desenvolvida para modelar o desempenho de respondentes da prova do Vestibular da UnB realizado pelo CEBRASPE (Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos), por meio de seus escores obtidos na mesma.

O banco de dados utilizado foi o de alunos que fizeram a prova do segundo semestre de 2014, foram consideradas somente as questões nas quais as possibilidades de resposta são apenas duas (questões do tipo A), ou seja, itens com variáveis dicotômicas, vale ressaltar que também existe a possibilidade do respondente não fazer o item.

Neste banco de dados há informações referentes ao evento em questão, a opção de curso marcada pelo respondente, o número de identificação de cada um, suas respectivas marcações para cada item, e por fim, os gabaritos dos itens, totalizando 237 questões do tipo A (já excluídas as anuladas) com 7232 indivíduos (respondentes), gerando 1.713.984 respostas. Com tais informações, já é possível o cálculo do escore total dos respondentes segundo a TCM e pela TRI a fim de fazer comparações.

O modelo da Teoria de Resposta ao Item que será usado é o dicotômico de 3 paramêtros, que são: dificuldade(b_i) , discriminação(a_i) e a probabilidade de o sujeito acertar o item ao acaso(c_i). Este seguirá a seguinte ordenação das categorias de resposta: "erro"e "acerto", sendo que em "erro"estarão os erros e as não respostas. Sua expressão matemática está descrita na seção 3.2.1.

Além disso, serão feitas duas comparacões, uma entre as estimativas dos θs (proficiências) para este e para o Modelo de Resposta Gradual(MRG), e outra para a diferença entre elas e a proporção de não respostas, para ver o efeito da não resposta no cálculo da proficiência, sendo que na MRG serão consideradas as três categorias anteriormente citadas, sem aglutinação da não resposta como erro. Já o Modelo de Resposta Gradual é descrito na seção 3.2.4.

De forma ordenada, essas serão as etapas que irão formar a metodologia empregada nesse projeto:

1) Tratamento dos dados quanto a possíveis dados faltantes, anulações de itens, dupla marcação de itens, o que os torna descartáveis, e demais alterações que forem julgadas necessárias.

2) Procura e preparação de programação usada para obtenção dos escores totais no software livre R e possível teste com dados disposto na internet ou simulados, para verificar o bom funcionamento do mesmo.

3) Execução da programação para obtenção dos parâmetros e o cálculo dos escores.

4) Comparações entre os escores obtidos.

5 Resultados

Neste capítulo, serão apresentados os resultados obtidos a partir das análises do banco de dados do CESBRASPE, que foi descrito no capítulo anterior. O programa que foi utilizado para gerar os resultados foi o software R, e as funções utilizadas para gerar o modelo de 3 parâmetros, gerar o modelo de resposta gradual, estimar os parâmetros para ambos os modelos, calcular os escores dos indivíduos para o modelo de 3 parâmetros e calcular os esocres dos indivíduos para o MRG foram, respectivamente: *tpm, mirt, coef, eap* e *fscores*, além da função *optim* utilizada no método alternativo para estimação de mais de um parâmetro e a função *monoreg* utilizada para obter a regressão isotônica.

5.1 Análise Descritiva

De acordo com o que foi apresentado no capítulo 4, foi disponibilizado pelo CEBRASPE um banco de dados contendo as respostas de 21.969 alunos às 300 questões aplicadas no vestibular da UnB de 2014/2. Pelo fato de terem sido aplicadas 3 tipos de provas, e por cada tipo ter 3 opções de língua estrangeira, nesta análise foram usadas somente as respostas de 1 dos tipos das provas, foram excluídas os 30 primeiros itens (itens esses que são referentes a língua estrangeira) e outros itens que ou foram anuladas pelo próprio Cespe, ou não são do interesse da nossa análise (itens politômicos).

Sendo assim, o nosso banco ficou constituído de de 7232 indivíduos que responderam a de 237 questões do tipo A (itens dicotômicos). Esses itens referiam-se as seguintes áreas de estudo : Matemática, Química, Física, Biologia, Língua Portuguesa, História, Geografia, Artes Visuais, Música, Filosofia e Antropologia.

Após o tratamento do banco de dados do vestibular da UnB, foi analisada a frequência de respostas dadas pelos indivíduos como um todo, sendo as 3 categorias : Errar (-1), Não responder (0) e Acertar (1).

Sobre os itens errados, temos algumas estatísticas a serem comentadas: A maioria dos indivíduos (34,21%) erraram somente em torno de 20% e 30% da prova, número que pode ser considerado relativamente baixo, podendo indicar uma boa preparação dos alunos. Observamos que nenhum aluno errou todas as questões da prova, e o indivíduo 6443 foi o que mais errou, com 78,9% da prova.

Em relação aos itens deixados em branco, apesar de ser possivelmente vantajoso não responder em caso de incerteza, a maioria deixa de responder até 40% da prova, em outras palavras, os alunos respondem mais da metade da prova. Também observamos que 5,93% (429 alunos) respondem tudo, ou seja, não deixam nenhum item sem marcação.

Sobre os itens certos, observamos que somente 0,11% dos alunos acertam acima de 80% da prova, e que nenhum aluno que acertou todos os itens da prova. O indivíduo 4711 foi o que acertou mais itens, 86,5% da prova, vale ressaltar que não necessariamente o maior índice de acerto corresponde a maior nota, já que há penalização para os itens respondidos de forma incorreta.

Na Tabela 5.1, estão apresentadas as frequências de respostas de cada categoria. Num total de 1.713.984 respostas, que é correspondente às respostas de 7232 indivíduos aos 237 itens da prova, foi observado que 28,69% dos itens foram marcados de forma incorreta, 33,14% dos itens foram deixados em branco e 38,15% dos itens foram marcadas corretamente.

Tabela 5.1: Frequências Globais de Respostas

Respostas	Frequência	Percentual	Função Acumulada
-1	491856	28.69665	491856
0	568135	33.14704	1059991
1	653993	38.15631	1713984

Podemos observar que as quantidades de respostas incorretas (-1), ausentes (0) e corretas (1) ficaram relativamente próximas, indicando equilíbrio entre a distribuição das mesmas. Essa distribuição de frequências é representada graficamente na Figura 5.1.

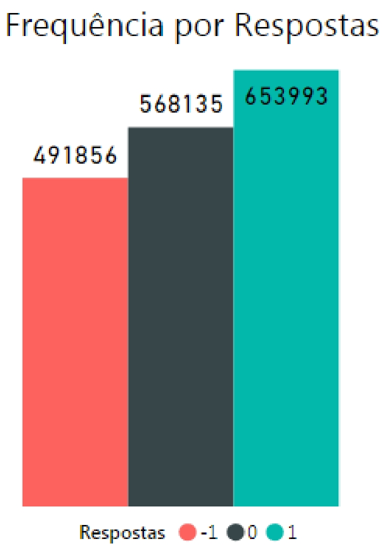


Figura 5.1: Gráfico de colunas das frequências de respostas

5.2 Análise dos Parâmetros dos Itens e Curva Característica dos Itens

5.2.1 Análise para o Modelo dicotômico

Depois do tratamento do banco de dados, foi aplicado o modelo dicotômico de 3 parâmetros e também o modelo politômico da Teoria de Resposta ao Item ou Modelo de Resposta Gradual (MRG), a fim de estimar as respectivas proficiências dos indivíduos. A partir da aplicação dos modelos, foram estimados os parâmetros dos itens e geradas as respectivas curvas características dos itens.

Após a estimação dos parâmetros do modelo dicotômico, foi notado que alguns valores encontrados nos 3 parâmetros em questão não eram ideais, o que pode ter sido causado por convergência no software, podendo ser também um problema no traço latente, pois estamos estimando o traço latente em diversas áreas de estudo, algo que possui certa dificuldade.

Partimos para a análise das respectivas Curvas Característica dos Itens (CCI), as quais representam a relação entre a probabilidade do indivíduo acertar o item dado a sua habilidade, $P(U_{ij} = 1 \mid \theta_j)$, e os parâmetros do modelo. Os gráficos da CCI o tem em seu eixo horizontal o valor do traço latente medido (habilidade) e o eixo vertical corresponde à probabilidade de o indivíduo ter sua resposta classificada como certa.

Abaixo temos a Figura 5.2 que ilustra a curva característica dos itens que vão do 108 ao 112, nele podemos observar que indivíduos com habilidade até aproximadamente -2 tem mais probabilidade de acertar o item 112, mas já um indivíduo com habilidade maior que 2 tem maior probabilidade de acertar o item 108. Um item poderia ser considerado bom quanto a sua discriminação caso ele tenha um formato que lembra um S, saindo de algo próximo a 0 e indo até próximo de 1, seria o oposto do observado no item 112, que se demonstra um item muito ruim quanto a discriminação, vale comentar também sobre a forma do item 110 que se comportam de maneira quase que linear, ou seja, a probabilidade de acerto se mantem praticamente a mesma independetemente da habilidade do respondente, o que sugere discriminação muito ruim.

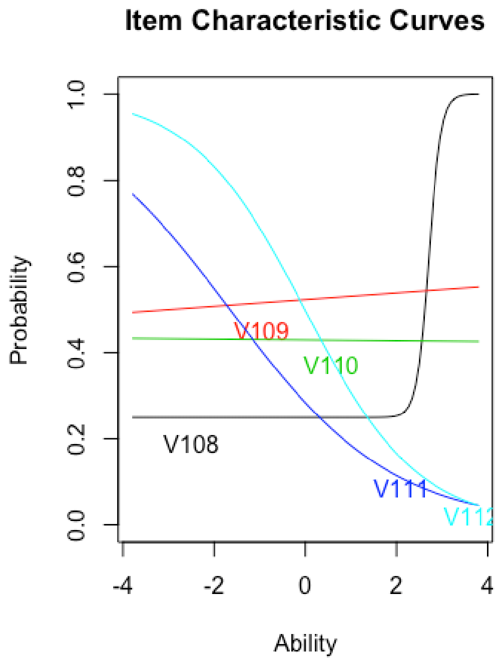


Figura 5.2: Curvas características dos itens 108 a 112

Na Figura 5.3 é apresentada a Curva de Informação do Teste (CIT), através desse gráfico podemos observar que o instrumento de medida tem maior informação para os valores da habilidade compreendidos entre aproximadamente 2,5 e 3, e também pode se ressaltar que há certa informação para os valores em torno de -1. Dessa maneira, este resultado indica que a prova é mais adequada para avaliar habilidades com valores contidos nesse pequeno intervalo, ou seja, mais um indício de que o método utilizado possui problemas, pois avalia bem poucas das habilidades. A curva ideal seria uma curva próxima ao formato de uma distribuição normal padrão, em que sua amplitude esteja incorporando a maior parte das habilidades dos indivíduos.

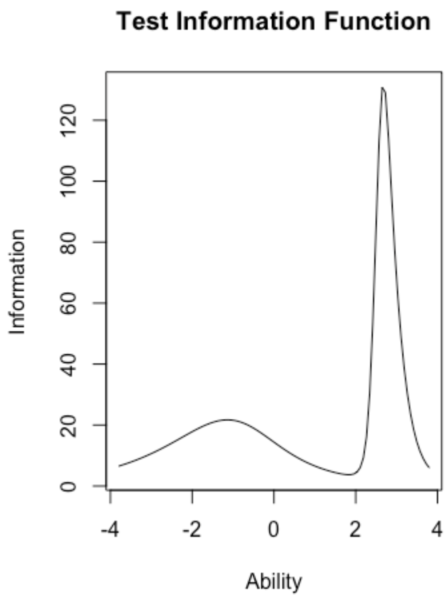


Figura 5.3: Curva de Informação do Teste (CIT)

5.2.2 Análise para o Modelo dicotômico pelo Método Alternativo

Averiguados os problemas de convergência, citado na seção 3.5, utilizamos a técnica de regressão isotônica a fim de e obter novos parâmetros e novas estimativas de proficiências e verificar se o erro foi sanado, ou pelo menos amenizado.

Tabela 5.2: Medidas descritivas dos parâmetros para o modelo dicotômico de 3 parâmetros por método alternativo

Medidas Descritivas	Dscrmn	Dfcld	Acrtocs
Mínimo	0.3683	-3.0655	0.1059
1º Quartil	0.4231	-0.4152	0.1815
Mediana	0.4714	0.8732	0.1904
Média	0.5089	0.9319	0.1787
3º Quartil	0.5768	2.2701	0.1938
Máximo	0.7532	4.2770	0.2107

Na tabela acima são apresentadas algumas medidas descritivas dos parâmetros dos itens estimados a partir do modelo dicotômico de 3 parâmetros de acordo com o método alternativo utilizado para estimação. Observamos que os valores apresentados nessa tabela estão mais condizentes com o esperado.

Em relação ao parâmetro de discriminação (a_i ou "Dscrmn"), o valor mínimo já não é mais negativo (0.3683), e percebe-se uma menor variação na discriminação dos itens, sendo que o valor máximo foi de 0.75 e que a média está em torno de 0.5, ou seja, pelos parâmetros apresentados, os itens tem baixa discriminação, sendo considerados baixos os valores abaixo de 1.

Quanto ao parâmetro de dificuldade (b_i ou "Dfcld"), notamos que ainda há itens que estão fora do intervalo em que a literatura indica (-2 a +2), mas em média os valores desse parâmetro estão dentro do habitual.

Notamos que a probabilidade de acerto ao acaso (c_i ou "Acrtocs") possui média de aproximadamente 0.17, ou seja, em média, para os candidatos que chutarem determinado item, a chance de acerto está em torno de 17%, o que faz com que o indivíduo tenha menos propenção ao chute.

Dessa maneira, podemos perceber que os resultados descritivos obtidos pelo método da regressão isotônica são razoáveis para seguir nossa análise, somente com a ressalva de que a prova não contém itens com boa discriminação.

A seguir temos um exemplo das estimativas das curvas características de um determinado item, sendo que o normal é que a curva vá crescendo de acordo com o escore padronizado dos respondentes.

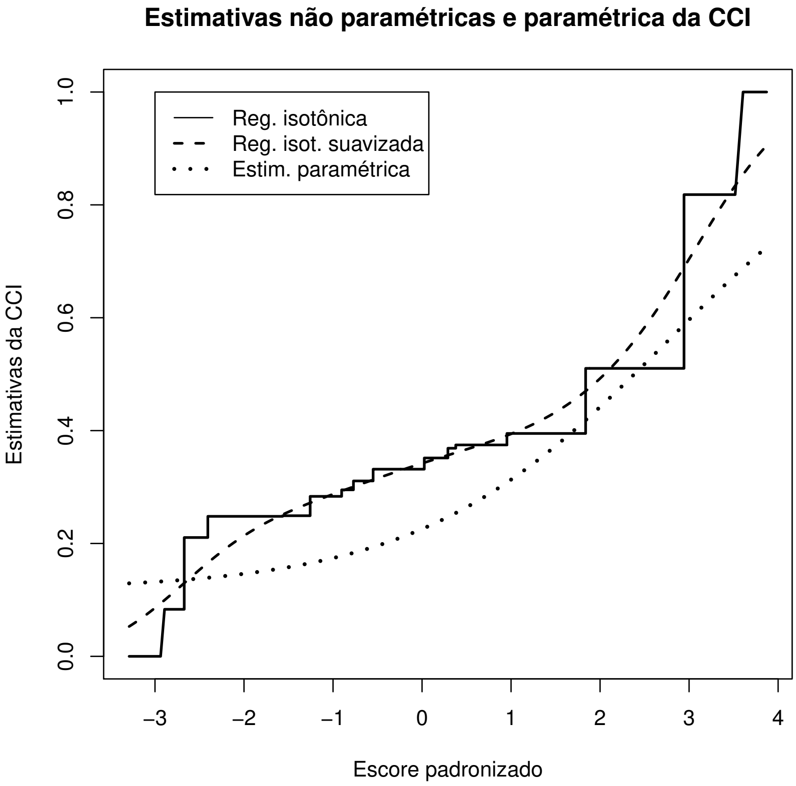


Figura 5.4: Estimativas não paramétricas e paramétrica da CCI de um item.

Podemos notar que essa curva não apresenta comportamento linear e que quanto maior for o escore padronizado do respondente, maior será a probabilidade de acerto. Devemos ressaltar também que cada curva dependerá exclusivamente dos parâmetros do seu item.

5.2.3 Análise para o Modelo de Resposta Gradual

Fizemos o mesmo para o Modelo de Reposta Gradual (MRG) para verificar se esse modelo é melhor que o de 3 parâmetros, foram estimados os parâmetros dos itens e geradas as respectivas curvas características dos itens.

Na tabela 5.3, são apresentadas algumas das medidas descritivas dos parâmetros dos itens estimados a partir do MRG. A partir disso, podemos observar que o valor mínimo para o parâmetro de discriminação (representado por Dscrmn) foi -0,88, o que não é esperado, pois, como já vimos, o valor do parâmetro de discriminação deve ser sempre maior que 0, e caso seja negativo, há algo de errado com o item. Porém, temos itens no último quartil que discriminam bem os indivíduos (Dscrmn>1), sendo o maior deles com o valor de 2,26.

Se tratando dos parâmetros de dificuldade dos itens (Dfcld1 e Dfcld2), podemos observar que há itens com categorias com baixa dificuldade, com o mínimo de -2,92, e com alta dificuldade, com máximo de 2,76. Esses valores resultaram um pouco fora do intervalo em que a literatura indica (-2 a +2). Entretanto, em média e boa parte dos valores desses parâmetros estão bastante próximos do intervalo.

Levando em consideração o obtido no modelo dicotômico de 3 parâmetros, podemos dizer que apesar de ambos possuírem itens que são pouco informativos e que talvez pudessem ser retirados da prova em questão, para o MRG, há mais itens com mais informação.

Tabela 5.3: Medidas descritivas dos parâmetros para o MRG

Medidas Descritivas	Dscrmn	Dfcld1	Dfcld2
Mínimo	-0.8823	-0.9634	-2.9245
1º Quartil	-0.0256	0.5273	-1.0133
Mediana	0.0059	1.0722	-0.5675
Média	0.1499	1.0160	-0.5519
3º Quartil	0.1581	1.5370	-0.1002
Máximo	2.2690	2.7682	1.6855

A Figura 5.5 ilustra a CCI do item 112, sendo a probabilidade de escolha do item errado representada pela curva azul, a de não resposta é a curva rosa, e a de escolha do item certo é a curva na cor verde. Podemos observar que indivíduos com habilidade até aproximadamente -2 tem mais probabilidade de errar o item, "escolhendo", portanto, a categoria -1. Já os indivíduos com habilidade aproximadamente de -2 a 0,9 tem maior probabilidade de não responder o item. Portanto, os indivíduos com habilidade acima de 0,9, aproximadamente, tem maior chance de acertar o item. Podemos verificar que o parâmetro de discriminação para este item é de 0,88, o que diz que ele é um item que tem boa discriminação, condizendo com o bom espaçamento entre as 3 curvas.

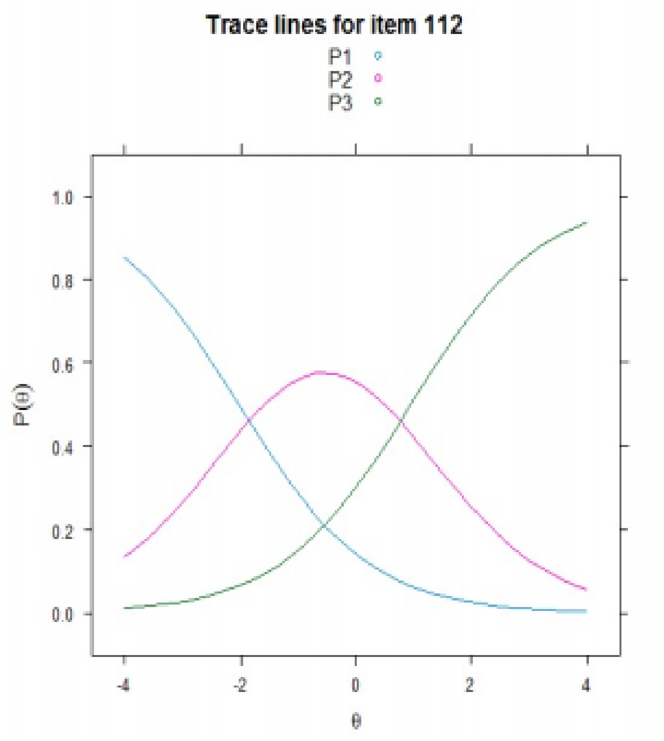


Figura 5.5: Curva Característica do Item 112

Na Figura 5.6, é apresentada a Curva de Informação do Teste (CIT), na qual podemos observar que o instrumento de medida tem maior informação para os valores da habilidade compreendidos entre aproximadamente -4,7 a 2,5, ou seja, isso indica que a prova é mais adequada para avaliar habilidades com valores que estão dentro desse intervalo, que, diga-se de passagem, é bem maior e possui forma adequada do que o encontrado na modelo dicotômico.

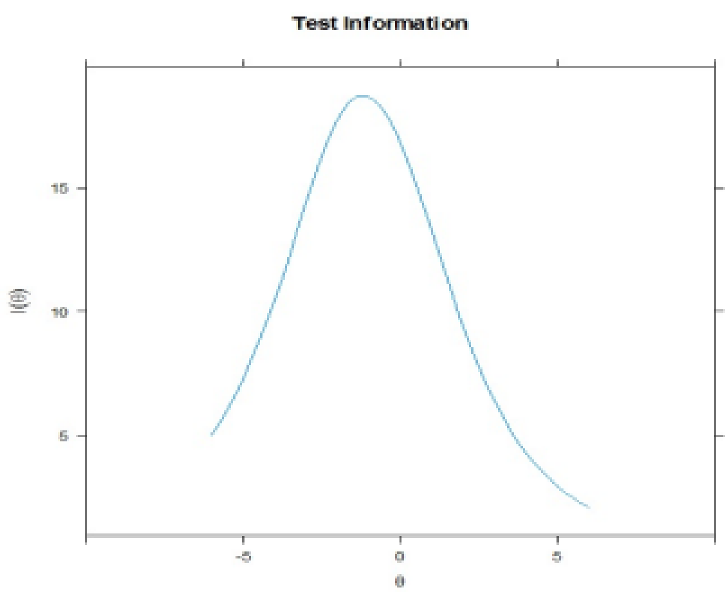


Figura 5.6: Curva de Informação do Teste

5.3 Comparação entre as notas das provas corrigidas pelo Método Convencional, pelo MRG e pelo Modelo Dicotômico de 3 parâmetros (Método Alternativo)

Conforme já dito anteriormente, o principal objetivo deste trabalho é calcular os escores dos indivíduos pelos 3 métodos e compará-los. Como citado na seção 3.6, para o MRG, acertar uma questão fácil é diferente de acertar uma questão difícil, o que não ocorre no método convencional. Dado isso, calculamos os escores para os indivíduos que responderam a prova em questão segundo o Método convencional, padronizando os mesmos a fim de comprar com o escore obtido pelo MRG e fizemos o seguinte gráfico para visualizar essa comparação.

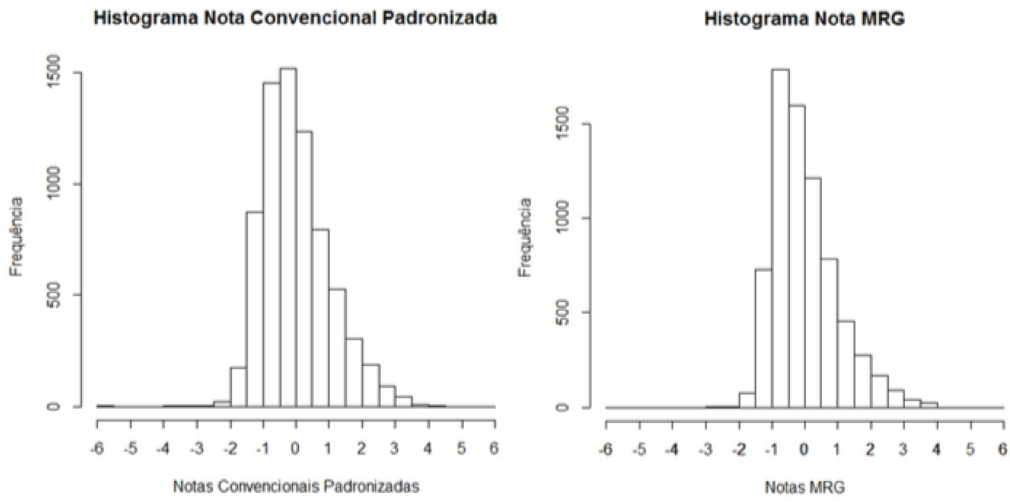


Figura 5.7: Histograma das Notas Convencionais Padronizadas e das Notas pelo MRG

Na Figura 5.7, são apresentados os histogramas das notas convencionais padronizadas e das notas do MRG, nele observamos que as notas convencionais padronizadas ficaram concentradas entre -1 e 0,5, e que as notas obtidas pelo MRG ficaram mais distribuídas próximas de -1 e 1. Para ter um melhor entendimento, plotamos o gráfico de dispersão entre essas notas, que notavelmente apresenta uma correlação linear alta (0,96) como vemos na figura 5.8.

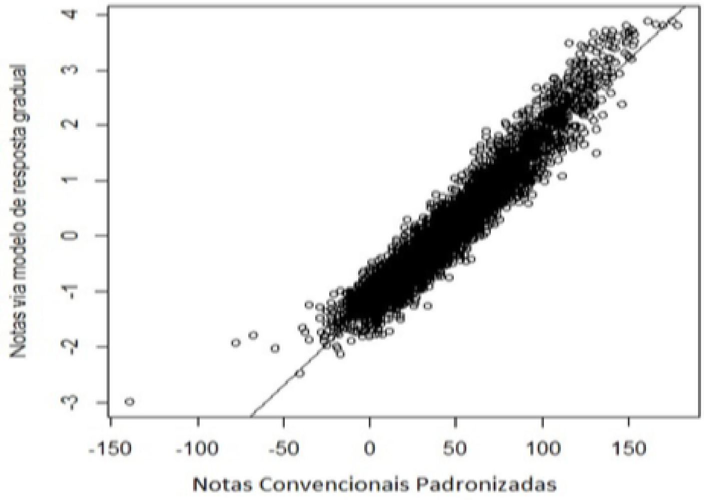


Figura 5.8: Gráfico de dispersão das Notas pelo método convencional e pelo MRG

Portanto, podemos concluir que a diferença nas notas mais elevadas e mais inferiores não aparentam ter tanta relevância ao se comparar as notas convencionais padronizadas e as notas obtidas pelo MRG.

Abaixo temos os histogramas das proficiências estimadas pelo modelo dicotômico e pelo modelo politômico.

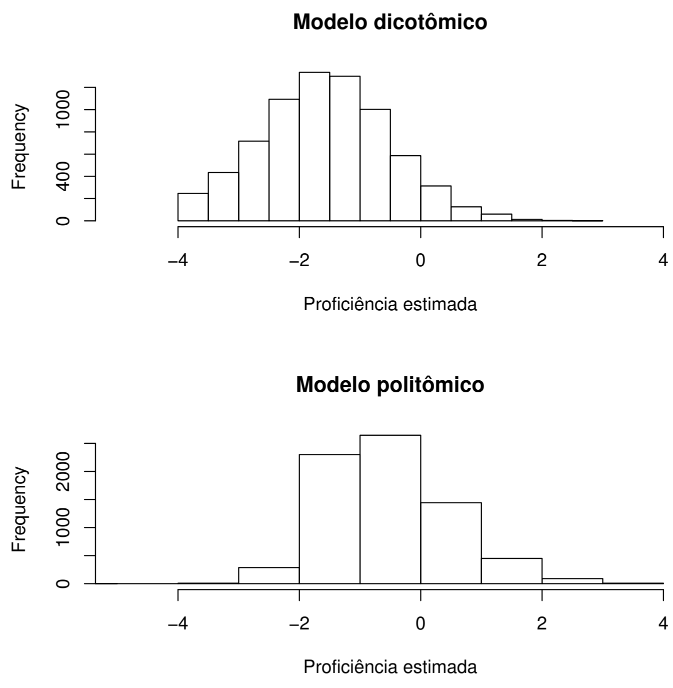


Figura 5.9: Histograma das proficiências estimadas

Percebemos por esses gráficos que a maior parte das proficiências estimadas estão no intervalo de -2 e -0.5, com algumas proficiências discrepantes em ambos, sendo o modelo politômico o que tem maior amplitude e maior média, o que já era esperado sendo que a não resposta considerada como erro aumenta a punição, consequentemente diminuindo a estimativa pelo modelo dicotômico.

Dessa maneira, já temos calculados os 3 escores objetivados nesse trabalho: As notas pelo método convencional, as notas pelo MRG e as notas pelo modelo dicotômico de 3 parâmetros. E o próximo passo é fazer a comparação das estimativas das proficiências dos 2 modelos calculados (dicotômico x politômico).

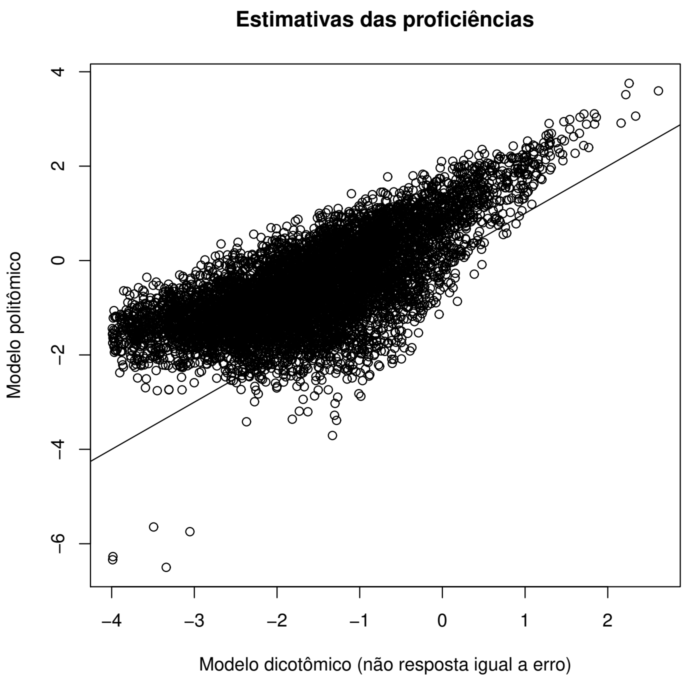


Figura 5.10: Gráfico de dispersão das proficiências estimadas pelos 2 modelos

Na figura acima observamos melhor tal comparação entre os modelos, e é notável que a estimativa é maior para o modelo politômico para a maior parte dos respondentes, além disso, percebemos que as estimativas têm uma boa correlação, com coeficiente de 0.71, o que nos diz que a correlação é positiva e forte.

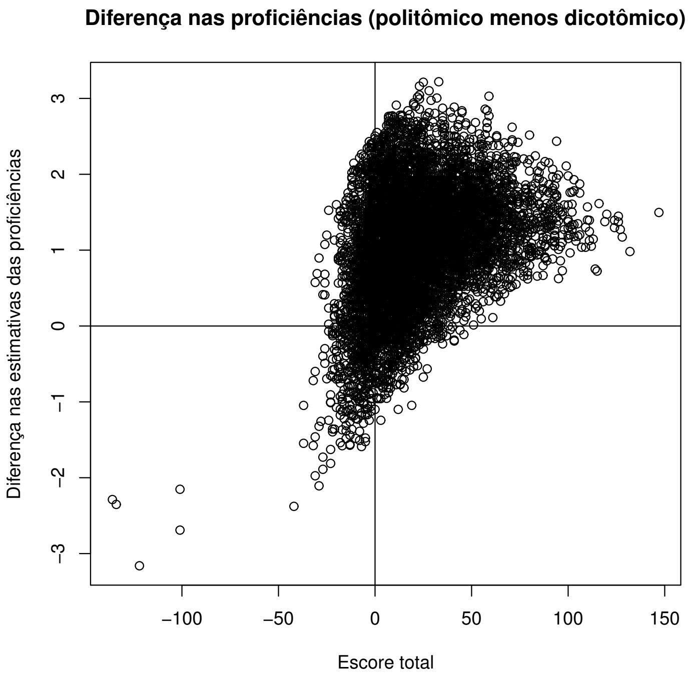


Figura 5.11: Gráfico de dispersão do escore total versus a diferença nas proficiências estimadas pelos 2 modelos

Na figura 5.11 temos um gráfico de dispersão entre a diferença das proficiências obtidas pelos 2 modelos (politômico menos o dicotômico) e o escore total obtido pelos indivíduos que realizaram a prova, no qual podemos perceber que há uma tendência de que o modelo politômico favorece os indivíduos com maior escore, ou seja, a presença da categoria não resposta faz com que os escores sejam aumentados, podemos ver que escores acima de 50 possuem diferença positiva nas estimativas.

6 Conclusão

Neste trabalho, foi realizada uma análise das notas de alunos que responderam à uma prova do vestibular da UnB, elaborada pelo CEBRASPE. Análise que tinha como base o cálculo das notas e comparação destas dentre os modelos e posteriormente verificar se a aglutinação da categoria "não respoder"teria efeito sobre o escore. Para fazê-lo, foi utilizado o banco de dados do próprio CEBRASPE, que continha originalmente mais de 20 mil indivíduos que responderam as 300 questões dispostas na prova em questão.

Analisando os dados das respostas dos indivíduos ficou claro que esta é uma prova que tem certa dificuldade dado que mais de 33% das respostas foram a não resposta. Foi verificado também, que alguns itens se mostraram não muito informativos para avaliar a discriminação do traço latente conhecimento em múltiplas áreas de estudo, mas essas inconsistências nas estimativas dos parâmetros dos itens podem ter sido causadas por termos considerado todos os itens das provas, sem uma separação por área. Porém, existem outros itens que demonstram ser bons no quesito de avaliação segundo os valores de seus parâmetros.

Feitos os cálculos dos escores, chegamos a conclusão que as notas obtidas pelo método convencional e pelo MRG estão bem próximas entre si, apresentando alto nível de correlação. Quando comparamos as estimativas segundo o MRG e o modelo dicotômico de 3 parâmetros, sendo que nesse último a não resposta foi considerada como erro, a correlação foi pouco maior que 0.7, resultando, após análise da figura 5.11, na conclusão de que o fator da não resposta quando não aglutinada ao erro, favorece os indivíduos com maior escore total.

Portanto, podemos concluir que os resultados finais demonstram que o modelo de resposta gradual poderia sim ser utilizado para o cálculo das notas dos itens do tipo A (itens dicotômicos) do Vestibular da UnB, porque, além de trazer resultados bem próximos do método convencional, ele pode comparar indivíduos de populações diferentes que realizaram provas ou testes que contenham itens em comum, além de poder analisar o traço latente de forma mais fidedigna, diferenciando se um indivíduo possui ou não maior conhecimento sobre o avaliado.

Referências

[1] Andrade, F. D., Tavares, H.R. e Valle, R. C - *Teoria de Resposta ao Item: Conceitos e Aplicações*, ABE, 2000, Caxambu

[2] Baker, F. - *Item Response Theory*, Marcel Dekker, 1992, New York.

[3] Baker, F.B., Kim, S-H. -*The Basics of Item Response Theory Using R*, 2017

[4] Demars, C. - *Item response theory*, 2010, New York.

[5] Guilford, J. P. - *Psychometric methods*, 1954, New York.

[6] Gulliksen, H. - *Theory of mental tests*, 1950, New York.

[7] Ostini, Nering - *Polytomous Item Response Models*, 2006

[8] Samejima, F. - *Estimation of latent ability using a response pattern of graded scores*, 1969

[9] van der Linden, Hambleton - *Handbook of Modern Item Response Theory*, 1997

[10] Brunk,H. D. , Barlow,R. E. , Bartholomew,D. J. , Bremner,J. M. - *Statistical Inference under Order Restrictions. (The Theory and Application of Isotonic Regression)*, 1972

[11] Robertson, T., Wright, F.T. and R.L. Dykstra. - *:Order Restricted Statistical Inference*, 1988

[12] Darrell Bock, R. Psychometrika - *:Estimating item parameters and latent ability when responses are scored in two or more nominal categories*, 1972

[13] Andrich, D. Psychometrika - *:A rating formulation for ordered response categories* , 1978.