



UNIVERSIDADE FEDERAL DO PARÁ

**FUNCIONAMENTO DIFERENCIAL DO ITEM NO MODELO  
MULTIDIMENSIONAL DA TRI: UMA ABORDAGEM BAYESIANA**

**Fádua Jorbelha Caldas Ferreira Antônio**

**DISSERTAÇÃO APRESENTADA  
AO  
PROGRAMA DE PÓS-GRADUAÇÃO  
EM  
MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE FEDERAL DO PARÁ  
PARA  
OBTENÇÃO DO GRAU DE MESTRE  
EM  
ESTATÍSTICA**

**Orientadora: Profa. Dra. Maria Regina Madruga Tavares**  
**Co-orientador: Prof. Dr. Héilton Ribeiro Tavares**

*Durante a elaboração deste trabalho a autora recebeu apoio financeiro do CNPq*

**Belém  
2017**

**Fádua Jorbelha Caldas Ferreira Antônio**

**FUNCIONAMENTO DIFERENCIAL DO ITEM NO MODELO  
MULTIDIMENSIONAL DA TRI: UMA ABORDAGEM BAYESIANA**

Este exemplar corresponde a redação  
da dissertação defendida por Fádua  
Jorbelha Caldas Ferreira Antônio.

Orientadora: **Profa. Dra. Maria Regina Madruga Tavares**

Co-orientador: **Prof. Dr. Héilton Ribeiro Tavares**

**Belém**

**2017**

Dados Internacionais de Catalogação-na-Publicação (CIP)  
UFPA - Biblioteca Central

Antônio, Fádua Jorbelha Caldas Ferreira, 1988-

Funcionamento diferencial do item no modelo multidimensional da TRI: uma abordagem bayesiana / Fádua Jorbelha Caldas Ferreira Antônio - 2017.

Orientadora: Maria Regina Madruga Tavares

Co-orientador: Héilton Ribeiro Tavares

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Matemática e Estatística, Belém, 2017.

1. Estatística matemática. 2. Teoria de Resposta ao Item. 3. Teoria bayesiana de decisão estatística. I. Título.

CDD 23 ed. 519.5

*Aos meus pais, Geraldo Antônio e Leonetti Caldas.*

Fádua Jorbelha Caldas Ferreira Antônio

**FUNCIONAMENTO DIFERENCIAL DO ITEM NO MODELO  
MULTIDIMENSIONAL DA TRI: UMA ABORDAGEM BAYESIANA**

Esta Dissertação, foi julgada e aprovada, para a obtenção do grau de Mestre em Estatística, no Programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará.

Belém, 29 de junho de 2017

João Marcelo Brazão Protázio

Prof. Dr. João Marcelo Brazão Protázio  
(Coordenador(a) do Programa de Pós-Graduação em Matemática e Estatística - UFPA).

**Banca Examinadora**

Maria Regina Madrugada Tavares  
Profa. Dra Maria Regina Madruga Tavares  
PPGME/UFPA  
**Orientador(a)**

Heliton Ribeiro Tavares  
Prof. Dr Heliton Ribeiro Tavares  
PPGME/UFPA  
**Co-orientador(a)**

Marinalva Cardoso Maciel  
Profa. Dra Marinalva Cardoso Maciel  
Faculdade de Estatística/UFPA  
**Examinador(a) Externo**

# Agradecimentos

À Deus por ter me dado saúde para nunca desistir.

À minha orientadora, Maria Regina Madruga Tavares, pela inestimável experiência, conselhos e paciência, desde a graduação e por mais esses dois anos da construção da presente dissertação e ao co-orientador Héilton Ribeiro Tavares pelas valiosas discussões e opiniões de grande valia para a minha formação e dessa dissertação.

Aos meus pais, Geraldo e Leonetti, pela personalidade, caráter e por todo apoio sentimental e financeiro. Amo vocês!

Aos colegas e amigos do PPGME, em especial Shirlaine Souza, Roger Noronha, Thiago Castro, Miguel Souza, Danilo Andrade, Thamara Rubia, pelos diversos artigos cedidos e discussões sobre otimização, entre outros assuntos não menos interessantes.

Aos meus amigos, que sempre compreenderam a importância do mestrado para mim. Em especial Anna Rafaella, Mariana Araújo, Isabella Couto Reis, Wecsley Prates, que mesmo longe me deram apoio.

Finalmente, gostaria de agradecer à UFPA pelo ensino gratuito de qualidade, ao PPGME e à CAPES, sem os quais essa dissertação dificilmente poderia ter sido realizada e a todos mais que eu não tenha citado nesta lista de agradecimentos, mas que de uma forma ou de outra contribuíram não apenas para a minha dissertação, mas também para eu ser quem eu sou.

*”Existir é sobreviver a escolhas injustas.”*

*The OA*

*”As nuvens mudam sempre de posição, mas são sempre nuvens no céu. Assim devemos ser todo dia, mutantes, porém leais com o que pensamos e sonhamos; lembre-se, tudo se desmancha no ar, menos os pensamentos.”*

*Paulo Beleki*

---

# Resumo

---

ANTÔNIO, Fádua Jorbelha Caldas Ferreira. Funcionamento Diferencial do Item no Modelo Multidimensional da TRI: Uma Abordagem Bayesiana. 2017. Dissertação para obtenção do grau de mestre em Estatística, UFPA, Belém - Pará, Brasil.

Neste trabalho pretende-se apresentar o Funcionamento Diferencial do Item (DIF) como uma importante ferramenta estatística na identificação de itens que sejam favoráveis a determinado grupo, em detrimento de outros usando o Modelo Multidimensional Logístico de 3 parâmetros da Teoria de Resposta ao Item (TRIM) via Abordagem Bayesiana. Um item apresenta DIF quando indivíduos de diferentes grupos que possuem o mesmo traço latente não têm a mesma probabilidade de acertarem o item. Ajustou-se o Modelo Multidimensional Logístico de 3 parâmetros com inclusão dos parâmetros de DIF. Todas as etapas do algoritmo e respectiva fundamentação matemática apresentam-se e ilustram-se com recurso a computação desenvolvida em R. Para testar o algoritmo proposto utilizam-se dados simulados, considerando que o traço latente afere 2 dimensões.

**Palavras-chave:** Estatística matemática. Teoria de Resposta ao Item. Teoria bayesiana de decisão estatística.



---

# Abstract

---

ANTÔNIO, Fádua Jorbelha Caldas Ferreira. Bayesian Analysis on the Differential Item Functioning in the Multidimensional Model of IRT. 2017. Dissertation presented for obtaining the master's degree in Statistics, UFPA, Belem - Para, Brazil.

In this work we present the Differential Function of the Item (DIF) as an important statistical tool in the identification of items that are favorable to a certain group over others ones using the Multidimensional 3 Parameters Logistic Model of Item Response Theory, via a Bayesian Approach. An item displays DIF when individuals from different groups that have the same latent trait are not as likely to give a correct response to the item. The 3 Parameter Logistic Multidimensional Model was adjusted with inclusion of DIF parameters. All the steps of the algorithm and its mathematical foundation are presented and illustrated using the computation developed in R. To test the proposed algorithm we used simulated data, considering that the latent trait measure only 2 dimensions.

**Keywords:** Mathematical Statistics, Item Response Theory, Bayesian Estimation.

---

# Sumário

---

<b>Resumo</b>	<b>viii</b>
<b>Abstract</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>Lista de Figuras</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Aspectos gerais . . . . .	1
1.2 Elaboração de Itens . . . . .	2
1.3 Justificativa e Importância da Dissertação . . . . .	4
1.4 Objetivos . . . . .	5
1.4.1 Objetivo Geral . . . . .	5
1.4.2 Objetivos Específicos . . . . .	5
1.5 Sumário da Dissertação . . . . .	5
<b>2 DIF no Modelo Logístico Unidimensional da TRI</b>	<b>6</b>
2.1 Introdução . . . . .	6
2.2 Funcionamento Diferencial do Item . . . . .	9
2.3 Modelo para detecção de DIF considerando unidimensionalidade no traço latente . . . . .	13
<b>3 DIF no Modelo Multidimensional da TRI</b>	<b>16</b>
3.1 Introdução . . . . .	16
3.2 Modelo para DIF Multidimensional . . . . .	18
3.3 Estimação bayesiana dos parâmetros . . . . .	20
3.3.1 Função de Verossimilhança . . . . .	21
3.3.2 Distribuições <i>a priori</i> . . . . .	22
3.3.3 Distribuição <i>a posteriori</i> conjunta . . . . .	23
3.3.4 Equações de estimação dos parâmetros dos itens . . . . .	24
3.4 Adequação do modelo multidimensional . . . . .	27
<b>4 Dados Simulados do DIF no Modelo Multidimensional da TRI</b>	<b>28</b>
4.1 Introdução . . . . .	28
4.2 Aplicação a dados simulados . . . . .	28

4.2.1 Algoritmo . . . . .	28
4.3 Estimativas via Inferência Bayesiana . . . . .	28
4.3.1 Definição das <i>prioris</i> . . . . .	32
<b>5 Conclusões e Trabalhos Futuros</b>	<b>36</b>
5.1 Considerações finais . . . . .	36
5.2 Recomendações para trabalhos futuros . . . . .	37
<b>Bibliografia</b>	<b>38</b>

---

# Lista de Tabelas

---

4.1	Valores adotados para os parâmetros dos itens na simulação dos dados. . .	31
4.2	Estimativas médias dos parâmetros dos itens para os dados simulados, via método bayesiano. . . . .	33
4.3	Desvios-padrão das estimativas bayesianas dos parâmetros dos itens para os dados simulados. . . . .	34
4.4	Erros na estimação bayesiana dos parâmetros dos itens para os dados simulados. . . . .	35

---

# Lista de Figuras

---

2.1	Exemplo de uma Curva Característica do Item. . . . .	8
2.2	Representação de um item com DIF uniforme. . . . .	12
2.3	Representação de um item com DIF não-uniforme. . . . .	13

---

# Capítulo 1

## Introdução

---

### 1.1 Aspectos gerais

A Teoria de Resposta ao Item (TRI) surgiu a partir dos trabalhos de Lord (1952) e Rasch (1960), propondo modelos para estimar traços (ou características) latentes de indivíduos a partir das respostas dadas por eles a um conjunto de itens relacionados com o traço latente de interesse. Traços latentes são quantidades (variáveis) que não podem ser medidas diretamente, tais como: intensidade de uma dor, nível de estresse, proficiência (conhecimento) em alguma área, qualidade de um produto, nível sócio-econômico, etc...

A TRI vem sendo largamente utilizada em diferentes áreas, tais como Psicometria, Avaliação Educacional, Saúde, Marketing, entre outras. A aplicação mais conhecida da TRI é nas avaliações educacionais de grande escala, em que um número grande de alunos precisam ser avaliados em algumas áreas do conhecimento, a fim de se ter um diagnóstico do sistema educacional em que os mesmos estão inseridos. Nessas avaliações, os alunos respondem provas contendo itens (questões) relacionados a conteúdos de disciplinas da área do conhecimento sob avaliação, e suas proficiências (nível de conhecimento) ou habilidades são estimadas.

A idéia básica da TRI é considerar os itens como elemento central da análise, diferentemente da Teoria Clássica do Teste (TCT) que baseia-se no escore total obtido pelo aluno, e tem como ênfase o instrumento de medida (teste, prova, questionário) como um todo, ou seja, as análises e interpretações estão associadas sempre ao instrumento. Uma das vantagens da TRI, em comparação à TCT, é possibilitar a construção de vários testes constituídos por alguns itens exclusivos e alguns itens em comum, e assegurar que o nível de dificuldade de todos os testes seja aproximadamente igual. Outra vantagem é possibilitar o cálculo do traço latente de um indivíduo que não responde a todos os itens, ou seja, na presença de dados faltantes, sem a necessidade de utilizar técnicas para respostas *missing*.

A TRI já é utilizada em avaliações educacionais de vários países, onde os programas de avaliação mais conhecidos são: o National Assessment of Educational Progress (NAEP), o Educational Testing Service (ETS) e o Graduate Management Admission Test (GMAT), todos nos Estados Unidos da América, GAOKAO na China, o Programa Internacional de Avaliação de Estudantes (PISA), desenvolvido e coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE), com a participação de 72 países.

No Brasil, um dos principais programas de avaliação que utiliza a TRI é o SAEB, que desde sua criação em 1995 a utiliza para estimar a proficiência em Matemática e em Português dos alunos do Ensino Fundamental e Médio das escolas públicas brasileiras, inicialmente através de amostragem do universo desses alunos. A partir de 2005 este programa buscou avaliar de forma censitária toda a rede pública do país, mas apenas no Ensino Fundamental, com aplicações no 5º e 9º Anos. Outra avaliação de extrema importância no Brasil e que utiliza a TRI é o ENEM (Exame Nacional do Ensino Médio), reformulado a partir de 2009 de forma a garantir a comparabilidade das notas entre diferentes edições.

O ENEM tem por objetivo avaliar o desempenho dos estudantes egressos do Ensino Médio, e é aplicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), órgão vinculado ao Ministério da Educação (MEC). O ENEM acontece uma vez no ano e mobiliza milhões de estudantes de todo o país, dentre os quais alguns estão terminando o ensino médio e outros já concluíram. Inscrevem-se no certame alunos da rede pública e da rede privada do país, realizando as provas em dois dias. No primeiro dia respondem 90 itens, sendo 45 de Ciências da Natureza e 45 de Ciências Humanas, e no segundo dia devem responder 45 itens de Linguagens e 45 itens de Matemática, além de elaborar uma redação.

Por todo esse trabalho será considerado o uso dos modelos da TRI na área de avaliação educacional, em que os itens são questões de prova, e o traço latente de interesse será a proficiência (nível de conhecimento) dos alunos.

## 1.2 Elaboração de Itens

Nos modelos estatísticos usuais da TRI cada item tem suas próprias características, representadas por 1, 2 e/ou 3 parâmetros: o *parâmetro de discriminação*  $a$ , que mede o poder do item em discriminar os indivíduos de alto e baixo traço latente, o *parâmetro*

de dificuldade  $b$ , que mede o grau de dificuldade do item, e o *parâmetro de acerto casual*  $c$ , que mede a probabilidade de acerto ao item por indivíduos de baixo traço latente. Esses modelos podem ter apenas 1 parâmetro (dificuldade), ou 2 parâmetros (dificuldade e discriminação) ou os 3 parâmetros.

Ao se ajustar esses modelos, além de estimar as proficiências  $\theta$  para cada aluno deseja-se, também, estimar esses parâmetros dos itens, o que é conhecido como *calibração dos itens*. Nesse processo de estimação é adotada uma escala (ou métrica) para os valores de  $\theta$ , e o parâmetro  $b$  também assume valores nessa escala. Assim, para um determinado nível de proficiência estimado para um aluno, é possível identificar nessa escala qual(is) o(s) conteúdo(s) de domínio desse aluno.

Na área de avaliação educacional, a elaboração desses itens segue critérios técnicos rigorosos e devem ser construídos com base na *matriz de referência* do exame. Por exemplo, a matriz de referência do ENEM (ver [www.inep.gov.br](http://www.inep.gov.br)). Os elaboradores de itens são especialistas na respectiva área do conhecimento a ser avaliada e são treinados para seguir os critérios definidos. Após a etapa de elaboração, os itens são revisados tecnicamente por um revisor *ad hoc*, que avalia e atesta a qualidade técnico-pedagógica do item, podendo aprovar ou não as questões analisadas. Assim, o item pode voltar ou não para o elaborador fazer as alterações.

Vencida a etapa de revisão de conteúdo, o item irá para uma etapa de revisão de forma para se adequar aos padrões de apresentação de itens, com posterior reapresentação ao elaborador para verificar se o sentido do item se mantém. Com a aprovação do elaborador, o item vai para a etapa de pré-teste, onde ele comporá testes aplicados a um número razoável de respondentes (geralmente em torno de 400 para resultados preliminares ou acima de 1000 para calibração dos itens), em que uma parte deles terá o aval para compor os testes principais.

O INEP, a partir de pré-testagens nacionais e de algumas das aplicações iniciais do ENEM, obteve os valores dos parâmetros  $a$ ,  $b$  e  $c$  de milhares de itens, montando assim um *banco de itens* para cada uma das quatro áreas avaliadas. A partir do conhecimento dos valores desses parâmetros as questões são posicionadas na escala (régua) permitindo assim sua interpretação pedagógica. O posicionamento de cada questão nos níveis da escala (régua) se dá a partir de critérios probabilísticos que garantem que, somente participantes com proficiência igual ou maior do que aquele nível possuem alta probabilidade de responder corretamente as questões que estão naquele nível e em níveis inferiores.



## 1.3 Justificativa e Importância da Dissertação

Apesar de todo o cuidado técnico na elaboração desses itens há uma preocupação crescente nas avaliações de grande escala, que envolvem alunos de diferentes regiões geográficas, com diferentes características culturais, econômicas e sociais, sendo submetidos a uma mesma prova, ou seja, respondendo um mesmo conjunto de itens. Será que as características desses itens é a mesma para esses diferentes grupos culturais e/ou sociais? Um determinado item pode ser mais difícil para alunos de uma região do que para alunos de outra região geográfica, mesmo que estes dois grupos tenham o mesmo nível de proficiência?

Diante de respostas afirmativas aos questionamentos acima, surge a necessidade de estudar formas de avaliar esse problema, uma vez que as estimativas das proficiências desses alunos e dos parâmetros destes itens podem ter algum viés (CAMILLI & SHEPARD, 1994). Na literatura, esse desempenho distinto do item entre 2 ou mais grupos é chamado de Funcionamento Diferencial do Item (DIF), a sigla adotada vem do termo em inglês *Differential Item Functioning*. Portanto, a identificação de itens com DIF é de grande importância no ajuste de modelos da TRI, pois essa diferença sistemática pode comprometer toda a inferência realizada, violar processos seletivos que dependam das notas dessas avaliações, levar ao planejamento de políticas educacionais equivocadas, entre outros problemas.

Portanto, a existência de DIF em um item significa que dois indivíduos, pertencentes a subgrupos distintos da população estudada, que tenham o mesmo nível de traço latente, tem probabilidades distintas para dar certa resposta para este item, quando seria esperado que a probabilidade fosse igual. Vários fatores podem estar associados à presença de DIF em testes educacionais. Segundo Gonçalves (2006):

Estudos conduzidos pelo Educacional Testing Service (ETS), nos Estados Unidos, apontam que o DIF, num contexto de avaliação em larga escala, pode ser causado por três fatores: a familiaridade com o conteúdo do item, que pode ser associada à exposição ao tema ou a um fator cultural; o interesse pessoal naquele dado conteúdo e a reação emocional negativa provocada pelo conteúdo.

A detecção e mensuração do DIF é essencial nas avaliações, uma vez que poderá contribuir com a identificação de diferenças nas ênfases dadas aos conteúdos em diferentes

regiões, municípios, e/ou escolas. O DIF pode ocorrer não apenas em extratos geográficos, mas também em diferentes gêneros, etnias, níveis sociais, entre outros.

## 1.4 Objetivos

### 1.4.1 Objetivo Geral

O objetivo principal desta dissertação é o de apresentar e discutir com detalhes a detecção do DIF via Estimação Bayesiana em modelos multidimensionais da TRI.

### 1.4.2 Objetivos Específicos

- i) Propor um modelo multidimensional da TRI que considere a existência do DIF;
- ii) Implementar um processo de estimação Bayesiana no modelo proposto;
- iii) Realizar estudos de simulação para analisar a eficácia do modelo proposto e da metodologia utilizada no seu ajuste.

## 1.5 Sumário da Dissertação

Este trabalho encontra-se dividido em 5 capítulos, a saber:

- No Capítulo 1 é feita uma introdução aos conceitos de TRI e DIF, são abordados os aspectos gerais, justificativa e importância do trabalho, os objetivos geral e específicos, e o sumário da dissertação.
- No Capítulo 2 é feita uma revisão bibliográfica da modelagem do DIF em modelos unidimensionais da TRI.
- No Capítulo 3 são apresentados estudos do DIF no modelo multidimensional da TRI que incorpore a detecção de DIF na discriminação e na dificuldade do item, propondo-se um modelo e uma abordagem Bayesiana no seu processo de estimação.
- No Capítulo 4 o modelo proposto no Capítulo 3 é avaliado, através de resultados obtidos com sua implementação em dados simulados.
- No Capítulo 5 serão apresentadas as considerações finais e recomendações para trabalhos futuros.

---

## Capítulo 2

# DIF no Modelo Logístico Unidimensional da TRI

---

### 2.1 Introdução

Este capítulo dedica-se a uma revisão bibliográfica sobre o Funcionamento Diferencial do Item (DIF) no modelo logístico unidimensional de 3 parâmetros, com a inclusão de dois parâmetros para representarem o DIF no parâmetro de discriminação e no de dificuldade do item. Basicamente, serão abordados alguns conceitos sobre TRI e DIF, o modelo proposto para detecção do DIF e a estimação dos parâmetros via abordagem Bayesiana (GONÇALVES, 2006).

Segundo Araujo et al. (2009), os primeiros modelos para traços latentes foram apresentados nos trabalhos de Lawley (1943), Guttman (1950) e Lazarsfeld (1950), mas cuja formulação teórica era ainda introdutória. Eram modelos em que se considerava que uma única habilidade, de um único grupo, estava sendo medida por um teste onde os itens eram corrigidos de maneira dicotômica. Foi com as publicações de Frederic Lord, na década de 50, que a TRI nasceu como uma área de pesquisa, pois ele deu início ao desenvolvimento formal e contribuiu com a criação de programas computacionais fundamentais para colocar essa teoria em prática. Lord (1952) foi o primeiro a desenvolver o modelo unidimensional de 2 parâmetros de natureza acumulativa para respostas dicotômicas, ainda baseado na distribuição normal (ogiva normal). Entretanto, o próprio Lord sentiu a necessidade da incorporação de um parâmetro que tratasse do problema do acerto casual, desenvolvendo o atual modelo logístico de 3 parâmetros. Alguns anos mais tarde, Birnbaum (1968) forneceu uma contribuição muito importante a estes modelos, ao sugerir a substituição, em ambos os modelos, da função ogiva normal pela função logística, matematicamente mais conveniente (ANDRADE, TAVARES & VALLE, 2000).

Dentre os modelos para itens com respostas dicotômicas ou itens de múltiplas es-

colhas, corrigidos de forma dicotômica, destaca-se o Modelo Logístico de 1 Parâmetro (ML1P), também conhecido como modelo de Rasch. Utiliza-se esse modelo quando não existe a possibilidade de acerto casual e todos os itens tem o mesmo poder de discriminação. Uma extensão natural é o Modelo Logístico de 2 Parâmetros (ML2P), utilizado quando não existe a possibilidade de acerto ao acaso, mas as discriminações podem diferir entre os itens. Ele teve sua forma explícita estabelecida por Lord (1952).

Esses dois modelos podem ser facilmente obtidos a partir de um caso mais geral, que é o Modelo Logístico de 3 Parâmetros (ML3P), atualmente o mais utilizado em avaliações educacionais, graficamente apresentado na Figura 2.1 e cuja expressão é dada por

$$P(Y_{ji} = 1 | \theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (2.1)$$

com  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, n$ . Neste modelo,  $Y_{ij}$  é uma variável dicotômica que assume o valor 1, com probabilidade dada em (2.1), quando o indivíduo  $j$  acerta o item  $i$ , ou 0, em caso contrário. A quantidade  $\theta_j$  representa o traço latente do  $j$ -ésimo indivíduo,  $a_i$  é o parâmetro de discriminação do item  $i$ ,  $b_i$  é o parâmetro de dificuldade do item  $i$ ,  $c_i$  é o parâmetro do item que representa a probabilidade de indivíduos com baixa habilidade responderem corretamente o item  $i$  e  $D$  é um fator de escala igual a um ou 1,7 (quando se deseja valores próximos da ogiva normal).

Assim, a expressão em (2.1) representa a probabilidade do indivíduo  $j$ , com traço latente  $\theta_j$ , acertar o item  $i$ , e é chamada de *função de resposta do item*. Sua representação gráfica é denominada “Curva Característica do Item” (CCI), ela tem uma forma de “S” tal que quanto maior o valor do traço ou habilidade  $\theta_j$ , maior é a probabilidade de acertar o item.

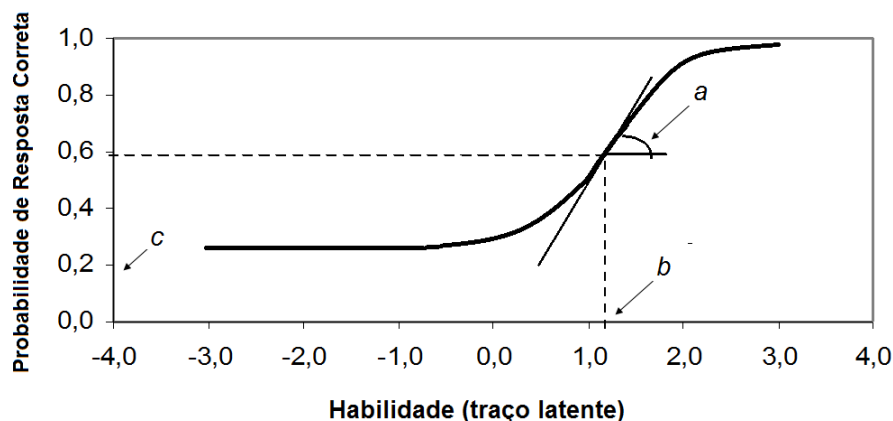
O parâmetro  $a_i$  controla a inclinação da curva em torno do  $b_i$ . Assim, itens com  $a_i$  negativo não são esperados sob esse modelo, uma vez que indicariam que a probabilidade de responder corretamente o item diminui com o aumento da habilidade. Quanto maior  $a_i$ , maior o nível de discriminação do item (maior será a inclinação máxima da curva) fazendo com que esse item discrimine melhor os indivíduos em dois grupos: os que possuem habilidade abaixo do valor do parâmetro  $b_i$  e os que possuem habilidade acima do valor do parâmetro  $b_i$ . O parâmetro  $b_i$  é medido na mesma escala da habilidade, e representa a habilidade necessária para uma probabilidade de acerto igual a  $(1 + c_i)/2$ . Assim, quanto maior o valor de  $b_i$ , mais difícil é o item, e vice-versa (ANDRADE, TAVARES & VALLE, 2000).

Pode-se observar que no modelo (2.1) a probabilidade de acerto tende a zero quando a habilidade do indivíduo tende a  $-\infty$ , ou seja,  $\lim_{\theta_j \rightarrow -\infty} P(Y_{ij} = 1 | \theta_j, a_i, b_i, c_i) = 0$ . O parâmetro  $c_i$  representa a probabilidade de um indivíduo com baixa habilidade responder corretamente ao item, e como tal, assume valores entre 0 e 1. Ele é muitas vezes referido como a probabilidade de acerto ao acaso ou probabilidade de “chute”. Quando não é permitido “chutar”,  $c_i$  é igual a 0.

Há duas suposições básicas para a aplicação dos modelos usuais da TRI: a unidimensionalidade e a independência local. Por unidimensionalidade entende-se que o teste esteja medindo apenas um traço latente, que pode representar uma proficiência ou uma composição de proficiências dos indivíduos, responsável pela realização de todos os itens do teste. Por independência condicional entende-se que a dependência entre os itens é perfeitamente explicada pelo traço latente dos indivíduos, ou seja, para um dado traço latente as respostas aos diferentes itens são independentes.

Nos modelos da TRI adota-se uma escala arbitrária para as habilidades, usualmente oriunda de uma distribuição Normal com média 0 e desvio padrão 1. Segundo Andrade et al. (2000) o importante são as relações de ordem existentes entre os pontos dessa escala e não necessariamente sua magnitude. A Figura 2.1 é um exemplo de CCI, e representa graficamente o ML3P apresentado em (2.1), tem forma de “S” com inclinação e deslocamento na escala da habilidade definidos pelos parâmetros do item.

Figura 2.1 *Exemplo de uma Curva Característica do Item.*



Pode-se observar na Figura 2.1 que quando  $\theta = b_i$ , a probabilidade de se acertar o item é igual a  $0,6 = \frac{1+c_i}{2} = \frac{1+0,2}{2}$ , isto é, se a habilidade do indivíduo é maior que o

parâmetro de dificuldade do item, ele terá maior probabilidade de acertar do que errar esse item. Além disso, o ponto de maior inclinação da curva característica é quando  $\theta = b_i$  e o valor da inclinação nesse ponto é diretamente proporcional ao parâmetro  $a_i$ . Assim, quanto maior for a discriminação do item maior será a inclinação da sua curva característica e, conseqüentemente, maior será a diferença entre as probabilidades de acerto de indivíduos com diferentes proficiências, ou seja, maior será a capacidade do item de discriminar os indivíduos. Já o parâmetro  $b_i$  determina o ponto central da curva, resultando no ponto de inflexão da mesma. Ou seja, é um parâmetro de localização (locação), e quanto maior o valor de  $b_i$  mais à direita estará a curva característica do item  $i$ . Assim, um item com  $b = 2$  pode ser considerado difícil e um item com  $b = -2$  ser considerado fácil para um indivíduo com proficiência mediana, considerando uma distribuição normal padrão para a habilidade.

## 2.2 Funcionamento Diferencial do Item

Estudos visando identificar itens que sejam favoráveis a determinado grupo, em detrimento de outros, ganham destaque na comunidade acadêmica, pois contribuem para assegurar que os instrumentos de medida sejam imparciais. Historicamente, a preocupação com o DIF esteve vinculada ao desejo de que os itens fossem elaborados de forma que a probabilidade dos indivíduos dar uma certa resposta dependesse apenas do traço latente do indivíduo, e não fossem afetados por características culturais e econômicas dos grupos submetidos aos instrumentos de avaliação educacional. Segundo Camilli e Shepard (1994) os primeiros trabalhos nessa área surgiram em 1910.

A existência do DIF em um item ocorre quando este item se comporta de maneira diferente entre dois ou mais grupos com relação aos parâmetros de discriminação, dificuldade e, até mesmo, acerto casual. Ou seja, indivíduos com mesmo nível de habilidade e de grupos diferentes podem ter probabilidades distintas de acertar um item, quando seria esperado que essas probabilidades fossem iguais. Assim, a preocupação com DIF está no fato de não se considerar o DIF quando este existe, levando a obtenção de resultados errôneos com relação aos parâmetros dos itens e, possivelmente, com relação às proficiências dos indivíduos também.

Segundo Dorans e Holland (1993) é importante distinguir entre o impacto e o DIF (ou viés) de um item. O impacto de um item ocorre quando a proporção de acertos ao

item é diferente entre os grupos, mas isso está relacionado à diferenças na distribuição do traço latente dos grupos. Já o DIF (ou viés) de um item caracteriza-se por um comportamento diferente do item condicionado a grupos de mesmo traço latente, mas diferentes com respeito a algum fator étnico, cultural ou sócioeconômico. Inicialmente, nos estudos de DIF, usava-se o termo *item bias (viés do item)*, mas devido ao momento altamente politizado em que estes estudos ocorriam, substituiu-se pelo termo *DIF - Differential item functioning* (ZUMBO, 2007).

Os estudos para desenvolver formas de identificação e mensuração do DIF nos itens tinham por objetivo detectar se os instrumentos de medida não possuíam nenhum tipo de viés, ou seja, se eram compostos por itens nos quais indivíduos de habilidades iguais, mas de culturas diferentes, não tinham a mesma probabilidade de acertar estes itens (ANGOFF, 1973; LINN et al., 1981; SHEPARD et al., 1993; IRONSON, 1993; LINN e DRASGOW, 1993). Muitos pesquisadores começaram a se dedicar ao estudo sistemático das diferenças entre os grupos étnicos, com o objetivo de tentarem encontrar explicações convincentes para as grandes diferenças de rendimento, observadas entre os diversos grupos étnicos e socioeconômicos, que refletiam, na realidade, disparidades nas oportunidades educacionais e se mostravam injustos, ao exigirem tarefas estranhas às culturas de algumas minorias.

A detecção de itens com DIF não é a única etapa nesses estudos, é importante também classificar o tipo de DIF e estudar possíveis fatores associados à sua ocorrência. Nesse contexto, é natural considerar modelos de regressão que associam à magnitude do DIF algumas covariáveis relacionadas aos itens (GONÇALVES, 2006). Essas covariáveis representariam possíveis fatores associados ao DIF, de tal forma que os resultados da análise de regressão permitiriam confirmar ou não a significância desses fatores para a ocorrência do DIF.

Existem métodos clássicos para considerar e identificar DIF, mas todos com alguma grave deficiência. Um dos mais conhecidos, que é também utilizado pelo *ETS - Educational Testing Service* nos Estados Unidos, baseia-se na estatística de Mantel-Haenszel (HOLLAND e THAYER, 1988; ANGOFF, 1993; DORANS e HOLLAND, 1993; ZWICK et al., 1999), e consiste na comparação das chances de acerto dos 2 grupos em análise, em  $M$  níveis da escala da habilidade, através da construção de  $M$  tabelas de contingência 2x2 contendo as proporções de acerto e erro ao item em cada grupo. As críticas a esse método consistem no fato da identificação do DIF ocorrer somente na dificuldade do item

e de considerar o escore total como traço latente, não utilizando nenhum modelo da TRI na sua estimação.

Dentre os métodos baseados em modelos da TRI, ou seja, que utilizam algum modelo da TRI para a estimação do traço latente, destacam-se: a comparação das curvas características do item (CCI) entre os grupos sob análise, testes de hipóteses para comparar os parâmetros dos itens entre os grupos e a regressão logística:

- *Comparação das CCI's*

A comparação se dá através da medida da área entre essas CCI's (RUDNER et al., 1980; LINN e HARNISCH, 1981; RAJU, 1988; WAINER, 1993). Para isso, o modelo da TRI é aplicado para cada grupo individualmente, gerando assim uma curva característica estimada para cada grupo. Este método apresenta problemas, pois fica difícil identificar o tipo de DIF (se na dificuldade, discriminação ou acerto casual). Para tentar resolver estes problemas, são feitas suposições de mesmo acerto casual para os diferentes grupos ou mesma discriminação. Uma crítica a este método é o fato de se ajustar o modelo da TRI separadamente para os grupos, não garantindo a mesma escala para as estimativas obtidas em cada grupo.

- *Comparação dos parâmetros*

Consiste na realização de testes de hipóteses para a comparação dos parâmetros estimados para cada item e em cada grupo individualmente (THISSEN et al., 1993; MUNIZ, 1997). As estatísticas utilizadas para o teste ou baseiam-se em distribuições normais assintóticas ou na distribuição qui-quadrado, por exemplo, o teste  $\chi^2$  de Lord (LORD, 1980). Esse método também recebe críticas por comparar quantidades que não estão na mesma escala.

- *Regressão logística*

Consiste em uma generalização do modelo logístico de dois parâmetros e, além de não considerar a possibilidade de acerto casual, supõe que as proficiências dos examinados são conhecidas (SWAMINATHAN e ROGERS, 1990; ZUMBO, 1999). A idéia é ajustar um modelo de regressão logística para a probabilidade de acerto ao item considerando como covariáveis o fator grupo e a interação entre o traço latente e o grupo. Outros autores propuseram incluir uma estrutura hierárquica no modelo

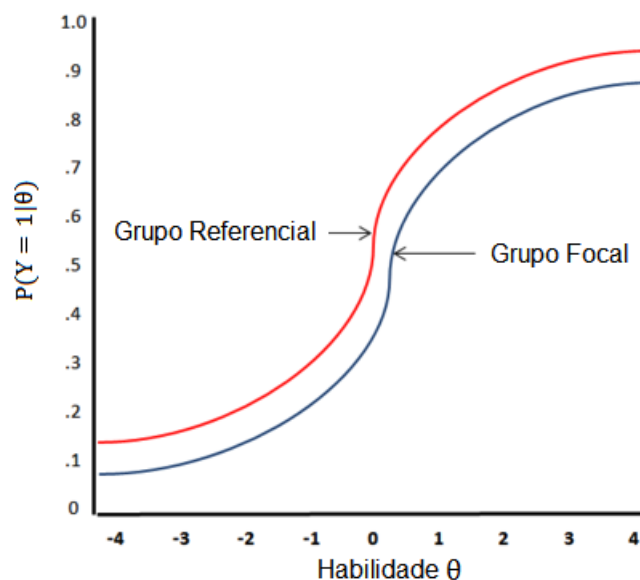


de regressão logística, considerando diferentes formas de aplicação dos testes, e covariáveis associadas ao item e aos respondentes (LONGFORD et al., 1993; ROGERS e SWAMINATHAN, 2000; SWANSON et al., 2002).

No contexto da TRI um item não tem DIF quando sua CCI é idêntica para os grupos, comparados em um mesmo nível ou magnitude da variável latente medida através do item (MELLENBERGH, 1989). Geralmente, os estudos para a determinação do DIF utilizam dois grupos, denominados de *grupo referência (GR)* e *grupo focal (GF)*. Os distintos métodos para a detecção do DIF foram desenvolvidos baseados em dois tipos de DIF.

O primeiro tipo é denominado *DIF uniforme ou consistente*, é quando as CCI's dos grupos de referência e focal são diferentes, porém não se cruzam. Em outras palavras, quando há uma vantagem relativa para um dos grupos investigados, e essa vantagem tem valor constante ao longo de todo o intervalo da escala de habilidades. Este caso ocorre quando o parâmetro de discriminação tem o mesmo valor nas duas CCI's, isto é, quando as CCI's são curvas paralelas (como mostrado na Figura 2.2).

Figura 2.2 Representação de um item com DIF uniforme.

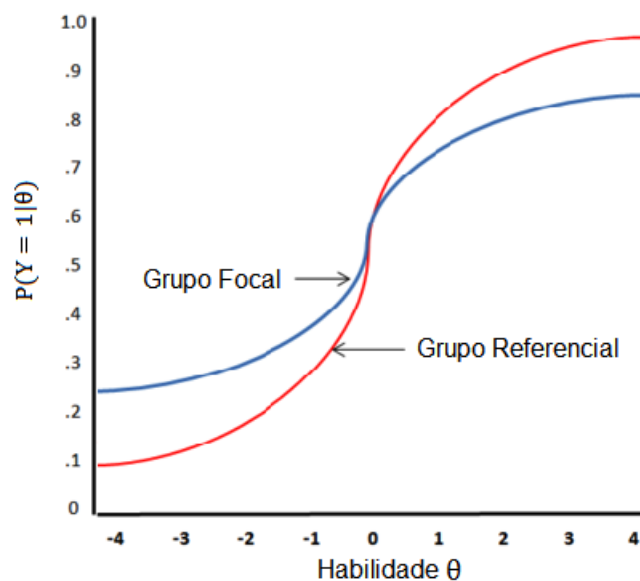


Fonte: Differential item functioning. (2016, December 18). In Wikipedia, The Free Encyclopedia. Acesso em 26/06/2017.

O segundo tipo de DIF é denominado *DIF não uniforme ou inconsistente*, e se observa quando as CCI's do item estudado, com respeito aos grupos de referência e focal, são diferentes e, ademais, se cruzam em algum ponto do intervalo da escala de habilidades

(ver Figura 2.3). Em outras palavras, quando há uma vantagem relativa para um dos grupos em um intervalo da escala da habilidade, mas essa vantagem é variável, com o outro grupo passando a ter a vantagem em outro intervalo. Este caso ocorre quando os parâmetros  $a$ ,  $b$  ou  $c$  têm valores distintos nas duas CCI's, isto é, quando as CCI's não são curvas paralelas.

Figura 2.3 *Representação de um item com DIF não-uniforme.*



Fonte: Differential item functioning. (2016, December 18). In Wikipedia, The Free Encyclopedia. Acesso em 26/06/2017.

## 2.3 Modelo para detecção de DIF considerando unidimensionalidade no traço latente

O modelo aqui apresentado foi proposto por Gonçalves (2006) e representa uma generalização do ML3P, com a inclusão de parâmetros para representar a presença de DIF na discriminação e na dificuldade do item. Ele assume que se conhece quais itens possuem DIF, considerando os demais itens como âncoras. Adicionalmente, ele propõe uma estrutura de regressão para esses parâmetros do DIF, tornando possível avaliar os fatores responsáveis pelo DIF ao mesmo tempo em que se estima o traço latente (diferentemente da proposta de May (2006)).

O modelo é dado por:

$$P(Y_{jig} = 1 | \theta_j, a_i, b_i, c_i, d_{ig}^a, d_{ig}^b) = c_i + (1 + c_i) \frac{1}{1 + e^{-De^{d_{ig}^a a_i (\theta_j - b_i + d_{ig}^b)}}}, \quad (2.2)$$

onde  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  e  $g = 1, \dots, G$ , com  $I$  representando o total de itens,  $n$  o total de indivíduos e  $G$  o total de grupos. Os parâmetros incluídos no modelo são  $\delta_{ig}^a$  e  $\delta_{ig}^b$ , e representam o funcionamento diferencial do item  $i$  no grupo  $g$  com relação a discriminação e a dificuldade, respectivamente. Os demais parâmetros são os mesmos já definidos no ML3P.

Note que se os novos parâmetros forem iguais a zero o modelo se reduz ao modelo logístico de três parâmetros com relação ao item  $i$ . Se  $d_{ig}^a = 0$  ( $d_{ig}^b = 0$ ), o item  $i$  não tem funcionamento diferencial com relação à discriminação (dificuldade). Para garantir a identificabilidade do modelo, é fixada uma distribuição normal padrão para as proficiências do grupo de referência (grupo 1) e supõe-se que as proficiências do grupo  $g$  têm distribuição normal com média  $\mu_g$  e variância  $\sigma_g^2$ .

Adicionalmente, para os itens com DIF Gonçalves (2006) propôs a seguinte estrutura de regressão para os parâmetros

$$d_{ig}^h = \gamma_{0g}^h + \sum_{k=1}^{K^h} \gamma_{kg}^h W_{(i)k+1}^h + \eta_{(i)g}^h, \quad h = a, b. \quad (2.3)$$

Os termos  $W_{(i)k+1}^h$  representam covariáveis associadas aos fatores que podem influenciar o DIF, e formam a matriz regressora  $W^h$ ,  $\gamma_{kg}^h$  é o coeficiente da variável preditora  $k$ , usada para explicar o DIF do parâmetro  $h$  no grupo  $g$ , e  $\eta_{(i)g}^h$  é o erro da regressão, com  $\eta_g^h \sim N_J(0, (\tau_g^h)^2 \mathbf{I}_{L^h})$ , onde  $\mathbf{I}_l$  é a matriz identidade de ordem  $l$  e  $L^h$  representa o número de itens para os quais se assume ter DIF no parâmetro  $h$ .

Assim, considerando (2.2) e (2.3), tem-se as seguintes notações para o modelo paramétrico:  $\lambda = (\lambda_1, \dots, \lambda_G)$ , com  $\lambda_g = (\mu_g, \sigma_g^2)$ ;  $d = (d^a, d^b)$ , com  $d^a = (d_{11}^a, \dots, d_{IG}^a)$  e de forma análoga para  $d^b$ ;  $\gamma = (\gamma^a, \gamma^b)$ , com  $\gamma^a = (\gamma_{02}^a, \dots, \gamma_{K^a G}^a)$  e de forma análoga para  $\gamma^b$ ;  $\tau^2 = ((\tau^a)^2, (\tau^b)^2)$ , com  $(\tau^a)^2 = ((\tau_2^a)^2, \dots, (\tau_G^a)^2)$  e de forma análoga para  $(\tau^b)^2$ ;  $\theta = (\theta_1, \dots, \theta_J)$ ;  $\beta = (a, b, c)$ , com  $a = (a_1, \dots, a_I)$ ,  $b = (b_1, \dots, b_I)$ ,  $c = (c_1, \dots, c_I)$ ; e os dados  $(Y, W)$ , com  $Y$  sendo o vetor de respostas e  $W = (W^a, W^b)$  a matriz de variáveis regressoras observadas.

Gonçalves (2006) apresenta em detalhes o processo de estimação dos parâmetros do modelo apresentado, sob uma abordagem Bayesiana, com o uso de métodos *MCMC - Markov Chain Monte Carlo*. Em particular, ele utiliza o *Gibbs Sampling* com passos de *Metropolis* (MULLER, 1991), uma vez que apenas algumas distribuições condicionais completas são conhecidas.

No próximo capítulo é proposta uma extensão do modelo apresentado e discutido acima (GONÇALVES, 2006), para a modelagem do DIF considerando multidimensionalidade no traço latente.

---

## Capítulo 3

# DIF no Modelo Multidimensional da TRI

---

### 3.1 Introdução

Segundo Fragoso (2008) a Teoria de Resposta ao Item Multidimensional (TRIM) pode ser considerada um tipo de análise fatorial para dados observados não contínuos. Os estudos teóricos na área de análise fatorial da TRIM, que impulsionaram seu desenvolvimento, iniciou-se a partir dos anos 70 fundamentados nos trabalhos de Paul Horst (1965), Roderick McDonald (1967), Bengt Muthén (1978), Fumiko Samejima (1974), entre outros. Esses modelos multidimensionais depende da natureza das dimensões do traços latentes, dos itens e da associação deles dentro do instrumento de medida. Além da análise fatorial, que normalmente é usada para encontrar um pequeno número de traços latentes capazes de explicar a variabilidade de um número maior de variáveis observadas (FRAGOSO, 2008), estes modelos baseiam-se, também, na extensão dos modelos de resposta ao item unidimensional. A TRIM envolve modelos onde as respostas aos itens podem está associadas a múltiplos traços latentes e a sua utilização é fundamental para identificar quantas dimensões são necessárias pra representar o nível de traço latente nessas diferentes dimensões.

Nos casos em que o instrumento de medida do traço latente é multidimensional, deve-se determinar quais itens pertencem a quais fatores. De acordo com Vieira (2016) há diferentes abordagens, a primeira delas é relacionada às análises exploratórias e confirmatórias. Na análise exploratória, todos os traços latentes podem ser relacionados a todos os itens. Em geral, quando se usa a análise exploratória, não se sabe quantas dimensões de fato o instrumento avaliativo está medindo. Para definir qual modelo melhor se ajusta aos dados é comum começar com o caso mais simples, o unidimensional, e, gradativamente, aumentar o número de dimensões. No entanto, apesar de normalmente o

modelo ficar melhor explicado ao se adicionar uma dimensão, ele fica mais complexo. Já na análise confirmatória, os traços latentes estão relacionados a um grupo de itens escolhidos previamente. Ela pode ser vista como um caso especial de uma exploratória em que os parâmetros dos itens são forçados a pertencer a uma das dimensões para representar as diferentes subescalas supostas a priori (BERNINI; MATTEUCCI; MIGNAI, 2015).

Outro modelo, considerado de análise fatorial confirmatória, é o bifatorial. Esse modelo possui um fator comum que explica as intercorrelações de todos os itens, e possui também fatores de cada grupo que tentam capturar a covariância dos itens que é independente da covariância devida ao fator comum (REISE; MORIZOT; HAYS, 2007). Em Gibbons e Hedeker (1992) tem-se que o modelo bifatorial força cada item  $i$  a ter um índice de discriminação diferente de zero na dimensão principal,  $a_{i1}$ , e um segundo índice,  $a_{ik}$ ,  $k = 2, \dots, K$ , em apenas um dos  $K - 1$  fatores, supondo que os dois primeiros estejam em uma subdimensão e os dois últimos em outra subdimensão, para quatro itens (VIEIRA, 2016).

Como base no trabalho de Reckase (2009), ainda temos os modelos chamados *compensatórios e não compensatórios*. De acordo com Vieira (2016) nos modelos compensatórios a probabilidade de resposta depende de uma soma ponderada dos traços latentes. Escores baixos em uma dimensão podem ser compensados por escores altos em outra dimensão para obter uma alta probabilidade de resposta.

McKinley & Reckase (1980) propõem o modelo logístico de três parâmetros com expressão dada por

$$P(X_i = 1 | \boldsymbol{\theta}, \mathbf{a}_i, d_i, c_i) = c_i + (1 - c_i) \frac{1}{1 + \exp[-\sum_{k=1}^p a_{ik} \theta_k + d_i]} \quad (3.1)$$

em que  $i = 1, \dots, I$ , sendo  $I$  o número total de itens. Os parâmetros dos itens são representados por  $\mathbf{a}_i, d_i, c_i$ , onde os  $a_{ik}$ ,  $k = 1, \dots, K$  e  $c_i$  são definidos como em (2.1), e  $d_i$  é um parâmetro relacionado à dificuldade geral do item.

Já os modelos não compensatórios permitem interações entre as variáveis latentes de modo que os escores devem ser elevados em todas as dimensões para manter uma probabilidade alta de resposta.

Whitely (1980) e Sympson (1977) propõem o modelo logístico não compensatório de três parâmetros dado por

$$P(X_{ij} = 1 | \boldsymbol{\theta}, \mathbf{a}_i, d_i, c_i) = c_i + (1 - c_i) \prod_{k=1}^p \frac{1}{1 + \exp(-D a_{ik} \theta_{jk} + d_{ik})} \quad (3.2)$$

## 3.2 Modelo para detecção de DIF considerando multidimensionalidade no traço latente

Na aplicação deste modelo, os traços latentes estão relacionados a um grupo de itens escolhidos previamente, definido como grupo de referência, para o qual os parâmetros do DIF são iguais a zero para todos os itens, e estima-se os parâmetros do DIF dos grupos restantes, chamados de grupos focais, para avaliar o funcionamento diferencial do item em cada grupo com relação ao grupo de referência.

O modelo multidimensional logístico de 3 parâmetros proposto nesta dissertação, para detecção de DIF nos parâmetros de discriminação e/ou dificuldade (intercepto) do item, é uma extensão do modelo logístico compensatório proposto por McKinley & Reckase (1980) que se baseia numa combinação linear das coordenadas dos traços latentes e considera que existem vários traços latentes, representando diferentes habilidades dos indivíduos por grupo, mas apenas um parâmetro de dificuldade. O modelo é dado por

$$P(Y_{jig} = 1 | \boldsymbol{\theta}_{jg}, \mathbf{a}_{ig}, d_{ig}, c_{ig}, \boldsymbol{\delta}_{ig}^a, \delta_{ig}^d) = c_{ig} + (1 - c_{ig}) \frac{1}{1 + e^{-\left\{ \sum_{k=1}^K (a_{igk} + \delta_{igk}^a) \theta_{jgk} + (d_{ig} + \delta_{ig}^d) \right\}}}, \quad (3.3)$$

onde  $i = 1, \dots, I$ , sendo  $I$  o número total de itens;  $j = 1, \dots, J$ , onde  $J$  representa o total de indivíduos;  $k = 1, \dots, K$ , onde  $K$  representa o número de dimensões do traço latente;  $g = 2, \dots, G$ , onde  $G$  representa o total de grupos. A variável  $Y_{jig}$  representa a resposta do indivíduo  $j$ , do grupo  $g$ , ao item  $i$  e  $\boldsymbol{\theta}_{jg} = (\theta_{jg1}, \dots, \theta_{jgK})$  é um vetor ( $1 \times K$ ) dos  $k$  traços latentes do  $j$ -ésimo indivíduo, do grupo  $g$ .

Em (3.3)  $P(Y_{jig} = 1 | \boldsymbol{\theta}_{jg})$  é a probabilidade de um indivíduo  $j$ , do grupo  $g$ , com vetor de traços latentes  $\boldsymbol{\theta}_{jg}$ , responder corretamente ao item  $i$  ( $Y_{jig} = 1$ ). O vetor  $\mathbf{a}_{ig} = (a_{ig1}, \dots, a_{igK})$  tem dimensão ( $1 \times K$ ) e contém os parâmetros de discriminação do item  $i$  nas várias dimensões, do grupo  $g$ . O parâmetro  $d_{ig}$  é um escalar associado indiretamente à dificuldade do item  $i$ , do grupo  $g$ . Normalmente ele é representado por  $d_{ig} = -b_{ig} \|\mathbf{a}_{ig}\| = -b_{ig} \sqrt{\sum_{k=1}^K a_{igk}^2}$  uma vez que nos modelos unidimensionais temos o equivalente  $d_{ig} = -a_{ig} b_{ig}$ , onde o parâmetro de dificuldade do item  $i$  é o ponto de inflexão da CCI e é medido na mesma escala do traço latente. Nos modelos multidimensionais,  $d_{ig}$  não pode ser interpretado como um ponto de inflexão da CCI, então faz-se um ajuste

nesse parâmetro para medir a distância de origem até o ponto de maior inclinação da superfície característica do item  $i$  (FRAGOSO, 2008).

Para detecção de itens com DIF são incluídos parâmetros para representar o funcionamento diferencial do item  $i$  nos parâmetros de discriminação e dificuldade onde,  $\delta_{i.k}^a = (\delta_{i1k}^a, \dots, \delta_{iGk}^a)$ , DIF no parâmetro de discriminação do item  $i$ , e  $\delta_{i.}^d = (\delta_{i1}^d, \dots, \delta_{iG}^d)$ , DIF no parâmetro de dificuldade do item  $i$ . De forma a manter a identificabilidade do modelo, adota-se o grupo 1 como referência (sem DIF), e o demais serão comparados a ele. Assim,

$$\delta_{i1k}^a = 0, \quad \forall k = 1, \dots, K, \quad \text{e} \quad \delta_{i1}^d = 0. \quad (3.4)$$

Note que se esses parâmetros que representam o DIF nos grupos forem todos nulos, o modelo se reduz ao modelo logístico de três parâmetros, ou seja, se para  $g \geq 2$  tivermos  $\delta_{igk}^a = 0, \forall k$ , e  $\delta_{ig}^d = 0$ , o item  $i$  não tem DIF com relação aos parâmetros de discriminação e dificuldade.

Para identificabilidade do modelo é fixada, também, uma distribuição normal multivariada padrão de dimensão  $K$ ,  $NM_K(\mathbf{0}, \mathbf{I}_K)$ , para os traços latentes do grupo de referência e supõe-se que os traços latentes do grupo  $g$  têm distribuição normal multivariada com vetor de médias  $\boldsymbol{\mu}_g = (\mu_{g1}, \dots, \mu_{gK})'$  e matriz de covariâncias  $\boldsymbol{\Sigma}_g$ , com dimensão  $K \times K$ . Ou seja,

$$\boldsymbol{\theta}_{g(j)} | \boldsymbol{\eta}_g \sim NM_K(\boldsymbol{\mu}_{g(j)}, \boldsymbol{\Sigma}_{g(j)}),$$

onde  $g(j)$  indica o grupo a que pertence o indivíduo  $j$  e  $\boldsymbol{\eta}_g = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ , para  $g = 2, \dots, G$ , além disso,  $\boldsymbol{\eta}_1 = (\mathbf{0}, \mathbf{I})$ .

Diferentemente da proposta de Gonçalves (2006), aqui não será considerada uma estrutura de regressão nos parâmetros do DIF. Um processo bayesiano de estimação para os parâmetros do modelo em (3.3) será apresentado na próxima seção.

Para avaliar os itens que sejam favoráveis a determinados grupos, em detrimento de outros, alguns testes de hipóteses podem ser elaborados para detectar a existência do DIF no item  $i$ . O interesse é verificar se há diferença entre as estimativas dos parâmetros de discriminação e dificuldade do item  $i$  para os diferentes grupos. Assim, a hipótese nula  $H_0$  é que não existe um funcionamento diferencial do item  $i$  entre grupo de referência e o grupo focal. É confrontada pela hipótese alternativa  $H_1$ , de que existe um funcionamento diferencial do item  $i$  entre os grupos.



Para testar o DIF no parâmetro de discriminação, testa-se as hipóteses

$$H_0 : \delta_{ig}^a = 0 \text{ versus } H_1 : \delta_{ig}^a \neq 0, \quad (3.5)$$

com  $i = 1, \dots, I$  e  $g = 2, \dots, G$ .

Para testar o DIF no parâmetro de dificuldade, testa-se as hipóteses

$$H_0 : \delta_{ig}^d = 0 \text{ versus } H_1 : \delta_{ig}^d \neq 0, \quad (3.6)$$

com  $i = 1, \dots, I$  e  $g = 2, \dots, G$ .

A decisão sobre a aceitação ou rejeição da hipótese  $H_0$  pode ser baseada no intervalo de credibilidade (ICr) construído para o parâmetro sob teste. No caso das hipóteses acima, pode-se construir o intervalo de credibilidade para cada diferença  $\delta_{ig}^a$  e  $\delta_{ig}^d$  e verificar se o valor zero pertence ou não ao intervalo, ou seja, se aceitamos ou rejeitamos  $H_0$ . Os intervalos para as diferenças são dados por

$$IC(\delta_{ig}^a; \gamma\%) = [\widehat{\delta}_{ig}^a - z_{\alpha/2}EP(\widehat{\delta}_{ig}^a); \widehat{\delta}_{ig}^a + z_{\alpha/2}EP(\widehat{\delta}_{ig}^a)] \quad (3.7)$$

e

$$IC(\delta_{ig}^d; \gamma\%) = [\widehat{\delta}_{ig}^d - z_{\alpha/2}EP(\widehat{\delta}_{ig}^d); \widehat{\delta}_{ig}^d + z_{\alpha/2}EP(\widehat{\delta}_{ig}^d)]. \quad (3.8)$$

### 3.3 Estimação bayesiana dos parâmetros

Para o processo de estimação dos parâmetros de interesse do modelo (3.3) será utilizado a *Estimação Bayesiana Marginal* proposta por Mislevy (1986a). Para isso, seja o vetor de parâmetros de interesse definido por  $\zeta_i = (\mathbf{a}_{ig}, d_{ig}, c_{ig}, \delta_{ig}^a, \delta_{ig}^d)$ , onde  $\zeta_i = (\zeta_{i1}, \dots, \zeta_{iI})$  é o vetor dos parâmetros de interesse do do item  $i$ , com  $i = 1, \dots, I$ , sendo  $I$  o número total de itens;  $g = 2, \dots, G$ , onde  $G$  representa o total de grupos. Nas próximas subseções são apresentadas a função de verossimilhança para os dados observados, as prioris especificadas para os parâmetros de interesse a serem estimados e a distribuição posterior conjunta.

### 3.3.1 Função de Verossimilhança

Denotando por  $P_{ig}(\boldsymbol{\theta}_{jg})$  a probabilidade expressa no modelo em (3.3), sendo  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_J)$ , com  $\mathbf{y}_j = (y_{j1g}, \dots, y_{jIg})$ , a matriz de respostas observadas dos  $J$  indivíduos, do grupo  $g$ ;  $\mathbf{a}$  é o vetor que representa todos os parâmetros  $\mathbf{a}_{ig}$  para os  $I$  itens, do grupo  $g$ ;  $\mathbf{d}_{ig}$  é o vetor dos parâmetros de dificuldade para os  $I$  itens, do grupo  $g$ ;  $\mathbf{c}_{ig}$  o vetor de parâmetros de acerto casual para os  $I$  itens, do grupo  $g$ ;  $\boldsymbol{\delta}_{ig}^a$  é o vetor dos parâmetros para detectar DIF na discriminação dos  $I$  itens, do grupo  $g$ ;  $\boldsymbol{\delta}_{ig}^d$  é o vetor de parâmetros para detectar DIF na dificuldade dos  $I$  itens, do grupo  $g$ . Assim, a função de verossimilhança obtida a partir desse modelo é dada por

$$\begin{aligned} L(\mathbf{Y}|\boldsymbol{\zeta}) &= \prod P(\mathbf{Y}_{\dots} = \mathbf{y}_{\dots}|\boldsymbol{\theta}, \boldsymbol{\zeta}) \\ &= \prod_{j=1}^J P(\mathbf{Y}_{j\cdot} = \mathbf{y}_{j\cdot}|\boldsymbol{\theta}_{j\cdot}, \boldsymbol{\zeta}) \\ &= \prod_{j=1}^J \prod_{i=1}^I P(Y_{ji} = y_{ji}|\boldsymbol{\theta}_{j\cdot}, \boldsymbol{\zeta}) \\ &= \prod_{j=1}^J \prod_{i=1}^I \prod_{g=2}^G [P_{ig}(\boldsymbol{\theta}_{jg})]^{y_{jig}} [1 - P_{ig}(\boldsymbol{\theta}_{jg})]^{1-y_{jig}}. \end{aligned}$$

onde  $P_{ig}(\boldsymbol{\theta}_{jg}) = P(Y_{jig} = 1|\boldsymbol{\theta}_{jg}, \mathbf{a}_{ig}, d_{ig}, c_{ig}, \boldsymbol{\delta}_{ig}^a, \boldsymbol{\delta}_{ig}^d)$ . Segue de (3.9) que a log-verossimilhança pode ser escrita como:

$$\log L(\mathbf{Y}|\boldsymbol{\zeta}) = \sum_{j=1}^J \sum_{i=1}^I \sum_{g=1}^G (y_{jig}) \log P_{ig}(\boldsymbol{\theta}_{jg}) + (1 - y_{jig}) \log(1 - P_{ig}(\boldsymbol{\theta}_{jg})). \quad (3.9)$$

Assumindo que os traços latentes são conhecidos, o interesse é apenas estimar os parâmetros dos itens, assim serão estabelecidas prioris para os componentes de  $\boldsymbol{\zeta}_i$ .

Os estimadores de máxima verossimilhança (EMV) de  $\boldsymbol{\zeta}_i$ ,  $i = 1, \dots, I$ , são os valores que maximizam a função de verossimilhança ou, equivalentemente, são as soluções das equações de estimação

$$\frac{\partial l(\mathbf{y}|\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}_i} = 0. \quad (3.10)$$

Assim, substituindo  $P_{jig}$  por sua expressão apresentada em (3.3), e resolvendo (3.10) individualmente para cada um dos parâmetros  $\mathbf{a}_{ig}$ ,  $d_{ig}$ ,  $c_{ig}$ ,  $\boldsymbol{\delta}_{ig}^a$  e  $\boldsymbol{\delta}_{ig}^d$ , as equações de estimação podem ser expressas por

$$\mathbf{a}_{ig} : (1 - c_{ig}) \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) \boldsymbol{\theta}_j W_{jig} = 0, \quad (3.11)$$

$$d_{ig} : (1 - c_{ig}) \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) W_{jig} = 0, \quad (3.12)$$

$$c_{ig} : \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) \frac{W_{jig}}{P_{jig}^*} = 0, \quad (3.13)$$

$$\boldsymbol{\delta}_{ig}^a : (1 - c_{ig}) \sum_{j=1}^J (y_{jig} - P_{jig}) \boldsymbol{\theta}_j W_{jig} = 0, \quad \forall g \in \{2, \dots, G\}, \quad (3.14)$$

$$\boldsymbol{\delta}_{ig}^d : (1 - c_{ig}) \sum_{j=1}^J (y_{jig} - P_{jig}) W_{jig} = 0, \quad \forall g \in \{2, \dots, G\}, \quad (3.15)$$

com  $W_{jig} = \frac{P_{jig}^* Q_{jig}^*}{P_{jig} Q_{jig}}$ ,  $P_{jig}^* = \{1 + e^{-(a_{i1g} + \delta_{i1g}^a) \theta_{j1g} + \dots + (a_{iKg} + \delta_{iKg}^a) \theta_{jKg} + (d_{ig} + \delta_{ig}^d)}\}^{-1}$  e  $Q_{jig}^* = 1 - P_{jig}^*$ . Para evitar o problema de falta de identificabilidade, as equações (3.14) e (3.15) estão sujeitas às restrições em (3.4).

De forma geral, as equações (3.11) a (3.15) não possuem solução explícita, sendo necessário utilizar algum método iterativo para a obtenção das soluções.

### 3.3.2 Distribuições *a priori*

Como dito anteriormente, a distribuição das proficiências dos indivíduos do grupo de referência é fixada na normal padrão. Assume-se *a priori* que:

$$p(\boldsymbol{\zeta}_i) = p(\mathbf{a}_{ig}) \times p(d_{ig}) \times p(c_{ig}) \times p(\boldsymbol{\delta}_{ig}^a) \times p(\boldsymbol{\delta}_{ig}^d) \quad (3.16)$$

Para cada componente de  $p(\boldsymbol{\zeta}_i)$  assume-se as seguintes distribuições *a priori*

$$p(a_{ikg}) \equiv p(a_{ikg} | \mu_a, \sigma_a^2) = \frac{1}{\sqrt{2\pi a_{ikg} \sigma_a}} \exp\left[-\frac{1}{2\sigma_a^2} (\log a_{ikg} - \mu_a)^2\right] \quad (3.17)$$

$$p(d_{ig}) \equiv p(d_{ig} | \mu_d, \sigma_d^2) = \frac{1}{\sqrt{2\pi \sigma_d}} \exp\left[-\frac{1}{2\sigma_d^2} (d_{ig} - \mu_d)^2\right], \quad i = 1, \dots, I, \quad (3.18)$$

$$p(c_{ig}) \equiv p(c_{ig} | \alpha_c, \beta_c) = \frac{\Gamma(\alpha_c + \beta_c - 2)}{\Gamma(\alpha_c - 1) \Gamma(\beta_c - 1)} c_{ig}^{\alpha_c - 2} (1 - c_{ig})^{\beta_c - 2}, \quad i = 1, \dots, I, \quad (3.19)$$

$$p(\boldsymbol{\delta}_{ig}^a) \equiv p(\boldsymbol{\delta}_{ig}^a | \boldsymbol{\mu}_{ig}^{\delta^a}, \boldsymbol{\Sigma}_{ig}^{\delta^a}) = \frac{1}{(2\pi)^{k/2} |\boldsymbol{\Sigma}_{\delta^a}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\delta}_{ig}^a - \boldsymbol{\mu}_{ig}^{\delta^a})(\boldsymbol{\Sigma}_{ig}^{\delta^a})^{-1}(\boldsymbol{\delta}_{ig}^a - \boldsymbol{\mu}_{ig}^{\delta^a})'\right],$$

$$i = 1, \dots, I, \quad k = 1, \dots, K, \quad (3.20)$$

$$p(\delta_{ig}^d) \equiv p(\delta_{ig}^d | \mu_{\delta^d}, \sigma_{\delta^d}^2) = \frac{1}{\sqrt{2\pi}\sigma_{\delta^d}} \exp\left[-\frac{1}{2\sigma_{\delta^d}^2}(\delta_{ig}^d - \mu_{\delta^d})^2\right], \quad i = 1, \dots, I, \quad (3.21)$$

onde

$a_{ikg} \sim LN(\mu_a, \sigma_a^2)$ ,  $\forall_i = 1, \dots, I$ ; onde  $LN(\cdot, \cdot)$  representa a distribuição *Log-normal* com média  $\mu_a$  e variância  $\sigma_a^2$ .

$d_{ig} \sim N(\mu_d, \sigma_d^2)$ ,  $\forall_i = 1, \dots, I$ ; onde  $N(\cdot, \cdot)$  representa a distribuição *Normal* com média  $\mu_d$  e variância  $\sigma_d^2$ .

$c_{ig} \sim Beta(\alpha_c, \beta_c)$ ,  $\forall_i = 1, \dots, I$ ; onde  $Beta(\cdot, \cdot)$  representa a distribuição *Beta* com parâmetros  $\alpha_c$  e  $\beta_c$ .

$\boldsymbol{\delta}_{ig}^a \sim NM(\boldsymbol{\mu}_{ig}^{\delta^a}, \boldsymbol{\Sigma}_{ig}^{\delta^a})$ ,  $\forall_i = 1, \dots, I$ ; onde  $NM(\cdot, \cdot)$  representa a distribuição *Normal Multivariada* com vetor média  $\boldsymbol{\mu}_{ig}^{\delta^a}$  e matriz de covariâncias  $\boldsymbol{\Sigma}_{ig}^{\delta^a}$  ( $\boldsymbol{\mu}$  é um vetor com  $k$  zeros e  $\boldsymbol{\Sigma}$  é a matriz identidade  $I_k$   $k \times k$ ,  $\forall_k = 1, \dots, K$ . Então o  $\det(\boldsymbol{\Sigma}_{ig}^{\delta^a}) = 1$ .

$\delta_{ig}^d \sim N(\mu_{\delta^d}, \sigma_{\delta^d}^2)$ ,  $\forall_i = 1, \dots, I$ ; onde  $N(\cdot, \cdot)$  representa a distribuição *Normal* com com média  $\mu_{\delta^d}$  e variância  $\sigma_{\delta^d}^2$ .

### 3.3.3 Distribuição *a posteriori* conjunta

Para o processo bayesiano de inferência é necessário obter a distribuição *a posteriori* conjunta de todos os parâmetros do modelo e, através dessa, produzir as estimativas (pontuais e intervalares) de cada um deles. Através do Teorema de Bayes, obtém-se a seguinte densidade para a distribuição *a posteriori* conjunta:

$$p(\boldsymbol{\zeta} | \mathbf{Y}) = \frac{p(\mathbf{Y} | \boldsymbol{\zeta})p(\boldsymbol{\zeta})}{\int \dots \int p(\mathbf{Y} | \boldsymbol{\zeta})p(\boldsymbol{\zeta})d\boldsymbol{\zeta}}. \quad (3.22)$$

onde:

$$\begin{aligned}
p(\mathbf{Y}|\boldsymbol{\zeta})p(\boldsymbol{\zeta}) &= p(\mathbf{Y}|\boldsymbol{\zeta})p(\mathbf{a}_{ig})p(\mathbf{d}_{ig})p(\mathbf{c}_{ig})p(\boldsymbol{\delta}_{ig}^a)p(\boldsymbol{\delta}_{ig}^d) \\
&= \prod_{j=1}^J \prod_{i=1}^I \prod_{g=1}^G P(Y_{jig} = 1 | \boldsymbol{\theta}_{jg}, \mathbf{a}_{ig}, d_{ig}, c_{ig}, \boldsymbol{\delta}_{ig}^a, \boldsymbol{\delta}_{ig}^d) \times \\
&\quad \prod_{i=1}^I p(\mathbf{a}_{ig} | \mu_a, \sigma_a) p(d_{ig} | \mu_d, \sigma_d^2) p(c_{ig} | \alpha_c, \beta_c) \times \\
&\quad \prod_{i=1}^I \prod_{g=1}^G p(\boldsymbol{\delta}_{ig}^a | \boldsymbol{\mu}_{\delta^a}, \boldsymbol{\Sigma}_{\delta^a}) p(\boldsymbol{\delta}_{ig}^d | \mu_{\delta^d}, \sigma_{\delta^d}^2),
\end{aligned}$$

que pode ser simplificada para

$$p(\boldsymbol{\zeta}|\mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\zeta}) \times p(\boldsymbol{\zeta}), \quad (3.23)$$

uma vez que o denominador em (3.23) não depende do vetor  $\boldsymbol{\zeta}$ .

A estimação bayesiana será feita pela moda da posteriori (MAP: maximum a posteriori) que consiste em obter o máximo de 3.23. Para facilitar os cálculos iremos trabalhar com o logaritmo da posteriori

$$\log p(\boldsymbol{\zeta}|\mathbf{Y}) \propto \log L(\mathbf{Y}|\boldsymbol{\zeta}) + \log p(\boldsymbol{\zeta}). \quad (3.24)$$

Logo, os estimadores Bayesianos são obtidos pela solução do sistema de equações

$$\frac{\partial \log p(\boldsymbol{\zeta}|\mathbf{Y})}{\partial \boldsymbol{\zeta}} = \frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} = 0. \quad (3.25)$$

### 3.3.4 Equações de estimação dos parâmetros dos itens

Nesta seção serão apresentadas as equações de estimação para a obtenção dos estimadores bayesianos para cada componente do vetor  $\boldsymbol{\zeta}_i = (\mathbf{a}_{ig}, d_{ig}, c_{ig}, \boldsymbol{\delta}_{ig}^a, \boldsymbol{\delta}_{ig}^d)$ , com base em (3.25):

$$\frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial \mathbf{a}_{ig}} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial \mathbf{a}_{ig}} = 0, \quad (3.26)$$

$$\frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial d_{ig}} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial d_{ig}} = 0, \quad (3.27)$$

$$\frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial c_{ig}} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial c_{ig}} = 0, \quad (3.28)$$

$$\frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial \boldsymbol{\delta}_{ig}^a} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial \boldsymbol{\delta}_{ig}^a} = 0, \quad (3.29)$$

$$\frac{\partial \log L(\mathbf{Y}|\boldsymbol{\zeta})}{\partial \delta_{ig}^d} + \frac{\partial \log p(\boldsymbol{\zeta})}{\partial \delta_{ig}^d} = 0. \quad (3.30)$$

As primeiras parcelas nas equações (3.26) a (3.30) foram obtidas na Seção 3.3, na estimação via Método da Máxima Verossimilhança, e são apresentadas no membro esquerdo das equações (3.11) a (3.15), respectivamente. A abordagem bayesiana adiciona uma nova parcela relativa a distribuição *a priori* em cada equação de estimação.

De (3.17) a (3.21), tem-se que

$$\frac{\partial \log p(\boldsymbol{\zeta})}{\partial \mathbf{a}_{ig}} = \frac{\partial \log p(\mathbf{a}_{ig})}{\partial \mathbf{a}_{ig}} = -\frac{1}{\mathbf{a}_{ig}} \left[ 1 + \frac{\log \mathbf{a}_{ig} - \mu_a}{\sigma_a^2} \right], \quad (3.31)$$

$$\frac{\partial \log p(\boldsymbol{\zeta})}{\partial d_{ig}} = \frac{\partial \log p(d_{ig})}{\partial d_{ig}} = -\frac{d_{ig} - \mu_d}{\sigma_d^2}, \quad (3.32)$$

$$\frac{\partial \log p(\boldsymbol{\zeta})}{\partial c_{ig}} = \frac{\partial \log p(c_{ig})}{\partial c_{ig}} = \frac{\alpha_c - 2}{c_{ig}} - \frac{\beta_c - 2}{1 - c_{ig}}, \quad (3.33)$$

$$\frac{\partial \log p(\boldsymbol{\zeta})}{\partial \boldsymbol{\delta}_{ig}^a} = \frac{\partial \log p(\delta^a)}{\partial \boldsymbol{\delta}_{ig}^a} = \boldsymbol{\mu}_{ig}^{\delta^a} - \boldsymbol{\delta}_{ig}^a, \quad (3.34)$$

$$\frac{\partial \log p(\boldsymbol{\zeta})}{\partial \delta_{ig}^d} = \frac{\partial \log p(\delta^d)}{\partial \delta_{ig}^d} = -\frac{\delta_{ig}^d - \mu_{\delta^d}}{\sigma_{\delta^d}^2}. \quad (3.35)$$

Assim, substituindo (3.11) a (3.15) e (3.31) a (3.35) no sistema de equações (3.26) a (3.30), as equações de estimação para a obtenção dos estimadores bayesianos dos parâmetros  $\mathbf{a}_{ig}$ ,  $d_{ig}$ ,  $c_{ig}$ ,  $\boldsymbol{\delta}_{ig}^a$  e  $\delta_{ig}^d$  são, respectivamente,

$$\mathbf{a}_{ig} : (1 - c_{ig}) \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) \boldsymbol{\theta}_{jg} W_{jig} - \frac{1}{\mathbf{a}_{ig}} \left[ 1 + \frac{\log \mathbf{a}_{ig} - \mu_a}{\sigma_a^2} \right] = \mathbf{0}, \quad (3.36)$$

$$d_{ig} : -(1 - c_{ig}) \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) W_{jig} - \frac{d_{ig} - \mu_d}{\sigma_d} = 0, \quad (3.37)$$

$$c_{ig} : \sum_{g=2}^G \sum_{j=1}^J (y_{jig} - P_{jig}) \frac{W_{jig}}{P_{jig}^*} + \frac{\alpha_c - 2}{c_{ig}} - \frac{\beta_c - 2}{1 - c_{ig}} = 0, \quad (3.38)$$

$$\boldsymbol{\delta}_{ig}^a : (1 - c_{ig}) \sum_{j=1}^J (y_{jig} - P_{jig}) \boldsymbol{\theta}_{jg} W_{jig} + (\boldsymbol{\mu}_{ig}^{\delta^a} - \boldsymbol{\delta}_{ig}^a) = \mathbf{0}, \quad (3.39)$$

$$\delta_{ig}^d : (1 - c_{ig}) \sum_{j=1}^J (y_{jig} - P_{jig}) W_{jig} - \frac{\delta_{ig}^d - \mu_{\delta^d}}{\sigma_{\delta^d}^2} = 0. \quad (3.40)$$

O sistema (3.36) a (3.40) não tem solução analítica, e algum processo numérico é necessário para sua solução.

Após a obtenção das estimativas bayesianas, o próximo passo é verificar se os parâmetros  $\boldsymbol{\delta}_{ig}^a$  e  $\delta_{ig}^d$  são significativos, ou seja, se existe efeito-posição nos parâmetros de discriminação e/ou dificuldade de algum item em alguma posição. Para isso, testam-se as hipóteses (3.5) e (3.6), apresentadas na Seção 3.2. Na abordagem bayesiana, a decisão sobre a rejeição ou não das hipóteses nulas baseia-se nos intervalos de credibilidade (ICr) para os respectivos parâmetros, verificando se os mesmos contém ou não o valor zero.

Os intervalos com  $\gamma\%$  de credibilidade são obtidos na distribuição posterior marginal dos parâmetros  $\boldsymbol{\delta}_{ig}^a$  e  $\delta_{ig}^d$  através dos percentis de ordem  $\frac{(1-\gamma)}{2}$  e  $\frac{(1+\gamma)}{2}$ .

Assim, o  $ICr(\boldsymbol{\delta}_{ig}^a; \gamma\%) = (l_{inf}^a, l_{sup}^a)$ , onde  $l_{inf}^a$  e  $l_{sup}^a$  são, respectivamente, os percentis de ordem  $\frac{(1-\gamma)}{2}$  e  $\frac{(1+\gamma)}{2}$  na distribuição

$$p(\boldsymbol{\delta}_{ig}^a | \mathbf{Y}) = \int p(\boldsymbol{\zeta} | \mathbf{Y}) d\boldsymbol{\zeta}_{(-\boldsymbol{\delta}_{ig}^a)}. \quad (3.41)$$

E o  $ICr(\delta_{ig}^d; \gamma\%) = (l_{inf}^d, l_{sup}^d)$ , onde  $l_{inf}^d$  e  $l_{sup}^d$  são, respectivamente, os percentis de ordem  $\frac{(1-\gamma)}{2}$  e  $\frac{(1+\gamma)}{2}$  na distribuição

$$p(\delta_{ig}^d | \mathbf{Y}) = \int p(\boldsymbol{\zeta} | \mathbf{Y}) d\boldsymbol{\zeta}_{(-\delta_{ig}^d)}. \quad (3.42)$$

Em (3.41) o vetor  $\boldsymbol{\zeta}_{(-\boldsymbol{\delta}_{ig}^a)}$  representa o vetor  $\boldsymbol{\zeta}$  sem o parâmetro  $\boldsymbol{\delta}_{ig}^a$  e em (3.42) o vetor  $\boldsymbol{\zeta}_{(-\delta_{ig}^d)}$  representa o vetor  $\boldsymbol{\zeta}$  sem o parâmetro  $\delta_{ig}^d$ .

### 3.4 Adequação do modelo multidimensional

Para analisar a dimensionalidade de um instrumento de medida uma das técnicas mais utilizadas é a análise fatorial. Como nesta dissertação os dados são provenientes de exames aplicados na área educacional, em que normalmente os itens são dicotômicos os pressupostos da análise fatorial clássica podem não ser atendidos. Dessa forma, as dificuldades apresentadas foram contornadas por Bock, Gibbons e Muraki (1988) ao proporem adaptações ao modelo, surgindo assim a Análise Fatorial de Informação Completa (AFIC), uma metodologia baseada na TRI que permite verificar a dimensionalidade e interpretar os fatores do instrumento, e evita o problema de calcular as correlações tetracóricas entre os itens.

Outros métodos também são utilizados para verificar o ajuste do modelo aos dados e sua dimensionalidade, ao selecionar modelos é preciso levar em consideração que não existem modelos verdadeiros, e sim modelos que aproximam-se da realidade causando algum tipo de perda de algumas informações. Desse maneira, é necessário utilizar algum método para identificar qual o “melhor” modelo, dentre aqueles que foram ajustados, para explicar o fenômeno sob estudo.

Akaike (1974) utilizou a Informação de Kullback-Leibler para testar se um dado modelo é adequado. Porém seu uso é limitado, pois depende da distribuição  $g$  (modelo verdadeiro), que é desconhecida. Mostrou que o viés é dado assintoticamente por  $p$ , em que  $p$  é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como:

$$AIC_{Modelo} = -2\hat{\theta} + 2(p). \quad (3.43)$$

O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) é dado por:

$$BIC_{Modelo} = -2 \log f(\mathbf{y}|\boldsymbol{\theta}) + p \log(n), \quad (3.44)$$

em que  $f(\mathbf{y}|\boldsymbol{\theta})$  é o modelo escolhido,  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra.



---

## Capítulo 4

# Dados Simulados do DIF no Modelo Multidimensional da TRI

---

### 4.1 Introdução

Neste capítulo apresentam-se as análises obtidas com os dados simulados para o modelo (3.3) proposto para modelagem do DIF. Na Seção 4.2 apresenta-se a estrutura do algoritmo utilizado que foi programado no *software R*. Na Seção 4.3 apresenta-se os resultados obtidos através da estimação por máxima verossimilhança

### 4.2 Aplicação a dados simulados

#### 4.2.1 Algoritmo

A aplicação do algoritmo ao modelo 3.3 proposto para modelagem de DIF. Foi utilizado alguns pacotes do *software R*: *mirt*, *irttoys*, *maxLik*, *readxl*, *plotrix*, *foreign* e *bbmle*. A posteriori (3.22) foi otimizada utilizando a função `optim` do R, com limitadores inferior e superior para cada parâmetro. Como estimativas iniciais dos parâmetros  $a_i$ ,  $b_i$  e  $c_i$  adotou-se os valores obtidos pelo `mirt` (). Para os parâmetros  $\delta_*^a$  e  $\delta^b$  adotou-se o valor nulo como inicial. As priors adotadas são apresentadas na Seção 3.3.2.

### 4.3 Estimativas via Inferência Bayesiana

O procedimento de estimação proposto foi aplicado a  $R = 100$  conjuntos de dados simulados (réplicas) obtidos com os parâmetros dos itens constantes na Tabela 4.1. Foi considerado que o traço latente afere duas dimensões ( $K = 2$ ). Em particular, os dados referem-se as resposta de 10.000 indivíduos a 45 itens, dentre os quais 15 foram escolhidos para terem DIF nos parâmetros de discriminação e/ou dificuldade. Foram considerados dois grupos para fins de simulação ( $G = 2$ ), referência e focal, de forma que os DIFs

associados a cada parâmetro ( $a_{i1}$ ,  $a_{i2}$  e  $d_i$ ) serão as alterações nos parâmetros obtidos com o grupo focal, relativamente ao grupo referência. Assim, teremos apenas parâmetros de DIF associados ao grupo focal, representados por  $\delta_i$ . Nos cenários abaixo especifica-se a divisão dos itens.

a) Itens 1, 8 e 15, com DIF na primeira dimensão do parâmetro de discriminação.

Item 1:  $\delta_1^a = -0.4$

Item 8:  $\delta_1^a = -0.2$

Item 15:  $\delta_1^a = 0.4$

b) Itens 31, 38 e 45, com DIF na segunda dimensão do parâmetro de discriminação.

Item 31:  $\delta_2^a = -0.2$

Item 38:  $\delta_2^a = 0.1$

Item 45:  $\delta_2^a = 0.5$

c) Itens 26, 36 e 44, com DIF nas duas dimensões do parâmetro de discriminação.

Item 26:  $\delta_1^a = -0.3$ ;  $\delta_2^a = -0.2$

Item 36:  $\delta_1^a = 0.1$ ;  $\delta_2^a = 0.3$

Item 44:  $\delta_1^a = 0.4$ ;  $\delta_2^a = 0.6$

d) Itens 4, 11 e 19, com DIF no parâmetro de dificuldade.

Item 4:  $\delta^b = 0.2$

Item 11:  $\delta^b = 0.5$

Item 19:  $\delta^b = 0.8$

e) Itens 5 e 12, com DIF no parâmetro de discriminação e dificuldade.

Item 5:  $\delta_1^a = -0.1$ ;  $\delta^b = -0.3$

Item 12:  $\delta_1^a = 0.4$ ;  $\delta^b = -0.2$

f) Itens 20 e 27, com DIF no parâmetro de discriminação e dificuldade.

Item 20:  $\delta_2^a = 0.3$ ;  $\delta^b = -0.1$

Item 27:  $\delta_2^a = -0.4$ ;  $\delta^b = 0.2$

g) Itens 37 e 43, com DIF nas duas dimensões do parâmetro de discriminação e dificuldade.

Item 37:  $\delta_1^a = -0.1$ ;  $\delta_2^a = 0.1$ ;  $\delta^b = 0.2$

Item 43:  $\delta_1^a = 0.2$ ;  $\delta_2^a = -0.2$ ;  $\delta^b = 0.3$

Nas colunas  $\{a_1, a_2, b, c\}$  da Tabela 4.1, constam os valores (verdadeiros) dos parâmetros dos itens adotados para a simulação das respostas do grupo referência, e nas colunas posteriores temos os DIFs para o grupo focal. Naturalmente espera-se que as estimativas obtidas no processo de estimação sejam muito próximas desses valores. Caso isto aconteça pode-se afirmar que o processo de estimação está funcionando bem, produzindo erros pequenos de estimação. Para o parâmetro de discriminação da primeira e segunda dimensão foram considerados os valores: 0,75, 1,00 e 2,00. Para os parâmetros de dificuldade foram adotados os seguintes valores: -2,00, -1,00, 0,00, 1,00 e 2,00. Para os parâmetros do acerto casual adotou-se 0,00, 0,10 e 0,20.

Tabela 4.1 *Valores adotados para os parâmetros dos itens na simulação dos dados.*

Item	$a_1$	$a_2$	$b$	$c$	$\delta_1^a$	$\delta_2^a$	$\delta^b$
Item 1	0,75	1,00	-2,00	0,00	-0,40	0,00	0,00
Item 2	0,75	0,75	-2,00	0,10	0,00	0,00	0,00
Item 3	0,75	1,00	-2,00	0,20	0,00	0,00	0,00
Item 4	0,75	2,00	-1,00	0,00	0,00	0,00	0,20
Item 5	0,75	2,00	-1,00	0,10	-0,10	0,00	-0,30
Item 6	0,75	1,00	-1,00	0,20	0,00	0,00	0,00
Item 7	0,75	1,00	0,00	0,00	0,00	0,00	0,00
Item 8	0,75	2,00	0,00	0,10	-0,20	0,00	0,00
Item 9	0,75	0,75	0,00	0,20	0,00	0,00	0,00
Item 10	0,75	0,75	1,00	0,00	0,00	0,00	0,00
Item 11	0,75	1,00	1,00	0,10	0,00	0,00	0,50
Item 12	0,75	0,75	1,00	0,20	0,40	0,00	-0,20
Item 13	0,75	1,00	2,00	0,00	0,00	0,00	0,00
Item 14	0,75	0,75	2,00	0,10	0,00	0,00	0,00
Item 15	0,75	2,00	2,00	0,20	0,40	0,00	0,00
Item 16	1,00	1,00	-2,00	0,00	0,00	0,00	0,00
Item 17	1,00	2,00	-2,00	0,10	0,00	0,00	0,00
Item 18	1,00	0,75	-2,00	0,20	0,00	0,00	0,00
Item 19	1,00	1,00	-1,00	0,00	0,00	0,00	0,80
Item 20	1,00	0,75	-1,00	0,10	0,00	0,30	-0,10
Item 21	1,00	1,00	-1,00	0,20	0,00	0,00	0,00
Item 22	1,00	2,00	0,00	0,00	0,00	0,00	0,00
Item 23	1,00	0,75	0,00	0,10	0,00	0,00	0,00
Item 24	1,00	0,75	0,00	0,20	0,00	0,00	0,00
Item 25	1,00	0,75	1,00	0,00	0,00	0,00	0,00
Item 26	1,00	1,00	1,00	0,10	-0,30	-0,20	0,00
Item 27	1,00	0,75	1,00	0,20	0,00	-0,40	0,20
Item 28	1,00	0,75	2,00	0,00	0,00	0,00	0,00
Item 29	1,00	1,00	2,00	0,10	0,00	0,00	0,00
Item 30	1,00	0,75	2,00	0,20	0,00	0,00	0,00
Item 31	2,00	2,00	-2,00	0,00	0,00	-0,20	0,00
Item 32	2,00	0,75	-2,00	0,10	0,00	0,00	0,00
Item 33	2,00	2,00	-2,00	0,20	0,00	0,00	0,00
Item 34	2,00	2,00	-1,00	0,00	0,00	0,00	0,00
Item 35	2,00	2,00	-1,00	0,10	0,00	0,00	0,00
Item 36	2,00	2,00	-1,00	0,20	0,10	0,30	0,00
Item 37	2,00	2,00	0,00	0,00	-0,10	0,10	0,20
Item 38	2,00	1,00	0,00	0,10	0,00	0,10	0,00
Item 39	2,00	2,00	0,00	0,20	0,00	0,00	0,00
Item 40	2,00	2,00	1,00	0,00	0,00	0,00	0,00
Item 41	2,00	1,00	1,00	0,10	0,00	0,00	0,00
Item 42	2,00	1,00	1,00	0,20	0,00	0,00	0,00
Item 43	2,00	1,00	2,00	0,00	0,20	-0,20	0,30
Item 44	2,00	0,75	2,00	0,10	0,40	0,60	0,00
Item 45	2,00	2,00	2,00	0,20	0,00	0,50	0,00

Fonte: o próprio autor

### 4.3.1 Definição das *prioris*

No processo de estimação bayesiana dos parâmetros para os dados simulados, as *prioris* utilizadas foram:

- $a_{ik} \sim LN(\mu_a, \sigma_a^2)$ , com  $\mu_a = 1$  e  $\sigma_a^2 = e^{0.5}$ .
- $b_i \sim N(\mu_b, \sigma_b^2)$ , com  $\mu_b = 0$  e  $\sigma_b^2 = 2$ .
- $c_i \sim \text{Beta}(\alpha - 1, \beta - 1)$ , com  $\alpha = 5$  e  $\beta = 17$ .
- $\delta_{ig}^a \sim N(0, 1)$ , com  $\Sigma_{\delta^a}^2 = I$ .
- $\delta_{ig}^b \sim N(0, \sigma_{\delta^b}^2)$ , com  $\sigma_{\delta^b}^2 = 0, 1$ .

As *prioris* acima especificadas são frequentemente adotadas na literatura (BAKER e KIM, 2004).

A Tabela 4.2 apresenta as estimativas médias dos parâmetros dos itens obtidas nas 100 réplicas, via metodologia bayesiana. Na Tabela 4.3 consta o desvio-padrão das estimativas bayesianas dos parâmetros dos itens. Aqui também observa-se que os desvios-padrão são pequenos.

As diferenças entre as estimativas médias e os valores (verdadeiros) adotados para os parâmetros dos itens, ou seja, os erros na estimação, são apresentados na Tabela 4.4. Observa-se que os erros foram pequenos, e a maior diferença em valor absoluto observada foi 0,304, associada a estimação do parâmetro  $b$  no item 1.

Assim, tem-se que o processo de estimação bayesiano também está funcionando bem, pois produziu estimativas precisas e pequenos erros de estimação.

No processo de bayesiano, o tempo total de execução do algoritmo para as 100 réplicas foi cerca de 10 horas.

Tabela 4.2 *Estimativas médias dos parâmetros dos itens para os dados simulados, via método bayesiano.*

Item	$a_1$	$a_2$	b	c	$\delta_1^a$	$\delta_2^a$	$\delta^b$
Item 1	0,761	0,760	-1,991	0,001	-0,409	-0,007	-0,012
Item 2	0,743	2,017	-1,994	0,100	0,011	-0,014	-0,003
Item 3	0,751	0,996	-1,999	0,199	0,000	0,015	0,002
Item 4	0,759	0,754	-1,004	0,002	-0,002	0,001	0,199
Item 5	0,762	0,759	-1,006	0,102	-0,107	0,001	-0,293
Item 6	0,750	1,008	-0,996	0,200	0,002	0,005	0,005
Item 7	0,763	0,757	-0,010	0,005	-0,001	-0,001	-0,004
Item 8	0,752	1,000	0,002	0,100	-0,194	0,002	-0,001
Item 9	0,751	0,754	0,004	0,199	0,002	-0,002	-0,003
Item 10	0,757	1,007	0,989	0,006	-0,003	0,002	0,000
Item 11	0,758	2,015	0,997	0,102	0,003	0,008	0,498
Item 12	0,751	1,000	0,998	0,201	0,409	0,003	-0,197
Item 13	0,756	2,044	1,979	0,011	0,003	-0,019	0,009
Item 14	0,758	0,756	1,979	0,109	-0,006	-0,001	0,007
Item 15	0,754	0,760	1,993	0,198	0,406	0,001	0,003
Item 16	1,009	1,018	-1,990	0,000	0,005	-0,011	-0,007
Item 17	1,004	0,748	-1,993	0,099	-0,010	-0,004	-0,007
Item 18	0,991	1,009	-1,996	0,200	0,019	-0,009	0,000
Item 19	1,005	2,017	-1,001	0,000	0,004	-0,003	0,801
Item 20	1,003	0,998	-1,001	0,100	0,005	0,302	-0,098
Item 21	0,996	0,992	-0,997	0,199	0,008	0,007	0,004
Item 22	1,005	0,762	-0,006	0,003	0,005	-0,008	0,000
Item 23	1,001	0,750	-0,001	0,100	0,000	-0,003	0,007
Item 24	1,006	2,010	-0,001	0,201	-0,008	0,001	0,001
Item 25	1,011	1,008	0,993	0,005	0,003	-0,004	-0,005
Item 26	1,006	2,012	1,001	0,103	-0,296	-0,210	-0,003
Item 27	1,010	0,754	0,993	0,204	-0,003	-0,394	0,197
Item 28	1,006	0,764	1,969	0,021	0,005	-0,003	-0,005
Item 29	1,027	2,035	1,991	0,102	-0,003	-0,001	0,004
Item 30	1,009	2,043	1,988	0,204	0,022	-0,009	-0,001
Item 31	2,029	2,037	-1,994	0,000	-0,010	-0,207	-0,004
Item 32	2,009	0,756	-2,004	0,100	0,011	0,005	0,003
Item 33	2,026	2,015	-1,999	0,200	0,036	0,045	0,007
Item 34	2,020	1,010	-1,001	0,000	-0,009	-0,003	0,003
Item 35	2,017	1,000	-0,999	0,100	-0,010	0,006	0,003
Item 36	2,024	2,023	-0,997	0,199	0,078	0,279	-0,002
Item 37	2,018	1,013	-0,002	0,001	-0,109	0,091	0,200
Item 38	2,005	2,013	-0,002	0,100	-0,001	0,100	0,003
Item 39	1,999	0,758	0,004	0,200	0,009	-0,007	-0,004
Item 40	2,014	1,005	0,999	0,002	0,013	0,004	-0,006
Item 41	2,019	2,017	0,998	0,101	0,006	-0,013	-0,001
Item 42	2,024	2,025	0,997	0,201	0,004	0,006	-0,001
Item 43	2,027	1,020	1,987	0,015	0,234	-0,196	0,287
Item 44	2,029	0,768	1,988	0,107	0,410	0,616	0,004
Item 45	2,058	2,044	1,990	0,210	-0,005	0,516	0,002

Fonte: Dados Simulados.

Tabela 4.3 *Desvios-padrão das estimativas bayesianas dos parâmetros dos itens para os dados simulados.*

Item	$a_1$	$a_2$	b	c	$\delta_1^a$	$\delta_2^a$	$\delta^b$
Item 1	0,035	0,033	0,046	0,001	0,050	0,044	0,073
Item 2	0,071	0,112	0,039	0,003	0,103	0,155	0,055
Item 3	0,058	0,082	0,052	0,006	0,083	0,090	0,065
Item 4	0,031	0,032	0,031	0,003	0,044	0,045	0,041
Item 5	0,039	0,044	0,039	0,009	0,051	0,052	0,044
Item 6	0,054	0,060	0,038	0,010	0,070	0,070	0,038
Item 7	0,030	0,033	0,025	0,007	0,042	0,044	0,031
Item 8	0,037	0,043	0,043	0,014	0,046	0,052	0,035
Item 9	0,039	0,039	0,056	0,016	0,050	0,053	0,046
Item 10	0,036	0,033	0,032	0,010	0,044	0,046	0,032
Item 11	0,045	0,064	0,031	0,019	0,067	0,091	0,033
Item 12	0,044	0,044	0,051	0,020	0,067	0,055	0,037
Item 13	0,057	0,072	0,035	0,015	0,073	0,089	0,043
Item 14	0,040	0,049	0,124	0,050	0,055	0,059	0,064
Item 15	0,050	0,049	0,130	0,053	0,074	0,074	0,079
Item 16	0,049	0,045	0,032	0,001	0,072	0,061	0,047
Item 17	0,073	0,059	0,055	0,005	0,090	0,076	0,075
Item 18	0,070	0,086	0,047	0,006	0,108	0,107	0,067
Item 19	0,040	0,052	0,016	0,000	0,050	0,071	0,020
Item 20	0,044	0,052	0,029	0,006	0,065	0,064	0,034
Item 21	0,060	0,056	0,032	0,008	0,073	0,066	0,030
Item 22	0,033	0,030	0,020	0,004	0,044	0,047	0,028
Item 23	0,042	0,039	0,038	0,014	0,051	0,041	0,032
Item 24	0,053	0,071	0,024	0,009	0,068	0,092	0,022
Item 25	0,032	0,035	0,024	0,008	0,045	0,057	0,029
Item 26	0,049	0,058	0,026	0,015	0,060	0,077	0,029
Item 27	0,051	0,047	0,076	0,030	0,060	0,058	0,052
Item 28	0,047	0,039	0,066	0,029	0,061	0,059	0,060
Item 29	0,064	0,083	0,051	0,033	0,097	0,124	0,045
Item 30	0,080	0,100	0,062	0,034	0,105	0,135	0,057
Item 31	0,094	0,082	0,024	0,000	0,123	0,115	0,034
Item 32	0,105	0,086	0,041	0,003	0,163	0,109	0,057
Item 33	0,137	0,147	0,029	0,004	0,201	0,182	0,039
Item 34	0,057	0,035	0,015	0,000	0,076	0,052	0,024
Item 35	0,079	0,051	0,017	0,004	0,096	0,070	0,021
Item 36	0,091	0,093	0,018	0,005	0,120	0,132	0,024
Item 37	0,048	0,041	0,015	0,001	0,058	0,060	0,020
Item 38	0,064	0,069	0,014	0,006	0,089	0,087	0,017
Item 39	0,065	0,046	0,021	0,008	0,085	0,067	0,023
Item 40	0,055	0,043	0,016	0,004	0,073	0,067	0,027
Item 41	0,069	0,083	0,020	0,010	0,095	0,105	0,025
Item 42	0,084	0,079	0,024	0,013	0,108	0,112	0,027
Item 43	0,094	0,057	0,043	0,022	0,130	0,103	0,055
Item 44	0,082	0,069	0,052	0,030	0,129	0,105	0,049
Item 45	0,094	0,116	0,042	0,029	0,149	0,165	0,039

Fonte: Dados Simulados.

Tabela 4.4 *Erros na estimação bayesiana dos parâmetros dos itens para os dados simulados.*

Item	$a_1$	$a_2$	b	c	$\delta_1^a$	$\delta_2^a$	$\delta^b$
Item 1	0,011	0,010	0,009	0,001	-0,009	-0,007	-0,012
Item 2	-0,007	0,017	0,006	0,000	0,011	-0,014	-0,003
Item 3	0,001	-0,004	0,001	-0,001	0,000	0,015	0,002
Item 4	0,009	0,004	-0,004	0,002	-0,002	0,001	-0,001
Item 5	0,012	0,009	-0,006	0,000	-0,007	0,001	0,007
Item 6	0,000	0,008	0,004	0,000	0,002	0,005	0,005
Item 7	0,013	0,007	-0,010	0,005	-0,001	-0,001	-0,004
Item 8	0,002	0,000	0,002	0,000	0,006	0,002	-0,001
Item 9	0,001	0,004	0,004	-0,001	0,002	-0,002	-0,003
Item 10	0,007	0,007	-0,011	0,006	-0,003	0,002	0,000
Item 11	0,008	0,015	-0,003	0,002	0,003	0,008	-0,002
Item 12	0,001	0,000	-0,002	0,001	0,009	0,003	0,003
Item 13	0,006	0,044	-0,021	0,011	0,003	-0,019	0,009
Item 14	0,008	0,006	-0,021	0,009	-0,006	-0,001	0,007
Item 15	0,004	0,010	-0,007	-0,002	0,006	0,001	0,003
Item 16	0,009	0,018	0,010	0,000	0,005	-0,011	-0,007
Item 17	0,004	-0,002	0,007	-0,001	-0,010	-0,004	-0,007
Item 18	-0,009	0,009	0,004	0,000	0,019	-0,009	0,000
Item 19	0,005	0,017	-0,001	0,000	0,004	-0,003	0,001
Item 20	0,003	-0,002	-0,001	0,000	0,005	0,002	0,002
Item 21	-0,004	-0,008	0,003	-0,001	0,008	0,007	0,004
Item 22	0,005	0,012	-0,006	0,003	0,005	-0,008	0,000
Item 23	0,001	0,000	-0,001	0,000	0,000	-0,003	0,007
Item 24	0,006	0,010	-0,001	0,001	-0,008	0,001	0,001
Item 25	0,011	0,008	-0,007	0,005	0,003	-0,004	-0,005
Item 26	0,006	0,012	0,001	0,003	0,004	-0,010	-0,003
Item 27	0,010	0,004	-0,007	0,004	-0,003	0,006	-0,003
Item 28	0,006	0,014	-0,031	0,021	0,005	-0,003	-0,005
Item 29	0,027	0,035	-0,009	0,002	-0,003	-0,001	0,004
Item 30	0,009	0,043	-0,012	0,004	0,022	-0,009	-0,001
Item 31	0,029	0,037	0,006	0,000	-0,010	-0,007	-0,004
Item 32	0,009	0,006	-0,004	0,000	0,011	0,005	0,003
Item 33	0,026	0,015	0,001	0,000	0,036	0,045	0,007
Item 34	0,020	0,010	-0,001	0,000	-0,009	-0,003	0,003
Item 35	0,017	0,000	0,001	0,000	-0,010	0,006	0,003
Item 36	0,024	0,023	0,003	-0,001	-0,022	-0,021	-0,002
Item 37	0,018	0,013	-0,002	0,001	-0,009	-0,009	0,000
Item 38	0,005	0,013	-0,002	0,000	-0,001	0,000	0,003
Item 39	-0,001	0,008	0,004	0,000	0,009	-0,007	-0,004
Item 40	0,014	0,005	-0,001	0,002	0,013	0,004	-0,006
Item 41	0,019	0,017	-0,002	0,001	0,006	-0,013	-0,001
Item 42	0,024	0,025	-0,003	0,001	0,004	0,006	-0,001
Item 43	0,027	0,020	-0,013	0,015	0,034	0,004	-0,013
Item 44	0,029	0,018	-0,012	0,007	0,010	0,016	0,004
Item 45	0,058	0,044	-0,010	0,010	-0,005	0,016	0,002

Fonte: Dados Simulados.



---

## Capítulo 5

# Conclusões e Trabalhos Futuros

---

### 5.1 Considerações finais

A Teoria de Resposta ao Item é uma das técnicas mais utilizada nas áreas de avaliação educacional e psicologia, sendo munida de ferramentas estatísticas para a produção de escalas quantitativas em ambas as áreas. A medida em que cresce a aplicação da TRI, surgem novos fatores e questionamentos de grande importância prática, e o funcionamento diferencial do item é um deles. A necessidade de generalizar a TRI, onde o instrumento de medida do traço latente é multidimensional, para incorporar modelos que se ajustem adequadamente a situações que não são convencionadas pelos modelos usuais, como a existência de DIF, levou à proposta do modelo descrito nesta dissertação, sendo uma generalização do modelo logístico de três parâmetros para o caso multidimensional, com a inclusão de parâmetros para detecção de itens com DIF relativos a cada dimensão.

Foi apresentada uma abordagem Bayesiana para se fazer inferência em tal modelo baseada na moda da posteriori (MAP: maximum a posteriori). Através de estudos simulados, mostrou-se que o modelo se adequa bem para a modelagem do funcionamento diferencial do item e que a metodologia Bayesiana adotada é muito eficiente para se fazer inferência sobre tal modelo. Os estudos simulados também mostraram questões importantes a respeito da convergência dos algoritmo proposto, de características do funcionamento diferencial existente nos dados e da escolha de distribuições a priori.

Como continuidade deste trabalho, o modelo poderá aplicado, por exemplo, a um conjunto de dados reais do Exame Nacional do Ensino Médio - ENEM. O Enem é composto por quatro provas de múltipla escolha e uma redação. No exame os candidatos são submetidos a uma prova objetiva dividida em quatro áreas de conhecimento, *Ciências da Natureza e suas tecnologias*, *Ciências Humanas e suas tecnologias*, *Linguagens, códigos e suas tecnologias* e *Matemática e suas tecnologias*, com 45 questões cada, totalizando 180 questões. Os dados que serão utilizados para aplicação do modelo será referente ao caderno

amarelo na área de conhecimento *Matemática e suas tecnologias* no ano de 2015. Para esta área os indivíduos serão divididos em dois grupos, o grupo de referência é formado pelo gênero masculino e o grupo focal pelo gênero feminino. Segundo o INEP, o ENEM 2015 teve 8.478.096 candidatos inscritos, será considerada uma amostra de 20.000 indivíduos, com 10.000 em cada grupo.

O algoritmo apresentado foi implementado no software R, devido à sua complexidade, gastam um tempo razoável para ser rodado. O modelo de detecção de DIF com 100 iterações, por exemplo, gasta em torno de 10 horas em um computador Intel Core i5, 2.40GHz, 8,00GB de RAM. Os gráficos foram feitos também no software R.

## 5.2 Recomendações para trabalhos futuros

Em trabalhos futuros, pretende-se estudar:

1. Aplicar esta metodologia aos dados do Exame Nacional do Ensino Médio em relação ao gênero;
2. Aplicar esta metodologia aos dados com mais de 2 grupos;
3. Estudar e propor critérios de decisão para classificação de itens quanto ao funcionamento diferencial.

---

# Bibliografia

---

- Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, Boston, v.19, n.6, p.716-723, Dec. 1974.
- Andrade, D.F., Tavares, H.R. e Valle, R.C. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicações*. Associação Brasileira de Estatística: São Paulo.
- Angoff, W. H. (1993). Perspectives on differential item functioning. Em P.W. Holland & H. Wainer (Orgs.), *Differential item functioning* (pp. 3-24). New Jersey: Lawrence Erlbaum.
- Araújo, E. A. C.; Andrade, D. F.; Bortolotti, S. L. V. Teoria da Resposta ao Item. *Revista da Escola de Enfermagem da USP, São Paulo*. Vol. 43, 2009.
- Bock, R. D.; Gibbons, R.; Muraki, E. Full-information item factor analysis. *Psychological Measurement*, v. 12, n. 3, p. 261 – 281, 1988.
- Bernini, C.; Matteucci, M.; Mignani, S. Investigating heterogeneity in residents' attitudes toward tourism with an mirt multidimensional approach. *QUALITY & QUANTITY*, v. 49, n. 2, p. 805–826, 2015.
- Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord & M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Camilli, G. & Shepard, L. A. (1994). *MMSS. Methods for identifying biased test items*. California: Sage.
- Chalmers, P. (2013) - Package mirt: Multidimensional Item Response Theory. Version 9.0 <http://cran.r-project.org/web/packages/mirt/mirt.pdf>
- Costa, P.; Soares, T.; Ferrão, M. Estimação Bayesiana e MCMC em Modelo Multidimensional: CONGRESSO XLII SBPO. Bento Gonçalves. Rio Grande do Sul. 2010. p. 2760.
- Dorans, N. J. & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and Standardization. Em P. W. Holland & H. Wainer (Orgs.). *Differential item*

- 
- functioning (pp. 35-66). New Jersey: Lawrence Erlbaum.
- Fragoso, T. M. (2010). Modelos Multidimensionais da Teoria de Resposta ao Item. Tiago de Miranda Fragoso. São Carlos, SP: USP, ICMC, 20110.
- Ginnons, R. D.; HEDEKER, D. R. Full-information item bi-factor analysis. *Psychometrika*, v. 57, n. 3, p. 423 – 436, 1992.
- Gonçalves, F.B. Análise Bayesiana da Teoria de Resposta ao Item: uma abordagem generalizada. Flávio Bambirra Gonçalves. Rio de Janeiro: UFRJ, IM, DME, 2006.
- Horst, P., Factor analysis of data matrices, Holt, Rinehart & Winston, New York, 1965.
- INEP. Microdados para download. 2015 [http://inep.gov.br/microdados]. Acessado em 03/06/2017.
- Lawley, D.N. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh, Series A*, 61, 273-287, 1943.
- Lazarsfeld, P.F. The logical and mathematical foundation of latent structure analysis. In S.A. Stauffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen (Eds.), *Measurement and prediction*. Princeton, NJ: Princeton University Press, 1950.
- Linn, R. L. & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement*, 18, 109-118.
- Lord, F. M. (1952). A theory of test scores (No. 7). *Psychometric Monograph*.
- McDonald, R.P. (1967), *Nonlinear factor analysis*, *Psychometric Monograph*, n. 15, 1-167.
- Muthén, B. (1978), Contributions to factor analysis of dichotomous variables, *Psychometrika*, 43, 551-560.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen : Danish Institute for Educational Research.
- Reise, S. P.; Morizot, J.; HAYS, R. D. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, v. 16, n. 1, p. 19 – 31, 2007.
- Reckase, M. D. *Multidimensional Item Response Theory*. USA: Springer, 2009.
- Vieira, N.N. (2016). *As Provas das Quatro Áreas do Enem Vistas como Prova Única na Ótica de Modelos da Teoria da Resposta ao Item Uni e Multidimensional*. Florianópolis, SC, 2016.

Samejima, F. (1974), Normal ogive model on the continuous response level in the multidimensional space, *Psychometrika*, 39, 111-121.

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense of Canadá.

Zwick, R., Thayer, D. T. & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, 36, 1-28.